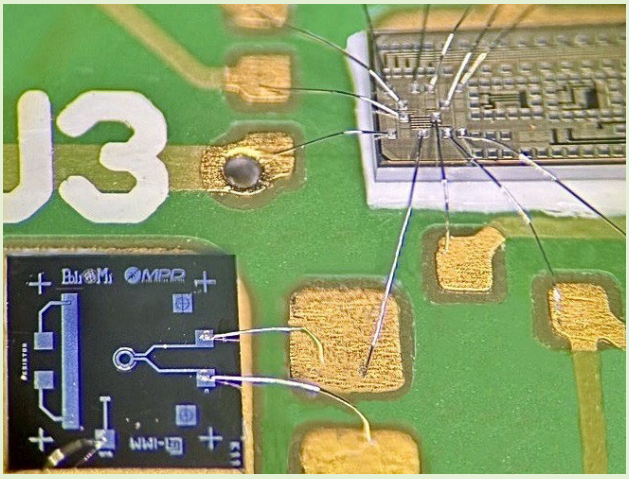
Double-Terminal Quenching Topology for Threefold After-Pulsing Reduction: Model and Experimental Validation

Francesco Malanga, *Graduate Student Member, IEEE*, Gennaro Fratta, *Graduate Student Member, IEEE*, Gabriele Laita, *Graduate Student Member, IEEE*, Angelo Gulinatti, *Member, IEEE*, Giulia Acconcia, *Member, IEEE*, and Ivan Rech, *Senior Member, IEEE*

***Abstract*—Single-photon avalanche diodes (SPADs) have emerged as crucial devices across a multitude of applications, ranging from fluorescence lifetime imaging (FLIM) to quantum technologies and light detection and ranging (LiDAR) systems. The increasing demand for fastening the acquisition rate of these applications has spurred significant interest in minimizing the SPAD dead time (DT) to a few nanoseconds. However, attempts to minimize its duration often exacerbate the after-pulsing (AP) phenomenon, posing a significant challenge in optimizing system performance. In this article, we propose a novel strategy to address this trade-off. We introduce a method that exploits passive or active quenching at the cathode terminal of SPADs, combined with an active quenching circuit (AQC) at the anode node. This combined approach aims at mitigating AP effects while simultaneously minimizing DT. We developed a comprehensive model and validation methodology to rigorously evaluate the effectiveness of this strategy. Finally, we demonstrate that how it is possible to achieve a strong reduction in AP compared with standard approaches.**



***Index Terms*— Active quenching circuit (AQC), after-pulsing (AP), Cadence Spectre, circuit model, dead time (DT), single-photon avalanche diode (SPAD).**

I. Introduction

THE remarkable features of single-photon avalanche diodes (SPADs) are today profitably exploited in a multitude of single-photon applications. For example, their usage is widely adopted in high-resolution fluorescence lifetime imaging (FLIM) and fluorescence spectroscopy (FS) [1], [2], [3], [4], [5], single-photon-based quantum applications [6], [7], [8], [9], [10], and time-of-flight (ToF) measurements in light detection and ranging (LiDAR) systems [11], [12], [13], [14], [15].

For single-photon detection, the SPAD is reverse biased above its breakdown voltage, operating in what is called the "Geiger mode," by a certain quantity defined as overvoltage. In such conditions, a macroscopic avalanche current occurs if a photon is absorbed in the active region of the junction. To stop the macroscopic avalanche current, the SPAD must be brought (*quenching*) and kept (*hold-off*) below its breakdown voltage and, afterward, reset to its initial bias condition (*reset*). The overall time needed to perform a photon-detection cycle is defined as dead time (DT), and it is the sum of quenching, hold-off, and reset times. A dedicated front-end circuit is used to control and execute different phases of a photon-detection cycle. Such a circuit can be generally referred to as a quenching circuit (QC), and it can be made using a passive, active, or mixed active–passive topology [16]. Nowadays, most

Manuscript received 1 July 2024; accepted 30 July 2024. Date of publication 7 August 2024; date of current version 13 September 2024. This work was supported in part by the Italian Ministero dell'Università e della Ricerca under Grant PRIN2022 (Proactive) and in part by the European Union (ERC, HÈRMES) under Grant 101116943. The associate editor coordinating the review of this article and approving it for publication was Dr. Chun Zhao. *(Francesco Malanga and Gennaro Fratta are co-first authors.) (Corresponding author: Francesco Malanga.)*

The authors are with the Department of Electronics, Information and Bioengineering (DEIB), Politecnico di Milano, 20133 Milan, Italy (e-mail: francesco.malanga@polimi.it).

Digital Object Identifier 10.1109/JSEN.2024.3436830

QCs use a mixed passive–active topology [active QC (AQC)], because it combines the immediate passive-quenching after avalanche ignition as well as a reliable and adjustable hold-off time of active-quenching. An AQC can be placed either at the anode or cathode terminal of a SPAD. Throughout this work, we will refer to these configurations as single-terminal (1T) topologies, which will be used as a point of comparison for our proposed architecture.

In recent years, the demanding requirements made by the most advanced applications led to a great interest in DT minimization, shortening it down to the range of a few nanoseconds [17], [18], [19], [20]. It is fundamental to underline how the DT cannot be shrunk indefinitely due to the limit imposed by the after-pulsing (AP) phenomenon. During an avalanche event, a fraction of the charge carriers flowing through the device can be trapped in energy states in the bandgap. If the SPAD is biased above breakdown before the traps have fully emptied, a detrapped carrier can initiate another avalanche, called an after-pulse [21]. This avalanche will be strictly correlated with the previously observed photon that triggered the AQC in the first place. On the contrary, in the hold-off phase, the released trapped carriers in the SPAD space-charge region will not generate secondary pulses because the biasing voltage is below the breakdown potential. Therefore, a long hold-off time is necessary to lower the AP, but it leads to a longer DT and thus to a reduced count rate of the overall structure.

Another possibility to lower the AP without resorting to a long hold-off time consists of limiting the total amount of current flowing through the device during the avalanche. So far, the great majority of efforts made to reduce the charge flowing after photon detection have focused on improving the AQC design to speed up the quenching phase. However, such an approach faces the limit posed by the SPAD parasitic elements at the quenching node and the chosen overvoltage [21]. Overcoming this limitation can be extremely challenging especially if devices with near-infrared capabilities and large overvoltage are considered [22], [23].

A possible solution to break the AP and DT trade-off has been presented in [24] with promising results. In principle, the solution consists of quenching both SPAD terminals to significantly reduce the charge flowing during the avalanche. However, that pioneering study focused only on a fully passive and differential model having all equal parasitic contributions. This is an unlikely situation if we consider the intrinsic dissimilarity between anode and cathode parasitisms, and the total absence of active elements in the model that are instead present in all AQCs.

In this article, we present and justify with an in-depth analysis all advantages and drawbacks related to the usage of an anode AQC, combined with a quenching resistance (R_{KQ}) at the cathode terminal (Fig. 1). We describe a comprehensive model and a mathematical approach to obtain the time-domain closed-form expressions of both anode and cathode nodes for a complete photon detection cycle. The presented model is then validated by means of Cadence Spectre[1] simulator and experimental characterization, demonstrating how it is suitable for comparing 1T and double-terminal (2T) architectures. Finally, we highlight the primary limitation of adopting the 2T architecture and demonstrate the effectiveness of a potential solution to address it.

[1]Registered Trademark.

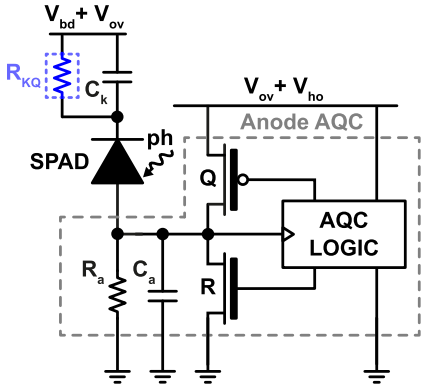


Fig. 1. 2T architecture with an AQC at the anode node combined with a quenching resistance (R_{KQ}) at the cathode node.

This article is organized as follows. Section II presents and demonstrates the model. In Section III, the model is validated in different scenarios. Section IV explains the working principle of the aforementioned solution and conclusion are then drawn in Section V.

II. Photon-Detection Cycle Modeling

A complete model is crucial to comprehend why resistance diminishes AP, how to dimension it effectively, the trade-offs involved, and the circumstances under which this solution proves advantageous. Our model accurately reproduces voltage trends at nodes, facilitating a straightforward evaluation of how key parameters like charge flowing into the SPAD and quench time depend on the electrical variables at play. To extract these results, first, a complete photon-detection cycle must be analyzed. This can be described by focusing on the 2T AQC architecture depicted in Fig. 1. Similar considerations can also be done for 1T architectures. In the figure, it is possible to identify passive elements deriving from parasitisms (C_k and C_a), passive quench (PQ) resistances (R_{KQ} and R_a), transistors for active quenching and reset, and AQC logic managing different phases. The two passive elements shown in the figure are a strong simplification to take into account all parasitic contributions. These will be better addressed at the end of this section.

