

Article

Statistical Modelling of SPADs for Time-of-Flight LiDAR

Alfonso Incoronato *, Mauro Locatelli and Franco Zappa * 

Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Piazza Leonardo da Vinci 32, I-20133 Milano, Italy; mauro2.locatelli@mail.polimi.it

* Correspondence: alfonso.incoronato@polimi.it (A.I.); franco.zappa@polimi.it (F.Z.); Tel.: +39-02-2399-6149 (F.Z.)

Abstract: Time-of-Flight (TOF) based Light Detection and Ranging (LiDAR) is a widespread technique for distance measurements in both single-spot depth ranging and 3D mapping. Single Photon Avalanche Diode (SPAD) detectors provide single-photon sensitivity and allow in-pixel integration of a Time-to-Digital Converter (TDC) to measure the TOF of single-photons. From the repetitive acquisition of photons returning from multiple laser shots, it is possible to accumulate a TOF histogram, so as to identify the laser pulse return from unwelcome ambient light and compute the desired distance information. In order to properly predict the TOF histogram distribution and design each component of the LiDAR system, from SPAD to TDC and histogram processing, we present a detailed statistical modelling of the acquisition chain and we show the perfect matching with Monte Carlo simulations in very different operating conditions and very high background levels. We take into consideration SPAD non-idealities such as hold-off time, afterpulsing, and crosstalk, and we show the heavy pile-up distortion in case of high background. Moreover, we also model non-idealities of timing electronics chain, namely, TDC dead-time, limited number of storage cells for TOF data, and TDC sharing. Eventually, we show how to exploit the modelling to reversely extract the original LiDAR return signal from the distorted measured TOF data in different operating conditions.

Keywords: Single Photon Avalanche Diode (SPAD); Silicon Photo-Multipliers (SiPM); Light Detection and Ranging (Lidar); Time-of-Flight (TOF) measurements; Monte Carlo simulations



Citation: Incoronato, A.; Locatelli, M.; Zappa, F. Statistical Modelling of SPADs for Time-of-Flight LiDAR. *Sensors* **2021**, *21*, 4481. <https://doi.org/10.3390/s21134481>

Academic Editor: Marco Leo

Received: 13 June 2021
Accepted: 28 June 2021
Published: 30 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Light Detection and Ranging (LiDAR) is an optical technique widely used to measure the distance of a target and to acquire 3D depth-resolved maps of a scene, and is employed in several fields of science and everyday life, such as automotive applications [1], gesture recognition [2], 3D scanners for virtual prototyping [3], and security surveillance.

In Time-of-Flight (TOF) LiDAR, the target is illuminated with light and a camera collects backscattered photons to evaluate the distance of the target; depending on the light modulation, this can be performed in different ways [4,5]. The so-called 'direct' TOF, employs a pulsed laser and a time measurement circuit to directly measure laser pulse's TOF and then the distance D from the object, according to $D = \frac{1}{2} \cdot \text{TOF} \cdot c$, where c is the speed of light through the medium. In this approach, the TOF measurement jitter directly affects the distance precision: for 1 cm single-shot distance precision, the time measurement system must have uncertainty below 70 ps.

In many applications, eye-safety considerations place limits on the wavelength, optical power, and optical system (e.g., focal length, numerical aperture, etc.) used to illuminate the scene [6]; moreover, at long distances, light attenuation and beam divergence can reduce the light intensity that can be received by the TOF sensor. Therefore, Single Photon Avalanche Diode (SPAD) detectors and SPAD arrays can be employed as TOF imagers, thanks to their single-photon sensitivity and ability to integrate one Time-to-Digital Converter (TDC) in each pixel, allowing to reconstruct high frame rate 3D images [7]. Their gating capability can be exploited to image defined regions of interest in the distance plane, suppress direct leakage between illumination laser beam and sensor (e.g., when both are physically placed

in the same enclosure), or even to suppress specific reflections from objects inside the scene, allowing for applications such as around-the-corner imaging [8], where the content of a scene hidden from direct line of sight can be reconstructed exploiting multiple reflections of a light beam. A drawback of using a SPAD in a LiDAR system is its extreme sensitivity, since it gets triggered by at least one photon, with no chance to signal how many concurrent photons composed the light pulse. This causes a strong susceptibility to background light (i.e., light naturally occurring in the environment, uncorrelated with the TOF illumination source), whose photons can saturate either the SPAD or the TDC conversions or even the electronics resources (e.g., TOF storage cells, digital readout speed, etc.), thus preventing the detection of just the desired signal.

A Silicon Photomultiplier (SiPM) is a solid-state single-photon detector which is constituted by many microcells all connected in parallel, for a single pixel detector. Each microcell consists in one SPAD and one passive quenching resistor, to provide a self-quenched analog current pulse when at least one photon hits the SPAD. Being the microcells all connected in parallel [9], the SiPM output consists of an analog current pulse, whose intensity is proportional to the number of concurrent photons hitting the SiPM (i.e., all microcells' SPADs): for this reason, SiPM are photon-number resolved compared to single SPADs; hence, they allow to discriminate a bunch of signal photons (e.g., the laser pulse return) from few randomly arriving background ones [5]. Detailed electrical models of SPADs [10] and SiPMs [11] have already been reported for proper optimization of the front-end sensing electronics.

A single SPAD or a single SiPM can be used in single-spot LiDAR measurements, but also in 3D scene acquisitions by means of 2D scanning of the scene [5]. Instead, SPAD arrays and SiPM arrays (where each SiPM acts as an independent pixel) allow to acquire 3D scenes with no moving part, with a much shorter acquisition time. For example, Bronzi et al. [12] report a SPAD array where the independent SPAD pixels can be used in photon-counting for acquiring 3D scenes in automotive environment, but with no background rejection. In Ref. [13] an array of SPADs with multi-TDCs is proposed, to provide strong background rejection and localization of the laser spot across the array. Instead, in Ref. [14] Sony presented a complete LiDAR system based on a 189×600 pixels SPAD array, in a 3D-stacked technology, where each pixel is a microcell composed of 3×3 or 4×4 SPADs.

The proper design of a TOF LiDAR system, from the selection of the detector, to the array sizing, the design of electronics and optics, requires a deep understanding of the working operation and limitations of both the detector (being either a SPAD, a SiPM, or an array of them) and the TOF timing electronics, the TDC, the storage memory, and the post-processing. In order to provide a valuable tool for the sizing of these components and the simulation of the expected acquisition waveforms of signal and background photons, we have developed a detailed statistical modelling for SPAD-based LiDAR systems.

In this paper, we present the analytical statistical modelling and we compare the results with those of a Monte Carlo simulator in very different operating conditions with very high background levels. We show how the two approaches lead to perfectly matching results and can be interchangeably used to predict LiDAR systems performance and to select design choices for both the SPAD detectors and the TOF processing electronics. In particular, moving from Monte Carlo simulations to a compact and comprehensive statistical model drastically reduces the computational time and enables more and faster comparisons in different operating scenarios. In the statistical modelling of the SPAD detectors, we take into consideration non-idealities such as hold-off time, afterpulsing, and crosstalk, which cause a strong distortion of the measurements due to the pile-up of multiple photons. Moreover, we also take into consideration the non-idealities of the timing electronics chain, namely, the TDC dead-time, the limited number of storage cells for the TOF data, and the acquired TOF histogram, all of which cause the saturation of measurements when a strong ambient background overwhelms signal photons.

Beyond the statistical modelling of the SPAD, we also present how to model the acquisition electronics needed to suppress background, by discussing general-purpose

methods such as coincident detection of multiple photons, time-gated detector operation, and multi-TDC architectures. Hence, we show how both modellings, starting from the actual LiDAR return waveform, can predict the final measured TOF histogram distribution in very different operating conditions. Eventually, we show how the reported modelling can be profitably exploited in an inverse way, starting from the acquired measurements, to reversely extract the original LiDAR return waveform, thus compensating for most distortions that could affect a real LiDAR system.

In Section 2 we present the statistical modelling of SPADs affected by all possible non-idealities, such as hold-off time, afterpulsing, and cross-talk. In Section 3 we model different acquisition techniques, such as for single- and multi-photon detections, coincident detections, time-gated approach and multi-TDCs architectures. Eventually, in Section 4 we correct the measured TOF data, by applying the modelling for both single-photon Time-Correlated Single Photon Counting setups and multi-photon acquisitions.

2. Statistical Modelling of the SPAD

At first, we start modelling SPAD arrays and digital SiPMs without taking into consideration the acquisition electronics and measurements technique. SPAD-based systems can experience saturations and distortions of measurements due to fact that they provide a digital pulse every time one photon (or more concurrent photons) hits the active area, triggers an avalanche current multiplication process [15], and gets sensed by a suitable quenching circuit, which provides a digital pulse [16]. Therefore, more photons hitting the SPAD at the same time provide just one pulse, such as if only one photon were detected. Moreover, after each triggering, the quenching circuit must quench the SPAD, wait a hold-off time, and then reset the SPAD back to operation (i.e., biased above its breakdown voltage); the overall minimum elapsed time between one triggering and the following potential one is called hold-off time or dead-time, during which the SPAD cannot detect other photons [17]. Both effects cause the so-called “pile-up” distortion in both photon-counting applications (i.e., when the actual number of incoming photons must be counted within a given frame time) and photon-timing acquisitions (i.e., when the actual waveform of an optical signal or multiple TOF measurements must be computed).

We start considering the case of non paralyzable dead-time, i.e., with a fixed duration from the first ignition, during which the SPAD is insensitive to photons: this is the way a classic active quenching circuit operates [16]. Then, we introduce other SPAD phenomena, such as afterpulsing (i.e., the probability to have a secondary avalanche ignition after a primary one, due to carriers trapping and subsequent release) and crosstalk (i.e., the probability to ignite a spurious avalanche due to the current flowing through another SPAD).

In order to compute the detection probability $P(DET_{xy}(t))$ of each SPAD of the array in position x and y , we start from the incoming photon rate Φ_{in} . The actual avalanche rate undergoes losses due to Photon Detection Probability (i.e., the product of quantum efficiency and avalanche triggering probability [15,17]) of SPADs, fill-factor (FF) and photon detection probability (PDP) of the imager: $\lambda_{detector}(t) = \Phi_{in}(t) \cdot FF \cdot PDP$. Such a rate is usually composed by two contributions $\lambda_{detector}(t) = \lambda_{signal}(t) + \lambda_{noise}(t)$: the former is given by the signal photon rate (usually concentrated in one or a few peaks), while the latter is given by background light and dark counts, which are assumed to be randomly scattered over time. Eventually, $\lambda_{detector}$ is considered to be equally distributed among the SPADs of the array; hence, in each SPAD we have $\lambda_{spad}(t) = \lambda_{detector}(t) / N_{spad}$. For the discrete-time approach, we employ a quantization time step T_{step} shorter than all transients involved in the measurements. Then, we normalize all time quantities with respect to such a step, e.g., $\Delta_{HO} = T_{HO} / T_{step}$.

In the following, each analytical estimation is cross-checked to MATLAB Monte Carlo simulations assumed as the golden reference when measurements are not available. Such simulations rely on the generation of random photon arrival times, according to Poisson statistics, and include all deterministic effects of SPAD detectors and related electronics.

Finally, multiple repetitions of the laser shot and the frame acquisition are implemented, so as to improve the measurements Signal-to-Noise Ratio, at different levels of signal, background and dark counts, and to check the model in extremely different situations.

