SPAD Pixel With Sub-NS Dead-Time for High-Count Rate Applications

Fabio Severini, *Student Member, IEEE*, Iris Cusini, Davide Berretta, Klaus Pasquinelli, Alfonso Incoronato, *Student Member, IEEE*, and Federica Villa, *Member, IEEE*

***Abstract*—Single-photon avalanche diode (SPAD) exploitation in high-flux applications is often hindered by the trade-off between the SPAD dead-time and afterpulsing probability. In this paper, we present the architecture and the experimental characterization of two chips including a novel SPAD sensing, and readout scheme designed to minimize dead-time (1.78 ns and 0.93 ns respectively) and afterpulsing probability (0.14% maximum). We have coupled this architecture with high-performance SPADs obtaining an extremely stable dead-time (6.44 ps_{rms} jitter) that can be easily regulated through an external voltage. Thanks to its compact size, this novel pixel architecture can be easily integrated within high-resolution SPAD arrays for GHz applications.**

***Index Terms*—Afterpulsing, dead-time, high-flux operation, photon counting, single-photon avalanche diode (SPAD).**

I. Introduction

SINGLE-PHOTON Avalanche Diodes (SPADs) are single-photon detectors employed in many applications which benefit from their sensitivity at single-photon level [1], timing resolution [2], and fast-gating capability (i.e., the possibility to be enabled and disabled with sub-nanosecond edges) [3]. Furthermore, SPADs can be integrated, together with the front-end and processing electronics, in CMOS technologies, thus enabling the possibility to develop SPAD arrays (e.g., for imaging applications), both in planar (SPAD and electronics on the same wafer) [4] and 3D-stacked processes (two tiers vertically integrated, one for SPADs and one for electronics) [5].

SPADs are intrinsically digital detectors that, working in a metastable region above their breakdown voltage, can generate a macroscopic current upon photon detection. The main drawback of SPADs with respect to linear detectors, which feature an output proportional to the intensity of the incoming light (e.g., Avalanche Photodiodes - APDs), is the presence of a dead-time after each photon detection, that limits the maximum count rate.

To quench the avalanche current, the SPAD must be brought below its breakdown voltage and, afterward, reset to the initial condition. Since some avalanche charges can be trapped by defects in the SPAD junction and released at later times (*afterpulsing phenomena*), the SPAD is kept below the breakdown voltage for a time interval called hold-off time, before being reset. Afterpulsing noise is usually the limiting factor in applications requiring very short dead-time [6], and mostly depends on the quality (number of defects) of the lattice [7]. Furthermore, afterpulsing probability depends on the SPAD structure itself, specifically on its associated junction capacitance (C_j) and other stray capacitances at the quenching node (C_{stray}), and on its excess bias (V_{EX}), which corresponds to the voltage range at the quenching node. Indeed, the total number of carriers involved in the avalanche (defined avalanche charge or pulse charge) can be approximated as in [8]:

$$Q_{pc} = V_{EX} \cdot (C_j + C_{stray}) . \tag{1}$$

At the circuital level, afterpulsing can be reduced by designing front-ends able to minimize the pulse charge. Indeed, fast quenching/reset circuits can minimize the charge across the SPAD thus decreasing the probability of trapping electrons [9]. Beyond limiting the afterpulsing probability, reducing the pulse charge is a fundamental step to contain the power consumption, which is one of the critical points of developing dense SPAD arrays, especially when working at high-photon fluxes (i.e., many pixels active at the same time). The power consumption due to a photon detection, i.e., power per detection (P_{pd}) can be approximated as reported in [8]:

$$P_{pd} = Q_{pc} \left(V_{BD} + \frac{V_{EX}}{2} \right) n_t \tag{2}$$

where, V_{BD} is the SPAD breakdown voltage and n_t is the total photon rate. Thus, when working at high photon rates, conceiving quenching circuits for reducing the pulse charge is an effective way for reducing the power consumption.

The SPAD front-end favors either the counting rate maximization or the afterpulsing minimization depending on the chosen topology. Indeed, quenching and reset operation can be performed either with Passive or Active Quenching Circuits (PQCs or AQCs) [8]. PQCs architectures are favored to minimize the afterpulsing effect, since quenching is usually faster thanks to the presence of the quenching resistance, whose value influences the dead-time, defined here as the exponential discharge of parasitic capacitances on the quenching resistance itself. On the other hand, for SPADs with AQCs the overall dead-time is the sum of quenching, hold-off, and reset times. AQCs are preferred

Manuscript received June 7, 2021; revised October 1, 2021; accepted October 29, 2021. Date of publication November 4, 2021; date of current version November 18, 2021.

The authors are with the Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, 20133 Milano, Italy (e-mail: fabio.severini@polimi.it; iris.cusini@mail.polimi.it; davide.berretta@polimi.it; klaus.pasquinelli@polimi.it; alfonso.incoronato@polimi.it; federica.villa@polimi.it).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JSTQE.2021.3124825.

Digital Object Identifier 10.1109/JSTQE.2021.3124825

when short and well-controlled dead-times are required, since these architectures swiftly quench and reset the SPAD, and the hold-off can be fixed by design, being independent from stray capacitances. Conversely, in AQCs part of the avalanche charge is discharged to ground before the actual active quenching process, yielding to worsened afterpulsing performance with respect to PQCs. Mixed passive-active approaches (like the one herein presented) combine the advantages of PQCs and AQCs, while overcoming their drawbacks: as the passive quenching begins right after avalanche ignition, the active action introduces a reliable and adjustable hold-off time before the fast reset.

