



Delft University of Technology

## Dynamic range extension for photon counting arrays

Antolovic, Ivan Michel; Bruschini, Claudio; Charbon, Edoardo

**DOI**

[10.1364/OE.26.022234](https://doi.org/10.1364/OE.26.022234)

**Publication date**

2018

**Document Version**

Final published version

**Published in**

Optics Express

**Citation (APA)**

Antolovic, I. M., Bruschini, C., & Charbon, E. (2018). Dynamic range extension for photon counting arrays. *Optics Express*, 26(17), 22234-22248. <https://doi.org/10.1364/OE.26.022234>

**Important note**

To cite this publication, please use the final published version (if applicable). Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.



# Dynamic range extension for photon counting arrays

IVAN MICHEL ANTOLOVIC,<sup>1,2,\*</sup> CLAUDIO BRUSCHINI,<sup>2</sup> AND EDOARDO CHARBON<sup>1,2</sup>

<sup>1</sup>*Applied Quantum Architecture Laboratory (AQUA), Technical University Delft 2628CD Delft, Netherlands*

<sup>2</sup>*Advanced Quantum Architecture Laboratory (AQUA), Ecole Polytechnique Fédérale de Lausanne (EPFL), 2002 Neuchâtel, Switzerland*

\*[i.m.antolovic@tudelft.nl](mailto:i.m.antolovic@tudelft.nl)

**Abstract:** Confocal microscopes use photomultiplier tubes and hybrid detectors due to their large dynamic range, which typically exceeds the one of single-photon avalanche diodes (SPADs). The latter, due to their photon counting operation, are usually limited to an output count rate to  $1/T_{dead}$ . In this paper, we present a thorough analysis, which can actually be applied to any photon counting detector, on how to extend the SPAD dynamic range by exploiting the nonlinear photon response at high count rates and for different recharge mechanisms. We applied passive, active event-driven and clock-driven (i.e. clocked, following *quanta image sensor* response) recharge directly to the SPADs. The photon response, photon count standard deviation, signal-to-noise ratio and dynamic range were measured and compared to models. Measurements were performed with a CMOS SPAD array targeted for image scanning microscopy, featuring best-in-class 11 V excess bias, 55% peak photon detection probability at 520 nm and >40% from 440 to 640 nm. The array features an extremely low median dark count rate below 0.05 cps/ $\mu\text{m}^2$  at 9 V of excess bias and 0°C. We show that active event-driven recharge provides  $\times 75$  dynamic range extension and offers novel ways for high dynamic range imaging. When compared to the clock-driven recharge and the *quanta image sensor* approach, the dynamic range is extended by a factor of  $\times 12.7$ -26.4. Additionally, for the first time, we evaluate the influence of clock-driven recharge on the SPAD afterpulsing.

© 2018 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

**OCIS codes:** (170.1790) Confocal microscopy; (180.2520) Fluorescence microscopy; (030.5260) Photon counting; (030.5290) Photon statistics.

## References and links

1. “Hamamatsu hybrid photodetector R10467U-40 datasheet,” [https://www.hamamatsu.com/resources/pdf/etd/HPD\\_TPMH1361E.pdf](https://www.hamamatsu.com/resources/pdf/etd/HPD_TPMH1361E.pdf).
2. J. B. Pawley, *Handbook of Biological Confocal Microscopy*, 3<sup>rd</sup> ed. (Springer, 2006).
3. C. J. Sheppard, “Super-resolution in confocal imaging,” *Optik (Stuttg.)* **80**, 53–54 (1988).
4. C. B. Müller and J. Enderlein, “Image scanning microscopy,” *Phys. Rev. Lett.* **104**(19), 198101 (2010).
5. C. J. R. Sheppard, S. B. Mehta, and R. Heintzmann, “Superresolution by image scanning microscopy using pixel reassignment,” *Opt. Lett.* **38**(15), 2889–2892 (2013).
6. M. Castello, C. J. R. Sheppard, A. Diaspro, and G. Vicidomini, “Image scanning microscopy with a quadrant detector,” *Opt. Lett.* **40**(22), 5355–5358 (2015).
7. J. Huff, “The Airyscan detector from ZEISS: confocal imaging with improved signal-to-noise ratio and super-resolution,” *Nat. Methods* **12**(12), 1–2 (2015).
8. I. Gyongy, N. Calder, A. Davies, N. A. W. Dutton, R. R. Duncan, C. Rickman, P. Dalgarno, and R. K. Henderson, “A 256x256, 100-kfps, 61% fill-factor SPAD image sensor for time-resolved microscopy applications,” *IEEE Trans. Electron Dev.* **65**(2), 547–554 (2018).
9. I. Gyongy, A. Davies, B. Gallinet, N. A. W. Dutton, R. R. Duncan, C. Rickman, R. K. Henderson, and P. A. Dalgarno, “Cylindrical microlensing for enhanced collection efficiency of small pixel SPAD arrays in single-molecule localisation microscopy,” *Opt. Express* **26**(3), 2280–2291 (2018).
10. I. M. Antolovic, S. Burri, C. Bruschini, R. A. Hoebe, and E. Charbon, “SPAD imagers for super resolution localization microscopy enable analysis of fast fluorophore blinking,” *Sci. Rep.* **7**, 44108 (2017).
11. I. M. Antolovic, S. Burri, C. Bruschini, R. Hoebe, and E. Charbon, “Analyzing blinking effects in super