2.1. Hold-Off

Defects inside the active volume of SPADs can trap some charge during an avalanche current flow; the following carrier release can trigger a second avalanche injection, to give a spurious event (either count or TOF). To reduce such an afterpulsing effect, the SPAD has to be promptly quenched and kept below breakdown (so becoming blind) for a hold-off time, during which trapped charges get released without triggering any avalanche. The necessary hold-off time to almost empty traps depends on the fabrication technology and also on the quenching electronics. The SPAD dead-time is usually defined as the hold-off time plus the quenching and resetting time of the electronics. To distinguish the SPAD dead time from other dead-time sources (e.g., TDC), it will be called hold-off time (HO).

Now, let us consider the distortion caused by the hold-off time. The probability to have a detection event $PH(k)$ at time index $k = t/T_{step}$ given an avalanche rate λ_{spad} is:

$$P(PH(k)) = 1 - e^{-\lambda_{spad}(k) \cdot T_{step}}$$

This event corresponds to an actual detection $DET(k)$ only if the SPAD is not in hold-off status ($\overline{HO}(k)$), without considering other avalanche triggering causes (e.g., afterpulsing and crosstalk). Thereby, detection probability is the intersection between the two independent events $PH(k)$ and $\overline{HO}(k)$:

$$P(DET(k)) = P(PH(k) \cap \overline{HO}(k)) = P(PH(k)) \cdot P(\overline{HO}(k))$$

A SPAD is in hold-off condition at the time index k if in the previous $T_{HOLDOFF}$ time interval, at least another detection occurred, that is:

$$P(HO(k)) = P\left(\overline{\bigcap_{j=k-\Delta_{HO}}^{k-1} \overline{DET}(j)}\right) = P\left(\bigcup_{j=k-\Delta_{HO}}^{k-1} DET(j)\right)$$

Since all the j events are mutually exclusive in the interval $[k - \Delta_{HO}, k - 1]$ (there can be only one detection), the previous equation can be rewritten as:

$$P\left(\bigcup_{j=k-\Delta_{HO}}^{k-1} DET(j)\right) = \sum_{j=k-\Delta_{HO}}^{k-1} P(DET(j))$$

By imposing the initial conditions (e.g., by either turning on the SPAD at $t = 0$ or by assuming stationary conditions), the probability of having a detection is given by:

$$P(DET(k)) = \left(1 - e^{-\lambda_{spad}(k) \cdot T_{step}}\right) \cdot \left(1 - \sum_{j=k-\Delta_{HO}}^{k-1} P(DET(j))\right) \quad (1)$$

In stationary conditions, where the probability of having a detection is constant, this equation can be rewritten as $P(DET(k)) = P(DET)$:

$$P(DET) = \frac{1 - e^{-\lambda_{spad} \cdot T_{step}}}{1 + \Delta_{HO} \cdot \left(1 - e^{-\lambda_{spad} \cdot T_{step}}\right)}$$

The equivalent rate in the continuous-time domain is $P(DET)/T_{step}$ for $T_{step} \rightarrow 0$, hence:

$$\lambda_{detection} = \lim_{T_{step} \rightarrow 0} \frac{1}{T_{step}} \cdot \frac{1 - e^{-\lambda_{spad} \cdot T_{step}}}{1 + \Delta_{HO} \cdot \left(1 - e^{-\lambda_{spad} \cdot T_{step}}\right)} = \frac{\lambda_{spad}}{1 + \lambda_{spad} \cdot T_{HOLDOFF}}$$

The result converges to others reported in literature [18,19] and is shown also in Figure 1, which compares this model with a Monte Carlo simulation, in case of 80 background pho-

tons within the simulation time (500 ns) and 20 signal photons within a 5 ns FWHM gaussian shape pulse, centred at a 300 ns. All simulations were performed with $T_{step} = 100$ ps quantization step and 100,000 repetitions. It can be noticed the heavy pile-up distortion due to the 10 ns hold-off time: the expected peak at 300 ns is shifted toward longer arrival times and gets compressed. Furthermore, in stationary conditions (constant λ_{spad}) the actual detection rate is lowered by a factor $1 + T_{HO} \cdot \lambda_{spad}$. As can be seen, the model perfectly matches simulations since no approximation have been introduced.

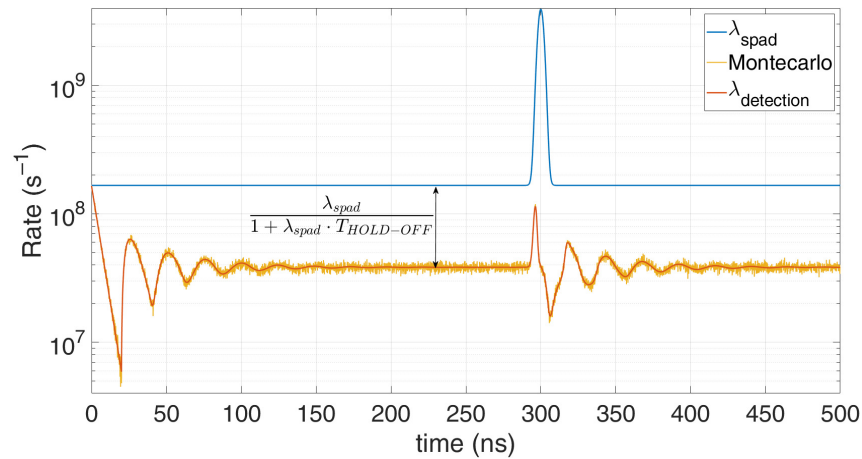


Figure 1. Comparison between a Monte Carlo simulation (yellow) and the proposed model (red), in case of 80 background photons and 20 signal photons within 500 ns simulation time. Note the perfect match; the discrepancy compared to the actual distribution (blue) is due to the SPAD hold-off time.

2.2. Afterpulsing

Now let us add afterpulsing into the model. A SPAD triggering ($DET(k)$) can have two distinct causes: either a photon detection ($PH(k)$) or an afterpulsing event ($AP(k)$). These two events can occur at the same time index k , thus they are not mutually exclusive:

$$\begin{aligned} P(DET(k)) &= P((PH(k) \cup AP(k)) \cap \overline{HO}(k)) = \\ &= P(PH(k) \cap \overline{HO}(k)) + P(AP(k) \cap \overline{HO}(k)) - P(PH(k) \cap \overline{HO}(k)) \cdot P(AP(k) \cap \overline{HO}(k)) \end{aligned}$$

The probability of an afterpulsing event $AP(k)$ requires the knowledge of the afterpulsing probability temporal distribution of the detector, i.e., $P(AP(k)|DET_{xy}(k-i))$, for i ranging from Δ_{HO} to Δ_{APmax} , this last corresponding to the time interval after a photon detection, beyond which the afterpulsing probability is negligible. Such time distribution can be experimentally measured, by observing the time interval between two consecutive SPAD ignitions with no light illumination, after subtracting the dark counts. An example of probability distribution is shown in Figure 2, adapted from [20].

An afterpulsing event at time k can take place only if a detection occurred at instant $k - \Delta$ and no detection happened in all following instants:

$$AP(k) = \bigcup_{i=\Delta_{HO}}^{\Delta_{APmax}} DET(k-i) \cap AP(k) \Big| DET(k-i) \cap \bigcap_{j=k-i+1}^{k-1} \overline{DET}(j)$$

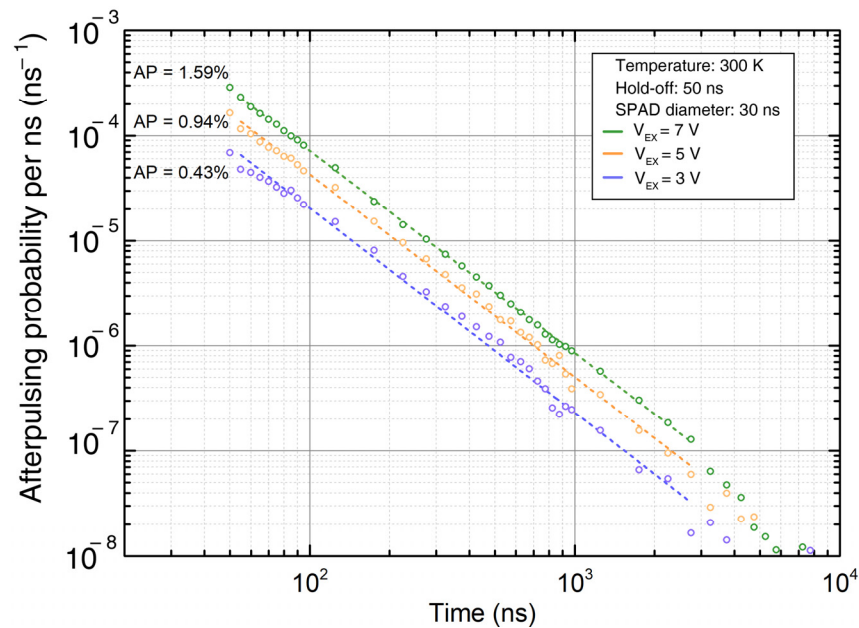


Figure 2. Example of experimental measurements of afterpulsing probability distributions.

Since $\cap_{j=k-i+1}^{k-1} \overline{DET}(j)$ implies $\overline{HO}(k)$, then $AP(k) \cap \overline{HO}(k) = AP(k)$. Hence, being all events $\overline{DET}(j)$ mutually exclusive, the probability to have an afterpulsing is:

$$P(AP(k) \cap \overline{HO}(k)) = \sum_{i=\Delta_{HO}}^{\Delta_{APmax}} P(DET(k-i)) \cdot P(AP(k)|DET(k-i)) \cdot P\left(\cap_{j=k-i+1}^{j-1} \overline{DET}(j) \mid AP(k) \cap DET(k-i)\right)$$

Notice that: $P(DET(k-i))$ is known, being computed in one of the previous times; $P(AP(k)|DET(k-i))$ is an experimental data (e.g., see Figure 2); and the last term:

$$\cap_{j=k-i+1}^{j-1} \overline{DET}(j) \mid AP(k) \cap DET(k-i)$$

is the probability that no photon has been detected from $k-i+1$ to $k-1$. Namely:

- Given $DET(k-i)$, the SPAD is held-off for $k-i+1 < j \leq k-i+\Delta_{HO}$;
- Being $AP(k)$ the afterpulsing effect, the detection in $k-i$ does not generate afterpulsing within $k-i < j < k$;
- For every j , given $\cap_{j=k-i+1}^{j-1} \overline{DET}(j)$, the SPAD is not held-off for $k-i+\Delta_{HO} < j < k$.

These considerations result in:

$$\cap_{j=k-i+1}^{k-1} \overline{DET}(j) \mid AP(k) \cap DET(k-i) = e^{(-\sum_{j=k-i+\Delta_{HO}+1}^{k-1} \lambda_{spad}(j) \cdot T_{step})}$$

So, the detection probability in Equation (1) can be written as the sum of two items:

$$P(AP(k) \cap \overline{HO}(k)) = \sum_{i=\Delta_{HO}}^{\Delta_{APmax}} \left(P(DET(k-i)) \cdot P(AP(k)|DET(k-i)) \cdot e^{(-\sum_{j=k-i+\Delta_{HO}+1}^{k-1} \lambda_{spad}(j) \cdot T_{step})} \right)$$

and: $(PH(k) \cap \overline{HO}(k)) = (1 - e^{-\lambda_{spad}(k) \cdot T_{step}}) \cdot (1 - \sum_{j=k-\Delta_{HO}}^{k-1} P(DET(j)))$ The third term of Equation (1) can be neglected in case of small T_{step} .