Short and well-defined dead-time is required in applications with high dynamic range (e.g., automotive Light Detection and Ranging - LiDAR [10]), high-count rate (e.g., Optical Wireless Communication – OWC [11], Quantum Key Distribution - QKD [12]), and high number of collected photons (e.g., fluorescence lifetime spectroscopy, optical time-domain reflectometry through Time-Correlate Single-Photon Counting - TCSPC [13]). In high dynamic range applications, short dead-times are important to increase the maximum count rate before saturation, but, at the same time, single-photon sensitivity is required for detecting very dim optical signals. Automotive is a typical application in which high dynamic range is required to cover a wide range of scene brightness conditions (e.g., dark tunnel or strong solar illumination). In OWC both long dead-times and afterpulsing have detrimental effects on the bit-error rate of the communication [14], thus the trade-off between these two parameters must be carefully evaluated. Comparing receivers containing SPADs or APDs in OWC, it results that SPADs are more sensitive in the dark and the APD receivers are more sensitive in normal indoor operating conditions, but SPAD arrays have the potentiality to outperform both single SPADs and APDs [15]. In QKD very simple but highly effective attacks exploit the dead-time of SPADs to gain information about the generated keys without being detected [16]. In TCSPC applications each measurement must be repeated many times to collect the arrival times histogram, which shape is equivalent to the detected optical waveform. To reduce the overall measurement time, high repetition rate lasers and detector with short dead-times are required. In [10] it is demonstrated that long dead-times are equivalent to a reduction of detection efficiency, proportionally to the laser repetition rate (i.e., the faster the repetition rate the higher the efficiency loss for a given dead-time), and consequently an increase of the measurement duration to achieve a given precision. In [17] fast TCSPC measurements are achieved by precisely matching the SPAD dead-time with the laser period, to avoid the so-called pile-up effect. In this case, dead-times with a short and stable duration are necessary to avoid distortion in the reconstructed waveform. The stability of the dead-time duration is also important to effectively implement saturation correction algorithms [18].

Given the applications push for high-count rate SPAD detectors, many groups are aiming to develop SPAD pixels with short dead-time and low afterpulsing. Under the push of the telecommunications market, a variety of high-speed gating techniques to reduce the charge per avalanche and power consumption has been developed for InGaAs/InP SPADs, improving, beyond the count rate, afterpulsing probability and detection efficiency [19], [20]. On the other hand, thick reach-through silicon SPADs are not designed for high-speed applications but, nonetheless, photon rates above 100 Mcps have been reached [21]. Regarding low-voltage CMOS SPADs, various fast quenching circuits have been recently proposed. In [22], [23] and [24] short dead-time (6, 5.4 and 6 ns respectively) and low afterpulsing probability have been achieved with active front-end circuits in combination with small size SPADs (8 and 10 μm diameter). The pixel presented [24] exploits a backside illuminated SPAD and 3D-stacking with a scaled 90 nm CMOS technology for electronics, allowing to reduce the size of the AQC (AQC dimensions are not reported in the paper, but the overall pixel area is 10×10 μm^2). The AQC presented in [25] features 6.2 ns dead-time with a larger SPAD (50 μm diameter) developed in custom technologies, with relatively low afterpulsing (5%). Nevertheless, this AQC occupies a large area of 364×90 μm^2, being not applicable to high density SPAD arrays. The shortest deadtime presented in literature (4 ns) is achieved in [26], but no information about the afterpulsing is provided in the paper.

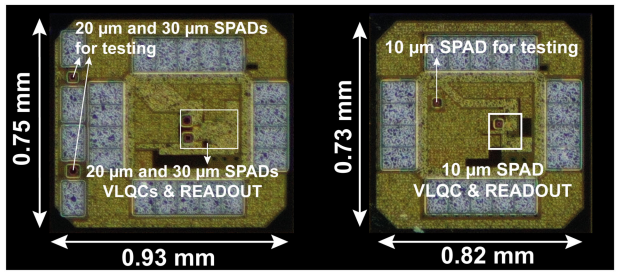

Fig. 1. Complete micrograph (including bonding pads) of the 3FF-chip (left) and of the 6-FF chip (right).

The aim of this work is to describe the architecture and the experimental characterization of a new AQC developed in a 160 nm Bipolar CMOS DMOS (BCD) technology, which features sub-nanosecond dead-time and very low afterpulsing (about 0.14%). To overcome the maximum output rate limit imposed by the maximum switching frequency of the chip wire bonding pads, a peculiar architecture with multiplexed outputs, namely a frequency dividing readout scheme, has been devised and will be presented in Section II.B. The presented architecture has been integrated along with low DCR and high PDP SPADs (with 'deep' p/n+ junction) thoroughly described in [27].

In Section V, Table I summarizes the main parameters of state-of-the-art AQCs featuring short dead-time for comparison with the work presented in this paper.

II. Chip Architecture

This Section introduces the architecture of three circuits embedding SPADs with different dimensions and with the same sensing/quenching circuit, but different frequency dividing readout circuits. In particular, the first two architectures are included in the same chip, hereafter named 3FF-chip (Fig. 1, left), and feature a 20 μm squared SPAD and a 30 μm circular one respectively, integrated with a divide-by-3 readout scheme. The third one, hereafter named 6FF-chip, (Fig. 1, right) features a 10 μm circular SPAD with a divide-by-6 readout scheme. Note that both

TABLE I
COMPARISON OF THE MAIN PERFORMANCE PARAMETERS AMONG STATE-OF-THE-ART HIGH-COUNT RATE PIXELS

Reference	Dead-time (ns)	Afterpulsing probability (%)	AQC size (μm^2)	SPAD Excess Bias (V)	Pixel technology	SPAD size (μm)	Peak PDP (% @ λ nm)
Niclass 2010 [22]	6	≈ 0	8×75	3.5	180 nm CMOS	8 (diameter)	20 @ 470
Eisele 2011 [23]	5.4	1.3	8×17	2.6	130 nm CMOS	8 (diameter)	N.A.
Ceccarelli 2019 [25]	6.2	5	364×90	5	Custom thin SPAD/ 180 nm CMOS (electronics)	50 (diameter)	50 @ 550
Katz 2019 [26]	4	N.A.	37×37	1.8	180 nm CIS	25 (side)	47 @ 580
Kumagai 2021 [24]	6	0.1	<10×10	N.A.	90 nm SPAD / 40 nm CMOS	10 (side)	22 @ 905
This Work, 6FF	0.9	0.14	37×35	5	160 nm BCD	10 (diameter)	48 @ 490
This Work, 3FF	1.8	0.09	37×35	5	160 nm BCD	20, 30 (diameter)	64 @ 490

