INTERNATIONAL JOURNAL OF CIRCUIT THEORY AND APPLICATIONS
Int. J. Circ. Theor. Appl. 2016; **44**:917–928
Published online 7 July 2015 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/cta.2113

Compact CMOS active quenching/recharge circuit for SPAD arrays

Ion Vornicu*,†, Ricardo Carmona-Galán, Belén Pérez-Verdú and Ángel Rodríguez-Vázquez

Instituto de Microelectrónica de Sevilla-CNM (CSIC-Universidad de Sevilla), Avenida Americo Vespucio s/n, Parque Científico y Tecnológico de la Cartuja, 41092 Sevilla, Spain

SUMMARY

Avalanche diodes operating in Geiger mode are able to detect single photon events. They can be employed to photon counting and time-of-flight estimation. In order to ensure proper operation of these devices, the avalanche current must be rapidly quenched, and, later on, the initial equilibrium must be restored. In this paper, we present an active quenching/recharge circuit specially designed to be integrated in the form of an array of single-photon avalanche diode (SPAD) detectors. Active quenching and recharge provide benefits like an accurately controllable pulse width and afterpulsing reduction. In addition, this circuit yields one of the lowest reported area occupations and power consumptions. The quenching mechanism employed is based on a positive feedback loop that accelerates quenching right after sensing the avalanche current. We have employed a current starved inverter for the regulation of the hold-off time, which is more compact than other reported controllable delay implementations. This circuit has been fabricated in a standard 0.18 μm complementary metal-oxide-semiconductor (CMOS) technology. The SPAD has a quasi-circular shape of 12 μm diameter active area. The fill factor is about 11%. The measured time resolution of the detector is 187 ps. The photon-detection efficiency (PDE) at 540 nm wavelength is about 5% at an excess voltage of 900 mV. The break-down voltage is 10.3 V. A dark count rate of 19 kHz is measured at room temperature. Worst case post-layout simulations show a 117 ps quenching and 280 ps restoring times. The dead time can be accurately tuned from 5 to 500 ns. The pulse-width jitter is below 1.8 ns when dead time is set to 40 ns. Copyright © 2015 John Wiley & Sons, Ltd.

Received 28 October 2014; Revised 18 May 2015; Accepted 1 June 2015

KEY WORDS: active quenching/recharge (AQR) circuit; tunable dead time; single-photon avalanche diode (SPAD); Geiger mode; afterpulsing reduction

1. INTRODUCTION

Single-photon avalanche diodes (SPAD) are capable of detecting very weak light signals [1]. Owing to their ability of responding to single photon events, they are employed in applications like fluorescence-lifetime imaging microscopy [2] and positron emission tomography (PET) [3]. In the first case, a high temporal resolution is crucial in the visualization of the exponential decay of fluorescence in the sample. Time-correlated single-photon counting is employed to realize a histogram of detection times in order to accurately depict this decay [4]. In the case of PET-time of flight imaging, better time resolution is required to reduce the uncertainty of positioning the positron annihilation along the line of response while the detector spatial resolution counts in the accurate evaluation of the depth of interaction. Both parameters are crucial to enhance the quality of PET images [5]. Usually, SPADs are organized into arrays in order to perform some type of 2D or 3D imaging of very low light scenes.

*Correspondence to: I. Vornicu, Instituto de Microelectrónica de Sevilla-CNM (CSIC-Universidad de Sevilla) Avenida Americo Vespucio s/n, Parque Científico y Tecnológico de la Cartuja, 41092 Sevilla, Spain.
†E-mail: ivornicu@imse-cnm.csic.es

To implement SPAD arrays of practical size, the basic detector cell must be compact. This cell is composed of the avalanche photodiode, the quenching/recharge circuit, and the readout circuitry. Any reduction of the area occupation and power consumption of the basic detector cell results in a larger number of pixels in the same chip, which means larger image size and resolution. Also, reducing the area of the circuits around the SPAD enhances the fill factor, that is, the amount of area of the basic detector cell that is sensitive to the incoming photons.

Different quenching/recharge techniques have been reported in literature [6]. The simplest employs just a ballast resistor [7]—known as a passive quenching circuit. The major drawbacks of this approach are (i) the lack of a well-defined quenching time (as it is statistically defined by the avalanche process), (ii) a very slow recharge, and (iii) the lack of control on afterpulsing. In order to speed-up the recharge phase and to introduce afterpulsing countermeasures, like an adjustable hold-off time, active quenching circuits (AQCs) were introduced [8]. In this approach, the avalanche onset is detected, and additional circuitry acts on the diode bias, actually lowering it below breakdown. Quenching time does not depend now on the statistical nature of the avalanche, but on the delay of the reacting circuit. Also, the duration of the voltage pulses can be adjusted, as well as the hold-off time. One of the disadvantages of this approach is the complexity of the circuit, in terms of number of transistors and feedback loops, and consequently the larger area occupation.

Regarding the avalanche detection mechanism, the circuit reported in [8] contains an integrated large ballast resistor that biases the sensing node to a voltage that situates the SPAD above the breakdown voltage. The triggering of the avalanche starts a passive-quenching action, rapidly followed by an active-quenching action accelerated by a positive feedback loop. After a pulse of an adjustable duration, a recharge transistor brings the SPAD back to the quiescent condition. The same architecture is implemented in [9] in which timing control is implemented with two monostables, based on a current-starved inverter. One of the monostables is employed for adjusting the quenching time, while the other determines the hold-off. A similar circuit [10] employs a high-speed voltage comparator in order to detect the drop at the ballast resistor. A different circuit [11] eliminates the ballast resistor and employs a variable load implemented by a metal-oxide-semiconductor (MOS) transistor. Hold-off time is controlled by the *RC* load of the NMOS inverter employed to detect the avalanche. When a MOS transistor is employed for passive quenching purposes, usually it must be a high-voltage (HV) device because it needs to support the complete swing of the avalanche photodiode (APD). A HV PMOS transistor for quenching Geiger-mode APDs has been reported in [12, 13], built in a 0.18 μm-1.8 V CMOS process. A lightly doped deep triple well module is employed. An additional proposal is a circuit that employs a monostable whose delay is fixed by design [14]. The latest developments report a miniature actively recharged, passively quenched single-photon detector [15]. The design is implemented in a HV-CMOS 0.18 μm technology and employs an innovative dual-threshold quenching and recharge scheme. The dead time is defined by discharging the parasitic capacitance of the sensing node. The outstanding low-noise features of this process, which has extremely low-defect density, are used to obtain a single-photon detector with very low dark-count rate and free of afterpulsing effects. The circuit utilizes 16 transistors and consumes 15 μA of quiescent current.

