Characterization-Based Modeling of Retriggering and Afterpulsing for Passively Quenched CMOS SPADs

Manuel Moreno-García[iD], *Member, IEEE*, Lucio Pancheri[iD], *Member, IEEE*, Matteo Perenzoni[iD], *Member, IEEE*, Rocío del Río, *Member, IEEE*, Óscar Guerra Vinuesa, and Ángel Rodríguez-Vázquez, *Fellow, IEEE*

Abstract— The current trend in the design of systems based on CMOS SPADs is to adopt smaller technological nodes, allowing the co-integration of additional electronics for the implementation of complex digital systems on chip. Due to their simplicity, a way to reduce the area occupied by the integrated electronics is the use of passive quenching circuits (PQCs) instead of active (AQCs) or mixed (MQCs) ones. However, the recharge phase in PQCs is slower, so the device can be retriggered before this phase ends. This paper studies the phenomena of afterpulsing and retriggering, depending on the characteristics of the SPADs and the working conditions. In order to do that, a test chip containing SPADs of different size has been characterized in several operating environments. A mathematical model has been proposed for fitting afterpulsing phenomenon. It is shown that retriggering can be also described in terms of this model, suggesting that it is linked to carriers trapped in the shallow levels of the semiconductor and that should be taken into account when considering the total amount of afterpulsing events.

Index Terms— CMOS, SPAD, passive quenching, afterpulsing, retriggering, modeling.

I. INTRODUCTION

AVALANCHE photodiodes (APDs) working in the Geiger mode can be used to detect single photons. These are also called Single Photon Avalanche Diodes (SPADs) [1]. In these operating conditions, the electric field is so high that a single carrier in the depletion region can originate a self-sustained avalanche current in the milliamperes range [2].

SPADs can be implemented in CMOS technologies [2]–[4]. Thus, they can be embedded with readout, timing, communication and control circuits into SPAD-based CMOS imagers. These imagers are capable to detect single photon events and hence, suitable for applications involving low-light detection and Time-of-Flight (ToF) measurements [5].
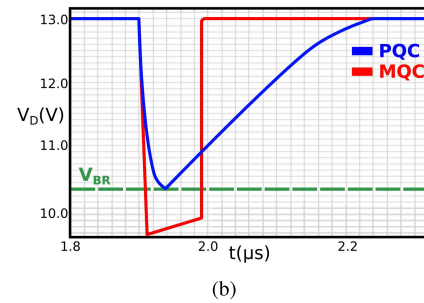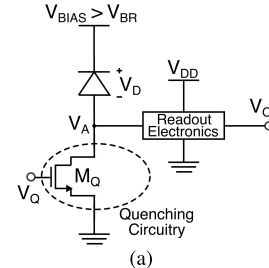
Fig. 1. (a) Conceptual diagram of the SPAD and its accompanying circuitry. (b) Transient response of the SPAD voltage during both the quenching and recharge phases for passive and mixed quenching circuits.

In order to extinguish the avalanche current and enable the device for further event detections, SPADs are connected to an appropriate quenching circuitry. The quenching circuit may just employ a ballast resistor or a transistor for quenching and recharging –referred to as Passive Quenching Circuit [6] (PQC, see Fig. 1a)–, helping to reduce the *pitch* and enlarge the *fill factor*. However, a PQC yields poorer control of the photodiode biasing. As shown in Fig. 1b in blue, this kind of circuit is characterized by a slow reset phase and by the fact that the device is biased above its breakdown voltage, V_{BR}[1], during this stage. Thus, any incident photon –or carriers that are thermally generated or released by internal traps– could retrigger the device before the complete recharge. This causes the variation of the dead time t_{dt} or the reduction of the Photon Detection Efficiency [7]. Since many SPAD applications rely

---

[1]The amount of voltage above the breakdown is referred to as excess bias voltage: $V_{EXC} = V_{BIAS} - V_{BR}$. V_{BR} is defined positive and SPAD breakdown occurs for $V_D > V_{BR}$, being V_D the voltage applied to the photodiode.

on time correlation, these spurious events may have significant impact on the performance of SPAD-based imagers.

The alternatives to PQC combine several transistors and passive components [6] to achieve fast and sharp transitions between sensitive and non-sensitive states (see Fig. 1b, red). However, this approach, referred to as Active or Mixed Quenching Circuits (AQC or MQC, respectively), results into enlarged pixel pitch and reduced fill factor, thereby limiting the resolution and responsivity of SPAD-based imagers.

This paper attempts to extract a model for the triggering events that characterize SPADs recharge when they are accompanied by PQCs, based on measurements made on a test chip designed in a high-voltage 0.35-μm CMOS process. The chip contains SPADs of different sizes which are characterized under several working conditions, in terms of dead time, excess bias voltage and temperature. Also, measurements are processed to derive conditions for a reduced impact of triggering phenomena during the recharge phase. This characterization can be of interest in applications requiring a high dynamic range, where retriggering can lead to a significant distortion of the pixel output at high light intensities. To the best of our knowledge, there are not previous models that do a complete description of triggering phenomena in passively quenched SPADs.[2]

The paper is organized as follows: Section II provides details about the design of the test chip and describes the measurement method and considerations to be taken into account for data processing, Section III presents the experimental results, Section IV proposes a mathematical model for fitting the results obtained in the measurements, and Section V draws the final conclusions of this work.

II. TEST CHIP DESCRIPTION AND BASIC CONSIDERATIONS

A. Test Chip Description

The test chip has been designed in a HV 0.35-μm CMOS process and contains a 4×60 SPAD array [9]. All SPADs are p+/deep n-tub devices, with a low doped p-region around their borders to implement the guard ring and avoid the premature edge breakdown. All SPADs have square shape with smoothed edges to lower the high electric field in these regions. They differ only in their active area side length, L, which can take four different values: 8, 12, 20 and 25μm. Fig. 2 illustrates the cross section of the SPADs.

NMOS transistors have been chosen for passively quenching SPADs, allowing the characterization of the devices at different dead times by externally modifying the applied gate voltage.

B. Triggering Phenomena in PQC and MQC Circuits

Fig. 3 illustrates the occurrence of triggering events during SPAD recharge –which could be due to the release of a trapped carrier– when using a PQC (Figs. 3a and 3b), and in the case of an MQC with active recharge (Fig. 3c). The waveforms for the voltage at the anode V_A (top) and the readout output V_O (bottom) are shown.

[2]The only experimental results found in the bibliography regarding retriggering can be seen in [8].
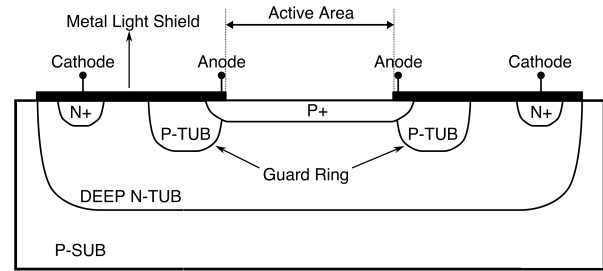

Fig. 2. Cross-section of the p+/deep n-tub SPAD structures implemented in this test chip.

