Distortions from Multi-photon Triggering in a Single CMOS SPAD

Matthew W. Fishburn, and Edoardo Charbon,

Both authors are with Delft University of Technology, Delft, the Netherlands

ABSTRACT

Motivated by the need for correct characterization and operation of single-photon avalanche diodes (SPADs), three methods for discriminating between single- and multi-photon triggering in a single CMOS SPAD are compared. The first method, utilizing a measurement of the avalanche's quench time, correctly distinguishes between avalanches initiated by one photon or 100 incident photons with $p > 0.80$ in single-shot measurements. The second method, which modulates the detector efficiency, correctly distinguished streams of pulses with single photons or 100 photons with $p > 0.99$, but is unable to provide a single-shot measurement. The third method, which examines distortions to the timing jitter's diffusion tail, requires knowledge of the incident timing jitter *a priori* and is unlikely to be useful in most systems. All compared methods are independent of one another, and show promise for distinguishing how many photons seed an avalanche.

Keywords: Geiger-Mode Avalanche Photodiodes; Single-Photon Avalanche Diodes; Quantum Key Distribution

1. INTRODUCTION

Recent attacks against quantum key distribution (QKD) systems have relied on avalanche photodiodes' inability to distinguish whether an avalanche is triggered by a quantum effect, such as a single photon, or a classical effect like a burst of bright light.[1] Additionally, in areas such as three-dimensional imaging, knowledge of the reflected power from a laser beam can help with distance estimation.[2] Previous work on photon-number-resolving (PNR) avalanche photodiodes has focused on using multiple devices to overcome these problems[3] and studying effects in single diodes;[4,5] in the present work, we compare three methods for estimating or classifying the number of photons triggering avalanches in a single single-photon avalanche diode (SPAD) fabricated in a standard CMOS process.

2. METHODOLOGY AND EXPERIMENTAL SETUP

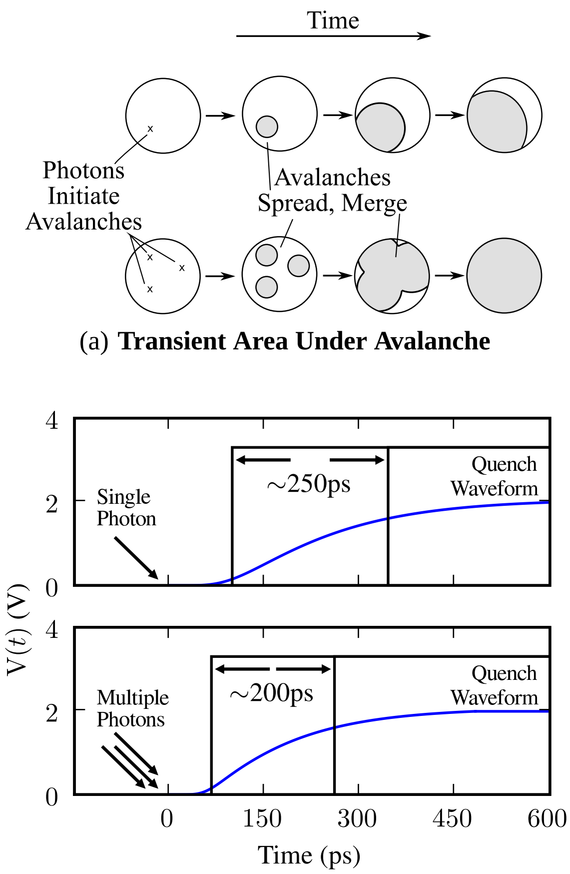
This section describes three methods for distinguishing the number of simultaneously incident photons on a single CMOS SPAD, along with an experimental setup capable of using all three methods simultaneously.