In our proposal integrated in a standard UMC 0.18 μm CMOS technology, we make use of a positive feedback loop to speed-up quenching without using a ballast resistor. Thus, both afterpulsing probability and power consumption decrease by limiting the magnitude and the duration of the avalanche [11]. Several design strategies are combined to be able to operate a SPAD detector in harsh conditions such as high dark-count rate (DCR). Small area and power consumption make the proposed AQC a good candidate to build large array. The detector ensemble is aimed for ToF and photon counting. The ensemble of SPAD detector and AQC is fully integrated. Instead of using an *RC* delay circuit for the single-shot generator [16], we have implemented the control of the hold-off time by a current-starved inverter. By reducing the delay of the quenching action, which is a principle that is employed even with the finest technologies [17], afterpulsing probability is certainly reduced. Afterpulsing is theoretically smaller than the one that can be observed in passively quenched SPADs, because of the reduction in the tail of the avalanche current and the introduction of a dead time where the SPAD is off [11, 18]. Moreover, actively quenched SPADs can allow

DOI: 10.1002/cta

higher count rates, as high as 185 MHz [19] for high illumination conditions. A high dynamic range as 139 dB is therefore reachable by using active quenching.

2. ACTIVE QUENCHING AND RECHARGE CIRCUIT

The active quenching and recharge circuit has been tailored to requirements posed by the use of CMOS standard technologies. SPAD detectors in these technologies exhibit larger DCR and lower PDE figures than customized technologies [15] and motivate the search for control circuits capable of dealing with these limitations and to mitigate as much as possible afterpulsing effects. Besides this, the proposed circuitry is intended to minimize power consumption and area and to operate in time-of-flight and photon counting scenarios when configured into an array structure. Results in this paper confirm these advantages as compared with previous AQCs.

The AQC presented here (Figure 1(a)) operates in two different phases: quenching and recharge. Prior to the avalanche and the quenching phase, the complete circuit is reset by a transition to 0 of the reset signal, RST. If V_A was not already 0, this negative pulse in RST makes V_{reset} high. As V_{reset} drives the gate of M4, V_A is pulled down to 0. Therefore, prior to any avalanche being triggered, V_A is at its lowest possible value. The excess voltage, V_e, that is, the extent to which the diode is pushed beyond breakdown voltage V_{BD}, is given by

$$V_{SPAD+} - V_A = V_{BD} + V_e \tag{1}$$

thus, operation starts with the diode biased beyond V_{BD} to an extent indicated by V_{SPAD+}. This state is maintained as long as no avalanche is triggered. A low V_A makes V_{sense} high and, because V_{sense} drives the gate of *n*-type transistor M2, node V_A is pulled down to ground, as long as no significant amount of current passes through the SPAD. In parallel with that, V_{sense} also drives the gate of *p*-type transistor M3. It will have later the mission of pulling up node V_A, but right now it is off, letting V_A at ground level. Given that the circuit has been at rest for enough time, V_2 is high, RST is up again, and therefore M4 is off.

When the avalanche starts, V_A rises because of the avalanche current building a voltage across M1, which is connected as a diode. Then two positive feedback mechanisms are provided to ensure the extinction of the avalanche. First, rising V_A makes V_{sense} to go down to 0. As V_{sense} drives the gate of *n*-type transistor M2, its pull is stopped, therefore allowing V_A to grow higher, thus reducing the excess voltage. At the same time, V_{sense} drives the gate of M3, which is *p*-type; therefore, V_{sense} going down to 0 allows M3 to pull up node V_A, another positive feedback action, until

$$V_A = V_{SPAD+} - V_{BD} \tag{2}$$

At this point, $V_e = 0$, and the avalanche is quenched. During this phase, M4 has been kept off. Because of the delay introduced by the current starved inverted M5-M6-M7 and Inv_3, the changes in V_{sense} are not noticed at V_2 until later on. In fact, it is the bias voltage $V_{\text{hold\_off}}$ that determined the magnitude of this delay—apart from the delays introduced by Inv_3 and Nand_1— resulting in a first approximation as follows:

$$T_{DEAD} \approx \frac{C_{M8} \cdot ln2}{\mu_p \cdot \text{Cox} \cdot \frac{W_{M5}}{L_{M5}} \cdot \left(\text{VDD} - V_{\text{hold\_off}} - \left|V_{TH,M5}\right|\right)} \tag{3}$$

where c_{M8}, μ_p, c_{ox}, W_{M5}, L_{M5}, and $V_{TH,M5}$ are the equivalent capacitance of the transistor M8, holes mobility, gate oxide capacitance, width, length, and threshold voltage of the transistor M5. Once the change in V_{sense}, which went to 0 because of the voltage built up at V_A by the avalanche current, is propagated to V_2, V_{reset} rises. This allows M4 to bring node V_A back to zero, V_{sense} up again, M2 turns on, and M3 turns off. After a delay of the magnitude already described, V_2 goes up and V_{reset}

