CMOS SPADs with up to 500 µm diameter and 55% detection efficiency at 420 nm

Federica Villa[a], Danilo Bronzi[a], Yu Zou[a], Carmelo Scarcella[a], Gianluca Boso[a], Simone Tisa[b], Alberto Tosi[a], Franco Zappa[a]*, Daniel Durini[c], Sascha Weyers[c], Uwe Paschen[c] and Werner Brockherde[c]

[a]*Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Piazza Leonardo Da Vinci 32, 20133 Milano, Italy;* [b]*Micro Photon Device S.r.l., Via Stradivari 4, 39100, Bolzano, Italy;* [c]*Fraunhofer IMS, Finkenstr. 61, 47057 Duisburg, Germany*

(*Received 19 August 2013; accepted 5 November 2013*)

1. Introduction

Single-photon sensitivity in the visible and near-infrared (NIR) ranges is required in many different applications, including single-molecule imaging [1], fluorescence lifetime imaging [2], fluorescent decays and luminescence in physics, chemistry, and biology [3], time-resolved spectroscopy [4], optical fiber characterization [5], quantum cryptography [6], quantum mechanics [7], astronomy [8], non-invasive testing of very-large-scale integrated circuits [9], and characterization of non-classical light sources such as single-photon generators [10]. In order to detect extremely faint signals, some single-photon detectors have been developed, such as photomultiplier tubes, microchannel plates, hybrid photo-detectors, superconducting single-photon detectors, single-photon avalanche diodes (SPADs), and silicon photomultipliers.

Semiconductor devices are strongly preferred for practical reasons: they require a relatively low bias voltage (typically 15–70 V), can be operated with no cooling, are insensitive to external magnetic fields, have miniature size, and are rugged, reliable and easy-to-use. SPADs are solid-state devices with single-photon sensitivity, able to provide a 'digital' output, unlike avalanche photodiodes, and with no readout noise, unlike charge-coupled devices. Silicon SPAD detectors are silicon p–n junctions, reverse-biased above the breakdown voltage in order to exploit the fast and intense avalanche build-up triggered by the absorption of a single optical photon. SPADs can be produced either in custom technology, aimed at optimizing detection performance through the custom tailoring of dopant concentrations and diffusion depth [11], or in standard complementary metal–oxide–semiconductor (CMOS) processes together with digital and analog circuitries, making them suitable for large detector arrays with on-chip image pre-processing.

Many applications demand large area single-photon detectors, for instance to simplify the setup alignment in single-molecule spectroscopy and microscopy [12], or to use them in quantum photometer for large telescopes [13], in application in space missions [14], and in photon starving applications, such as time-resolved diffuse spectroscopy [15], where a very large collection area is a must. Therefore, the aim of this work was to develop new SPADs in CMOS technology with large active area diameter up to 500 µm.

This paper is organized as follows. Section 2 focuses on the structure of the SPADs we have developed in a high-voltage standard CMOS 0.35 µm technology, integrated together with the quenching and active reset electronics. Section 3 describes the main parameters of SPADs, i.e. breakdown voltage, electric field uniformity, photon detection efficiency (PDE), dark counting rate

*Corresponding author. Email: franco.zappa@polimi.it

(DCR), afterpulsing probability, timing response, and crosstalk, and provides the complete SPAD characterization at different diameters (from 10 µm to 500 µm). Finally, Section 4 summarizes the results.

2. Large area CMOS SPAD with integrated quenching circuit

SPADs are essentially p–n junctions biased above the breakdown voltage (V_{BD}), in the so-called Geiger mode (in analogy with the gas counters of ionizing radiation), that require avalanche quenching and recharge mechanisms [16]. The voltage above breakdown is commonly called excess bias (V_{EX}). Upon photon detection that triggers an avalanche multiplication process through the SPAD junction, the quenching circuitry stops the avalanche, thus preventing too high current flow and power dissipation, and keeps it quenched for a fixed hold-off time ($T_{\text{hold-off}}$) in order to release trapped charges, thus avoiding the triggering of spurious avalanches (the so-called afterpulsing issue). Finally, the recharge circuitry brings the SPAD back to operation for the next detection cycle, by raising the bias voltage again above breakdown [17].

The exploitation of a custom technology gives the opportunity to tailor dopants concentration and diffusions width for optimizing SPAD performance [1]. On the other hand, very scaled CMOS technologies make it possible to develop very dense arrays of SPADs with in-pixel pre-processing electronics [18–20]. We developed SPADs in a 0.35 µm high-voltage technology with high level of cleanness and controlled substrate to reduce the number of defects that could deteriorate SPAD performance. Such a not-scaled technology node is a good compromise to achieve not just large arrays of smart pixels but also high performing SPADs.

Figure 1 represents the SPAD cross-section and a simplified representation of the electric field across the device. The p+ diffusion and the n-enrichment define the active absorption and avalanche region; the p-guard ring avoids premature breakdown. The bias voltage of the p-substrate influences the width of the neutral region in the HV-nwell, hence the timing performance of the detector. A thin passivation layer and an antireflection coating for the near-UV enhance the efficiency at wavelengths shorter than 500 nm. SPADs with different diameters (10 µm, 20 µm, 30 µm, 50 µm, 100 µm, 200 µm, and 500 µm) have been designed, fabricated, and characterized. To the best of our knowledge, these are the first CMOS SPADs with 500 µm active area diameter ever reported in the literature.

A mixed passive-active quenching circuit has been integrated close to the detector. This kind of quenching circuit gathers the advantages of the passive quenching with those of the active quenching circuit. In fact, a transistor acting as passive quenching resistor fast quenches the avalanche with consequent reduction of the number of carriers that cross the device, and an active circuit completes the quenching, also assuring a fixed duration of the hold-off time and fast quench and reset transient. The quenching circuit is connected to the anode, because, compared to the cathode, it shows lower parasitic capacitance, thus resulting in faster avalanche quenching and better photon timing resolution. The cathode is biased at $V_C = V_{BD} + V_{EX}$ ($V_{BD} \approx 25$ V, $V_{EX} \approx$ 2–6 V, $V_C \approx$ 30 V), the anode, connected to the circuit, varies from V_A = 0 V (SPAD ready to detect photons) and $V_A = V_{EX}$ (SPAD quenched). Since the substrate is in common with the CMOS electronics, it must be biased at $V_{bulk} = 0$ V, therefore the cathode–substrate junction is strongly reverse biased ($V_{rev} \approx 30$ V). The quenching and active reset circuit is described in [21] and a simplified schematic is shown in Figure 2. This circuit aims at speeding up the sensing of detector ignition and at promptly quenching the avalanche current build-up in order to reduce charge trapping causing afterpulsing. The transistor M_S is a variable-impedance sensing load, M_R is the reset transistor, R_{HOLD} is a variable resistance, whose value can be adjusted with an external voltage to

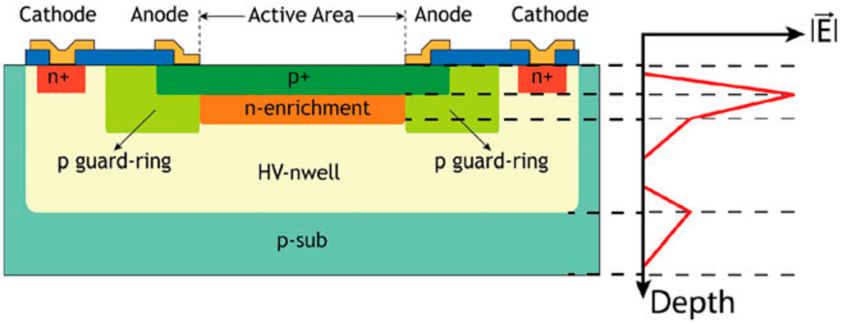

Figure 1. Cross-section of the SPAD developed in 0.35 µm high-voltage CMOS technology and the simplified electric field along the center of the device. The p+ and n-enrichment define the absorption and avalanche regions. (The colour version of this figure is included in the online version of the journal.)

