A 130nm CMOS Single Photon Avalanche Diode

Cristiano Niclass, Marek Gersbach, Robert Henderson*, Lindsay Grant**, and Edoardo Charbon
Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
*University of Edinburgh, Edinburgh, Scotland
**ST Microelectronics, Edinburgh, Scotland

ABSTRACT

We report on the first implementation of a single photon avalanche diode (SPAD) in 130nm CMOS technology. The SPAD is fabricated as p+/nwell junction with octagonal shape. A guard ring of p-well around the p+ anode is used to prevent premature discharge. To investigate the dynamics of the new device, both active and passive quenching methods have been used. Single photon detection is achieved by sensing the avalanche using a fast comparator. The SPAD exhibits a maximum photon detection probability of 41% and a typical dark count rate of 100kHz at room temperature. Thanks to its timing resolution of 144ps (FWHM), the SPAD has several uses in disparate disciplines, including medical imaging, 3D vision, bio-photonics, low-light-illumination imaging, etc.

Keywords: CMOS single photon detector, single photon avalanche diode, SPAD, Geiger mode of operation.

1. INTRODUCTION

The design of avalanche photodiodes in DSM CMOS technology involves additional challenges than in larger feature size technologies. In order to operate in the so-called Geiger mode, a SPAD requires a design configuration that supports a planar and uniform multiplication region extending laterally and vertically underneath the area of the SPAD as much as possible [1]. Even though this requirement is mandatory to allow the creation of a reasonably large photosensitive or active area, it is not sufficient in general. For example, reference [2] reports the design of a SPAD fabricated in 0.18μm CMOS that implements a planar multiplication region, according to simulations, but exhibits DCR levels of 1MHz, unacceptable in most applications.

Noise performance becomes a major issue for SPADs in deep-submicron CMOS technologies. It is therefore very important to keep a strict notion of noise performance when assessing potential design structures. The main sources of noise in SPADs are more significant in deep-submicron due to (i) higher doping levels, (ii) reduced annealing and drive-in diffusion steps, and (iii) the presence of shallow-trench isolation (STI).

Higher doping levels increase the effects of tunneling-induced dark counts and increase the parasitic capacitance. The increase of parasitic capacitance increases the number of carriers involved in an avalanche discharge and thus worsens afterpulsing probability.

Driven by miniaturization, state-of-the-art fabrication processes reduce the strength and duration of annealing and drive-in diffusion steps to a minimum. The lack of effective annealing steps increases the concentration of impurities that introduce carrier recombination-generation and trapping centers, thus worsening both thermally-generated dark counts and afterpulsing effects [1].

At and below the 0.25μm mark, standard CMOS processes feature STI compulsorily. It is known that STI may dramatically increase the density of deep-level carrier generation centers at its interface [3],[4]. When a STI is close to or in contact with the multiplication region of a SPAD, such as in [2], one can expect high dark count rates.

Unfortunately, very often designers do not have enough flexibility to change or adapt a process parameter in order to better fit the SPAD requirements in CMOS technology. In order to address the issues described above, designers are left with a number of design layers, models and rules. It was the aim of this work to design, test, and characterize high-quality SPADs based on an existing and fixed 130nm CMOS technology. This approach is obviously beneficial in terms design time and fabrication costs.

2. SINGLE PHOTON AVALANCHE DIODES

A SPAD is generally implemented as a pn junction biased *above* breakdown. In this regime of operation, known as Geiger mode, photo-generated carriers may cause an avalanche by impact ionization. The number of carriers generated as a result of the absorption of a single photon determines the optical gain of the device, which in the case of SPADs may be virtually infinite.

An avalanche in the multiplication region causes a current pulse of appreciable amplitude but it needs to be stopped. This is generally accomplished via a *quenching circuit*. The avalanche current pulse may be converted into a digital voltage pulse by proper design techniques, thus enabling the direct conversion of photons onto digital signals compatible with CMOS low-voltage circuitries. There exist several types of quenching circuits, divided in two main categories: active quenching and passive quenching. In active quenching, the avalanche is sensed and a feedback circuit provides a mechanism to force the reverse bias of the pn junction below breakdown. The same circuit is generally used to actively recharge the device to its initial state, above breakdown, so as to enable the next detection cycle. In passive quenching, the avalanche current is used to directly act on the reverse bias voltage by lowering it towards breakdown voltage, which eventually quenches the current. If this is achieved, for example, using a resistance in series to the photodiode, the effective capacitance of the junction must be passively recharged through the quenching resistance. In SPADs, the detection cycle requires a total time known as *dead time*, which includes quenching and recharge. The dead time is also responsible for the upper limit of photon flux detectable by a SPAD.