Consider first Fig. 3a, which corresponds to a PQC case. Following the first photon arrival and subsequent quenching, recharge starts in a gradual process during which V_D remains larger than V_{BR} (this behavior was previously illustrated in Fig. 1b). Fig. 3a shows the signal waveforms when a new avalanche is triggered before the anode voltage intersects the threshold voltage V_{TH} of the readout comparator. In this situation, a pulse is lost at the output and the externally measured dead time is extended. This will be called *retriggering* and is characterized by the pileup of two or more events that are actually signaled by a single pulse.

Fig. 3b also illustrates V_A and V_O for a PQC, but in this case the second triggering event takes place after the previous avalanche has been signaled. In this occasion, two different dead times are computed and pileup is avoided. This will be defined as an *afterpulsing* event.

Comparing these two cases, it can be seen that the two phenomena are different from the electrical point of view: in retriggering we have an enlarged output pulse, while in afterpulsing we have for each triggering event different output pulses.

Consider now the case in which an MQC is used, represented in Fig. 3c. The first photon produces an avalanche which is signaled by a positive output signal edge. Once the avalanche gets quenched, recharge happens in a process where the voltage applied to the device remains below V_{BR} until sensitive conditions are set again. Therefore, the second event has no influence, and the dead time is set by the characteristics of the quenching circuit. The end of the recharging phase is signaled by a negative edge thereby delimiting an output pulse whose width defines the dead time.

C. Considerations for Data Acquisition and Processing

The test chip was designed to evaluate the effect of several parameters that impact retriggering and afterpulsing, and whose values have been modified during the characterization process. First of all, the dead time has been considered as a variable for describing triggering phenomena during the recharge. PQCs, as those used in our case for doing the quenching, do not allow to fix precisely the off-time of SPADs (this behavior has been illustrated in Fig. 3), but the pulse width provides an equivalent indication and will be considered as the value of t_{dt}. For this kind of circuit, the length of the quenching phase is mainly determined by the SPAD series resistance and cannot be varied externally [10]. However,
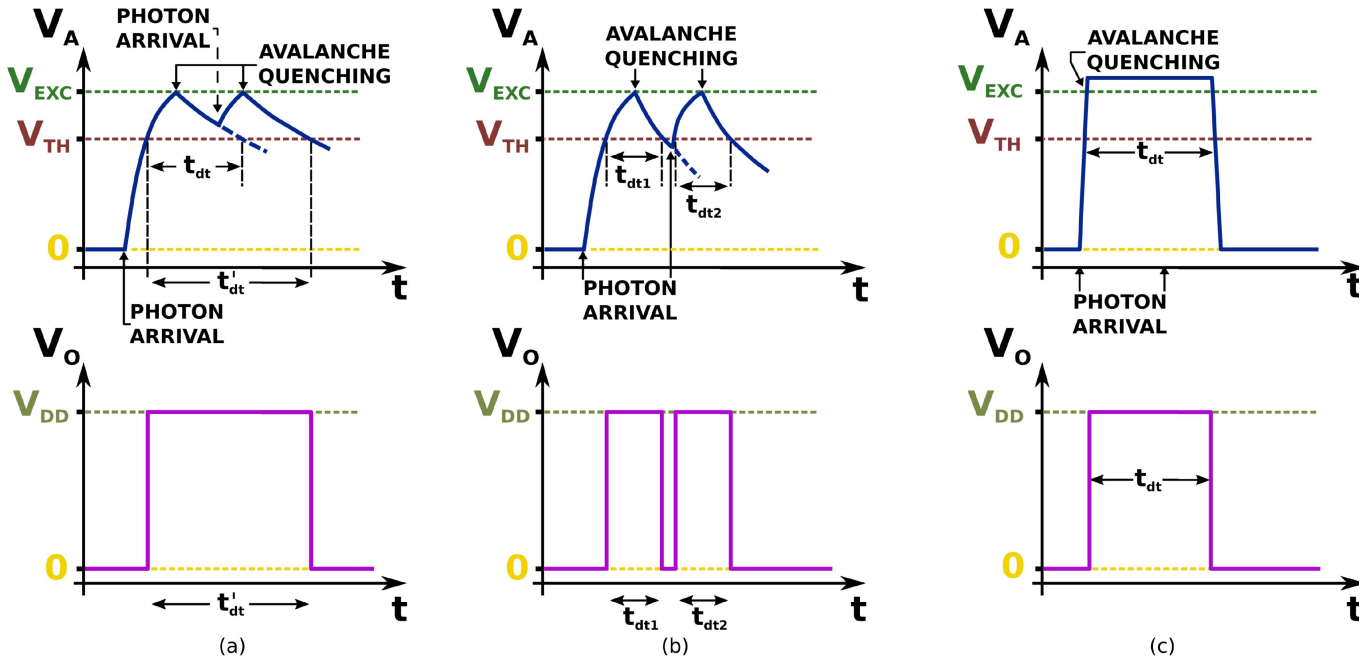
Fig. 3. Evolution of the voltage at the anode (V_A, top) and at the output of the readout electronics (V_O, bottom); (a) and (b) correspond to cases in which a PQC is used (in (a), the second triggering event happens before the first one has been measured externally, while in (b) both are measured separately), whereas in (c) an MQC is utilized.

the duration of the recharge depends strongly on the quenching resistance [10], that can be varied through the gate voltage of the quenching transistor, V_Q (see Fig. 1a).

Data have been taken for different excess bias voltages. In a typical application, the choice of V_{EXC} is bound to the optimization of system Signal-to-Noise Ratio (SNR), that depends on Photon Detection Probability and DCR, in addition to correlated noise. The larger V_{EXC}, the higher probability of avalanche triggering.

Besides this, temperature T and the size of the SPADs are other parameters that have been considered for the analysis of afterpulsing and retriggering.

The effect of those variables have been characterized through the study of the distributions of interarrival times between consecutive pulses (afterpulsing) and pulse widths (retriggering).

Fig. 4 shows a typical interarrival time distribution. The probability of afterpulsing P_{AP} is computed as the difference in area between the experimental data and the Poisson fit curve (red line) that defines the Dark Count Rate (DCR) [11].

Regarding retriggering, Fig. 5 (left) shows the pulse width distribution for SPAD test structures with MQC that were included as a check. In this case, as expected, we do not have enlarged pulses due to retriggering and the curve clearly follows a Gaussian distribution. However, when using PQCs, pulse pileup affects this distribution, causing a shift of the measured pulse widths towards larger values and distorting the distribution on its right side (see Fig. 5, right). Hence, only the data before the effect of pulse pileup starts to be noticeable have been taken into account for doing the Gaussian fit. The difference in area between the experimental and fitting curves defines the probability of retriggering P_R.
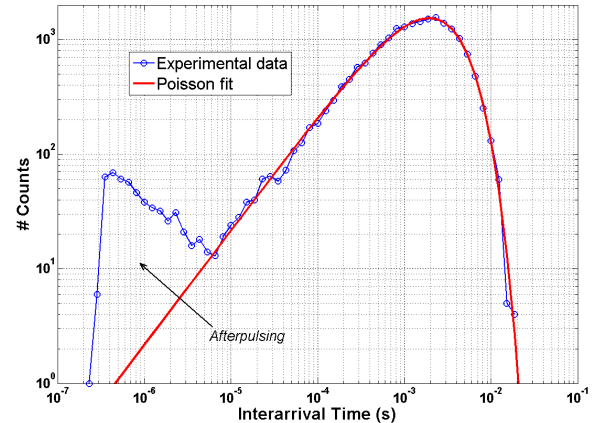

Fig. 4. Histogram of the interarrival time between consecutive pulses [11].