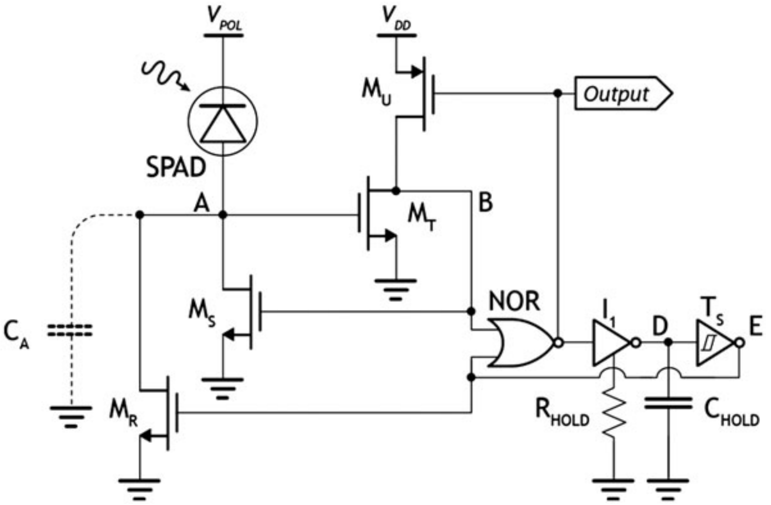

Figure 2. Simplified schematic of the integrated quenching circuit. C_A includes all (anode-to-cathode, anode-to-ground) stray capacitances.

modify the hold-off duration. Different circuits with properly sized transistors have been designed for each SPAD active area, in order to properly drive the different parasitic capacitances. Suitable CAD tools able to predict the electrical behavior of the ensemble of SPAD with integrated quenching circuit have been used [22].

3. SPAD characterization

This section provides a complete characterization of the SPADs considering all parameters that must be taken into account for high performing SPAD detectors. The excess bias voltage above breakdown (V_{EX}) heavily influences the performance of the SPAD, above all its photon detection efficiency (PDE), dark counting rate (DCR), i.e. the internal noise measured as the rate of spurious pulses due to carriers generated either thermally or by tunneling, time jitter (or photon-timing resolution) in acquiring the photon arrival time. Other parameters like afterpulsing probability, uniformity within the active area and crosstalk among adjacent SPADs play an important role in some applications. This section describes in detail all these parameters, the techniques used to measure them and the results we have obtained.

3.1. Breakdown voltage (V_{BD})

Since it is important to quote all performances at the same V_{EX}, as a first characterization we measured the breakdown voltage of our SPADs as a function of temperature.

In order to measure V_{BD}, it is necessary to reverse bias the SPAD and to trace the linear mode current–voltage (*I-V*) characteristics. When the SPAD is illuminated it is possible to clearly measure the 'on' avalanche current branch; instead when the SPAD is kept in dark it is important to check the presence of a faint flickering due to the toggling of the current between the 'on' (avalanche ignited) and the 'off' (device in quiescence, with no current flow) branches above breakdown [16]. The breakdown voltage $\mathrm{V_{BD}}$ is defined as the voltage corresponding to the intersection between the 'on' *I-V* characteristic photocurrent above breakdown and the 'off' (dark current) *I-V* characteristic below breakdown. This concept is clearly shown in [16] (Figure 1) and [17] (Figure 2), which show the ideal SPAD *I-V* curve, the typical *I-V* curve measured with an analog *I-V* tracer, and the extracted V_{BD}. Compared to SPADs developed with custom processes, CMOS SPADs have typically lower V_{BD}, of just few tens of volts, or even lower in more scaled technologies (because of the higher dopant concentrations). The V_{BD} has a linear dependence with temperature; in fact lattice vibrations become stronger when temperature increases, thus augmenting the probability that carriers interact with the crystal before gaining enough energy to cause impact ionization.

We acquired the current–voltage characteristic by means of a programmable electrometer (model 617 by Keithley Instruments Inc.). According to an extensive experimental characterization, breakdown voltage depends neither on the device under test nor on its diameter. In fact, the breakdown voltage spread is lower than 50 mV among SPADs fabricated within the same production run even with different diameters. Instead SPADs fabricated in different production runs present breakdown voltage variations up to 1 V, again independently of SPAD diameter. The measured dependence of V_{BD} on temperature and the extrapolated linear fit are shown in Figure 3; the temperature coefficient is 36.2 mV/°C, therefore by varying the temperature from −50° C to

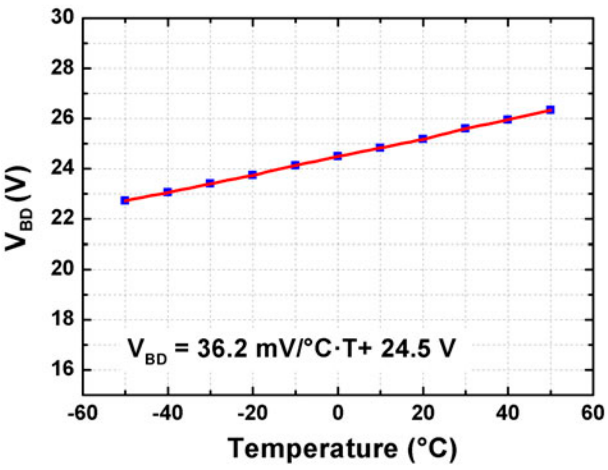

Figure 3. Breakdown voltage vs. temperature: experimental data (diamonds) and linear fitting (line). The same trend has been measured for all the SPADs, independently from the active area diameter. (The colour version of this figure is included in the online version of the journal.)

+50° C, the breakdown voltage just slightly moves, from 23.5 V to 26.5 V, with a variation of ±6% compared to room temperature (20° C with V_{BD} = 25 V).

3.2. *Active area uniformity*

An important issue in SPAD fabrication is electric field uniformity within the active region in order to guarantee uniform sensitivity. Guard rings are commonly employed to smooth the electric field at the edges and to avoid premature lateral breakdown. Ideally, a uniform electric field across the active area gives a carrier the same triggering probability anywhere it is photo-generated. Hence, the resulting photon detection efficiency should be constant over the active area, while it should sharply drop outside its periphery. Instead, edge effects and alignment tolerances in the manufacturing process generate localized electric field peaks at the periphery, while thin and low doped bulk diffusions cause high series resistance, which causes an almost self-quenching of the avalanche process in the center of active area. In fact, as shown in [23], the SPAD resistance in the center of the device is higher than in the edges, and thus in large area SPADs the avalanches ignited in the center could self-quench before properly being detected by the front-end electronics. These are indeed the most common causes of non-uniformity in SPADs.

In order to estimate the uniformity of a SPAD, a laser is focused into a diffraction limited spot within the active area, by means of an objective; by scanning the entire active area and counting the avalanches in each position, a uniformity map of the SPAD response is obtained. In this paper, we quantify the uniformity error with the following equation:

$$\varepsilon_U = \frac{C_{\max} - C_{\min}}{\bar{C}}.100, \tag{1}$$

i.e. the difference between maximum C_{max} and minimum C_{min} counts measured within the active area, normalized by the average counts of the whole active area. The smaller the value, the better the uniformity within the active area.

The uniformity of the electric field was tested by focusing a 2 μm laser spot within the active area through a 50× objective and then counting the photon ignition rate. Figure 4 reports the typical uniformity map of a small (20 μm diameter) and of a large (500 μm) SPAD. The results highlight a good uniformity within the active area, and no premature edge breakdown is visible. The guard ring effectively reduces the electric field at the SPAD edges, but also the effective radius by about 2.5 μm, i.e. the effective diameter is 5 μm smaller than the drawn nominal one.

The central part of the 500 μm SPAD has a slightly lower (detection) efficiency in respect to the peripheral region. This phenomenon is perfectly explained by the typical SPAD resistance map reported in [23], in which a higher series resistance is present at device center compared to its edges. That causes a partial self-quench of avalanches ignited far away from the SPAD periphery, where the cathode contact is located.

Eventually we computed also the uniformity error for all SPADs, by means of (1), considering just the effective diameter, i.e. putting aside the peripheral region influenced by the guard ring. For all SPADs such uniformity error is better than 10%, as shown in Figure 5.