The first of the proposed method observes the quench time of the photodiode, which is known to vary as a function of the number of incident photons.[6] When multiple photons are simultaneously incident on a SPAD, more than one avalanche may be triggered, and the quench time of the avalanche will be reduced from the faster current build-up. Fig. 1 illustrates the mechanism of multi-photon avalanching. By sampling the quench time with a time-to-digital converter (TDC) coupled to a SPAD via two comparators, as shown later, quench times can be observed as a function of the incident power. This method has been previously used in custom silicon processes,[4] and a related method has been recently used with InGaAs/InP diodes.[5]

The second method relies on modulating the detector responsivity to single photons and seeing whether the event rate caused by incident light decreases as expected. Since triggering on a single photon is probabilistic,[7] when a large number of photons are simultaneously incident on the detector, the detector responds identically, no matter the efficiency. However, if single photons are triggering the detector, fewer avalanches will be triggered. Comparing avalanches as a function of efficiency allows an estimation of whether avalanches are being triggered by one photon or many photons.

Finally, the third method observes whether or not the timing response of the detector exhibits a diffusion tail. This method is similar to previous work showing the timing jitter FWHM can change as a function of the number of simultaneously incident photons,[8] though this work focuses on the diffusion tail rather than the FWHM jitter. In an avalanche diode based on an n-well, such as the ones in the present work, carriers created deep in the n-well are known to cause a

Author contact information: E-mail: e.charbon@tudelft.nl, Telephone: +31-15-278-3667



(b) **Measured Quench Time**

Figure 1. **Expected Quench Waveforms for Single vs. Multiple Photons**

diffusion tail in the diode's timing response.[9] When multiple photons are simultaneously incident on the diode, there is an increase in the probability that one of the carriers promptly triggers an avalanche and masks the diffusion tail. As more photons are simultaneously incident, the diffusion tail will become less pronounced and eventually disappear. Fig. 2 shows the processes which can mask the diffusion tail.

To compare these methods, an FPGA-based read out system acquires data from a CMOS chip integrating a SPAD, quenching circuitry, comparators, and a TDC in a 0.35µm CMOS process. Circular SPADs with a 20µm diameter were used, whereas an NMOS transistor acts to passively quench these SPADs; a passive quenching scheme with a dead time of 300ns was used for the sake of simplicity, though in retrospect an active quenching scheme would have been preferable for the reduction of afterpulsing. The NMOS transistor and SPAD are in series between 0V and a single positive voltage. The SPADs were coupled to an on-chip, short-range, Vernier-delay-line-based TDC via two comparators with adjustable thresholds, as shown in Fig. 3. Figures of merit for the TDCs and SPADs, along with information about SPAD active area distortions caused by the CMOS process, have been previously reported.[10]

The SPADs were assembled on a daughterboard that also contains a commercial TDC with a resolution of 61ps, and this TDC was coupled to a buffered output from the SPAD, along with the synchronous, electrical output of a laser. The on-chip TDC is only able to measure the quench time of the avalanche or the jitter, not both; an external TDC allows measurements of timing jitter and quench time simultaneously. After being routed through zero or more neutral density filters (NDFs), with a total optical density ranging between 0 and 4.5, the optical output of the laser shone on the SPADs. The distance between the laser and the chip was also varied to allow a greater increase in optical power. The number of simultaneously incident photons, E[n], is extrapolated for each laser distance based on the optical density of the current set of NDFs from the SPAD's count rate above its dark count rate (DCR) compared to the laser frequency when using an

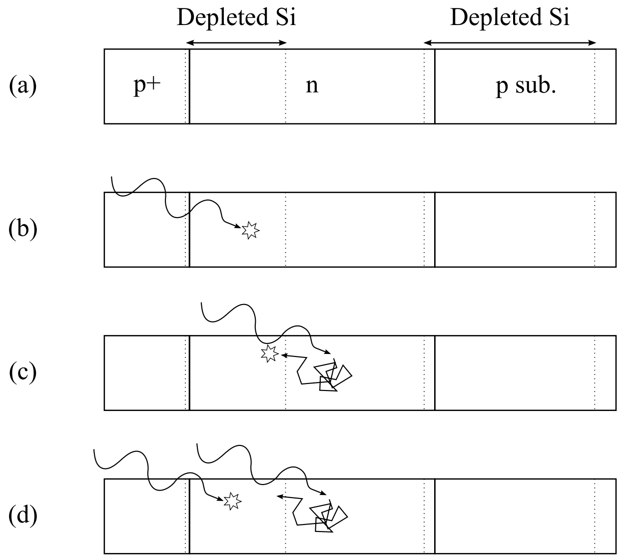


Figure 2. **Processes Creating the Diffusion Tail** — (a) shows the diode regions, in (b) a photon-generated carrier immediately causing an avalanche, in (c) a photon-generated carrier must diffuse before triggering an avalanche, and (d) shows how an immediate carrier can mask a diffusing carrier when multiple photons are simultaneously incident on a SPAD