Different fitting criteria have been studied, with similar results. Thus, this analysis can be considered to provide relevant data in absolute and relative terms.

In order to obtain these distributions, a fast digitizing oscilloscope interfaced with a PC was used, developing a program for on-line data analysis that computes the time difference between two consecutive rising edges of the reading circuitry, and a rising and a consecutive falling edge through half maximum of the measured output pulse, respectively. Multiple oscilloscope acquisitions have been performed, until collecting 25000 samples. The reason for choosing this value is that we gradually increased the number of samples and beyond 25000, the variation of results after post-processing was negligible. All the experiments have been repeated on a number (60) of different SPADs of each size: the values
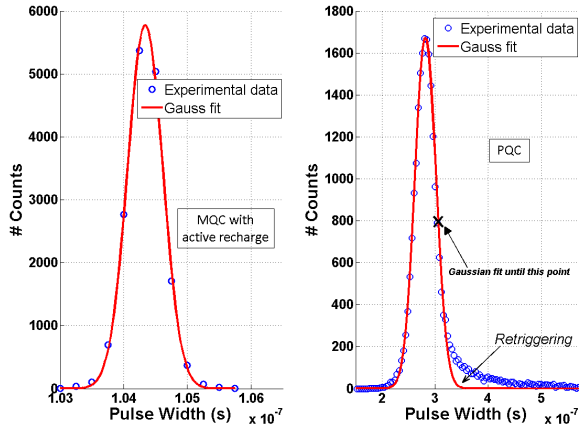
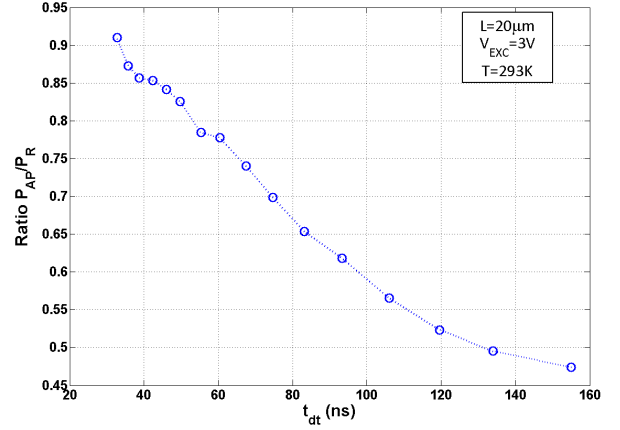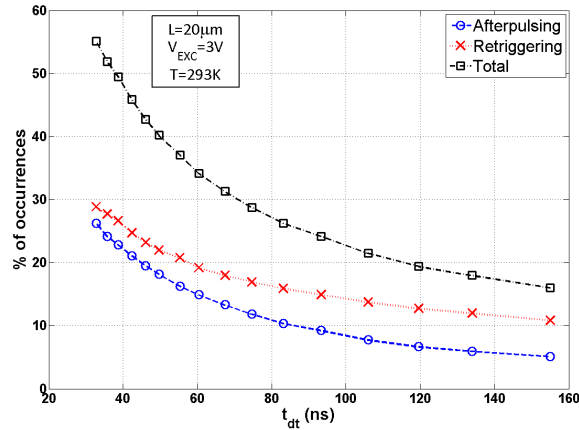Fig. 5. Pulse width distributions for SPADs with MQC (left) and PQC (right).

Fig. 7. Variation of the ratio P_{AP}/P_R with t_{dt}.

Fig. 6. Variation of the probability of afterpulsing, retriggering and their sum with the dead time.

presented throughout the paper correspond to the median of the collected statistics.

III. EXPERIMENTAL RESULTS

In this section, the results obtained in the characterization of the test chip designed in a HV 0.35-μm CMOS technology are shown. A complete characterization of afterpulsing and retriggering will be presented, considering different values of t_{dt}, V_{EXC}, operating temperature and SPAD size. SPADs are passively quenched, using a circuit similar to the one shown in Fig. 1a.

A. Dependence on Dead Time

For the characterization of P_{AP} and P_R as a function of the dead time, the group of SPADs whose active area side length is 20μm has been chosen. For this set of measurements, V_{EXC} has been kept constant at 3V, while the temperature was 293K[3]. The experimental results are illustrated in Fig. 6.

First of all, it is to note the large contribution of afterpulsing and retriggering to the total noise of the SPADs in this specific

---

[3]Variations of V_{EXC} or temperature would lead to different release conditions for the trapped carriers.

technology, and also the similar behavior of both phenomena. In terms of their variation with the dead time, as expected, they decrease when increasing t_{dt}. By further increasing t_{dt} it will be possible to reduce both P_{AP} and P_R, although this would limit the maximum photon count rate.

It can be also of interest not only to study the variation of P_{AP} and P_R with t_{dt}, but also the ratio of both quantities in terms of the dead time. In Fig. 7, the variation of this ratio with t_{dt} is illustrated.

Considering that these measurements have been performed at a fixed temperature and excess bias, reducing t_{dt} would mean a larger probability that the carriers are released during the afterpulsing phase, after the SPAD anode voltage has decreased below the comparator threshold. This would explain the increase in the proportion of afterpulsing events with regard to retriggering for reduced dead times, with the latter becoming less than two times the amount of afterpulsing for dead times below 130ns.

B. Dependence on Excess Bias Voltage

For the same group of SPADs previously characterized in terms of the dead time ($L = 20\mu$m), a study of the variation of P_{AP} and P_R has been carried out as a function of the excess bias. During this set of measurements, T has been fixed constant at 293K. The dead time has also been kept constant to 90ns. The results obtained for the probability of afterpulsing, retriggering and their sum are represented in Fig. 8.

Both afterpulsing and retriggering increase almost linearly with the excess bias voltage in the range considered here. There is a larger dependence on V_{EXC} for retriggering phenomena. Moreover, it can be seen that extrapolating afterpulsing probability with voltage, it tends to saturate and the intersection with the x-axis is far below 0. These issues will be discussed in Sect. IV.

An increase of the excess bias voltage means a larger number of carriers flowing through the device, which leads to a larger probability that some carriers are trapped and later released by trap states. Increasing V_{EXC}, the released carriers have also a larger probability of triggering a new avalanche.
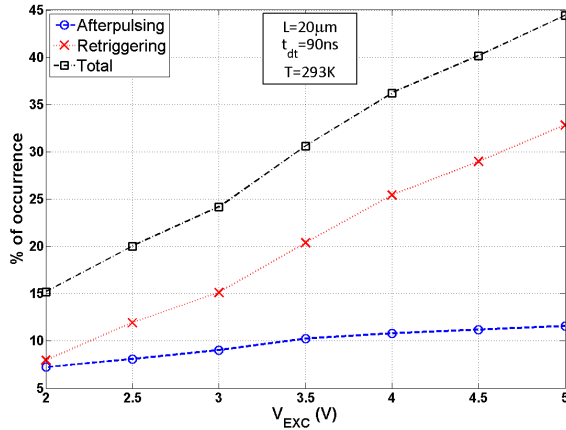
Fig. 8. Probability of occurrence of afterpulsing, retriggering events and their sum depending on the excess bias voltage.
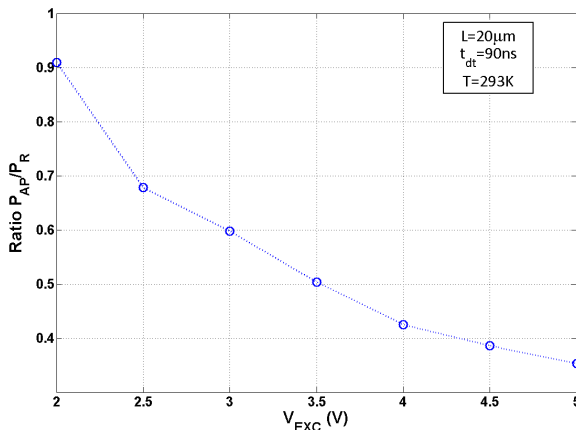

Fig. 9. Variation of the ratio between the probability of afterpulsing and retriggering with the excess bias voltage.