Noise performance of SPADs is mainly characterized by spurious pulses in the dark, i. e. *dark counts*. Dark counts, quantified in terms of a frequency, or dark count rate (DCR), are caused by thermally or tunneling generated carriers [1]. The relative impact of the two effects can generally be evidenced with device analysis as a function of temperature. DCR is also strongly dependent upon the *excess bias voltage*, i.e. the voltage in excess of breakdown at which the SPAD is biased.

The sensitivity is characterized in SPADs as the probability of a photon impinging the device's surface to cause a pulse. It is known as *photon detection probability* (PDP) and it is strong function of photon wavelength and excess bias voltage.

The uncertainty of the time delay between photon impingement and the leading edge of the pulse generated by the sensor is known in the literature as timing resolution or *timing jitter*. In a small SPAD, the timing jitter mainly depends on the time a photogenerated carrier requires to be swept out of the absorption zone into the multiplication region. In large devices, timing jitter is also caused due to the fluctuations of the avalanche propagation across the active area.

Trapping centers in the multiplication region tend to capture carriers generated during an avalanche. As trapping centers are characterized by finite lifetimes, trapped carriers are released at a random later time, thus potentially re-triggering a subsequent avalanche [1]. Such phenomenon causes so-called *afterpulses*, i.e. spurious pulses correlated to previous Geiger pulses. The parameter characterizing this effect is known as *afterpulsing probability*, or probability of afterpulsing, and it is also function of the number of carriers involved in an avalanche, which in turn depends on the SPADs parasitic capacitance. In addition to the correlated noise introduced by afterpulsing, this phenomenon may limit the maximum rate of detectable photons as one photon may generate in average more than a single event.

3. DEVICE STRUCTURE

Fig. 1 shows the cross-section of the proposed SPAD. It consists of a p+ anode within an n-well cathode where p+ and n-well are the implantations of source/drain and bulk, respectively, of standard 1.2V PMOS transistors. This configuration allows for a full isolation of the p+ anode from the p-substrate. In addition, the configuration enables coupling relatively high bias voltages necessary in SPADs to low-voltage CMOS logic, similarly to [5] and [6].

The planar multiplication region was enabled by means of a p-well guard ring [5], where p-well is the bulk of isolated 1.2V NMOS transistors. A useful feature of this technology is the availability of a buried n-type isolation layer that allows for a full isolation of p-well within n-well from p-substrate. This layer was used to prevent a punch-through of the

p-well guard-ring to p-substrate. The combination of n-well and buried n-isolation layer was the lowest doping concentration feasible in this technology for the cathode.
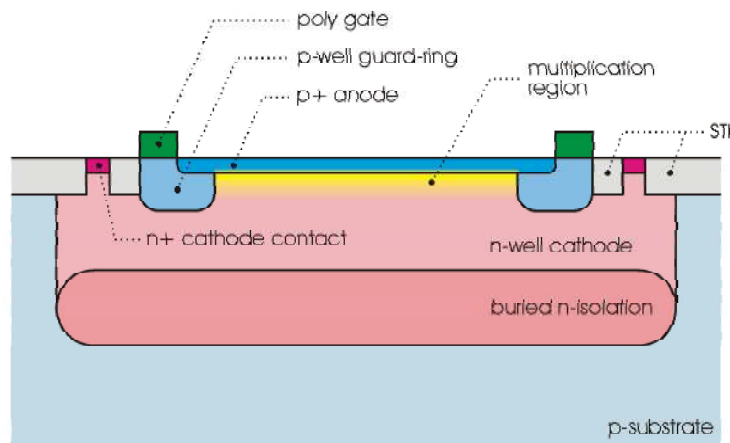

Fig. 1. SPAD cross-section: p+ anode within n-well cathode (not to scale).

