PAPER

Measuring count rates free from correlated noise in digital silicon photomultipliers

IOP Publishing Measurement Science and Technology
Meas. Sci. Technol. **32** (2021) 025105 (11pp) https://doi.org/10.1088/1361-6501/abba4b

Measuring count rates free from correlated noise in digital silicon photomultipliers

Frédéric Vachon, Samuel Parent, Frédéric Nolet, Henri Dautet, Jean-François Pratte and Serge A Charlebois

Institut Interdisciplinaire d'Innovation Technologique (3IT), Université de Sherbrooke, 3000 Boulevard de l'Université, Sherbrooke, J1K OA5 Québec, Canada

E-mail: frederic.f.vachon@usherbrooke.ca

Received 23 May 2020, revised 9 September 2020
Accepted for publication 21 September 2020
Published 25 November 2020

Abstract

The characterization of nuisance parameters in digital silicon photomultipliers (SiPMs) is important to their understanding and future development. Methods able to distinguish the types of events are necessary to obtain fair and legitimate measurements. In this work, the zero photon probability (ZPP) method and the time delay (TD) method are used to measure the dark noise of digital SiPMs free from the contribution of correlated noise such as afterpulsing and crosstalk. It highlights the unique features of digital SiPMs such as the holdoff delay, the digital output signal, and the embedded processing (e.g. the selection of the interval sampling width). The two methods correctly separate the correlated and uncorrelated events in digital SiPMs and therefore the determination of a true photon detection efficiency (PDE) is possible. The ZPP method is also implemented inside a digital SiPM using embedded digital signal processing.

Keywords: silicon photomultiplier (SiPM), single photon avalanche diode (SPAD), characterization, dark noise, afterpulsing, time delay method, zero photon probability method

(Some figures may appear in colour only in the online journal)

1. Introduction

Silicon photomultipliers (SiPMs) are solid-state devices capable of single photon detection. They consist in a parallel array of hundreds of single photon avalanche diodes (SPADs) operated above breakdown voltage, each connected to its own quenching resistor. SiPMs are good alternatives to photomultiplier tubes for low-light applications such as positron emission tomography [1], high energy physics detectors [2], range-finding systems [3] and quantum optics [4]. They are made in either low-cost complementary metal-oxide semiconductor (CMOS) technologies or in dedicated foundries, are far more compact, operate at lower bias voltages and are insensitive to magnetic fields [5].

In recent years, the growing interest in SiPMs has prompted a lot of companies to manufacture them. Several characterization methods have been developed across the community without having an established measurement standard. The comparison from one SiPM to another therefore becomes non-trivial. This is especially true for digital SiPMs with embedded data processing and integrated functionalities where usual SiPM characterization methods do not always apply.

The measurement of the SiPM photon detection efficiency (PDE) illustrates this lack of standards. Several methods exist to measure the PDE: the photocurrent method [6], the zero photon probability (ZPP) method [7] and the time delay (TD) method [8]. They all determine the absolute PDE with good agreement, reliability and repeatability [9]. In each case, photogenerated events must be distinguished from correlated noise sources, such as afterpulsing and crosstalk, to extract a true PDE value. Although no real consensus has been reached as to which method should be favored, these methods apply mainly to analog SiPMs.

In contrast, digital SiPMs differ from conventional analog SiPMs in terms of output signal. The nature of digital circuitry

and the access to on-chip processing allow numerous architectures to be implemented and thus enable numerous ways to process and transfer the information to the user. A purely digital output can limit access to fundamental characteristics to fully understand the SiPM's behavior. Hence, special attention must be taken during the design of digital SiPMs in anticipation of the characterization needs.

In this paper, we elaborate on the relevance of the aforementioned methods to extract the uncorrelated event rate in digital SiPMs, namely the dark count rate or the photon count rate. In particular, we present an analysis of the Poisson process statistics applied to the digital SiPM's specific parameters such as the holdoff delay, the digital processing dead time, and the interval sampling width. The methods are first applied to a single SPAD inside a digital SiPM with an analog-like output. A similar method is then implemented inside a digital SiPM to measure count rates free from correlated noise, thereby demonstrating the benefits of the digital on-chip approach for SiPM characterization.

2. Features of digital SiPMs

SiPMs can be grouped into two families: analog or digital. They are nevertheless based on the same photosensitive component, the SPAD. Their difference lies in the way the avalanche-generated current is processed. In both cases, SPADs share a common supply voltage and are biased above breakdown voltage. Operating a SPAD in its non-linear regime takes advantage of the impact ionization property of silicon to initiate carrier avalanches, thus creating a very fast and large current pulse. After triggering, the SPAD is quenched and returned to its initial ready state.

In analog SiPMs, each SPAD is coupled to a passive quenching resistor. As illustrated in figure 1(a), the SPAD-resistor pairs are connected in parallel to a common reading node where the signals are collected before being amplified, shaped and digitized by external front-end electronics.

The analog SiPM architecture allows for an easy measurement of the breakdown voltage through current–voltage (I–V) characteristics. The signal amplitude can also be analyzed to determine the SiPM's gain and individual cell capacitance needed to achieve single photon resolution [10]. Further analysis of the signal shape provides useful insights into the behavior of SiPMs at cryogenic temperatures [11], the presence of defects in the crystal lattice and the hardness to radiation damage [12, 13]. In this way, the analysis of I–V curves and pulse shapes provides the basic characterization methods of analog SiPMs as standalone components.

State-of-the-art analog SiPMs such as those outlined in [10] achieve a PDE above 50% in the visible spectrum, have dark noise rates lower than 50 kHz mm^{-2}, and achieve a single photon timing resolution (SPTR) of 70 ps FWHM [14].

In digital SiPMs, each SPAD is individually coupled to an internal CMOS quenching circuit (QC). This circuit detects the early formation of an avalanche, quenches the avalanche and, in due time, starts the recharge cycle (figure 1(b)). The digital SiPM can either quench the avalanche passively with a transistor maintained in a high resistance state or actively by forcing the high-voltage bias below breakdown [15]. The time a SPAD is held below breakdown is defined as the holdoff delay (t_h).

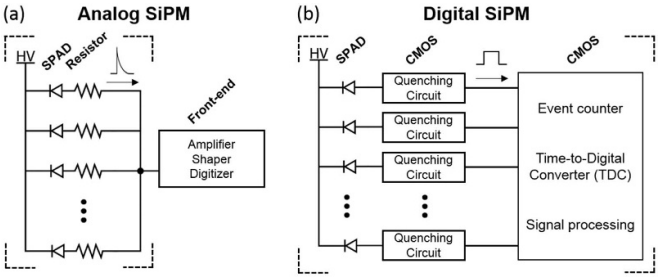

Figure 1. Representation of (a) an analog SiPM and (b) a digital SiPM. The dashed line corners mark out the components that form the integrated circuit.

CMOS QCs help in achieving optimal timing characteristics and minimal detection dead time. In advanced digital SiPM architectures, QC outputs can be read individually, summed or further processed with time-to-digital converters (TDCs) and event counters [16–19]. Depending on the application requirements, other utility circuits can be implemented to enhance the SiPM performances and characterization possibilities [20].

The following sections present two digital SiPM architectures developed by our group with which a count rate free from correlated events can be extracted.

2.1. The analog monitor architecture

The analog monitor architecture provides an output signal similar to that of an analog SiPM. When a QC senses a SPAD avalanche, a current source generates a pulse of configurable amplitude and duration.

