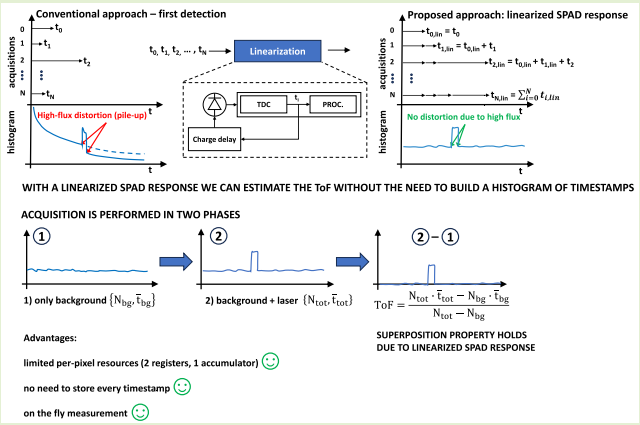
Histogram-Less LiDAR Through SPAD Response Linearization

Alessandro Tontini, *Member, IEEE*, Sonia Mazzucchi, Roberto Passerone, *Member, IEEE*, Nicolò Broseghini, and Leonardo Gasparini, *Member, IEEE*

***Abstract*—We present a new method to acquire the 3-D information from a Single-photon avalanche diode (SPAD)-based direct-Time-of-Flight (d-ToF) imaging system which does not require the construction of a histogram of timestamps and can withstand high-flux operation regime. The proposed acquisition scheme emulates the behavior of an SPAD detector with no distortion due to dead time, and extracts the ToF information by a simple average operation on the photon timestamps ensuring ease of integration in a dedicated sensor and scalability to large arrays. The method is validated through a comprehensive mathematical analysis, whose predictions are in agreement with a numerical Monte Carlo model of the problem. Finally, we show the validity of the predictions in a real d-ToF measurement setup under challenging background conditions well beyond the typical pile-up limit of 5% detection rate up to a distance of 3.8 m.**



***Index Terms*— Light detection and ranging (LiDAR), single-photon avalanche diode (SPAD).**

I. Introduction

SPATIAL perception enabled by 3-D imaging techniques is constantly gaining interest for industrial [1], automotive [2], space [3], and consumer applications. As an example, the required level of self-awareness of autonomous driving vehicles demands for 3-D imaging systems with high resolution and high frame rate. Unfortunately, these requirements are in conflict, constraining engineers to performance-limiting tradeoffs. In this article, we focus on light detection and ranging (LiDAR)/direct-Time-of-Flight (d-ToF) measurement based on single-photon avalanche diodes (SPADs), which is one of the most promising among active techniques [4], [5]. In an SPAD-based d-ToF measurement, the distance is extracted by measuring the traveling time of a pulse of light projected from the source and reflected back by the target to a detector that consists of an SPAD operating as a photon-to-edge converter, coupled to a photon timestamping circuit [usually a time-to-digital converter (TDC) or a time-to-amplitude converter (TAC)] [6], [7], [8]. Due to hardware limitations, such as the detector dead time and the statistical nature of photons, together with the presence of uncorrelated background light, a number of observations are usually accumulated into a histogram memory to enhance the signal-to-noise ratio (SNR) and extract the target distance by means of signal processing techniques [9], [10], [11].

The increased interest in advanced driver assistance systems (ADAS), where 3-D vision is a pillar, is focusing the attention of researchers in developing techniques to increase the robustness of such systems against the effect of background light and against possible interference from similar devices. Concerning the problem of background light, one of the most effective and widespread techniques is known as *photon coincidence*, which exploits the temporal proximity of photons belonging to the reflected laser pulse to filter out unwanted background photons which are more likely to be temporally sparse [12], [13]. With a different method, based on a smart accumulation technique by Yoshioka et al. [14], the SNR is increased by merging the information from pixels observing similar regions of the scene. A different approach has been recently proposed by Manuzzato et al. [15], where a per-pixel circuit is able to automatically decrease the SPAD sensitivity reducing the probability of saturation in case of high

Manuscript received 28 November 2023; revised 5 December 2023; accepted 8 December 2023. Date of publication 27 December 2023; date of current version 13 February 2024. The associate editor coordinating the review of this article and approving it for publication was Prof. Yongqiang Zhao. *(Corresponding author: Alessandro Tontini.)*

Alessandro Tontini and Leonardo Gasparini are with Fondazione Bruno Kessler, 38123 Trento, Italy (e-mail: tontini@fbk.eu).

Sonia Mazzucchi is with the Department of Mathematics, University of Trento, 38123 Trento, Italy.

Roberto Passerone is with the Department of Information Engineering and Computer Science, University of Trento, 38123 Trento, Italy.

Nicolò Broseghini is with the Department of Physics, University of Trento, 38123 Trento, Italy.

Digital Object Identifier 10.1109/JSEN.2023.3342609

background intensities, favoring the detection of laser photons. Yet another technique is known as time-gating, where by means of a search procedure several subranges of the scene are measured, increasing the SNR at the expense of an increased acquisition time [16]. Regarding the problem of mutual interference, Ximenes et al. [17] propose a spread-spectrum-based technique, where the laser emission time is randomized from device to device, spreading any other interference below the level of the signal of interest. Another solution is based on the emission of two laser pulses, whose temporal relationship is different from device to device and used to actively discard unwanted interference [18], [19], [20].

These techniques, however, cannot cope with very high fluxes because of two fundamental problems. First, the histogram of timestamps appears distorted due to the dead-time of SPAD detectors and timestamping circuits, which translates into a nonlinear response of the system to the incident flux of photons over time. It is a widely held belief that the upper photon flux limit that still results in a negligibly distorted histogram is given by 5% of detected photons per laser cycle [21]. Second, the amount of data generated by the sensor is too large to scale to a large number of pixels.

To overcome the former problem, several solutions have been proposed in the literature. In the work from Rapp et al. [22], the histogram is linearized through a Markov chain model of the photon detection times. By using this method, the system can cope with up to five photoelectrons per illumination cycle. Another approach, which is implemented at the system level, is to rely on the so-called multievent or continuous sampling TDCs [23], [24]. With this solution, the time-to-digital conversion is not limited to a single event per laser cycle, as in standard approaches, but allows for a continuous sampling of the incoming flux of photons, limited by either the gate delay of the TDC architecture [23] or by the fixed amount of memory that stores the continuously sampled timestamps [24]. These solutions proved to be effective in mitigating the pile-up distortion, as they allow for a multitude of timestamps to be recorded from each laser pulse. Therefore, to a certain extent, the linearization of the SPAD response is possible and well approximated. However, due to the necessity to employ an asynchronous SPAD driving approach, true linearization cannot be achieved. Continuous sampling approaches also match well with innovative techniques aimed at extending the range of SPAD-based LiDARs. For instance, Patanwala et al. [25] propose the synchronous summation technique (SST) which attains the highest throughput compared to existing pulse combining techniques [12], [13], [23], [24], maximizing the effectiveness of such TDC approaches. Gupta et al. [26] address the problem of histogram distortion due to high background light with a gated acquisition approach, where the gating signal is progressively scanned over the whole acquisition window to produce an approximate version of a true linearized histogram. While this approach can average out the effect of pile-up, the histogram linearization is only approximated and is more effective with a high number of temporal gate shifts, resulting in an increased acquisition time and complexity. In another work, the same authors address the problem of optimal attenuation for an SPAD-based LiDAR to be operated with [27]. With this approach, the typical fixed limit of 5% detection rate is overcome by computing an optimal flux attenuation factor which requires an estimate of the background flux of photons. Despite being effective in improving the per-pixel signal-to-background ratio (SBR), this article does not state clearly how the level of attenuation can be adjusted in real time on a pixel basis.

Concerning the amount of generated data, Gyongy et al. [28] propose a mitigation technique by upscaling a low-resolution depth image based on a high-resolution intensity image. A similar approach is used by Ruget et al. [29], where the native resolution of a depth image from an SPAD camera is increased by means of a deep neural network. Still the bottleneck is represented by the necessity to build and transfer a histogram of timestamps for each pixel in the image from which the ToF is extracted.

One strategy to reduce the bandwidth requirement on the amount of data which is transferred from the chip to the controller (usually an FPGA or μC) is to integrate the histogram, or part of it, directly on chip. Several solutions are proposed in the literature to integrate histogramming capability on-chip [24], [30], [31], [32], [33], [34], [35], [36], [37], [38], but despite the advantage in bandwidth performance compared to other solutions where the histogram is built off-chip [12], [13], [15], [16], [17], still many limitations are present. In general, the on-chip realization of either a *partial* or *full* histogram requires additional area, which can be obtained by either reducing the fill factor or by using expensive 3-D-stacked solutions.

With the so-called *partial* approach [18], [32], [33], [34], [36], [37], [38], a reduced histogram memory is available on-chip, therefore requiring a search procedure to identify the location of the ensemble of histogram bins containing the laser peak. In the literature, two techniques have been described to implement a *partial* histogram behavior. With the so-called *zooming* technique [18], [32], [33], [37], at the beginning of the measurement the reduced set of histogram bins is spread across the entire distance range. By counting the number of photons detected on each bin, the reduced set of histogram bins is concentrated over several iterations on a shorter range, thus achieving the desired resolution on the estimated target distance. With the other technique, called *sliding*, the subset of histogram bins is already set to the desired resolution, thus covering only a small portion of the range. Again by means of several iterations, the subset of histogram bins slides across the entire range, and the number of photons at each iteration is used to estimate the target distance. Despite the intrinsic differences between the two methods, for both of them, as outlined by Taneski et al. [39], a laser power penalty occurs as more laser shots are required to find the laser peak location.

A *full* histogram approach is possible with resource sharing, e.g., by reallocating the same histogram circuitry to several pixels, as described by Kumagai et al. [35]. However, resource sharing resulted in only $\approx$27% of the chip area dedicated to the SPAD array, requiring a very high clock frequency of 500 MHz and the design of a 3-D stacked solution with a complex scanning illumination approach.

In the literature, other techniques to reduce the amount of data generated and handled by SPAD-based d-ToF systems have been proposed. Ingle and Maier [40] propose a novel approach based on race logic which can provide an equi-depth histogram where the photon counts from several SPAD detections are divided into quantiles (in particular, the median value is used). By cascading several binner circuits, it is possible to reconstruct the equi-depth histogram of photon arrival times from which the ToF is extracted. This approach, however, has been only tested with simulated data and no physical realization of the binner circuit implemented with race logic has been provided. In the work from Gutierrez-Barragan et al. [41], an on-the-fly compressive histogram approach is presented. In this work, the amount of memory reduction is the same as for partial histogram approaches with the advantage of ensuring a better accuracy in the localization of the histogram peak. However, it requires a high in-pixel compute effort, since several multiply-add operations must be executed in real time for each new timestamp. Sheehan et al. [42] propose a sketching approach based on the Fourier transform, which does not scale with either the number of photons or the timestamp resolution and does not effectively build a histogram of timestamps. However, this approach, which also was not validated with hardware, requires the real-time computation of the Fourier features which is hardly integrable at the pixel level.