A major improvement in this design is the physical separation of the STI interface from the SPAD multiplication region, thus having a beneficial impact on DCR. In standard DSM CMOS, it is not possible to prevent STI by means of a drawn layer. As a general rule, STI is etched everywhere so that all the p+ and n+ implantations are surrounded by STI to improve isolation. It is possible however to draw a polysilicon gate of a standard transistor that represents a stop mask for n+ and p+ implantations. STI can therefore be effectively separated from the surrounding of the anode by drawing a superposition of polysilicon, thin-gate-oxide, p+, and diffusion layers around the p+ anode. In order to prevent the formation of a high-electric field within the thin-gate-oxide layer, the polysilicon gate is kept at the same potential as the p+ and p-well layers, ensured by ohmic contacts.

Since the polysilicon gate prevents the p+ to be implanted, the result of the fabrication process is a p-well extension of p+ completely free of STI, whose extension can be adjusted as desired. Around the p-well guard-ring, there is still a STI ring. This STI interface, in particular at the depletion region between the p-well guard-ring and n-well cathode, may induce a large density of generation centers. Nonetheless, the p-well guard-ring lowers the electric field around the SPAD sufficiently to prevent impact ionization but it is enough to collect most of the carriers generated at the STI/p-well interface. As a result, this structure allows a small parasitic current to flow from cathode to anode without triggering avalanche events, thus reducing DCR.

4. EXPERIMENTAL RESULTS

The photomicrograph of the proposed device is shown in Fig. 2. The structures visible in the figure include the octagonal anode, guard ring, and metal interconnect. The additional function of the metal is that of preventing the guard ring to be exposed to light for characterization purposes. The anode measures 10μm in the picture. 30μm structures were also integrated in the same technology for characterization purposes.
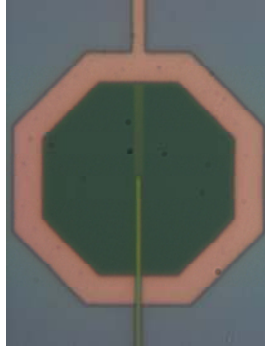
Fig. 2. Photomicrograph of the SPAD structure.

The diode was tested in a number of ways. First, the I-V characteristic was measured statically using a standard semiconductor analyzer. Fig. 3 shows the I-V characteristics of the diode in reverse bias. The picture shows that the reverse current close to breakdown voltage approaches 600pA. This relatively large current would suggest that DCR tends to be high. For instance, if we suppose that all the carriers were collected by the multiplication region, the device would not properly operate in Geiger mode as its DCR would be of the order of 3-4 GHz. In this section, it will be shown that the structure properly operates in Geiger mode and exhibits acceptable levels of DCR. As described in Section 3, most of the reverse current is expected to be generated at the periphery of the SPAD, at the STI/p-well/n-well interface, where impact ionization is prevented by the p-well guard-ring.
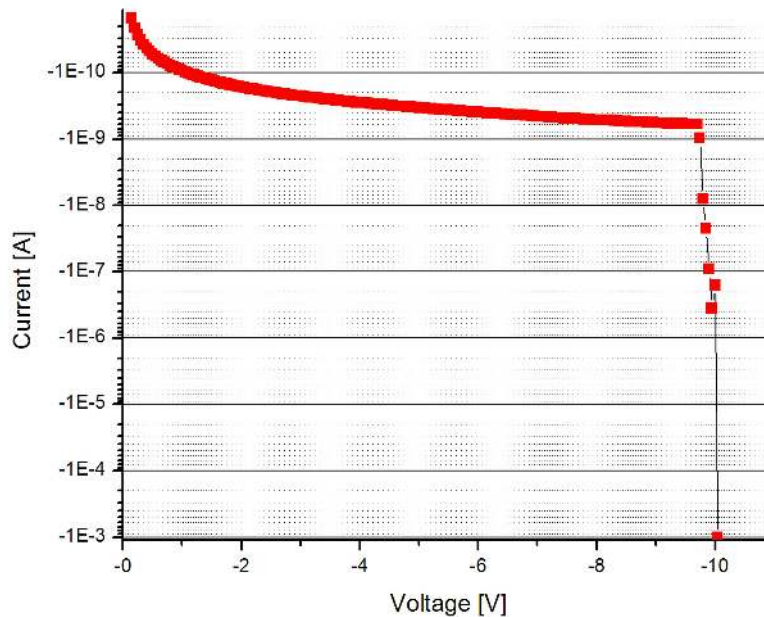

Fig. 3. I-V Characteristic of proposed structure at room temperature.