As shown in figure 2, all the SPAD readout circuits are connected to a common reading node performing the analog sum of the currents. The controller can enable or disable any SPADs of the array in order to study a SPAD individual characteristics. The output capacitance of the analog monitor architecture matches that of the sum of the CMOS transistor drains which is much lower than the SPADs' own capacitance. The benefits of this architecture include a low output capacitance (5.6 pF mm^{-2}), a high dynamic range (20 dB) and a high amplitude uniformity across all SPADs (peak-to-valley ratio > 100) [21].

2.2. Digital SiPM architecture with embedded digital signal processing

In the design of digital SiPMs, more complex architectures exist where QCs are connected to individual TDCs or event counters (figure 3). It allows for advanced digital signal processing of SPAD events inside the application-specific integrated circuit (ASIC) and for digital transmission of data. Digital communication is less prone to data losses and signal degradation compared to an analog signal transmission. Furthermore, the early digitization of the signal achieves high timing optimization such as sub-10 ps FWHM SPTR at the

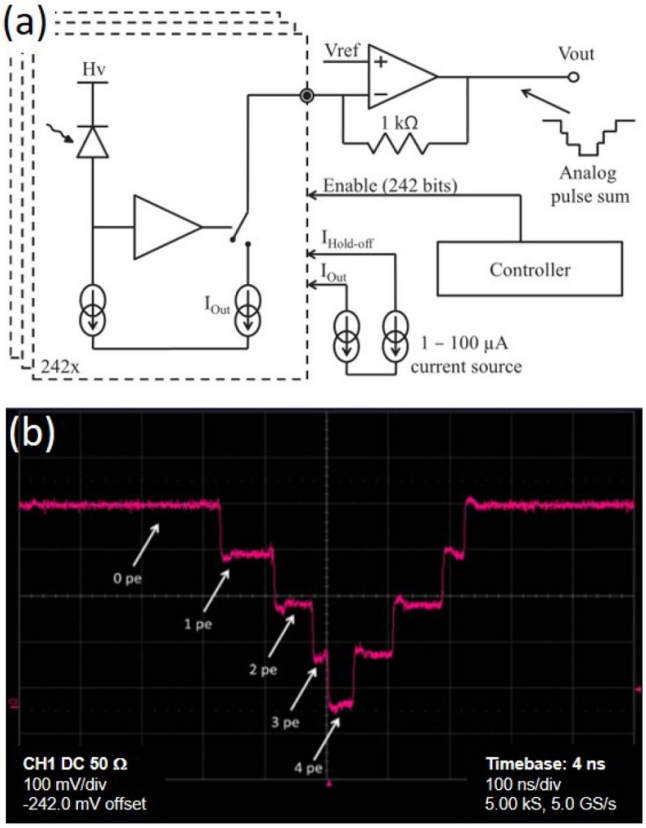

Figure 2. (a) Top-level block diagram of the analog monitor architecture. (b) Output signal showing the analog sum of four different SPADs triggering successively. The signal falling edges correspond to QC triggers. © 2020 IEEE. Reprinted, with permission, from [21].

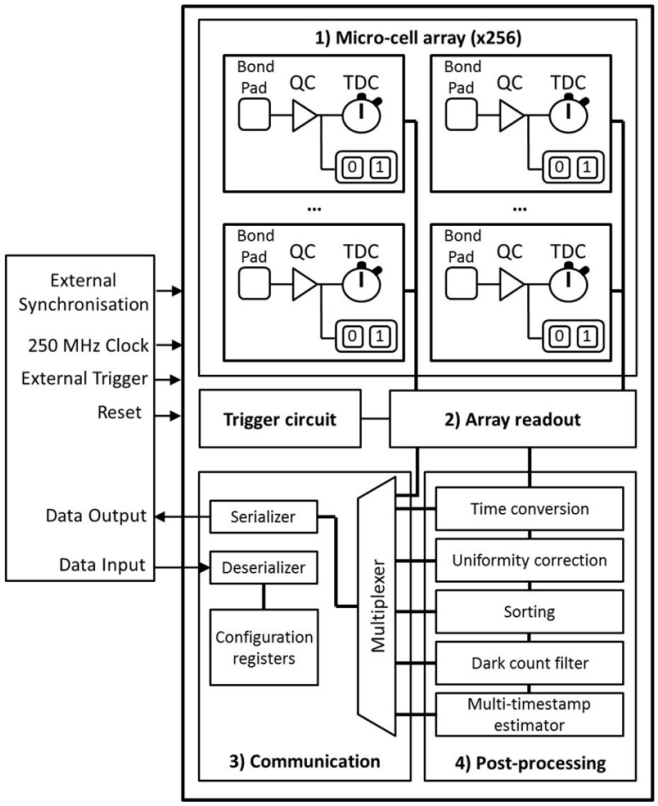

Figure 3. The architecture scheme of a digital SiPM developed for medical imaging. Individual SPADs are intended to be integrated in 3D and connected to the bond pad. Reprinted from [23], with permission from Elsevier.

SPAD level [22] and less than 20 ps RMS array timing jitter [23].

3. PDE measurement

The PDE describes the ability of a SiPM to produce a readable electric signal from an incident light beam. It is a measurement of the sensitivity to photons of a given wavelength while the SiPM is biased at a specific voltage (V_{ov}) over its breakdown voltage (V_{BD}). Contributions to the PDE are [24, 25]

$$PDE \sim QE(\lambda,\theta) \times \mathcal{P}_{BD}(E) \times FF, \tag{1}$$

where QE is the global quantum efficiency and depends on the wavelength (λ) and the angle of incidence (θ) of the light beam. $\mathcal{P}_{BD}$ is the breakdown initiation probability. It is related to the electric field across the avalanche region (E), and thus V_{ov}. FF is the fill factor, the ratio of photosensitive area to the total device area.

Experimentally, the PDE is defined as the ratio between the number of detected photons (N_{ph}) and the number of photons impinging on the detector (N_{ref}). The discrepancy between the two quantities can be attributed to photon losses in non-photosensitive areas (FF), photon absorption outside the depletion region (QE), or photons that did not initiate a sustainable avalanche ($\mathcal{P}_{BD}$). Due to the SiPM's intrinsic dark noise (N_{dn}), the PDE has to be corrected as follows:

$$PDE = \frac{N_{ph+dn} - N_{dn}}{N_{ref}}. \tag{2}$$

In addition to dark noise and detected photons, called herein *uncorrelated events*, SPAD-based detectors generate other minority carriers that also produce readable signals. They are called *correlated events* and are commonly referred to afterpulses and optical crosstalk [26]. In other words, one photon may result in more than one measured event during the characterization of the PDE. Using only the total number of events produced will inevitably result in its overestimation. The challenge lies in the separation of correlated and uncorrelated events.

Methods described in the literature [6–9, 27–29] are based on the assumption that the generation of thermal carriers and incident photons are Poisson point processes; providing the opportunity to be distinguished from afterpulsing and optical crosstalk events that are always correlated to a primary event.

A description of the TD method (section 3.1) and the ZPP method (section 3.2) is given below. Specific parameters for the characterization of the digital SiPM are also introduced.