In this article, we propose for the first time a robust method supported by a rigorous mathematical model to extract the 3-D information from a set of acquired timestamps *without the need to build a histogram*, which can also sustain high photon fluxes, enabling the possibility to operate beyond the standard limit of 5% detection rate [21] for pile-up distortion. The method can be implemented using only two registers and one accumulator for each pixel. With such a low amount of resources, the per-pixel memory requirement is reduced by more than three orders of magnitude compared to standard d-ToF architectures (off-chip histogram) [16], [17], and by a factor of $\simeq$5 compared to architectures with on-chip, full histogramming capability [43]. We reach the goal in two steps. First, we propose and evaluate an algorithm to efficiently extract the target distance from a set of timestamps based on a simple on-the-fly average operation, which does not require the allocation of a histogram memory. Then, since the proposed algorithm works on the assumption that the detector response is linear, we present two acquisition schemes that can be easily implemented on chip and emulate the behavior of a single-photon detector with no dead time, providing the desired linear response to the input flux of photons. The proposed method is supported by analytical and numerical (Monte Carlo) models and has been validated experimentally up to a distance of 3.8 m (mainly limited by the sensor used for characterization [44]) under a background light equivalent to 85 klx and beyond the standard 5% rule for pile-up distortion [21].

This article is organized as follows. In Section II, a typical d-ToF acquisition system with the most important parameters of concern is described and preliminary considerations on a histogram-less acquisition approach are provided with a first validation by means of a Monte Carlo model. In Section III, we provide the analytical proof of the proposed acquisition method, while in Section IV we explain in detail two acquisition schemes which are needed to emulate the response of a linear detector, together with a comparison against state-of-the art sensors in terms of memory requirement, scalability, and tolerance to high background flux. In Section V, we provide measurement results from an existing SPAD-based d-ToF sensor, showing that the proposed acquisition and extraction schemes are capable of successfully computing the ToF without the need for a histogram of timestamps. Finally, we discuss future perspectives to advance the results found in this work in Section VI.

II. Preliminary Validation

In this section, we present the principle of operation of an SPAD-based d-ToF system with preliminary considerations and Monte Carlo simulations on the histogram-less approach which will be further developed in this article.

A. Typical d-ToF Operation

A typical d-ToF image acquisition requires a pulsed laser and a time-resolved, single-photon image sensor with photon timestamping capabilities. It works by sending periodic laser pulses and then measuring the arrival time, or *timestamp*, of the first detected photons reflected by the target following each pulse. Due to space and bandwidth limitations, the number of photon timestamps generated per laser pulse is typically limited to one.

In principle, a single laser pulse, and thus a single timestamp, would be sufficient to estimate the ToF. However, due to the presence of uncorrelated background events [from both external light sources or internal SPAD dark count rate (DCR)] and of shot noise, the first detected photon may not be from the laser pulse, so that several repetitions are needed to discriminate the different contributions. To do so, the timestamps measured during the acquisition process are collected in a histogram memory that records how many times each timestamp has been observed. This provides a convenient representation of the temporal distribution of the arrival times, as shown in the example of Fig. 1. In a system capable of acquiring only one photon (the first), the distribution of the arrival times is a piecewise exponential curve, where each segment is described by

$$P_i(t) = A_i \cdot e^{-\lambda_i \cdot t} \tag{1}$$

whose *intensity* (rate) λ_i depends on the intensity of background light, dark counts, and laser echo. Table I summarizes the most important parameters of the detection process.

The ToF is typically estimated from the histogram by finding the location of its peak or of a sharp rising edge, which likely belongs to the reflected laser pulse. The histogram of timestamps contains all the relevant information to properly estimate the ToF, and represents the gold standard processing technique in the field of SPAD-based d-ToF systems. Unfortunately, the histogram requires a considerable amount of resources in terms of memory, bandwidth, and power, as it

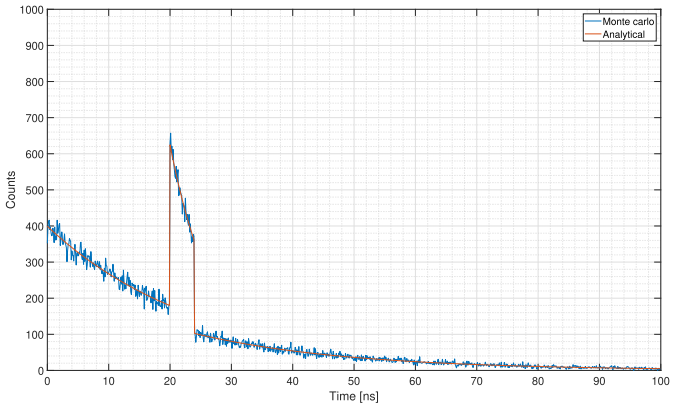


Fig. 1. Simulated distribution of timestamps in a typical d-ToF system able to record one photon per acquisition, with a ToF of 20 ns and a laser pulse duration T_W of 4 ns. The histogram is composed of 10^5 timestamps, with a bin size of 100 ps. Superimposed to the Monte Carlo simulation, we show also the analytical exponential distribution.

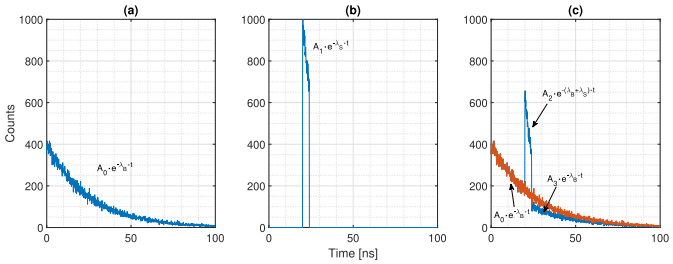


Fig. 2. Simulated distribution of timestamps in a typical d-ToF acquisition. (a) and (b) Distributions of background-only events (with rate λ_B) and laser-only events (with rate λ_S), respectively, are shown. (c) Distribution of the combination of background and laser events is reported, graphically showing that the superposition property does not hold due to the nonlinear behavior of the detection process. In particular, the portion of background events after the laser peak is underestimated, as only one photon per acquisition can be detected. For each contribution, the amplitude terms of the exponential (A_0, A_1, A_2 and A_3) are reported, with $A_3 < A_0$ due to the SPAD nonlinearity.

TABLE I
LIST OF PARAMETERS FOR A TYPICAL d-ToF ACQUISITION

Parameter	Unit	Description
λ_B	s^{-1}	Background events rate
λ_S	s^{-1}	Reflected laser events rate
T_W	s	Laser pulse duration
ToF	s	Time of flight

requires the readout of every timestamp from the sensor for processing by an external controller (FPGA or μC). Even with the latest implementations where the histogram is available on-chip, the required amount of resources is considerable. As an example, a 10-m-range 128 $\times$ 128 LiDAR system with a 100-ps time resolution and 8-bit histogram depth requires approximately 10 MB of memory.

B. Histogram-Less Approach

Intuitively, if no background events are present, and neglecting the width of the laser pulse, we could estimate the ToF without the need to build a histogram. This can be achieved by simply calculating the average of the continuous stream of laser-only timestamps. To extend the above method to scenarios where background events are also present, we need to eliminate their contribution to the average. Again, intuitively, this can be accomplished by dividing the measurement into two acquisitions. The first is performed with the laser turned off, and is used to estimate the contribution of the background light only, by computing the average $\bar{t}_{\text{bg}}$ of the recorded timestamps. The same operation is repeated in the second acquisition with the joint contribution of background and laser timestamps, resulting in a total average time $\bar{t}_{\text{tot}}$. In principle, the ToF could be estimated as a linear combination of the averages[1] as

$$\text{ToF} \propto k_{\text{tot}} \cdot \bar{t}_{\text{tot}} - k_{\text{bg}} \cdot \bar{t}_{\text{bg}} \tag{2}$$

with the contribution of the background canceling out. The coefficients k_{tot} and k_{bg} depend on the number of detected photons, hence they are affected by shot noise, with a direct impact on the measurement precision.

This approach, however, relies on the superposition property which does not hold, as SPADs are nonlinear detectors. More specifically, the problem lies in the dead time of the detection process, which depends upon the SPAD dead time itself, on the limited bandwidth of the timestamping circuit and also on the limited memory available, blinding the measurement channel for some time after each detection. With this limitation, the detection of a photon belonging to the laser echo does prevent a later photon from being potentially detected, resulting in a distortion of the statistics. In particular, the amount of background photons which contribute to $\bar{t}_{\text{tot}}$ in the second acquisition (with the laser turned on) is underestimated. This behavior can be observed on the histograms of Fig. 2, where the distribution of timestamps in different scenarios is compared, emphasizing the estimation error. Furthermore, the greater the laser echo intensity, the higher the number of background photons that are underestimated.

This can also be seen analytically by computing explicitly the cumulative distribution function F of the random variable T associated with the first photon detection time, defined as $F(t) := \mathbb{P}(T \in [0, t])$. If $t \in [0, \text{ToF}]$ then the incoming photon necessarily belongs to the background, hence

$$F(t) = \mathbb{P}(T \in [0, t]) = 1 - e^{-\lambda_B t}.$$

If $t \in (\text{ToF}, \text{ToF} + T_W]$, then

$$\begin{aligned} F(t) &= \mathbb{P}(T \in [0, \text{ToF}]) + \mathbb{P}(T \in (\text{ToF}, t]) \\ &= 1 - e^{-\lambda_B \text{ToF}} + e^{-\lambda_B \text{ToF}}\left(1 - e^{-(\lambda_S + \lambda_B)(t - \text{ToF})}\right). \end{aligned}$$

Finally, for $t > \text{ToF} + T_W$, we have

$$\begin{aligned} F(t) &= \mathbb{P}(T \in [0, \text{ToF}]) \\ &\quad + \mathbb{P}(T \in (\text{ToF}, \text{ToF} + T_W]) \\ &\quad + \mathbb{P}(T \in (\text{ToF} + T_W, t]) \\ &= 1 - e^{-\lambda_B \text{ToF}} \\ &\quad + e^{-\lambda_B \text{ToF}}\left(1 - e^{-(\lambda_S + \lambda_B) T_W}\right) \\ &\quad + e^{-\lambda_B \text{ToF}} e^{-(\lambda_S + \lambda_B) T_W}\left(1 - e^{-\lambda_B (t - \text{ToF} - T_W)}\right). \end{aligned}$$

[1]For the precise formulation, see (6).