3.3. *Photon detection efficiency (PDE)*

Photon detection efficiency (PDE) is the ratio between the number of avalanche pulses and the number of photons that reached the detector's active area. PDE depends on the two main processes involved in single-photon detection: photon absorption and avalanche multiplication [16]. The absorption efficiency η is defined as the ratio between absorbed photons and the whole number of photons that cross the active area of the junction. Therefore [16]:

$$\eta = (1 - R) \cdot e^{-\alpha D} \cdot (1 - e^{-\alpha W}), \tag{2}$$

where R is the optical power reflection coefficient, α the silicon absorption coefficient, D the depth at which the depleted layer starts (i.e. the thickness of the top neutral region), and W the depleted layer thickness. The power loss, due to the reflection at the interface between air and silicon, can be reduced by means of an antireflection coating tailored for the desired optical wavelength. The avalanche triggering probability is defined as the probability of a pair of photogenerated carriers to trigger

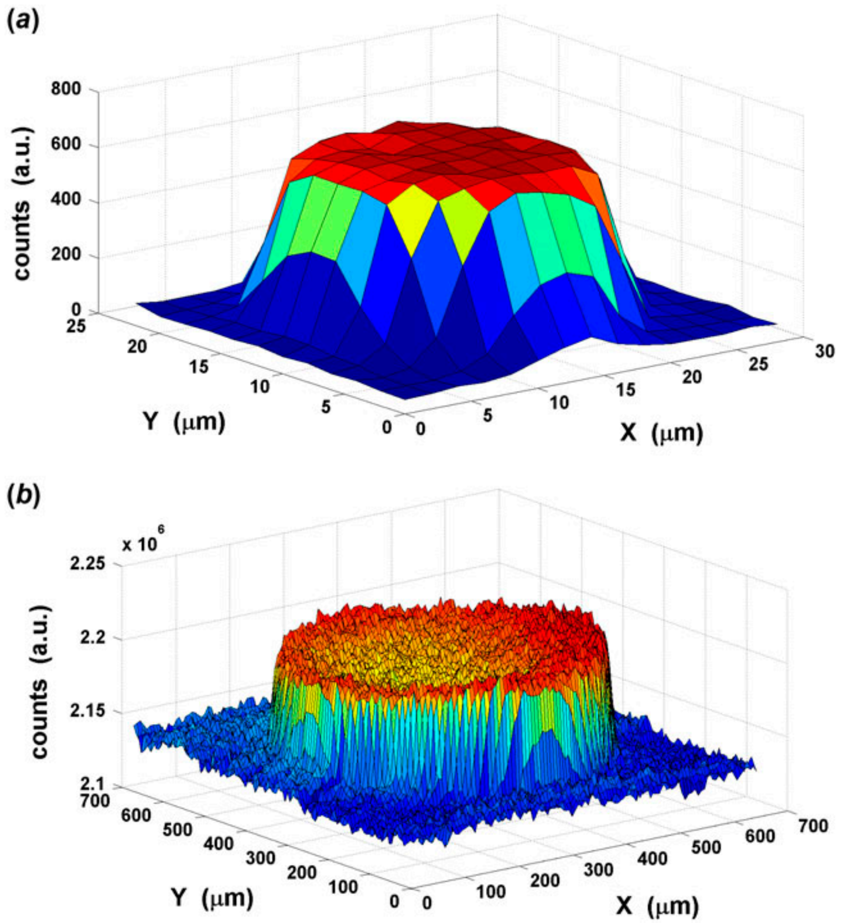

Figure 4. Detection uniformity of a 20 μm (*a*) and a 500 μm (*b*) SPAD, measured by scanning a light spot with 2 μm and 6 μm steps, respectively. Note the drop in photon detection efficiency at the center of the large-area SPAD compared to its edges, as discussed in the text. (The colour version of this figure is included in the online version of the journal.)

a self-sustaining avalanche multiplication process; it can be approximated as [24]:

$$P_T = 1 - e^{-\frac{V_{ex}}{V_C}}, \qquad (3)$$

where V_C is a parameter dependent on depleted layer thickness and on electron and hole ionization coefficients. As can be seen, PDE strongly depends on excess bias.

The optical set-up we used to measure the PDE at different wavelengths is based on a broadband and stable light source, a monochromator for the wavelength range of interest (from near UV to near IR), optical filters, and an integrating sphere to obtain a planar light beam to shine onto the SPAD under test.

We measured the PDE of each SPAD at three different excess bias voltages (2 V, 4 V, and 6 V) in the 300–1100 nm wavelength range, with 5 nm steps. Figure 6 shows the dependence of PDE on photon wavelength of a 50 μm and a 200 μm SPAD; all other SPADs exhibit very similar PDE curves. The large oscillations visible in Figure 6 are due to the SiO_2 and Si_3N_4 that cover the CMOS chips and strongly depend on the incoming light's angle. The peak PDE is about 55% at 450 nm, is still 20% at 300 nm, and is 5% at 850 nm, as shown in Figure 7. Figure 7 also highlights that different excess biases cause strong PDE variations, as expected from (2) and (3). With respect to other CMOS and custom SPADs [1,18–20], the PDE is enhanced in the near-UV, thanks to a thinner Si_3N_4 passivation layer, optimized for near-UV that covers the whole chip.

SPADs from 10 μm to 200 μm diameter show a slight increase in PDE at larger diameter, whereas the 500 μm SPAD exhibits a PDE lower that the 200 μm one. Both these facts can be explained considering the uniformity measurements presented in the previous section: by increasing the SPAD diameter, the guard ring effect of reducing the effective active area becomes negligible, whereas the 500 μm has a lower efficiency in the central area, as already reported in Figure 4.

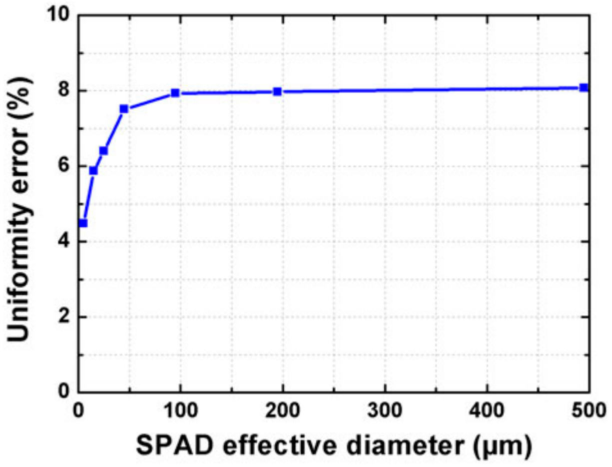

Figure 5. Detection uniformity error vs. SPAD effective diameter, which is 5 µm shorter than the nominal one for all the SPADs. (The colour version of this figure is included in the online version of the journal.)

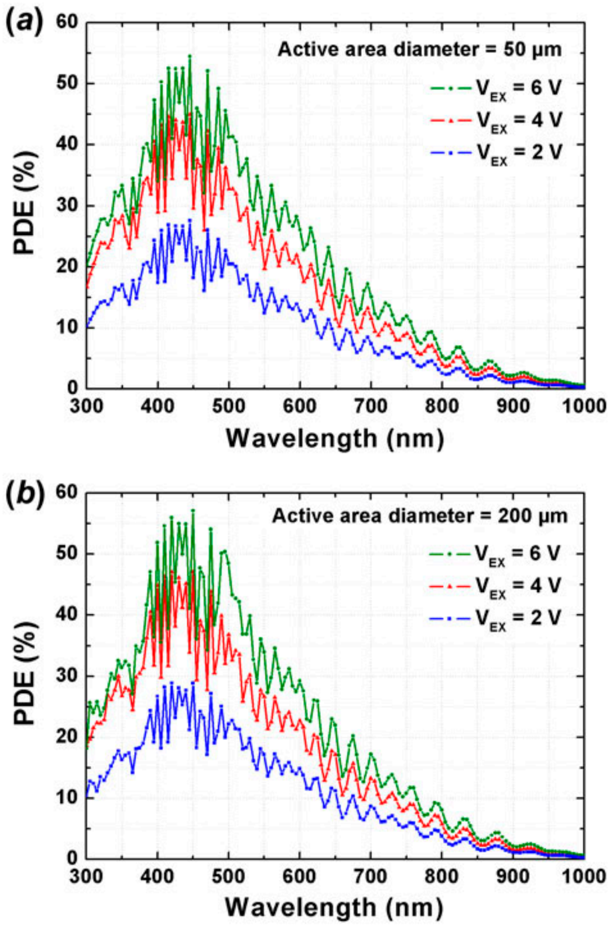

Figure 6. Photon detection efficiency vs. wavelength at 2 V, 4 V, and 6 V excess bias, for a 50 µm (*a*) and a 200 µm (*b*) SPAD. The peak PDE is about 55% at 450 nm, and still 5% at 850 nm. (The colour version of this figure is included in the online version of the journal.)

3.4. *Dark counting rate (DCR)*

A SPAD's noise is usually called dark count, i.e. an avalanche ignition not due to a photon, though undistinguishable from the useful signal, that is triggered by thermal generation (dominant at high temperature) or tunneling (dominant below about −5 °C). DCR depends on processing quality (mainly defects density), technological parameters (processes and doping doses), layout (active area dimension and shape), and on external operating conditions (excess bias and temperature).

The DCR is measured by keeping the SPAD in a dark environment while counting the ignition rate. Since the device is kept quenched for the hold-off time duration after each ignition, the number of measured ignition DCR_{meas} should be corrected in order to estimate the real DCR, through the following equation:

$$DCR = \frac{DCR_{meas}}{1 - DCR_{meas} \cdot T_{hold-off}}. \qquad (4)$$

The hold-off duration ($T_{hold\text{-}off}$) has been kept long enough to result in negligible afterpulsing. According to the dependence of afterpulsing probability on hold-off described in the next subsection (see Figure 12), we chose a 60 ns hold-off duration for small (10–50 µm diameter) SPADs, but 150 ns for large (100–500 µm diameter) SPADs.