- resolution localization microscopy with single-photon SPAD imagers,” in *Proc. SPIE 9714* (International Society for Optics and Photonics, 2016), p. 971406.
12. C. Bruschini, H. Homulle, and E. Charbon, “Ten years of biophotonics single-photon SPAD imager applications: retrospective and outlook,” in *Proc. SPIE 10069* (International Society for Optics and Photonics, 2017), 10069–21.
  13. D. Bronzi, F. Villa, S. Tisa, A. Tosi, and F. Zappa, “SPAD figures of merit for photon-counting, photon-timing, and imaging applications: a review,” *IEEE Sens. J.* **16**(1), 3–12 (2016).
  14. J. Kröger, T. Ahrens, J. Sperling, W. Vogel, H. Stolz, and B. Hage, “High intensity click statistics from a  $10 \times 10$  avalanche photodiode array,” *J. Phys. At. Mol. Opt. Phys.* **50**(21), 214003 (2017).
  15. J. Ma, S. Masoodian, D. A. Starkey, and E. R. Fossum, “Photon-number-resolving megapixel image sensor at room temperature without avalanche gain,” *Optica* **4**(12), 1474 (2017).
  16. E. Fossum, “The quanta image sensor (QIS): concepts and challenges,” in *Imaging and Applied Optics* (OSA, 2011), p. JTUe1.
  17. E. R. Fossum, “Modeling the performance of single-bit and multi-bit quanta image sensors,” *IEEE J. Electron Devices Soc.* **1**(9), 166–174 (2013).
  18. X. Hesong, L. Pancheri, G.-F. D. Betta, and D. Stoppa, “Design and characterization of a p<sup>+</sup>/n-well SPAD array in 150nm CMOS process,” *Opt. Express* **25**, 77–83 (2017).
  19. D. Bronzi, F. Villa, S. Bellisai, B. Markovic, S. Tisa, A. Tosi, F. Zappa, S. Weyers, D. Durini, W. Brockherde, and U. Paschen, “Low-noise and large-area CMOS SPADs with timing response free from slow tails,” in *2012 Proceedings of the European Solid-State Device Research Conference (ESSDERC)* (IEEE, 2012), 230–233.
  20. S. Mandai, M. W. Fishburn, Y. Maruyama, and E. Charbon, “A wide spectral range single-photon avalanche diode fabricated in an advanced 180 nm CMOS technology,” *Opt. Express* **20**(6), 5849–5857 (2012).
  21. M.-J. Lee, P. Sun, and E. Charbon, “A first single-photon avalanche diode fabricated in standard SOI CMOS technology with a full characterization of the device,” *Opt. Express* **23**(10), 13200–13209 (2015).
  22. C. Veerappan and E. Charbon, “A low dark count p-i-n diode based SPAD in CMOS technology,” *IEEE Trans. Electron Dev.* **63**(1), 65–71 (2016).
  23. S. Lindner, S. Pellegrini, Y. Henrion, B. Rae, M. Wolf, and E. Charbon, “A high-PDE, backside-illuminated SPAD in 65/40 nm 3D IC CMOS pixel with cascaded passive quenching and active recharge,” *IEEE Electron Device Lett.* **38**(11), 1547–1550 (2017).
  24. S. Lindner, C. Zhang, I. M. Antolovic, J. M. Pavia, M. Wolf, and E. Charbon, “Column-parallel dynamic TDC reallocation in SPAD sensor module fabricated in 180nm CMOS for near infrared optical tomography,” in *International Image Sensor Workshop* (2017).
  25. E. A. G. Webster, L. A. Grant, and R. K. Henderson, “A high-performance single-photon avalanche diode in 130-nm CMOS imaging technology,” *IEEE Electron Device Lett.* **33**(11), 1589–1591 (2012).
  26. M. W. Fishburn, “Fundamentals of CMOS single-photon avalanche diodes,” Delft University of Technology (2012).
  27. I. M. Antolovic, S. Burri, C. Bruschini, R. Hoebe, and E. Charbon, “Nonuniformity analysis of a 65-kpixel CMOS SPAD imager,” *IEEE Trans. Electron Dev.* **63**(1), 57–64 (2016).
  28. S. H. Lee and R. P. Gardner, “A new G-M counter dead time model,” *Appl. Radiat. Isot.* **53**(4-5), 731–737 (2000).
  29. G. F. Knoll, *Radiation Detection and Measurement* (Wiley, 2000).
  30. S. Burri, H. Homulle, C. Bruschini, and E. Charbon, “LinoSPAD: a time-resolved 256×1 CMOS SPAD line sensor system featuring 64 FPGA-based TDC channels running at up to 8.5 giga-events per second,” in F. Berghmans and A. G. Mignani, eds. (International Society for Optics and Photonics, 2016), p. 98990D.
  31. S. Burri, C. Bruschini, and E. Charbon, “LinoSPAD: A compact linear SPAD camera system with 64 FPGA-based TDC modules for versatile 50 ps resolution time-resolved imaging,” *Instruments* **1**(1), 6 (2017).
  32. L. Sbaiz, F. Yang, E. Charbon, S. Susstrunk, and M. Vetterli, “The gigavision camera,” in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing* (IEEE, 2009), 1093–1096.
  33. S. Chick, R. Coath, R. Sellahewa, R. Turchetta, T. Leitner, and A. Fenigstein, “Dead time compensation in CMOS single photon avalanche diodes with active quenching and external reset,” *IEEE Trans. Electron Dev.* **61**(8), 2725–2731 (2014).
  34. N. Dutton, T. Al Abbas, I. Gyongy, and R. Henderson, “Extending the dynamic range of oversampled binary SPAD image sensors,” in *International Image Sensor Workshop* (2017).
  35. A. Eisele, R. Henderson, B. Schmidtke, T. Funk, L. Grant, J. Richardson, and W. Freude, “185 MHz count rate 139 dB dynamic range single-photon avalanche diode with active quenching circuit in 130 nm CMOS technology,” in *International Image Sensor Workshop* (2011).
  36. D. Stoppa, D. Mosconi, L. Pancheri, and L. Gonzo, “Single-photon avalanche diode CMOS sensor for time-resolved fluorescence measurements,” *IEEE Sens. J.* **9**(9), 1084–1090 (2009).
  37. S. H. Lee and M. Jae, “Non-Poisson counting statistics of a hybrid G-M counter dead time model,” *Nucl. Instruments Methods Phys. Res. Sect. B Beam Interact. with Mater.* **263**, 46–49 (2007).
  38. S. V. Polyakov and A. L. Migdall, “High accuracy verification of a correlated-photon-based method for determining photoncounting detection efficiency,” *Opt. Express* **15**(4), 1390–1407 (2007).
  39. M. Stipčević, B. G. Christensen, P. G. Kwiat, and D. J. Gauthier, “Advanced active quenching circuit for ultra-fast quantum cryptography,” *Opt. Express* **25**(18), 21861–21876 (2017).
  40. T. Al Abbas, N. A. W. Dutton, O. Almer, S. Pellegrini, Y. Henrion, and R. K. Henderson, “Backside illuminated

- SPAD image sensor with 7.83 $\mu\text{m}$  pitch in 3D-stacked CMOS technology,” in *2016 IEEE International Electron Devices Meeting (IEDM)* (IEEE, 2016), p. 8.1.1–8.1.4.
41. E. A. G. Webster, J. A. Richardson, L. A. Grant, D. Renshaw, and R. K. Henderson, “A single-photon avalanche diode in 90-nm CMOS imaging technology with 44% photon detection efficiency at 690 nm,” *IEEE Electron Device Lett.* **33**(5), 694–696 (2012).
  42. D. Bronzi, S. Tisa, F. Villa, S. Bellisai, A. Tosi, and F. Zappa, “Fast sensing and quenching of CMOS SPADs for minimal afterpulsing effects,” *IEEE Photonics Technol. Lett.* **25**(8), 776–779 (2013).
  43. D. Bronzi, F. Villa, S. Tisa, A. Tosi, F. Zappa, D. Durini, S. Weyers, and W. Brockherde, “100 000 frames/s 64  $\times$  32 single-photon detector array for 2-D imaging and 3-D ranging,” *IEEE J. Sel. Top. Quantum Electron.* **20**(6), 354–363 (2014).

## 1. Introduction

Confocal microscopes use photomultiplier tubes (PMTs), hybrid detectors [1] or single-photon avalanche diodes (SPADs). These microscopes can achieve a sub-diffraction limited lateral resolution by using a point-like illumination and detection pinhole [2]. In practice, a small detection pinhole reduces the number of collected photons and the signal-to-noise ratio (SNR), thus forcing the use of larger pinholes with coarser lateral resolution. Sheppard, and later on Muller *et al.*, showed an increase in SNR and lateral resolution by employing a detector array instead of a pinhole [3,4]. This approach was termed image scanning microscopy (ISM) [4–6].

A major drawback of PMTs and hybrid detectors is the fact that they are discrete sensors (one pixel), rarely implemented as 2D arrays. As a consequence, systems intended for ISM applications (i.e. requiring a 2D array) employ either a combination of discrete sensors, or linear arrays coupled through fibers and microlenses in order to form a 2D array at the microscopy output [7]. Custom-made SPADs are manufactured as discrete sensors as well. In contrast, SPADs designed in a CMOS process are reproducible and can be easily integrated in large arrays [8–12]; however, they usually feature a lower photon detection probability (PDP) and higher dark count rate (DCR) than custom-made SPADs [13].

PMTs and hybrid detectors can be operated in analog mode, generating an amplified current which is directly related to the detected photon flux; this enables flexibility in amplification and increases their dynamic range. Integrated SPADs are usually designed to operate only in Geiger mode, i.e. in a pulsed or click regime, with an output count rate limited to  $1/T_{dead}$ , where  $T_{dead}$  corresponds to the SPAD dead time. Confocal microscopes require detectors to measure up to 50 Mcps for conventional, and 1 Gcps for very bright samples. We therefore examine the possibility and limitations of extending the SPAD dynamic range, for different recharge mechanisms, by utilizing the exponential distribution of photons in time, which results in a nonlinear photon response. This methodology can actually be extended to any photon counting detector. Note that we define photon counting as the ability to detect and count single photon-generated carriers. Due to the detection efficiency, we cannot count all the impinging photons. Additionally, SPAD detectors cannot distinguish between one and more simultaneously impinging photons within  $T_{dead}$ . Due to this reason, the detectors are sometimes termed as click detectors [14].