A stationary regime model can be computed as:

$$\lambda_{detection} = \lim_{T_{step} \rightarrow 0} \frac{P(DET)}{T_{step}} = \frac{\lambda_{spad}}{1 + \lambda_{spad} \cdot T_{HO} - \int_{T_{HO}}^{t_{APmax}} P(AP(t)|DET(t - \Delta t)) \cdot \exp(-\lambda_{spad} \cdot (\Delta t - T_{HO})) d\Delta t}$$

A good approximation is to consider the afterpulsing event right after the end of the hold-off time, i.e., a delta-like afterpulsing probability in time $P(AP(t)|DET(t - \Delta t)) = P_{AP} \cdot \delta(t - T_{HO})$ [21], to obtain the detection rate:

$$\lambda_{detection} = \frac{\lambda_{spad}}{1 + \lambda_{spad} \cdot T_{HO} - P_{AP}}$$

Such an approximation holds better when afterpulsing probability distribution becomes much shorter than the inverse of the average photon rate, so that $e^{-\lambda_{spad} \cdot (\Delta t - T_{HO})} \simeq 1$. In this case, we get the total afterpulsing probability as:

$$\int_{T_{HO}}^{t_{APmax}} P(AP(t)|DET(t - \Delta t)) \cdot e^{-\lambda_{spad} \cdot (\Delta t - T_{HO})} d\Delta t \simeq P_{AP}$$

To highlight the afterpulsing effect, in Figure 3 we set low background (1.5 photons) and narrow pulse signal (10 photons in 1 ns FWHM), with 5 ns hold-off time.

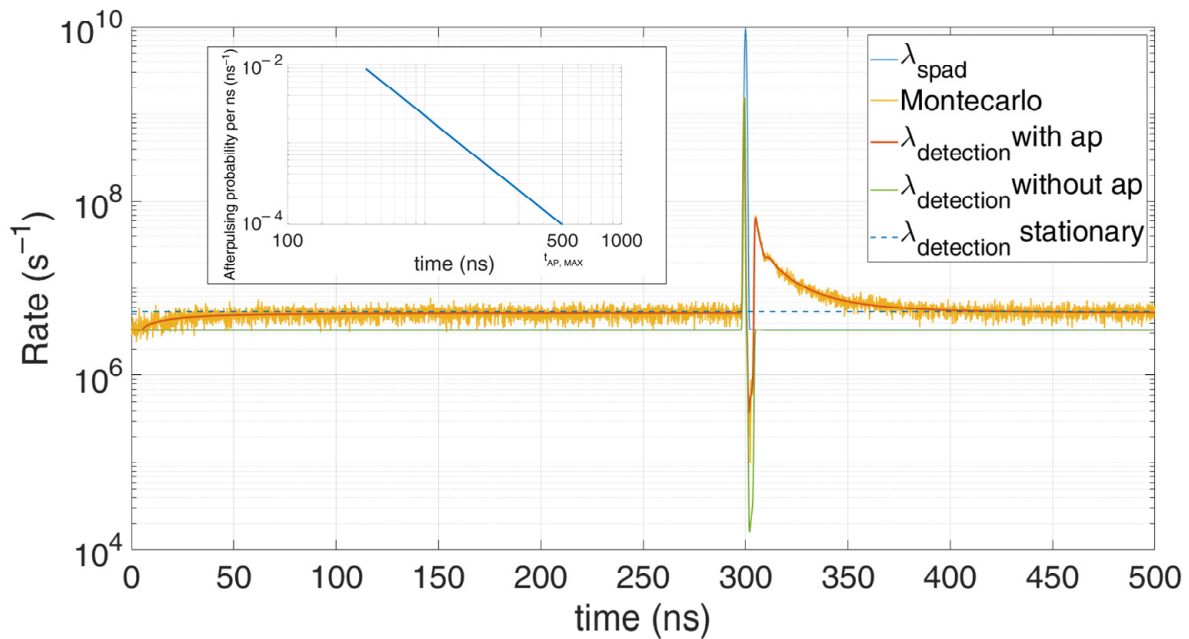


Figure 3. Comparison between Monte Carlo simulations and the model taking into consideration the afterpulsing probability (shown in the inset). Note their perfect match; the distortion compared to the actual expected distribution (blue) is due to the SPAD afterpulsing.

2.3. Crosstalk

SPAD arrays involve many detectors on the same chip being triggered independently. Such an activity can spuriously trigger neighboring SPADs through optical crosstalk. This phenomenon is defined as the probability to induce a false ignition in a victim SPAD due to a triggering in another aggressor SPAD, after an average delay time T_{XT} linked to the avalanche build-up time. In order to fully characterize crosstalk, a probability matrix needs to be assessed [22], by triggering one specific SPADs (the aggressor) and observing the ignition rates of neighboring victims. Figure 4 shows such a matrix.

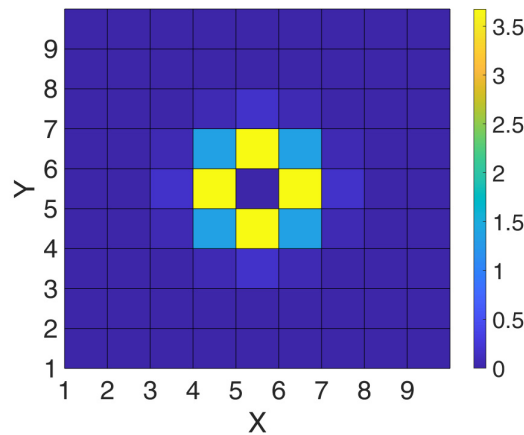


Figure 4. Example of crosstalk probability matrix P_{XT} in a 9×9 SPAD array.

The generation event due to crosstalk is defined as XT_{xy} and depends on all the neighbour SPADs activities:

$$P(XT_{xy}(k)) = P(\cup_{i,j} XT_{xy \leftarrow i,j}(k)) = 1 - P(\cap_{i,j} \overline{XT}_{xy \leftarrow i,j}(k)) = 1 - \prod_{adj} P(\overline{XT}_{xy \leftarrow i,j}(k))$$

each of them has a certain probability (P_{XT}) to generate an avalanche in the xy SPA.

$$P(\overline{XT}_{xy \leftarrow i,j}(t)) = 1 - P_{XT_{i,j}} \cdot P(DET_{i,j}(t - T_{XT}))$$

Now, two events can trigger an avalanche (PH and XT_{xy}) if the SPAD is not in hold-off, so the probability to have a detection at time index k is now:

$$P(DET_{xy}(k)) = P((PH_{xy}(k) \cup XT_{xy}(k)) \cap \overline{HO}(k))$$

$XT_{xy}(k)$ and $\overline{HO}(k)$ are not independent, since this second event is related to a possible crosstalk induced from the SPAD to its neighbors (i, j) that can generate a second crosstalk to the SPAD (XT_{XY}) in the Δ_{HO} . Such a double crosstalk event can be considered negligible for small crosstalk probabilities ($P_{XT} < 10\%$), so the two events may be independent.

$$P(DET_{xy}(k)) = P(PH_{xy}(k) \cup XT_{xy}(k)) \cdot P(\overline{HO}(k))$$

and

$$P(PH_{xy}(k) \cup XT_{xy}(k)) = P(PH_{xy}(k)) + P(XT_{xy}(k)) - P(PH_{xy}(k)) \cdot P(XT_{xy}(k))$$

Similar to the afterpulsing model, the last product is negligible for small time steps.

3. Modelling of the Acquisition Technique

Many applications rely on time-resolved photon-counting techniques (also known as photon-timing) to attain high timing resolution and precision, from biological applications such as Fluorescence Lifetime Imaging (FLIM) [23], near-infrared spectroscopy [24], to VLSI chip inspection [25], to looking-around the corner vision [26] to LiDAR [26]. In this section, we investigate two main photon-timing techniques: classical Time-Correlated Single Photon Counting (TCSPC), usually being single-hit and with SPAD reset synchronous with the excitation laser pulse; and multi-hit Time-of-Flight (TOF) measurements, with free-running SPAD and multiple detections of photons within the frame time. The former is single-hit because it registers the timing information of just the first photon and, by creating a histogram, it reconstructs the incoming light waveform in photon starving application, when at the most one photon per laser shot is expected. Instead, the latter technique can be exploited at high counting rates, when more than one photon is received per laser

excitation. Ideally, all photons should be acquired, though this is impossible due to the SPAD dead-time and the dead-time of the timing electronics (e.g., the TDC conversion time) and other resources limitations (e.g., on-chip memory availability and readout speed). The following subsections model some techniques copying those effects.

3.1. Single-Hit Detection

TCSPC is the key technique to acquire very faint signals down to the picosecond time scale; however, particular care on the count rate must be taken. Since the technique relies on the detection and timestamping of the first photon's timing information, this gives rise to a strong "pile-up" distortion when more than one photon reaches the detector, but only the first one is recorded. This effect is negligible for count rates lower than 5% of the laser frequency [27,28], i.e., when the probability to have more than one photon is less than 0.1% for each laser excitation, as predicted by Poisson statistics. In fact, the probability that a photon be detected at a given time, if no photon has been detected yet, is:

$$P(PH(t)) = 1 - e^{-\lambda_{spad}(t) \cdot T_{step}}$$

In case of single-hit technique, the detection at time t is possible only if an avalanche is generated *and* no photon has been revealed previously since the start of the observation time (namely, the last laser excitation). Hence:

$$P(\cap_{k=0}^t \overline{DET}(k)) = e^{-\sum_{k=0}^t \lambda_{spad}(k) \cdot T_{step}}$$

These two events being independent [29], we get:

$$P(DET(t)) = e^{-\sum_{k=0}^t \lambda_{spad}(k) \cdot T_{step}} \cdot \left(1 - e^{-\lambda_{spad}(t) \cdot T_{step}}\right)$$

The same result [30] can be achieved by exploiting the recursive expression as in Equation (1), by prolonging the hold-off to the whole observation time before time index k :

$$P(\cap_{k=0}^t \overline{DET}(k)) = P(1 - \cup_{j=0}^{k-1} DET(j)) = 1 - \sum_{j=k-\Delta_{HO}}^{k-1} P(DET(j))$$

So, to get:

$$P(DET(k)) = \left(1 - e^{-\lambda_{spad}(k) \cdot T_{step}}\right) \cdot \left(1 - \sum_{j=0}^{k-1} P(DET(j))\right) \quad (2)$$

Figure 5 shows how this model perfectly matches the Monte Carlo simulation: an average of 0.5 signal photons and four noise photons are considered within the 1 μ s observation time. The typical exponential decay due to pile-up distortion is clearly visible.