The DCR has been measured for all SPADs at different temperatures (from −50°C to +50°C), and at different V_{EX} (from 4 V to 6 V). The DCR for the 10 µm diameter SPAD is as low as 1 cps at room temperature. The 50 µm SPAD, with about 100 cps DCR at room temperature is comparable with the best-in-class custom SPADs [11,25]. Also the 500 µm, with a DCR of 100,000 cps at room temperature is suitable in many applications requiring very large areas and no cooling. Note that DCR values around 100 kcps have been reported in the last few years, but for CMOS SPADs with just 18 µm or smaller devices, instead of the present 500 µm ones. Figure 8 shows the dependence of DCR on temperature for all fabricated SPADs at 4 V and 6 V excess bias. The smaller SPADs exhibit a change in the slope of the characteristic at about 0°C, since for lower temperatures the dominant effect in the DCR is tunneling generation and no longer the thermal generation. Instead large SPADs show a DCR always dominated by thermal generation in the considered temperature range. A possible explanation can be inferred from Figure 9, which shows the trend of DCR when changing the SPAD diameter at room temperature, at different excess bias. For small diameters (<50 µm) DCR increases linearly with area (i.e. quadratic with diameter), while for larger SPADs it is worst (i.e. steeper) because the probability to have at least one localized microplasma (i.e. an extended defect such as metal precipitates, dislocations, etc.) inside the active

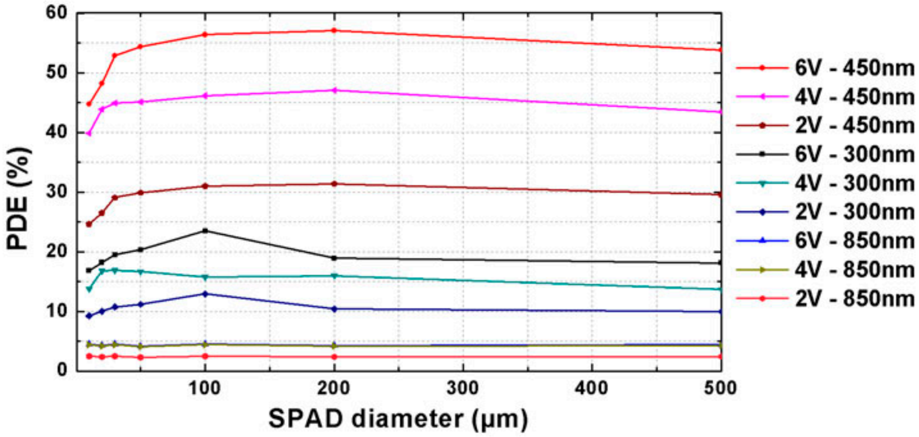

Figure 7. PDE vs. SPAD diameter at 2 V, 4 V, and 6 V excess bias, and at 300 nm, 450 nm (peak), and 850 nm wavelengths. (The colour version of this figure is included in the online version of the journal.)

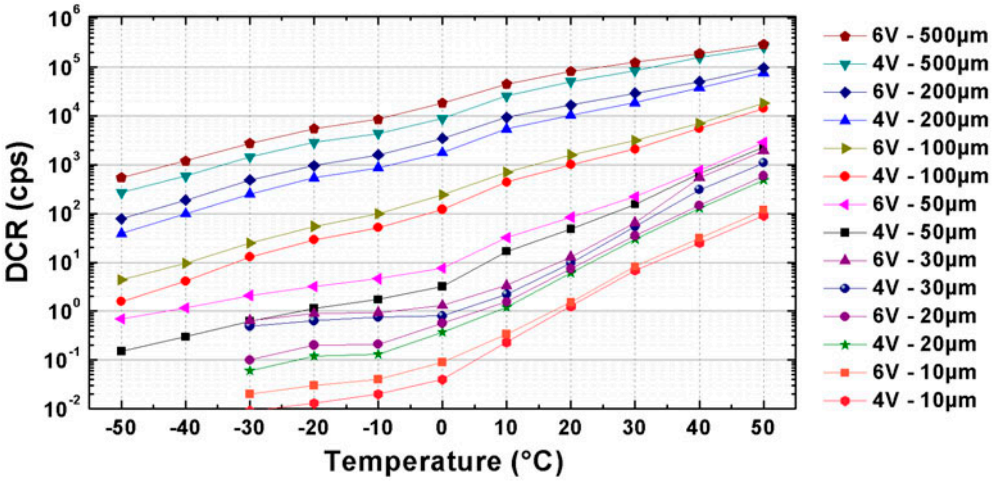

Figure 8. Dark counting rate vs. temperature for SPADs with diameter ranging from 10 µm to 500 µm at 4 V and 6 V of excess bias. (The colour version of this figure is included in the online version of the journal.)

area increases. In Figure 8 it is possible to note that the slope of DCR vs. temperature of largest SPADs differs from the smallest SPADs. Probably microplasmas have lower activation energies than normal defects, thus causing higher DCR, which masks thermal generation and tunneling contributions.

Finally, we investigated the uniformity of DCR over many SPADs, which is a signature of the reliability and reproducibility of the fabrication process. To this aim, we measured the DCR at room temperature of some arrays of SPADs consisting of 2048 pixels with 30 µm SPADs, and 512 pixels with 100 µm SPADs. The cumulative distribution function shown in Figure 10 for the two arrays confirms that the larger the diameter, the lower the yield. In fact, if we define as 'hot' those SPADs with a DCR higher than the average value of the best (i.e. those with lower DCR), the 30 µm devices show less than 5% hot pixels, whereas the 100 µm array has almost 30% of hot ones. Anyway note that in our case 'hot' refers always to devices with DCR lower than 100 kcps, while CMOS SPADs reported in literature often show hot devices with DCR higher than 10^6 cps, which must then be permanently switched off by in-pixel electronics in order not to impair the overall array operation.

3.5. *Afterpulsing*

Another source of noise, peculiar to SPADs, is the afterpulsing. Unlike dark counts, afterpulsing causes a non-linear distortion of the measured signal, since it is correlated with the signal itself. It is caused by different

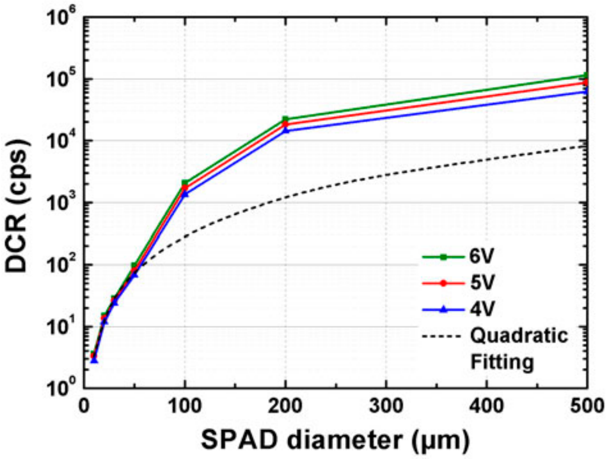

Figure 9. DCR vs. SPAD diameter at room temperature and 4 V, 5 V, and 6 V of excess bias. The quadratic fitting shows that the small-area SPADs have a quadratic trend with diameter, while large ones show a more than quadratic trend. (The colour version of this figure is included in the online version of the journal.)

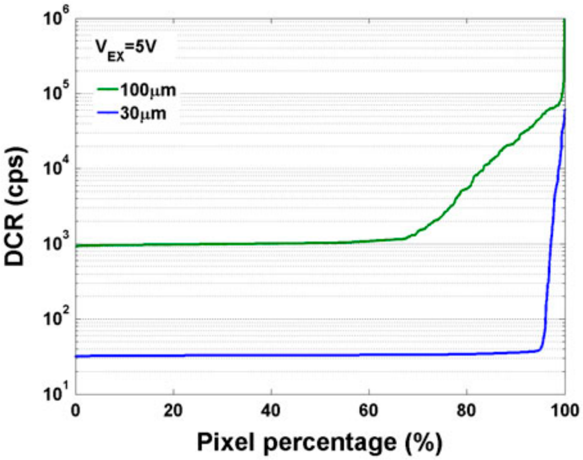

Figure 10. Cumulative distribution function of DCR for two arrays of 2048 pixels (30 μm diameter) and 512 pixels (100 μm diameter). Note that larger SPADs show higher DCR and also higher percentage of 'hot' devices. (The colour version of this figure is included in the online version of the journal.)

local defects in the depletion layer, which capture some carriers during the avalanche current flow and then release them after a considerable delay (from nanoseconds up to hundreds of nanosecond in silicon). During the hold-off time the SPAD is kept off, but if the release occurs when the SPAD voltage is driven above breakdown, the carrier can retrigger an avalanche process, thus causing a spurious ignition. Therefore afterpulsing is not just a boost in the DCR, since it is proportional to the overall counting of photons plus dark counts. By cooling the detector, the DCR decreases but afterpulsing increases, since the release time constant becomes longer. By improving the sensing and quenching electronics, it is possible to speed up the avalanche detection and to minimize the charge flow, thus reducing afterpulsing probability [17].