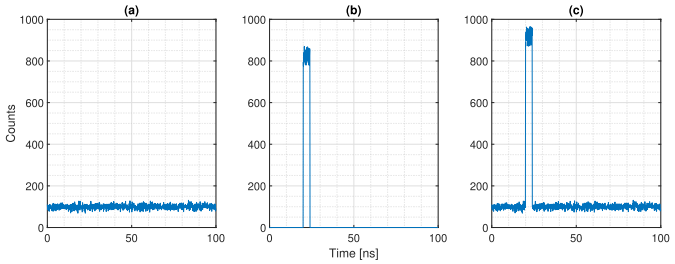


Fig. 3. Distribution of timestamps obtained with a linear detection process, i.e., with no dead-time limitation. The distributions are uniform, since we are now considering the absolute arrival time of detected photons with respect to the beginning of the acquisition window. This approach enables the secure subtraction of the background contribution in (a) from the combined measurement in (c). (b) Result is the isolated contribution from the laser light alone, which carries the ToF information. As the superposition property holds in this case, there is no longer an under-weighting condition of background counts after the laser pulse peak in the histogram (c).

The probability density f of the distribution of T, where $F(t) = \int_{-\infty}^{t} f(u)du$, can be easily computed by means of the formula $f(t) = F'(t)$, yielding

$$f(t) = \begin{cases} \lambda_B e^{-\lambda_B t}, & t \in [0, \text{ToF}] \\ (\lambda_S + \lambda_B) e^{\lambda_S \text{ToF}} e^{-(\lambda_S+\lambda_B)t}, & t \in (\text{ToF}, \text{ToF} + T_W] \\ \lambda_B e^{-\lambda_S T_W} e^{-\lambda_B t}, & t > \text{ToF} + T_W \end{cases}$$

while the average first arrival time is given by

$$\begin{aligned} \mathbb{E}[T] &= \int_0^{+\infty} t f(t) dt \\ &= \frac{1 - e^{-\lambda_B \text{ToF}}}{\lambda_B} + \frac{e^{-\lambda_B \text{ToF}}(1 - e^{-(\lambda_S+\lambda_B)T_w})}{\lambda_S + \lambda_B} \\ &\quad + \frac{e^{-\lambda_S T_W} e^{-\lambda_B(\text{ToF}+T_W)}}{\lambda_B}. \end{aligned} \tag{3}$$

As shown in (3), the average detection time depends nonlinearly on two parameters (λ_S and ToF). This confirms that it is not possible to uniquely extract the ToF from the aforementioned acquisition method, since the average of the timestamps acquired in the second acquisition, $\bar{t}_{\text{tot}}$, which is governed by the probability of detection of laser photons, also depends on the λ_S parameter, which depends not only on the ToF, but also on the target reflectivity. One could try to compensate for the error; however, this requires measuring also the intensity of the received laser light (which affects the error), introducing an extra variable which is hard to estimate, invalidating the procedure.

Conversely, a linear detector, with no dead time, is able to timestamp every photon which falls within the acquisition window. In this case, the histograms shown in Fig. 2 become linear in time, as shown in Fig. 3.

One can then extract the ToF with the proposed two-step procedure. In the first step, we measure the total number of *background* events, N_{bg}, and their average absolute arrival time, $\bar{t}_{\text{bg}}$. In the second step, with the combination of both background and laser events, we measure the *total* number of events and their average absolute arrival time, denoted N_{tot} and $\bar{t}_{\text{tot}}$. Because the superposition property holds, the difference $N_{\text{tot}} - N_{\text{bg}}$ is equal to the amount of photons from the reflected laser source. We can then extract the ToF by properly weighting each average timestamp measurement with the relative photon count contribution

$$N_{\text{tot}} \cdot \bar{t}_{\text{tot}} = N_{\text{bg}} \cdot \bar{t}_{\text{bg}} + (\text{ToF} + \bar{t}_l) \cdot (N_{\text{tot}} - N_{\text{bg}}) \tag{4}$$

where $\bar{t}_l$ is the average arrival time of the laser photons referred to the laser emission time in the absence of background light (i.e., $\lambda_B = 0$ and $\text{ToF} = 0$). The value $\bar{t}_l$ is a characteristic parameter of the laser source, which can be experimentally estimated by means of an initial calibration. The proposed extraction method does not require the allocation of histogram memory, and needs only two counters to store N_{bg} and N_{tot}, and two accumulators to compute $\bar{t}_{\text{bg}}$ and $\bar{t}_{\text{tot}}$ reducing the memory requirements by more than three orders of magnitude compared to recent long-range high-resolution d-ToF sensors [16], [17]. We can further reduce the amount of resources down to a single accumulator, needed to store $\bar{t}_{\text{tot}}$, and two counters for N_{bg} and N_{tot}, because a constant background throughout the acquisition window leads to an average background time $\bar{t}_{\text{bg}}$ of $T_{\text{acq}}/2$. In this case, (4) turns into the simpler form

$$N_{\text{tot}} \cdot \bar{t}_{\text{tot}} = N_{\text{bg}} \cdot \frac{T_{\text{acq}}}{2} + (\text{ToF} + \bar{t}_l) \cdot (N_{\text{tot}} - N_{\text{bg}}) \tag{5}$$

which yields

$$\text{ToF} = \frac{N_{\text{tot}} \bar{t}_{\text{tot}} - N_{\text{bg}} \frac{T_{\text{acq}}}{2}}{N_{\text{tot}} - N_{\text{bg}}} - \bar{t}_l. \tag{6}$$

We have simulated this extraction method with a Monte Carlo simulator [45] by sweeping the parameters λ_S and λ_B in the range $[10^6 - 10^8]$ and $[10^5 - 10^9]$, respectively. For each pair of λ_S and λ_B values, 10^4 measurements have been acquired with the ToF value set to 25 ns. The resulting ToF, obtained from (4), is shown in Fig. 4 with the correct estimation over a wide range of λ_B, λ_S pairs, failing only when the λ_S/λ_B ratio is too low even for a classic histogram-based approach.

In Section III, we provide a rigorous mathematical analysis which proves the results briefly introduced with (6) from the underlying statistical distribution of photons.

III. Mathematical Analysis

This section proves the validity of the method described above analytically. In the following, we shall denote the duration of the acquisition window by T_{acq}, assume the laser echo to entirely occur within it, i.e., $T_{\text{acq}} \gg \text{ToF} + T_W$, and denote the time-dependent intensity of the laser pulse by the function $\lambda_S : [0, T_W] \to \mathbb{R}$. The flux of photons can be modeled by a counting process $(N_t)_{t\in[0,T_{\text{acq}}]}$ obtained as the sum of two independent Poisson processes: $(N_t^B)_{t\in[0,T_{\text{acq}}]}$, with intensity λ_B, describing the background flux of photons, and $(N_t^S)_{t\in[\text{ToF},\text{ToF}+T_W]}$, describing the signal. In particular, the process $(N_t^S)_{t\in[\text{ToF},\text{ToF}+T_W]}$ is modeled through an inhomogeneous Poisson process with time-dependent intensity $(\tilde{\lambda}_S(t))_{t\in[\text{ToF},\text{ToF}+T_W]}$ given by $\tilde{\lambda}_S(t) := \lambda_S(t - \text{ToF})$. Hence, the overall flux of photons reaching the SPAD is given by

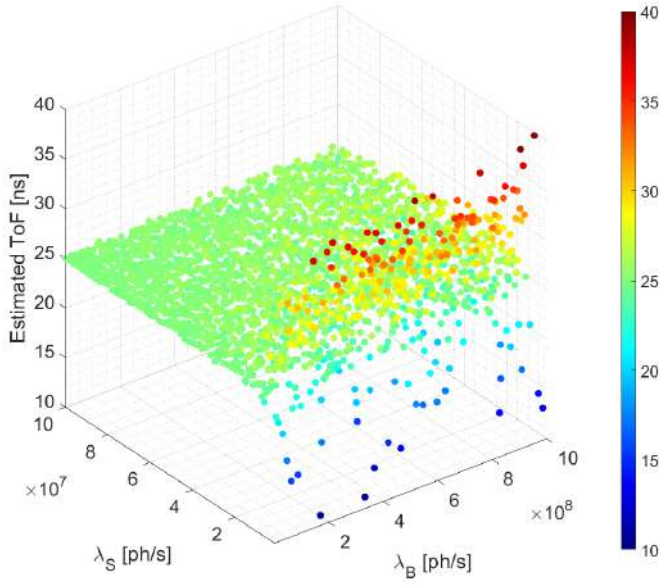


Fig. 4. Preliminary Monte Carlo simulation results showing the ToF computed with the proposed acquisition method with the hypothesis of an ideal linear detector over a full scale range (acquisition window) of 100 ns. The ToF can be properly estimated over a wide range of λ_B, λ_S pairs. With the minimum λ_S value considered for the simulation (10^6 events/s), the ToF estimation becomes very noisy, but still centered around the correct value of 25 ns. If the signal intensity were zero, the ToF estimation would be centered around zero, as the contributions from the two measurements (with and without background) cancel out reciprocally.

an inhomogeneous Poisson process $(N_t))_{t\in[0,T_{\text{acq}}]}$ with varying intensity λ given by

$$\lambda(t) = \begin{cases} \lambda_B, & t \in [0, \text{ToF}) \cup [\text{ToF} + T_W, T_{\text{acq}}] \\ \lambda_B + \tilde{\lambda_S}(t), & t \in [\text{ToF}, \text{ToF} + T_W). \end{cases}$$

Hence, we can prove that, given n photons detected in the interval $I = [0, T_{\text{acq}}]$, the n detection times are independent and distributed on $[0, T_{\text{acq}}]$ with a distribution density function $f : [0, T_{\text{acq}}] \to \mathbb{R}$ given by

$$f(t) = \begin{cases} \dfrac{\lambda_B}{\lambda_B T_{\text{acq}} + \int_0^{T_W} \lambda_S(u)du}, \\ \qquad\qquad t \in [0, \text{ToF}) \cup [\text{ToF} + T_W, T_{\text{acq}}] \\ \dfrac{\lambda_B + \lambda_S(t - \text{ToF})}{\lambda_B T_{\text{acq}} + \int_0^{T_W} \lambda_S(u)du}, \\ \qquad\qquad t \in [\text{ToF}, \text{ToF} + T_W). \end{cases} \tag{7}$$