Measurements were performed with a CMOS SPAD array targeted for image scanning microscopy, applying different recharge mechanisms – passive, active event-driven and clock-driven (following *quanta image sensor* response [15–17]) – directly to the SPADs. The CMOS SPAD array performance, which is comparable to custom made SPADs and actually outperforms available CMOS SPADs [18–21], PMTs and hybrid detectors, is presented in Section 2. Section 3 concentrates on the dynamic range extension. Section 4 investigates the linearity differences (with inter-arrival time histograms) and proposes a method to estimate afterpulsing in active clock-driven recharge. Concluding remarks are provided in Section 5.

## 2. CMOS high-performance SPAD array

The pixel schematic is shown in Fig. 1. The SPAD sensor is based on a previous low noise SPAD design [22], implemented in a 180 nm CMOS technology. The impinging photon flux  $i$

translates into the detected count rate  $n$  (loss-degraded, i.e. lowered by the photon detection efficiency PDE, whereby  $\text{PDE} = \text{PDP} \times \text{fill factor}$ ). The measured count rate (or click rate)  $m$  is equal to  $n$  only at low count rates. Quenching and recharge are implemented with a novel cascode transistor combination (transistors T1 and T2 in Fig. 1(a)) [23] to enable high excess bias  $V_{\text{EX}}$ , over and beyond what is usually permissible in the target technology. T2 performs both the quenching and recharge, and is biased by  $V_{\text{Q}}$ , the quenching voltage. While the SPAD fires, the anode voltage increases to  $V_{\text{EX}}$  and the inverter input reaches maximally  $3.3 \text{ V} - V_{\text{T1}}$ , where  $V_{\text{T1}}$  is the T1 threshold voltage. During the SPAD pulse, the source voltage of T1 increases and the body effect increases  $V_{\text{T1}}$ . This allows a higher  $V_{\text{EX}}$  with reliable operation which, as defined by the foundry, limits the maximum applicable voltage  $V_{\text{Tmax}}$  on the drain, source and gate of the transistors to 3.6 V. Simulations show indeed that for  $V_{\text{EX}} = 5.2$ , all transistor voltages are below 3.6 V [24]. We have however measured continuous SPAD operation at up to  $V_{\text{EX}} = 11 \text{ V}$  for a duration of one week without observed changes in the dead time (that would indicate a resistance change).

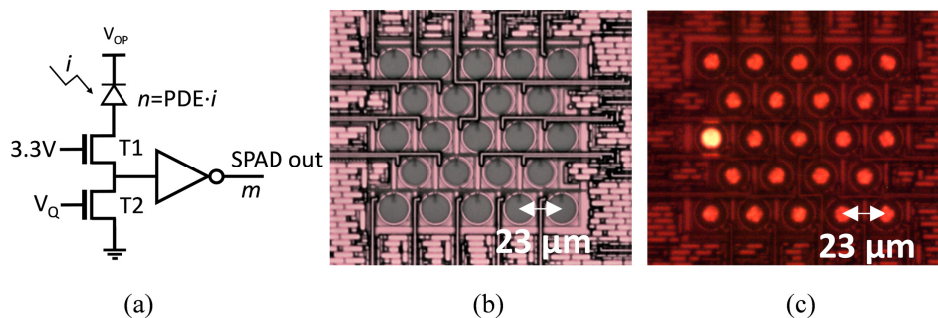


Fig. 1. (a) Pixel schematic with cascode transistors to allow high  $V_{\text{OP}}$  values. (b) Chip micrograph with external illumination. (c) Chip micrograph with SPADs operating with low quenching resistance to enable simultaneous avalanching and light emission. Light is emitted from the active area, in the pixel centers. The bright outlier on the right is a “hot” pixel.

$V_{\text{Q}}$  is a global signal connected to all pixels. Passive recharge is implemented with a constant bias, while other recharge mechanisms (event-driven or clock-driven) can be performed by driving  $V_{\text{Q}}$  with a FPGA.

The inverter performs capacitive isolation so as to reduce afterpulsing and timing jitter. SPAD output signals are fed into digital outputs with programmable output strength. Pixels are arranged in a hexagonal pattern to accommodate a round shape (desirable in ISM) and increase the fill factor (the vertical pitch is shorter than in a rectangular arrangement). The radial pixel pitch is  $23 \mu\text{m}$ , and the radius of the active area is  $5.85 \mu\text{m}$ . This translates into a native fill factor of 23.5%. Chip micrographs are shown in Fig. 1. Panel (c) shows a uniform light emission test.

Figure 2 shows the two optical setups for the experiments conducted in this paper. The pulsed red laser is only used for the jitter timing measurement, while the CW light path is used for the other experiments.

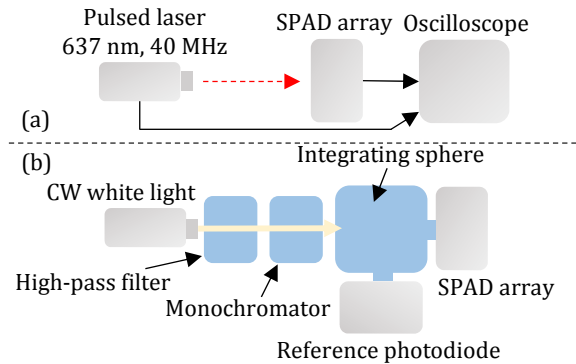


Fig. 2. Optical setups. (a) The pulsed laser is directly coupled with the SPAD array whose rising edge position in time is compared to the laser reference clock. We can estimate the jitter by creating a histogram of the time difference between these two signals. (b) A PDP setup with an integrating sphere creating uniform, dispersed light outputs for the reference photodiode and the SPAD array.

Figure 3 shows PDP measurements, with the peak PDP increased by 10% compared to the previous design [22]. The PDP measurement setup is illustrated in Fig. 2(b). We used anode connections at the border of the SPAD active area, whereas the previous design featured a  $6 \times 2 \mu\text{m}$  metal shielding on top of the  $113 \mu\text{m}^2$  active area. Additionally, p-well diffusion might increase the effective active area, thus falsely yielding higher PDP values for smaller active areas (original  $6 \mu\text{m}$  radius compared to  $5.85 \mu\text{m}$  in the current design). This CMOS SPAD array outperforms state-of-the-art CMOS SPAD arrays in terms of peak PDP and features a wider spectral response with an enhanced red sensitivity [18–21]. This red enhancement is particularly difficult to achieve with substrate isolated SPADs that can be used as arrays. Substrate shared SPADs with higher PDP do exist [25], but are not demonstrated as large 2D arrays. Substrate sharing induces high crosstalk and limits the applicability of such SPADs in arrays.

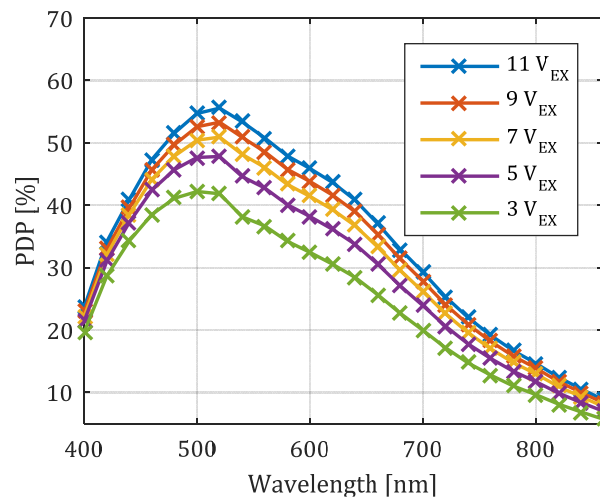


Fig. 3. PDP measurements at different  $V_{EX}$ . The peak PDP at 520 nm is 55%, while PDP is higher than 40% between 440 and 640 nm.