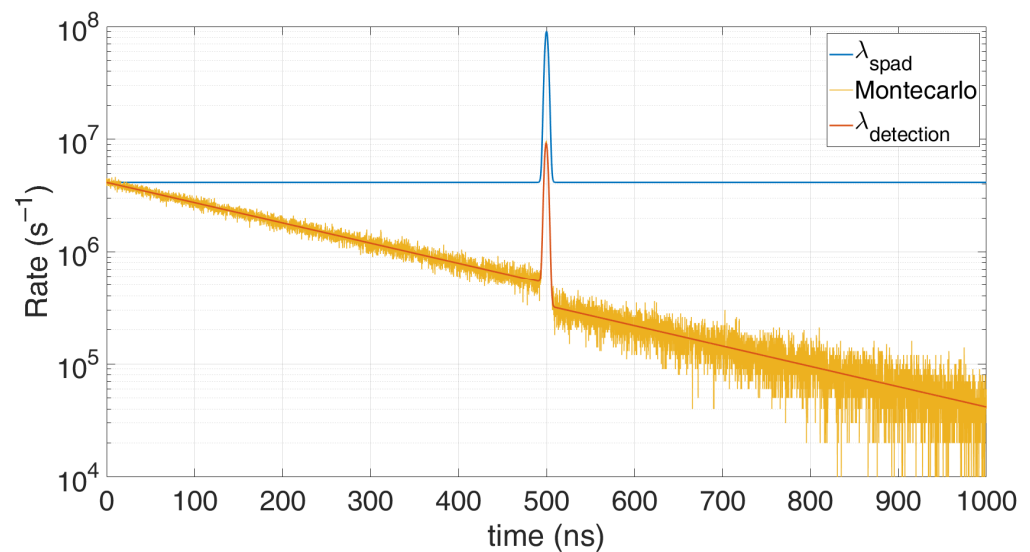


Figure 5. Comparison of the TCSPC model (red) with the Monte Carlo simulation (yellow), in the case of 0.5 signal photons and 4 noise photons on average within the 1 μ s observation time. Note their perfect match; the decaying trend compared to the actual expected distribution (blue) is due to pile-up caused by the single-hit acquisition.

3.2. Multi-Hit Detection

Most TOF applications cannot trust on single photon regime, for example due to the presence of high-level background light, which can represent a huge source of noise, even up to 100 klx in many automotive environments. While luminance refers to the amount of light that passes through, emits from, or reflects off an object and falls within a given solid angle (and is measured in candela per square meter, cd/m^2), illuminance refers to the total amount of light incident onto a given surface, per unit area, and is measured in lux (lx, i.e., lumens per squared meter) and is wavelength-weighted to correlate with human brightness perception. Illuminance is often called brightness, but the latter is meant just for nonquantitative references to physiological sensations and perceptions of light. Actually, in the LiDAR field, background light is often quoted in terms of lux, both by companies and by literature. Other typical requirements in TOF LiDAR are high laser pulse energy and high dynamic range, needed to reveal object from few meters up to hundreds of meters.

Therefore, all these applications cannot rely on the detection of just the first returning photon after each laser shot, such as in classical single-hit TCSPC techniques, because measurements would show a huge pile-up distortion (as shown in Figure 5). The ideal solution would be to have no dead-time due to either the detector (e.g., the SPAD hold-off time) or the timing electronics (e.g., the TDC conversion time), and the possibility to record the highest number of TOF acquisitions (ideally all photons' TOFs) within the frame-time (e.g., the maximum distance full-scale range) after each laser shot. Of course, this is not possible, due to SPAD hold-off time (usually some ns), TDC dead-time (from few ns to some μ s, depending on the architecture), limited readout bandwidth (above all with laser repetition exceeding some MHz) and limited storage availability. Short hold-off SPADs, fast re-triggerable TDCs, and fast timing architectures are being studied to achieve shorter and shorter dead-times. Furthermore, histogram generation and accumulation and TOF centroid evaluation and readout are being implemented on-chip to avoid the bottleneck of chip's readout bandwidth and off-chip data storage and processing. Some studies in literature try to model the combined dead-time effect [31]; however, that is feasible when considering a specific architecture, e.g., a SPAD with a TDC [32]. Instead, for more complex architectures that modelling becomes very difficult to implement. In the following, we develop a statistical model able to describe the timing electronics dead-time

(T_{DEAD}) and also the limited number of TOF data to be store in a frame for the case of a multi-hit architecture.

Figure 6 shows a generic detection architecture employing either a single SPAD or a combination of SPADs (such as in a SiPM) to provide one output event a time. The detector provides a defined event rate (i.e., single photons, coincidence photons, or logic-OR combination among SPADs, etc.) that trigger one single TDC, resulting in the TOF time-stamp to be stored into a register. As shown in Figure 6, a time-stamp is generated when a TRIGGER event triggers the timing electronics, which was quiescent for at least the previous T_{DEAD} time.

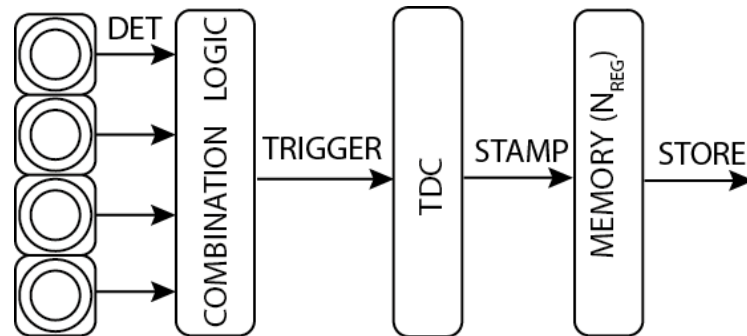


Figure 6. Typical photon-timing chain, from photon detection to Time-of-Flight storage.

In order to simplify the model, we introduce an approximation: the trigger probability is considered to be independent of the previous time instant, since the analytical dependency on what happened in the previous T_{DEAD} is not straightforward to compute. Therefore, similarly to the model with just hold-off (see Par. 2), a TDC conversion (STAMP event) occurs whether a trigger event takes place and no conversion happened in the last TDC dead-time:

$$\begin{aligned}
 P(STAMP(k)) &= P\left(TRIGGER(k) \cap \bigcap_{j=k-\Delta_{DEAD}}^{k-1} \overline{STAMP(j)}\right) \\
 &= P(TRIGGER(k)) \cdot P\left(\bigcap_{j=k-\Delta_{DEAD}}^{k-1} \overline{STAMP(j)} \mid TRIGGER(k)\right) \quad (3) \\
 &\simeq P(TRIGGER(k)) \cdot \left(1 - \sum_{j=k-\Delta_{DEAD}}^{k-1} P(STAMP(j))\right)
 \end{aligned}$$

For single SPAD architectures (i.e., no combination logic), if the detector dead-time is dominant, i.e., $T_{HO} \geq T_{DEAD}$, we have:

$$P\left(\bigcap_{j=k-\Delta_{DEAD}}^{k-1} \overline{STAMP(j)} \mid TRIGGER(k)\right) = 1$$

and $P(STAMP(k)) = P(TRIGGER(k)) = P(DET(k))$, as in Equation (1). In this case, the first model described by Equation (1) fits perfectly, since only one time stamp can occur after one detection. However due to the approximation, the model in Equation (3) gives the error shown in Figure 7. Conversely, in case the SPAD hold-off time is negligible compared to the TDC dead-time, the dependency of TRIGGER event with the previous instants is very weak, so the approximation holds better, as shown in Figure 8.

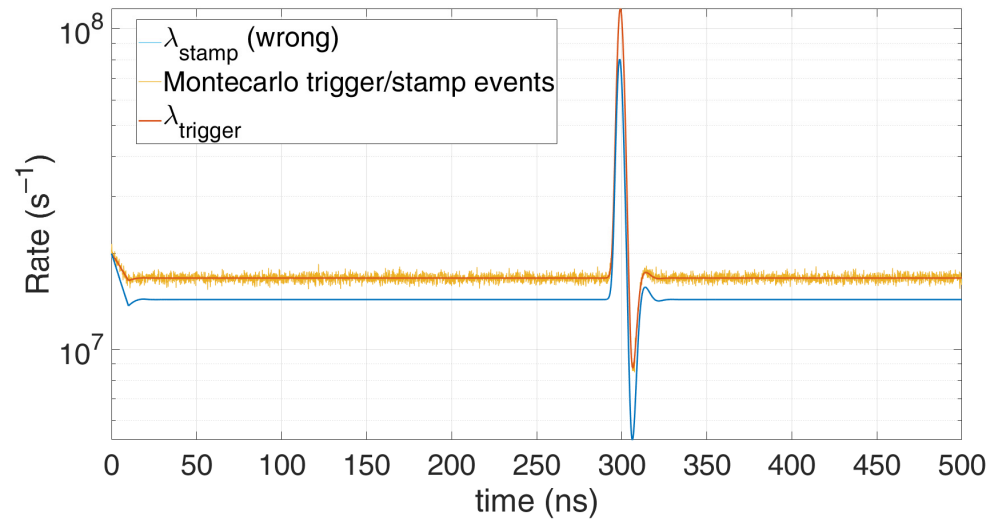


Figure 7. Comparison between Monte Carlo simulation and model with timing electronics dead-time, single SPAD, $T_{HO} \geq T_{DEAD} = 10$ ns (worst case, when the approximation does not hold, so Equation (1) should be used instead of Equation (3)), with 1 signal photon and 10 noise photons on average per repetition.

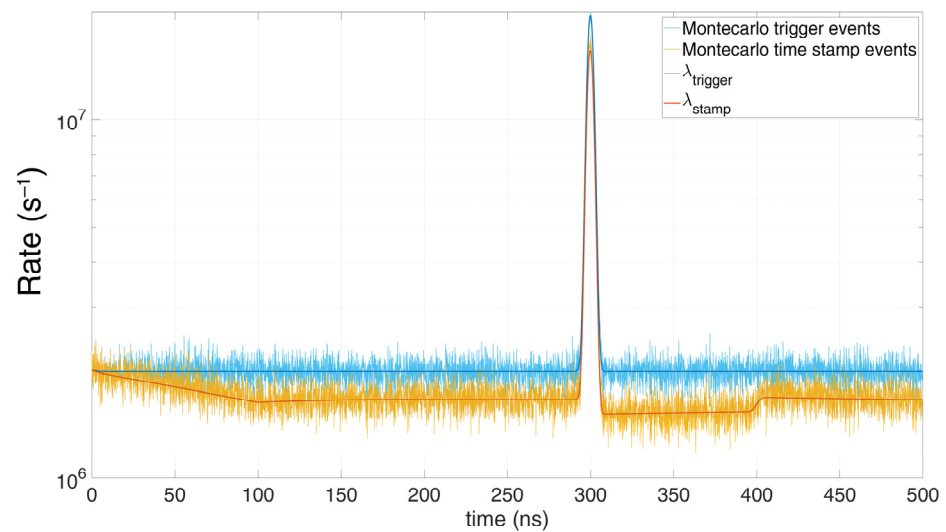


Figure 8. Comparison between Monte Carlo simulation and model with timing electronics dead-time, single SPAD, $T_{HO} > T_{DEAD}$, with 0.1 signal photon and 1 noise photon on average per repetition ($T_{HO} = 5$ ns, $T_{DEAD} = 100$ ns).