Indeed, by considering a partition $\{I_j\}_{j=1,\dots,m}$ of the interval $I = [0, T_{\text{acq}}]$, the independence of the increments of the inhomogeneous Poisson processes $(N_t)_{t\in[0,T_{\text{acq}}]}$ yields

$$\begin{aligned} &\mathbb{P}(N(I_1) = n_1, \dots, N(I_m) = n_m \mid N(I) = n) \\ &\quad = \frac{\prod_{j=1}^m e^{-\lambda_{I_j}} \frac{(\lambda_{I_j})^{n_j}}{n_j!}}{e^{-\lambda_{[0,T_{\text{acq}}]}} \frac{\left(\lambda_{[0,T_{\text{acq}}]}\right)^n}{n!}} \end{aligned} \tag{8}$$

where, for an interval $I = [a, b] \subset \mathbb{R}$, we adopted the notation $\lambda_I := \int_a^b \lambda(t)dt$ and we assumed $\sum_{j=1}^m n_j = n$. By a straightforward computation, the right-hand side of (8) can be written as

$$\begin{aligned} &\mathbb{P}(N(I_1) = n_1, \dots, N(I_m) = n_m \mid N(I) = n) \\ &\quad = \frac{n!}{n_1!, \dots, n_m!} \prod_{j=1}^m \left(\frac{\lambda_{I_j}}{\lambda_{[0,T_{\text{acq}}]}} \right)^{n_j} \end{aligned}$$

the latter being equivalently obtained in terms of n independent and identically distributed continuous random variables $T_1, \dots, T_n$ with density f given by (7). In other words, the arrival times $\{T_i\}_{i=1,\dots,n}$ of the n photons provide a statistical sample for the distribution (7). The distribution (7) has a mean μ given by

$$\begin{aligned} \mu &= \int_0^{T_{\text{acq}}} t f(t)dt \\ &= \frac{1}{\lambda_B T_{\text{acq}} + \int_0^{T_W} \lambda_S(t)dt} \\ &\quad \cdot \left(\frac{\lambda_B T_{\text{acq}}^2}{2} + \text{ToF} \int_0^{T_W} \lambda_S(t)dt + \int_0^{T_W} t\lambda_S(t)dt \right). \end{aligned} \tag{9}$$

Clearly, since μ is a linear function of ToF, it can easily be inverted. By denoting with α the ratio

$$\alpha := \frac{\lambda_B T_{\text{acq}}}{\lambda_B T_{\text{acq}} + \int_0^{T_W} \lambda_S(t)dt} \tag{10}$$

we get an equation providing the ToF as a function of the other characteristic parameters of the process

$$\text{ToF} = \frac{1}{1-\alpha} \left(\mu - \alpha \frac{T_{\text{acq}}}{2} \right) - \bar{t}_l \tag{11}$$

where $\bar{t}_l = (\int_0^{T_W} t\lambda_S(t)dt)/(\int_0^{T_W} \lambda_S(t)dt)$ is the average arrival time of the laser photons referred to the laser emission time in the absence of background light (i.e., $\lambda_B = 0$ and $\text{ToF} = 0$).

If n photons are detected within the interval $[0, T_{\text{acq}}]$, the sample mean $\bar{t}_{\text{tot}}$ of their detection times provides an unbiased estimator for μ. The main issue with this approach is the estimation of the parameter α, which depends on λ_B and λ_S, the latter being affected by a high level of uncertainty since it is related to the intensity of the laser echo. Because the total number of photons detected in the time interval $[0, T_{\text{acq}}]$ is a Poisson random variable with average $\lambda_B T_{\text{acq}} + \int_0^{T_W} \lambda_S(t)dt$ and, analogously, the number of background photons detected in the time interval $[0, T_{\text{acq}}]$ is a Poisson random variable with average $\lambda_B T_{\text{acq}}$, we can estimate both parameters by observing a realization of both processes. More precisely, let us first switch off the laser source and collect the number N_{bg} of photons arriving during the interval $[0, T_{\text{acq}}]$, then let us switch on the laser source and collect the total number N_{tot} of photons arriving during the interval $[0, T_{\text{acq}}]$. The observed values N_{bg} and N_{tot} are, respectively, an estimate for $\lambda_B T_{\text{acq}}$ and $\lambda_B T_{\text{acq}} + \int_0^{T_W} \lambda_S(t)dt$, while their ratio is an estimate $\hat{\alpha} := (N_{\text{bg}}/N_{\text{tot}})$ for the parameter α defined in (10). By replacing the parameters μ and α with their estimates

$\bar{t}_{\text{tot}}$ and $\hat{\alpha}$ in formula (11), we obtain the following estimator for ToF:

$$\widehat{\text{ToF}} = \frac{1}{1-\hat{\alpha}}\left(\bar{t}_{\text{tot}} - \hat{\alpha}\frac{T_{\text{acq}}}{2}\right) - \bar{t}_l \tag{12}$$

$$= \frac{N_{\text{tot}}\bar{t}_{\text{tot}} - N_{\text{bg}}\frac{T_{\text{acq}}}{2}}{N_{\text{tot}} - N_{\text{bg}}} - \bar{t}_l \tag{13}$$

which coincides with (6).

IV. Acquisition Schemes

The simulation results obtained in Section II-B are based on the assumption that the photon detection process is ideal, i.e., with no dead time and with a linear response over the incoming flux of photons. In a real-world scenario, however, detectors are limited by the dead time between subsequent detections, resulting in a nonlinear response. To implement the proposed extraction method, we propose a novel SPAD acquisition scheme which emulates the behavior of a linear detector. More in detail, we propose two ways to obtain a linearized SPAD response from a real SPAD. Both methods are based on the assumption that the underlying statistical processes are stationary and ergodic. In particular, we assume that there are no major fluctuations of the characteristic parameters of the process during the acquisition time. Similar to an equivalent-time sampling oscilloscope, both methods rely on repeating the observation multiple times to emulate the response of an SPAD detector with no dead time. In Sections IV-A and IV-B, we describe the working principle of each method and propose a possible implementation. Then, in Section IV-C, we provide the mathematical proof that both acquisition methods are capable of correctly sampling the distribution of photon arrival times.

A. Acquisition Scheme #1: Acquire or Discard

The first acquisition scheme relies on a simple (albeit inefficient) mechanism, which requires no additional resources in terms of the SPAD driving circuit. The acquisition works over multiple runs, each requiring multiple observations. The first timestamp of every run is considered valid, memorized, and used to increment either N_{bg} or N_{tot}, depending on the current phase of the acquisition, and update $\bar{t}_{\text{tot}}$. Then, in the next observations, timestamps are considered valid, and used to update the algorithm parameters, only if they are higher than the largest previous timestamp, otherwise they are discarded. The procedure is iterated until the last photon within the acquisition window T_{acq} is timestamped. The necessary condition to conclude the current linearization cycle is the absence of detected photons before T_{acq}. This implies that the next photon will occur after the acquisition window, thereby validating that the last recorded photon marks the conclusion of the current linearization cycle. The process is then repeated multiple times to increase statistics. Fig. 5 shows an example of a run, including all the discarded events, and a possible implementation. While the implementation is straightforward, the method is inefficient because the majority of the detected photons may end up being discarded.

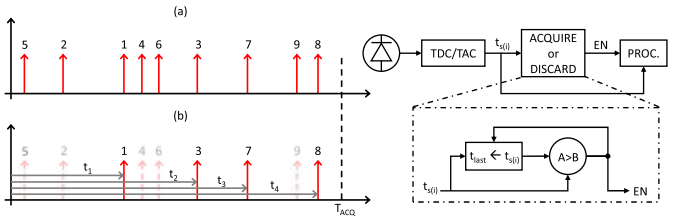


Fig. 5. Example of SPAD response linearization with the *acquire or discard* acquisition method. (a) Each photon arrival time is represented by a red arrow and the order of arrival is indicated. The first run starts with the acquisition of photon #1, resulting in timestamp t_1. Photon #2 is discarded, since its arrival time is earlier than photon #1. The next recorded information comes instead from photon #3, which is later than photon #1, and sets the new minimum time. The run proceeds with the same criteria resulting in the stream of photon arrival times t_1, t_2, t_3 and t_4 from photons #1, #3, #7 and #8, which is a single realization of the emulated response of the linearized SPAD detector (b). On the right, a principle schematic is proposed, showing the lightweight usage of resources, with only one comparator and one register required on top of the processing circuit. The *acquire or discard* acquisition method is simple but inefficient, as most of the photons arrival times are discarded, resulting in longer acquisition times.

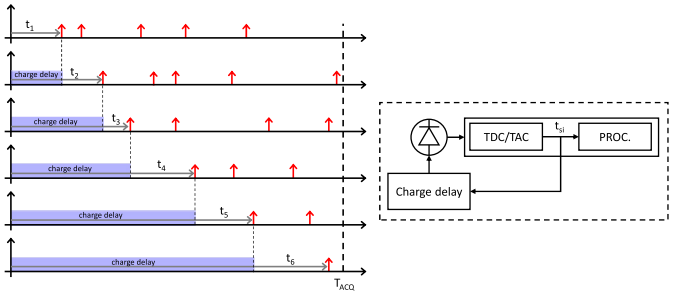


Fig. 6. Example of SPAD response linearization with the *time-gated* acquisition method. With this approach, no photon timestamp is discarded thanks to the delayed activation of the SPAD for each timing measurement. During the *charge delay* phase, the SPAD front end is forced OFF, thus photons cannot be detected. For each measurement, the first timestamp is detected and used to increment either N_{bg} or N_{tot} and update the average time $\bar{t}_{\text{tot}}$. At the same time, the *charge delay* phase value is updated accordingly for the next measurement. As opposed to the *acquire or discard* method, more hardware resources are needed to build the delay element which controls the activation of the SPAD.

B. Acquisition Scheme #2: Time-Gated

The time-gated acquisition scheme works by delaying the activation of the SPAD to start from the previously acquired timestamp, until no photon is detected before the end of the acquisition window.

With this approach, there is no need to discard timestamps, allowing for a faster acquisition. This, however, comes at the expense of a more complex hardware implementation, which needs a time-activated gating scheme, for instance using a programmable delay line. An example of acquisitions is shown in Fig. 6, together with a possible implementation.