DOI: 10.1002/cta

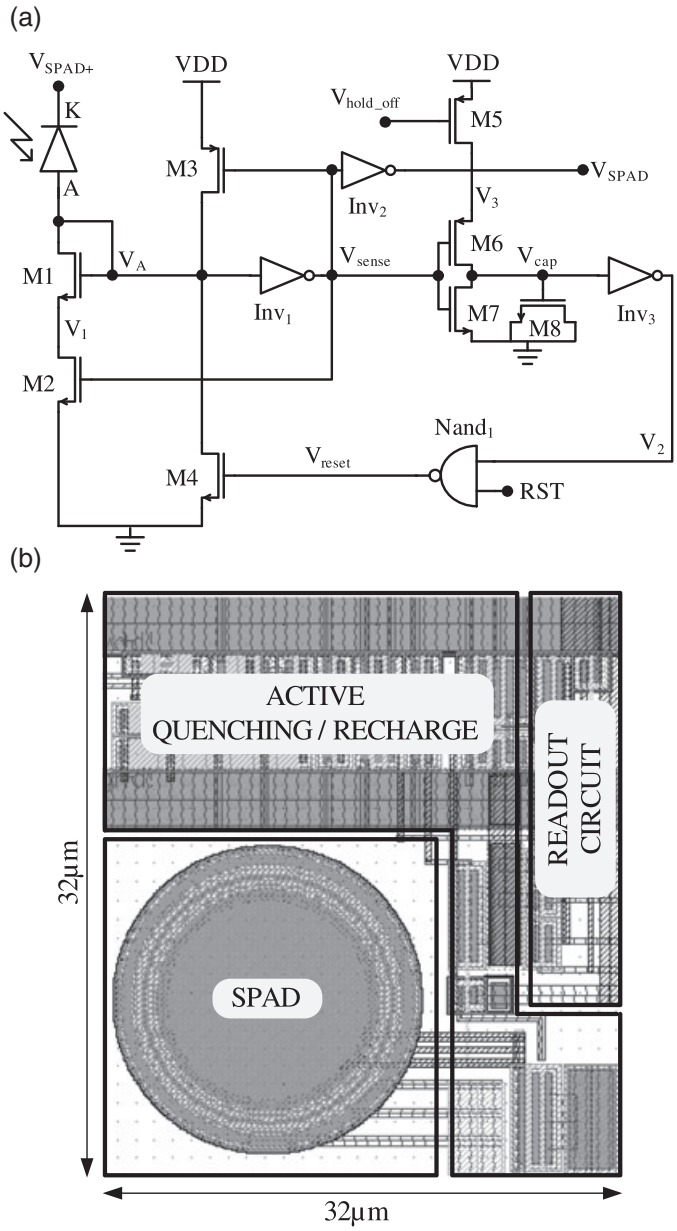

Figure 1. Single-photon avalanche diode (SPAD) and active quenching circuits: (a) schematic and (b) layout.

goes back to zero, leaving the circuit ready for the detection of a new avalanche. This circuit acts as a monostable; a rising edge in V_A is seen as a pulse in V_{SPAD} of a controlled duration.

The benefits of this scheme are several. In the first place, quenching of the avalanche occurs at a much higher pace. This limits the maximum current, thus reducing power consumption and the probability of carriers getting trapped and triggering afterpulses later. Concerning the recharge phase, active recharge is much faster than passive recharge, usually through the same large ballast resistor employed for quenching. In addition to this, duration of the output pulse is relatively stable, and controlled by the user. This allows for a fine control of the dead-time and its influence in reducing pile-up and afterpulsing.

Concerning the sizing of the transistors, M1 and M2 need to be large enough to sense small current peaks. The sizing of the transistors in this design is summarized in Table I. Keeping the avalanche current as small as possible avoids unwanted effects like the increase of afterpulsing, overheating, and crosstalk. Post-layout simulations have been performed using an accurate model of the SPAD

DOI: 10.1002/cta

Table I. Transistor sizes (in μm)—W/L.

M1, M2	M3, M4	M5	M6	M7	M8	Inv1-PMOS; NMOS	Inv2-PMOS; NMOS	Inv3-PMOS; NMOS	Nand1-PMOS; NMOS
5/0.18	2/0.18	0.33/3	3/0.25	1/1.54	4/2	1/0.18; 5/0.18	12.56/0.18; 5/0.18	2.5/0.18; 0.5/0.18	1.25/0.18; 0.5/0.18

that merges a statistical model [20] with the device characteristic [21] (Figure 2). The hold-off time is dynamically switched between 3.3 ns and 30 ns.

Owing to its simplicity, this scheme can be easily incorporated to every pixel in a SPAD sensor array. We have included this AQC in a prototype chip targeted for time-of-flight measurements and focal-plane calculation of light-spot statistics. The sensor has been fabricated in a standard 0.18 μm CMOS technology. It does not support the HV option. Also it is not compliant with low-noise design which requires very low-defect density by high-quality material and low-doped epilayer [15]. The transistors M1, M3, and M4 and inverter Inv_1 are not high voltage. This can be performed thanks to the fast active quenching scheme that limits the swing in node A to 1.8 V. Figure 1(b) depicts the compact layout of a single detector. The pixel pitch is 32 μm, and the fill factor is around 11%. The SPAD has a quasi-circular shape with a 12 μm-diameter active area. The structure of the device can be seen in Figure 3. The TWELL implant is a *p*-type guard ring meant to uniformly distribute the electrical field across the edges of p+ implant, in order to avoid undesired avalanches below the defined breakdown voltage.

This guard-ring structure has the drawback of increasing the jitter because of electron–hole pairs generated by photon absorption in areas with low field. Further optimization of this structure with a technology-CAD tool leads to an improvement of the SPAD jitter. An alternative that can reduce the jitter even more is replacing the *p*-type structure with a SiO_2 shallow-trench-isolation guard-ring [22]. This greatly reduces the diffusion tail of the instrument response function.

3. MEASUREMENTS

3.1. Dark-count rate, afterpulsing and dead-time

The first magnitude to be evaluated, in order to characterize the operation of the SPAD and the proposed AQC, is the influence of dark counts (DC), which are current avalanches that are not

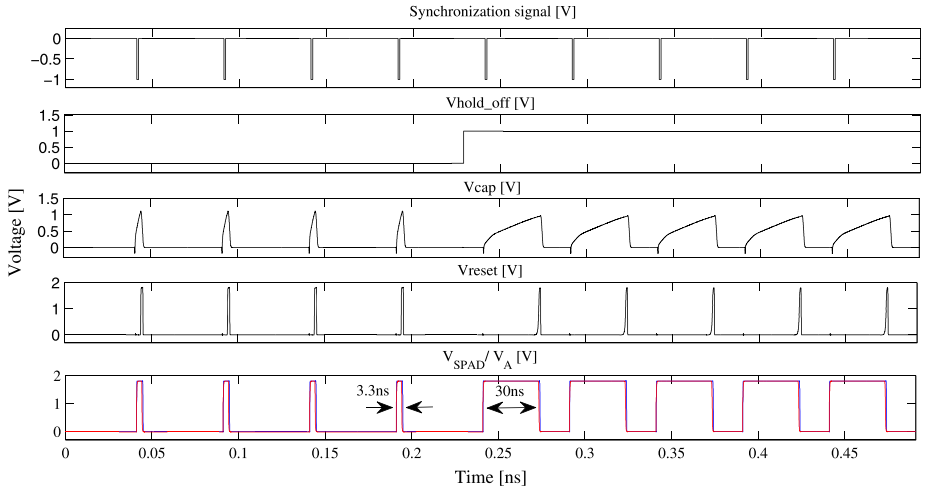

Figure 2. Post layout simulation of the quenching and recharge times—signals chronogram.