As previously mentioned, to work in "Geiger mode," the SPAD is biased above breakdown voltage (V_{bd}) by a certain quantity defined as over-voltage (V_{ov}). At first, during the PQ phase, the current generated after photon detection flows through the anode and cathode impedance while both transistors are disabled. This results in a rising voltage at the anode node and a falling voltage at the cathode node. In 2T architectures, the passive quenching phase reduces the SPAD voltage difference acting at both terminals. Next, the anode rising voltage triggers the active quench (AQ) phase, and the corresponding quenching pMOS is activated forcing an abrupt transition of the SPAD below the breakdown of a certain quantity. This sub-breakdown voltage can be evaluated

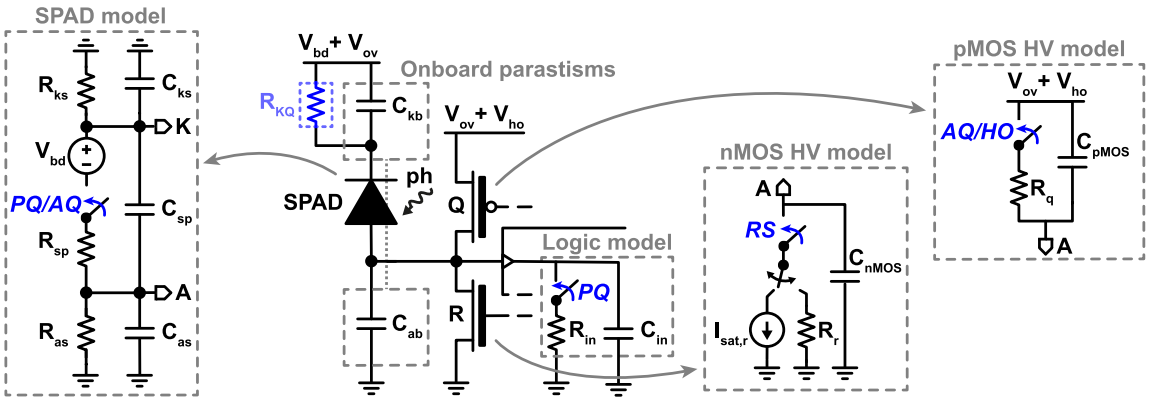


Fig. 2. 2T AQC equivalent circuit explanation. All major parasitisms are considered: onboard parasitisms (C_{ab} and C_{kb}), input logic equivalent impedance (C_{in} and R_{in}), quench and reset MOS introduced capacitances, and all parasitic effects toward the substrate for the SPAD. Every switch is closed only during the associated phase.

as the V_{ho} voltage. Due to capacitive coupling along the terminals, the cathode voltage rises following the anode voltage. The active quenching suppresses the photon-generated current flowing inside the device, fully quenching it. The AQC logic keeps the SPAD below breakdown for the total amount of the hold-off time (HO phase). Finally, the quenching pMOS is turned-off and the Reset nMOS is activated, forcing the anode back to the initial-bias ground potential (RS phase). In this phase, the cathode terminal is initially pulled down due to capacitive coupling and then passively restored by the R_{KQ} resistance to its bias potential.

As a first step in modeling a complete photon-detection cycle, all parasitic contributions must be considered. In Fig. 2, the model used is depicted, elucidating the cause of parasitic contributions for each element. To represent the circuital behavior of the SPAD, we employed a well-established electrical model found in the literature [25], [26]. For the sake of completeness, all parasitic effects toward the substrate are taken into account. Onboard parasitisms are considered through C_{kb} and C_{ab}. The AQC input logic equivalent impedance is represented by the elements C_{in} and R_{in}. Models for the quenching pMOS and the reset nMOS are provided. The notable difference between these two elements lies in the possibility of considering them to operate either solely in the linear region (pMOS) or in both the saturation and linear regions (nMOS) when activated. This distinction will be clarified later in the discussion in Section II-D. Finally, all switches are closed only during the associated phase.

To simplify the time-domain discussion based on the opening and closing phases of switches, our model strategically uses step-like functions, allowing for the study of the circuit in the Laplace domain. The time-domain expressions for the signals of interest, i.e., anode and cathode voltage waveforms, are then obtained by applying the Laplace antitransform operator. The key element to ensure the effectiveness of the model is to recreate the boundary conditions for two subsequent phases. In this way, the amplitude of the input step-like generators is modified according to the phase.

To this aim, in any phase, the transfer function features two poles and one zero for both cathode and anode, represented as follows:

$$\text{TF}(s) = K \cdot \frac{cs+1}{as^2+bs+1} = P \cdot \frac{(s\tau_z+1)}{(s\tau_l+1)(s\tau_h+1)} \tag{1}$$

where a, b, and c are the coefficients of the singularities, P and K are the proportionality coefficients, where $P = (K/a)$, τ_l and τ_h are the time constants of the high-frequency and low-frequency poles, respectively, once the expression has been solved, and τ_z is the zero time constant. The antitransform expression for a step-like input stimulus is given by

$$\begin{aligned} v(t) &= \mathcal{L}^{-1}\left\{\frac{V_0}{s}\cdot \text{TF}(s)\right\} \\ &= V_0 \cdot P \cdot \left(1 - Ae^{-\frac{t}{\tau_h}} - Be^{-\frac{t}{\tau_l}}\right) \end{aligned} \tag{2}$$

where (V_0/s) is the amplitude of the step-like input signal, and A and B are the two exponential coefficients. These coefficients are directly dependent on the values of the singularities, and it can be easily demonstrated that they can be expressed as follows:

$$A = \frac{\tau_l - \tau_z}{\tau_l - \tau_h} \quad B = \frac{\tau_z - \tau_h}{\tau_l - \tau_h} \tag{3}$$

and it is always true how $A + B = 1$. Once the strategy to evaluate a complete photon-detection cycle has been clarified, the phases that compose it can be studied individually.

A. PQ Phase

The equivalent circuit responsible for this phase is illustrated in Fig. 3(a). Starting from all the electrical elements presented in Fig. 2, several parasitisms persist throughout every phase and can be simplified as follows:

$$R_k = R_{KQ} \parallel R_{ks} \simeq R_{KQ} \tag{4}$$

$$C_k = C_{ks} + C_{kb} \tag{5}$$

$$C_a = C_{as} + C_{ab} + C_{\text{in}} + C_{n_{\text{MOS}}} + C_{p_{\text{MOS}}}. \tag{6}$$

Here, R_k represents the total cathode resistance, and C_k and C_a denote the overall parasitic capacitances at the cathode and anode nodes, respectively. In our study, R_{KQ} will never reach a value comparable with the cathode-substrate resistance (R_{ks}) of the SPAD, simplifying the expression in (4). Indeed, the

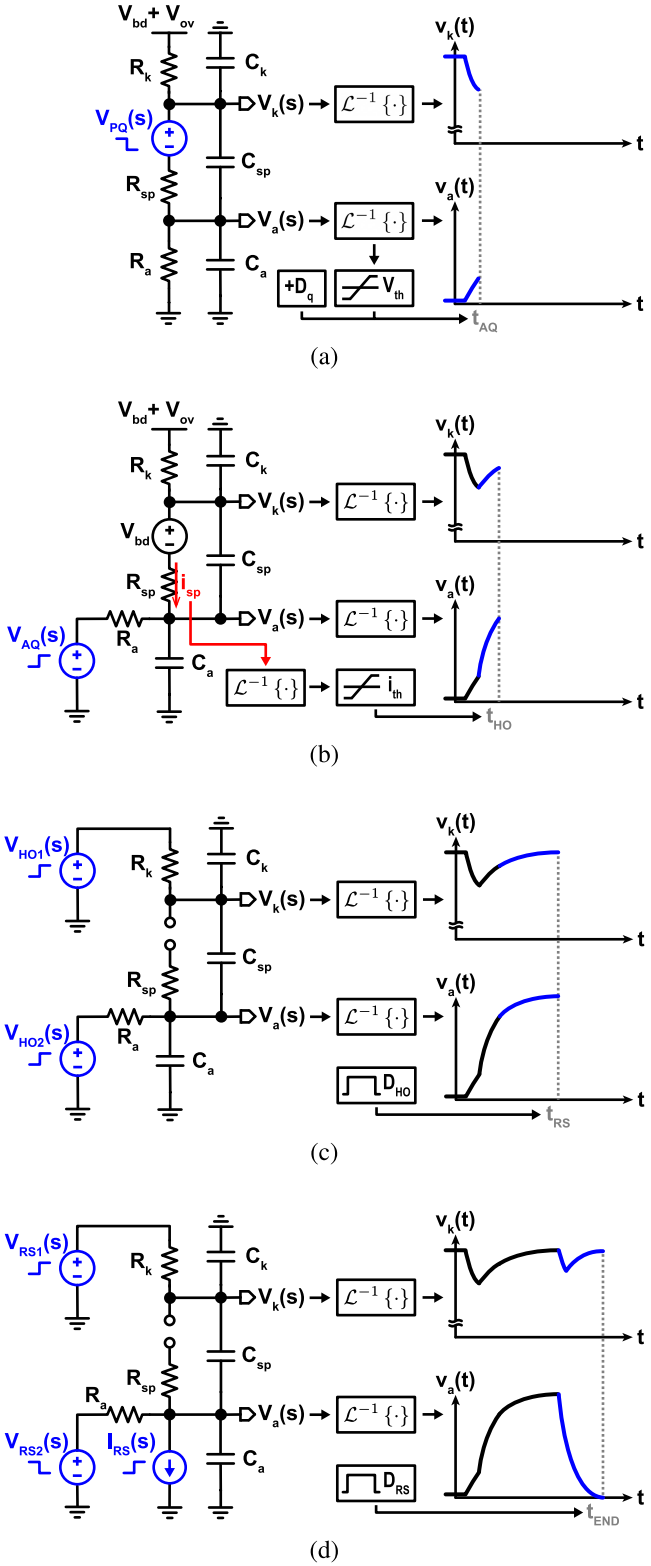


Fig. 3. Complete AQC cycle description in the proposed switching-topology Laplace domain model. (a) In PQ, the photon-detection is represented by the depicted voltage source. (b) AQ phase enabling the quench transistor working in the linear region. (c) Hold-off phase after the SPAD has been fully quenched. (d) Reset phase enabling reset transistor working both in saturation and linear regions.