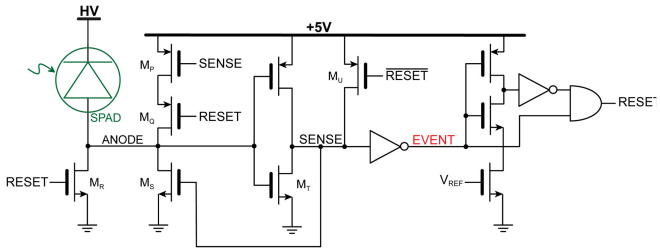

Fig. 2. Schematic representation of the VLQC circuit with the current-starved inverter used for regulating the hold-off time through V_{REF}.

chips include additional separate SPADs with anode and cathode dedicated bonding pads for proper device characterization (i.e., breakdown voltage, Photon Detection Probability, PDP, and Dark Count Rate, DCR).

A. Free Running Variable Load Quenching Circuit

When operated in free-running mode, SPADs are always kept above breakdown by a quantity called excess bias (V_{EX}), and they are always ready to detect photons, unless the detector is in hold-off condition.

We have readapted the Variable Load Quenching Circuit (VLQC) presented in [9], so to purposely fit the free-running operation mode, to promptly quench the avalanche current, fast enough to make the afterpulsing effect negligible [28] (see Section IV.D), and to generate an *EVENT* signal every time an avalanche is triggered.

To reduce afterpulsing probability, the VLQC must be able to quench the avalanche as soon as possible. Through a feedback action, the circuit helps in charging the parasitic capacitances, reducing the SPAD quenching time. Despite this feedback action, the unavoidable temporal limits are the time required to charge and discharge the capacitances and the propagation delay of the electronics.

The circuit schematic is shown in Fig. 2 and makes use of thick oxide 5 V MOS transistors that, being able to support $V_{EX} = 5$ V, are needed for sensing, activation, and quenching. Beyond that, having a single-rail stage allows avoiding level-shifting stages, resulting in a faster signal propagation for resetting the SPAD, reducing the dead-time. In the initial condition, the SPAD anode is kept at ground through M_S, operating in ohmic regime, ready to sense an avalanche current. When an avalanche occurs, the current in M_S passively increases the anode voltage, switching on M_T, that discharges *SENSE* node down to ground. M_S is consequently switched-off and the quenching resistance is maximized, thus speeding up the quenching process and minimizing the avalanche current charge, eventually assisted by the positive feedback mechanism involving the M_P-M_Q transistors series (anode reaches $V_{EX} = 5$ V in about 100 ps). Indeed, the measured pulse charge, evaluated as the ratio between average current consumption and photon rate, results below 400 fC, much smaller than the values previously reported in literature [9], [29]. According to eq. (2) the corresponding SPAD energy dissipation is 11.6 pJ per avalanche.

Upon photon detection, *EVENT* signal rises as soon as the positive feedback loop brings *SENSE* node to ground, simultaneously keeping the SPAD quenched for the corresponding hold-off time, after which *RESET* signal swiftly restores the initial conditions, by rapidly switching-on M_R. Reset phase is also assisted by M_U which, switched on by $\overline{RESET}$, recharges *SENSE* node, thus helping the anode to completely discharge. Special effort was put in tailoring the anode connected transistors, because of their contribution to the overall anode capacitance which trades-off with the prompt avalanche quenching (and consequently the avalanche charge) and the fast SPAD reset. The variable hold-off is achieved through a voltage-controlled current-starved inverter, through which a non-linear dead-time against V_{REF} characteristic is achieved.

B. Frequency Dividing Readout Scheme

When the processing electronics is outside the chip (as it typically is for single pixel detectors), one of the main limitations to reach very high-count rates (i.e., from hundreds of Mcps to Gcps) is the limited bandwidth of the chip output signals. Even introducing a high-frequency integrated readout architecture, the sum of pads, bonding wires, and Printed Circuit Board (PCB) interconnections generate large stray capacitances (up to 10 pF), inducing a cut-off frequency on the output signal at roughly 100 MHz.

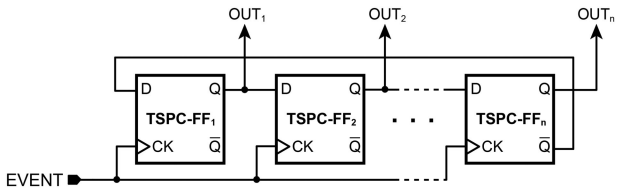

Fig. 3. Proposed read-out scheme with n-stage serial shift register (n equal to 3 in the 3-FF chip, 6 in the 6-FF chip).

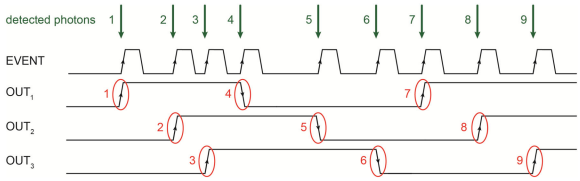

Fig. 4. Time scheme of the main signals of the 3-FF chip (right).

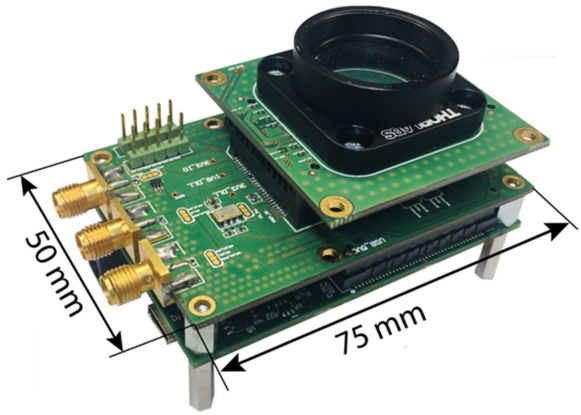

Fig. 5. Complete testing system based on an Opal Kelly XEM7310-A75 (lower board). The middle board generates the needed power supplies and interfaces the chip with the FPGA and the external instrumentation, while the top one hosts the chip and flange for optical component connection.