DOI: 10.1002/cta

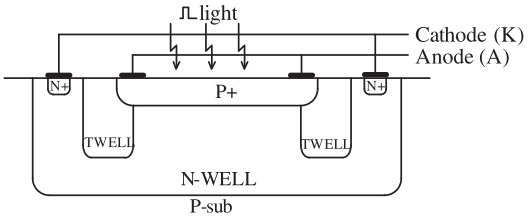

Figure 3. Schematic cross section of the integrated single-photon avalanche diode—notice that depths of the implants are not drawn to scale.

ignited by a triggering photon. These DC are considered noise in the operation of the SPAD. They are usually caused by thermal noise. A different type of phenomenon, also resulting in false photon count, is afterpulsing. It is due to nonzero lifetime of the carriers trapped in the silicon imperfections. SPAD noise becomes more important when dealing with lower illumination rates, what means low signal-to-noise ratio (SNR). The number of dark counts can be optimized for a specific application by tuning the active area of the SPAD (at the design phase), the excess voltage (V_e), and the operating temperature. Notice that minimizing the DCR by playing with the aforementioned parameters, the PDE lowers as well. Therefore, a trade-off needs to be made between PDE and DCR (Section 3.2). Afterpulsing is also strongly dependent on the temperature, excess voltage, and the quality of the material used in the process. For a certain density of defects, afterpulsing can be reduced by limiting the magnitude and duration of the avalanche current. This is why the speed of the quenching action is very important. Besides, an appropriate dead-time allows lowering the probability of spurious triggering after an avalanche has been ignited by light or thermal noise. Figure 4 depicts the output of the SPAD detector exposed to a pulsed light source of 2.5 MHz. The hold-off time of the SPAD is set to 3 ns (middle subset) and 80 ns (bottom subset).

The output voltage of the detector, V_{SPAD}, has been routed to a digital output pad and measured with an oscilloscope or driven into one of the channels of a time tagging instrument (PicoHarp 300 from PicoQuant). In which case, the other channel of the instrument is connected to the synchronization signal of a pulsed laser. Usually the dark-count rate can be measured by counting the pulses over a certain time window. Another way to infer the contribution of this noise is to measure the distribution of time intervals between consecutive noisy pulses (Figure 5), which is called the inter-avalanche time method. This alternative relies on the exponential nature of the occurrence probability of uncorrelated noise [23]. Thus the DCR is inferred from the slope of the best fit curve of the histogram plotted in logarithmic scale. At the same time, this approach is effective to evaluate

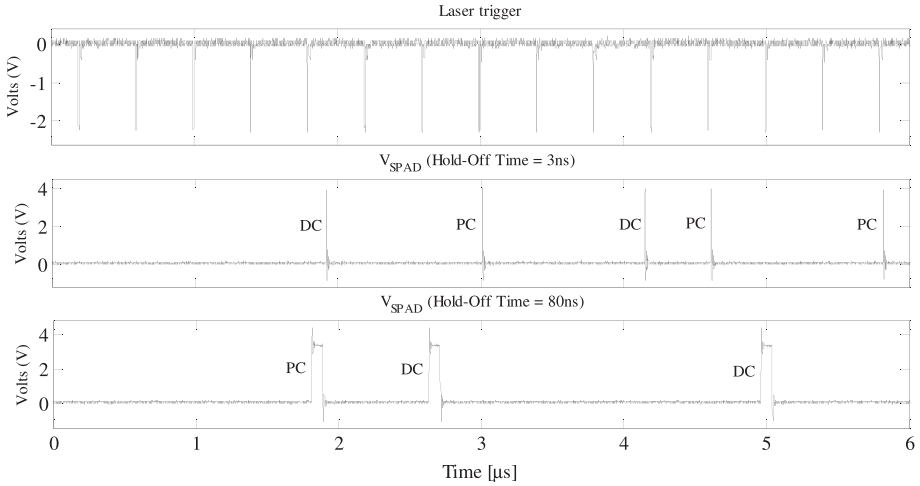

Figure 4. Measured active quenching circuit output when the detector is exposed to a pulsed laser, at two different hold-off time configurations: 3 and 80 ns. PC, photon count; DC, dark count.

DOI: 10.1002/cta

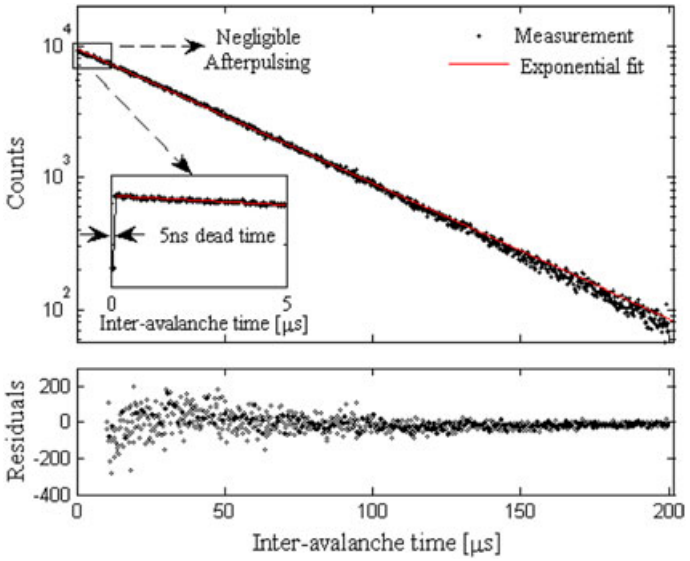

Figure 5. Dark-count rate and afterpulsing measurements.