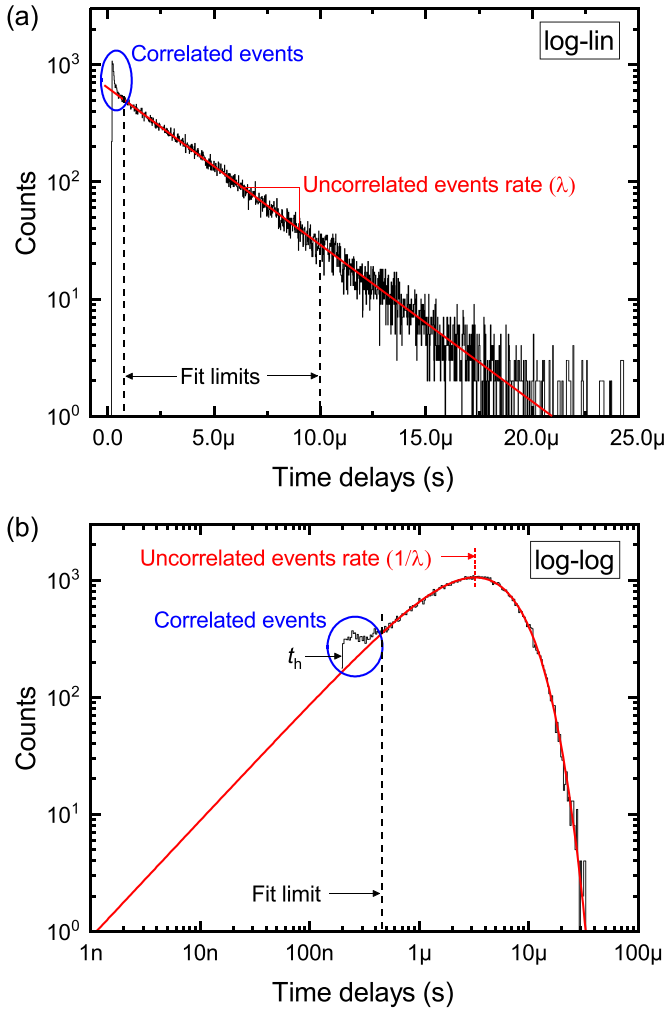

Figure 4. (a) Log-linear and (b) log–log representations of the same time delay distribution acquired with the analog monitor architecture. The log–log representation gives more emphasis to shorter time delays and allows one to visualize the structure of the correlated events time distribution.

3.1. The TD method

The TD method relies on the statistics of delays between consecutive events that are not correlated. The absence of memory between two events is described by the Poisson process theory and results in an exponential distribution. Even in the presence of a dead time (t_h), the memoryless property is preserved and the probability density function (PDF) takes the form

$$\mathcal{P}(t) = \lambda e^{-\lambda(t-t_h)} \tag{3}$$

where λ is the average rate of uncorrelated events. The variable t_h normalizes the PDF to unity considering that no events can be measured during the holdoff delay.

Figure 4 introduces two representations of the same time delay distribution acquired with a single SPAD. The correlated events depicted are therefore only afterpulsing and not optical crosstalk.

In the log-linear representation (figure 4(a)), the exponential behavior of time delays gives a linearly decreasing distribution with its slope being λ. At short time delays, counts that do not follow the exponential distribution are associated with correlated events.

The extraction of the uncorrelated event rate is done by applying two fit limits to the regression analysis where counts depart from the expected exponential behavior. The limit at large time delays rejects bins with too few counts (e.g. <10 counts). The lower limit is more difficult to define. At the moment, it is set qualitatively well within the linear region in order to exclude any afterpulsing events. Future works will implement a robust fitting method.

Another way of representing the exponential distribution is by sorting time delays in logarithmically spaced bins (figure 4(b)). In this representation, the exponential distribution has a maximum at $1/\lambda$ and the fit function takes the form

$$f(t) = N_\lambda \cdot \ln(10) t \cdot \delta x \lambda e^{-\lambda(t-t_h)} \tag{4}$$

where the x-axis has been logarithmically transformed through a change of variable $x = \log_{10}(t)$ [30, 31].

N_λ is the area under the fit curve and corresponds to the total number of uncorrelated events:

$$N_\lambda = N_\Sigma \cdot (1 - f_{AP}) \tag{5}$$

where f_{AP} is the fraction of afterpulsing to the total number of events N_Σ. δx defines the granularity of the log x-axis. In figure 4(b), t_h is visible as a truncation of the exponential distribution.

The PDE can be measured with the TD method by acquiring two sets of time delay distributions. In dark conditions, to extract the uncorrelated dark noise rate λ_{dn}, and in light conditions to extract the photon rate λ_{ph+dn} at any given wavelength. The PDE is calculated from equation (2) with the reference photon rate λ_{ref} measured separately.

3.2. The ZPP method

Similar to the TD method, the goal of the ZPP method is to provide a proper estimation of λ isolated from correlated events. It uses the statistics of events occurring during fixed time intervals instead of time delays. That is, in a given time interval of width Δt, the average number of uncorrelated events μ occurring at a rate λ is

$$\mu = \lambda \Delta t. \tag{6}$$

The probability of observing an interval containing k events is given by the Poisson probability mass function (PMF):

$$\mathcal{P}(k) = \frac{\mu^k e^{-\mu}}{k!}. \tag{7}$$

The probability of measuring 0 events during a given interval is independent of correlated events because no events are measured. From equation (7), the probability becomes

$$\mathcal{P}(0) = e^{-\mu}. \tag{8}$$

The case $k = 0$ relates experimentally to the number of intervals without events (N_0) and the total number of intervals (N_{total}) as

$$N_0 = N_{\text{total}} \cdot \mathcal{P}(0). \tag{9}$$

Assuming that dark noise (μ_{dn}) and incident photons (μ_{ph}) are Poisson point processes, the average photon detected per interval can be written as

$$\mu_{\text{ph}} = \mu_{\text{ph+dn}} - \mu_{\text{dn}} \tag{10}$$

$$= -\ln\left(\frac{N_0^{\text{ph+dn}}}{N_{\text{total}}^{\text{ph+dn}}}\right) + \ln\left(\frac{N_0^{\text{dn}}}{N_{\text{total}}^{\text{dn}}}\right) \tag{11}$$

where the contribution of dark noise to detected photons is removed.

To calculate the PDE, the incident photon flux μ_{ref} is to be measured from an independent calibrated reference photodetector. From a statistical point of view, each photon undergoes a Bernoulli trial in the SiPM, that is, the probability of detecting a photon is either 0 or 1. Considering that the incident photons follow a Poisson process, its mixture with a Bernoulli trial simplifies to a Poisson distribution of argument $\mu_{\text{ref}} \times PDE$ and therefore

$$PDE = \frac{\mu_{\text{ph}}}{\mu_{\text{ref}}}. \tag{12}$$

4. Setup and experimental procedures

The analog monitor architecture and the digital SiPM with embedded digital signal processing are characterized using the same test bench. They are mounted and wirebonded on dedicated daughterboards [21, 23]. They are both connected to their respective motherboards which provide the necessary supply voltages. All SPADs from a single SiPM share a common operating voltage and can be enabled or disabled on demand.

4.1. Offline procedures for the digital SiPM with analog monitor architecture

This section presents an offline implementation of the TD method (section 3.1) and the ZPP method (section 3.2) acquired with a single SPAD using the analog monitor architecture (section 2.1).

Events from the SPAD are acquired with a frequency counter (Keysight 53 230A [32]). It reads the output of the analog monitor and measures time delays between consecutive events. The frequency counter triggers on rising edges. Its threshold level is set to half of the pulse amplitude (500 mV) where the slope of the signal trace is the steepest and free from electrical noise. It ensures an accurate triggering time from one pulse to another. The width of each pulse (corresponding to t_{h}) is kept constant for each pulse of the same experiment.