maximum value considered for R_{KQ} will be around 10 kΩ, while the cathode-substrate resistance R_{ks} is estimated to start from at least 100 kΩ. The value of R_a varies depending on the phase under consideration and it must be studied separately. During passive quenching, it is

$$R_a = R_{\text{in}} \parallel R_{as}. \tag{7}$$

The coefficients of the common denominator in the transfer function can be computed with the time-constant method [27] as follows:

$$a = \frac{R_k R_a R_{\text{sp}}}{R_k + R_{\text{sp}} + R_a} \cdot (C_a C_k + C_a C_{\text{sp}} + C_k C_{\text{sp}}) \tag{8}$$

$$b = C_a R_a \parallel (R_{\text{sp}} + R_k) + C_k R_k \parallel (R_{\text{sp}} + R_a) + C_{\text{sp}} R_{\text{sp}} \parallel (R_a + R_k) \tag{9}$$

while the proportionality coefficients and the zero time constant τ_z change according to the selected output. The complete voltage expressions of the anode and cathode in the Laplace domain as a function of the input step-like voltage source $V_{PQ}(s)$ can be written as follows:

$$V_k(s)|_{PQ} = \frac{V_{\text{ov}} + V_{\text{bd}}}{s} - V_{PQ}(s) \cdot \frac{R_k}{R_k + R_{\text{sp}} + R_a} \cdot \frac{1 + sC_a R_a}{as^2 + bs + 1} \tag{10}$$

$$V_a(s)|_{PQ} = V_{PQ}(s) \cdot \frac{R_a}{R_k + R_{\text{sp}} + R_a} \cdot \frac{1 + sC_k R_k}{as^2 + bs + 1}. \tag{11}$$

It is worth noticing how any term expressed as (V/s) will set the bias starting point, while $(V/s) \cdot e^{-s\cdot\delta}$ term is going to be the input step generator, delayed by a time δ.

Before the photon arrival, the voltage source V_{PQ} starts from $V_{\text{ov}} + V_{\text{bd}}$, resulting into no current through R_{sp}, and the anode and cathode voltages equal to $V_k = V_{\text{ov}} + V_{\text{bd}}$ and $V_a = 0$ V, respectively. When a photon is detected, there is an instantaneous voltage drop of V_{PQ} to V_{bd} that generates the current flowing into the SPAD. Hence, the Laplace-domain expression of the step voltage generator is as follows:

$$V_{PQ}(s) = \frac{V_{\text{ov}} + V_{\text{bd}}}{s} - \frac{V_{\text{ov}}}{s} \cdot e^{-s\cdot t_{\text{ph}}} \tag{12}$$

where t_{ph} is the generic time instant of a photon striking the SPAD. Following the same procedure explained in (2) and (3), it is possible to derive the time-domain behavior of the anode and cathode nodes in this phase.

The passive phase lasts until the active one is triggered when a certain threshold V_{th} is crossed by the anode node voltage. After a logical delay D_q, the active quenching starts, ending the passive one at the time instant t_{AQ}. In 2T architectures, there is a higher delay in triggering the next phase compared with regular 1T architectures. This is caused by a reduction in amplitude at the anode node in 2T approaches due to capacitive partition. This delay can reach the maximum value of

$$\Delta T|_{PQ} \simeq \frac{V_{\text{th}}}{V_{\text{ov}}} \cdot R_{\text{sp}} \cdot \frac{C_a C_{\text{sp}}}{C_k} \tag{13}$$

above a certain value of the R_k resistor considered. This result will be clarified later in the discussion in Section III.

During PQ, the majority of the current flows through the photodetector. The SPAD current waveform is controlled by a decreasing exponential expression as follows:

$$i_{\text{sp}}(t)\big|_{PQ} \simeq \frac{V_{\text{ov}}}{R_{\text{sp}}} \cdot e^{-\frac{t - t_{\text{ph}}}{\tau_h|_{PQ}}} \cdot h\big(t - t_{\text{ph}}\big) \tag{14}$$

where $h(t - t_{\text{ph}})$ represents the Heaviside step function, and $\tau_h|_{PQ}$ denotes the fast time constant (τ_h) as described in (1). This expression is applicable to both 1T and 2T architectures, and the effect of the slow pole is negligible if the threshold voltage V_{th} is sufficiently small, and the input logic delay D_q is short enough. In this article and selecting a value of $R_a \gg R_{\text{sp}}$ and R_k to minimize the passive quenching duration, $\tau_h|_{PQ}$ can be expressed according to the case as follows:

$$\tau_h|_{PQ} \simeq \begin{cases} R_{\text{sp}} \cdot (C_{\text{sp}} + C_a), & \text{1T arch} \\ R_{\text{sp}} \cdot (C_{\text{sp}} + C_a \parallel C_k), & \text{2T arch.} \end{cases} \tag{15}$$

Once the parasitisms are set, the current time-constant $\tau_h|_{PQ}$ ranges from the maximum value for 1T approaches down to the minimum possible value obtainable in 2T architectures with a value of R_k greater than $5 \cdot R_{\text{sp}}$ resistor. As a consequence, the charge (therefore the AP) present during passive quenching in a 2T approach is always lower than the one flowing in a 1T architecture. This model also confirms the AP reduction reported in [24].

B. AQ Phase

When the AQC is triggered, the active quenching phase starts. All resistances at the anode node are entirely masked by the linear-region equivalent resistance R_q of the quenching pMOS. Moreover, in several AQCs present in the literature [17], [18], the input logic resistance R_{in} is increased during quench to accelerate the passive phase, becoming even more negligible when compared to R_q. The behavior of the pMOS can be represented as operating solely in the linear region as the drain–source voltage starts from a point close to pinch-off due to the rise in the anode voltage during passive quenching.

In Fig. 3(b), the active quenching model is represented. The model is almost equal to the one represented in Fig. 3(a), thus the transfer-function denominator coefficients are exactly the same as the ones reported in (8) and (9). The sole difference is accounted by considering

$$R_a \simeq R_q. \tag{16}$$

If we explicit the zero time-constant, the Laplace domain expressions of anode and cathode nodes in this phase are

$$V_k(s)|_{AQ} = \frac{v_k|_{PQ}(t_{AQ})}{s} + V_{AQ}(s) \cdot P_k|_{AQ} \cdot \frac{1 + sC_{\text{sp}}R_{\text{sp}}}{as^2 + bs + 1} \tag{17}$$

$$V_a(s)|_{AQ} = \frac{v_a|_{PQ}(t_{AQ})}{s} + V_{AQ}(s) \cdot P_a|_{AQ} \cdot \frac{1 + s(C_{\text{sp}} + C_k)(R_{\text{sp}} \parallel R_k)}{as^2 + bs + 1} \tag{18}$$

where $V_{AQ}(s)$ is the input voltage source depicted in Fig. 3(b), $v_x|_{PQ}(t_{AQ})$ is the starting point value sampled as the last time instant of the previous phase (t_{AQ}), and $P_x|_{AQ}$ is the proportionality coefficient. The input signal $V_{AQ}(s)$ can be computed as a quenching step signal, with the following expression:

$$V_{AQ}(s) = \frac{V_{\text{ov}} + V_{\text{ho}} - v_a|_{PQ}(t_{AQ})}{s} \cdot e^{-s \cdot t_{AQ}} \tag{19}$$

in this case, the amplitude of the input signal has been modified to take into account also the previous PQ contribution up to this phase. This translates into diminishing the amplitude of $V_{\text{ov}} + V_{\text{ho}}$ by a $v_a|_{PQ}(t_{AQ})$ quantity. Adopting such an approach preserves the boundary conditions despite the change of topology while passing from one phase to the next one.

A fundamental consideration must be made from now on. While using a single step-like input generator, it can be easily demonstrated how the transfer function will lead to an erroneous steady-state value of the anode and cathode nodes. This is due to the fact that the constant voltage sources are not taken into account as contributions in this phase. However, these are restoring the capacitances, charged in the previous phase, back to their initial bias point. To solve this conundrum, it is possible to exploit a generalization of the final-value theorem [28], thus calculating both proportionality coefficients to match the expected steady-state value in this phase. Being the expected steady-state values the following:

$$V_k|_{AQ}^{\infty} = \frac{V_{\text{ov}} \cdot (R_q + R_{\text{sp}}) + V_{\text{ho}} \cdot (R_k)}{R_q + R_{\text{sp}} + R_k} + V_{\text{bd}} \tag{20}$$

$$V_a|_{AQ}^{\infty} = \frac{V_{\text{ov}} \cdot (R_q) + V_{\text{ho}} \cdot (R_{\text{sp}} + R_k)}{R_q + R_{\text{sp}} + R_k} \tag{21}$$

the proportionality coefficients can be derived

$$P_k|_{AQ} = \frac{V_k|_{AQ}^{\infty} - v_k|_{PQ}(t_{AQ})}{V_{\text{ov}} + V_{\text{ho}} - v_a|_{PQ}(t_{AQ})} \tag{22}$$

$$P_a|_{AQ} = \frac{V_a|_{AQ}^{\infty} - v_a|_{PQ}(t_{AQ})}{V_{\text{ov}} + V_{\text{ho}} - v_a|_{PQ}(t_{AQ})}. \tag{23}$$

The signals can now be antitransformed to obtain the time-domain expressions. The current flowing through the SPAD inner resistance R_{sp} is obtained from the anode and cathode expressions. This is needed because the condition which starts the hold-off phase at the time instant t_{HO} is triggered by the SPAD current falling below a certain current threshold I_{th}. This value changes according to the considered photodetector; in our example, we can set it to the most typical value of 100 μA.