Now, let us consider the limitations in reading out and storing the TOF data: one measure can be stored (STORE event) when a time-stamp occurs (STAMP event) and the memory is not saturated. At first, we consider a trigger event corresponding to a time-stamp, i.e., no TDC dead-time for now, then:

$$P(\text{STORE}(k) = 1) = P(\text{STAMP}(k) \cap \overline{\text{SAT}}(k))$$

A measurement saturates ($\text{SAT}(k) = 1$) if the number of time stamps before the i -th bin is higher or equal to the available number of storage registers N_{REG} :

$$P(\overline{\text{SAT}}(k)) = P(N_{\text{STAMP}}(k-1) < N_{\text{REG}})$$

Since $P(STAMP)$ depends on the photon rate, it is not constant over time. Thus, the exact computation of $P(\overline{SAT}(k))$ would require the computation of all possible combinations with $N_{STAMP}(k-1) < N_{REG}$. Within a T_{obs} observation time, in the worst case there are $\binom{N_{TOTAL\ BIN}}{N_{REG}}$ possible combinations, whose analytic computation is not feasible. Moreover, it would be also necessary to know the dependency between different $STAMP(k)$ events. In order to simplify the analysis, we made the following assumptions:

- All $STAMP(k)$ events are independent;
- Being $P(STAMP(k))$ almost constant within the whole observation period apart from the peak duration, the probability distribution in the previous $k-1$ windows ($N_{STAMP}(k-1)$) is considered a binomial distribution with a number of trials $n(k) = k-1$ and a probability per trial equal to the average probability: $p(k) = \frac{\sum_{j \in [1, k-1]} P(STAMP(j))}{n}$.

The binomial probability is defined as:

$$B(n, i, p) = \binom{n(k)}{i} \cdot p(k)^i \cdot (1 - p(k))^{n(k)-i}$$

By applying the assumptions, $P(\overline{SAT}(k))$ and $P(STORE(k))$ can be computed as:

$$\begin{aligned} P(\overline{SAT}(k)) &= P(N_{coinc}(k-1) < N_{sat}) = \sum_{i=0}^{N_{sat}-1} B(n(k), i, p(k)) \\ P(STORE(k)) &= P(STAMP(k) \cap \overline{SAT}(k)) = P(STAMP(k)) \cdot P(\overline{SAT}(k)) \end{aligned} \quad (4)$$

Figure 9 shows the pile-up distortion due to the limited number of registers, the first measurements (at shorter time) have higher probability to be stored since storage registers are still empty. The assumptions made to simplify the model can still provide a good approximation. In the following, in order to avoid such a heavy distortion, we investigate three different solutions, namely, photon coincidence, time gating, and multi-TDCs.

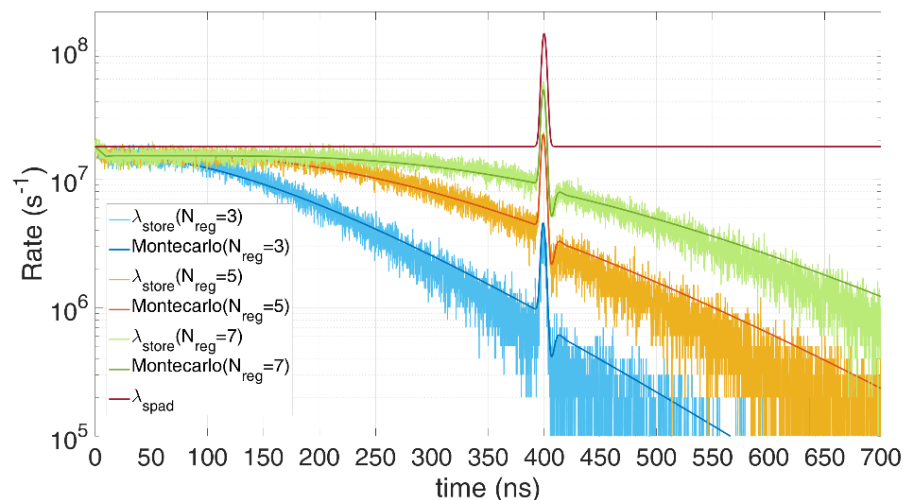


Figure 9. Comparison between Monte Carlo simulations and model in case of register limitations (3, 5, or 7 maximum TOF storable values), single SPAD, with 0.6 signal photons and 12 noise photons on average per repetition ($T_{HO} = 10$ ns). Note the perfect matches; the distortion compared to the actual expected distribution (blue and red top lines) is due to the limited available resources.

3.3. Coincidence Detection

In order to reduce the impact of the detector dead-time and to provide a sort of background suppression, one possibility is to exploit multi-SPAD pixel architectures and photon coincidence [30]. Even if background photons trigger a large number of SPADs within the sensor at a certain rate, these events are usually random and uncorrelated;

instead, signal photons arrive at a precise moment in time, the TOF of the laser pulse, so they are coincident within the laser pulse width duration. Therefore, as soon as a defined number of SPADs fire in such a coincidence time window, the coincidence logic triggers the timing electronics to compute the signal TOF, while being almost insensitive to background. Hence, this solution reduces the probability to saturate measurements and resources, because the timing electronics is triggered only in case of useful signal photons.

Let us now consider one single pixel with one TDC and an array of $m \times n$ SPADs, acting just as a single active area (unlike an x and y imager). The SPADs outputs are combined into an AND-like logic gate, which provides a pulse to the TDC only if at least $i \geq N_t$ SPADs trigger concurrently. Here, we compute the probability $P(COINC(j))$ that N_t photons trigger the TDC in a coincidence time window j .

At first, we neglect any crosstalk so all SPAD events can be considered independent of each other. Two different coincidence approaches are investigated: fixed-window and moving window. The former is used in synchronous systems, where the sum of SPADs contributions is observed at the end of each coincidence time window, by checking if the N_t threshold is exceeded. Instead, the latter is implemented with a suitable asynchronous coincidence electronics.

3.3.1. Fixed Coincidence Window

With a uniform photon distribution among the $m \times n$ SPADs in the same pixel, the probability that a photon hits one specific SPAD (of coordinated x, y) is given by Equation (1). Since detection events in a SPAD are mutually exclusive (considering $T_{window} < T_{HO}$), the probability of having a detection in a time window j (whose width index is $\Delta_{window} = T_{window}/T_{step}$) is given by:

$$P(DET_{xy, \Delta_{window}}(j)) = P(\cup_{k \in \Delta_{window}} DET_{xy}(k)) = \sum_{k \in \Delta_{window}} P(DET_{xy}(k))$$

A coincidence event $COINC_{N_t}(j)$ occurs when at least N_t detections arise in the same coincidence window ($DET_{xy, T_{window}}(j)$). Indicating with $comb$ one of the possible N_{comb} combinations to have N_{DET} SPADs triggered among N_{SPAD} available ones, the probability of having a coincidence in the window j is:

$$P(COINC_{N_t}(j)) = P\left(\cup_{N_{DET}=N_t}^{N_{spad}} \left(\cup_{comb_x=1}^{N_{comb}} \left(\cap_{spad_{index}=1}^{N_{spad}} EVENT(spad_{index})\right)\right)\right)$$

where $N_{comb} = \binom{N_{spad}}{N_{DET}}$ and

$$EVENT(spad_{index}) = \begin{cases} DET_{spad_{index}, \Delta_{window}}(j) & \text{if } spad_{index} \in combination \\ \overline{DET_{spad_{index}, \Delta_{window}}(j)} & \text{if } spad_{index} \notin combination \end{cases}$$

In the case under exam, the probability of a detection event can be considered constant among different SPADs, thus the expression can be simplified to:

$$P(COINC_{N_t}(j)) = \sum_{N_{DET}=N_t}^{N_{spad}} B\left(N_{spad}, N_{DET}, P(DET_{T_{window}}(j))\right)$$

Figure 10 shows the clear benefit of photon coincidence in a group of 16 SPADs: the probability to collect background events compared to the signal gets significantly lower when setting a threshold N_t of 2 and 4 photons, by a factor 10 and 10,000, respectively.

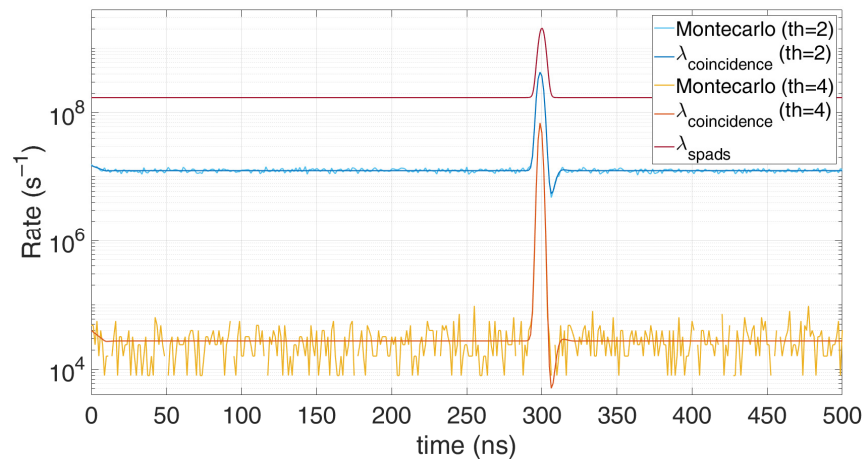


Figure 10. Comparison between Monte Carlo simulation and model for fixed-window photon coincidence, at two photon thresholds ($N_t = 2$ or 4), with 10 signal photons and 100 noise photons on average per repetition ($T_{HO} = 10$ ns). Note the effective rejection of the background baseline by a factor of about 10,000 in case of TOF measured only when at least 4 SPADs trigger concurrently.

3.3.2. Moving Coincidence Window

The second approach exploits a sliding coincidence window and can be implemented in asynchronous systems. Conversely to the fixed coincident case, which requires to sample the counting electronics at every time window step, this method can be implemented by shaping the SPAD digital pulse with a width equal to the desired coincidence time window interval. Then, all these digital pulses are summed in intensity and a coincidence occurs every time the total intensity crosses a given analog threshold, proportional to the desired photon threshold N_t .

The probability to have a coincidence at time index k is defined as the probability to have i detections in the previous coincidence window (indexes from $k - \Delta_{window}$ to $k - 1$) and the probability to have $N_t - i$ detections in the current time index k :

$$P(\text{COINC}(k)) = \sum_{i=0}^{N_t-1} P(N_{\text{DET}}(k - \Delta_{window}, k - 1) = i) \cdot P(N_{\text{DET}}(k) \geq N_t - i \mid N_{\text{DET}}(k - \Delta_{window}, k - 1) = i)$$

Such probabilities follow the binomial distribution:

$$P(\text{COINC}(k)) = \sum_{i=0}^{N_t-1} \left(B(N_{\text{spad}}, i, \sum_{j=k-\Delta_{window}-1}^{k-1} P(\text{DET}(j))) \cdot \sum_{m=N_t-i}^{N_{\text{spad}}} B(N_{\text{spad}} - i, m, P(\text{DET}(k))) \right)$$

If T_{step} is very short, the probability to have more than one detection among the N_{spad} at time index k is almost nil; therefore, in such a case the coincidence probability can be rewritten as the probability that one photon is detected by the group of SPAD at time k and $N_t - 1$ have been detected within the $\Delta_{window} - 1$ previous time indexes. We get:

$$P(\text{COINC}(k)) = B(N_{\text{spad}}, \sum_{j=k-\Delta_{window}-1}^{k-1} P(\text{DET}(j)), N_t - 1) \cdot (1 - (1 - P(\text{DET}(k))))^{N_{\text{spad}} - (N_t - 1)}$$

As for fixed coincidence window, the higher the threshold, the better the noise filtering, as shown in Figure 11. A high coincidence threshold lets just fewer measures to trigger the TDC and occupy a memory cell. However, too high threshold causes the probability to detect signal photons to lower as well.