These facts are reflected in the results shown in Fig. 8, where both afterpulsing and retriggering rise with the excess bias.

Just as it was done with the dead time, the variation of the ratio between afterpulsing and retriggering has been studied depending on V_{EXC}. Fig. 9 shows the evolution of the quantity P_{AP}/P_R in terms of the excess bias. As it happened for larger dead times, the ratio between P_{AP} and P_R decreases for increasing values of V_{EXC}.

C. Dependence on Temperature

A detailed characterization in temperature has been carried out for all the SPADs sizes. For the measurements presented in this section, t_{dt} has been kept constant to 250ns. It is well known that P_{AP} increases at lower temperatures [12]. Therefore, to avoid P_{AP} and P_R to become too large while reducing the temperature, t_{dt} was set to 250ns, a value larger than in the previous section. Moreover, for each temperature different values of the excess bias voltage have been investigated (3, 5 and 7V).

Fig. 10 illustrates the variation of the probability of afterpulsing and retriggering with temperature for SPADs with different active areas and for several V_{EXC}. In all the cases, it can be seen that increasing temperatures lead to lower

values of both P_{AP} and P_R. This can be explained by the smaller release time by the traps, which lose the trapped carriers sooner in the recharge phase, when the probability of triggering a new avalanche is lower.

It can also be seen that for larger temperatures, there is a trend towards similar values of P_{AP} and P_R, regardless of the size of the SPADs or the value of V_{EXC}. Thus, at large temperatures, both quantities tend to be dominated by the trap release time, while they are not affect by the different numbers of trapped carriers, that are related to V_{EXC} and to the size of the SPADs.

Variations between SPADs of diverse active areas or working at different excess bias voltages can be specially noted at lower temperatures, when the trapped carriers are released later. In this situation, the probability of triggering a new correlated avalanche increases, so having more carriers flowing through the device and a larger probability of trapping them results in an increase of both P_{AP} and P_R.

As in the previous sections, the ratio P_{AP}/P_R has been studied in terms of the operating temperature. The values that have been obtained are shown in Fig. 11 and correspond to the nine different cases illustrated separately in Fig. 10.

It can be seen that regardless of the size of the SPADs and V_{EXC}, the dependence of the ratio with the operating temperature is very similar.

D. Dependence on SPAD Size

In the previous section, SPADs with different sizes were compared. There, the increase of both P_{AP} and P_R with the active area of the SPAD (see in Fig. 10) could be linked to a larger number of traps and a larger current flowing through the devices for SPADs of larger size. To focus on this behavior, the ratio P_{AP}/P_R for a fixed operating temperature (293K) and two different dead times has been studied in terms of V_{EXC} and the size of the SPADs. The experimental results are shown in Fig. 12. As it was expected, reducing t_{dt}, more carriers are released in the afterpulsing phase, leading to larger values of the ratio between P_{AP} and P_R.

It is to note that this plot illustrates completely different behaviors depending on t_{dt}. For the longer dead time, the variation of the ratio P_{AP}/P_R corresponds to what was also observed previously in Fig. 11: the ratio is almost independent on the size of the SPAD and the applied excess bias. In this range of dead times, both afterpulsing and retriggering are probably dominated by the same type of traps, which have a relatively long time constant, in the hundreds of nanoseconds range.

Fig. 12 also suggests that, at short dead times, retriggering should be affected by trap levels with shorter release time, whose effect can be appreciated only when speeding up the recharge phase. Afterpulsing, being caused by the release of carriers after the dead time, is less affected by these fast trapping centers and therefore, their ratio becomes unbalanced, as shown in Fig. 12.

IV. MATHEMATICAL MODEL

In this section we will focus on trying to model the experimental data that have been presented in the previous
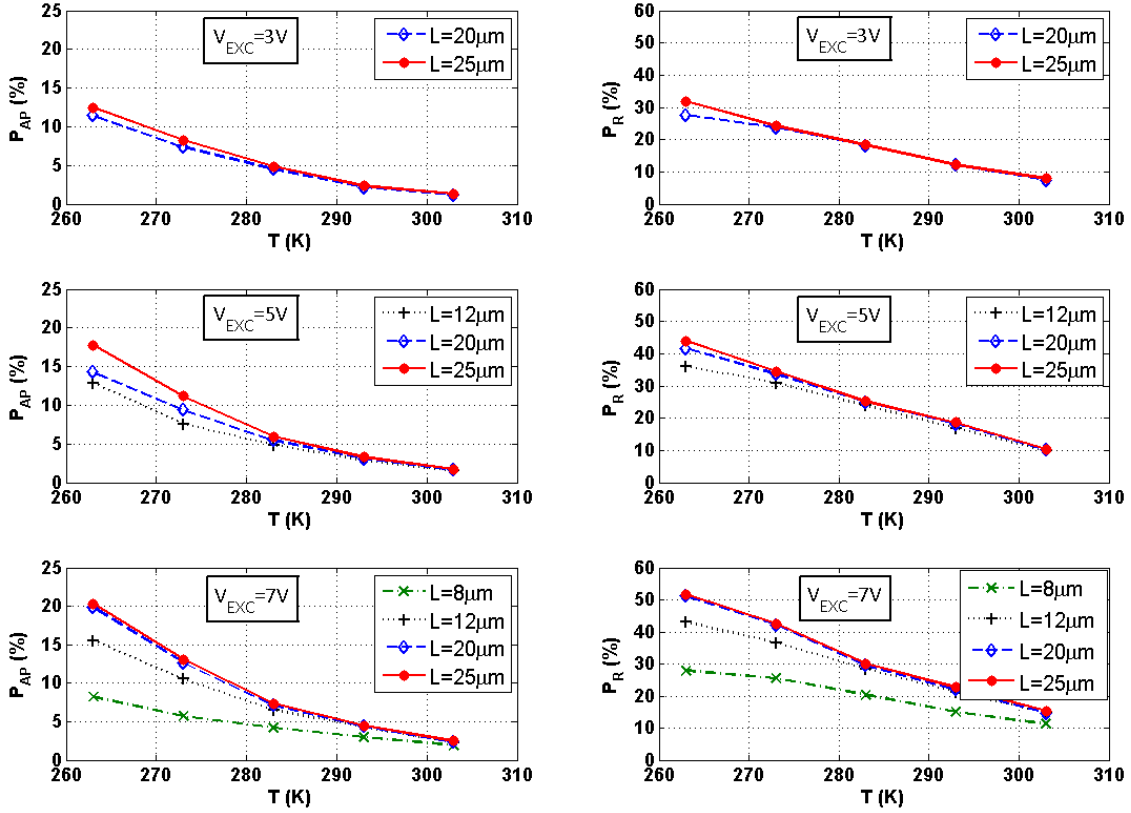
Fig. 10. Probability of afterpulsing and retriggering depending on temperature, considering different V_{EXC} and active areas.
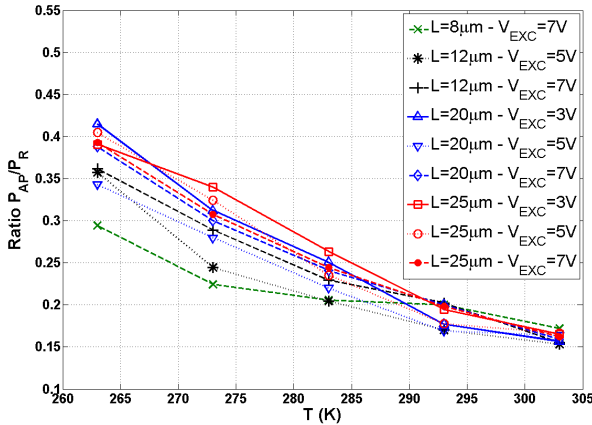