The time delays read by the frequency counter can be readily sorted in histograms and analyzed with the TD method (section 3.1).

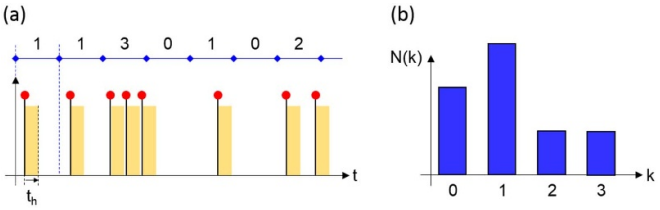

Figure 5. (a) Timestamps waveform containing time of arrival of events generated by the SPAD. Intervals of fixed width sample the waveform. For a single SPAD, no events can occur during the holdoff delay (t_{h}). (b) The N(k) histogram is constructed from the number of events contained in each interval.

Alternatively, the offline ZPP method identifies the time of arrival of each event. The resulting data file is a sequence of timestamps where each timestamp corresponds to the rising edge of a pulse. The timestamp waveform is then divided in N_{total} uniform and sequential time intervals as in figure 5(a). The number of intervals containing k events forms the N(k) histogram of figure 5(b).

The average number of events per interval (μ) is calculated from the number of intervals containing 0 events (N_0) and the total number of intervals acquired (N_{total}). The uncorrelated event rate (λ) is computed with equation (6) where Δt is the chosen interval width. The contribution of afterpulses (f_{AP}) for a single SPAD is calculated from λ and the acquisition's total count rate (N_Σ/T_Σ) as:

$$f_{\text{AP}} = 1 - \lambda\frac{T_\Sigma}{N_\Sigma} \tag{13}$$

where T_Σ is the acquisition's total duration.

In the literature [7, 9, 10, 27, 29], the ZPP method is generally used in pulsed mode. Photon events are counted from intervals of fixed width following a LED flash with limited repetition rate. Dark noise is measured in a similar fashion with intervals positioned before the expected LED flash. However, the selection criteria and the effect of the interval width on the extracted uncorrelated event rate are not thoroughly studied nor explained.

In this manuscript, the photon rate is measured with a continuous light source. Unlike the pulsed mode, the intervals are contiguous. The dark noise rate is measured from a second acquisition where the light source is turned off.

For the ZPP method to be valid, in pulsed mode or continuous mode, the light source must be Poissonian and its intensity must be low enough to avoid SiPM saturation. As shown in [27] for the pulsed mode and in the present manuscript for the continuous mode, the interval width also needs to be chosen correctly.

The offline procedure presented above allows to study the effect of the interval width on the extraction of the uncorrelated event rate. The selection of the interval width is presented in section 5.2 using the analog monitor architecture, and in section 6 using the digital SiPM with embedded signal processing.

4.2. Implementation of the ZPP method into the digital SiPM with embedded signal processing

In order to take full benefit of the digital processing capabilities of digital SiPMs (section 2.2), the ZPP method is implemented on-chip using a time-driven event counter.

Similar to the procedure described in section 4.1, an internal counter measures the number of events k for each interval and each SPAD. The ratio between the 0-event bin (N_0) and the total number of intervals recorded (N_{total}) gives the uncorrelated event rate of each individual SPAD. The uncorrelated event rate of the whole SiPM is measured either by summing all events of intervals recorded at the same time or by summing the uncorrelated event rate of all SPADs.

In this specific architecture, data transmission is done by serial communication and adds a 1 µs dead time after each interval. The time during which the counter records events is set by the sampling interval width and ranges from 4 ns to 4 µs. The holdoff delay is the same for all SPADs and can be configured from 5 ns to 100 ns.

5. Digital SiPM with analog monitor architecture and offline procedures

In section 5, the uncorrelated event rate of the digital SiPM with the analog monitor architecture is estimated with the TD method (section 5.1) and the ZPP method (section 5.2). The effect of temperature and the holdoff delay are analyzed using a single square shaped SPAD (36× 36 µm^2) with $V_{\text{BD}} = 22.1$ V at room temperature (22 °C–24 °C). More details about the SPAD technology can be found in [33].

The extraction of the uncorrelated event rate with the ZPP method and the digital SiPM architecture with embedded digital signal processing is presented in section 6.

5.1. TD method

In order to study distinct time delay distributions, the SPAD is measured at three different temperatures (300 K, 167 K and 87 K) in dark conditions. As the temperature varies, V_{ov} is maintained at 4 V by adjusting the applied bias according to the breakdown voltage temperature coefficient of ∼0.1%/K. The uncorrelated event rates λ and afterpulsing contribution f_{AP} are extracted from equations (4) and (5) respectively. The holdoff delay t_{h} is 200 ns. Fit limits, represented by arrows on the x-axis, are placed qualitatively at the inflection point where counts depart from the exponential distribution.

In figure 6, the uncorrelated event rate of the SPAD decreases as a function of temperature, ranging from 1048 ± 6 s^{-1} at 300 K down to 2.29 ± 0.06 s^{-1} at 87 K. This leads to the distribution peaks shifting toward higher time delays and separating from the afterpulsing contribution.

The TD method also highlights the distribution of the traps lifetime in silicon. At 300 K, afterpulsing events represent 4.3 ± 0.4% of the SPAD's total noise where its time delay distribution extends to ∼30 µs and overlaps that of the uncorrelated dark events. At 87 K, thermal noise is reduced significantly but afterpulsing events represent 34 ± 2% of the total noise.

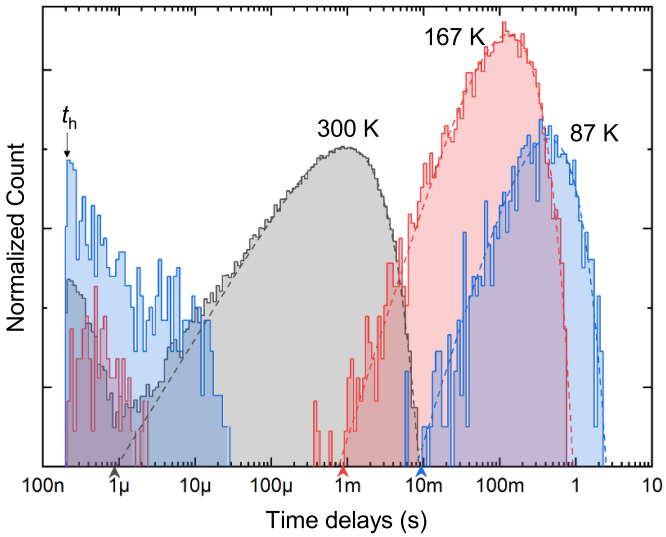

Figure 6. The effect of temperature on the time delay distributions. The fit function from equation (4) is represented by the dashed lines. The extracted uncorrelated event rate (λ) and afterpulsing (f_{AP}) are shown in table 1. Arrows on the x-axis define the fit limits.

Despite this large contribution to noise, the uncorrelated event rate can still be adequately extracted because the two distributions are well defined.

5.1.1. Effect of the holdoff delay on the TD method. Digital SiPMs offer a unique way to mitigate the contribution of afterpulsing by controlling the time during which a SPAD is held off. Depending on the duration of the holdoff delay (t_{h}), an event generated shortly after a previous one will not be detected because the SPAD is kept below its breakdown voltage.

To showcase this effect, a single SPAD is biased at $V_{\text{ov}} =$ 4 V while t_{h} varies. A source of light is turned on to reach an event rate of the order of thousands of counts per second.