afterpulsing. It consists in extracting correlation between the time instants in which the leading edges of the avalanche pulses occur. Afterpulsing happens at time scales of hundreds of nanoseconds at room temperature. This means that the histogram of the inter-avalanche arrival time must show multi-exponential behavior up to 1 μs, as explained in [24]. The afterpulsing probability for a specific value of the dead-time can be found by taking the inter-avalanche time histogram, fitting an exponential to the uncorrelated noise source, and then finding the fraction of events above the fit but below the experimental curve. The afterpulsing probability is the area between two curves, divided by the area under the experimentally acquired curve. Figure 5 shows the inter-avalanche time histogram for an excess voltage of 1 V, at room temperature and with a dead-time set to 40 ns. DCR is 26 kHz at 1 V excess voltage.

Using V_e close to V_{BD}, DCR could be drastically decreased [25]. However, in order to have a reliable estimation, DCR needs to be reported together with the PDE achieved at the same excess voltage and temperature. This implementation proved to have a DCR of 19 kcps at room temperature and 0.9 V excess voltage while the maximum value of the PDE in these conditions is 2% at 447 nm illumination source wavelength. The inset zooms in the histogram up to 5 μs. Because no multi-exponential behavior appears, we can claim that the proposed AQC provides negligible afterpulsing. Thanks to the fast quenching mechanism, the avalanche current does not take long enough to fill the traps of the silicon surface [26].

The maximum excess voltage of this detector is 1 V. This excess voltage is kept when the temperature varies from −10°C to 80°C. For higher excess voltages, the anode of the SPAD is stuck to 1.8 V because of the mechanism described in the succeeding discussions and found also reported in [9]. Let us assume the SPAD junction is recharged, that is, the SPAD is turned on. When the first avalanche current rises, the anode is pulled-up to VDD. At this point, the SPAD should be quenched, and the avalanche current brought back to zero. But if the excess voltage goes beyond 1 V, the behavior observed is different. What happens is that the current through the SPAD does not decrease to zero. This explains the increase of the current consumption measured at the power supply source. In these conditions, the recharge transistor M4 is not strong enough to fight against the resistance of the SPAD. As a result, the anode is stuck to VDD and the SPAD cannot be recharged. We have employed transistors supplied at 1.8 V in order to lower the power consumption and to improve the quenching/recharging time.

A magnitude that displays the robustness of the implementation of the dead time is the jitter of the output pulse width. It is lower than 1.8 ns for dead time up to 40 ns (Figure 6(a)). Moreover, the variability of the hold-off time across multiple SPAD detectors needs to be evaluated. The measurements for an array of 8 × 8 SPADs are depicted in Figure 6(b). The hold-off time is tuned

DOI: 10.1002/cta

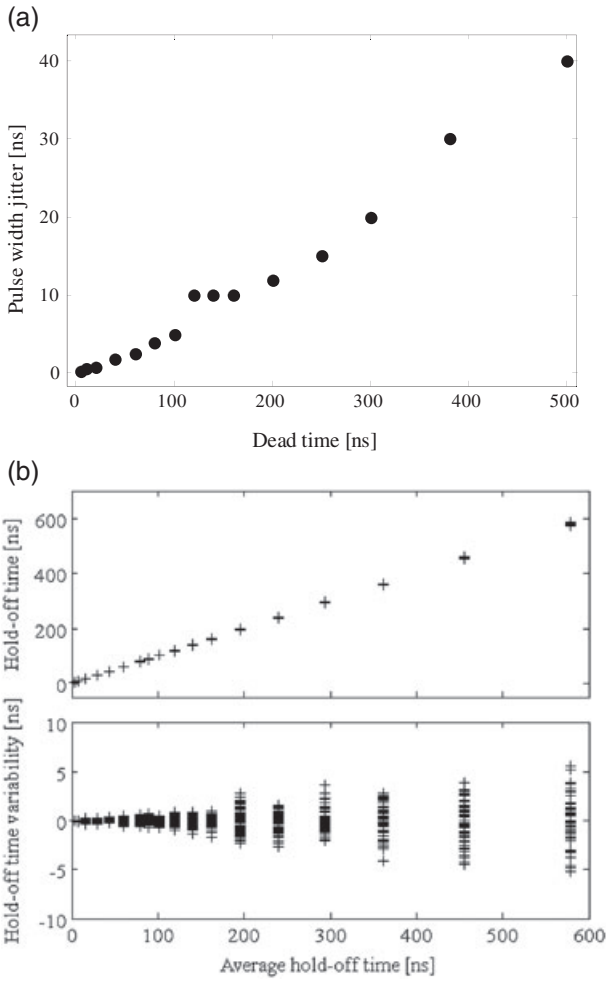

Figure 6. (a) Hold-off time stability represented by the jitter of the output pulse-width and (b) hold-off time variability across 8×8 array of single photon avalanche diodes.

from 3 ns to less than 600 ns. Notice that a hold-off time of 30 ns has a maximum deviation of less than ±300 ps.

3.2. Photon detection efficiency

In order to properly evaluate the operation of the proposed AQC, the PDE must be measured. The PDE is the photon-count rate minus the dark-count rate, divided by the expected arrival rate of incoming photons. If the proper illumination conditions are applied, the measured PDE should be independent on the optical power of the source. In order to realize these measurements, we need to comply with the conditions for single photon detection. To prove this, we have tuned the irradiance to the range of nW/mm^2, as explained in [27]. In a first measurement, in which we try to confirm the independence of the measured PDE on the optical power, we have employed a continuous light source of 447 nm wavelength. The integration time is 500 ms. The excess voltage is 0.9 V and the breakdown voltage being 10.3 V. The DCR for this excess voltage is about 19 kHz. The measurements have been performed at room temperature. The experimental results confirm that the measured PDE do not present any dependence on the illumination. In addition, the PDE have been measured for different wavelengths. These measurements have been performed at room temperature, with an excess voltage of 0.55, 0.7, and 1 V (Figure 7).