During the active quenching phase, the cathode voltage moves in the same direction as the anode quenched terminal. Compared with a 1T architecture, this could be seen as a disadvantage because it delays the time instant when the photocurrent goes under threshold. However, if a large enough value of R_k is considered ($\geq 5 \cdot R_{\text{sp}}$), it is possible to obtain a closed-form first-order approximation of the time instant at which the next phase starts. Both anode and cathode are dominated by the high-frequency pole contribution and their time-domain expressions become

$$v_k|_{AQ}(t) \simeq v_k|_{PQ}(t_{AQ}) + \frac{A_k|_{AQ} P_k|_{AQ}}{\tau_h|_{AQ}} \cdot t \tag{24}$$

$$v_a|_{AQ}(t) \simeq v_a|_{PQ}(t_{AQ}) + \frac{A_a|_{AQ} P_a|_{AQ}}{\tau_h|_{AQ}} \cdot t \tag{25}$$

where $A_x|_{AQ}$ is the high-frequency proportionality coefficient reported in (3). This leads to the following equation for the

time instant when the current goes under the threshold:

$$\Delta T|_{AQ} = \frac{R_{sp} I_{th} - \Delta V|_{PQ} - V_{bd}}{A_k|_{AQ} P_k|_{AQ} - A_a|_{AQ} P_a|_{AQ}} \cdot \tau_h|_{AQ} \tag{26}$$

where $\Delta V|_{PQ}$ is the voltage difference between the anode and cathode starting point during active quenching phase. If this value is compared with the 1T architecture approach, it leads to a time increment only for a small interval of R_k values ($\simeq R_{sp}$). Also, this consideration will be better appreciated in Section III-A.

C. Hold-Off Phase

During the hold-off phase, the switch inside the SPAD model shown in Fig. 2 is opened. Anode and cathode terminals are connected only through C_{sp} capacitance, while they are reaching the bias point fixed by the two input voltage source generators depicted in Fig. 3(c). Due to the change of topology in the circuit, both numerator and denominator coefficients of all transfer functions have to be re-evaluated. In this phase, denominator coefficients can be expressed as follows:

$$a = R_a R_k \cdot (C_a C_k + C_a C_{sp} + C_k C_{sp}) \tag{27}$$

$$b = R_a \cdot (C_a + C_{sp}) + R_k \cdot (C_k + C_{sp}) \tag{28}$$

where R_a is preserving the same value reported in (16). The expressions of anode and cathode nodes in this phase are a one-output–two-input type as the following one:

$$\begin{bmatrix} V_k(s)|_{HO} \\ V_a(s)|_{HO} \end{bmatrix} = \begin{bmatrix} 1 + sR_a(C_a + C_{sp}) & sC_{sp}R_k \\ sC_{sp}R_a & 1 + sR_k(C_k + C_{sp}) \end{bmatrix} \times \begin{bmatrix} V_{HO1}(s) \\ V_{HO2}(s) \end{bmatrix} \cdot \frac{1}{D(s)} + \begin{bmatrix} v_a|_{AQ}(t_{HO}) \\ v_a|_{AQ}(t_{HO}) \end{bmatrix} \tag{29}$$

where $V_{HOX}(s)$ is the input voltage source signals, $v_x|_{AQ}(t_{HO})$ is the bias starting point computed from the previous phase, and $D(s)$ is the expression of a two-pole denominator reported in (1).

The amplitude of the input signals is modified again to consider the starting bias point from the previous phase. The expression of the input signal generators are

$$V_{HO1}(s) = \frac{V_{bd} + V_{ov} - v_k|_{AQ}(t_{HO})}{s} \cdot e^{-s \cdot t_{HO}} \tag{30}$$

$$V_{HO2}(s) = \frac{V_{ov} + V_{ho} - v_a|_{AQ}(t_{HO})}{s} \cdot e^{-s \cdot t_{HO}} \tag{31}$$

while the cathode voltage is reaching its bias point back to the original condition, the anode is continuing moving toward the $V_{ov} + V_{ho}$ potential. No changes to the proportionality coefficients are needed here because the steady-state value computed in the Laplace domain coincides with the expected value for both anode and cathode. This is due to the fact that all voltage sources are considered.

As mentioned in the introduction, one of the greatest advantages of active AQCs is to control the duration of the hold-off phase. For this reason, the time instant when this phase ends (t_{RS}) is externally controlled to have an overall D_{HO} duration. For this reason, there is no relevant time analysis to be done in this phase. However, it is worth noting how the model is robust and perfectly works even in the scenario when a D_{HO} duration too short is selected, thus either the anode or cathode node do not reach the steady-state value.

D. Reset Phase

This last phase begins with the activation of the Reset nMOS and the deactivation of the quenching pMOS to restore ground potential at the anode node. The MOS is operating both in linear and saturation regions due to a wide voltage range across its terminals. This enhances the modeling accuracy to better represent the Reset phase, and a similar consideration may be done also in the active quenching phase, if a large V_{ho} subbreakdown voltage is considered. In Fig. 3(d), all relevant input stimuli generators are depicted.

At first, the constant current generator is the only one active. It is possible to ignore all other generators being completely irrelevant compared to this dominant contribution. The transfer function from the input current generator to anode and cathode terminals differs with respect to the one reported in (1). However, the expressions for the SPAD terminals signals in the first part of the reset phase can be derived as follows:

$$V_k(s)|_{RS} = \frac{v_k|_{HO}(t_{RS})}{s} - I_{RS}(s) \cdot \frac{C_{sp}}{C_a + C_{sp}} \cdot \frac{1}{1 + s(C_k + C_{sp} \parallel C_a)R_k} \tag{32}$$

$$V_a(s)|_{RS} = \frac{v_a|_{HO}(t_{RS})}{s} - \frac{I_{RS}(s)}{s} \cdot \frac{1}{C_a + C_{sp}} \cdot \frac{1 + sC_kR_k}{1 + s(C_k + C_{sp} \parallel C_a)R_k} \tag{33}$$

where $I_{RS}(s)$ is the input current generator and $v_x|_{HO}(t_{RS})$ is the terminal bias starting point. The single pole of the transfer function is $\tau|_{RS} = (C_k + C_{sp} \parallel C_a)R_k$. The Laplace-domain expression of the step current generator is

$$I_{RS}(s) = \frac{I_{n_{SAT}}}{s} \cdot e^{-s \cdot t_{RS}} \tag{34}$$

where $I_{n_{SAT}}$ is the saturation current of the Reset transistor, and t_{RS} is the time instant triggered by the end of the previous phase. When the nMOS works in saturation, this first part of the Reset phase can be seen as just a current generator discharging a capacitance. The time-domain expression of anode and cathode terminal can be derived by applying the $\mathcal{L}^{-1}\{\cdot\}$ operator to the expression in (32) and (33). These become

$$v_k|_{RS}(t) = v_k|_{HO}(t_{RS}) + I_{n_{SAT}} \cdot \frac{C_{sp}}{C_a + C_{sp}} \cdot R_k \cdot e^{-\frac{t - t_{HO}}{\tau|_{RS}}} \tag{35}$$

$$v_a|_{RS}(t) = v_a|_{HO}(t_{RS}) - I_{n_{SAT}} \cdot \frac{1}{C_a + C_{sp}} \cdot \left[t - R_k \cdot (C_{sp} \parallel C_a) \cdot \left(1 - e^{-\frac{t - t_{HO}}{\tau|_{RS}}} \right) \right]. \tag{36}$$

The anode voltage waveform is given by the difference of two independent contributions: the linear discharge of the anode equivalent capacitance by the saturation current and a minimal charging effect given by the R_k resistance on this node. According to the selected value of R_k (i.e., R_{KQ}), these two effects change their weight in the previous expression. The greatest drawback in adopting a 2T architecture is due to the discharge of the cathode terminal occurring during the first

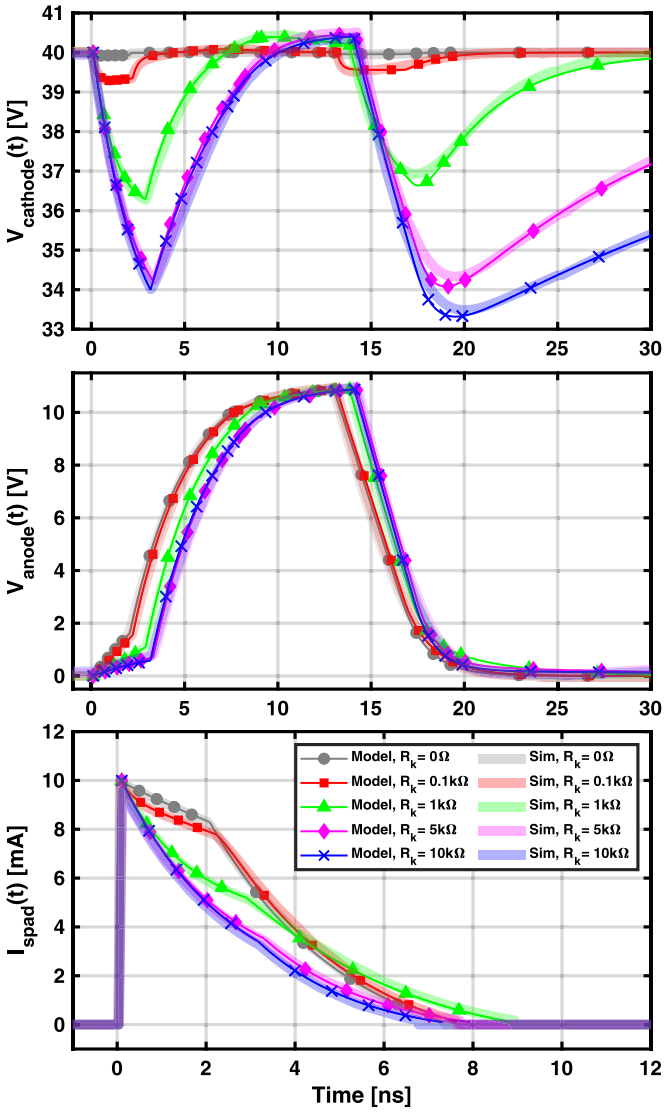


Fig. 4. Model validation though Cadence Spectre[1] simulation on a sweep of R_k resistor values. Cathode and anode voltages are shown for the whole duration of a photon-detection cycle. Photodetector current is shown only during quench.