Afterpulsing can be measured by means of time-correlated carrier counting (TCCC) technique [26], based on the collection into a histogram of the time intervals between two successive avalanche ignitions. The theoretical histogram with no afterpulsing (i.e. a simple exponential decay) is subtracted to the experimental histogram, in order to highlight just the effect of afterpulses. The afterpulsing probability is computed as the integral sum of the obtained histogram divided by the number of valid events in the experimental histogram (Figure 11).

We measured the afterpulsing probability at room temperature by means of a multichannel analyzer (Varro 16k, by Silena), at different hold-off durations (from 20 ns to 150 ns) and different excess bias (4 V, 5 V, 6 V). Results are shown in Figure 12. As expected, afterpulsing probability increases with shorter hold-off time (because carrier release becomes more effective in triggering a spurious ignition), higher excess bias and larger SPAD area (because more carriers get trapped due to the higher number of carriers flowing during each avalanche process). SPADs with small diameter (≤50 μm) have almost identical trends of afterpulsing probability vs. $T_{holf\text{-}off}$, and afterpulsing becomes negligible (probability <1%) at hold-off durations shorter than 40 ns. This result is excellent, taking into consideration the large diameter compared to all other CMOS SPADs so far reported in literature. Furthermore, even if 'no afterpulsing' SPADs are sometimes claimed in literature, those measurements are often not correct, since they are often based on correlation function with minimum time slot of 80 ns or longer, thus hiding any actual afterpulsing present with shorter time decays.

The good afterpulsing performance we achieved is due to both the high cleanness of the CMOS processing employed in the Fraunhofer IMS foundry and the fast mixed passive-active quenching performed by the front-end circuit shown in Figure 2 (see [21]). Large-area SPADs have higher afterpulsing probability, because of the larger parasitic capacitance at the sensing (anode) node that delays the intervention of the quenching circuit. Nevertheless the afterpulsing probability is quite low (lower than 1% with 150 ns hold-off for 500 μm SPAD) even compared with other 0.35 μm CMOS SPADs with much smaller area [27–29], which require $T_{hold\text{-}off}$ longer than 500 ns to reach less than 1% afterpulsing probability.

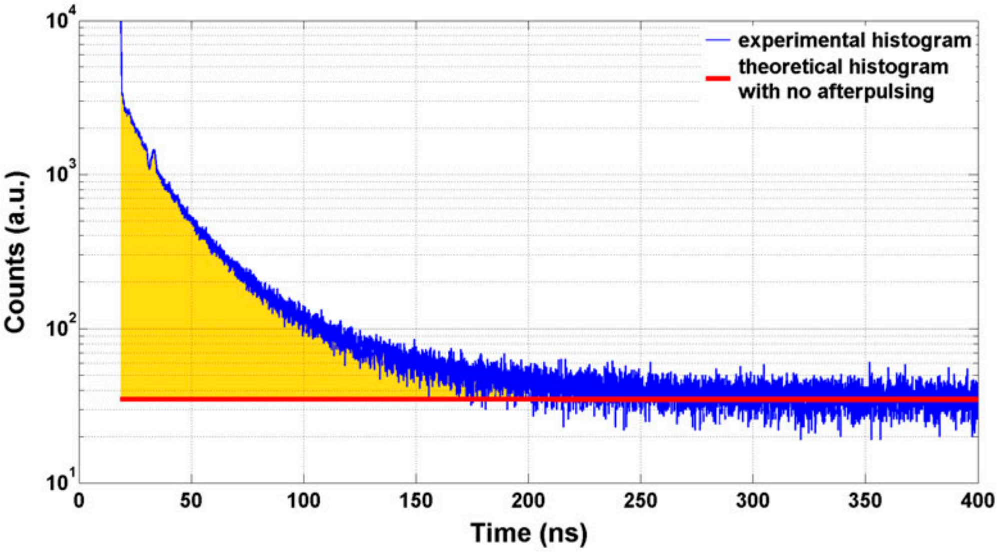

Figure 11. Experimental histogram in blue, compared to the theoretical histogram in red. The afterpulsing probability corresponds to the integral sum of the difference between the two histograms (area in yellow), normalized to the total number of the events in the experimental histogram. (The colour version of this figure is included in the online version of the journal.)

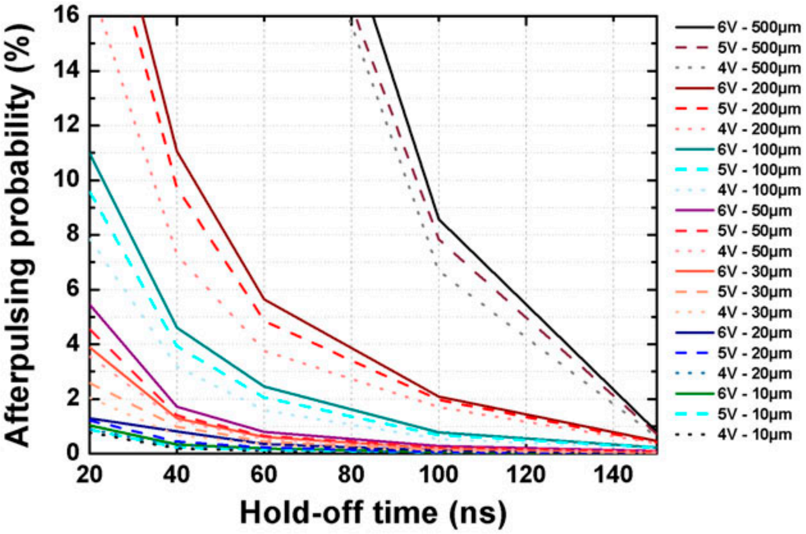

Figure 12. Afterpulsing probability vs. hold-off duration at room temperature with 4 V, 5 V, and 6 V of excess bias. For small-area SPADs the afterpulsing probability is negligible (<1%) with a hold-off time longer than 40 ns, while large SPADs require at least $T_{\text{hold-off}} = 100\,\text{ns}$. (The colour version of this figure is included in the online version of the journal.)

3.6. Timing jitter

The time precision, or photon-timing jitter, is the SPAD quality in the identification of the photon arrival time. Usually it is measured in terms of the full-width at half maximum (FWHM) of the time distribution of arrival times to a repetitive collection of fast laser pulses. In order to achieve excellent timing performance (i.e. jitter of few tens of picosecond), it is mandatory to perform a low threshold sensing of the avalanche current [21]. Nevertheless, the intrinsic timing jitter depends on the SPAD itself, namely its series resistance and spurious capacitive loadings. High excess bias generally improves timing performance because the avalanche current is proportionally more intense and the triggering is more precisely detected by the sensing electronics [16],[17]. Moreover, besides the Gaussian distribution of arrival time, which is caused by the avalanche build-up statistics, SPADs usually show also an exponential tail in their timing response, due to photons absorbed in the neutral region that slowly diffuse into the avalanche region [16],[17].

Both FWHM precision and diffusion tail's time constant can be measured by means of the time-correlated single-photon counting (TCSPC) technique [30], which collects the histogram of the time delays between repetitive sharp laser pulses shone to the SPAD and its triggering. Since the SPADs require a hold-off time after each avalanche ignition and also since standard timing boards can measure only one time interval for each laser fire, less than one photon must be detected by the SPAD for each laser pulse in order to reconstruct the timing waveform without distortion [30].

Photon timing responses were characterized in different conditions, by means of TCSPC technique, though the SPC-130 timing board by Becker & Hickl. First of all, we tested the entire ensemble of SPAD and integrated quenching circuit at three different wavelengths (λ = 390 nm, 520 nm, and 780 nm), by using high repetition rate (80 MHz) mode-locked lasers (Menlo Systems, TC-1550), because different excitation wavelengths cause different time responses of the SPAD. The whole system had an overall jitter of 19 ps. The timing responses for all the SPADs at λ = 520 nm and V_{EX} = 6 V are shown in Figure 13, whereas Figure 14 reports the FWHM vs. SPAD diameter at 5 V of V_{EX} and λ = 390 nm, 520 nm, and 780 nm.