The diode was operated in Geiger mode using both passive and active quenching circuitries. The schematic setup of the passive quenching configuration is shown in Fig. 4.
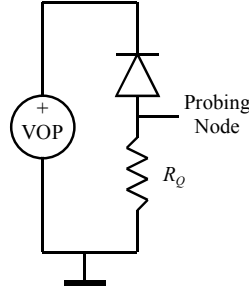
Fig. 4. Passive quenching setup. A probe senses the voltage across quenching resistor R_Q.

The 20kΩ quenching resistance R_Q, placed at the anode of the pn junction, causes an increase of its potential in case of avalanche. If the reverse bias voltage across the junction decreases towards breakdown voltage, the avalanche current is reduced to a level in the order of tens of microamperes and eventually stops. Avalanche quenching is followed by an exponential recharge to allow the voltage across the junction to return to its initial value of *VOP*. This voltage satisfies the following condition

$$VOP = |V_{BD}| + V_e,$$ (1)

where V_{BD}, and V_e are the breakdown and excess bias voltage, respectively.

The plot of Fig. 5 shows the recharge phase of the probed voltage as a function of time for different values of *VOP*. The simple exponential behavior is due to the RC recharge. R accounts for the resistive path to ground and C for the overall capacitance at the probing node.
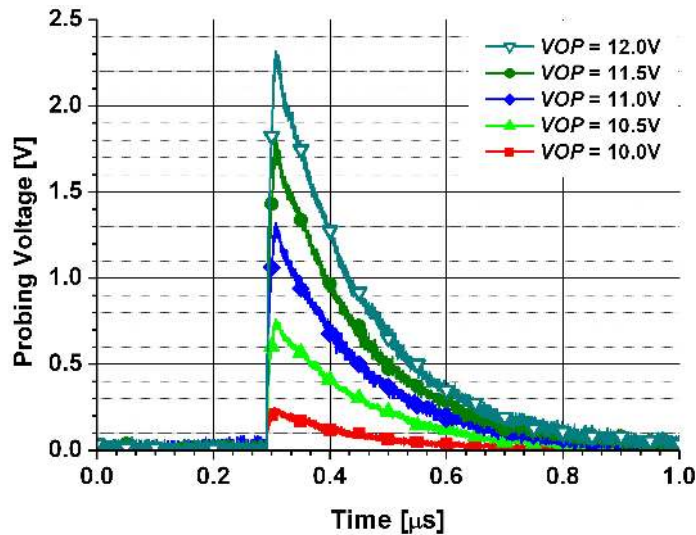

Fig. 5. Plot of the time-dependent output of the probe voltage as a function of time for several values of *VOP*.

Due to the fact that this device does not have integrated quenching circuitry, the term C is dominated by the parasitic capacitance of external components. It has been estimated to be 10pF, a factor 70 to 100 larger than the expected SPAD junction capacitance. The dead-time under this condition is estimated to be 450ns. As described in Section 2, afterpulsing probability depends independently on dead time, due to trap lifetimes, and on the parasitic capacitance as it increases the number of carries traversing the multiplication region, thus filling up traps. Thus, a characterization of afterpulsing probability under this condition is irrelevant, since it gives no insight on the true potential of the device and of its internal

capacitance when the SPAD is monolithically integrated with its quenching and recharge circuit. We assume that afterpulsing characterization under the present condition would be incorrect and thus irrelevant.

In order to precisely investigate DCR and PDP in this work independently of dead time and afterpulsing effects, an alternative setup involving the use of an external gated active recharge circuit combined with TCSPC was used. This technique is often used in the characterization of III-V SPADs, which exhibit significantly higher DCR and afterpulsing effects [8]. In most active quenching and recharge setups, an active circuit replaces the quenching resistance, thus allowing one to reduce the recharge time to a few tens of nanoseconds. Our experimental setup is based on a commercially available gated active recharge circuit [7] and is described as follows.