Fig. 11. Variation of P_{AP}/P_R with the operating temperature.
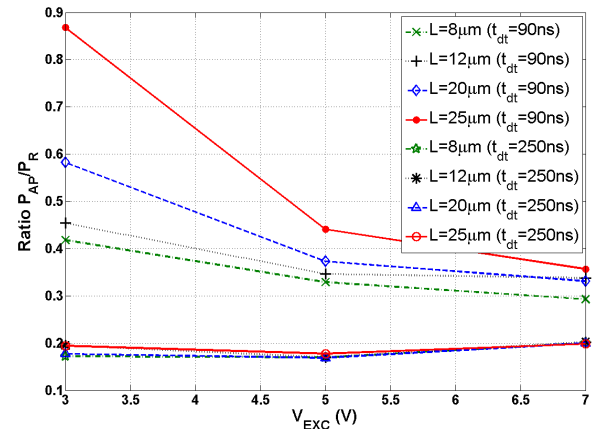

Fig. 12. Dependence of the ratio P_{AP}/P_R on the excess bias for SPADs of different size and for two fixed dead times.

sections regarding both the probability of afterpulsing and retriggering. Considering these data and previous works [8], we consider both quantities different manifestations of the same physical phenomena. Thus, the equations proposed in this section will have the objective to fit the data obtained for afterpulsing and retriggering. We will also study the behavior of their ratio in different working conditions.

The probability of occurrence of afterpulsing and retriggering phenomena can be modeled as the product of:

- Probability of capture by the traps: It depends on the number of traps and on the number of carriers flowing through the SPADs. The latter varies with V_{EXC}.

- Probability of release by the traps: This term depends on the dead time, but also on the operating temperature, since the release time by the traps varies with this parameter.

- Probability of triggering: Not all the released carriers will trigger an avalanche. The probability of triggering a new avalanche depends on the voltage applied to the SPAD, which varies with the time as illustrated in Fig. 1b (blue curve).

This will allow us to describe the behavior of P_{AP}, P_R and their ratio in terms of the excess bias, dead time, temperature
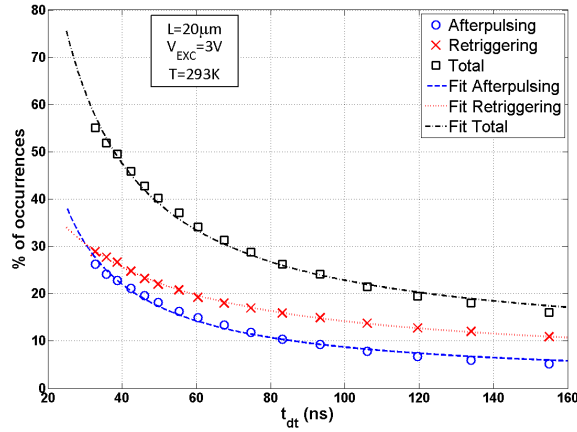
Fig. 13. Experimental and fitting results obtained when characterizing afterpulsing and retriggering in terms of dead time.

and SPAD size:

$$P_{AP,R} \propto P_{cap}(L, V_{EXC}) \cdot P_{rel}(T) \cdot P_{trig}(V(t_{dt})) \quad (1)$$

A. Dependence on Dead Time

First, we will focus on the dependence on t_{dt}. When the SPADs are biased above V_{BR}, the probability of triggering a new avalanche can be described as [13],

$$P_{trig} = 1 - \exp(\beta \cdot V(t_{dt})) \quad (2)$$

where β is a negative empirical constant [13] and $V(t_{dt})$ represents the voltage applied to the SPADs during the recharge phase. For PQCs, this voltage can be written with the following equation [10],

$$V(t_{dt}) = (V_{BR} + V_{EXC}) - \gamma \cdot \exp\left(-\frac{\epsilon}{t_{dt}}\right) \quad (3)$$

being γ and ϵ constants that depend on the SPAD and the conditions in which it operates. Eq. 3 is such that if t_{dt} increases, $V(t_{dt})$ decreases, as the recharge to the original bias conditions is slower.

Considering the previous equations and the fact that the other terms in Eq. 1 do not depend on t_{dt}, the probability of afterpulsing in terms of the dead time can be written as,

$$P_{AP,R}(t_{dt}) = \nu \cdot \left(1 - \exp\left(\eta - \psi \cdot \exp\left(-\frac{\epsilon}{t_{dt}}\right)\right)\right) \quad (4)$$

with ν, η and ψ being constants. η is determined by the empirical parameter β and the excess bias voltage, while ψ is a constant given by boundary conditions[4] and also by β. Those are the most critical parameters to get the fitting curves.

Eq. 4 has been used to describe the variation of P_{AP} and P_R with t_{dt} (see Fig. 13), with good fit of the model to the experimental data.

---

[4]It is necessary that the voltage applied to the SPAD at the end of the quenching phase –described by an equation different to Eq. 3– is equal to the SPAD voltage at the beginning of the recharge.

B. Dependence on the Excess Bias Voltage

Regarding the dependence on V_{EXC}, both the probability of triggering an avalanche and the probability of carrier capture by the traps should be considered. The last term, as stated before, is related to the current flowing through the SPADs, which, in the range of interest, is almost proportional to the applied voltage. Apart from this, we have to consider that the number of traps is finite and would saturate if the current flowing is high enough. Thus, this factor is modeled as,

$$P_{cap}(V_{EXC}) = (1 - \exp(-\chi \cdot V_{EXC})) \quad (5)$$

where χ is a constant. This term will tend to 0 when V_{EXC} is very small, and to 1 if V_{EXC} becomes very large. It will also be proportional to the excess bias voltage for small values of this parameter (considering the Taylor series expansion of Eq. 5). Therefore, it is possible to describe the proportional and saturation behaviors that have been observed depending on the value of V_{EXC}.