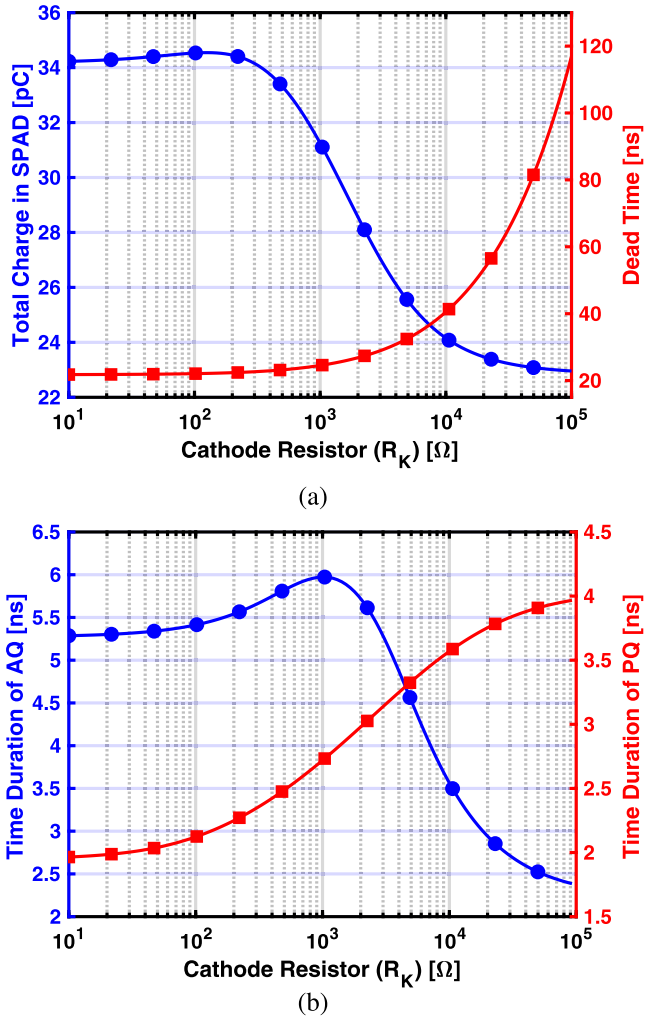


Fig. 5. Relevant parameters versus a sweep for R_k values. (a) Total charge integrated during avalanche is severely reduced if the R_k value increases, but a long recovery time after reset is needed. (b) Whole duration for PQ is worsened by the increasing of the R_k resistor, while the AQ duration benefits from it.

part of the Reset. This effect is minimized reducing the size of R_k, which is in contrast with the desire of having a large value of R_k for AP reduction.

The second part of the Reset phase is governed by the same expression reported in (29). The sole difference relies on updating the R_a value to

$$R_a \simeq R_r \tag{37}$$

where R_r is the equivalent of resistance of the reset transistor, and adjusting the input signal generators as follows:

$$V_{\mathrm{RS1}}(s) = \frac{V_{\mathrm{bd}} + V_{\mathrm{ov}} - v_k|_{\mathrm{RS}}(t_{\mathrm{RS}})}{s} \cdot e^{-s \cdot t_{\mathrm{PO}}} \tag{38}$$

$$V_{\mathrm{RS2}}(s) = -\frac{v_a|_{\mathrm{PO}}}{s} \cdot e^{-s \cdot t_{\mathrm{PO}}} \tag{39}$$

where $v_k|_{\mathrm{RS}}(t_{\mathrm{RS}})$ is the voltage sampled at the end of the saturation part in Reset phase, $v_a|_{\mathrm{PO}}$ is the pinch-off voltage of the reset transistor, and t_{PO} is the time instant when the transistor enters in the linear region. The charge of the cathode terminal back to its bias point is governed by a slow time constant equal to

$$\tau_l|_{\mathrm{RS}} = C_a R_a + C_k R_k + C_{\mathrm{sp}} \cdot (R_a + R_k). \tag{40}$$

Two approaches widely adopted exist to terminate this phase. The Reset MOS can be deactivated when the anode node goes under a certain threshold or after a controlled time. Our model is validated both in simulation and measurement characterization using an AQC that uses the latter approach. After a D_{RS} duration starting from t_{RS}, the reset MOS is deactivated.

It may happen that a photon strikes into the SPAD during the reset phase, causing a new avalanche which will trigger a new photon-detection cycle. However, in this case, a photo-generated current flows into the SPAD while the R_k resistor is attached to the cathode terminal, resulting in a lower potential at the cathode node. Our model is suitable to represent also this scenario exploiting the same routine described so far. Referring to Fig. 3(a), the sole difference for a during-reset photon detection will be to consider a smaller overvoltage at the beginning of the next passive quenching phase, due to the

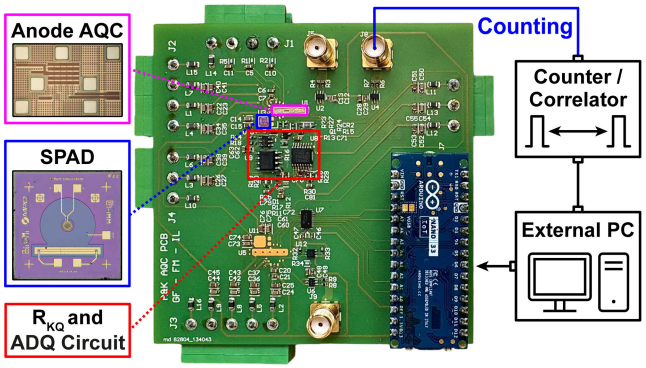


Fig. 6. Setup used for model validation and characterization. The die micrograph of both anode AQC and SPAD are depicted.

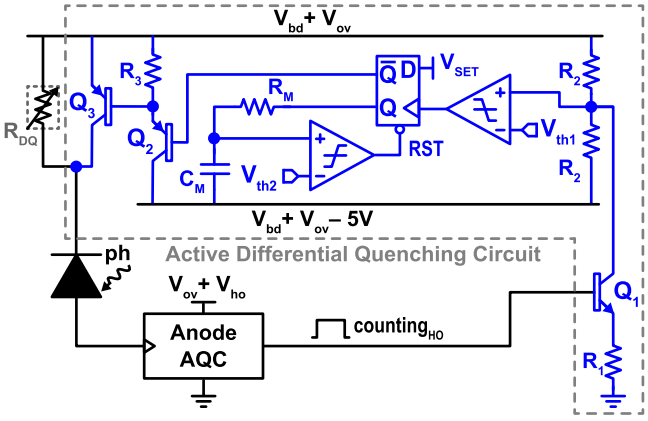


Fig. 7. ADQ circuit schematic and working principle. The whole circuit is designed with a 5-V ECL, at the same potential as the cathode terminal. V_{th1} and V_{th2} thresholds are used for upshifting and fixing the pulsewidth of the controlling signal of Q_3.

cathode voltage value starting from

$$V_k = (V_{ov} + V_{bd}) - V_{ov} \cdot \frac{R_k}{R_k + R_{sp}}. \tag{41}$$

III. Model Validation

After elucidating the analysis of an entire photon-detection cycle, a 2T topology is implemented to validate the model calculations. The schematic of a 2T approach follows the structure in Fig. 1. This one is constituted by an anode AQC combined with a quenching resistor (R_{KQ}) at the cathode terminal. An AQC similar to the one presented in [17] is exploited. The AQC was designed in an HV 150-nm technology. The SPAD used for model validation shares the same characteristics to the one reported in [29]. The photodetector has a diameter of 100 μm, a breakdown voltage $V_{bd} = 34.8$ V, and an R_{sp} inner resistance of 1 kΩ. The SPAD was fabricated in a custom technology process. The capacitance contribution deriving from all circuital elements is estimated to be

$$C_k \simeq 1 \text{ pF} \quad C_{sp} \simeq 2 \text{ pF} \quad C_a \simeq 8 \text{ pF}. \tag{42}$$

The smaller parasitisms present at the cathode node are the principal reason for AP reduction. Besides the SPAD differential capacitance C_{sp}, all other parasitic contributions derives from several factors as reported in (5) and (6). The cathode-substrate and anode-substrate parasitic values were obtained from device characterization, the AQC parasitism were estimated from postlayout simulation with Cadence Spectre,[1] and the onboard parasitism with Saturn PCB Toolkit.[1] The greatest parasitic contribution at the anode node is due to the anode-substrate capacitance.