VOP is maintained below V_{BD} at the beginning of each event measurement cycle. *VOP* is then quickly increased to its nominal level, according to Equation (1), so as to recharge the SPAD. The time interval between the SPAD recharge signal and the moment a first Geiger event occurs is measured using a high precision time-to-digital converter (TDC). *VOP* is subsequently kept below V_{BD} during a hold-off time of the order of 500µs. This hold-off time is chosen large to prevent any afterpulse. As this measurement cycle is repeated a large number of times, a histogram is built conforming to the TCSPC technique. The resulting histogram shows an exponential decay similarly to a typical florescence lifetime measurement. The inverse of the mean value of the histogram provides the desired counting rate. Any timing offset between full SPAD recharge and Geiger pulse leading edge is removed prior to computing the counting rate. The active recharge circuit conveniently performs fast active recharge and also provides a trigger signal used to compute time-interval measurements as described above. As detector dead time and afterpulsing do not impair the measurement even at high counting rates, this technique is used to measure DCR as well as PDP.

In order to correctly characterize the measurements presented hereafter, the breakdown voltage was firstly measured for the structure as a function of temperature. Hence, *VOP* was set for a given temperature to reflect the correct excess bias voltage according to Equation (1). Fig. 6 shows a plot of the breakdown voltage as a function of temperature.
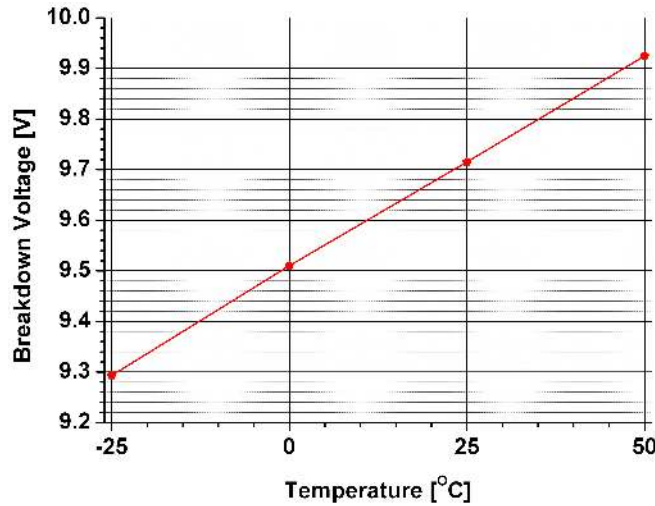

Fig. 6. Breakdown voltage vs. temperature.

The PDP was measured for two excess bias voltages for the entire spectrum of interest (350–1000nm). Fig. 7 shows a plot of the PDP at room temperature. PDP outperformed our expectations as measurements showed values in the range of previously reported SPADs in near-micron CMOS technologies [5],[6], whose multiplication regions are wider and deeper. The result of a shallower multiplication region was a shift of the maximum of detection probability from 550nm in [5] to 450nm in this work. We believe that this relatively good PDP performance partially resulted from the use of enhanced dielectrics for optical detection available in this imaging CMOS technology. In Fig. 2, it is possible to notice a darker region in the middle of the SPAD, where optimized dielectrics was used, if compared to the remaining area of the picture, where only partial optimization was used. This darker region suggests that the light reflection coefficient at the

center of the SPAD was noticeably lower than it would have been if we utilized non optimized passivation layers. Notice that, as DCR measurements were performed previously to the PDP characterization, the mean value of DCR contribution was suppressed from each counting rate used in the measurement of PDP. As a result, DCR is not responsible for an artificially increased PDP.
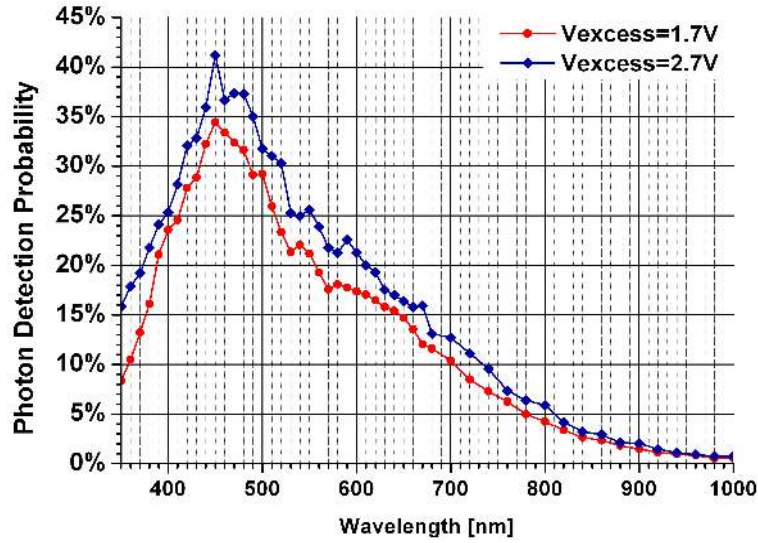

Fig. 7. Photon detection probability (PDP) as a function of wavelength for two values of excess bias voltage.