For SPADs with small dimension (<50 μm diameter) the FWHM is better than 120 ps and as low as 56 ps in the best conditions (10 μm diameter, λ = 780 nm, V_{EX} = 6 V). The diffusion tail is almost negligible (always shorter than 100 ps). Overall, the timing resolution is not as good as custom SPADs [25], but it is comparable to other CMOS SPADs [27].

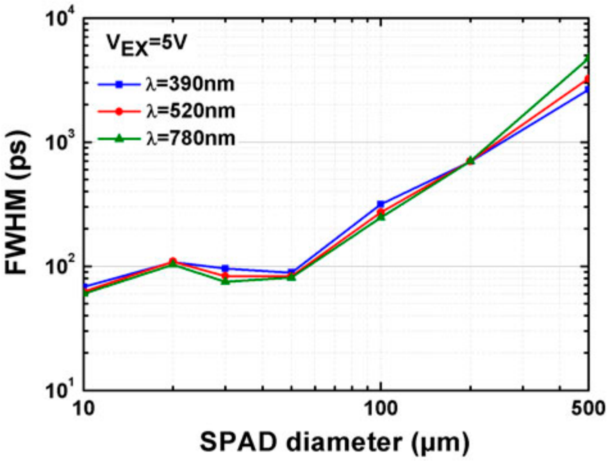

Figure 14. Photon timing jitter vs. SPAD diameter at 390 nm, 520 nm, and 780 nm, at 5 V excess bias. (The colour version of this figure is included in the online version of the journal.)

Large-area SPADs exhibit a much wider timing. Reasons could be manifold: the avalanche build-up has wider statistical or systematic spread; avalanches ignited by photons absorbed in different radial positions are sensed by the electronics with different time delays due to the avalanche propagation and also the time dependent variable voltage fluctuation through the radial extension

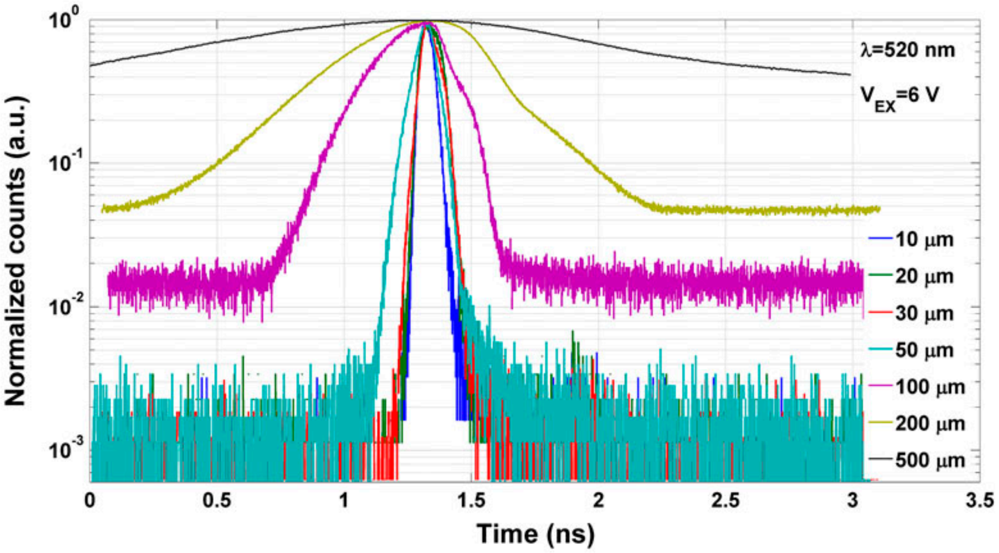

Figure 13. Photon timing responses at λ = 520 nm of SPADs with different diameters, operated by on-chip integrated quenching circuits at 6 V excess bias. (The colour version of this figure is included in the online version of the journal.)

of the well; the integrated quenching circuit is driven by a very faint signal (since the avalanche current has to charge very large stray capacitances).

In order to investigate the timing spread due to photon absorption in different position over the active area, we performed a set of measurements focusing a laser beam (Antel Optronics PS-820F pulsed laser with 10 ps FWHM) in a diffraction limited spot with a 50× objective. By scanning the SPAD from the center to the periphery, we measured point by point the FWHM of the timing response (Figure 15) and the absolute time delay of the peak in the timing response. We noticed that for small SPADs both the FWHM and the maximum of the time response are constant for all spot positions and timing jitter is almost identical to the one measured with un-focused light. Instead, for large SPADs, the FWHM improves when moving away from the center and the time delay corresponding to the maximum of the response gets shorter when the avalanche is triggered in the periphery, as we can see in Figure 16. This behavior is well visible in the 500 µm SPAD, whose FWHM when focused in the center is 4.2 ns while it shrinks to 1.6 ns in the periphery; and also the peak of the response in the periphery has 5 ns in advance.

These remarks highlight that the cause of the worsening of photon timing jitter in large SPADs could be due to

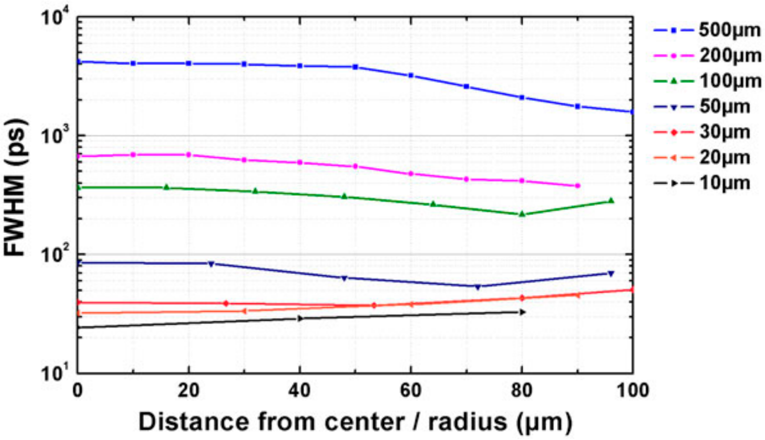

Figure 15. Photon timing response vs. radial position of illumination spot (expressed in % of the radius). (The colour version of this figure is included in the online version of the journal.)

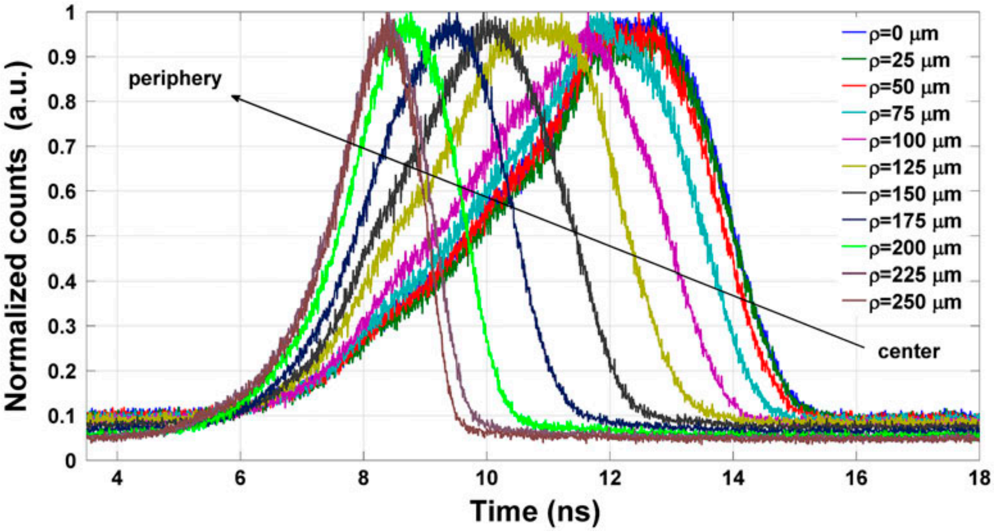

Figure 16. Photon timing response of a 500 µm SPAD at different radial positions of the illumination spot (ρ) from 0 µm (center) to 250 µm (periphery). Curves have been normalized to the same peak. (The colour version of this figure is included in the online version of the journal.)

the high series resistance caused by depleted region (see Figure 1), especially when photons are absorbed in the center of the active area. In fact in CMOS technology, usually the SPAD active junction is fabricated within an n-well, in order to preserve the isolation between the detector and the surrounding electronics. In this case, the bias between n-well and p-substrate strongly influences the timing performance: in fact, if the cathode-substrate junction is strongly reverse-biased the neutral region is reduced with the effect of increasing the series resistance of the device, thus worsening the FWHM value. On the other hand a thinner neutral region results in a lower number of photons absorbed in this region and so a reduced diffusion tail.