C. Discussion on Implementation, Expected Performance, and Mathematical Analysis

While providing the same result, it is clear that the implementation cost and the performance of the two acquisition schemes are different. With the *acquire or discard* scheme, almost no hardware modification is required to an already existing SPAD sensor. However, because of the decimation

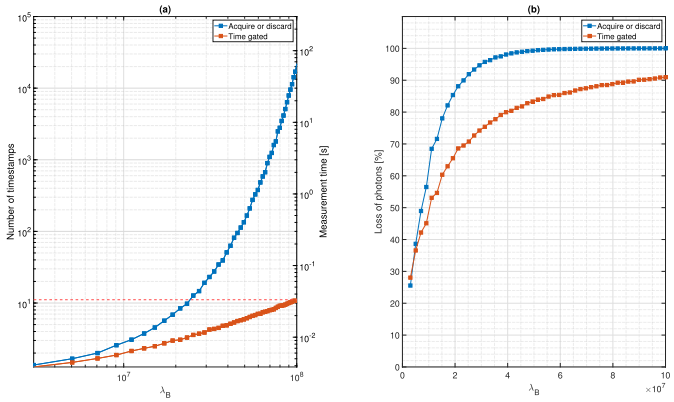


Fig. 7. Result of Monte Carlo simulation comparing the two proposed acquisition schemes in terms of efficiency for increasing values of background light flux (λ_B) in the range $[10^6, 10^8]$ events/s. (a) We compare the acquisition schemes in absolute terms. On the left axis, we show the number of timestamps required to linearize the SPAD response over an acquisition window T_{acq} of 100 ns. On the right axis, we show the total time required for the two methods to collect $N = 3 \cdot 10^4$ measurements to average the linearized response of the SPAD N times. The horizontal line indicates a limit of ≈33.3 ms, for an equivalent operation frame rate of 30 FPS. With the *time-gated* scheme, the required frame rate can be guaranteed over the entire range of background light flux, while considering the *acquire or discard* scheme, the maximum sustainable flux is limited to $\approx 2.4 \cdot 10^7$ events/s. (b) We compare the acquisition schemes in relative terms against an ideal system composed by an ideal SPAD with an ideal TDC, showing the percentage of photons lost during the linearization process over the acquisition window T_{acq}. The loss of photons occurs as both acquisition schemes works with a real SPAD, which in this case can timestamp at most one photon per acquisition.

process, the efficiency of the acquisition could be very low. This also depends on the intensity of the incoming rate of events: the higher the rate, the higher the probability to have smaller timestamps which block the detection process. On the other hand, the *time-gated* scheme requires a delay line and the SPAD-gating, but the efficiency is much higher since no decimation process occurs. To show the difference in terms of efficiency of the two proposed acquisition schemes, we run a Monte Carlo simulation with background light flux in the range $[10^6, 10^8]$ events/s with the aim to linearize the SPAD response over an acquisition window T_{acq} of 100 ns. As shown in Fig. 7(a), at the highest rate of 10^8 events/s, the amount of timestamps to be acquired to cover the acquisition window T_{acq} for the *acquire or discard* scheme is more than three orders of magnitude higher compared to the more efficient *time gated* scheme. From the simulation, we can also identify the maximum number of measurements which can be executed by the two acquisition schemes to sustain an operation frame rate of 30 frames/s (FPS). The *time-gated* scheme can average the linearized SPAD response up to $N = 3 \cdot 10^4$ times over the whole range of background light event rate. On the other hand, with the same number N of acquisitions, the *acquire or discard* scheme can only support up to $\approx 2.4 \cdot 10^7$ events/s of background rate. Fig. 7(b) shows a comparison in relative terms against an ideal SPAD coupled to an ideal TDC, thus being capable of timestamping all photons impinging on the sensor surface over the acquisition window at once. Both the *acquire or discard* and *time-gated* schemes are meant to work with a real SPAD, which in this case can timestamp at most one photon per acquisition window. For such a reason, if the acquisition schemes are compared against an ideal system, a loss of photons occurs, which can be limited by the *time-gated* scheme at ≈90% in the worst case of photon flux.

Concerning the implementation complexity, the *time-gated* scheme requires a finely controlled delay element, which may be expensive on a per-pixel basis. Proper balance between the linearization efficiency and implementation complexity can be obtained with a mixture between the two acquisition schemes. This can be accomplished by employing a coarse but simple delay architecture for the gating mechanism (e.g., a clock-based counter), which activates the SPAD one clock cycle before the last recorded timestamp. Then, the linearization process is concluded by relying on the *acquire or discard* scheme to cover the remaining amount of time. The efficiency is still guaranteed, as the *acquire or discard* scheme starts at most one clock cycle before the last recorded timestamp, and not from the beginning as in the standalone implementation.

From a mathematical point of view, both acquisition methods allow sampling the correct distribution of the photon arrival times $(T_n)_{n\geq 1}$. Let T_n denote the occurrence time of the nth event, i.e., the arrival of the nth photon. This can be defined as the infimum of the set of times t such that the number of arrivals N_t in the interval $[0, t]$ is greater than or equal to n

$$T_n := \inf\{t \geq 0 \,:\, N_t \geq n\}, \quad n \geq 1.$$

By definition, $N_{T_n} = n$. We can therefore derive an equivalent representation of the random variables T_n. Indeed, for $n \geq 2$, the time T_n of occurrence of the nth event can be obtained as the infimum of the set of times t greater than T_{n-1} [the time of occurrence of the $(n-1)$th event] such that the increment $N_t - N_{T_{n-1}}$, the number of events occurring in the interval $(T_{n-1}, t]$, is greater than 1. Therefore,

$$T_1 := \inf\{t \geq 0 \,:\, N_t \geq 1\}$$
$$T_n := \inf\{t \geq T_{n-1} \,:\, N_t - N_{T_{n-1}} \geq 1\}, \quad n \geq 2$$

which corresponds to the results of the acquisition schemes described previously.

D. Comparison With State-of-the-Art

In this section, we compare our histogram-less acquisition method with state-of-the-art SPAD-based LiDAR sensors in terms of memory requirement, scalability, and tolerance to high background light flux. For all comparisons in this section, we consider for our method 16 bits of counters depth (i.e., up to 65 535 counts for N_{bg} and N_{tot}, respectively) and then three times the number of TDC bits (1 xTDC bits required for the TDC word itself, and then 2 xTDC bits to properly size the accumulator memory). First, we compare against standard sensors, i.e., sensors that require the raw timestamps to be read out to build the necessary histogram of timestamps off-chip. To provide a fair comparison, we do not consider the sensor resolution, which changes from chip to chip, but only the amount of memory required to build the histogram for one pixel. In the works we consider for our comparison [2], [12], [15], [16], [17], we extrapolate

TABLE II
MINIMUM, AVERAGE, AND MAXIMUM MEMORY REDUCTION FACTOR OF THE PROPOSED HISTOGRAM-LESS ACQUISITION METHOD AGAINST STANDARD d-ToF SENSORS (OFF-CHIP HISTOGRAM) [2], [12], [15], [16], [17] AND SENSORS WITH ON-CHIP FULL HISTOGRAM CAPABILITY [23], [43], [46]

Standard sensors			Full on-chip histogram		
Min.	Avg.	Max.	Min.	Avg.	Max.
69.4 [2]	2129.5	6553.6 [15]	4.7 [43]	135.9	331.3 [46]

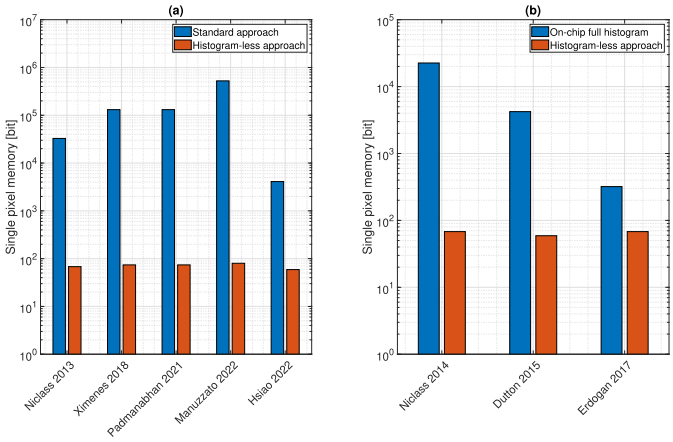


Fig. 8. Comparison of the amount of per-pixel memory required by our histogram-less acquisition method against histogram-based d-ToF sensors. (a) We consider standard sensors where every timestamp is read out and the histogram is built off-chip [2], [12], [15], [16], [17]. (b) We consider sensors with full on-chip histogram capability [23], [43], [46].

the total amount of per-pixel memory based on the number of reported TDC bits and on an 8-bit histogram depth for all of them. We then compare our solution to sensors that offer full on-chip histogram capability [23], [43], [46]. Also in this case, we consider the amount of memory reported in each work necessary to build the histogram of timestamps for one pixel. Results are reported in Fig. 8, where an average memory reduction factor of ≈2129 and ≈136 for standard and full on-chip histogram sensors, respectively, is obtained. More comparison details, including minimum and maximum memory reduction factors, are reported in Table II.

Similar to our method, partial histogram approaches [24], [32], [33], [38] are also quite effective in reducing the memory requirements. Nevertheless, our approach not only outperforms them with a memory reduction that ranges from 67% [24] to 3% [38], but also performs better in many other important aspects. In fact, unlike previous work, our approach does not have any of the following needs.

1) The need to find the laser peak in time using a zooming or a sliding search procedure, which is at the basis of every partial histogram approach [39].
2) The need to share hardware resources (TDCs, memory) among pixels in the same column [32] or in the same cluster [38], to reduce the area usage.
3) The need for area consuming processors to manage the algorithm underlying the partial histogram technique and that can only be implemented using advanced 3-D integrated technologies with single-pixel access [24].

All these translate into higher measurement time and laser power penalty as more acquisitions are needed than a standard full-histogram approach [39], and higher costs.

Our method, given the very limited amount of required memory resources, is also advantageous in terms of scalability to higher sensor resolutions, and also in terms of range extension. As an example, a standard histogram-based sensor with 15-bit TDC requires memory to store up to 32 767 histogram bins per pixel. If the measurement range is doubled, the additional TDC bit results in an increase of 100% on the memory requirement. Conversely, with our approach, the amount of memory increase to double the range is limited to only ≈3.9%.

Concerning the tolerance to high background light flux, both detection processes can sustain very high flux regimes, with a limit determined by the finite resolution T_{TS} of the timestamping circuit. This limit translates to the requirement $(1/(\lambda_B + \lambda_S)) \gg T_{TS}$, i.e., having a low probability that more than one photon fall into the same time bin. By considering a threshold on this probability, we can extract the maximum flux of photons λ_{max} which can be sustained by our detection process. The probability to have more than one photon per time bin is expressed as $P(n > 1) = 1 - e^{-\lambda_{max} \cdot T_{TS}} \cdot (1 + \lambda_{max} \cdot T_{TS})$. By setting a threshold of less than 1%, and considering $T_{TS} =$ 100 ps, the maximum photon flux that can be sustained is equal to $\lambda_{max} \simeq 1.48 \cdot 10^9$ ph/s. Compared to the maximum flux required by a standard system which must comply with the 5% rule, and with the hypothesis of an acquisition window T_{acq} of 100 ns, our detection process can sustain a photon flux ≃3000 times higher.

As the ToF is extracted without direct access to the complete statistics of photon timestamps (i.e., the histogram), the contributions to the ToF due to multiple targets or multipath reflections cannot be distinguished as they linearly contribute to the average timestamp value extracted after the linearization process. A foreseen countermeasure is the implementation of an additional system-level gating scheme (thus, on top of the one required for the SPAD linearization), where by means of an approximate knowledge of the target(s) location, the timing localization of the LiDAR acquisition is properly steered to avoid the mixture of multiple contributions. On the other hand, the two acquisition schemes proposed in this work do not prevent the construction of a histogram of (linearized) timestamps, supporting the complete distinction of multiple targets and multipath reflections.