Figure 7 shows the suppression of afterpulsing events as larger holdoff delays are set. At $t_{\text{h}} = 158$ ns, the contribution of afterpulsing to the total noise is almost zero (figure 11(c)).

As long as the holdoff delay is kept below the global maximum of the exponential distribution, it is possible for the TD method to extract λ and f_{AP}. As a general rule for digital SiPMs, the holdoff delay is chosen to be much shorter than the average time between two events ($t_{\text{h}} < \frac{1}{\lambda}$).

As shown in figure 6 for 167 K and 87 K, a holdoff delay as long as 100 µs can shadow the contribution of afterpulsing without affecting the uncorrelated events. However, when the two contributions overlap (300 K), a holdoff delay larger than 1 µs will suppress part of the uncorrelated event distribution. The optimal value of the holdoff delay is to be chosen considering the specific requirements of an application.

5.2. ZPP method

Figure 8 illustrates the effect of the interval width (Δt) on N(k) histograms computed from a single data set. For a small Δt, intervals contain mostly zero-photon events. As the Δt widens,

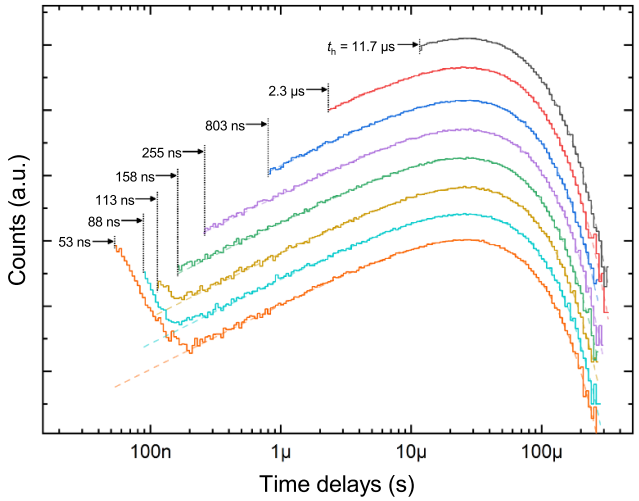

Figure 7. The capability of the digital SiPM to control the holdoff delay (t_h). The time delay distributions are offset to highlight the effect of the holdoff delay. Dashed lines correspond to the fit of uncorrelated events (equation (4)).

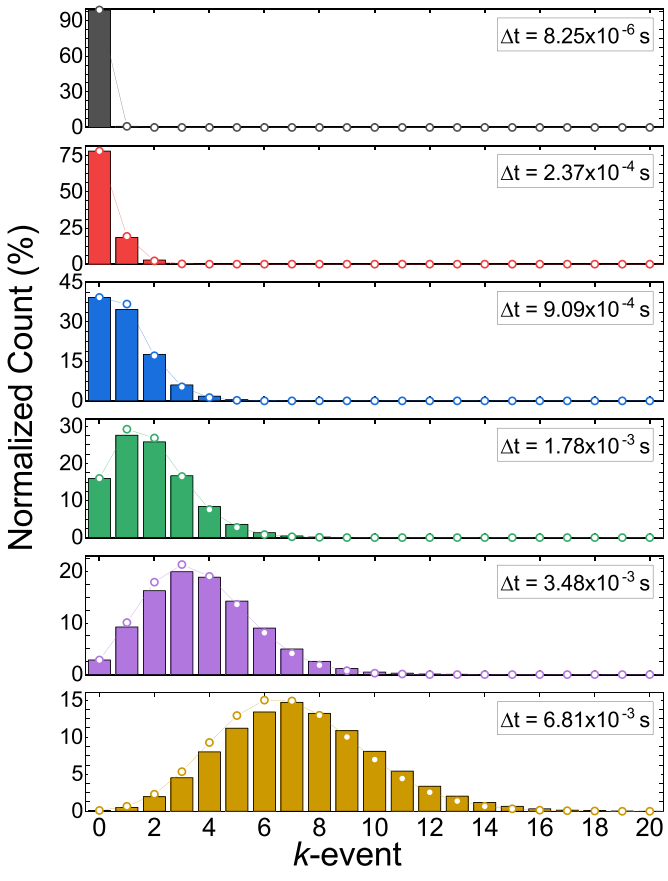

Figure 8. N(k) histograms computed using various interval widths (bars). Superimposed are the Poisson probability mass functions for corresponding µ values (open circles).

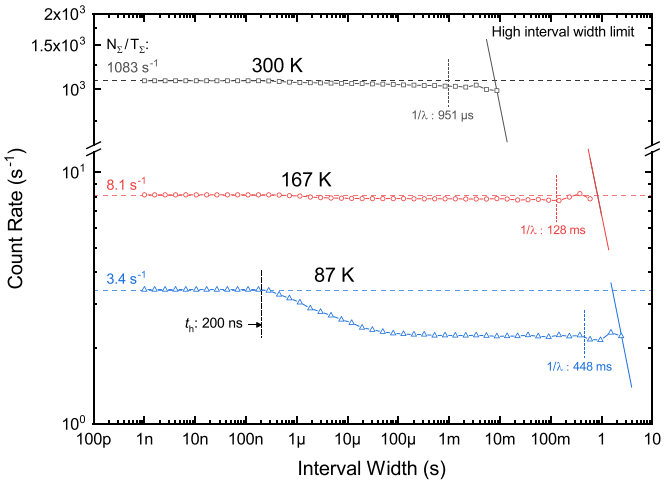

Figure 9. The uncorrelated event rates as a function of the interval width using the ZPP method. The total count rate (N_Σ/T_Σ) of each acquisition is shown as a dashed line.

more events are included per interval. This shifts the mode of the distribution to higher k-indices.

Also shown in figure 8 are the theoretical Poisson PMFs (equation (7)) obtained with the corresponding µ values. The pure Poisson distributions exhibit slight differences with respect to the experimental values that are affected by correlated events.

The construction of N(k) histograms with digital SiPMs provides very distinct bins. Even with the analog monitor, there is no ambiguity in defining how many events occurred in each interval due to the pulses' steep edges and constant amplitude. The determination of N_0 and N_{total} is therefore implicit.

With analog SiPMs, similar approaches use the pedestal peak of the charge spectrum to measure N_0 [29]. Others record 0-event intervals when the noise baseline is measured [10]. The characterization of analog SiPMs with this method requires a notably good adjustment of the signal amplifier because individual SPADs rarely generate the same amount of charges per event. The 0-event peak must be separated from higher order peaks, otherwise leading to increased variability in the determination of a true uncorrelated event rate.

An important parameter in extracting the uncorrelated event rate with the ZPP method is the choice of the interval width (Δt). As shown in figure 8, the average number of events per interval (µ) changes with Δt. However, when Δt is chosen between the limits described below, values of λ and f_{AP} are consistent with those obtained using the TD method (section 5.3).

The data sets previously analyzed with the TD method (figure 6) are now processed with the ZPP method. The resulting uncorrelated event rates are presented in figure 9 as a function of the interval width. Values of λ and f_{AP} are reported in table 1 and correspond to an average value calculated on the plateau region. The plateau region lies between two limits whose origins are presented in the following.

High interval width limit. The uncorrelated event rate has an upper limit at large Δt when each interval contains at least one event ($N_0 = 0$). The value at which this occurs depends on the event rate (λ) and the total number of samples acquired (N_{total}).

Table 1. Comparison of the uncorrelated count rate (λ) and afterpulsing (f_{AP}) extracted with the TD method and the ZPP method for three temperatures.