The parameter χ depends on the size of the SPAD, as the absolute number of traps varies with the dimensions of the device. Moreover, it is also function of the type of trap (fast or slow), since their concentration –and thus, their saturation level– can be different. Thus, χ is a critical parameter when fitting the experimental data to differentiate between afterpulsing and retriggering events.

For the probability of triggering an avalanche, we can refer to Eq. 2. Thus, combining Eqs. 2, 3 and 5, the probability of afterpulsing as a function of the excess bias could be defined as:

$$P_{AP,R}(V_{EXC}) \propto \left[1 - \exp(-\chi \cdot V_{EXC})\right] \\ \times \left[1 - \exp(\beta \cdot V_{EXC} - \kappa)\right] \quad (6)$$

where κ is derived from Eqs. 2 and 3:

$$\kappa = \beta \cdot \left(\gamma \cdot \exp\left(-\frac{\epsilon}{t_{dt}}\right) - V_{BR}\right) \quad (7)$$

Hence, we have studied the variation of $P_{AP,R}$ depending on the excess bias voltage based on three parameters $-\chi$, β and $\kappa-$ which define the different behaviors that have been observed for afterpulsing and retriggering in Fig. 8.

Fig. 14 shows that Eq. 6 describes the dependence on V_{EXC} of P_{AP}, P_R and their sum. In the case of afterpulsing, the results suggest that there is a saturation of the type of traps involved in the occurrence of this phenomenon.

Eqs. 4 and 5 (see also Fig. 15 as illustration, where E_T and τ represent the energy level and release time of the trap, respectively) can be used to explain the behavior that was shown in Figs. 11[5] and 12. Considering that afterpulsing can be described by a multiple exponential, if the dead time is increased, both P_{AP} and P_R will be reduced because only the slow components of the exponential (associated to the deeper levels) contribute. As will be seen in more detail later when studying the dependence on temperature, the deeper traps saturate for all the SPAD sizes and this leads to a constant

---

[5]In Fig. 11, the effect of temperature is changing the release time by the traps for a fixed t_{dt}. This would be equivalent to fix the operating temperature and vary t_{dt}, as in Fig. 12.
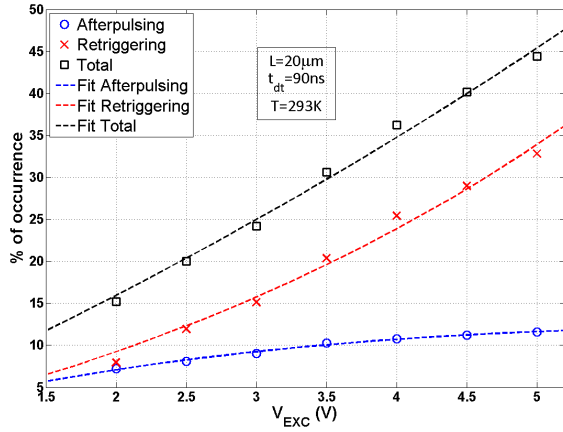
Fig. 14. Experimental and fitting results for the probability of afterpulsing and retriggering depending on the excess bias voltage.
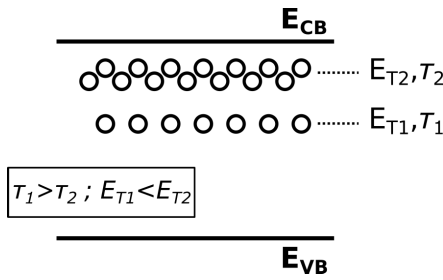

Fig. 15. Illustration of the different traps involved in the occurrence of afterpulsing and retriggering phenomena.

Fig. 16. Fit of the experimental results obtained for the ratio between P_{AP} and P_R versus V_{EXC} (red) and t_{dt} (black).

ratio P_{AP}/P_R for long enough dead times. This is what happens in Fig. 11 and the second case shown in Fig. 12, in which t_{dt} is equal to 250ns.

Moreover, if we consider the first set of measurements of Fig. 12 (dead time of 90ns), more traps would come into play. For the smaller devices ($L = 8, 12\mu$m), as the number of traps is lower, there is some impact of these additional traps that now contribute, but rapidly saturate with V_{EXC}. In the case of larger SPADs ($L = 20, 25\mu$m), this number of new traps contributing to P_{AP} and P_R is higher, and thus it is necessary to further increase V_{EXC} to reach the saturation state. In our case, for V_{EXC} at 7V, P_{AP}/P_R behaves similarly regardless of the SPAD size.

Eqs. 4 and 6 have also been used to describe the ratio of afterpulsing and retriggering in terms of V_{EXC} and t_{dt} (see Fig. 16 –in red and black, respectively–).

Analyzing more in detail Fig. 16 (red), we know that increasing V_{EXC} means a larger number of carriers flowing through the SPADs and a higher probability of triggering. Considering the latter, the differences due to a variation of the excess bias are maximum in the afterpulsing phase (see Eqs. 2 and 3). Consequently, the results in Fig. 16 (red) suggest that the fact which mainly determines the variation of P_{AP}/P_R for increasing V_{EXC} is the larger number of trapped/released carriers in the retriggering phase. This is in line with the saturation at smaller bias voltages of the traps characterizing afterpulsing that was previously highlighted.
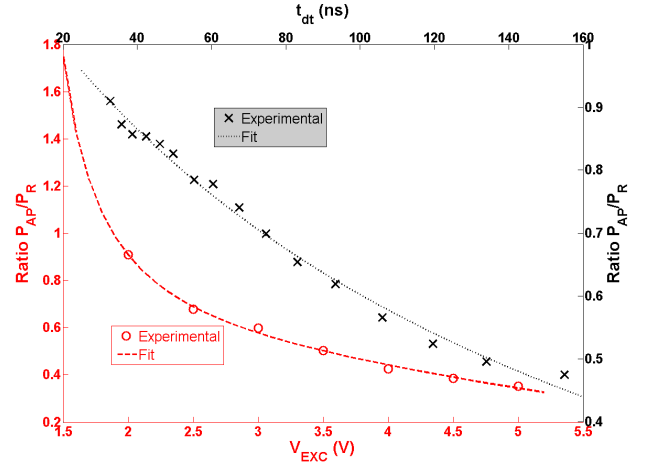
C. Dependence on Temperature

Next, we will focus on temperature, which will be included in the model through the probability of release by the traps. The release time by the traps can be described by [13],

$$\tau = \tau_0 \cdot \exp\left(\frac{E_{CB} - E_T}{k_B T}\right) \tag{8}$$

where E_{CB} and E_T represent the energy of the conduction band and of the trap, respectively (see Fig. 15). The other terms considered for modeling afterpulsing phenomena do not depend on the operating temperature. Thus, the probability of afterpulsing in terms of temperature can be written, based on Eq. 8, with the following expression,

$$P_{AP,R} = \Omega + \rho \cdot \exp\left(\frac{E_{CB} - \overline{E}_T}{k_B \cdot T}\right) \tag{9}$$

where Ω and ρ are constants that depend on the fixed working conditions (V_{EXC} and t_{dt}), k_B is Boltzmann's constant, E_{CB} is the conduction band energy and $\overline{E}_T$ (in eV) is a weighted linear combination that depends on the energy difference between the conduction band and the position of the trap levels in the band gap that contribute to both afterpulsing and retriggering.