V. MEASUREMENT RESULTS

The proposed acquisition scheme has been validated with measurements using real data from an existing single-point

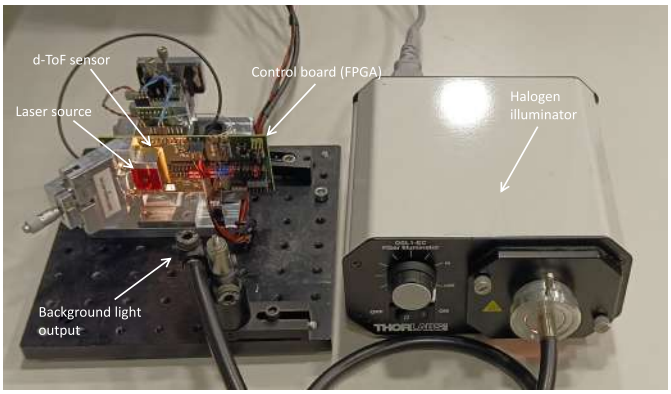


Fig. 9. Measurement setup with the FPGA control board, d-ToF system, and halogen illuminator for the generation of background light pointed directly toward the sensor.

SPAD-based d-ToF sensor, with an architecture similar to the one from Perenzoni et al. [44], which in addition offers on-chip histogramming capability. The sensor is fabricated in a standard 150-nm CMOS process with the SPAD technology developed in the work by Xu et al. [47]. The histogram features 1024 bins with 10-b depth, and a TDC resolution of 100 ps. The SPADs are enabled synchronously with the beginning of the acquisition window and the first measured timestamp for each acquisition increments the corresponding histogram bin. After a user-selectable number of acquisitions, the histogram is read out and unpacked. Then, the unpacked data is shuffled to recover a realization vector of the arrival times of the detected photons.

Background events were generated by means of a ≈180-W fiber-coupled halogen illuminator pointed directly toward the sensor, while a black matte panel with low ≈10% reflectivity was selected as target, with a distance range from 1 up to 3.8 m. A picture of the setup is shown in Fig. 9, with indications on the main components.

First, we focus on the validation of the linearization behavior of the proposed acquisition scheme by considering only background light. Then, we consider the combination of background and laser together, as in a real scenario, and we compute the ToF with the proposed histogram-less acquisition scheme.

A. Preliminary Considerations

As we base our measurements on the re-engineering of an existing d-ToF sensor, preliminary considerations are needed before providing further details on measurement results. The sensor measures the arrival time of the first detected photon for each laser pulse, as described in Section II, which is stored in an on-chip histogram memory. Since the sensor measures the arrival time of the first photon, the statistical distribution is exponential, thus we are considering *relative* arrival times. A statistically valid realization of the incoming timestamps is obtained by unpacking and randomly shuffling the content of the histogram memory. The obtained realization is a vector of *relative* arrival times, which is the starting point of our measurement analysis.

In Section IV, we described two possible acquisition schemes. The *acquire or discard* scheme, even though is intrinsically inefficient, can be straightforwardly used with our dataset as it requires no hardware modification over the already existing SPAD-based d-ToF system. The *time-gated* scheme, while more efficient, requires a time-gating circuit which is not implemented in our sensor.

The first set of measurements focuses on background events only. In this case, there is a single source of events with intensity λ_B, so we can apply the *time-gated* scheme by computing the cumulative sums of timestamps to obtain *absolute* arrival times from *relative* ones. On the other hand, when events from background and laser are combined, as in a real measurement scenario, it is not possible to mimic the behavior of the *time-gated* scheme by means of the cumulative sum operation. In that case, we rely on the *acquire or discard* scheme.

B. Measurements With Background Light Only

We set the intensity of background light from a minimum of $\approx 6.5 \cdot 10^6$ up to $\approx 133 \cdot 10^6$ events/s. This is the rate of events at the output of the SPAD, which therefore takes into account all physical parameters of concern of a typical d-ToF system [45]. Considering an acquisition window of 100 ns, specific from the sensor [44], the equivalent average number of detections within T_{acq} equals ≈0.65 and ≈13.3 for the minimum and maximum background light intensity, respectively. In both cases, this is much higher than the conventional limit of 5% events [21] (13 and 266 times higher, respectively), showing the high resistance of our method against pile-up distortion. By considering the equation which links the intensity of background events, λ_B, with the physical parameters of the system [45], it is possible to derive the equivalent background illumination level, in kilolux, up to a maximum of ≈85 klx. Measurement results are shown in Fig. 10, showing a relative deviation from the reference background intensity extracted from the exponential fit of the original histogram of less than ±0.5% over the whole range of values.

C. Measurements With Background and Laser Light and Extraction of the ToF

Our first goal is to show that the underestimation of background counts which occurs in a standard d-ToF system can be completely recovered with our acquisition scheme. This is demonstrated in the first measurement, displayed in Fig. 11, which compares a traditional acquisition with the *acquire or discard* scheme, qualitatively showing the linearization process by means of the linearized histogram of timestamps.

We then quantitatively evaluated the linearization process by estimating the intensity of background light from both portions of the histogram, i.e., before and after the laser peak. For this characterization, we used the ≈10% reflectivity target (black matte panel) at 2.5-m distance from the sensor. The results are depicted in Fig. 12, showing a relative deviation from the ground truth (estimated from an exponential fit on the original dataset) below ±4%.

In a different measurement, we verify the resistance of the proposed SPAD linearization method against pile-up distortion. To do so, we acquire several timestamps from the

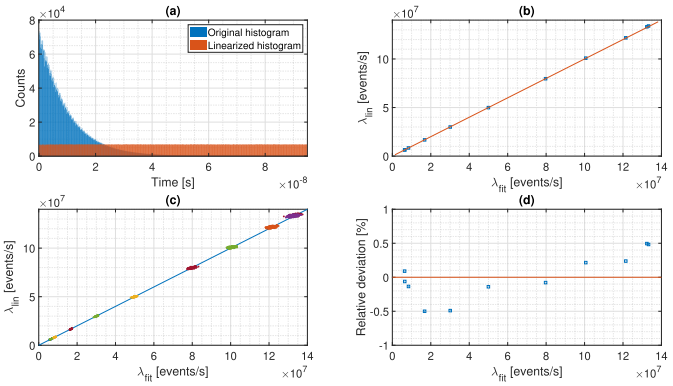


Fig. 10. Linearization of the SPAD response with background events only. For each value of background flux, $8 \cdot 10^6$ timestamps are acquired from the sensor. (a) Example of linearized histogram is shown together with the original one (exponentially distributed) for a background flux of $\approx 100 \cdot 10^6$ events/s. (b) Flux of background events estimated from the linearized histogram of timestamps, λ_{lin} against the flux estimated from an exponential fit on the original histogram of timestamps, λ_{fit}. (c) For each value of background flux, the entire dataset was split into 200 subsets to analyze the homogeneity of the linearization process. (d) Relative deviation from the background flux measured from the original histograms is shown, used as a reference, demonstrating a relative deviation below $\pm 0.5\%$ over all data subsets.

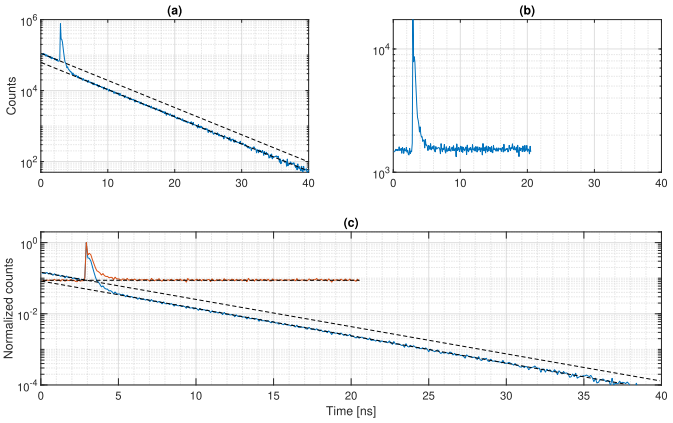


Fig. 11. Qualitative measurement showing the linearization process of the proposed acquisition scheme. (a) Original histogram of timestamps is shown in logarithmic scale, where the drop of counts which occurs after the laser peak is clearly visible. (b) Histogram obtained with the *acquire or discard* scheme proves the efficacy of the linearization process, which fully compensates for the nonlinearity of the detector. The length of the linearized histogram is shorter than the original dataset, as we decided to stop the linearization earlier to reduce the data loss which naturally occurs with the *acquire or discard* scheme. Due to the intrinsic inefficiency of this scheme, the histogram peak in (b) is attenuated by ≈ 34 dB with respect to the original dataset in (a). (c) Two histograms are shown together after normalization.

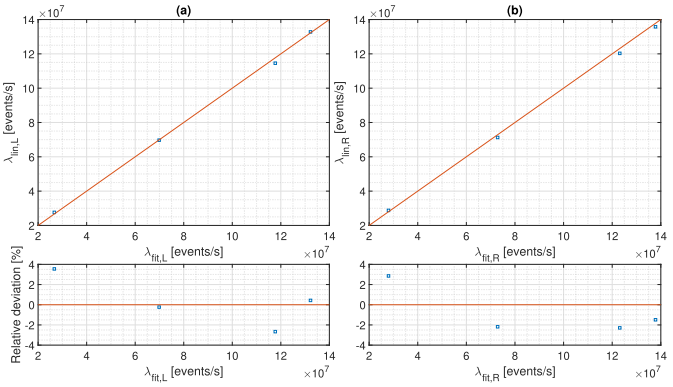


Fig. 12. Quantitative characterization of the linearization process considering four different values of background flux, from $\approx 27.6 \cdot 10^6$ events/s up to $\approx 133 \cdot 10^6$ events/s with a target distance of 2.5 m. For each value of background flux, $2.5 \cdot 10^6$ timestamps are acquired from the sensor. (a) Relationship between the background fluxes computed before the laser peak is shown, where $\lambda_{fit,L}$ comes from an exponential fit on the original histogram, while $\lambda_{lin,L}$ comes from the linearized histogram. (b) Same relationship is shown but considering the portion of background events after the histogram peak. For each portion, the relative deviation of the flux extracted from the linearized histogram of timestamps is shown, demonstrating an estimation error below $\pm 4\%$ over the whole range. The application of the *acquire or discard* acquisition scheme results in a data reduction factor of ≈ 7.5 and ≈ 165 for the minimum and maximum background light flux, respectively.