A. Schematic Simulation

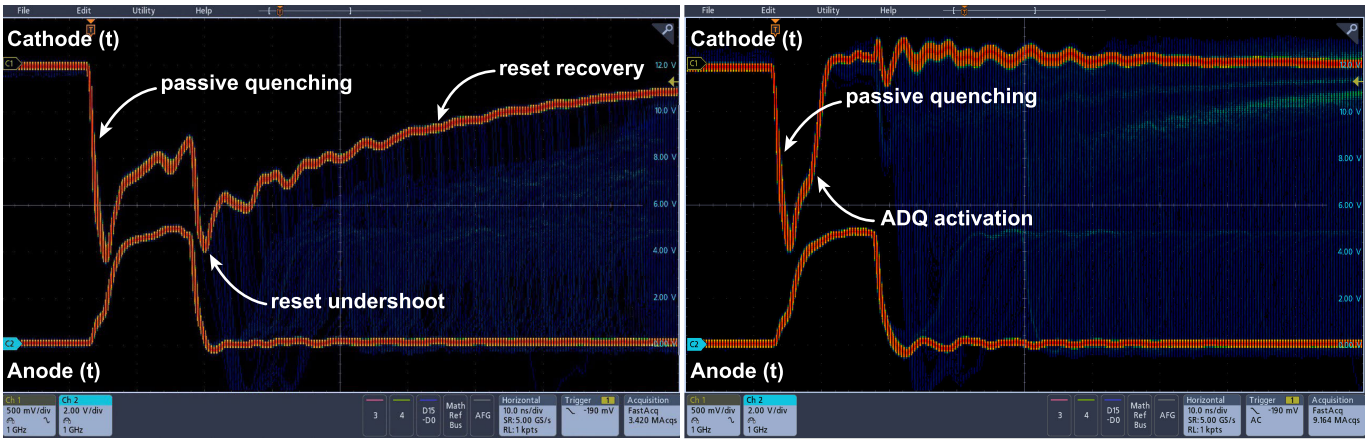
Fig. 4 shows the comparison between the simulated and modeled time-domain waveforms for anode and cathode terminals, and the inner SPAD current waveform during avalanche. The simulated waveforms were obtained via Cadence Spectre[1] simulation on schematic. The simulation were done considering a $V_{ov} = 10$ V and a $V_{ho} = 1$ V. For sake of simplicity, V_{bd} used during simulation has been rounded to $V_{bd} = 30$ V, being completely irrelevant for model validation. A logarithmic sweep has been done on the value of the R_{KQ} resistor.

Regardless the value of the R_{KQ} resistor used, the proposed model perfectly follows the transistor-level simulation. The model is thus suitable for comparing the 1T and 2T architectures to evaluate the differences. In 2T architectures, on one hand cathode node initially contributes to the PQ of the SPAD, strongly limiting the current flowing through it. On the other hand, the anode node starts the AQ phase delayed according to the presence of a larger value of R_{KQ} resistor. A hold-off of 6 ns was selected in this simulation per every value of R_{KQ}. During the Reset phase, the capacitive coupling causes a significant undershoot of the cathode terminal, with an amplitude that increases as the value of R_{KQ} increments as predicted in (35).

In Fig. 5(a), the effect of 2T approach can be evaluated compared with 1T architectures. With a 1T architecture (i.e., $R_{KQ} = 0$ Ω), the overall charge flowing during the whole quenching phase achieves its maximum value. This value diminishes until reaching a plateau around a $R_{KQ} \simeq 5 \cdot R_{sp}$ value. In these conditions, the DT increases according to the value of R_{KQ}. By defining the DT of the structure as the time need to restore the 99% of the starting bias voltage across the SPAD terminals, this one suffers as the R_{KQ} value is increased.

The equations presented in (13) and (26) can now be explained, focusing on Fig. 5(b). The AQ phase triggered at the anode node starts delayed if a large value of R_{KQ} is considered. The maximum delay along 1T and 2T architectures that can be reached with fixed parasitisms is shown in the figure. However, the overall duration of active quenching benefits from a large value of R_{KQ}, due to a lower amplitude of the photocurrent in the SPAD at the beginning of this phase. In the end, these two effects counterbalance, but with the fundamental difference of the possibility to further reduce the hold-off time, due to a severe AP reduction.

In Fig. 5(b), the active quenching duration behavior is only due to the set of parameters considered and the parsistims at the anode and cathode nodes. In principle, a higher value of the SPAD current amplitude at the beginning of active quenching is present in the case of 1T topology with respect to the amplitude in the 2T one. The active quenching duration reaches its minimum with a large value of the R_k resistor. The peak in the time duration observed in the graph occurs for a mid-range value for the R_k resistor. In this case, we lose the advantage of a differential passive quenching for lowering



(a) (b)

Fig. 8. Oscilloscope measures of anode and cathode waveforms in 2T and 2T-with-ADQ circuit architectures. The R_{KQ} resistance is equal to 4.7 kΩ in every case. $V_{ov} = 3$ V and $V_{ho} = 1$ V. (a) 2T architecture with anode AQC. (b) 2T-with-ADQ architecture anode AQC.

the SPAD current amplitude at the beginning of the active quenching phase, combined with a slight increment of the time constant considered for active quenching the SPAD. A visible justification of this effect is appreciable in the photocurrent waveform reported in Fig. 4.

B. Measurements and Characterization

The model is validated also through a full characterization on 2T and 1T architectures with the measurement setup reported in Fig. 6. The board hosts the aforementioned anode AQC and SPAD bond-wired to each other on a thin pad onboard. The cathode terminal can be shorted to power supply or attached to an R_{KQ} resistor according to the architecture taken under exam. The selected value for the cathode differential quenching resistor is $R_{KQ} = 4.7$ kΩ ($\simeq 5 \cdot R_{sp}$), in order to minimize the current flowing during passive quenching as reported from the total charge behavior in Fig. 5(a).

The two architectures are compared under the same low illumination conditions, with an average dark count rate of 6.13 kHz. As reported in [24], the presence of a resistance at the cathode terminal does not affect in any way the dark count rate. The AQC output is a counting signal that is triggered every time a photon-detection occurs. The pulsewidth of this signal can be selected to be as long as the D_{HO} or the D_{RS} time interval. This signal is then processed by a counter/correlator unit which extracts the AP correlation factor on a long-duration experiment. An Active Differential Quench (ADQ) circuit has also been designed, and its inner structure is depicted in Fig. 7. Its working principle will be discussed in the next section.

In Fig. 8(a), it is possible to appreciate the waveforms of ac-coupled cathode terminal, and anode terminal when a 2T architecture is considered. V_{ov} considered is equal to 3 V, while the sub-breakdown voltage V_{ho} is equal to 1 V. These values were chosen to limit the quantum efficiency of the detector during the characterization, thus obtaining a clearer view of the waveform measurements. Obviously, any result can be easily scaled up to higher V_{ov} values.

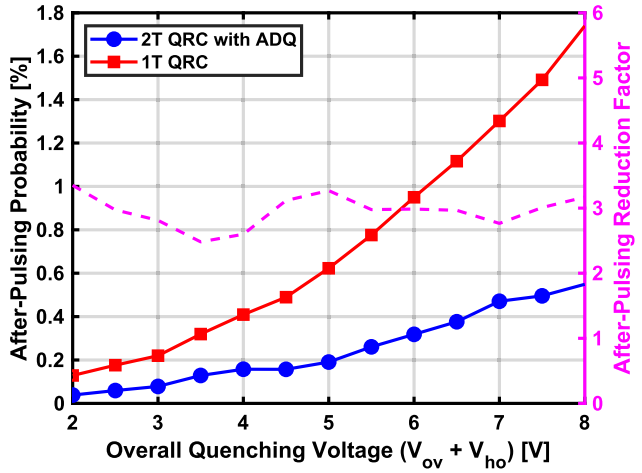
It is evident how the predicted behavior of the simulation and model is visible here. Naturally, the huge capacitive weight (4.95 pF) of the probe must be taken into account while observing these waveforms. Thus, at both terminals, steeper transitions can be expected, due to a capacitive lightening of both terminals when the probe is removed.

IV. Active Differential Quench

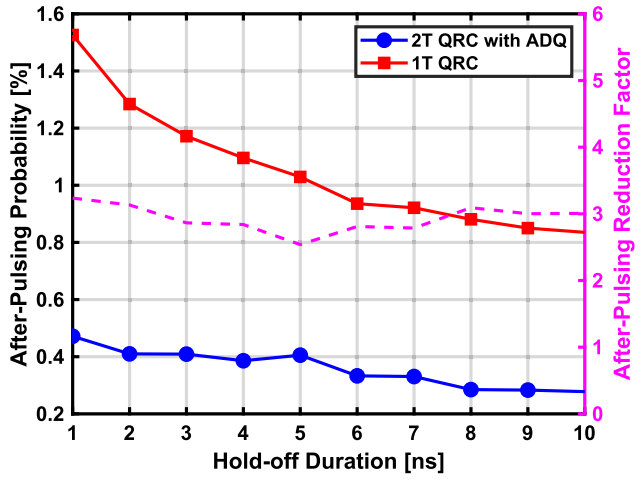
The long recovery time during reset represents the greatest drawback in adopting a 2T architecture. The goal is having a large value of R_{KQ} during the quench, but a small value for the whole duration of the Reset phase. A possible solution to solve the trade-off is closing a switch in parallel to R_{KQ} at the time instant when the hold-off phase begins, and reopen it when the reset ends.