NDF with an optical density of 1 and knowledge of the SPAD's PDP. The laser is operated at a frequency of 2.5MHz, with an output wavelength of 637nm and a timing FWHM of 320ps. A read out system based on a field-programmable gate array (FPGA) allows simultaneous acquisition of data from both TDCs, along with observation of the SPAD's event rate. The FPGA also controls the applied voltage to the SPADs, as output by an adjustable power supply. The power supply normally outputs a voltage of +20.5V, corresponding to an excess bias of 2.0V, but during the efficiency modulation method the supply switches between +20.5V and +21.0V to modulate the diode's detection efficiency. An ethernet link to a computer workstation allows any streamed data to be saved for offline analysis. The entire experimental setup is shown in Fig. 3.

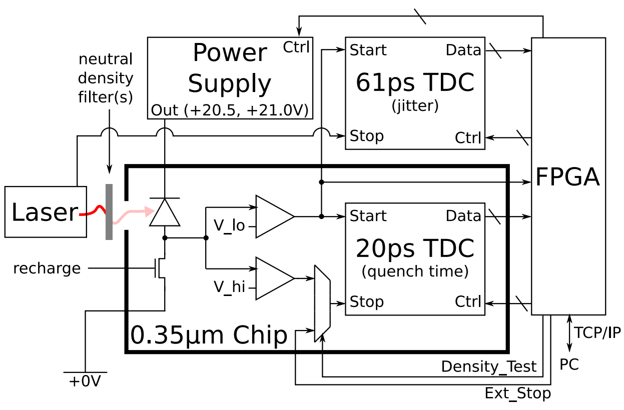


Figure 3. **Experimental Setup**

Utilizing this experimental setup E[n] can be modulated by switching between different neutral density filters. Estimations of whether single or multiple photons are triggering the avalanches in a single diode can be characterized with three different methods: using the avalanche's quench timing, observing the jitter of a SPAD, or acquiring the event rate as a function of the excess bias.

3. QUENCH TIME METHOD

Fig. 4 shows the results from observing the quench time as a function of E[n] when operating the SPAD at an excess bias of 2.0V, and sampling the time between the quench waveform's crossing of 0.1V and 1.6V. The graph shows four regions: a photon-starved region, in which the quench time does not visibly vary; a saturating region, in which the quench time decreases; a saturated regime, in which the quench time stays approximately uniform; and an over-saturated region, with an unknown effect causing the quench time to begin increasing again. In the over-saturated region, substrate noise may be a factor in the increase in rise time. In this region, a large number of free carriers are being created in the silicon substrate, and this injected substrate noise might be causing problems with either the on-chip TDCs or the SPAD-coupled comparators. The SPAD was passively recharged when using this scheme.

A simple criterion for classifying an avalanche that is either 1-photon- or 100-photon-initiated is to set a threshold quench time; any avalanche with a quench time lasting longer than this threshold is expected to be caused by a single photon, whereas a shorter quench time would signify a 100-photon-initiated. When a 125ps threshold is set for the quench time, >80% of 1-photon- or 100-photon-initiated avalanches are correctly classified.