In order to make a fair comparison with other detectors, DCR and PDE should be measured in the same conditions, that is, excess voltage and temperature. We have tuned the bias voltage of the SPAD

DOI: 10.1002/cta

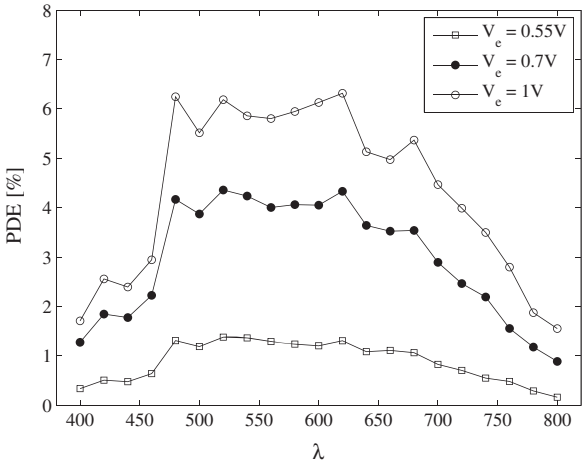

Figure 7. PDE versus wavelength of the incident light.

from 10.8 V to 11.3 V at room temperature and measured both DCR and PDE for 447 nm wavelength (Figure 8). The irradiance employed for PDE is about 3.6 nW/mm^2. The best PDE of 2% for this λ is achieved at 0.95–1 V excess voltage. Therefore, the optimum excess voltage that gives the best PDE at minimum DCR is 0.95 V.

3.3. Time resolution

An important parameter of a SPAD detector is the time resolution, which is the minimum time slot that can be resolved. The purpose of the proposed detector is to be incorporated in arrays of sensors for PET applications [3]. Therefore, a time bin of less than 200 ps is enough for this kind of measurement. For the experimental determination of the time resolution of the SPAD detector, we have employed a PicoHarp 300 from PicoQuant. The instrument response function can be collected at maximum 4 ps resolution. Because of the limitation imposed by the input channels and also to avoid possible inter-pulse pile-up effects, a 2.5 MHz laser repetition rate is used in this setup. The power of the laser spot has been set to 6 nW which means an irradiance of 1.46 nW/mm^2. Considering the laser repetition rate, the active area of the SPAD, and the energy of each photon with 447 nm wavelength, the estimated number of photons per pulse is 0.148 photons/pulse. The total number of counts is then 2.4% of the repetition rate. This will be enough to ensure that the time measurement is not affected by the pile-up. Last but not least, special caution is needed to attenuate the laser reflections. They can distort significantly the instrument response function. The experiment has been conducted

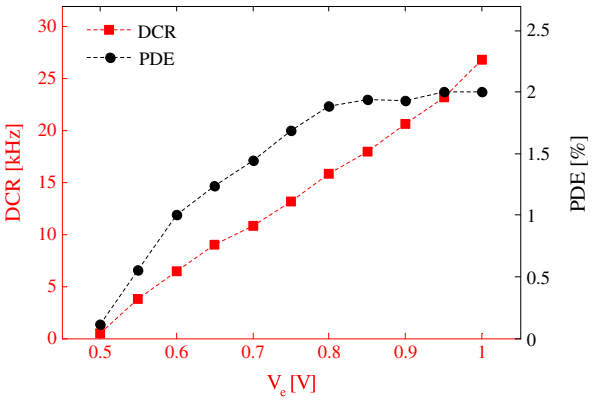

Figure 8. PDE and dark-count rate (DCR) versus V_e at 447 nm wavelength.

DOI: 10.1002/cta

at room temperature. In these conditions, time measurements obtained by the PicoHarp 300, with 8 ps time granularity, are depicted in Figure 9. The time resolution of the detector is obtained by subtracting the quadratic value of the 68 ps laser full width at half maximum (FWHM) and the 20 ps electronic time jitter of the synchronization signal from the 200 ps FWHM of the experimental curve. The so obtained time resolution is 187 ps.

4. DISCUSSION

The characteristics of the AQC presented here are the following: (i) an area of $12 \times 22\,\mu m^2$; (ii) a time delay between the actual arrival time of an impinging photon and the positive edge of the output pulse of 100 ps; (iii) a tunable dead time, T_{DEAD} ranging from 5 to 500 ns; and (iv) an average dynamic power consumption of 2.5 μW drawn from the 1.8 V supply at 1 MHz count rate. This evaluation is very conservative because the maximum allowed count rate without pile-up is 125 kcps, considering that the laser repetition rate is about 2.5 MHz. The detector does not consume static power. Notice that very low average current is drawn from V_{SPAD+} thanks to the quenching circuit that limits effectively the avalanche current. The time resolution measured at 0.9 V excess voltage is 187 ps. The DCR and PDE measured at room temperature, 0.9 V excess voltage and 540 nm wavelength are 19 kHz and 5%, respectively.

Although some of the newest works published in the area deal with topics like high-resolution 3D imagers with in-pixel TDC [28] or single pixel with remarkable performance [15], there are still some works in the quenching/recharge circuit itself. In fact, concerning afterpulsing, active quenching with its inherent limitation of the magnitude of the avalanche current is one of the more effective approaches. The more modern and finer technologies allow using passive quenching because of their intrinsic low-noise nature. However, when dealing with other—and much cheaper—technologies, an effort is necessary in order to reduce these unwanted effects. This work provides a useful active quenching/recharge circuit that upgrades the performance of the devices built in a relatively poor technology. The proposed detector is free of afterpulsing thanks to the very fast quenching/recharge mechanism.