To overcome this constraint, we have designed a readout scheme, represented in Fig. 3, based on an n-stage Shift Register (SR), that acts as a frequency divider on the event rate.

The circuit makes use of True Single-Phase Clock Flip-Flops (TSPC-FFs), exploiting the advantages of the domino logic, allowing to reach high-count rates with reduced propagation delay (below 400 ps, simulated) and power consumption. Unfortunately, given the intrinsic dynamic nature of the TSPC-FF architecture, there is a minimum required count rate to keep consistent data (around 5 kcps) since leakage currents discharge internal nodes in about 200 μs, making the architecture not suitable for low photon rate applications.

Considering the 3-FF chip, each of the two SPAD *EVENTs* is fed to its own independent readout, with a 3-stage readout, while the 6-FF chip employs a 6-stage readout. To better clarify the circuit behavior, the time scheme of the main signals in the 3-FF chip is depicted in Fig. 4. The *EVENT* signal, generated by the SPAD front-end circuitry at each photon detection (green arrows), acts as clock of all FFs. Being the $\bar{Q}$ of the last FF fed to the input of the first, each FF output changes its state once every three *EVENT*s. The three outputs (i.e., OUT_1, OUT_2 and OUT_3) are shifted between one another by the time interval between two subsequent photon detections. All in all, each photon event corresponds to an edge of one output signal (either rising or falling edge). The overall count rate can be extracted through post-processing computations without losing the information about single events.

Through the designed configuration, the 100 MHz limit is overcome, making the external readout reliable up to 600 Mcps detection rate in the 3-FF configuration and up to 1.2 Gcps in the 6-FF one.

III. System Design

To easily characterize both the proposed chips, a standalone system has been developed. The system provides the power supplies, manages the communication with the chips, and sends the acquired data to a remote computer through a USB 3.0 link. Fig. 5 shows the complete system made by three PCBs. The lower board, a commercial digital interface board (Opal Kelly XEM7310-A75), is equipped with a Xilinx Artix-7 Field Programmable Gate Array (FPGA). The middle board is designed in-house, making the chip interfacing with the FPGA and with external instrumentation through three SMA connectors. These connectors can act either as inputs or as outputs according to the setup requirements. The same board generates also all the needed power rails from a common external 5 V supply. In particular, the bias voltage of the SPADs can be regulated to compensate breakdown voltage variations, and V_{REF} can be set from 0 to 3.3 V regulating a Digital to Analog Converter (DAC) output. The upper board, provided with an SM1-threaded flange, hosts the chip and can be suitably interchanged. Each chip output, coupled with a low-jitter buffer able to drive 50 Ω impedance cables, can be either fed to the FPGA or to an SMA connector.

IV. Experimental Characterization

The developed system has allowed a thorough characterization of the two chips, both in terms of sensor characteristics and front-end behavior. After evaluating the SPAD breakdown voltage to be 26.5 V, we have tested the detector noise, the efficiency, the timing response, and the dead-time characteristics, along with the chips' response to high-photon fluxes. All the measurements have been carried out at room temperature without any cooling and at 5 V excess bias, offering the best trade-off in terms of efficiency against detector noise [27].

A. SPAD DCR and PDP

In SPAD detectors, noise is caused by all the avalanches not related to a photon detection. One of the major contributions are ignitions triggered by thermal generation processes or trap-assisted tunneling [30]. The average rate at which these phenomena happen is called Dark Count Rate (DCR). DCR depends mostly on Silicon lattice quality, SPAD area, operating temperature, and excess-bias voltage. For the three tested SPADs, we have measured an average DCR of 0.19 cps/μm^2 which translates into a negligible noise contribution in high-flux applications.

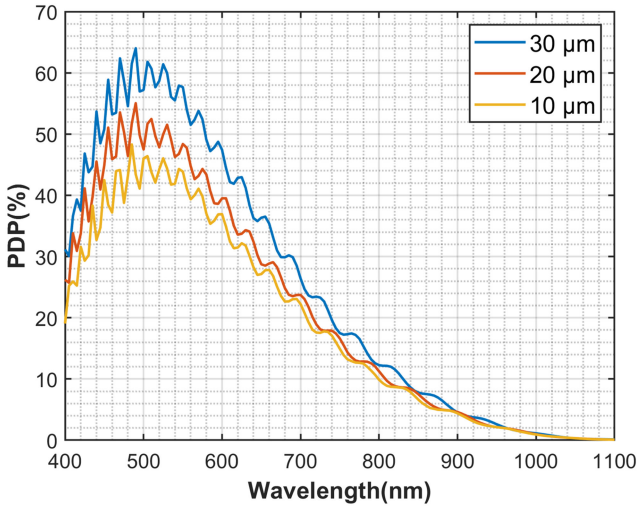

Fig. 6. Photon Detection Probability as a function of the wavelength for the three SPADs (at room temperature and 5 V excess bias).