Parameters	λ (s^{-1})		f_{AP} (%)	
Methods	TD	ZPP	TD	ZPP
300 K	1048 ± 6	1052 ± 11	4.3 ± 0.4	3 ± 1
167 K	7.7 ± 0.1	7.83 ± 0.04	4.7 ± 0.9	3.7 ± 0.5
87 K	2.29 ± 0.06	2.24 ± 0.01	34 ± 2	34.4 ± 0.3

Following [34, section 7.4.1.] and setting a confidence level of 95%, the high limit is given by

$$\lambda \Delta t > \ln\left(\frac{N_{\text{total}}}{3}\right). \tag{14}$$

It is represented in figure 9 by the oblique lines on the right-hand side of the graph.

Also seen on figure 9 is the count rate rolling off near the high limit. It starts to move away from the constant plateau region when the interval width becomes greater than the average uncorrelated event rate. That is, the probability of finding empty intervals rapidly reduces until $N_0 = 0$. Practically, the plateau region high limit rather becomes

$$\Delta t > 1/\lambda. \tag{15}$$

Low interval width limit. The uncorrelated event rate has a lower limit at small Δt when all intervals contain no events ($N_0 = N_{\text{total}}$). It is defined by the following inequality [34]:

$$\lambda \Delta t < -\ln\left(1 - \frac{3}{N_{\text{total}}}\right). \tag{16}$$

However, in the case of the ZPP method depicted in this section, N_0 never reaches N_{total} because the intervals are always contiguous in time and they sample the whole waveform. N_0 will be at most equal to $(N_{\text{total}} - N_{\text{event}})$. This happens when the Δt is smaller than t_h as the waveform is completely discretized and each event is contained in an independent interval. At that moment, the extracted λ values correspond to the acquisition's total count rate (N_Σ/T_Σ), that is, not free from the contribution of correlated events.

The plateau region is rather reached when the majority of afterpulsing events are included in an interval containing at least one primary event. Indeed, from the time delay distributions in figure 6, the afterpulsing lifetime extends up to ~30 µs. Consequently, at a similar interval width, the plateau region is reached in figure 9 using the ZPP method. The count rate departing from the plateau region toward N_Σ/T_Σ at small Δt is referred to as *correlated event mixing*.

5.2.1. Effect of the holdoff delay on the ZPP method. The data set previously analyzed with the TD method (figure 7) is now processed with the ZPP method and is shown in figure 10. The extracted λ and f_{AP} are shown in figure 11.

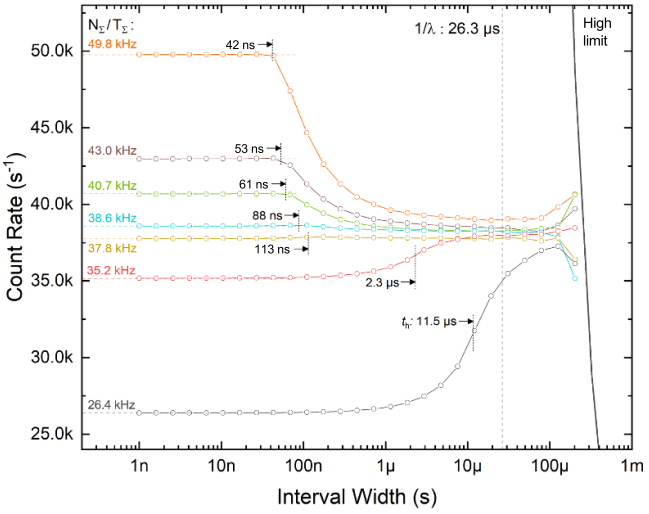

Figure 10. Effect of the holdoff delay on the count rate as a function of interval width using the ZPP method. The constant plateau region is affected by correlated event mixing for $t_h < 88$ ns and by the holdoff delay damping for $t_h > 2.3$ µs.

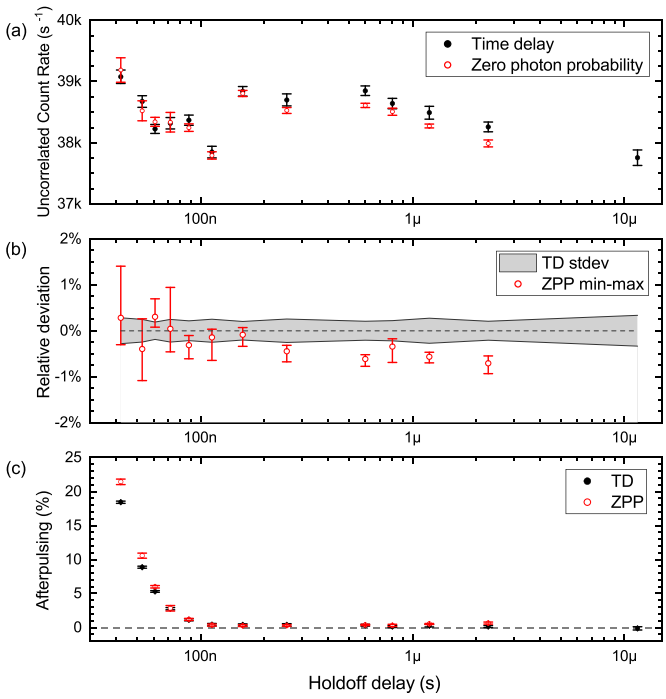

Figure 11. Comparison of the uncorrelated count rate (λ) and afterpulsing (f_{AP}) extracted with the TD method and the ZPP method as a function of the holdoff delay (t_h).

For small holdoff delays ($t_h < 88$ ns), the count rate departs from the plateau region at small interval widths due to correlated event mixing. A higher contribution of afterpulses inherently limits the plateau region to high interval widths.

For $t_h \sim 113$ ns, the plateau region spans over the entire interval widths range. The acquisition's total count rate (N_Σ/T_Σ) matches with the uncorrelated count rate (λ) as afterpulsing events are mostly shadowed by the holdoff delays.

For large holdoff delays ($t_h > 2.3$ µs), the count rate is damped to N_Σ/T_Σ at small interval widths. The premature masking of intervals by a holdoff delay of similar width artificially increases the number of intervals containing zero events (N_0) and thus lowers the count rate. As Δt is made larger, the holdoff delay following each event becomes less significant and the plateau region is eventually reached. Thus, true values of λ are achieved only when the interval width is larger than the holdoff delay ($\Delta t > t_h$).

5.3. Comparison of the TD method and the ZPP method

Table 1 compares λ and f_{AP} obtained with the TD method and the ZPP method from figures 6 and 9 respectively. Both methods lead to compatible results within their respective error bars. The differences highlight the uncertainty in the definition of an accurate fit limit for the TD method and a precise plateau region for the ZPP method.

A benefit of the TD method lies in the fact that correlated event lifetimes can be measured. In SiPMs, the TD method also allows the study of the correlated noise origins through a two-dimensional histogram with the pulse amplitude information [10, 35].

Figure 11 compares λ and f_{AP} as a function of the holdoff delay for both methods. For t_h ranging from 53 ns to 11.7 µs, a spread of about 5% is observed in the uncorrelated event rates while error bars remain under 0.4% for each data point. The variation in absolute values of λ in figure 11(a) is believed to come from faint light fluctuations between the different acquisitions.

In figure 11(b), values of λ for the ZPP method are normalized point by point with respect to the TD method to better show the relative deviation between the two methods and to circumvent variations of possible light fluctuation between acquisitions.