In this work, we present an architecture to implement this solution. This is called the ADQ circuit, whose schematic is depicted in Fig. 7. An emitter-coupled logic (ECL) shares the positive power supply terminal with the cathode bias, and it works on a 5-V voltage domain. A counting signal, synchronous with the t_{AQ} time instant and coming from the anode AQC, is up-shifted to the cathode voltage domain thanks to an HV n-p-n bipolar transistor (Q_1). The R_1 resistor sets 1-mA current value used for up-shifting the counting signal. This pulse is then regenerated by means of a high-speed comparator with V_{th1} as close as possible to the bias point to guarantee fast pulse recognition. The output of the comparator is monostabilized with an ECL D-flip-flop. The pulsewidth of the Q signal coming from the flip-flop is fed to an RC network which is close-to-linearly charging the C_M capacitor. The voltage threshold V_{th2} permits to select the pulsewidth of this signal to make it last until the end of the Reset phase. The ECL signal $\overline{Q}$ is connected to two following Q_{2-3} RF p-n-p bipolar transistors that short the R_{KQ} resistor imposing an overall R_k value of $\simeq 5$ Ω when closed. The parasitic capacitance introduced by the ADQ circuit plays a key role in minimizing the AP, for this reason, low-capacitance RF p-n-p transistors are used, directly attached to the photodetector.

The effectiveness of the circuit can be observed from Fig. 8(b). The circuit is tested in the same exact conditions reported in the previous section and observable in Fig. 8(a). It is evident that how the benefits of quenching



(a)



(b)

Fig. 9. AP probability evaluation. (a) AP probability versus quenching voltage in 1T and 2T-with-ADQ architectures. (b) AP probability versus hold-off duration in 1T and 2T-with-ADQ architectures.

are preserved, while the long recovery time after reset does not represent an issue anymore. As predicted, the onboard circuit solution broke the trade-off between AP reduction and long DT after reset. The monostabile pulsewidth has to be adjusted to guarantee that the p-n-p transistor reopens at the same time instant when the Reset phase is over.

In Fig. 9(a), a sweep of the AP probability versus the quenching voltage ($V_{ov} + V_{ho}$) is reported. The sweep has been made keeping fixed to 1 V V_{ho}, while increasing the overvoltage. As well known in the literature, the AP probability increases as the overvoltage increases. However, the 2T architecture with ADQ strongly limits the AP, reaching a maximum of 0.55% AP probability with a quenching voltage as high as 8 V. Compared with a standard 1T architecture, the AP reduction factor is constant and equal to 3 on the whole sweep.

In Fig. 9(b), a similar sweep of the AP is represented, this time against the hold-off phase duration. The sweep was made from 1 to 10 ns of hold-off duration, with an overvoltage equal to 5 V and a V_{ho} equal to 1 V. The AP probability reduces as the hold-off time increases. The AP reduction factor is constant and equal to 3 also in this case. An AP probability as low as 0.28% can be reached in a 2T AQC with ADQ topology, when a 0.92% AP probability was presented in the 1T architecture under the same testing conditions.

Finally, a fair comparison along all architectures shall take into account also the power consumption per every topology. For the sake of simplicity, this one is treated separately in the Appendix, exploiting the presented model reliability for calculation.

V. Conclusion

In this article, a comprehensive and accurate analytical model for 2T quenching architectures for SPADs is proposed, and its effectiveness is demonstrated for circuit design purposes. This model relies on a novel switching-topology Laplace domain analysis approach, to easily derive time-domain signal expressions of SPAD anode, cathode and avalanche current. The model has shown great accuracy with respect to both simulation and experimental measurements, demonstrating its effectiveness to compare 2T and 1T architectures. A clear view on AP reduction is then justified, widely expanding the promising results observed in [24]. The benefits and drawbacks of 2T architectures have been deeply addressed to offer a powerful tool for designing AQCs that aim to exploit this approach. Finally, a working solution to solve the AP versus DT trade-off in 2T topologies was presented, demonstrating a reduction factor up to three times of AP probability with resepct to common 1T architectures. Considering the adoption of HV BCD technologies, the work described here can be used as fundamental base to implement full-integrated SPAD front-end structures, exploiting 2T architectures with co-integrated active differential QCs.

Appendix
Power Consumption Confrontation

Power dissipation can be derived from the model, and it can be considered in the design choice between 1T and 2T architectures. Referring to the model represented in Fig. 2, the following analysis is derived. The power dissipation of both architectures per photon cycle can be computed as the power drained from the $V_{bd} + V_{ov}$ and $V_{ov} + V_{ho}$ supplies. The average power consumption per photon cycle for the first one is

$$\overline{P_{pol}} = \frac{V_{ov} + V_{bd}}{\Delta T_{ph}} \cdot \int_{t_{ph}}^{t_{END}} \frac{v'_k(t)}{R_k}\, dt \tag{43}$$

where $v'_k(t)$ is equal to $(V_{ov} + V_{bd}) - v_k(t)$, i.e., the voltage drop across the R_k resistor, and ΔT_{ph} is the duration of a whole photon cycle. The phases when the power consumption is not nil are the quenching and the reset phase, meaning that the following can be written:

$$\overline{P_{pol}} = \frac{V_{ov} + V_{bd}}{\Delta T_{ph}} \cdot \int_{t_{ph}}^{t_{AQ}} \frac{v'_k(t)|_{PQ+AQ}}{R_k}\, dt \tag{44}$$

$$+ \frac{V_{ov} + V_{bd}}{\Delta T_{ph}} \cdot \int_{t_{HO}}^{t_{RS}} \frac{v'_k(t)|_{RS}}{R_k}\, dt \tag{45}$$

obtaining the expressions for effective power consumption only during the phases of interest. Obviously, for the

1T topology, the previous expressions do not lead to an infinite value for the $R_k \to 0$ term at the denominator. In this case, the following expressions stand:

$$\overline{P_{\text{pol}}} \simeq \frac{V_{\text{ov}} + V_{\text{bd}}}{\Delta T_{\text{ph}}} \cdot \int_{t_{\text{ph}}}^{t_{PQ}} \frac{V_{\text{ov}}}{R_{\text{sp}}} \cdot e^{-\frac{t-t_{\text{ph}}}{\tau_h|_{PQ}}} \tag{46}$$

$$+ \frac{V_{\text{ov}} + V_{\text{bd}}}{\Delta T_{\text{ph}}} \cdot \int_{t_{PQ}}^{t_{AQ}} i_{\text{sp}}(t)\big|_{AQ}\, dt \tag{47}$$

$$+ \frac{V_{\text{ov}} + V_{\text{bd}}}{\Delta T_{\text{ph}}} \cdot \int_{t_{\text{HO}}}^{t_{\text{RS}}} I_{n_{\text{SAT}}} \cdot \frac{C_{\text{sp}}}{C_a + C_{\text{sp}}} \cdot e^{-\frac{t-t_{\text{HO}}}{\tau|_{\text{RS}}}}\, dt \tag{48}$$

where the considerations made in (14) and (35) were exploited to derive these equations. The term $i_{\text{sp}}(t)|_{AQ}$ represents the time-domain expression of the SPAD current, derived from the time-domain equations of anode and cathode nodes in this phase. To obtain the actual dynamic power consumption from the average power consumption per photon cycle, the following must be written:

$$P_{\text{pol}} = \overline{P_{\text{pol}}} \cdot \Delta T_{\text{ph}} \cdot f_{\text{ph}} \tag{49}$$

where f_{ph} is the average number of detected photons per second.

The power dissipation drained from the $V_{\text{ov}} + V_{\text{ho}}$ supply can be derived with a simpler analysis. Current coming from this supply charges a capacitive load only during the active quenching phase and once per every photon detection cycle. The power consumption can be expressed as follows:

$$P_{aqc} \simeq (V_{\text{ov}} + V_{\text{ho}})^2 \cdot f_{\text{ph}} \cdot \left[C_a + C_{\text{sp}} \parallel \left(C_k \cdot \frac{R_k}{R_k + R_{\text{sp}}} \right) \right] \tag{50}$$

where the initial rising of the anode node at the beginning of the active quenching phase is neglected.

These expressions allow a direct comparison between the power consumption for the 1T and 2T cases, as a function of the design parameters and parasitic elements. Considering the set of parasitic elements in (42), the power consumption for a 2T architecture is 13% smaller with respect to the 1T topology. Finally, when a 2T-with-ADQ is exploited the power consumption of the ADQ circuit must be added to the calculations.

Acknowledgment

The authors would like to thank Christian Riboldi for the precious help in designing and assembling the PCB and Angela Bovolenta for her contribution in reviewing the calculations.

References

[1] X. Liu et al., "Fast fluorescence lifetime imaging techniques: A review on challenge and development," *J. Innov. Opt. Health Sci.*, vol. 12, no. 5, Sep. 2019, Art. no. 1930003.

[2] D. Stoppa, D. Mosconi, L. Pancheri, and L. Gonzo, "Single-photon avalanche diode CMOS sensor for time-resolved fluorescence measurements," *IEEE Sensors J.*, vol. 9, no. 9, pp. 1084–1090, Sep. 2009.

[3] V. Zickus et al., "Fluorescence lifetime imaging with a megapixel SPAD camera and neural network lifetime estimation," *Sci. Rep.*, vol. 10, no. 1, p. 20986, Dec. 2020.