In order to prove such hypotheses, we measured the timing response with an external off-chip quenching circuit (inset of Figure 17), in order to be able to bias the substrate at different voltages compared to the n-well, representing the SPAD's cathode. Indeed, the voltage across p-substrate and n-well cathode modulates the width of the depleted region, hence the neutral region path through the n-well, thus changing the value of the series resistance. The anode is biased at a constant negative voltage ($V_A = -V_{BD} = -26.5\,\text{V}$), the substrate is biased at a variable voltage from 0 V to −30 V and the quenching circuit biases the cathode (V_C) at V_{EX}, senses the avalanches and quenches the detector lowering $\mathrm{V_C}$ to 0V. The most similar condition to the previous cases, when the SPAD was quenched by the integrated quenching circuit, is the one with $V_{bulk} = -30\,\text{V}$, since the cathode–substrate junction is reversely biased at about 30 V. In Figure 17 we observe that by increasing the cathode-substrate reverse bias the FWHM becomes wider whereas the diffusion tail time constant becomes shorter. This fact confirms that, by reverse biasing the cathode-substrate junction, the neutral region becomes thinner and consequently the number of photons absorbed in this region decreases (thus a shorter diffusion tail), but also the avalanche current becomes fainter with a consequent worsening of the timing jitter.

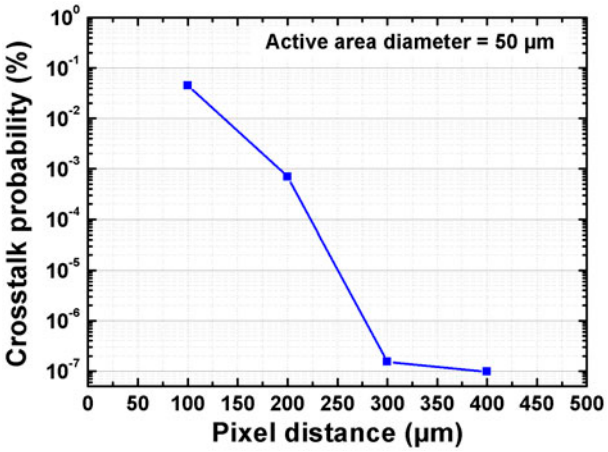

Figure 18. Crosstalk probability between 50 μm SPADs placed at different distances. (The colour version of this figure is included in the online version of the journal.)

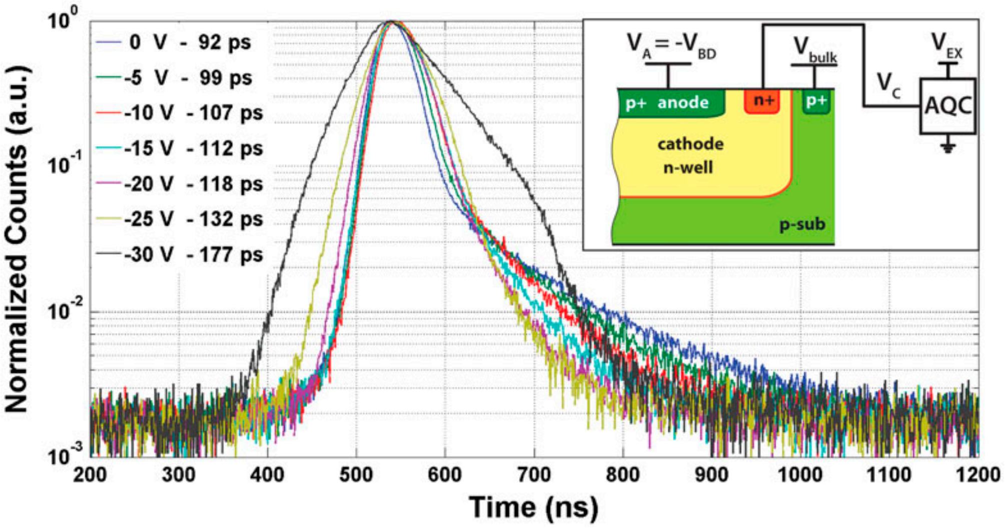

Figure 17. Photon timing responses of a 100 μm SPAD at λ = 850 nm, at different substrate voltage form 0 V (same voltage as the cathode) and −30 V (same voltage as the anode). Curves have been normalized to the same peak. For each curve the substrate voltage and the resulting FWHM are indicated. The inset shows the SPAD and external active quenching circuit. V_{EX} is the excess bias at which the SPAD is biased and V_{BD} is the breakdown voltage (that is about 26.5 V for the SPADs considered in this work). The bulk voltage can be modified through a pad connect to the p-sub of the wafer. (The colour version of this figure is included in the online version of the journal.)

Table 1. Performances of the SPADs presented in this work, at 6 V excess bias.

	10 µm	20 µm	30 µm	50 µm	100 µm	200 µm	500 µm	unit
Peak PDE	45	48	53	54	56	57	54	%
DCR (T = 25°C)	5	21	39	154	2,373	22,821	101,987	cps
$T_{\text{hold-off}}$ (afterpulsing < 1%)	20	40	60	60	100	150	150	ns
FWHM with integrated QC (λ = 780 nm)	56	80	76	82	290	608	4470	ps
FWHMwith external AQC (λ = 850 nm)					92			ps
ε_U*	4.5	5.9	6.4	7.5	7.9	8.0	8.1	%
Crosstalk (100 µm distance)				0.05				%

*The uniformity error has been computed with Equation (1).

Table 2. Comparison of the performances of the SPADs presented in this work (30 µm diameter) and the state-of-the-art SPAD in both custom and CMOS (0.35 µm, 0.12 µm, and 90 nm) technologies. For each peak PDE value the corresponding wavelength is reported in brackets. The afterpulsing probability (AP) achieved with the given hold-off time is also reported in brackets.

	PDE (%)					
	Peak	850 nm	DCR/area (cps/µm^2)	$T_{\text{hold-off}}$ (ns)	Timing FWHM (ps)	Diameter (µm)
this work	53 (450 nm)	4.5	0.055	40 (AP < 1%)	85	10 20 **30** 50 100 200 500
[11]	60 (650 nm)	29	0.051	n.a.	93	50
[33]	42 (450 nm)	4.5	3.979	600 (AP < 1%)	39	20
[18]	28 (500 nm)	6	0.249	100 (AP < 0.02%)	200	8
[19]	44 (690 nm)	21	3.466	15 (AP < 0.375%)	52	6.4

In conclusion, in order to obtain good timing performance with large-area SPADs, the substrate should be biased at about the same voltage of the cathode. To do that an external quenching circuit should be used because integrating electronics with SPADs forces to connect the bulk of the wafer (that is also the bulk of the SPAD) to ground. On the other hand SPADs with diameters smaller than 100 µm present not such a problem in photon timing applications, and can be effectively operated by in-pixel quenching circuits.

3.7. *Crosstalk*

When a silicon p–n junction operates in avalanche regime it emits photons due to hot-carrier [31]. In a monolithic detector array, photons emitted from a SPAD can trigger an avalanche into another detector, thus causing optical crosstalk between array pixels. Crosstalk can also have an electrical nature, due to unwanted couplings between neighboring pixels or quenching and sensing circuits. Crosstalk measurements can be performed exploiting a setup similar to that used in afterpulsing measurement based on TCCC. The pulse from an 'emitting' SPAD feeds the START of the timing board while the pulse from the 'detecting' SPAD provides the STOP and the time intervals between two successive START and STOP signals are collected in a histogram. If no crosstalk is present, the distribution of time delay should be an exponential, since the probability to detect photons is constant in time, but only the first detected photon is considered in the time intervals histogram. Otherwise, with the presence of crosstalk, the repetitive collection of the measurements produces a shape resembling the avalanche current waveform through the SPAD. Subtracting the ideal exponential without crosstalk to the measured time intervals histogram the remaining area is proportional to the total crosstalk probability.

The crosstalk between adjacent 50 µm SPADs has been tested with a Multichannel Analyzer (Varro 16k, by Silena) placed at 100 µm, 200 µm, 300 µm, and 400 µm distance. Figure 18 shows that even for the closest SPADs at 100 µm distance, the crosstalk is negligible (less than 0.1%) thus making these SPADs suitable for large detector arrays. Such crosstalk probability is comparable with data reported for other 0.35 µm CMOS SPADs with 20 µm diameter [32].