In figure 11(c), the afterpulsing decreases as a function of the holdoff delay. Once again, the differences between the TD method and ZPP method can be associated with the positioning of the fit limit and selection of the plateau region. While no values of λ and f_{AP} are estimated for $t_h = 11.5$ µs with the ZPP method (because of the lack of a plateau region), the TD method succeeds in fitting all the TD distributions of figure 7.

6. ZPP method implemented in the digital SiPM with embedded signal processing

The ZPP method was implemented inside the digital SiPM of section 2.2 using a time-driven event counter. Figure 12 presents λ as a function of Δt for a subset of 4 SPADs while the remaining SPADs of the array are turned off. SPADs are biased at an overvoltage of 0.5 V. A total of 10^5 intervals (N_{total}) are acquired for each SPAD. A period of 1 ms elapses between consecutive intervals. The holdoff delay is set to 20 ns.

The uncorrelated event rate of each SPAD follows the distinctive behavior of the plateau region at large interval widths as observed previously with the analog monitor architecture. It is limited at small interval widths by the holdoff delay damping and at large interval widths by N_{total}.

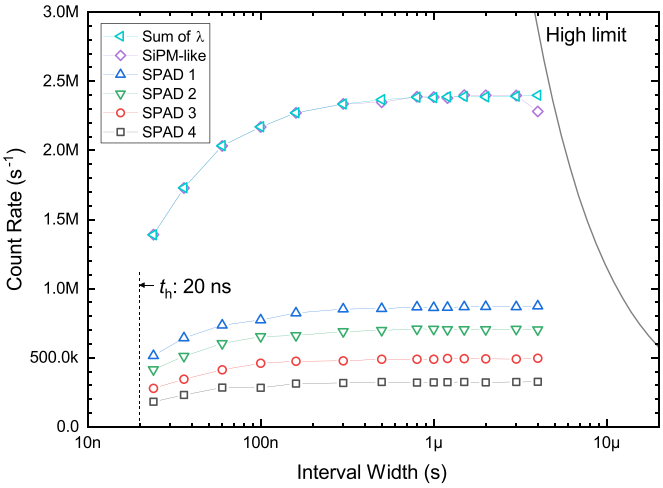

Figure 12. Uncorrelated event rate as a function of the interval time width for a 1× 4 SPAD subset of the digital SiPM acquired with an embedded time-driven event counter.

The uncorrelated event rates greatly differ from one SPAD to another, ranging from $(323 \pm 3) \times 10^3$ s^{-1} to $(865 \pm 7) \times 10^3$ s^{-1}. A high variation in dark noise rate is typical for SPADs fabricated in high density CMOS technology because the silicon surface and wells contain a high density of defects. Despite a high dark noise rate at room temperature, they have one of the highest single photon timing resolutions ever reported [22].

In figure 12, the *SiPM-like* curve represents the 4 SPADs treated as a small SiPM. That is, the ZPP method was performed on the sum of their signals within the readout ASIC. As in an analog SiPM, where the SPADs' individual addresses are not available, it is the sum of all events from intervals recorded during the same acquisition frame.

In contrast, the *Sum of* λ curve corresponds to the sum of all SPAD's individual uncorrelated event rates. Unlike the SiPM-like curve that departs from the plateau region near the high Δt limit, the sum of each individual λ circumvent the limit imposed by N_{total}. Indeed, as λ increases (with over voltage for example) or as more SPADs are turned on, intervals of the SiPM are filled quickly and reach a point where they all contain at least one event ($N_0 = 0$). This limits the total uncorrelated event rate that can be computed with the ZPP method if only the signal of the whole SiPM is available. Recording more intervals (higher N_{total}) becomes a time-consuming solution.

In this case, having access to SPAD addresses allows the characterization of individual SPAD's event rates and provides an estimation of the overall SiPM's uncorrelated event rate not limited by an excess number of intervals with more than one event or by noisy SPADs. Also, the interval width can be set larger when a SPAD is characterized individually as it is more likely to observe 0-event intervals than a full SiPM. Still, the characterization of large analog SiPMs using the ZPP method can be achieved with proper conditioning [10].

The TD method could also be implemented using on-chip TDCs where event time-stamping is implicit. However, sorting and fitting the time delays would be done offline considering the space needed to implement such modules inside a digital SiPM. By requiring only the count of empty intervals N_0 and the total number of intervals acquired N_{total}, the ZPP method is faster and more practical to implement on-chip for the characterization of the digital SiPM with embedded digital signal processing.

7. Conclusion

This paper reviews two characterization methods used to measure the uncorrelated event rate of digital SiPMs. The TD method and the ZPP method give similar results in identifying and correctly estimating the contribution of correlated events such as afterpulsing.

We provide guidelines to select the interval width for the ZPP method considering the holdoff delay and the event rate. For the TD method, we adapt the fitting procedure to consider the holdoff delay explicitly.

The two methods are tested on the digital SiPM with the analog monitor architecture. For both methods, reliable values of the uncorrelated event rate for single SPAD are obtained. We also demonstrate that the holdoff delay can filter out all or part of the afterpulsing contribution. The measurement of count rates free from correlated noise shows that the overestimation of the PDE can be avoided in digital SiPMs.

As a step forward, the ZPP method using a time-driven sampling procedure is successfully implemented inside the digital SiPM with embedded signal processing. Such an achievement is of high importance in the development of fully digital SiPMs. Indeed, while digital SiPMs benefit from on-chip processing, access to low-level parameters such as the noise rate is critical to a complete and proper characterization but not necessarily in the form of an analog signal.

Acknowledgments

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), by the Fonds de Recherche du Québec–Nature et Technologies (FRQNT), by the Regroupement Stratégique en Microsystèmes du Québec (ReSMiQ), by the Arthur B. McDonald Institute and by CMC Microsystems.

ORCID iDs

Frédéric Vachon https://orcid.org/0000-0002-6976-6624
Frédéric Nolet https://orcid.org/0000-0002-6750-7442
Jean-François Pratte https://orcid.org/0000-0002-8327-3842
Serge A Charlebois https://orcid.org/0000-0001-7857-5056