We tested the breakdown uniformity with the excess count rate method [26], with the setup from Fig. 2.  $V_{OP}$  was varied from 25.32 V to 25.78 V in steps of 20 mV, and the count rate linearly extrapolated to counts equal zero, which allows to determine, in principle, the

breakdown voltage  $V_{BD}$ . Measurements yield an average  $V_{BD}$  value of 21.95 V with a standard deviation of 56 mV. However, since the PDP is not linear with respect to  $V_{OP}$ , this procedure does in fact underestimate the breakdown voltage: data reveal a narrower distribution of counts at around 23  $V_{OP}$ , which was also measured to be the breakdown voltage for a testing SPAD without integrated electronics (using pulse emergence voltage as the breakdown). A 56 mV breakdown voltage variation yields a PDP change of 0.17%.

The photon response nonuniformity (PRNU) was tested using the setup in Fig. 2(b). At 1.5% of dynamic range (to avoid count saturation due to dead time, operating with passive recharge) and 7 V of excess bias, we measured a PRNU of 1.6%, expressed as the ratio between the standard deviation and the average count number. Active clock-driven recharge, detailed below, yields the same results.

Figure 4(a) represents the DCR distribution, which allows to quantify the percentage of “hot” pixels. At 3.3 V of excess bias, less than 4% of pixels have a DCR higher than 100 cps, whereas less than 2% have a DCR higher than 1 kcps. This data indicate very good DCR uniformity. Figure 4(b) indicates a higher band-to-band tunneling and/or higher trap-assisted tunneling for  $V_{EX} = 11$  V with respect to  $V_{EX} = 9$  V. While trap-assisted thermal generation dominates the DCR at temperatures higher than 20°C, band-to-band tunneling is limiting a further DCR decrease at temperatures below 0°C and  $V_{EX} = 9$  V, where DCR reaches a very low 5 cps. For  $V_{EX} = 11$  V, it seems that there is increased trap-assisted tunneling that can be further reduced while cooling below 0°C.

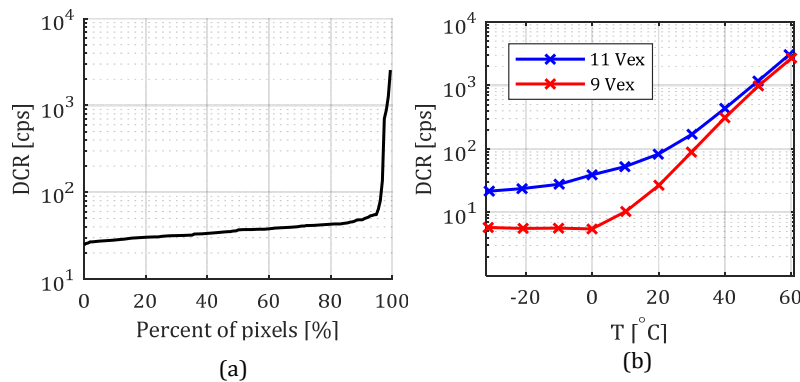


Fig. 4. (a) DCR distribution of 7 combined chips (161 SPADs in total), at  $V_{EX} = 3.3$  V. Better DCR uniformity is achieved by using smaller active areas or by cooling [27]. (b) DCR reduces exponentially with temperature, but reaches a plateau at around 0°C. DCR for this device then becomes dominantly generated by band-to-band tunneling, which is less affected by cooling.

The timing jitter FWHM is 134, 128 and 123 ps for 7, 9 and 11 V of excess bias, respectively. The measurements are done with the setup in Fig. 2(a). When compared to the previous implementation, the timing jitter increased from 100.8 ps [22] to the current 123 ps. A cascode transistor in series increases the output resistance and could increase the timing jitter due to a slower rising edge. Additionally, the SPAD rising edge is compared to a fixed inverter threshold, whereas Veerappan [22] used an oscilloscope with adjustable thresholding. The measurements are performed with a 40 MHz red laser (637 nm, PiL063F from Advanced Laser Diode Systems, Berlin, Germany), a SPAD dead time of 100 ns and a count rate at 100 kcps. These parameters were chosen to limit pile-up.

Afterpulsing was measured to be 0.1% with 50 ns dead time, using a SPAD operating at 11 V of excess bias and 0.75 V of quenching bias. Crosstalk between two adjacent pixels was measured to be 0.09% with the same biasing. We used an oscilloscope to construct inter-arrival time histograms between consecutive pulses in the same pixel (afterpulsing) and in adjacent pixels (crosstalk). A low afterpulsing is achieved due to capacitive isolation of the SPAD anode through an inverter. Deep trench isolation limited optical crosstalk, but also

reduced the fill factor. Increased correlations are present only in the first  $\Delta T$  bin, indicating optical coupling. If electrical coupling was present, increased correlations would be featured at longer  $\Delta T$  as well.

### 3. Recharge-dependent dynamic range extension

We used the setup from Fig. 2(b) to examine the dynamic range when applying different recharge mechanisms. We employed a fixed  $V_Q$  for passive recharge, and a FPGA-controlled  $V_Q$  for event-driven and clock-driven recharge (see Fig. 5). In passive recharge, the dead time is controlled by a constant resistance of transistor T2. In *event-driven recharge*, one keeps the resistance of the transistor T2 high for a fixed delay  $T_{dead}$  following a photon detection, and then applies a short pulse at  $V_Q$  creating a T2 resistance decrease and a prompt recharge of the SPAD. The duration of the applied  $V_Q$  pulse is 20 ns. *Clock-driven recharge* is entirely clock based (with a period of  $T_{dead}$  and with an 8% duty cycle), whereby the changes in T2 resistance are independent from the activity of the SPAD.

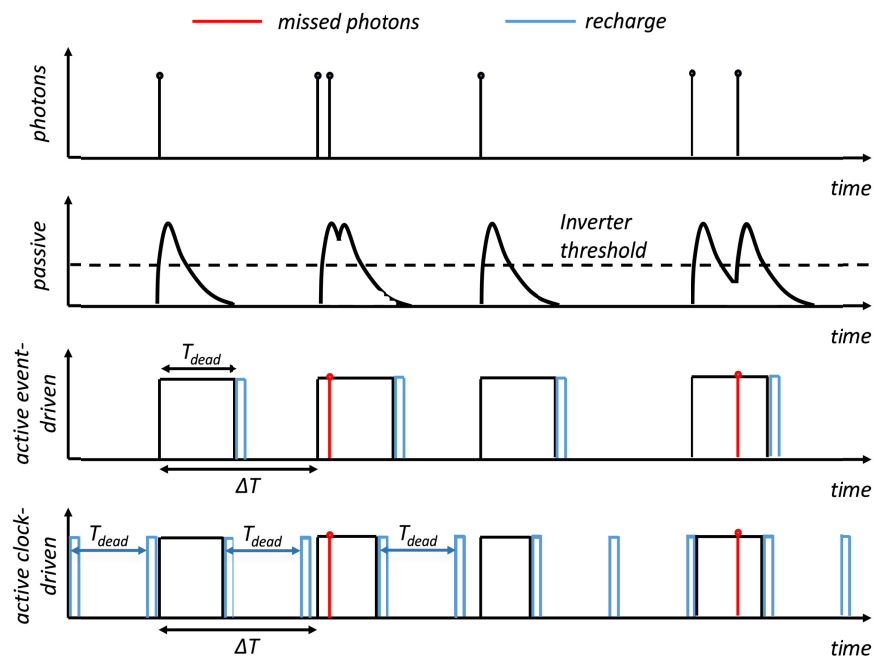


Fig. 5. Schematic representation of different recharging mechanisms. In the active event-driven mode, recharge is performed after a fixed delay  $T_{dead}$  following a photon detection. In clock-driven mode, the SPAD is periodically recharged. The period of this recharge is  $T_{dead}$ . Note that the time interval between two photon detections,  $\Delta T$ , can be shorter than  $T_{dead}$  in the clock-driven mode.