4. Conclusions

We have reported the characterization of novel large SPADs fabricated in CMOS technology, which represent the new state-of-art for SPADs. SPADs with 10 µm, 20 µm, 30 µm, 50 µm, 100 µm, 200 µm, and 500 µm diameters have been designed, fabricated, and characterized. This is the first time in the literature that CMOS SPADs with diameters larger than 50 µm have been reported. The performances in terms of DCR, afterpulsing, and crosstalk are comparable to those of the best in class custom SPADs. The PDE, lower than the one achievable by custom SPADs in the NIR, still outperform the other CMOS SPADs presented in the literature. Although the timing response of large-area SPADs is quite poor when operated by the on-chip integrated quenching circuit, it can be definitely improved by means of an external quenching circuit that leaves the chance to properly bias the substrate at the same voltage of the cathode, instead of pinching the n-well almost off. The electric field is uniform within the entire active area, with a non-uniformity lower than 10% in all the SPADs.

Table 1 summarizes the performances of all characterized SPADs, whereas Table 2 compares the performances of the SPADs presented in this work with other state-of-art devices, fabricated either in custom technology [11], 0.35 µm CMOS technology [33], or scaled [18],[19] CMOS technologies.

The large-area SPADs can be used stand-alone, in applications that require single pixel acquisitions and large active area in order to collect as much photon as possible of a very faint optical signal. For instance, in time domain photon migration experiments higher collection efficiency allows to investigate deeper regions in the tissue under analysis [34]. SPADs with smaller area (<200 µm) can be used in high performing arrays, together with counting or timing digital electronics [35], where the ultimate goal is high photon detection performance. If instead extreme pixel density and pixel count are a must, other more scaled technology should be investigated, probably at the expenses of poorer overall performances.

Funding

This work was supported by the 'MiSPiA' project, under the ICT theme of the EC 7th Framework Program (FP7, 2007–2013) [grant number 257646].

References

[1] Michalet, X.; Colyer, R.A.; Scalia, G.; Ingargiola, A.; Lin, R.Millaud, J.E.; Weiss, S.; Siegmund, O.H.; Tremsin, A.S.; Vallerga, J.V.; Cheng, A.; Levi, M.; Aharoni, D.; Arisaka, K.; Villa, F.; Guerrieri, F.; Panzeri, F.; Rech, I.; Gulinatti, A.; Zappa, F.; Ghioni, M.; Cova, S. *Philos. Trans. R Soc., B* **2012,** *368,* 20120035.

[2] Hogan, H. *Biophotonics Int.* **2007,** *14*, 54–55.

[3] Cova, S.; Longoni, A.; Adreoni, A.; Cubeddu, R. *IEEE J. Quantum Electron.* **1983,** *19*, 630–634.

[4] Dalla, Mora; Tosi, A.; Zappa, A.; Cova, F.; Contini, S.; Pifferi, A, D.; et al. *IEEE J. Sel. Top. Quantum Electron.* **2010,** *16*, 1023–1030.

[5] Bethea, C.; Levine, B.; Marchut, L.; Mattera, V.; Peticolas, L. *Electron. Lett.* **1985,** *22*, 302–303.

[6] Townsend, P.D.; Rarity, J.G.; Tapster, P.R. *Electron. Lett.* **1993,** *29*, 634–635.

[7] Rarity, J.; Tapster, P. *Phys. Rev. Lett.* **1990,** *64*, 2495–2498.

[8] Nightingale, N. *Exp. Astron.* **1991,** *1*, 407–422.

[9] Stellari, F.; Tosi, A.; Zappa, F.; Cova, S. *IEEE Trans. Instrum. Meas.* **2004,** *53*, 163–169.

[10] Brida, G.; Degiovanni, I.P.; Genovese, M.; Piacentini, F.; Traina, P.; Della Frera, A.; et al. Appl. Phys. Lett. **2012,** 101, 221112.

[11] Gulinatti, A.; Rech, I.; Panzeri, F.; Cammi, C.; Maccagnani, P. Ghioni. M.; Cova, S. *J. Mod. Opt.* **2012,** *59*, 1489–1499.

[12] Michalet, X.; Cheng, A.; Antelman, J.; Suyama, M.; Arisaka, K.; Weiss, S. Proc. SPIE **2008,** 6862, 68620F.

[13] Barbieri, C.; Naletto, G.; Capraro, I.; Occhipinti, T.; Verroi, E.; Zoccarato, E.; et al. Proc. SPIE **2010,** 7681, 768110.

[14] Moscatelli, F.; Marisaldi, M.; Maccagnani, P.; Labanti, C.; Fuschino, F.; Prest, F. *Nucl. Instrum. Methods Phys. Res., Sect. A* **2013,** *711*, 65–72.

[15] Pifferi, A.; Torricelli, A.; Spinelli, L.; Contini, D.; Cubeddu, R.; Martelli, F.; et al. *Phys. Rev. Lett.* **2008,** 100, 138101.

[16] Zappa, F.; Tisa, S.; Tosi, A.; Cova, S. *Sens. Actuators, A* **2007,** *140*, 103–112.

[17] Tisa, S.; Zappa, F.; Tosi, A.; Cova, S. *Sens. Actuators, A* **2007,** *140*, 113–122.

[18] Richardson, J.A.; Grant, L.A.; Henderson, R.K. *IEEE Photonics Technol. Lett.* **2009,** *21*, 1020–1022.

[19] Webster, E.A.G.; Richardson, J.A.; Grant, L.A.; Renshaw, D.; Henderson, R.K. *IEEE Electron Device Lett.* **2012,** *33*, 694–696.

[20] Gersbach, M.; Richardson, J.; Mazaleyrat, E.; Hardillier, S.; Niclass, C.; Henderson, R.; Grant, L.; Charbon, E.J. *Solid-State Electron.* **2009,** *53*, 803–808.

[21] Bronzi, D.; Tisa, S.; Villa, F.; Tosi, A.; Zappa, F. *IEEE Photonics Technol. Lett.* **2013,** *25*, 776–779.

[22] Zappa, F.; Tosi, A. Dalla Mora, A.; Tisa, S. *Sens. Actuators, A* **2009,** *152*, 197–204.

[23] Assanelli, M.; Ingargiola, A.; Rech, I.; Gulinatti, A.; Ghioni, M. *IEEE J. Quantum Electron*. **2011,** *47*, 151–159.

[24] Ghioni, M.; Cova, S.; Zappa, F.; Samori, C. *Rev. Sci. Instrum.* **1996,** *67*, 3440–3448.

[25] Ghioni, M.; Gulinatti, A.; Maccagni, P.; Rech, I.; Cova, S. *Proc. SPIE* **2006,** 6372, 63720R.

[26] Cova, S.; Lacaita, A.; Ripamonti, G. *IEEE Electron Device Lett.* **1991,** *12*, 685–687.

[27] Tisa, S.; Guerrieri, F.; Tosi, A.; Zappa, F. In Proceeding of the Solid-State Device Research Conference, Edinburgh, Scotland, United Kingdom, Sept 15–19, 2008, **2008;** pp 274–277.

[28] Niclass, C.; Sergio, M.; Charbon, E. *Proc. SPIE* **2006,** *6372*, 63720S.

[29] Stoppa, D.; Mosconi, D.; Pancheri, L.; Gonzo, L. *IEEE Sens. J.* **2009,** *9*, 1084–1090.

[30] Becker, W. *Advanced Time-Correlated Single Photon Counting Techniques*; Springer, Berlin, 2005.

[31] Lacaita, A.; Zappa, F.; Bigliardi, S.; Manfredi, M. *IEEE Trans. Electron Devices* **1993,** *40*, 577–582.

[32] Benetti, M.; Popleteeva, M.; Dalla Betta, G.; Pancheri, L.; Stoppa, D. In Proceedings of the 7th Conference on Ph.D. Research in Microelectronics and Electronics, PRIME 2011; Madonna di Campiglio, Trento, Italy, July 3–7, 2011, **2011;** pp 185–188.

[33] Guerrieri, F.; Tisa, S.; Tosi, A.; Zappa, F. *IEEE Photonics J.* **2010,** *2*, 759–774.

[34] Pifferi, A.; Torricelli, A.; Spinelli, L.; Contini, D.; Cubeddu, R. et al. *Phys. Rev. Lett.* **2008,** 100, 138101.

[35] Villa, F.; Bronzi, D.; Bellisai, S.; Boso, G.; Bahgat Shehata, A.; Scarcella, C.; Tosi, A.; Zappa, F.; Tisa, S.; Durini, D.; Weyers, S.; Brockherde, W. *Proc. SPIE* **2012,** *8542*, 85420G.