This equation can be used to explain the behavior illustrated in Fig. 10. At high temperatures, the shallow traps are too fast, the overall behavior is dominated by the deep levels, and both P_{AP} and P_R decrease and saturate for all the sizes. If the temperature is reduced, the release time by the traps increases, and thus the impact of the shallow levels is higher. In that situation, the probability of afterpulsing and retriggering increase and depend on the dimension of the SPADs.

Table I shows the results obtained from the fit in temperature of P_{AP} and P_R. Using Eq. 9, we have calculated the difference in energy between the conduction band and a weighted contribution of the different traps involved in afterpulsing and retriggering phenomena. As can be seen, for afterpulsing the contribution of deeper levels is larger, while retriggering is more affected by shallower levels with smaller release time.

TABLE I

ESTIMATION OF THE ENERGY DIFFERENCE E_{CB}-$\overline{E}_T$ FROM THE FIT IN TEMPERATURE OF AFTERPULSING AND RETRIGGERING

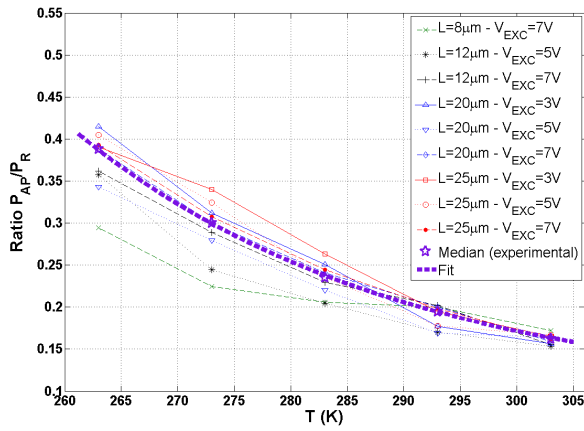|	E_{CB}-$\overline{E}_T$ (eV); AP	E_{CB}-$\overline{E}_T$ (eV); RET
L=8μm, V_{EXC}=7V	0.303	0.144
L=12μm, V_{EXC}=5V	0.305	0.191
L=12μm, V_{EXC}=7V	0.314	0.168
L=20μm, V_{EXC}=3V	0.330	0.181
L=20μm, V_{EXC}=5V	0.306	0.179
L=20μm, V_{EXC}=7V	0.304	0.196
L=25μm, V_{EXC}=3V	0.324	0.217
L=25μm, V_{EXC}=5V	0.318	0.228
L=25μm, V_{EXC}=7V	0.291	0.169

Fig. 17. Dependence of P_{AP}/P_R on the operating temperature for the different cases that have been studied. Together with this, the median value of the experimental data and its best fit curve.

For studying the ratio of afterpulsing and retriggering depending on the operating temperature, we have taken the median value (at each temperature) of the nine different case studies shown in Fig. 11. These data have been fitted using a ratio of expressions as the one shown in Eq. 9. The results obtained are illustrated in Fig. 17, with the best fit curve satisfying the following equation:

$$\frac{P_{AP}}{P_R} = \frac{0.28 + 2.56 \times 10^{-6} \cdot \exp\left(\frac{0.333}{k_B \cdot T}\right)}{1.97 + 5.17 \times 10^{-3} \cdot \exp\left(\frac{0.180}{k_B \cdot T}\right)} \quad (10)$$

Thus, when t_{dt} equals 250ns, Eq. 10 describes the general behavior in temperature of the proportion between afterpulsing and retriggering events for the SPADs studied in this paper, regardless of their dimension and V_{EXC}.

D. Dependence on SPAD Size

Finally, we will consider the size of the SPAD for modeling afterpulsing and retriggering phenomena. Looking at Eq. 1, we can see that the only factor which depends on the size of the SPADs is the probability of capture. This is an instantaneous probability which varies with the voltage applied to the SPAD, the type of trap and the area of the device as,

$$P_{cap}(L) \propto N_{trap}(L) \cdot \sigma_{trap} \quad (11)$$

with σ_{trap} representing the capture cross section, which is a function of the type of trap, and N_{trap} standing for the density of trapping centers which are occupied. N_{trap} depends on the device itself (through the total concentration of traps), but also on the current flowing through it. As it was stated previously (see Eq. 5), this quantity is a function of the excess bias voltage and also depends on the size of the SPAD (through the component of the capacitance which is function of L).

P_{cap} will increase with the number of carriers going through the device, till saturating when there are no more available unoccupied traps. If we reach this saturation state, we will not observe any dependence of the probability of afterpulsing on the size of the SPADs, as we have checked experimentally.

V. CONCLUSIONS

In this work, we have presented a test chip designed in a HV 0.35-μm CMOS process, containing SPADs of four different sizes ($L = 8, 12, 20, 25\mu$m). They are implemented as a 4×60 array and are accompanied by pure passive quenching circuits. We have performed a detailed description of afterpulsing and retriggering phenomena in terms of temperature, dead time, excess bias voltage and device area. This has been done through the study of the interarrival time and pulse width distributions, respectively. With the experimental results obtained, a mathematical model has been proposed for describing both P_{AP} and P_R. It is shown how retriggering phenomena can be described with the same model as afterpulsing. This result suggests that retriggering is linked to carriers trapped in the shallow levels of the semiconductor and that should be taken into account for a complete characterization of the SPADs in terms of the probability of afterpulsing. Moreover, this model might also be useful for predicting afterpulsing behavior when designing the electronics that will accompany the SPADs, considering also the fact that retriggering is significant in this technology and that this could lead to a considerable reduction of the dynamic range at high incident light intensities.

REFERENCES

[1] F. Zappa, S. Tisa, A. Tosi, and S. Cova, "Principles and features of single-photon avalanche diode arrays," *Sens. Actuators A, Phys.*, vol. 140, no. 1, pp. 103–112, Oct. 2007.
[2] G. F. Dalla Betta, L. Pancheri, D. Stoppa, R. Henderson, and J. Richardson, *Avalanche Photodiodes in Submicron CMOS Technologies for High-Sensitivity Imaging*. Rijeka, Croatia: InTech, 2011, ch. 11, pp. 225–248.
[3] J. A. Richardson, E. A. G. Webster, L. A. Grant, and R. K. Henderson, "Scaleable single-photon avalanche diode structures in nanometer CMOS technology," *IEEE Trans. Electron Devices*, vol. 58, no. 7, pp. 2028–2035, Jul. 2011.
[4] D. P. Palubiak and M. J. Deen, "CMOS SPADs: Design issues and research challenges for detectors, circuits, and arrays," *IEEE J. Sel. Topics Quantum Electron*, vol. 20, no. 6, pp. 409–426, Nov. 2014.
[5] D. E. Schwartz, E. Charbon, and K. L. Shepard, "A single-photon avalanche diode imager for fluorescence lifetime applications," in *Proc. IEEE Symp. VLSI Circuits*, Jun. 2007, pp. 144–145.
[6] A. Gallivanoni, I. Rech, and M. Ghioni, "Progress in quenching circuits for single photon avalanche diodes," *IEEE Trans. Nucl. Sci.*, vol. 57, no. 6, pp. 3815–3826, Dec. 2010.

[7] H. Finkelstein, "Shallow-trench isolation bounded single-photon avalanche diodes in commercial deep submicron CMOS technologies," Ph.D. dissertation, Univ. California, San Diego, San Diego, CA, USA, 2007.