Regarding the occupied area, further improvement can be achieved by limiting the range of the dead time. Thus we can use the parasitic capacitance of the starved inverter instead of the MOS capacitor, which has $60\,\mu m^2$. The size of transistors M1 and M2 in this circuit is critical. If linearity is not important, a larger equivalent resistance is preferred because the current that need to be supplied to the SPAD is then smaller, reducing the width of the metal tracks, the chance of overheating, the risk of electro-migration, and, of course, the power consumption. On the other side, the maximum equivalent resistance must be chosen small enough for the avalanche current to be larger than the

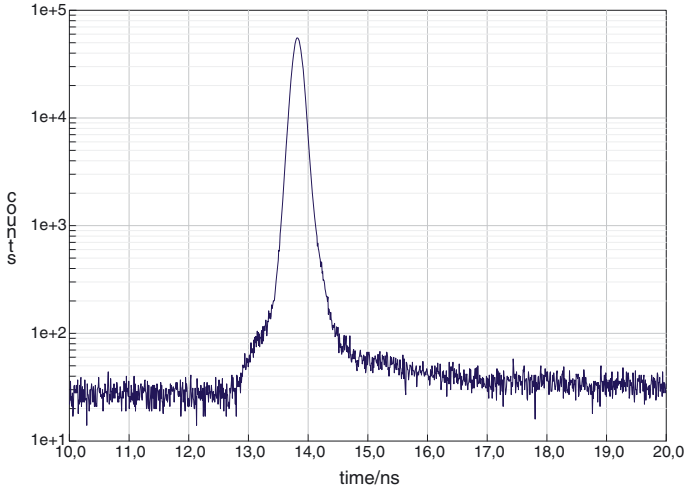

Figure 9. Time-interval histogram collected with 8 ps accuracy. Full width at half maximum is 187 ps.

DOI: 10.1002/cta

Table II. SPAD detector and quenching circuit characteristics.

Feat.\Ref.	[8]	[9]	[10]	[11]	[14]	[15]	This work
CMOS process	HV-0.8 µm	0.18 µm	Bi-2 µm	0.35 µm	0.13 µm	HV-0.18 µm*	0.18 µm
V_{DD}	5 V	1.8 V	5 V	3.3 V	2.5 V	>2.5 V	1.8 V
λ	833 nm	370–700 nm	—	820 nm	—	—	400–800 nm
T_{QNCH}	25 ns	—	1.5 ns	—	—	—	117 ps
T_{RST}	20 ns	—	—	≤2 ns	—	—	280 ps
T_{DEAD}	5–500 ns	30 ns	3 ns	40 ns–2 µs	7.5 ns	6 ns	5–500 ns
Time resol.	—	—	—	39 ps	—	80 ps at V_e of 3.5 V	187 ps at V_e of 0.9 V
Max. V_e	20 V	2 V	12 V	5 V	—	3.5 V	1 V
V_{BD}	50 V	10.2 V	20 V	24 V	—	20.3 V	10.3 V
AQC area	0.4×0.9 mm^2	80×80 µm^2	—	50×80 µm^2	15.6×8.9 µm^2	8×7.5 µm^2	12×22 µm^2
Ø SPAD	—	10–20 µm	100 µm	20 µm	16 µm	8.5 µm	12 µm
Fill factor	—	6%	—	36%	—	—	11%
P_{avg}	20 mW	—	5 mW	56 µW	—	~49 µA static power	2.5 µW at 1 MHz
I_{avg}	—	—	—	30 µA	—	—	350 nA at 125 kHz
DCR	—	60 kHz at 27°C	—	4.9 kHz at V_e of 5 V	—	180 Hz at V_e of 3.5 V	19 kHz at V_e of 0.9 V
PDE	—	5.5% at 470 nm and V_e of 2 V	—	42% at 430 nm and V_e of 5 V	—	20% at 470 nm and V_e of 3.5 V	6.2% at 480 nm and V_e of 1 V

V_{DD}, power supply voltage; T_{QNCH}, quenching time; T_{RST}, recharge time; T_{DEAD}, dead time; V_e, excess voltage; V_{BD}, breakdown voltage; P_{avg}, average power drawn from V_{DD}; I_{avg}, average current drawn from V_{SPAD+}; HV, high voltage; DCR, dark-count rate; SPAD, single-photon avalanche diode; AQC, active quenching circuit.
*This process is compliant with extremely low-defect density requirements by using high quality material and low doped epitaxial layer.

latching current ($I_{lat,}$ usually around 100 µA). Acting on the size and aspect ratio of M1, the capacitance of node A is also affected, thus it can be lowered to improve the quenching time.

Besides, we have achieved very good quenching and recharge times, which at the same time significantly decrease the average power consumption. This leads to high-speed photon detection using minimum dead time but still featuring a negligible afterpulsing rate. Also, the quick restore narrows the probability of false positives during the recharge phase, when the excess voltage goes from zero to 0.9 V. Although the evaluated time resolution is larger than that reported in a 0.18 µm technology [22], it is enough for PET which is the targeted application. The guard ring implant implemented into a standard CMOS technology degrades the time resolution because of the lateral drift and diffusion of the carriers generated by the photons absorption at the periphery of the active area. Of course these unwanted effects can be minimized by using TCAD tools. Least but not last, one has to consider that the fabrication costs are an issue if a PET scanner with a couple of tens of sensors arrays has to be built.

Table II displays the features of the proposed circuit in comparison with different reported AQCs. The table is not complete as not every data is available. Our circuit is one of the smallest and the one with the lowest power consumption reported, which are highly desirable properties for the elementary cell of a sensor array. Time resolution is not the best reported, but enough for the targeted application.

5. CONCLUSION

In order to build SPAD arrays of a practical size, quenching and restore circuits need to be of a reduced size and consume as low power as possible. Using active devices, the area occupation of an AQC without ballast resistor is usually smaller than the area required implementing other alternatives. In addition, the fast quenching mechanism and the direct disconnection of the current path allow for moderating the magnitude of the avalanche current, resulting in a significant power reduction. We have designed, fabricated, and tested an AQC based on these principles. It has been built in 0.18 µm standard CMOS which has no HV-option or low-noise requirements. It features one of the smallest area occupation reported and the lowest power consumption.

DOI: 10.1002/cta

ACKNOWLEDGEMENTS

This work has been funded by the Spanish Government through projects TEC2012-38921-C02 MINECO (European Region Development Fund, ERDF/FEDER), IPT-2011-1625-430000 MINECO and IPC-20111009 CDTI (ERDF/FEDER), by Junta de Andalucía through project TIC 2338-2013 CEICE and by the Office of Naval Research (USA) through grant N000141410355.