[4] A. T. Erdogan et al., "A high dynamic range 128×120 3-D stacked CMOS SPAD image sensor SoC for fluorescence microendoscopy," *IEEE J. Solid-State Circuits*, vol. 57, no. 6, pp. 1649–1660, Jun. 2022.

[5] J. Kekkonen, T. Talala, J. Nissinen, and I. Nissinen, "On the spectral quality of time-resolved CMOS SPAD-based Raman spectroscopy with high fluorescence backgrounds," *IEEE Sensors J.*, vol. 20, no. 9, pp. 4635–4645, May 2020.

[6] M. T. Rakher, L. Ma, O. Slattery, X. Tang, and K. Srinivasan, "Quantum transduction of telecommunications-band single photons from a quantum dot by frequency upconversion," *Nature Photon.*, vol. 4, no. 11, pp. 786–791, Nov. 2010.

[7] G.-J. Fan-Yuan et al., "Optimizing single-photon avalanche photodiodes for dynamic quantum key distribution networks," *Phys. Rev. Appl.*, vol. 13, no. 5, May 2020, Art. no. 054027.

[8] M.-J. Lee and E. Charbon, "Progress in single-photon avalanche diode image sensors in standard CMOS: From two-dimensional monolithic to three-dimensional-stacked technology," *Jpn. J. Appl. Phys.*, vol. 57, no. 10, Oct. 2018, Art. no. 1002A3.

[9] B. Patra et al., "Cryo-CMOS circuits and systems for quantum computing applications," *IEEE J. Solid-State Circuits*, vol. 53, no. 1, pp. 309–321, Jan. 2018.

[10] R. Sax et al., "High-speed integrated QKD system," *Photon. Res.*, vol. 11, no. 6, pp. 1007–1014, 2023.

[11] G. Chen, C. Wiede, and R. Kokozinski, "Data processing approaches on SPAD-based d-TOF LiDAR systems: A review," *IEEE Sensors J.*, vol. 21, no. 5, pp. 5656–5667, Mar. 2021.

[12] F. Villa, F. Severini, F. Madonini, and F. Zappa, "SPADs and SiPMs arrays for long-range high-speed light detection and ranging (LiDAR)," *Sensors*, vol. 21, no. 11, p. 3839, Jun. 2021.

[13] O. Kumagai et al., "7.3 A 189×600 back-illuminated stacked SPAD direct time-of-flight depth sensor for automotive LiDAR systems," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, vol. 64, Feb. 2021, pp. 110–112.

[14] F. Mattioli Della Rocca et al., "A 128×128 SPAD motion-triggered time-of-flight image sensor with in-pixel histogram and column-parallel vision processor," *IEEE J. Solid-State Circuits*, vol. 55, no. 7, pp. 1762–1775, Jul. 2020.

[15] R. K. Henderson et al., "5.7 A 256×256 40nm/90nm CMOS 3D-stacked 120dB dynamic-range reconfigurable time-resolved SPAD imager," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2019, pp. 106–108.

[16] 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. 1956–1976, 1996.

[17] A. Giudici, G. Acconcia, I. Labanca, M. Ghioni, and I. Rech, "4 ns dead time with a fully integrated active quenching circuit driving a custom single photon avalanche diode," *Review of Scientific Instruments*, vol. 93, no. 4, 2022, Art. no. 043103.

[18] F. Gramuglia, M.-L. Wu, C. Bruschini, M.-J. Lee, and E. Charbon, "A low-noise CMOS SPAD pixel with 12.1 ps SPTR and 3 ns dead time," *IEEE J. Sel. Topics Quantum Electron.*, vol. 28, no. 2, pp. 1–9, Mar. 2022.

[19] F. Severini, I. Cusini, D. Berretta, K. Pasquinelli, A. Incoronato, and F. Villa, "SPAD pixel with sub-NS dead-time for high-count rate applications," *IEEE J. Sel. Topics Quantum Electron.*, vol. 28, no. 2, pp. 1–8, Mar. 2022.

[20] M. D. Lakeh, J.-B. Kammerer, W. Uhring, J.-B. Schell, and F. Calmon, "An ultrafast active quenching circuit for SPAD in CMOS 28nm FDSOI technology," in *Proc. IEEE SENSORS*, Oct. 2020, pp. 1–4.

[21] 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 Sensors J.*, vol. 16, no. 1, pp. 3–12, Jan. 2016.

[22] 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.*, vol. 33, no. 11, pp. 1589–1591, Nov. 2012.

[23] A. Gulinatti et al., "New silicon SPAD technology for enhanced red-sensitivity, high-resolution timing and system integration," *J. Mod. Opt.*, vol. 59, no. 17, pp. 1489–1499, 2012.

[24] W. Jiang, R. Scott, and M. J. Deen, "Differential quench and reset circuit for single-photon avalanche diodes," *J. Lightw. Technol.*, vol. 39, no. 22, pp. 7334–7342, Nov. 15, 2021.

[25] A. Dalla Mora, A. Tosi, S. Tisa, and F. Zappa, "Single-photon avalanche diode model for circuit simulations," *IEEE Photon. Technol. Lett.*, vol. 19, no. 23, pp. 1922–1924, Dec. 2007.

[26] V. Savuskan, M. Javitt, G. Visokolov, I. Brouk, and Y. Nemirovsky, "Selecting single photon avalanche diode (SPAD) passive-quenching resistance: An approach," *IEEE Sensors J.*, vol. 13, no. 6, pp. 2322–2328, Jun. 2013.

[27] S. Haley and P. Hurst, "Pole and zero estimation in linear circuits," *IEEE Trans. Circuits Syst.*, vol. 36, no. 6, pp. 838–845, Jun. 1989.

[28] E. Gluskin, "Let us teach this generalization of the final-value theorem," *Eur. J. Phys.*, vol. 24, no. 6, p. 591, 2003.

[29] M. Ghioni, A. Gulinatti, I. Rech, P. Maccagnani, and S. Cova, "Large-area low-jitter silicon single photon avalanche diodes," in *Quantum Sensing and Nanophotonic Devices V*, vol. 6900. Bellingham, WA, USA: SPIE, 2008, pp. 267–279.

Francesco Malanga (Graduate Student Member, IEEE) was born in Avellino, Italy, in 1996. He received the B.S. and M.S. (cum laude) degrees in electronics engineering from the Politecnico di Milano, Milan, Italy, in 2018 and 2020, respectively, where he is currently pursuing the Ph.D. degree in information technology.

His current research interests include front-end circuits design for single-photon avalanche diodes and ultralow-noise timing circuits for time-correlated single-photon counting (TCSPC) and quantum applications.

Gennaro Fratta (Graduate Student Member, IEEE) was born in Naples, Italy, in 1999. He received the B.S. and M.S. (cum laude) degrees in electronics engineering from the Politecnico di Milano, Milan, Italy, in 2021 and 2023, respectively, where he is currently pursuing the Ph.D. degree in information technology with the Department of Electronics, Information and Bioengineering (DEIB).

His current research interests include high-performance electronic systems based on high-speed time-correlated single-photon counting (TCSPC) and single-photon detectors for applications in quantum imaging, astronomy, LiDAR, molecular biology, and diagnostics.

Gabriele Laita (Graduate Student Member, IEEE) received the B.Sc. (cum laude) and M.Sc. (cum laude) degrees in electronics engineering from the Politecnico di Milano, Milan, Italy, in 2019 and 2022, respectively, where he is currently pursuing the Ph.D. degree in information technology with the Dipartimento di Elettronica, Informazione e Bioingegneria.

His research interests are directed toward high-detection efficiency single-photon avalanche diodes (SPADs).

Angelo Gulinatti (Member, IEEE) received the M.Sc. (cum laude) degree in electronics engineering and the Ph.D. (cum laude) degree in information technology from the Politecnico di Milano, Milan, Italy, in 2003 and 2007, respectively.

From 2013 to 2017, he has been a Visiting Scientist with the Cornell NanoScale Science and Technology Facility, Cornell University, Ithaca, NY, USA. Since 2008, he has been with the Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, first, as an Assistant Professor and now as an Associate Professor. He has co-authored more than 100 papers published in peer-reviewed journals and conference proceedings. His research interests are focused on the development of custom technology single-photon avalanche diodes (SPADs) and related electronics for applications ranging from biochemistry-to-quantum information processing.

Giulia Acconcia (Member, IEEE) received the B.S. degree in engineering of computing systems, the M.S. degree in electronics engineering, and the Ph.D. (Hons.) degree in information technology from the Politecnico di Milano, Milan, Italy, in 2011, 2013, and 2017, respectively.

Since 2017, she has been with the Department of Electronics, Information and Bioengineering, Politecnico di Milano, where she is currently a Senior Researcher. She has co-authored more than 60 papers in international journals and conferences. Her current research interests include the development of integrated electronics for high-performance counting and timing with single-photon avalanche diodes (SPADs).

Dr. Acconcia was awarded the SPIE Rising Researcher in 2019 and the Young Investigator Award in 2020.

Ivan Rech (Senior Member, IEEE) received the M.S. degree in electronic engineering and the Ph.D. (Hons.) degree in information technology from the Politecnico di Milano, Milan, Italy, in 2000 and 2004, respectively.

He is currently an Associate Professor with the Politecnico di Milano. He was involved in several interdisciplinary research projects on molecular biology, quantum cryptography, adaptive optics in astronomy, and developing dedicated instrumentation. His current research interests include the development of single photon detectors and associated electronics for biomedical, genetic, and diagnostic applications.

Open Access funding provided by 'Politecnico di Milano' within the CRUI CARE Agreement