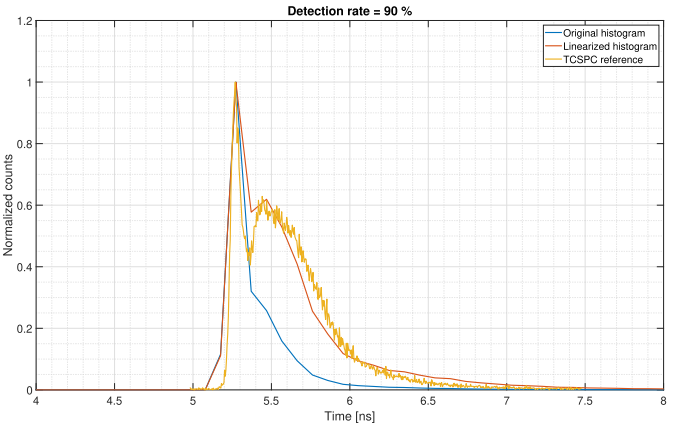


Fig. 13. Characterization of the behavior of the proposed SPAD linearization method under strong pile-up conditions. The histogram obtained from the linearized vector of timestamps is compared against the original histogram (built from the detection of the first arrival time) and against a reference measurement obtained with a conventional TCSPC setup. In the histograms obtained from our sensor timestamps, the bin width is 100 ps, while the reference measurement from the TCSPC setup has 4-ps timing resolution. The proposed SPAD linearization method allows us to recover the full shape of the laser envelope even if the detection rate is 18 times higher than the conventional limit of 5%.

reflected laser pulse with a detection rate of 90%, which is 18 times higher than the conventional limit of 5%. The results, shown in Fig. 13, proves the efficacy of our linearization method in challenging pile-up conditions where a standard sensor would fail. A reference measurement acquired with a conventional time-correlated single photon counting (TCSPC) setup is shown as reference.

The last set of measurements shows the extracted ToF without the need to build a histogram of timestamps. For each measured distance, we run the linearization algorithm 250 times to have sufficient statistics to compute accuracy and precision. For each run of the algorithm, we average the results from $N = 1.5 \cdot 10^4$ vectors of linearized SPAD timestamps, to emulate an equivalent 30-FPS operation rate, as outlined in Section IV-C and with Fig. 7. For all measurements, the same $\simeq 10\%$ reflectivity target (black matte panel) was used, in the range from 1 to 3.8 m, to emulate a challenging scenario for a typical SPAD-based d-ToF system. First, we evaluate the behavior of the ToF extraction process without background light. The results, depicted in Fig. 14, show good agreement between the extracted ToF and the ground truth. Then, we repeat the measurements with the inclusion of

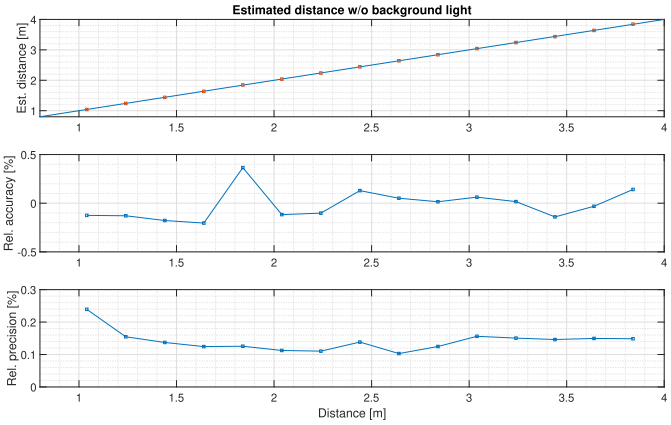


Fig. 14. Measurement results with no background light, showing the ToF extracted without the need to build a histogram of timestamps. The relative accuracy is below ±0.5%, while the relative precision is below 0.25% for all measurements.

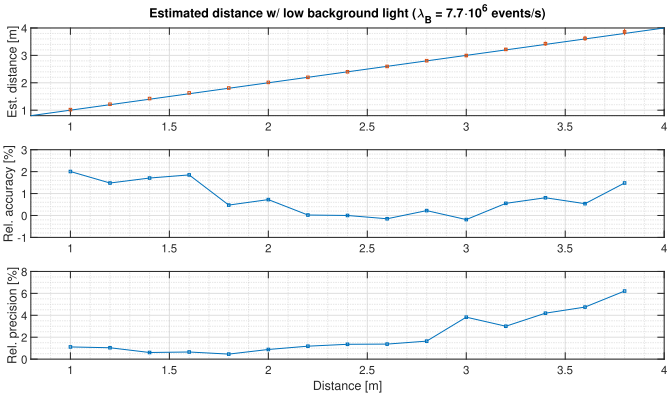


Fig. 15. Measurement results with low background light flux ($\lambda_B = 7.7 \cdot 10^6$ events/s), showing the extracted ToF without the need to build a histogram of timestamps. The relative accuracy is in the range $[-0.2, 2]$%, while the worst relative precision is 6% at the highest distance of 3.8 m.

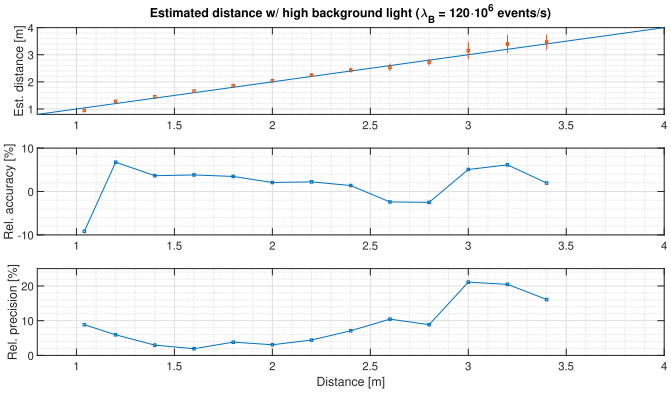


Fig. 16. Measurement results with high background light flux ($\lambda_B = 120 \cdot 10^6$ events/s), showing the extracted ToF without the need to build a histogram of timestamps. The relative accuracy is in the range $[-9, 6.7]$%, while the worst relative precision is 21% at 3 m. With this high background light flux (corresponding to ≈75 klx), and the decision to use a low reflectivity target (≈10%), the maximum achieved range decreased to 3.4 m.

background light by setting the halogen illuminator to generate a background light flux of $7.7 \cdot 10^6$ and $120 \cdot 10^6$ events/s. The values of background light flux are considered at the output of the SPADs of the sensor, and they correspond to an illumination level of ≃15 and ≃75 klx, respectively. Results are shown in Figs. 15 and 16, demonstrating the validity of the proposed histogram-less ToF estimation in a real setup.

VI. Conclusion

In this work, we demonstrate how to extract the ToF information in an SPAD-based d-ToF system without the need to build a resource and bandwidth-hungry histogram of timestamps. Moreover, the proposed method is resistant against high photon fluxes and can withstand detection rates three orders of magnitude higher than the conventionally recognized limit of 5%. The acquisition method, which is based on the linearization of the SPAD response, is suitable for integration in CMOS technology using low resources and is therefore scalable to large arrays, since it can be easily integrated per-pixel. The proposed extraction method has been completely characterized, first with Monte Carlo numerical simulations. The method is also mathematically justified, and we demonstrated its validity with real measurements, by repurposing an existing d-ToF sensor and using real data to extract the ToF. The proposed extraction method can be implemented at least in two ways, by means of the *acquire or discard* or *time-gated* detection schemes. While the *acquire or discard* scheme allows for the least usage of resources, it suffers from long integration times especially when the flux of photons is too high. On the other hand, the *time-gated* scheme can guarantee a more efficient acquisition at the expense of a per-pixel controllable delay element. A hybrid scheme, which implements a coarse but simple delay element to implement the *time-gated* acquisition and relies on the *acquire or discard* scheme for the remaining amount of time, can guarantee at the same time a compact hardware implementation and optimized acquisition time. Concerning the ToF extraction method, we demonstrated its validity by using an extremely low amount of resources, as only two counters and one accumulator are required.

References

[1] V. Sesta, K. Pasquinelli, R. Federico, F. Zappa, and F. Villa, "Range-finding SPAD array with smart laser-spot tracking and TDC sharing for background suppression," *IEEE Open J. Solid-State Circuits Soc.*, vol. 2, pp. 26–37, 2022.

[2] A.-T. Hsiao et al., "Real-time LiDAR module with 64×128-pixel CMOS SPAD array and 940-nm PCSEL," in *Proc. IEEE Sensors Appl. Symp. (SAS)*, Aug. 2022, pp. 1–6.

[3] 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, Jun. 2021.

[4] K. Morimoto et al., "Megapixel time-gated SPAD image sensor for 2D and 3D imaging applications," *Optica*, vol. 7, no. 4, pp. 346–354, Apr. 2020.

[5] F. Piron, D. Morrison, M. R. Yuce, and J.-M. Redouté, "A review of single-photon avalanche diode time-of-flight imaging sensor arrays," *IEEE Sensors J.*, vol. 21, no. 11, pp. 12654–12666, Jun. 2021.

[6] Z.-P. Li et al., "Single-photon imaging over 200 km," *Optica*, vol. 8, no. 3, pp. 344–349, Mar. 2021.

[7] A. Tontini, L. Gasparini, L. Pancheri, and R. Passerone, "Design and characterization of a low-cost FPGA-based TDC," *IEEE Trans. Nucl. Sci.*, vol. 65, no. 2, pp. 680–690, Feb. 2018.

[8] J. Richardson et al., "A 32×32 50ps resolution 10 bit time to digital converter array in 130 nm CMOS for time correlated imaging," in *Proc. IEEE Custom Integr. Circuits Conf.*, Sep. 2009, pp. 77–80.

[9] G. Chen, C. Wiede, and R. Kokozinski, "Data processing approaches on SPAD-based d-TOF LiDAR systems: A review," *IEEE Sensors J.*, vol. 21, no. 5, pp. 5656–5667, Mar. 2021.

[10] V. Sesta, A. Incoronato, F. Madonini, and F. Villa, "Time-to-digital converters and histogram builders in SPAD arrays for pulsed-LiDAR," *Measurement*, vol. 212, May 2023, Art. no. 112705.

[11] I. Gyongy, N. A. W. Dutton, and R. K. Henderson, "Direct time-of-flight single-photon imaging," *IEEE Trans. Electron Devices*, vol. 69, no. 6, pp. 2794–2805, Jun. 2022.

[12] C. Niclass, M. Soga, H. Matsubara, S. Kato, and M. Kagami, "A 100-m range 10-frame/s 340 ×96-pixel time-of-flight depth sensor in 0.18-μm CMOS," *IEEE J. Solid-State Circuits*, vol. 48, no. 2, pp. 559–572, Dec. 2013.

[13] M. Perenzoni, D. Perenzoni, and D. Stoppa, "A 64 × 64-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 J. Solid-State Circuits*, vol. 52, no. 1, pp. 151–160, Jan. 2017.

[14] K. Yoshioka et al., "A 20-ch TDC/ADC hybrid architecture LiDAR SoC for 240 × 96 pixel 200-m range imaging with smart accumulation technique and residue quantizing SAR ADC," *IEEE J. Solid-State Circuits*, vol. 53, no. 11, pp. 3026–3038, Nov. 2018.