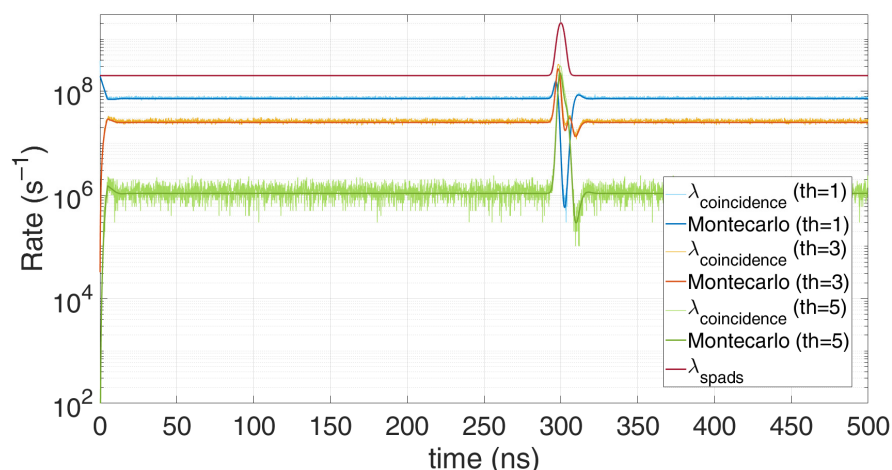


Figure 11. Comparison between Monte Carlo simulations and model for a moving photon coincidence window, at different thresholds, for 10 signal photons and 100 background photons on average per repetition ($T_{HO} = 10$ ns).

Figure 12 shows an example of coincidence photons acquisitions when background (62 photons on average per laser pulse repetition) overwhelms signal (only 0.6 photons on average per repetition). Only a sufficiently high photon coincidence threshold (e.g., 4 instead of 3 in the figure) and the availability of a sufficient number of registers to store the TOFs in each laser shot (e.g., 5 instead of 3 in the figure) allows to acquire a low-distorted histogram (line in purple color) with high signal level, with strong background rejection (about a factor 100) whilst limited signal loss.

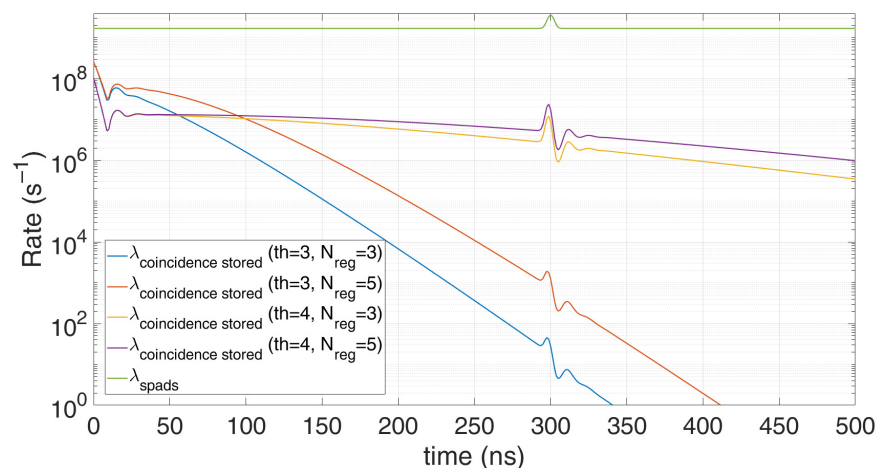


Figure 12. Example of modelling a limited number of TOF registers per laser pulse, with different photon threshold and number of registers, for 0.6 signal photons and 62 background photons on average per repetition ($T_{HO} = 10$ ns). Note how distortion increases for lower coincidence threshold and for lower TOF register availability per laser pulse repetition.

3.4. Time-Gated Detection

The possibility to turn on the detector within a defined enabling time window is a convenient way to avoid detecting and time-stamping useless photons. Such time-gating approach requires the knowledge of the time interval where the signal is expected, providing a GATE control signal to the detection electronics. In another approach, time gating can be exploited to perform repetitive measurements within a given gate window and then sliding it across the desired full-scale range of the acquisition.

The new model takes into account the saturation to N_{reg} measurements (see previous Section 3.2), to see the time-gate effect. The probability of not saturate the memory ($P(\overline{SAT}(k))$) is computed as in Equation (4). Let us define $GATE_{start}$ and $GATE_{stop}$ the time indexes of the starting (gate-on) and the ending (gate-off) time of the GATE, respectively.

The distribution describing $N_{TRIGGER}(k-1)$ events, i.e., photon coincidences measured before the window i , for $i \in [GATE_{start}, GATE_{stop}]$ becomes:

$$P(\overline{SAT}(k)) = \sum_{N_{coinc}=0}^{N_{sat}-1} \binom{n(k)}{N_{TRIGGER}} p(i)^{N_{TRIGGER}} \cdot (1-p(k))^{n(k)-N_{TRIGGER}},$$

$$P(MEAS(GATE_{start} \leq k \leq GATE_{stop})) = P(COINC_{N_t}(GATE_{start} \leq k \leq GATE_{stop})).$$

$$P(\overline{SAT}(k)) \cdot P(MEAS(k|GATE_{start} || k)GATE_{stop}) = 0$$

where $n(k) = k - GATE_{start}$ and $p(k) = \frac{\sum_{j \in [GATE_{start}, k-1]} P(COINC_{N_t}(j))}{n}$.

Figure 13 clearly shows the measurements improvement with time gating: much less distortion is achieved, since no memory storage gets wasted for background photons.

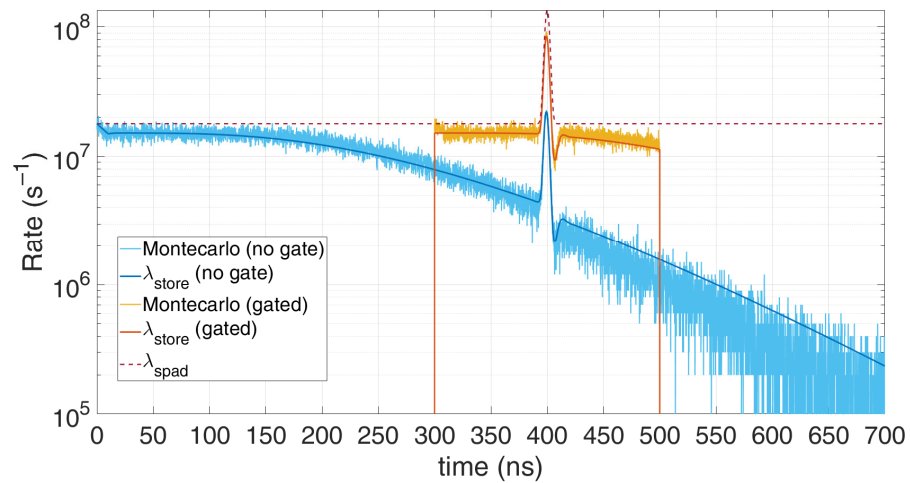


Figure 13. Comparison between Monte Carlo simulation and modelling for time-gated detection, in case of 7 registers and a single SPAD, with 0.6 signal photons and 12 background photons on average per repetition ($T_{HO} = 10$ ns). Note how time gating reduces distortion, since TOF memory registers are not wasted for background photons before the gate window.

3.5. Multi-TDCs Architectures

The last method relies on the possibility to measure the TOF through more TDCs per detector, in order to reduce the impact of the timing electronics dead-time, by routing the detector to the first available TDC [31]. A time stamp is possible if an event triggers the TDC and the number of TDC is not saturated (\overline{TDCSAT}). Again, for simplicity, the trigger probability is considered independent of what happened before, hence:

$$P(STAMP(k)) = P(TRIGGER(k) \cap \overline{TDCSAT}(k)) = P(TRIGGER(k)) \cdot P(\overline{TDCSAT}(k))$$

Similarly to the computation of $P(SAT)$ in Equation (4), the probability not to saturate the available number of TDCs (N_{TDC}) can be defined as:

$$P(\overline{TDCSAT}(k)) = P(N_{busy\ TDC}(k-1) < N_{TDC}) = \sum_{i=0}^{N_{TDC}-1} B(\Delta_{dead}, i, p(k))$$

with

$$p(k) = \frac{\sum_{j \in [k-\Delta_{dead}, k-1]} P(STAMP(j))}{\Delta_{dead}}$$

Of course, the higher the number of available TDC, the closest the measurement gets to the actual incoming rate. Figure 14 shows the excellent matching of the model with the Monte Carlo simulations, at different number of available multi-TDCs. The minor mismatches are due to the implemented simplification. It can be noticed the TDC dead-time effect, introducing a first pile-up until 150 ns and a second pile-up right after the laser pulse.

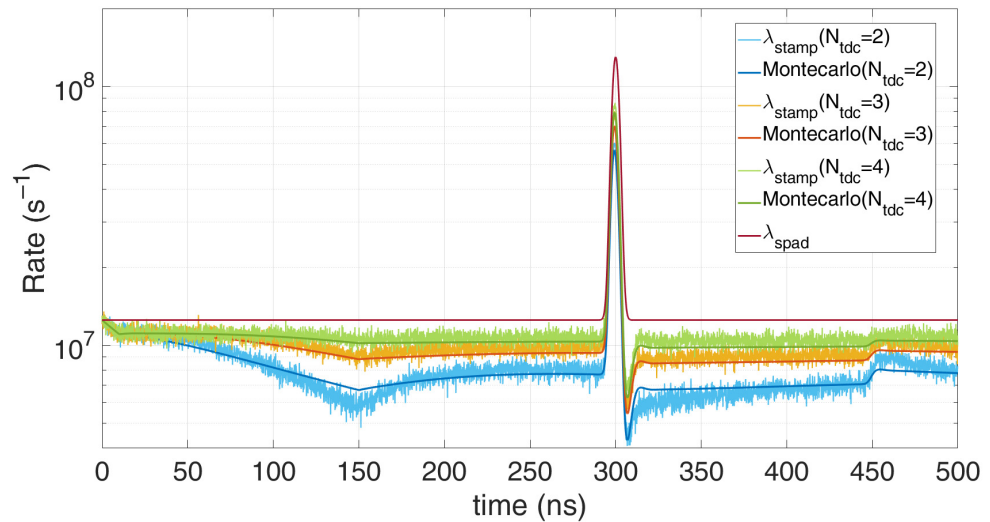


Figure 14. Comparison of Monte Carlo simulations and modelling for N_{tdc} multi-TDCs and single SPAD, with 0.6 signal photons and 6 background photons on average per repetition ($T_{HO} = 10$ ns, $T_{DEAD} = 150$ ns).

4. Distortion Correction

Up to now, we presented a detailed analytical modelling of different detector configuration (Section 2) and electronic acquisitions (Section 3). Now, we show how the developed statistical approach can be profitably used to predict the effect of non-idealities of SPAD-based detection systems. Furthermore, in some specific use-cases, the modelling can be exploited to correct the experimentally acquired histogram, featuring distortions introduced by the sensor, so as to extract a better estimation of the actual signal of interest. In the following subsections, we consider two typical situations.

4.1. TCSPC Correction

Time-Correlated Single Photon Counting (TCSPC) is usually applied with extremely faint photon sources, which provide less than one photon on average. Various works in literature deal with statistical ways to correct distorted TCSPC histograms in case of few photons, for example [27,33].