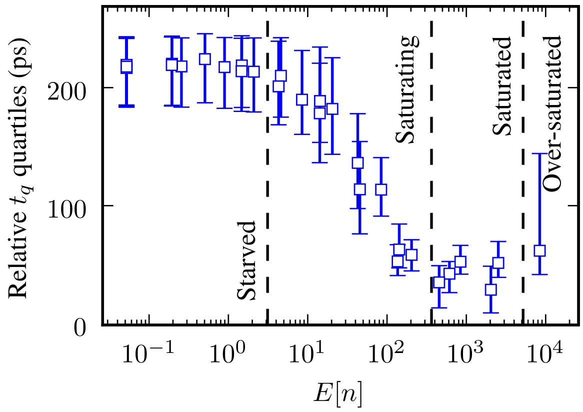


Figure 4. **Quench time vs. number of simultaneously incident photons**

4. EFFICIENCY MODULATION METHOD

In a photon-starved mode of operation, shifts in the excess bias will cause roughly a linear shift in the observed event rate.[11] For example, a count rate ratio slightly less than 1.25 is expected when the SPAD is modulated from an excess bias of 2.5V to 2.0V — in other words, the count rate at an excess bias of 2.5V is expected to be roughly 1.25 times that of the count rate at an excess bias 2.0V. When multiple photons are incident on the diode, the probability of avalanche/no avalanche is governed both by whether E[n] is zero and the PDE, rather than simply the PDE, giving the probability of triggering during a single pulse being

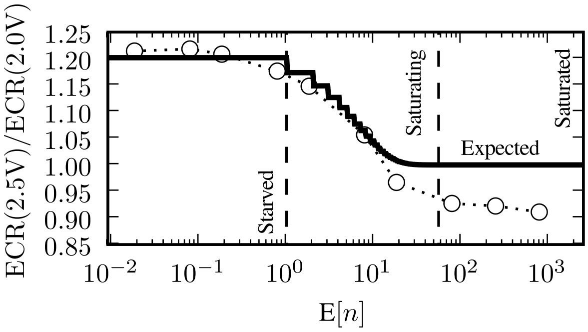
$$P(\text{Avalanche}) = 1-(1-\text{PDP}(\lambda,\text{V}_{\text{eb}}))^{\text{E[n]}}, \tag{1}$$

with λ the wavelength of the incident photons, V_{eb} the excess bias, and PDP the photon detection probability, which is the probability of an avalanche occurring when a single photon is incident on a SPAD. If a large number of photons per laser burst are incident on the diode, then the probability of an avalanche per burst will approach one. In this condition, the triggering probability is independent of the excess bias; the event rate of the avalanche diode will track the laser frequency. Ignoring noise, the ratio of the event rate between different excess biases V_{eb1} and V_{eb2} is expected to be

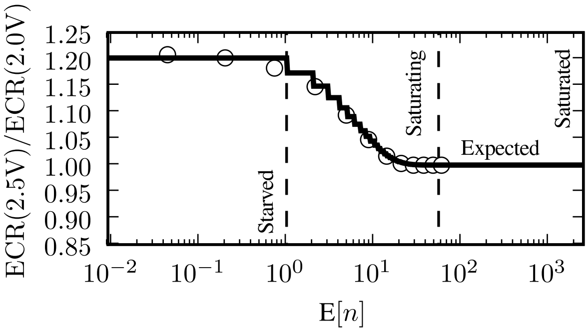
$$\frac{1-(1-\text{PDP}(\lambda,\text{V}_{\text{eb1}}))^{\text{E[n]}}}{1-(1-\text{PDP}(\lambda,\text{V}_{\text{eb2}}))^{\text{E[n]}}}. \tag{2}$$

However, correlated and uncorrelated noise, which are expected to be larger at higher excess biases, may actually cause the ratio to drop below 1 when the diode is passively recharged with a hold-off time similar to the laser period. If an afterpulse occurs during recharge in this scheme, the diode's quench voltage will not drop below the comparator's threshold voltage before the next laser pulse, which may trigger an avalanche; thus greater afterpulsing probability will actually decrease the ratio for when the number of simultaneously incident photons is large. The phenomenon is commonly called pile-up or charge pile-up. The effect will not occur when a SPAD is actively recharged.