[8] H. Xu, L. Pancheri, G. F. Dalla Betta, and D. Stoppa, "Design and characterization of p+/n-well SPAD array in 150 nm CMOS process," *Opt. Exp.*, vol. 25, no. 11, pp. 12765–12778, May 2017. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-25-11-12765

[9] L. H. C. Braga, L. Pancheri, L. Gasparini, R. K. Henderson, and D. Stoppa, "A mini-SiPM array for PET detectors implemented in $0.35\,\mu m$ HV CMOS technology," in *Proc. 7th Conf. Res. Microelectron. Electron.*, Jul. 2011, pp. 181–184.

[10] A. Rochas, "Single photon avalanche diodes in CMOS technology," Ph.D. dissertation, Ecole Polytechn. Federale Lausanne, Lausanne, Switzerland, 2003.

[11] C. Piemonte *et al.*, "Development of an automatic procedure for the characterization of silicon photomultipliers," in *Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf. Rec. (NSS/MIC)*, Oct. 2012, pp. 428–432.

[12] M. Liu, C. Hu, J. C. Campbell, Z. Pan, and M. M. Tashima, "Reduce afterpulsing of single photon avalanche diodes using passive quenching with active reset," *IEEE J. Quantum Electron.*, vol. 44, no. 5, pp. 430–434, Mar. 2008.

[13] G. Giustolisi, R. Mita, and G. Palumbo, "Behavioral modeling of statistical phenomena of single-photon avalanche diodes," *Int. J. Circuit Theory Appl.*, vol. 40, no. 7, pp. 661–679, Jul. 2012.

Manuel Moreno-García (M'16) received the five-year B.Sc. (with the National Award for Excellence in Academic Performance) degree from the Universidad de Sevilla, Spain, in 2009, and the M.Sc. degree in nuclear physics from WWW Münster, Germany, in 2009, the M.Sc. degree in microelectronics from the Universidad de Sevilla, Spain, in 2011, and the Ph.D. degree from the Instituto de Microelectrónica Sevilla, Sevilla, Spain, in 2015. Since 2015, he has been with the Integrated Radiation and Image Sensors (IRIS) Research Unit, Fondazione Bruno Kessler (FBK), Trento, Italy. His research interests include PPD and SPAD-based sensors, the simulation, design and characterization of semiconductor photon detectors for ToF imaging, and integrated circuits design.

Lucio Pancheri (M'11) received the Ph.D. degree from ICT in 2006. He is currently an Associate Professor of Electronics with the University of Trento, Italy. From 2006 to 2012, he was a Research Scientist with the Fondazione Bruno Kessler (FBK), Trento, Italy, where he has been involved in the development of optical and radiation sensors in CMOS technology for advanced imaging and biomedical applications. In this period, his research activity has been mainly dedicated to the development of single photon avalanche diodes and readout circuits for fluorescence lifetime detection and of electro-optical demodulating detectors for Time-of-Flight 3D image sensors. Since 2012, he has been a Researcher with the University of Trento. His research interests include the development of integrated optical radiation sensors for industrial and scientific applications.

Matteo Perenzoni (M'09) received the Laurea degree in electronics engineering from the University of Padova, Italy, in 2002. In 2002, he was with the University of Padova, where he worked on mixed-signal integrated circuit design for channel decoding. Since 2004, he has been a Researcher with the Center for Materials and Microsystems, Fondazione Bruno Kessler (FBK-CMM), Trento, Italy. He was collaborating on teaching courses of electronics and sensors for the NanoMicro Master, Trento, from 2006 to 2010. In 2011, he was the Director of the bi-annual Optoelectronics and Photonics Winter School co-organized with the University of Trento. In 2014, he was a Visiting Research Scientist with the THz Sensing Group, Technical University of Delft. Since 2017, he has been leading the Integrated Radiation and Image Sensors (IRIS) Research Unit, FBK. His research interests include the design of advance vision sensors for visible, X-ray, infrared, and terahertz, and modeling and optimization of analog integrated circuits. He has been a member of the Technical Program Committee of the European Solid-State Circuit Conference (ESSCIRC), since 2015 and the International Solid-State Circuit Conference (ISSCC), since 2017.

Rocío del Río (S'00–M'07) received the M.Sc. degree in electronic physics and the Ph.D. degree in microelectronics from the University of Seville, Spain, in 1996 and 2004, respectively. She joined the Department of Electronics and Electromagnetism, University of Seville, in 1995, where she is currently an Associate Professor. She has also been with the Instituto de Microelectrónica de Sevilla (IMSE-CNM, CSIC/University of Seville), since 1995, where she works with the Group of Analog and Mixed-Signal Microelectronics. She has co-authored more than 100 journal and conference papers, and has participated in the organizing committees of the diverse international conferences. Her main research interests include mixed-signal circuits (with special emphasis in switched-capacitor circuit techniques) and analog-to-digital converters, including analysis, behavioral modeling, and design automation. She has been serving as an Associate Editor of the *International Journal Circuit Theory and Application* (IJCTA, Wiley), since 2018.

Óscar Guerra Vinuesa received the M.Sc. degree in electronic physics and the Ph.D. degree in microelectronics from the University of Seville (US), Spain, in 1995 and 2002, respectively. Since 1997, he has been with the Instituto de Microelectrónica de Sevilla (IMSE-CNM, CSIC/University of Seville) within the Group of Analog and Mixed-Signal Microelectronics. In 2000, he joined the Department of Electronics and Electromagnetism, University of Seville, where he is currently an Associate Professor. He has authored or co-authored more than 70 articles in specialized publications (journals or conference papers) and has participated in the organizing committees of several international conferences. His main research interests include the design techniques of analog and mixed-signal integrated circuits, and modeling and design of communication front-ends in deep submicron CMOS technologies.

Ángel Rodríguez-Vázquez (F'99) co-founded the Instituto de Microelectrónica de Sevilla, a joint undertaken of Consejo Superior de Investigaciones Científicas (CSIC) and the Universidad de Sevilla, and started a Lab on Analog and Mixed-Signal Circuits for Sensors and Communications. In 2001, he was the main promoter of AnaFocus Ltd., where he served as CEO until 2009, when the company reached maturity as a worldwide provider of smart CMOS imagers. He also participated in the foundation of the Hungarian start-up AnaLogic Ltd. He has nine patents filed. AnaFocus was founded on the basis of his patents on vision chip architectures. His R&D production includes three generations of vision chips, analog front-ends for XDSL MoDems, ADCs for wireless communications, ADCs for automotive sensors, and complete MoDems for power-line communications, among many other mixed-signal SoCs. Many of these chips were state-of-the-art in their respective fields. Some of them entered massive production. He also produced teaching materials on data converters that were delivered to companies and received the Quality Label of EuroPractice. His publications have some 9,300 citations and several awards: the IEEE Guillemin-Cauer Best Paper Award, two Wiley's IJCTA Best Paper Awards, two IEEE ECCTD Best Paper Awards, one IEEE-ISCAS Best Paper Award, one SPIE-IST Electronic Imaging Best Paper Award, the IEEE ISCAS Best Demo-Paper Award, and the IEEE ICECS Best Demo-Paper Award. He has an h-index of 48 and an i10-index of 179 (Google Scholar).