It is well known that the output measured count rate  $m$  (i.e. the actual count rate at the SPAD output node in Fig. 1(a)) is reduced with respect to the detected count rate  $n$  [27,28], due to the dead time  $T_{dead}$ , whereby each SPAD recharge mechanism features a different  $m$  to  $n$  response.

Passive recharge follows  $m = n \times \exp(-nT_{dead})$ , featuring a so-called paralyzable response. Due to the dead time spread (up to 25% in standard deviation) over the array (caused by transistor T2 deviations in resistance) and the non-monotonic photon response ( $m$  firstly increases and then decreases), the corresponding useful dynamic range is limited to around 7.8 Mcps of detected count rate  $n$ , when using  $T_{dead} = 50$  ns. At higher detected count rates, the SNR decreases and the response over the array becomes nonuniform due to pixel-to-pixel dead time variations (see response curves in Fig. 6).



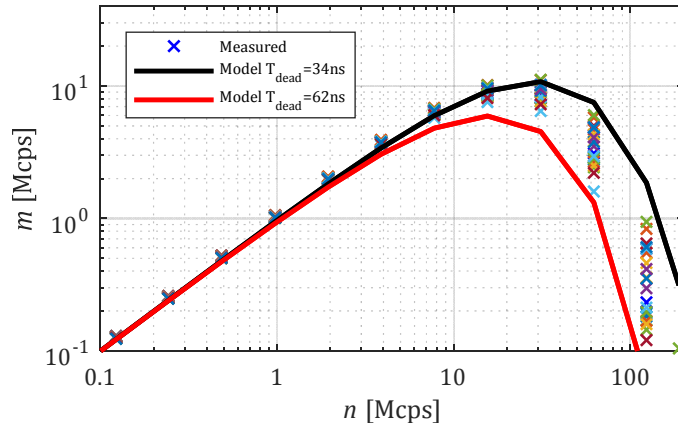


Fig. 6. Measured ( $m$ ) vs. detected ( $n$ ) count rate responses for each pixel over the array for passive recharge, following  $m = n \times \exp(-nT_{dead})$ . Up to  $n = 7.8$  Mcps,  $m$  response curves are uniform and indicate the useful dynamic range.  $T_{dead}$  is between 34 and 62 ns, for  $V_Q = 0.75$  V. At higher detection rates, the measured count rates  $m$  become largely nonuniform due to differences in  $T_{dead}$ .

Active recharge methods employ FPGAs or integrated electronics to control  $T_{dead}$  and feature a monotonic response. Active event-driven recharge features a non-paralyzable response with  $m = n/(1 + nT_{dead})$  [27,29]. Active clock-driven recharge follows  $m = (1 - \exp(-nT_{dead}))/T_{dead}$  [27]. A Spartan 6 FPGA system [30,31] was used to generate signals driving the quenching and recharge transistor T2 depicted in Fig. 1(a), based on a logic clock with 10 ns period. Figure 7 shows measured and corrected count rate responses compared to the presented theoretical models, with  $T_{dead}$  equal to 250 ns and an integration time  $t$  of 100 ms. The deviations of the measured curves (event-driven recharge) from the theoretical ones are likely due to the imperfections of the recharge implementation, which is not integrated at pixel level within the chip, but placed on a separated printed circuit board (PCB); this entails a relatively long recharge pulse of 20 ns. This recharge implementation also entails a certain probability that the SPAD is not going to be recharged during the recharge period. We believe that a more complex integrated recharge scheme would yield a precise matching of the event-driven measured data with the theory.

While  $m$  is limited by  $1/T_{dead}$ ,  $n$  required to saturate  $m$  is much larger than  $1/T_{dead}$ . This is a consequence of the exponential distribution of photon arrivals, where photons have a higher probability to be close in time and impinge onto the detector within  $T_{dead}$  after a primary detection. This increases the suppression of detected counts at high illumination levels (when the average photon inter-arrival time  $\Delta T$  is comparable to  $T_{dead}$ ).

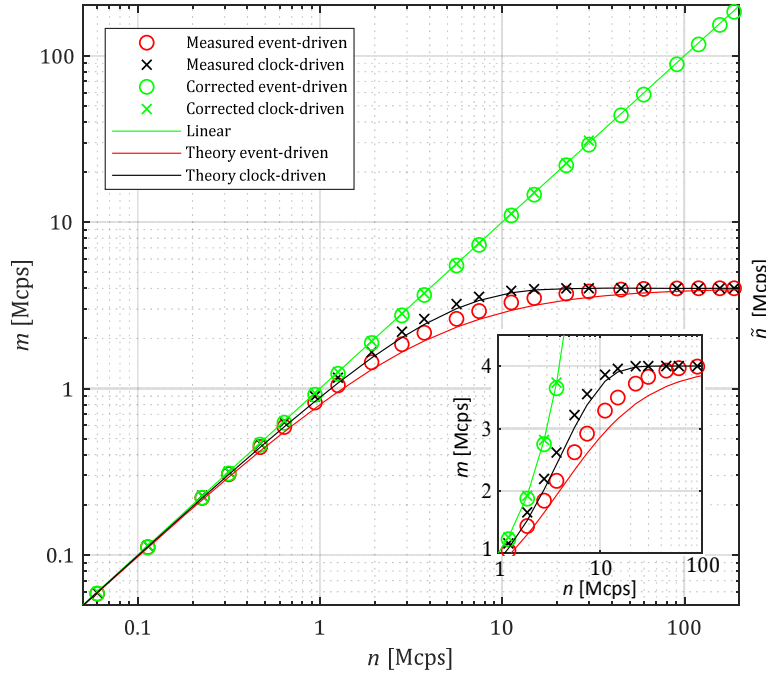


Fig. 7. Measured ( $m$ ) vs. detected ( $n$ ) count rate responses for active event-driven and clock-driven recharge with  $T_{dead} = 250$  ns and an integration time  $t$  of 100 ms. A longer  $T_{dead}$  than in the previously illustrated results is used to limit the recharge timing uncertainty with respect to the SPAD pulse to 2% ( $\pm 5/250$  ns). The responses are linearized back (i.e. measured  $m$  is corrected to  $\tilde{n}$ :  $m \rightarrow E(n) \rightarrow \tilde{n}$ , which is plotted on the right hand scale) so as to match the detected count rate  $n$  (green). Active clock-driven recharge saturates (i.e. reaches  $m = 4$  Mcps) at  $n = 44.86$  Mcps, and active event-driven recharge saturates at  $n = 224.1$  Mcps. Black and red curves show theoretical curves, whose behavior is calculated using the analytic models. The inset shows a detail of the curves with linear y axis.

While the aforementioned saturation behavior is sometimes seen as a limiting factor due to the loss of linearity at high count rates, researchers in the imaging community have started to use it to their advantage, in order to effectively *extend the dynamic range*, as proposed in the *gigavision* and *quanta image sensors* theory [16,32]. However, due to the clocked readout nature of charge accumulating CMOS imagers, theoretical work concentrated on active clock-driven recharge [16,17,32,33]. The SPAD community followed this approach [34]. However, there is no standardized quantitative measure for the extension of the SPAD dynamic range due to the suppressed count response, mainly citing  $1/T_{dead}$  [35] or the linearity of the corrected response [36] as limiting the dynamic range. Here, we propose a quantitative SNR measure for the extended dynamic range.