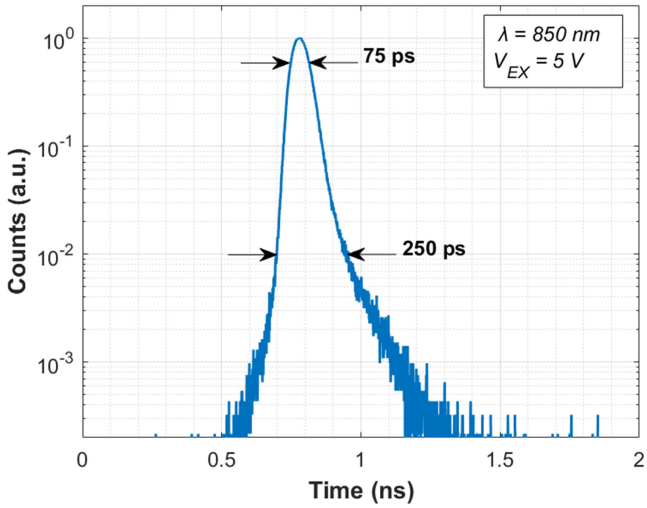

Fig. 7. Timing response of the 10 μm diameter SPAD when illuminated by a 45 ps FWHM pulsed laser (at room temperature and 5 V excess bias).

The Photon Detection Probability (PDP) of three tested SPADs has been measured, with the set-up illustrated in [27], in a wavelength range between 400 nm and 1100 nm, obtaining the results shown in Fig. 6. The peak PDP is 64% at 490 nm for the 30 μm SPAD, 55% for the 20 μm one, and 48% for the 10 μm one. Then, in all three cases, it decreases to about 6-8% at 850 nm. The interference etaloning shown in all the curves is due to the gap in the refractive index between air and chip back-ends.

B. Timing Response

Even if the proposed pixel architecture is designed and optimized for photon-counting applications, we have also investigated the temporal response of the chips. The temporal response of a SPAD-based system can be measured by illuminating the device with a pulsed laser with a narrow pulse width and reconstructing a histogram of the photon arrival times.

The Full Width at Half-Maximum (FWHM) of the histogram reflects the jitter contributions of the laser, the SPAD, the front-end electronics, and the instrumentation used for the measurement. We have performed the characterization using a pulsed diode laser at 850 nm, with 45 ps FWHM. We have acquired the distribution of the time delay between a laser sync and the output of the first flip-flop using an 8 ps FWHM SPC-630 TCSPC board (Becker&Hickl GmbH). Fig. 7 shows the temporal response of the 10 μm diameter SPAD, representative for all the tested chips. The resulting FWHM is 75 ps and the Full Width at 1% of the Maximum is 250 ps, with no significant difference in the timing response emerges spanning VREF from 1.8 to 3.3 V in both chips.

C. Dead-Time

As illustrated in Section IV, the minimum interval between two successive detections can be chosen by regulating an analog voltage (V_{REF}) between 0 and 3.3 V. To characterize the dead-time of our designs, we kept the chips under an intense light flux so that after each detection, as soon as the SPAD is active again, another photon triggers an avalanche. This means that the time delay between two output pulses from two consecutive flip-flops is constant and equal to the dead-time. By measuring the number of events in a fixed integration time, we have estimated the dead-time (T_D) as T_{INT}/N_{count} where T_{INT} is the integration time and N_{count} the average number of events during T_{INT}. Note, however, that theoretically the saturation point, which is a function of V_{REF}, can be reached only for infinitely high flux [18].

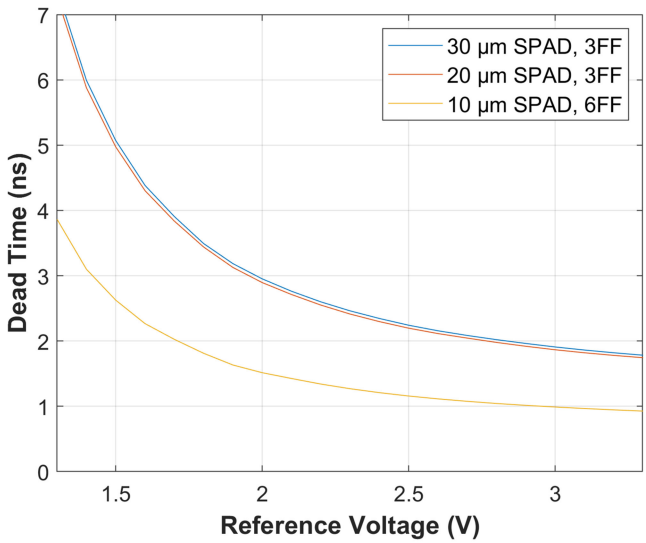

Fig. 8. Comparison of the dead time as a function of the reference voltage (V_{REF}) for the three tested pixels.

The results are graphed in Fig. 8. for the three tested pixels. Given the different readout schemes, the two chips present different dead-times while the small discrepancy (about 70 ps) between the results from the two 3-FF SPADs is due to the dissimilar anode parasitic capacitance, consequence of different SPAD areas. The dead-time can be as low as 1.78 ns in the 3-FF chip, and down to 0.93 ns in the 6-FF chip corresponding to 560 Mcps and 1.08 Gcps event rates at saturation, respectively.

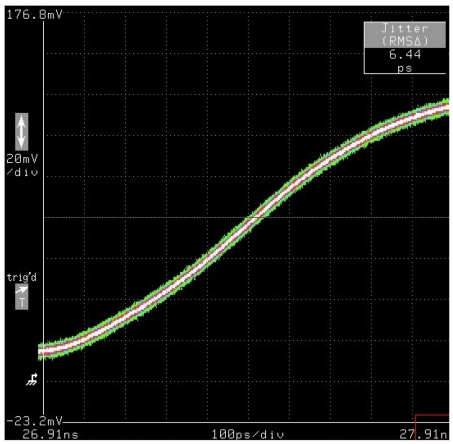

Fig. 9. Rising edge of OUT_2 when the SPAD is in saturation. The waveform has been sampled with a digital sampling oscilloscope employing OUT_1 as trigger.