References

[1] Roncali E and Cherry S R 2011 Application of silicon photomultipliers to positron emission tomography *Ann. Biomed. Eng.* **39** 1358–77
[2] Garutti E 2011 Silicon photomultipliers for high energy physics detectors *J. Instrumentation* **6** C10003
[3] Vinogradov S L *et al* 2017 Evaluation of performance of silicon photomultipliers in LIDAR applications *Photon Counting Applications 2017* (Int. Society for Optics and Photonics) SPIE Optics + Optoelectronics Prague, Czech Republic Calice/SiPM Collaboration **10229** 102290L
[4] Chesi G, Malinverno L, Allevi A, Santoro R, Caccia M, Martemiyanov A and Bondani M 2019 Optimizing silicon photomultipliers for quantum optics *Sci. Rep.* **9** 7433
[5] Dolgoshein B, Balagura V, Buzhan P, Danilov M, Filatov L, Garutti E, Groll M, Ilyin A, Kantserov V and Kaplin V *et al* 2006 Status report on silicon photomultiplier development and its applications *Nucl. Instrum. Methods Phys. Res.* A Calice/SiPM Collaboration **563** 368–76
[6] Bonanno G, Finocchiaro P, Pappalardo A, Billotta S, Cosentino L, Belluso M, Di Mauro S and Occhipinti G 2009 Precision measurements of photon detection efficiency for SiPM detectors *Nucl. Instrum. Methods Phys. Res.* A **610** 93–7
[7] Otte A N, Hose J, Mirzoyan R, Romaszkiewicz A, Teshima M and Thea A 2006 A measurement of the photon detection efficiency of silicon photomultipliers *Nucl. Instrum. Methods Phys. Res.* A **567** 360–3
[8] Piemonte C, Ferri A, Gola A, Picciotto A, Pro T, Serra N, Tarolli A and Zorzi N 2012 Development of an automatic procedure for the characterization of silicon photomultipliers *Nuclear Symp. and Medical Conf. (NSS/MIC), 2012 IEEE* (IEEE: Anaheim, CA, USA) pp 428–32
[9] Zappalà G, Acerbi F, Ferri A, Gola A, Paternoster G, Zorzi N and Piemonte C 2016 Set-up and methods for SiPM photodetection efficiency measurements. *J. Instrumentation* **11** P08014
[10] Otte A N, Garcia D, Nguyen T and Purushotham D 2017 Characterization of three high efficiency and blue sensitive silicon photomultipliers *Nucl. Instrum. Methods Phys. Res.* A **846** 106–25
[11] Akiba M, Tsujino K, Sato K and Sasaki M 2009 Multipixel silicon avalanche photodiode with ultralow dark count rate at liquid nitrogen temperature *Opt. Express* **17** 16885–97
[12] Xu C, Klanner R, Garutti E and Hellweg W L 2014 Influence of x-ray irradiation on the properties of the Hamamatsu silicon photomultiplier s10362-11-050c *Nucl. Instrum. Methods Phys. Res.* A **762** 149–61
[13] Garutti E and Musienko Y 2019 Radiation damage of SiPMs *Nucl. Instrum. Methods Phys. Res.* A **926** 69–84
[14] Gundacker S, Turtos R M, Kratochwil N, Pots R H, Paganoni M, Lecoq P and Auffray E 2020 Experimental time resolution limits of modern SiPMs and TOF-PET detectors exploring different scintillators and Cherenkov emission *Phys. Medicine Biol.* **65** 025001
[15] Cova S, Ghioni M, Lacaita A, Samori C and Zappa F 1996 Avalanche photodiodes and quenching circuits for single-photon detection *Appl. Opt.* **35** 1956–76
[16] Tétrault M-A, Lamy E D, Boisvert A, Fontaine R and and Pratte J-F 2013 Low dead time digital SPAD readout architecture for realtime small animal PET *2013 IEEE Nuclear Symp. and Medical Conf. (2013 NSS/MIC)* (IEEE: Seoul, South Korea) pp 1–6
[17] Lemaire W, Nolet F, Therrien A C, Pratte J-F and Fontaine R 2016 Design considerations for embedded real-time processing for 3D digital SiPMs with multiple TDCs *2016 IEEE Nuclear Symp., Medical Conf. and*

Room-Temperature Semiconductor Detector Workshop (NSS/MIC/RTSD) (IEEE: Strasbourg, France) pp 1–3
[18] Schaart D R, Charbon E, Frach T and Schulz V 2016 Advances in digital SiPMs and their application in biomedical imaging *Nucl. Instrum. Methods Phys. Res.* A **809** 31–52
[19] Villa F, Lussana R, Bronzi D, Tisa S, Tosi A, Zappa F, Mora A D, Contini D, Durini D, Weyers S and Brockherde W 2014 CMOS imager with 1024 SPADs and TDCs for single-photon timing and 3-D time-of-flight *IEEE J. Selected Topics Quantum Electron.* **20** 364–73
[20] Frach T, Prescher G, Degenhardt C, de Gruyter R, Schmitz A and Ballizany R 2009 The digital silicon photomultiplier—Principle of operation and intrinsic detector performance *Nuclear Symp. Conf. Record (NSS/MIC), 2009 IEEE* (IEEE: Orlando, FL, USA) pp 1959–65
[21] Nolet F, Rhéaume V-P, Parent S, Charlebois S A, Fontaine R and Pratte J-F 2016 A 2D proof of principle towards a 3D digital SiPM in HV CMOS with low output capacitance *IEEE Trans. Nucl. Sci.* **63** 2293–9
[22] Nolet F, Parent S, Roy N, Mercier M-O, Charlebois S A, Fontaine R and Pratte J-F 2018 Quenching circuit and SPAD integrated in CMOS 65 nm with 7.8 ps FWHM single photon timing resolution *Instruments* **2** 19
[23] Nolet F, Lemaire W, Dubois F, Roy N, Carrier S, Samson A, Charlebois S A, Fontaine R and Pratte J-F 2020 256 pixelated SPAD readout ASIC with in-pixel TDC and embedded digital signal processing for uniformity and skew correction *Nucl. Instrum. Methods Phys. Res.* A **949** 162891
[24] Kindt W J 1999 Geiger mode avalanche photodiode arrays: For spatially resolved single photon counting *PhDT* 471 (Delft University Press) (http://resolver.tudelft.nl/uuid:300807f8-97db-4e6b-a77e-a19f45ce6e40)
[25] Acerbi F and Gundacker S 2019 Understanding and simulating SiPMs *Nucl. Instrum. Methods Phys. Res.* A **926** 16–35
[26] Rosado J, Aranda V M, Blanco F and Arqueros F 2015 Modeling crosstalk and afterpulsing in silicon photomultipliers *Nucl. Instrum. Methods Phys. Res.* A **787** 153–6
[27] Vacheret A *et al* 2011 Characterization and simulation of the response of multi-pixel photon counters to low light levels *Nucl. Instrum. Methods Phys. Res.* A **656** 69–83
[28] Musienko Y, Reucroft S and Swain J 2006 The gain, photon detection efficiency and excess noise factor of multi-pixel Geiger-mode avalanche photodiodes *Nucl. Instrum. Methods Phys. Res.* A **567** 57–61
[29] Eckert P, Schultz-Coulon H-C, Shen W, Stamen R and and Tadday A 2010 Characterisation studies of silicon photomultipliers *Nucl. Instrum. Methods Phys. Res.* A **620** 217–26
[30] Sigworth F and Sine S M 1987 Data transformations for improved display and fitting of single-channel dwell time histograms *Biophys.* J. **52** 1047
[31] Nekola J C, Šizling A L, Boyer A G and Storch D 2008 Artifactions in the log-transformation of species abundance distributions *Folia Geobotanica* **43** 259–68
[32] Keysigh Technologies 53230A 350 MHz Universal Frequency Counter/Timer, 12 digits/s, 20 ps (https://www.keysight.com/en/pd-1893420-pn-53230a/350-mhz-universal-frequency-counter-timer-12-digits-s-20-ps?cc=CA&lc=eng) (Accessed: 5 April 2020)
[33] Parent S, Côté M, Vachon F, Groulx R, Martel S, Dautet H, Charlebois S A and Pratte J-F 2018 Single photon avalanche diodes and vertical integration process for a 3D digital SiPM using industrial semiconductor technologies *2018 IEEE Nuclear Symp. and Medical Conf. Proc. (NSS/MIC)* (IEEE: Sydney, Australia) pp 1–4
[34] D'Agostini G 2003 *Bayesian Reasoning in Data Analysis: A Critical Introduction* (Singapore: World Scientific) 9812383565
[35] Jamil A *et al* 2018 VUV-sensitive silicon photomultipliers for xenon scintillation light detection in nEXO *IEEE Trans. Nucl. Sci.* nEXO Collaboration **65** 2823–33