We propose another correction by starting from the modelling presented in Equation (2) and by computing an estimation of λ_{spad} . We start from the detection probability histogram $P(\text{DET}(k))$, given by the measured histogram divided by the number of pulses. An estimation $\hat{\lambda}_{spad}$ of the detection probability can be derived from the measured histogram, normalized to the number of pulses, by means of [34]:

$$\hat{\lambda}_{spad}(k) = -\ln\left(1 - \frac{h(k)}{N_{pulses} - \sum_{j=0}^{k-1} h(j)}\right)$$

Notice the correction (amplification) of the measured data, due to the denominator terms, so that the higher N_{pulses} gets, the better and less noisy the correction becomes, because the acquired histogram gets less noisy.

By implementing this correction, it is possible to break the TCSPC rule of thumb stating that the incoming rate should be less than 5% of the repetition frequency in order to

avoid excessive pile-up distortion due to the first photon masking all subsequent ones [29]. Unfortunately, such a constraint drastically increases the measurement time, since no photon is detected for at least 95% of the laser shot repetitions.

Concerning the model, if the incoming flux is too high, the probability to have no photon detections in a repetition tends to zero; therefore, for a given number of repetitions, $N_{pulses} - \sum_{j=0}^{k-1} h(j)$ vanishes and $\hat{\lambda}_{spad}$ diverges. Such a condition must be avoided, by limiting the incoming flux and by increasing the probability to have some repetitions with no photon arrivals, namely:

$$N_{pulses} \cdot p(0 \text{ photons}) = N_{pulses} \cdot \exp\left(\sum_{k=[1, N_{tot \text{ bins}}]} \lambda_{spad}(k) \cdot \Delta t_{bin}\right)$$

Thereby, in order to have more than k empty repetitions we must have $p(0 \text{ photons}) > \frac{k}{N_{pulses}}$.

We performed different simulations with different number of repetitions and average number of signal and background photons. For example, let us consider 0.1 signal photon and 1 background photon on average per laser shot repetition, so as to get $p(0 \text{ photons}) = 35\%$. Figure 15 shows that the higher the number of repetitions becomes, the less noisy the measured histogram gets and the better the correction results.

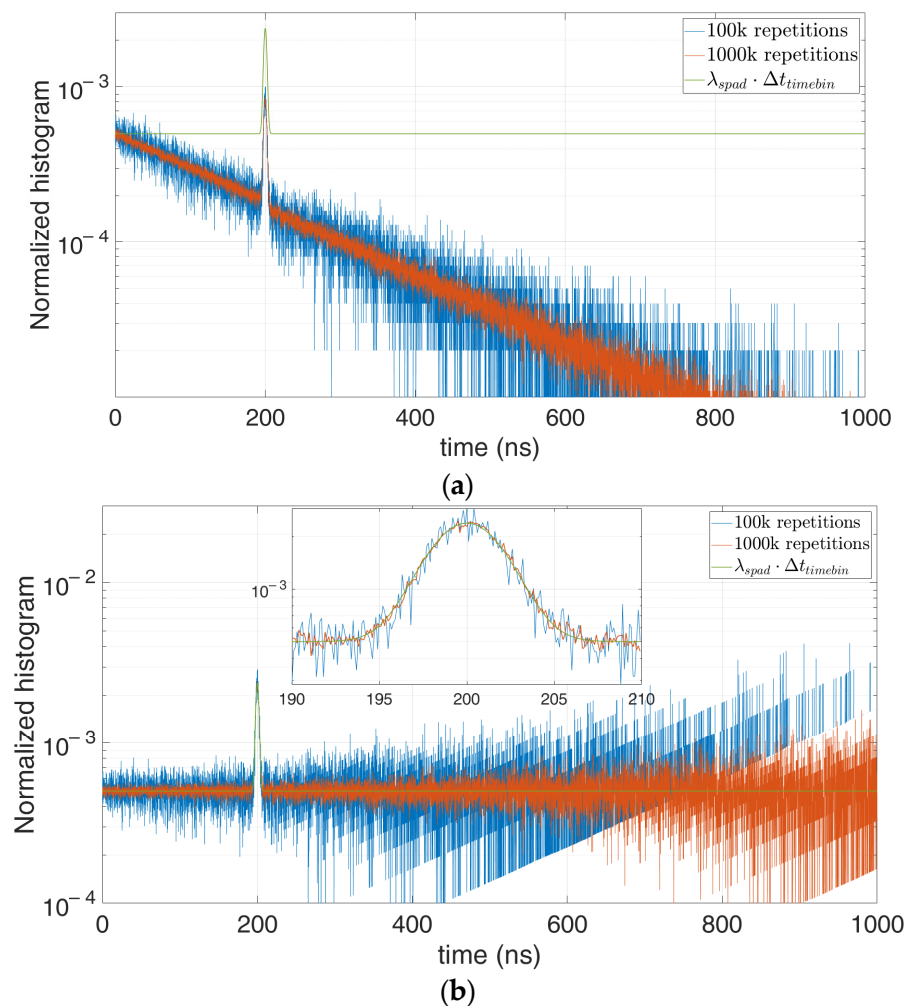


Figure 15. Distorted histogram in a single-photon TCSPC setup (a) and corrected estimation of the actual original signal (b), in case of 0.1 signal photon and 1 background photon on average per repetition, at 100,000 (blue) and 1,000,000 (red) repetitions.

Indeed, the number of k empty repetitions increases (from 35,000 to 350,000 in Figure 15, for the case with 100,000 and 1,000,000 repetitions, respectively) and the error at higher time bins decreases. We can conclude that thanks to the correction a lower $p(0 \text{ photons})$ can be allowed, giving the possibility to perform a measure with either higher photon flux or observation time (affording way less than 95% of $p(0 \text{ photons})$).

4.2. Multi-Photon Correction

The same approach can be applied to free-running applications with more photon detections within the observation time. The corresponding correction method can be employed both in high background applications and whenever a distortion due to fixed dead-time becomes severe. An example is a detection system employing a single SPAD with a hold-off time being the main limitation, or an architecture with a generic event that could trigger the TDC (a combination of SPADs) and the assumptions made in Equation (3) (trigger event independent of previous instants). In these cases, a good estimation is:

$$\hat{\lambda}_{spad}(k) = -\ln\left(1 - \frac{h(k)}{N_{pulses} - \sum_{j=k-\Delta_{dead}}^{k-1} h(j)}\right)$$

Figure 16b shows the correction of the distorted Figure 16a in case of a single SPAD feeding a single TDC, whose dead time is shorter than the SPAD hold-off time.

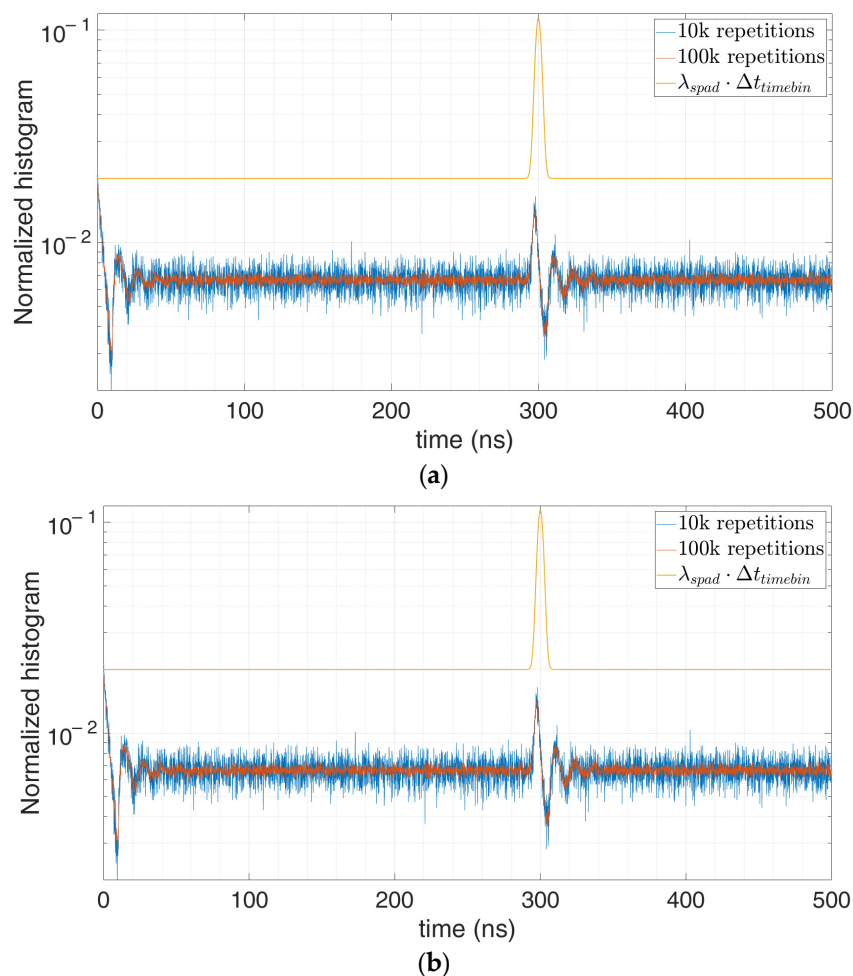


Figure 16. Distorted histogram in a multi-photon setup (a) and corrected estimation of the actual original signal (b) for single SPAD and single TDC architecture, in case of 5 signal photons and 100 background photons on average per repetition ($T_{HO} = 10 \text{ ns}$) at 100,000 (blue) and 1,000,000 (red) repetitions.

However, the same model can be used also when the timing electronics dead-time is longer and is the main limitation, by exploiting Equation (3). The more the trigger event is independent of previous instants, the more reliable the estimation becomes. An example could be when more SPADs feeding a combination logic drive one single TDC. Figure 17 shows the results in case of 100 SPADs in global OR triggering one TDC. In these cases, the more the SPADs the lower the dependency of the trigger event (OR among the SPADs).

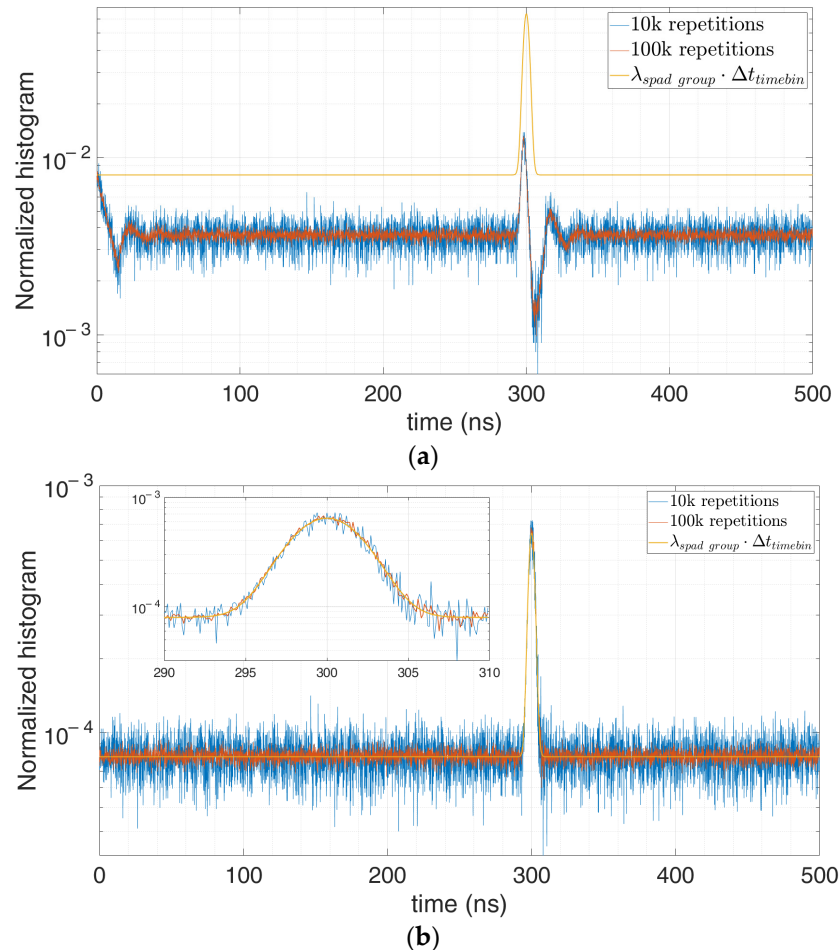


Figure 17. Distorted histogram in a multi-photon setup (a) and corrected estimation of the actual original signal (b) for 100 SPADs in a global OR combination, driving one TDC, with 3 signal photons and 40 background photons on average per repetition (on the cluster) ($T_{HO} = 10$ ns, $T_{DEAD} = 15$ ns), at 10,000 (blue) and 100,000 (red) repetitions.