As explained in Section I, the dead-time is one of the major drawbacks of SPADs in respect to other photodetectors since it causes a count rate not linearly proportional to the impinging photon flux. However, at any photon-flux level, the actual photon flux can be estimated from the measured count rate, assuming constant illumination during the integration time and stable dead-time [18]:

$$N_{ideal} = \frac{N_{count}}{1 - N_{count}T_D/T_{INT}} \tag{3}$$

Where N_{ideal} is the number of photons that would be detected without dead-time. The more stable and well-defined the dead-time, the lower the error introduced by this correction. For this reason, the dead-time jitter becomes an important parameter in high-flux applications. To measure the dead-time jitter, we connected two consecutive outputs (OUT_1 and OUT_2) to a digital sampling oscilloscope (11801C by Tektronix). A high and constant light flux has been employed to keep the chips at saturation, and the rising edge of OUT_1 has been employed as trigger to sample OUT_2. As an example, Fig. 9 shows OUT_2 of the 6-FF chip at 3.3 V V_{REF}, with a timing jitter equal to 6.44 ps. There is no significant variation for other V_{REF} values, and for the 3-FF chip.

D. Afterpulsing Probability

As seen in Section I, another source of noise in SPAD is the afterpulsing phenomenon. The afterpulsing probability of both chips has been measured by reconstructing the histogram of the inter-arrival times between consecutive output rising edges of OUT_1 and OUT_2. To correctly estimate the afterpulsing, it is necessary to work at low photon rates since it reduces the probability that a following detection masks an afterpulsing event. Thus, the measurement has been carried out using a constant light source to have a detection of about 15 kcps. An example of the measurement is shown in Fig. 10, where the blue line represents the inter-arrival time histogram. The contribution of the events not due to afterpulsing can be fitted with an exponential decay considering the histogram bins with negligible afterpulsing contribution (e.g., starting from 20 ns in Fig. 10). The exponential decay (orange line) can be subtracted from the histogram, so that only the events due to afterpulsing remain. Their probability is computed as the integral of this final histogram divided by the total number of events. The Time to Digital Converter (TDC) employed for this measurement has been directly implemented into the FPGA with 100 ps resolution and 1 ms Full Scale Range (FSR), which allows a reliable fitting of the exponential decay. The afterpulsing probability at various dead-times is plotted in Fig. 11, highlighting, as expected, how probability decreases with dead-time, with a maximum probability of 0.14% in correspondence of the minimum dead-time.

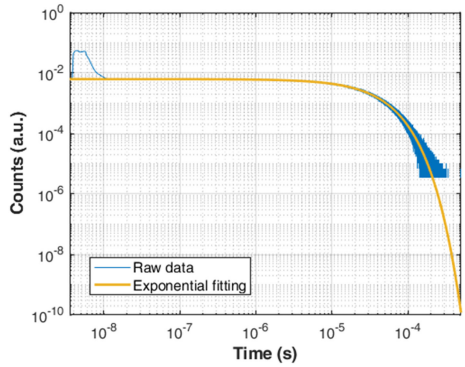

Fig. 10. Example of an afterpulsing measurement carried out with a FPGA-implemented TDC. The blue line represents the inter-arrival times between two subsequent detections, while the orange line is the histogram expected if the afterpulsing probability were zero.

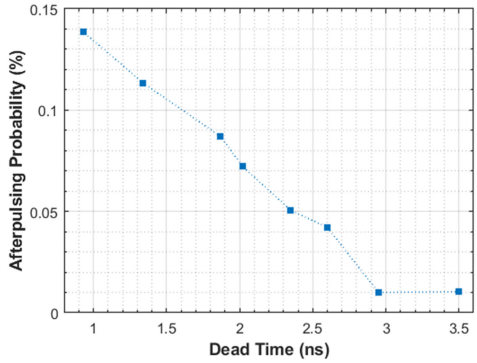

Fig. 11. Afterpulsing probability measurements at various dead-times.

A. Linearity and SNR

Due to the dead-time between subsequent detections, the response of a SPAD to a light flux is not linear, as instead happens in conventional linear sensors before saturation. At high fluxes, due to their quantized nature, SPADs "softly" saturate toward a value equal to $T_{INT}/T_{D.}$ The behavior of the 6-FF chip exposed to different light intensities is shown in Fig. 12 comparing the characteristic for three dead-times for T_{INT} equal to 10 μs. The measured data (circles) are graphed together with the simulated SPAD response to the same photon flux (continuous lines).

In conventional sensors, the photon noise increases as the square-root of the incident flux since it is dominated by shot

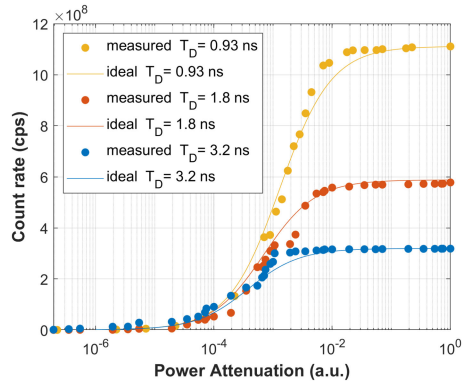

Fig. 12. SPADs response to an increasing photon flux for three different hold-off times (circles) and the count rate expected from eq. (3) (continuous lines). Saturation occurs at $1/T_D$.

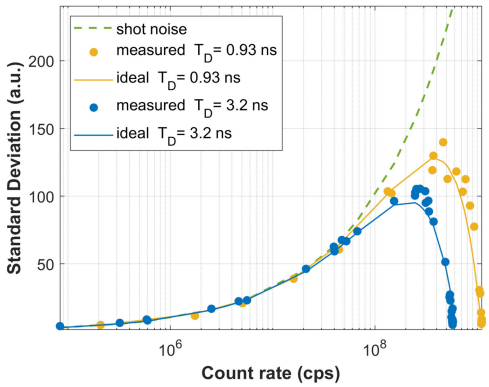

Fig. 13. Measured standard deviation (circles) compared to the theoretical one (continuous) and to the shot noise (dashed green line).

noise [31]. Then, when the full well capacity is reached, the sensor saturates. Differently, photon noise in SPADs firstly increases, reaches a maximum, and then decreases since, approaching saturation, the probability of having an event immediately after the dead-time increases. According to [18], the variance in the number of detected photons (N_{count}) at different fluxes can be approximated to:

$$Var\ [N_{ideal}] = \frac{N_{ideal}}{(1 + N_{ideal} \cdot T_{INT}/T_D)^3} \qquad (4)$$

Where N_{ideal} can be computed using (3). Fig. 13 shows in green the shot noise contribution (i.e., square root of mean count rate), the continuous lines are the theoretical standard deviation from (4), and the circles are the measured one, as a function of the impinging photon flux, for the 6-FF chip.