The SNR in shot noise limited detection (assuming no dark noise) can be defined as:

$$\text{SNR} = 20 \log_{10} \left( \frac{n_i}{\sqrt{n_i}} \right) = 20 \log_{10} (\sqrt{n_i}), \quad (1)$$

where  $n_i$  is the detected count number over an integration time  $t$ ,  $n_i = n \times t$ . Due to the count saturation, the measured count number  $m_i = m \times t$  has a lower shot noise than  $\sqrt{n_i}$ . The measured count number variance for event-driven passive recharge is [37]:

$$\sigma_{m_i}^2 = mt \left( 1 - 2mT_{dead} + \frac{mT_{dead}^2}{t} \right) \quad (2)$$

The variance for event-driven active recharge is [37]:

$$\sigma_{m_i}^2 = \frac{n \times t}{(1 + nT_{dead})^3} + \left(1 + \frac{2}{3}nT_{dead} + \frac{1}{6}(nT_{dead})^2\right) \times \frac{(nT_{dead})^2}{(1 + nT_{dead})^4} \quad (3)$$

Finally, for active clock-driven recharge the variance is given by the binominal distribution:

$$\sigma_{m_i}^2 = \frac{t}{T_{dead}} (1 - e^{-nT_{dead}}) e^{-nT_{dead}} \quad (4)$$

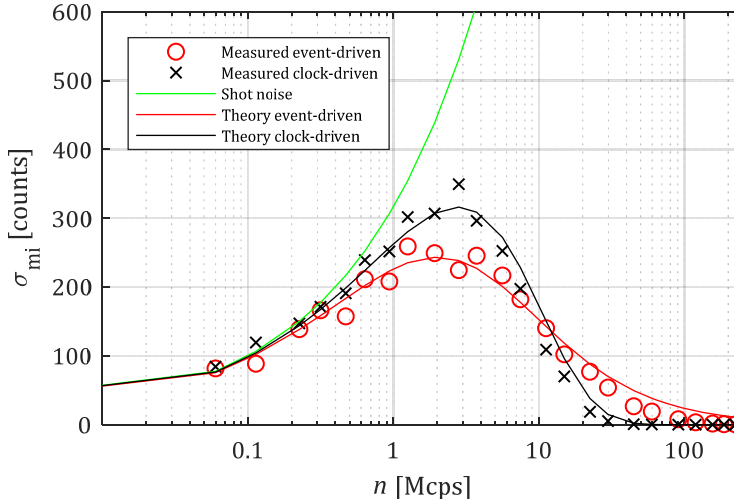


Fig. 8. Standard deviation  $\sigma_{m_i}$  of the measured count number  $m_i$  due to shot noise and the count saturation behavior, for active event-driven and clock-driven recharge,  $T_{dead} = 250$  ns and integration time  $t = 100$  ms. Green shows the reference, i.e. the standard deviation due to shot noise. The theoretical curves are modeled with Eqs. (3) and (4).

A clear difference between the active recharge mechanisms is presented in Fig. 8, which shows the standard deviation of 100 measurements of  $m_i$  with the same impinging photon flux. The standard deviation ( $\sigma_{m_i}$ ) of the measured counts  $m_i$  is lower than  $\sqrt{n_i}$  (shown with a green curve in Fig. 8). However, if we want to extend the dynamic range by linearizing back to  $n_i$ :

$$m_i \rightarrow E(n_i) \rightarrow \tilde{n}_i, \quad (5)$$

we should instead calculate  $\sigma_{\tilde{n}_i}$ , the standard deviation of  $\tilde{n}_i$ , determined by:

$$\sigma_{\tilde{n}_i}^2 = \left( \sigma_{m_i} \frac{\partial n_i}{\partial m_i} \right)^2 \quad (6)$$

Linearization is in our case done with a look-up-table from a calibrated  $f(n) = m$ , where  $n$  is measured with the reference photodiode from Fig. 2(b), and  $m$  is measured with the SPAD array. Equations (2)-(6) indicate a  $\sigma_{\tilde{n}_i}$  larger than  $\sqrt{\tilde{n}_i}$  when  $n$  is approaching  $1/T_{dead}$ , as shown in Fig. 9(a). A larger  $\sigma_{\tilde{n}_i}$  will, of course, reduce the signal to noise ratio of the estimated detected counts  $\text{SNR}\tilde{n}_i$ , as shown in Fig. 9(b). We assume a wide recharge pulse (and passive quenching) to be the cause for the event-driven recharge  $\text{SNR}\tilde{n}_i$  to drop at significantly lower  $n$  than modeled. A maximally allowed drop in SNR will be taken to estimate the dynamic range: we propose an SNR decrease of  $-3\text{dB}$  (as customary to estimate the frequency bandwidth of analog amplifiers) compared to the SNR of a hypothetical sensor with a linear

response, saturating at  $1/T_{dead}$ . The SNR at this point is defined as  $20\log_{10}[\sqrt{t/T_{dead}}]$ , i.e. 56 dB in the example. One can also choose to compare the SNR drop to the SNR of a sensor with unlimited dynamic range, i.e. the SNR defined by Eq. (1). We feel that a quantitative SNR drop is a better way to determine the maximum detectable count rate rather than taking a point at which  $m$  reaches a predefined 99% of the saturation value  $1/T_{dead}$  [17]. The maximum  $\tilde{n}_i$  at which the SNR drops by  $-3$ dB compared to a perfectly linear detector is then denoted as  $\tilde{n}_{i\max}$ :

$$\text{SNR}(\tilde{n}_{i\max}) = 20\log_{10}\left(\frac{\tilde{n}_{i\max}}{\sigma_{\tilde{n}_i\max}}\right) = 20\log_{10}\left(\sqrt{t/T_{dead}}\right) - 3\text{dB} \quad (7)$$