Fig. 5 compares experimental data from operating a SPAD at an excess bias of 2.0V and 2.5V with the theoretical result from (2). Excluding DCR, the count rate ratio is roughly 1.2, which is less than the expected 1.25 due to PDP saturation.[11] When the laser period is near the SPAD's dead time and the diode is passively recharged, which occurs when a constant bias is applied to the restore signal (see Fig. 3), as expected correlated noise due to afterpulsing causes significant distortions when the number of simultaneously incident photons is large. If the diode is instead actively recharged, by applying a square wave with a small duty cycle to the restore signal, this distortion does not occur and the ratio gives an excellent match to the predicted values, as Fig. 5(b) shows. The square wave was not synchronized to the laser, and recharge occurred whether or not an avalanche does.



(a) **Passive Restore**



(b) **Active Restore**

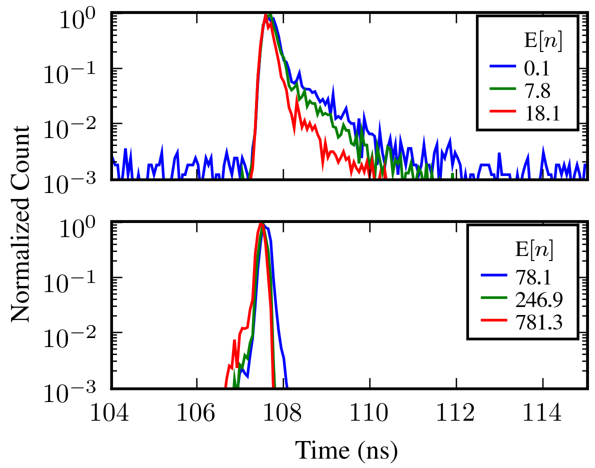
Figure 5. **Count rate ratio vs. number of simultaneously incident photons**

5. DIFFUSION TAIL METHOD

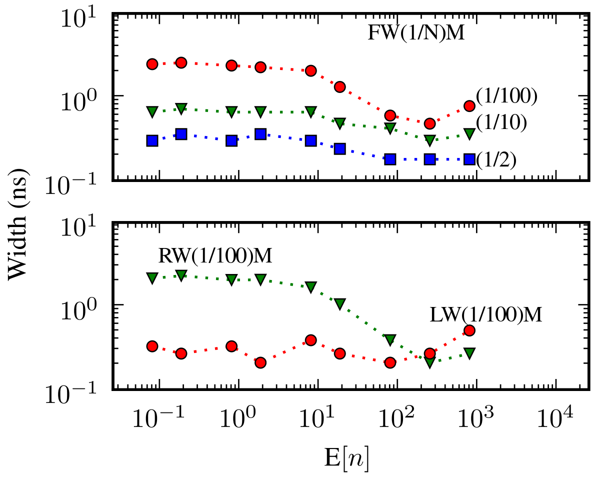
The diffusion tail's attenuation as a function of E[n] is shown in Fig. 6(a). When E[n] $\approx$ 15, the tail is still present, but less pronounced, than in the photon-starved regime, but when E[n] $>$ 60, the diffusion tail is, at most, three orders of magnitude smaller than the jitter peak. Whether the diffusion tail is observed can be quantified by examining the right-width one-hundredth maximum, RW(1/100)M, which is the width of the curve to the peak's right at a value that is 1/100 that of the peak. The RW(1/100)M is seen to decrease in Fig. 6(b) from several nanoseconds when E[n] $<$ 1 to below 500ps when E[n] $>$ 60. The SPAD was passively recharged when using this scheme.

6. METHOD COMPARISON AND DISCUSSION

While each presented method has several strengths and weaknesses, the efficiency modulation methods shows the greatest promise for distinguishing whether a detector is operating in a photon-starved regime or not. The diffusion tail method



(a) **Time-domain jitter curves**



(b) **FW(1/N)M, LW(1/100)M and RW(1/100)M**

Figure 6. **Jitter vs. number of simultaneously incident photons**

requires knowledge of the incident photon stream's timing jitter *a priori*; shifts in the incident photon stream's generation require complex timing measurements. Especially for QKD, where an attacker may have precise control over the incident photon's timing, this method is not likely to be useful. Additionally, this method is likely to be sensitive to shifts in the operating conditions, such as the excess bias or temperature, of the detector.