Notice that any other model discussed in this work does not offer the advantage to be reversed in order to estimate the actual original signal. However, all of them can be implemented in iterative algorithms, starting from an estimation of the external conditions, computing the histogram through the proposed modelling, and then reducing the error between initial guess and computed histogram, by updating parameters such as background, reflectivity, and distance of the object at every cycle. Once the model converges, the λ_{SPAD} is eventually estimated. Such methodology can be used in photon coincidence application, taking advantage of the applicable model among those discussed in the previous sections, namely, fixed and moving coincidence, gating and multi-TDC architectures.

5. Conclusions

SPAD based system often require studies that demonstrates the signal and noise behavior impinging on the sensor. A statistical model has been introduced to study the signal and background noise and their effects on the SPAD. This analysis starts from

the incoming photon flux, which depends on the optical system parameters (i.e., lens aperture, emitted pulse width, Field-of-View, optical efficiency) and optical power, and introduces some of the most important non-idealities (i.e., as hold-off time, afterpulsing, optical crosstalk and pile-up). Some acquisition techniques are also discussed as photon coincidence, time gating and multi- TDC. Every model is compared to the relative Monte Carlo simulation to highlight its quality. Finally, the reverse form of the model can be exploited to correct the pile-up distortion and estimate the incoming photon flux.

Author Contributions: Conceptualization, A.I. and M.L.; Methodology, A.I. and M.L.; Resources, A.I. and F.Z.; Validation, A.I. and M.L.; Writing—original draft preparation, A.I. and M.L.; Writing—review and editing, A.I. and F.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not Applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wei, X.; Phung, S.L.; Bouzerdoum, A. Pedestrian sensing using time-of-flight range camera. In Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR 2011), Colorado Springs, CO, USA, 20–25 June 2011; pp. 43–48.
2. Kollorz, E.; Penne, J.; Hornegger, J.; Barke, A. Gesture recognition with a time-of-flight camera. *Int. J. Intell. Syst. Technol. Appl.* **2008**, *5*, 334–343. [[CrossRef](#)]
3. Cui, Y.; Schuon, S.; Chan, D.; Thrun, S.; Theobalt, C. 3D shape scanning with a time-of-flight camera. In Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR 2010), San Francisco, CA, USA, 13–18 June 2010; pp. 1173–1180.
4. Donati, S. *Electro-Optical Instrumentation: Sensing and Measuring with Lasers*; Pearson Education: London, UK, 2004.
5. Villa, F.; Severini, F.; Madonini, F.; Zappa, F. SPADs and SiPMs Arrays for Long-Range High-Speed Light Detection and Ranging (LiDAR). *Sensors* **2021**, *21*, 3839. [[CrossRef](#)]
6. BSI—British Standard. *BS EN 60825-1:2007—Safety of Laser Products. Equipment Classification and Requirements*; BSI: London, UK, 2007.
7. Villa, F.; Lussana, R.; Bronzi, D.; Tisa, S.; Tosi, A.; Zappa, F.; Dalla Mora, A.; Contini, D.; Durini, D.; Weyers, S.; et al. CMOS Imager With 1024 SPADs and TDCs for Single-Photon Timing and 3-D Time-of-Flight. *IEEE J. Sel. Top. Quantum Electron.* **2014**, *20*, 364–373. [[CrossRef](#)]
8. Buttafava, M.; Zeman, J.; Tosi, A.; Eliceiri, K.; Velten, A. Non-line-of-sight imaging using a time-gated single photon avalanche diode. *Opt. Express* **2015**, *23*, 20997–21011. [[CrossRef](#)] [[PubMed](#)]
9. Eshkoli, A.; Nemirovsky, Y. Characterization and Architecture of Monolithic N⁺P-CMOS-SiPM Array for ToF Measurements. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–9. [[CrossRef](#)]
10. Zappa, F.; Tosi, A.; Mora, A.D.; Tisa, S. SPICE modeling of single photon avalanche diodes. *Sens. Actuators A Phys.* **2009**, *153*, 197–204. [[CrossRef](#)]
11. Villa, F.; Zou, Y.; Mora, A.D.; Tosi, A.; Zappa, F. SPICE Electrical Models and Simulations of Silicon Photomultipliers. *IEEE Trans. Nucl. Sci.* **2015**, *62*, 1950–1960. [[CrossRef](#)]
12. Bronzi, D.; Zou, Y.; Villa, F.; Tisa, S.; Tosi, A.; Zappa, F. Automotive Three-Dimensional Vision Through a Single-Photon Counting SPAD Camera. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 782–785. [[CrossRef](#)]
13. Sesta, V.; Severini, F.; Villa, F.; Lussana, R.; Zappa, F.; Nakamuro, K.; Matsui, Y. Spot Tracking and TDC Sharing in SPAD Arrays for TOF LiDAR. *Sensors* **2021**, *21*, 2936. [[CrossRef](#)]
14. Kumagai, O.; Ohmachi, J.; Matsumura, M.; Yagi, S.; Tayu, K.; Amagawa, K.; Matsukawa, T.; Ozawa, O.; Hirono, D.; Shinozuka, Y.; et al. A 189 × 600 Back-Illuminated Stacked SPAD Direct Time-of-Flight Depth Sensor for Automotive LiDAR Systems. In Proceedings of the 2021 IEEE International Solid-State Circuits Conference (ISSCC), San Francisco, CA, USA, 13–22 February 2021; pp. 110–112.
15. Zappa, F.; Tisa, S.; Tosi, A.; Cova, S. Principles and features of single-photon avalanche diode arrays. *Sens. Actuators A Phys.* **2007**, *140*, 103–112. [[CrossRef](#)]
16. Tisa, S.; Zappa, F.; Tosi, A.; Cova, S. Electronics for single photon avalanche diode arrays. *Sens. Actuators A Phys.* **2007**, *140*, 113–122. [[CrossRef](#)]
17. Ceccarelli, F.; Acconcia, G.; Gulinatti, A.; Ghioni, M.; Rech, I.; Osellame, R. Recent Advances and Future Perspectives of Single-Photon Avalanche Diodes for Quantum Photonics Applications. *Adv. Quantum Technol.* **2020**. [[CrossRef](#)]
18. Sarbazi, E.; Safari, M.; Haas, H. Statistical Modeling of Single-Photon Avalanche Diode Receivers for Optical Wireless Communications. *IEEE Trans. Commun.* **2018**, *66*, 4043–4058. [[CrossRef](#)]
19. Padmanabhan, P.; Zhang, C.; Charbon, E. Modeling and Analysis of a Direct Time-of-Flight Sensor Architecture for LiDAR Applications. *Sensors* **2019**, *19*, 5464. [[CrossRef](#)]

20. Sanzaro, M.; Gattari, P.; Villa, F.; Tosi, A.; Croce, G.; Zappa, F. Single-Photon Avalanche Diodes in a 0.16 μm BCD Technology with Sharp Timing Response and Red-Enhanced Sensitivity. *IEEE J. Sel. Top. Quantum Electron.* **2018**, *24*, 1–9. [[CrossRef](#)]
21. Straka, I.; Grygar, J.; Hlousek, J.; Jezek, M. Counting Statistics of Actively Quenched SPADs under Continuous Illumination. *J. Light. Technol.* **2020**, *38*, 4765–4771. [[CrossRef](#)]
22. Xu, H.; Pancheri, L.; Braga, L.H.; Dalla Betta, G.F.; Stoppa, D. Crosstalk Characterization of Single-photon Avalanche Diode (SPAD) Arrays in CMOS 150nm Technology. *Procedia Eng.* **2014**, *87*, 1270–1273. [[CrossRef](#)]
23. Datta, R.; Heaster, T.M.; Sharick, J.T.; Gillette, A.A.; Skala, M.C. Fluorescence lifetime imaging microscopy: Fundamentals and advances in instrumentation, analysis, and applications. *J. Biomed. Opt.* **2020**, *25*, 071203. [[CrossRef](#)]
24. Scholkmann, F.; Kleiser, S.; Metz, A.J.; Zimmermann, R.; Pavia, J.M.; Wolf, U.; Wolf, M. A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. *NeuroImage* **2014**, *85*, 6–27. [[CrossRef](#)]
25. Tosi, A.; Zappa, F.; Cova, S.; Stellari, F. Luminescence measurements for the investigation of VLSI circuit defects. *Sens. Microsyst.* **2003**, 521–527. [[CrossRef](#)]
26. Li, Y.; Ibanez-Guzman, J. Lidar for Autonomous Driving: The principles, challenges, and trends for automotive lidar and perception systems. *IEEE Signal Process. Mag.* **2020**, *37*, 50–61. [[CrossRef](#)]
27. O’connor, D.; Philips, D. *Time-Correlated Single Photon Counting*; Academic Press: London, UK, 1984.
28. GmbH, P. Time-Correlated Single Photon Counting. 2014. Available online: http://www.picoquant.com/images/uploads/page/files/7253/technote_tcspsc.pdf (accessed on 30 March 2021).
29. Pediredla, A.K.; Sankaranarayanan, A.C.; Buttafava, M.; Tosi, A.; Veeraraghavan, A. Signal Processing Based Pile-up Compensation for Gated Single-Photon Avalanche Diodes. *arXiv* **2018**, arXiv:1806.07437.
30. Beer, M.; Haase, J.F.; Ruskowski, J.; Kokozinski, R. Background Light Rejection in SPAD-Based LiDAR Sensors by Adaptive Photon Coincidence Detection. *Sensors* **2018**, *18*, 4338. [[CrossRef](#)]
31. Arlt, J.; Tyndall, D.; Rae, B.R.; Li, D.D.U.; Richardson, J.A.; Henderson, R.K. A study of pile-up in integrated time-correlated single photon counting systems. *Rev. Sci. Instrum.* **2013**, *84*, 103105. [[CrossRef](#)] [[PubMed](#)]
32. Rapp, J.; Ma, Y.; Dawson, R.; Goyal, V. High-flux single-photon lidar. *Optica* **2021**, *8*, 30–39. [[CrossRef](#)]
33. Rapp, J.; Ma, Y.; Dawson, R.M.A.; Goyal, V.K. Dead Time Compensation for High-Flux Ranging. *IEEE Trans. Signal Process.* **2019**, *67*, 3471–3486. [[CrossRef](#)]
34. Coates, P.B. The correction for photonpile-up’ in the measurement of radiative lifetimes. *J. Phys. E Sci. Instrum.* **1968**, *1*, 878–879. [[CrossRef](#)]