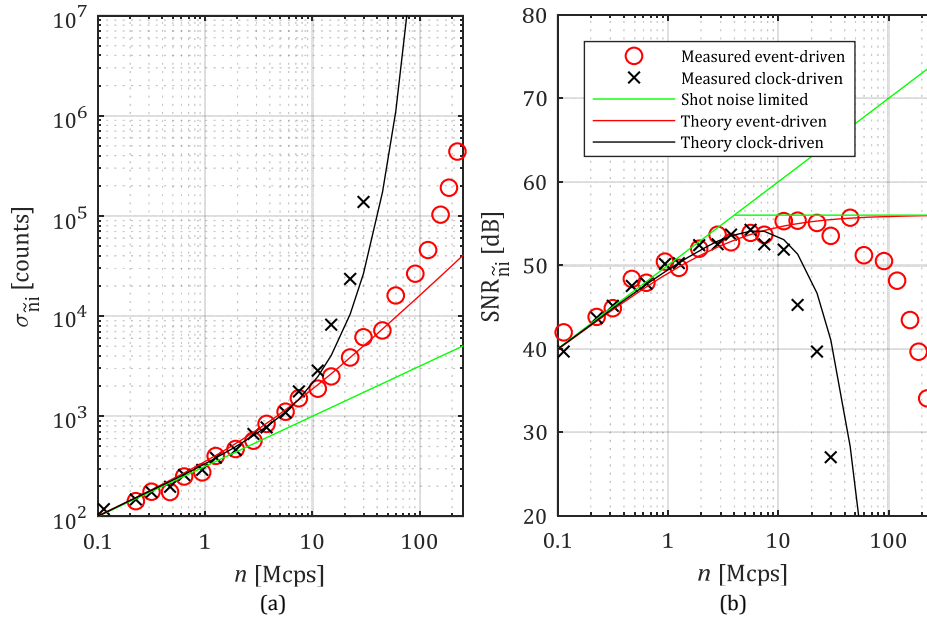


Fig. 9. (a) Standard deviation  $\sigma_{\tilde{n}_i}$  of the estimated detected count number  $\tilde{n}_i$  due to shot noise and the count saturation behavior, for active event-driven and clock-driven recharge, with  $T_{dead} = 250$  ns and integration time  $t = 100$  ms. We estimate  $\tilde{n}$  from  $m$  and calculate the standard deviation of  $\tilde{n}$ . (b) SNR  $\tilde{n}_i$  ( $\tilde{n}/\sigma_{\tilde{n}_i}$ ) of the detected count number  $\tilde{n}_i$ , which is, for both recharge mechanisms, comparable to the classical shot noise  $\sqrt{\tilde{n}_i}$  up to  $n \sim 1$  Mcps. SNR  $\tilde{n}_i$  is compared to a hypothetical sensor with linear response up to  $1/T_{dead} = 4$  Mcps, with a maximum SNR  $\tilde{n}_i$  of  $20\log_{10}[\sqrt{t/T_{dead}}] = 56$  dB. The dynamic range does extend to  $\tilde{n}_{i\max} = 7$  Mcps and  $\tilde{n}_{i\max} = 55$  Mcps for clock-driven and event-driven recharge, respectively.

The dynamic range is then defined as the ratio between the maximum and minimum detectable counts,  $\tilde{n}_{i\max}$  and  $\tilde{n}_{i\min}$ .  $\tilde{n}_{i\min}$  is limited by the variations in DCR.

$$\text{DR} = 20\log_{10}\left(\frac{\tilde{n}_{i\max}}{\tilde{n}_{i\min}}\right) = 20\log_{10}\left(\frac{\tilde{n}_{i\max}}{\sigma_{DCR}}\right) = 20\log_{10}\left(\frac{F \times t / T_{dead}}{\sqrt{t \times DCR}}\right), \quad (8)$$

where  $F$  is a factor for the count saturation in SPADs. Table 1 gives theoretical values for  $F$  for different recharge mechanisms. We included three different measures for the maximum detectable count rate. The first column indicates the maximum count rate if we allow for 10% nonlinearity without corrections. The second column indicates at which detected count rate  $n$  we saturate the SPAD output  $m$ , i.e.,  $m$  reaches  $0.99/T_{dead}$ . The last column is derived from Eqs. (2)-(6) and indicates  $n$  at which the SNR decreases by  $-3$ dB (as previously discussed).

**Table 1. Theoretical  $F$  factors to estimate  $\tilde{n}_{\max}$  for different recharge mechanisms and criteria. Factors are provided as  $F/T_{dead} = \tilde{n}_{\max}$ . Please note that Eq. (3) introduces an SNR dependent on  $t/T_{dead}$ , thus  $F$  (SNR decrease) for active event-driven is indicated for typical  $t/T_{dead}$  ratios (255 to 1023, so as to reach an 8 to 10 bit image depth).**

Recharge \ Criteria	10% nonlinearity <sup>a</sup>	Response saturation <sup>b</sup>	SNR decrease <sup>c</sup>
Passive	$0.105/T_{dead}$	$0.864/T_{dead}^d$	$0.39/T_{dead}^e$
Active event-driven	$0.111/T_{dead}$	$99/T_{dead}$	$36/T_{dead}$ to $75/T_{dead}^f$
Active clock-driven	$0.214/T_{dead}$	$4.605/T_{dead}$	$2.842/T_{dead}$

<sup>a</sup> $n$  at which  $m = 0.9n$ , <sup>b</sup> $n$  at which  $m = 0.99/T_{dead}$ , <sup>c</sup> $n$  at which SNR<sub>ni</sub> drops by 3dB, <sup>d</sup> $n$  at peak  $m = 1/eT_{dead}$ , since passive recharge never reaches  $0.99/T_{dead}$ , <sup>e</sup> $n$  at peak SNR $\tilde{n}_i$ , since passive recharge never reaches  $\sqrt{(t/T_{dead}/2)}$ , <sup>f</sup>for  $t$  from  $255 \times T_{dead}$  to  $1023 \times T_{dead}$ , to yield an 8- to 10-bit image.

#### 4. Afterpulsing and response linearity with clock-driven recharge

To learn more about the properties of the recharge mechanisms, we recorded FPGA-based photon inter-arrival ( $\Delta T$ ) histograms with a bin width of 10 ns. Figure 10 shows inter-arrival time histograms for different detection count rates (active recharge only). Each curve represents a measurement point in Figs. 7-9. A steeper curve with a high occurrence at short  $\Delta T$  indicates a higher detection count rate.

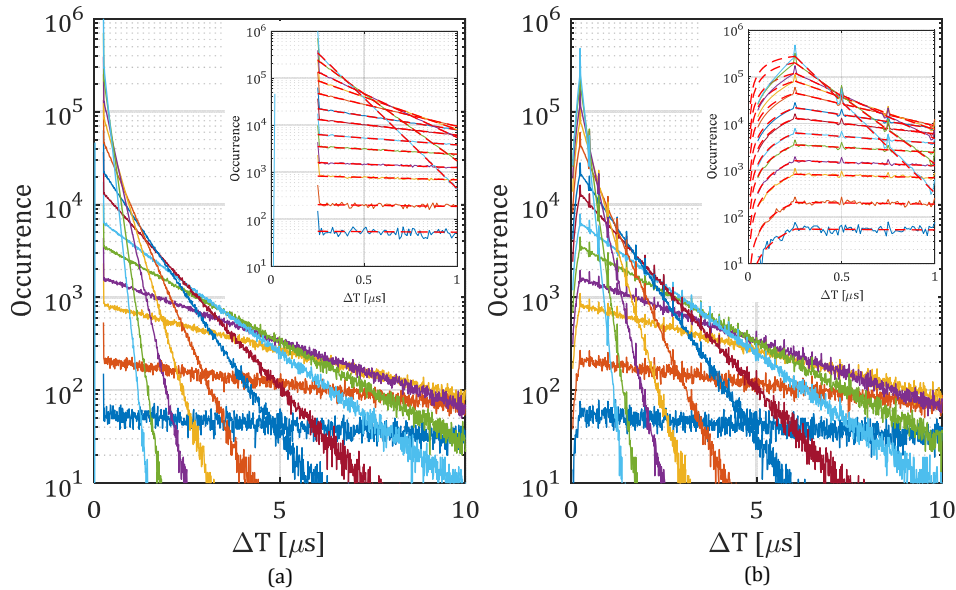


Fig. 10. Photon inter-arrival histograms for active (a) event-driven and (b) clock-driven recharge mechanisms with  $T_{dead} = 250$  ns dead time. Each curve is a histogram of the same pixel at different count rates ranging from  $n = 0.06$  Mcps to  $n = 7.5$  Mcps. The insets show a zoomed version of Fig. 10. The red discontinuous lines represent exponential fits to estimate afterpulsing.

Figure 10 shows that active recharge increases the occurrence of inter-arrival times at  $\Delta T = 250$  ns, the dead time. Figures 10(a) and (b) show data for active event-driven and clock-driven recharge, respectively. Due to FPGA-based recharge, we used relatively long recharge pulses of 20 ns, thus increasing the probability of a photon detection during the recharge itself (the avalanche is not being quenched until the end of the recharge). This effect is also known as the *twilight effect* [38,39]. The photons detected from 230 ns ( $T_{dead} - \text{recharge pulse width}$ )

to 250 ns will be transferred to the  $\Delta T = 250$  ns bin. The twilight effect falsely induces an increased afterpulsing estimate. True afterpulsing (that is caused by charge trapping) is constant over different detection rates, while twilight “afterpulsing” increases with detection rates. We measured a twilight “afterpulsing” increase from 0.3% to 5.6% for a detection count rate increasing from 0.11 Mcps to 2.75 Mcps for active event-driven recharge.