REFERENCES

1. Zappa F, Tisa S, Tosi A, Cova S. Principles and features of single-photon avalanche diode arrays. *Sensors and Actuators, A: Physical* 2007; **140**(1):103–112.
2. Li D *et al.*. Real-time fluorescence lifetime imaging system with a 32 × 32 0.13 um CMOS low dark-count single-photon avalanche diode array. *Optics Express* 2010; **18**(10):10257–10269.
3. Conti M. State of the art and challenges of time-of-flight PET. *Physica Medica* 2009; **025**(1):1–11.
4. Becker W. *Advanced Time-correlated Single Photon Counting Techniques*. Springer-Verlag: Berlin, 2005; 20–27.
5. Lerche CW *et al.*. Depth of γ-ray interaction within continuous crystals from the width of its scintillation light-distribution. *IEEE Transactions on Nuclear Science* 2005; **52**(3):560–572.
6. Gallivanoni A, Rech I, Ghioni M. Progress in quenching circuits for single photon avalanche diodes. *IEEE Transactions on Nuclear Science* 2010; **57**(6):3815–3826.
7. Niclass C, Gersbach M, Henderson R, Grant L, Charbon E. A single photon avalanche diode implemented in 130-nm CMOS technology. *IEEE Journal Selected Topics in Quantum Electronics* 2007; **13**(4):863–869.
8. Zappa F, Lotito A, Giudice AC, Cova S, Ghioni M. Monolithic active-quenching and active-reset circuit for single-photon avalanche detectors. *IEEE Journal Solid-State Circuit* 2003; **38**(7):1298–1301.
9. Faramarzpour N, Jamal Deen M, Shirani S, Fang Q. Fully integrated single photon avalanche diode in standard CMOS 0.18 μm technology. *IEEE Transaction on Electron Devices* 2008; **55**(3):760–767.
10. Mita R, Palumbo G, Fallica G. A fast active quenching and recharging circuit for single-photon avalanche diodes. *European Conference on Circuit Theory and Design (ECCTD 2005)* 2005; **3**:385–388.
11. Tisa S, Guerrieri F, Zappa F. Variable-load quenching circuit for single-photon avalanche diodes. *Optics Express* 2008; **16**(3):2232–2244.
12. Marwick MA, Andreou AG. A high voltage PMOS transistor for quenching of Geiger-mode avalanche photodiodes in deep submicron CMOS technologies. *International Device Research Symposium* (*ISDRS*), College Park, MD, 2007; 1–2.
13. Lin J, Andreou AG. A 32 × 32 single photon avalanche diode imager with delay-insensitive address-event readout. *IEEE International Symposium on Circuit and Systems* (*ISCAS*), Rio de Janeiro, 2011; 1824–1827.
14. Gronholm M, Poikonen J, Laiho M. A ring-oscillator-based active quenching and active recharge circuit for single photon avalanche diodes. *European Conference on Circuit Theory and Design* (*ECCTD 2009*), Antalya, 2009; 5–8.
15. Niclass C, Soga M. A miniature actively recharged single-photon detector free of afterpulsing affects with 6 ns dead time in a 0.18 μm CMOS technology. *IEEE International Electron Device Meeting* (*IEDM*), San Francisco, CA, 2010; 14.3.1–14.3.4.
16. Mita R, Palumbo G. High-speed and compact quenching circuit for single-photon avalanche diodes. *IEEE Transactions on Instrumentation and Measurement* 2008; **57**(3):543–547.
17. Bronzi D, Tisa S, Villa F, Bellisai S, Tosi A, Zappa F. Fast sensing and quenching of CMOS SPADs for minimal afterpulsing effects. *IEEE Photonics Technology Letters* 2013; **25**(8):776–779.
18. Pancheri L, Panina E, Dalla Betta GF, Gasparini L, Stoppa D. Compact analog counting SPAD pixel with 1.9% PRNU and 530 ps time gating. *Proceedings of the European Solid-State Circuit Conference* (ESSCIRC), Bucharest, 2013; 295–298.
19. Eisele A, Henderson R, Schmidtke B *et al.* 185 MHz count rate, 139 dB dynamic range single-photon avalanche diode with active quenching circuit in 130 nm CMOS technology. *Proceeding of International Image Sensor Workshop*, 2011; 278–280.
20. Giustolisi G, Mita R, Palumbo G. Behavioral modeling of statistical phenomena of single-photon avalanche diodes. *International Journal of Circuit Theory and Applications* 2011; **40**(7):661–679.
21. Zappa F, Tosi A, Dalla Mora A, Tisa S. Spice modeling of single photon avalanche diodes. *Sensors and Actuators, A: Physical* 2009; **153**:197–204.
22. Hsu MJ, Finkelstein H, Esener SC. A CMOS STI-bounded single-photon avalanche diode with 27-ps timing resolution and a reduced diffusion tail. *IEEE Electron Device Letters* 2009; **30**(6):641–643.
23. Morath CP, Vaccaro K, Buchwald W, Clark WR. Comparator-based measurement scheme for dark-count rate in single photon avalanche diodes. *IEEE Transactions on Instrumentation and Measurements* 2005; **54**(5):2020–2026.
24. Fishburn MW. Fundamentals of CMOS single-photon avalanche diodes. *Ph.D. Thesis*, *TU Delft*, 2012; 43–48.
25. Marwick MA, Andreou AG. Single photon avalanche photodetector with integrated quenching fabricated in TSMC 0.18 μm 1.8 V CMOS process. *IEEE Electronics Letters* 2008; **44**(10):643–644.
26. Giudice AC, Ghioni M, Cova S, Zappa F. A process and deep level evaluation tool: after-pulsing in avalanche junctions. *Proceedings of 33rd European Solid State Device Research Conference* (*ESSDERC*), Estoril, Portugal, 2003; 347–350.
27. Rochas A. Single photon avalanche diodes in CMOS technology. *Ph.D. Thesis*, *EPFL*, *Lausanne*, 2003; 131–136.
28. Kumar P, Charbon E, Staszewski RB, Borowski A. Low power time-of-flight 3D imager system in standard CMOS. *19th IEEE International Conference on Electronics*, *Circuits and Systems* (*ICECS*), Seville, 2012; 941–944.

DOI: 10.1002/cta