V. Conclusion

We have presented the design and the characterization of a novel SPAD sensing and readout scheme that allows the reduction of the dead-time down to 0.93 ns, enabling operations with photon fluxes in the giga count per second range. In particular, the proposed chips integrate three SPADs with different areas (10 μm diameter round SPAD, 20 μm squared SPAD, and 30 μm diameter round SPAD), a fast-quenching VLQC (devised for pulse charge minimization) and two different frequency-dividing readout schemes. Through an in-house designed system, we thoroughly characterized the chips both in terms of SPAD performance, which reflects what reported in [27], and front-end behavior. Table I compares the performance of the chips presented in this paper with other high-count rate SPAD pixels presented in literature in terms of minimum dead-time, afterpulsing probability, and SPAD and AQC area. The architecture presented in this work compares favorably both in term of dimension and overall performance. We believe that the compact size of the sensing and readout scheme easily allows their integration within high-resolution SPAD arrays. Furthermore, the obtained extremely short and stable dead-time, paired with minimal afterpulsing probability, can be favorably exploited in many single-photon applications that require either high count rates or to effectively implement saturation correction algorithms.

References

[1] M. Ghioni, A. Gulinatti, I. Rech, F. Zappa, and S. Cova, "Progress in silicon single-photon avalanche diodes," *IEEE J. Sel. Topics Quantum Electron.*, vol. 13, no. 4, pp. 852–862, Jul./Aug. 2007.

[2] A. Giudice *et al.*, "High-rate photon counting and picosecond timing with silicon-SPAD based compact detector modules," *J. Modern Opt.*, vol. 54, no. 2-3, pp. 225–237, 2007.

[3] G. Boso, A. Dalla Mora, A. Della Frera, and A. Tosi, "Fast-gating of single-photon avalanche diodes with 200ps transitions and 30ps timing jitter," *Sensors Actuators A: Phys.*, vol. 191, pp. 61–67, 2013.

[4] E. Charbon, "Single-photon imaging in complementary metal oxide semiconductor processes," *Philos. Trans. Roy. Soc. A: Math., Phys. Eng. Sci.*, vol. 372, no. 2012, 2014, Art. no. 20130100.

[5] E. Charbon, C. Bruschini, and M.-J. Lee, "3D-Stacked CMOS SPAD image sensors: Technology and applications," in *Proc. 25th IEEE Int. Conf. Electron., Circuits Syst.*, 2018, pp. 1–4.

[6] F. Zappa, S. Tisa, A. Tosi, and S. Cova, "Principles and features of single-photon avalanche diode arrays," *Sensors Actuators A: Phys.*, vol. 140, no. 1, pp. 103–112, 2017.

[7] A. C. Giudice, M. Ghioni, S. Cova, and F. Zappa, "A process and deep level evaluation tool: Afterpulsing in avalanche junctions," in *Proc. ESSDERC '03. 33rd Conf. Eur. Solid-State Device Res.*, 2003, pp. 347–350.

[8] S. Cova, M. Ghioni, A. Lacaita, C. Samori, and F. Zappa, "Avalanche photodiodes and quenching circuits for single-photon detection," *Appl. Opt.*, vol. 35, no. 12, pp. 1956, 1996.

[9] D. Bronzi *et al.*, "Fast sensing and quenching of CMOS SPADs for minimal afterpulsing effects," *IEEE Photon. Technol. Lett.*, vol. 25, no. 8, pp. 776–779, Apr. 2013.

[10] T. T. Ta *et al.*, "A 2D-SPAD array and read-out AFE for next-generation solid-state LiDAR," in *Proc. IEEE Symp. VLSI Circuits*, 2020, pp. 1–2.

[11] E. Sarbazi, M. Safari, and H. Haas, "The impact of long dead time on the photocount distribution of SPAD receivers," in *Proc. IEEE Glob. Commun. Conf.*, 2018, pp. 1–6.

[12] J. Zhang *et al.*, "2.23 GHz gating InGaAs/InP single-photon avalanche diode for quantum key distribution," in *Proc. Adv. Photon Counting Techn. IV*, 2010, Art. no. 76810Z.

[13] P. Peronio, G. Acconcia, I. Rech, and M. Ghioni, "Improving the counting efficiency in time-correlated single photon counting experiments by dead-time optimization," *Rev. Sci. Instruments*, vol. 86, no. 11, 2015, Art. no. 113101.

[14] E. Sarbazi, M. Safari, and H. Haas, "On the information transfer rate of SPAD receivers for optical wireless communications," in *Proc. IEEE Glob. Commun. Conf.*, 2016, pp. 1–6.

[15] L. Zhang *et al.*, "A comparison of APD- and SPAD-Based receivers for visible light communications," *J. Lightw. Technol.*, vol. 36, no. 12, pp. 2435–2442, 2018.

[16] H. Weier *et al.*, "Quantum eavesdropping without interception: An attack exploiting the dead time of single-photon detectors," *New J. Phys.*, vol. 13, no. 7, 2011, Art. no. 073024.