Estimating afterpulsing for active clock-driven recharge requires additional curve fitting, as shown in the inset displayed in Fig. 10(b). Since the recharge is totally clock-driven, SPAD pulses can be very short. Note here the discrepancy between SPAD pulse duration and  $T_{dead}$ : while in event-driven recharge they are equivalent, in clock-driven mode the SPAD pulse duration follows a probability function, and  $T_{dead}$  is merely the recharge period. The rising edge position of a SPAD pulse features an exponential distribution (photon arrival). The falling edge is determined by the periodic recharge. The inter-arrival histogram for very short  $\Delta T$  values is thus determined by the probability that one or more photons are going to be detected right after a recharge, corresponding to the cumulative probability density function of an exponential:

$$\text{pdf}(\Delta T) = 1 - e^{-\frac{\Delta T}{\tau}} \text{ for } \Delta T < T_{dead}, \quad (9)$$

where  $\tau$  is the average  $\Delta T$  (inter-arrival time between detected photons) determined by the detection rate. The actual occurrence needs to be normalized to reach the exponential distribution at  $T_{dead}$  and is:

$$\text{occurrence} = \frac{I_0 e^{-\frac{T_{dead}}{\tau}}}{1 - e^{-\frac{T_{dead}}{\tau}}} \left( 1 - e^{-\frac{\Delta T}{\tau}} \right) \quad (10)$$

At high count rates, measured inter-arrival histograms start to diverge from the models, forcing a periodic SPAD output with almost constant  $\Delta T$  at very high count rates. One would be interested in afterpulsing caused by the possibility of having short SPAD pulses. We did not observe such an afterpulsing behavior, i.e. occurrence higher than what forecasted by the Eq. (10) model, for  $\Delta T < 250$  ns. Firstly, SPAD afterpulsing in this SPAD is 0.1% for SPAD pulse durations from 50 ns onwards. Secondly, the afterpulsing probability for SPAD pulses below  $T_{dead}$  is multiplied by Eq. (9), thus reducing it.

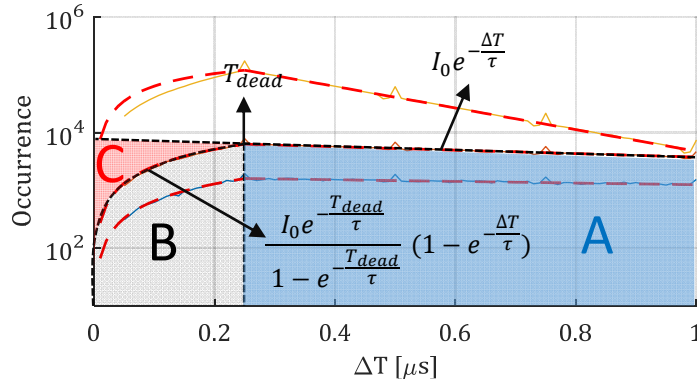


Fig. 11. A schematic explanation of the difference between active event-driven and clock-driven recharge. Event-driven recharge cuts all counts with  $\Delta T < T_{dead}$ , while clock-driven recharge still contains a part of counts with  $\Delta T < T_{dead}$  (B area).

Figure 11 provides an intuitive explanation why active clock-driven recharge has better linearity and saturates before active event-driven recharge. Light with a Poisson distribution (with exponentially distributed  $\Delta T$ ), i.e. detected counts  $n_i$  are contained in the area A + B +

C. Measured counts  $m_i$  with clock-driven recharge are contained in A + B, while event-driven counts  $m_i$  just in A. Dotted red curves are fitted following Eq. (10) up to  $\Delta T = T_{dead}$  and with  $I_0 e^{-T_{dead}/\tau}$  from  $\Delta T = T_{dead}$  onwards.

## 5. Conclusion

We introduced a full set of models for different recharge modes and included measurements. Table 1 offers a standardized set to evaluate the SPAD dynamic range when response corrections are applied, extending the usually referenced  $1/T_{dead}$  by a factor of  $\times 75$ . Event-driven recharge, in particular, gives a  $\times 12.7$ - $26.4$  ( $\times 7.9$  measured) higher dynamic range when compared to clock-driven (also called *quanta image sensor*). SPADs incrementing a 1-bit memory and resetting the memory synchronously to the SPAD activity (or an actively recharged SPAD driving a multibit memory) could thus be used for novel high dynamic range imaging, instead of the classical *quanta image sensor* approach [17,34]. An image sensor designed in [40], but supplemented with active event-driven recharge, could use the methodology proposed in our paper to implement a large photon-counting array with extended dynamic range.

Clock-driven recharge was shown to have better linearity. To the best of our knowledge, direct clock-driven SPAD recharge with less than 1  $\mu$ s period was never performed, mainly due to the possible increase in afterpulsing. In this work, we provide a novel methodology to estimate afterpulsing in such operating modes, directly from inter-arrival time histograms. We believe that the significance of this work also lies in the deep understanding of the afterpulsing and dynamic range trade-offs that the use of clock-driven and event-driven recharging implies. These trade-offs become even more significant when using alternative biasing techniques to boost the excess bias voltage, which is the trend in modern SPAD imaging designs.

Table 2 presents a comparison of state-of-the-art CMOS SPAD detectors. Using a cascode transistor combination for quenching and recharge, the SPAD array presented in this work operates at a high 11 V excess bias, with 55% peak PDP at 520 nm and  $>40\%$  from 440 to 640 nm. The array shows an extremely low DCR of 5 cps for 9 V excess bias and 0°C. This high-end CMOS SPAD array with integration flexibility endows different parallelization approaches in confocal microscopy. Further designs however need to increase the fill factor (with the use of microlenses).

**Table 2. A comparison of state-of-the-art CMOS SPAD detectors.**

Sensor	This work	[19]	[18]	[25]	[41]
Info	180 nm CMOS process	350 nm CMOS process	150 nm CMOS process	130 nm CMOS process	90 nm CMOS process
Detection efficiency maximum	55%@ 520nm	50%@ 420nm	31%@ 450nm	72%@560nm	44%@ 690nm
Spectral range (above 20%) [nm]	400-750	400-650	400-600	350-850	500-850
Native fill factor [%]	23.5	$\sim 13^a$	39.9	not demonstrated as large 2D array <sup>b</sup>	not demonstrated as large 2D array <sup>b</sup>
Size of active area [ $\mu\text{m}^2$ ]	107	78.5	97	50	32
DCR[cps/ $\mu\text{m}^2$ ]	0.32	0.07	0.4	0.36	0.31
Afterpulsing [%]	0.1	1.3 [42]	0.85	4	0.375
Timing jitter [ps]	120	77	52	60	51

<sup>a</sup>estimated from [43], <sup>b</sup>p-substrate-shared SPAD implementation, assumed not to be used as array due to large crosstalk.

**Funding**

Netherlands Organization for Scientific Research (NWO) (13916).

**Acknowledgments**

The authors would like to acknowledge Dr. Samuel Burri for the FPGA Spartan 6 readout board. Thanks to Chao Zhang and Scott Lindner for the assistance with the chip tapeout and the discussions about the SPAD response characteristics. Thanks to Dr. Chockalingam Veerappan for the discussions about the SPAD design and characterization. We are also grateful to Dr. Ron Hoebe for his suggestions on the design and application.