While the quench time method allows for a single-shot measurement, an advantage that neither of the other two methods share, this method may also be susceptible to shifts in the operating condition of the avalanche photodiode. This method has an additional shortcoming for QKD, which is that an attacker could stream single photons separated by the quench time of the photon-starved diode. In this mode of illumination, the diode would always measure a single photon incident on its active area; however, the diode would still be triggered with high probability.

Of the three methods, the efficiency modulation method shows the most promise for use in an actual system. The method is better resistant to shifts in operating point than the other two methods, and does not require a TDC. However, this method requires a shift in diode operating point, with the associated cost of more complex voltage supplies and operating the diode for a period of time at a lower detection efficiency than the optimal efficiency. Additionally, a minimum integration time is required for this method, though there will be a trade-off between the integration time and the uncertainty in whether the avalanches are being caused by a single photon or multiple photons.

The three methods, which are independent of one another and use the same setup, have the potential to be used at the same time to cover one each other's weaknesses. For example, utilizing both the diffusion tail and quench time methods would eliminate the mentioned avenues of attack against QKD systems using such methods, since one attack relies on an imprecise triggering method using precisely timed single photons while the other attack relies on precisely timing a large number of photons. In addition, the methods show compatibility with methods relying on observing correlations in avalanches across multiple diodes.[3] If such methods were to be used as preventative measures in a QKD system, using multiple methods is likely to create a more secure system, though at the cost of increasing system complexity and possible false alarms.

7. ACKNOWLEDGMENTS

The authors are grateful to Yuki Maruyama for help with the experimental setup, and to Xilinx, Inc. and Acam Messelectronic GmbH for hardware donations.

REFERENCES

[1] Lyderson, L., Wiechers, C., Wittmann, C., Elser, D., Skaar, J., and Makarov, V., "Hacking commercial quantum cryptography systems by tailored bright illumination," *Nat. Photonics* **4**, 686–689 (Aug. 2010).

[2] Niclass, C. L., *Single-Photon Image Sensors in CMOS: Picosecond Resolution for Three-Dimensional Imaging*, PhD thesis, École Polytechnique Fédérale de Lausanne (2008).

[3] Eraerds, P., Legré, M., Rochas, A., Zbinden, H., and Gisin, N., "SiPM for fast photon-counting and multiphoton detection," *Opt. Express* **15**, 14539–14549 (Oct 2007).

[4] Blazej, J., "Photon number resolving in geiger mode avalanche photodiode photon counters," *Journal of Modern Optics* **51**(9-10), 1491–1497 (2004).

[5] Kardynal, B. E., Yuan, Z. L., and Shields, A. J., "An avalanche-photodiode-based photon-number-resolving detector," *Nature Photonics* **2**(7), 425–428 (2008).

[6] Kirchner, G. and Koidl, F., "Compensation of spad time-walk effects," *Journal of Optics A: Pure and Applied Optics* **1**(2), 163 (1999).

[7] Spinelli, A. and Lacaita, A., "Physics and numerical simulation of single photon avalanche diodes," *Electron Devices, IEEE Transactions on* **44**, 1931–1943 (Nov. 1997).

[8] Blazej, J. and Prochazka, I., "Avalanche photodiode output pulse rise-time study," *Photon Counting Applications, Quantum Optics, and Quantum Information Transfer and Processing II* **7355**(1), 73550M, SPIE (2009).

[9] Ripamonti, G. and Cova, S., "Carrier diffusion effects in the time-response of a fast photodiode," *Solid-State Electronics* **28**(9), 925–931 (1985).

[10] Fishburn, M., Maruyama, Y., and Charbon, E., "Reduction of fixed-position noise in position-sensitive single-photon avalanche diodes," *Electron Devices, IEEE Transactions on* **58**, 2354–2361 (Aug. 2011).

[11] Cova, S., Ghioni, M., Lotito, A., Rech, I., and Zappa, F., "Evolution and prospects for single-photon avalanche diodes and quenching circuits," *Journal of Modern Optics* **51**(9), 1267–1288 (204).