[17] G. Acconcia, A. Cominelli, M. Ghioni, and I. Rech, "Fast fully-integrated front-end circuit to overcome pile-up limits in time-correlated single

photon counting with single photon avalanche diodes," *Opt. Exp.*, vol. 26, no. 12, 2018, Art. no. 15398.
[18] A. Ingle, A. Velten, and M. Gupta, "High flux passive imaging with single-photon sensors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 6753–6762.
[19] Z. Lu, W. Sun, J. C. Campbell, X. Jiang, and M. A. Itzler, "Pulsed gating with balanced InGaAs/InP single photon avalanche diodes," *IEEE J. Quantum Electron.*, vol. 49, no. 5, pp. 485–490, May 2013.
[20] A. Restelli, J. C. Bienfang, and A. L. Migdall, "Single-photon detection efficiency up to 50% at 1310 nm with AN INGAAS/INP avalanche diode gated at 1.25 GHz," *Appl. Phys. Lett.*, vol. 102, no. 14, 2013, Art. no. 141104.
[21] M. A. Wayne, J. C. Bienfang, and A. L. Migdall, "Low-noise photon counting above 100×10^6 counts per second with a high-efficiency reach-through single-photon avalanche diode system," *Appl. Phys. Lett.*, vol. 118, no. 13, 2021, Art. no. 134002.
[22] C. Niclass and M. Soga, "A miniature actively recharged single-photon detector free of afterpulsing effects with 6ns dead time in a 0.18 μm CMOS technology," in *Proc. Int. Electron Devices Meeting*, 2010, pp. 14.3.1–14.3.4.
[23] A. Eisele *et al.*, "185 MHz count rate 139 dB dynamic range single-photon avalanche diode with active quenching circuit in 130 nm CMOS technology," in *Proc. IISW*, 2011, pp. 6–8.
[24] O. Kumagai *et al.*, "7.3 A 189 $\times$ 600 back-illuminated stacked SPAD direct Time-of-Flight depth sensor for automotive LiDAR systems," in *Proc. IEEE Int. Solid- State Circuits Conf.*, 2021, pp. 110–112.
[25] F. Ceccarelli, G. Acconcia, A. Gulinatti, M. Ghioni, and I. Rech, "Fully integrated active quenching circuit driving custom-technology SPADs with 6.2-ns dead time," *IEEE Photon. Technol. Lett.*, vol. 31, no. 1, pp. 102–105, Jan. 2019.
[26] A. Katz, T. Blank, A. Fenigstein, T. Leitner, and Y. Nemirovsky, "Active-Reset for the N+P single-ended SPAD used in the NIR LiDAR receivers," *IEEE Trans. Electron Devices*, vol. 66, no. 12, pp. 5191–5195, Dec. 2019.
[27] M. Sanzaro *et al.*, "Single-Photon avalanche diodes in a 0.16 μm BCD technology with sharp timing response and red-enhanced sensitivity," *IEEE J. Sel. Topics Quantum Electron.*, vol. 24, no. 2, Mar./Apr. 2018, Art. no. 3801209.
[28] M. A. Wayne, A. Restelli, J. C. Bienfang, and P. G. Kwiat, "Afterpulse reduction through prompt quenching in silicon reach-through single-photon avalanche diodes," *J. Lightw. Technol.*, vol. 32, no. 21, pp. 4097–4103, 2014.
[29] S. Tisa, F. Guerrieri, and F. Zappa, "Variable-load quenching circuit for single-photon avalanche diodes," *Opt. Exp.*, vol. 16, no. 3, 2008, Art. no. 2232.
[30] A. Migdall, *Single-photon Generation and Detection*. Cambridge, Massachusetts, USA: Academic, 2013.
[31] S. W. Hasinoff, "Photon, Poisson noise," in *Proc. Comput. Vis.*, 2014, pp. 608–610111.

Fabio Severini (Student Member, IEEE) received the B.Sc. and M.Sc. degrees in electronics engineering in 2016 and 2018, respectively, from the Politecnico di Milano, Milan, Italy, where he is currently working toward the Ph.D. degree in information technology with the Dipartimento di Elettronica, Informazione e Bioingegneria. His research activity focuses on the design and development of CMOS circuits containing SPAD detectors, targeting quantum imaging and quantum communication applications.

Iris Cusini received the B.Sc. degrees in automation engineering (*summa cum laude*) both from Politecnico di Milano, Milan, Italy and from Tongji University, Shanghai, China, within a double degree project, and the M.Sc. degree (*summa cum laude*) in electronics engineering in 2019, from Politecnico di Milano. Since November 2019, she has been a Ph.D. Candidate in information technology with Politecnico di Milano. Her research interests include the design, development, and testing of systems with single-photon avalanche diodes (SPADs) and digital Silicon Photomultipliers (dSiPMs).

Davide Berretta received the B.Sc. degree in biomedical engineering and the M.Sc. degree in electronics engineering from Politecnico di Milano, Milan, Italy, in 2017 and 2020, respectively. He is currently a Ph.D. Student in information technology (Electronics) with the Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano. His research activity is focuses on the design of an innovative SPAD front end integrated circuit, targeting the power consumption minimization.

Klaus Pasquinelli received the B.Sc. and M.Sc. degrees in electronics engineering in 2016 and October 2018, respectively, from Politecnico di Milano, Milan, Italy, where he is currently working toward the Ph.D. degree in information technology. His research interests include the design, development, and testing of systems with single-photon avalanche diodes (SPADs) matrices and digital Silicon Photomultiplier (dSiPM).

Alfonso Incoronato (Student Member, IEEE) received the B.Sc. degree in biomedical engineering and the M.Sc. degree in electronics engineering from Politecnico di Milano, Milan, Italy, in 2016 and 2019, respectively. Since November 2019, he has been a Ph.D. Candidate in information technology (Electronics) with Politecnico di Milano. His research activity concerns the design of SPADs and SiPM arrays for automotive LiDAR and quantum communication.

Federica Villa (Member, IEEE) was born in Milan, in 1986. Since 2015, she has been an Assistant Professor with Electronics Engineering, Politecnico di Milano. Her academic research interests mainly include the design of microelectronic integrated sensors and circuits in CMOS and bipolar-CMOS-DMOS (BCD) technologies for the development of single photon detectors such as single photon avalanche diodes (SPADs) and silicon photomultipliers (SiPMs), photon-counting array imagers with single-photon sensitivity, and high-resolution and linearity time-to-digital converters (TDCs).