[15] E. Manuzzato, A. Tontini, A. Seljak, and M. Perenzoni, "A 64×64-pixel flash LiDAR SPAD imager with distributed pixel-to-pixel correlation for background rejection, tunable automatic pixel sensitivity and first-last event detection strategies for space applications," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, vol. 65, Feb. 2022, pp. 96–98.

[16] P. Padmanabhan et al., "A 256×128 3D-stacked (45 nm) SPAD FLASH LiDAR with 7-level coincidence detection and progressive gating for 100 m range and 10klux background light," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, vol. 64, Feb. 2021, pp. 111–113.

[17] A. R. Ximenes, P. Padmanabhan, M.-J. Lee, Y. Yamashita, D. N. Yaung, and E. Charbon, "A 256×256 45/65 nm 3D-stacked SPAD-based direct TOF image sensor for LiDAR applications with optical polar modulation for up to 18.6 dB interference suppression," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2018, pp. 96–98.

[18] H. Seo et al., "Direct TOF scanning LiDAR sensor with two-step multievent histogramming TDC and embedded interference filter," *IEEE J. Solid-State Circuits*, vol. 56, no. 4, pp. 1022–1035, Apr. 2021.

[19] H. Seo, G. Cho, S.-J. Kim, J.-H. Chun, and J. Choi, "Multievent histogramming TDC with pre–post weighted histogramming filter for CMOS LiDAR sensors," *IEEE Sensors J.*, vol. 22, no. 23, pp. 22785–22798, Dec. 2022.

[20] A. Tontini, L. Gasparini, and R. Passerone, "A SPAD-based linear sensor with in-pixel temporal pattern detection for interference and background rejection with smart readout scheme," in *Proc. Int. Image Sensor Workshop*, Crieff, U.K., May 2023.

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

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

[23] N. A. W. Dutton et al., "A time-correlated single-photon-counting sensor with 14 GS/S histogramming time-to-digital converter," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2015, pp. 1–3.

[24] S. W. Hutchings et al., "A reconfigurable 3-D-stacked SPAD imager with in-pixel histogramming for flash LiDAR or high-speed time-of-flight imaging," *IEEE J. Solid-State Circuits*, vol. 54, no. 11, pp. 2947–2956, Nov. 2019.

[25] S. M. Patanwala et al., "A high-throughput photon processing technique for range extension of SPAD-based LiDAR receivers," *IEEE Open J. Solid-State Circuits Soc.*, vol. 2, pp. 12–25, 2022.

[26] A. Gupta, A. Ingle, and M. Gupta, "Asynchronous single-photon 3D imaging," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 7908–7917.

[27] A. Gupta, A. Ingle, A. Velten, and M. Gupta, "Photon-flooded single-photon 3D cameras," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 6763–6772.

[28] I. Gyongy et al., "High-speed 3D sensing via hybrid-mode imaging and guided upsampling," *Optica*, vol. 7, no. 10, pp. 1253–1260, Oct. 2020.

[29] A. Ruget, S. McLaughlin, R. K. Henderson, I. Gyongy, A. Halimi, and J. Leach, "Robust super-resolution depth imaging via a multi-feature fusion deep network," *Opt. Exp.*, vol. 29, no. 8, pp. 11917–11937, Apr. 2021.

[30] S. Lindner, C. Zhang, I. M. Antolovic, M. Wolf, and E. Charbon, "A 252 × 144 SPAD pixel flash LiDAR with 1728 dual-clock 48.8 PS TDCs, integrated histogramming and 14.9-to-1 compression in 180 nm CMOS technology," in *Proc. IEEE Symp. VLSI Circuits*, Jun. 2018, pp. 69–70.

[31] F. M. D. Rocca et al., "A 128 × 128 SPAD motion-triggered time-of-flight image sensor with in-pixel histogram and column-parallel vision processor," *IEEE J. Solid-State Circuits*, vol. 55, no. 7, pp. 1762–1775, Jul. 2020.

[32] C. Zhang, S. Lindner, I. M. Antolovic, 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 J. Solid-State Circuits*, vol. 54, no. 4, pp. 1137–1151, Apr. 2019.

[33] B. Kim, S. Park, J.-H. Chun, J. Choi, and S.-J. Kim, "A 48×40 13.5 mm depth resolution flash LiDAR sensor with in-pixel zoom histogramming time-to-digital converter," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2021, pp. 108–110.

[34] D. Stoppa et al., "A reconfigurable QVGA/Q3VGA direct time-of-flight 3D imaging system with onchip depth-map computation in 45/40 nm 3D-stacked BSI SPAD CMOS," in *Proc. Int. Image Sensor Workshop*, 2021.

[35] O. Kumagai et al., "A 189×600 back-illuminated stacked SPAD direct time-of-flight depth sensor for automotive LiDAR systems," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, vol. 64, Feb. 2021, pp. 110–112.

[36] I. Gyongy, N. A. W. Dutton, H. Mai, F. M. D. Rocca, and R. K. Henderson, "A 200kFPS, 256 × 128 SPAD dToF sensor with peak tracking and smart readout," in *Proc. Ing. Image Sensor Workshop*, 2021.

[37] S. Park, B. Kim, J. Cho, J.-H. Chun, J. Choi, and S.-J. Kim, "An 80×60 flash LiDAR sensor with in-pixel histogramming TDC based on quaternary search and time-gated δ-intensity phase detection for 45m detectable range and background light cancellation," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, vol. 65, Feb. 2022, pp. 98–100.

[38] C. Zhang et al., "A 240 × 160 3D-stacked SPAD dToF image sensor with rolling shutter and in-pixel histogram for mobile devices," *IEEE Open J. Solid-State Circuits Soc.*, vol. 2, pp. 3–11, 2022.

[39] F. Taneski, T. A. Abbas, and R. K. Henderson, "Laser power efficiency of partial histogram direct time-of-flight LiDAR sensors," *J. Lightw. Technol.*, vol. 40, no. 17, pp. 5884–5893, Sep. 2022.

[40] A. Ingle and D. Maier, "Count-free single-photon 3D imaging with race logic," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, pp. 1–12, 2023, doi: 10.1109/TPAMI.2023.3302822.

[41] F. Gutierrez-Barragan, A. Ingle, T. Seets, M. Gupta, and A. Velten, "Compressive single-photon 3D cameras," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 17833–17843.

[42] M. P. Sheehan, J. Tachella, and M. E. Davies, "A sketching framework for reduced data transfer in photon counting LiDAR," *IEEE Trans. Comput. Imag.*, vol. 7, pp. 989–1004, 2021.

[43] A. T. Erdogan, R. Walker, N. Finlayson, N. Krstajic, G. O. S. Williams, and R. K. Henderson, "A 16.5 Giga events/s 1024 × 8 SPAD line sensor with per-pixel zoomable 50ps-6.4ns/bin histogramming TDC," in *Proc. Symp. VLSI Circuits*, Jun. 2017, pp. C292–C293.

[44] M. Perenzoni, N. Massari, L. Gasparini, M. M. Garcia, D. Perenzoni, and D. Stoppa, "A fast 50 × 40-pixels single-point DTOF SPAD sensor with photon counting and programmable ROI TDCs, with σ <4 mm at 3 m up to 18 klux of background light," *IEEE Solid-State Circuits Lett.*, vol. 3, pp. 86–89, 2020.

[45] A. Tontini, L. Gasparini, and M. Perenzoni, "Numerical model of SPAD-based direct time-of-flight flash LiDAR CMOS image sensors," *Sensors*, vol. 20, no. 18, p. 5203, Sep. 2020.

[46] C. Niclass, M. Soga, H. Matsubara, M. Ogawa, and M. Kagami, "A 0.18-μmCMOS SoC for a 100-m-range 10-frame/s 200 ×96-pixel time-of-flight depth sensor," *IEEE J. Solid-State Circuits*, vol. 49, no. 1, pp. 315–330, Jan. 2014.

[47] H. Xu, L. Pancheri, G.-F. D. Betta, and D. Stoppa, "Design and characterization of a p+/n-well SPAD array in 150 nm CMOS process," *Opt. Exp.*, vol. 25, no. 11, pp. 12765–12778, May 2017.

Alessandro Tontini (Member, IEEE) received the M.S. degree in telecommunications engineering from the University of Trento, Trento, Italy, in 2017, where he is currently pursuing the Ph.D. degree with the Information Engineering and Computer Science Doctoral School, focusing on advanced techniques for SPAD-based CMOS d-ToF systems.

Since 2017, he has been a Researcher with the Center for Sensors and Devices, Fondazione Bruno Kessler, Trento, focusing on the modeling and development of SPAD-based CMOS image sensors for Time-of-Flight/LiDAR, quantum random number generations and bio-/medicine applications.

Sonia Mazzucchi received the M.S. degree in physics and the Ph.D. degree in mathematics from the University of Trento, Trento, Italy, in 1999 and 2003, respectively.

She was an Alexander von Humboldt Fellow at the Institute of Applied Mathematics, University of Bonn, Bonn, Germany. She is currently an Associate Professor in Probability and Mathematical Statistics at the Department of Mathematics, University of Trento. Dr. Mazzucchi has authored numerous research articles and monographs in the area of stochastic analysis and its applications to quantum mechanics, quantum information, and quantum technologies.

Roberto Passerone (Member, IEEE) received the M.S. degree in EECS from the Politecnico di Turino, Turin, Italy, in 1994, and the Ph.D. degree in EECS from the University of California at Berkeley, Berkeley, CA, USA, in 2004.

He was a Research Scientist at Cadence Design Systems, San Jose, CA, USA. He is currently an Associate Professor of Electronics at the Department of Information Engineering and Computer Science, University of Trento, Trento, Italy. Dr. Passerone has authored numerous research papers on international conferences and journals in the area of design methods for systems and integrated circuits, formal models, and design methodologies for embedded systems.

Dr. Passerone has served as the general, program, and track chair for various editions of SIES and ETFA and was the Track Chair for DATE. He has participated to several European, national, and industrial projects, and was a Local Coordinator for the University of Trento for ArtistDesign, COMBEST, and Ciphers.

Nicolò Broseghini received the B.S. degree in physics from the University of Trento, Trento, Italy, in 2021, where he is currently pursuing the M.S. degree in physics. His studies are focusing on integrated photonics and optoelectronics. He is currently working on his thesis project on the characterization of heralded single-photon source.

Leonardo Gasparini (Member, IEEE) received the Ph.D. degree in information and communication technologies from the University of Trento, Trento, Italy, in 2011, with a focus on ultralow-power camera systems.

Since 2010, he has been a Researcher with the Center for Sensors and Devices, Fondazione Bruno Kessler, Trento, where he is currently the Head of the Integrated Readout-ASICs and Image Sensors (IRIS) Research Unit. Here, he is engaged in the development of CMOS image sensors, including single-photon detector arrays for time-of-flight/LiDAR systems, bio-medicine, particle physics, and quantum optics.

Open Access funding provided by 'Università degli Studi di Trento' within the CRUI CARE Agreement