In near-micron CMOS SPAD implementations, DCR can be as low as a few tens of Hertz [5],[6] and it is a strong function of temperature and of excess bias voltage.
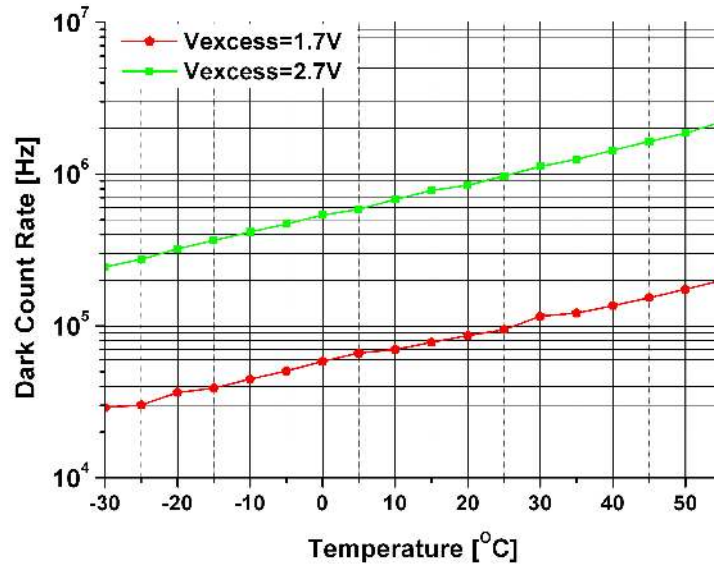

Fig. 8. DCR dependence on temperature.

Fig. 8 shows a plot of DCR as a function of temperature for two excess bias voltages, measured using the TCSPC method as described above. Besides its higher absolute values if compared to [5], DCR also exhibits weaker dependence on temperature. This suggests that DCR has a non negligible tunneling contribution [9]. This behavior was expected due to relatively higher doping levels of both p+ and n-well layers available in this CMOS technology. Fig. 9 shows a plot of DCR as a function of V_e for four different temperatures based on the TCSPC measurement. It also shows a curve of DCR

measured using the passive quenching setup of Fig. 4 for a temperature of 25°C. Since the measurements based on the passive quenching setup strongly saturate due to dead time, the errors in those measurements compared to the TCSPC method are significant for any DCR higher than a few tens of kHz.

As can be seen in Fig. 9, DCR reaches prohibitive levels as V_e is chosen higher than 2V. Depending on the amount of parasitic light in a given application, higher levels of DCR may be tolerated. For instance, noise in a 3D image sensor, based on the time-of-flight principle [6] is in general dominated by the parasitic background light when it operates outdoor. In such cases, in order to improve PDP and increase overall signal-to-noise ratio, higher values of V_e may be recommended.
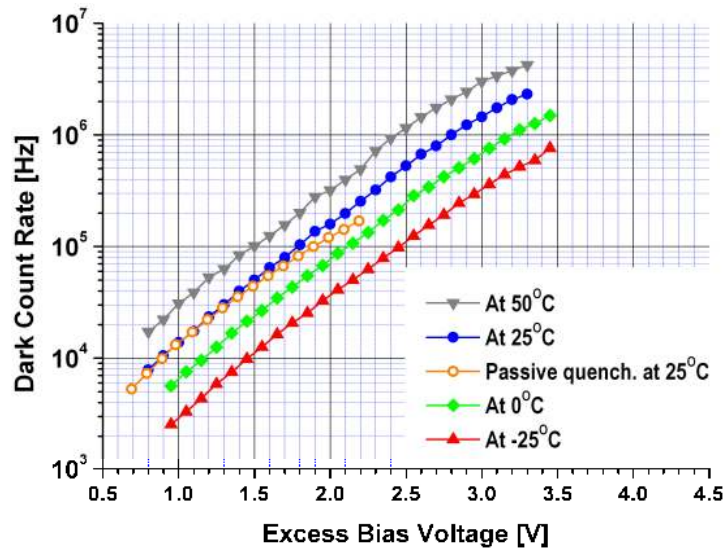

Fig. 9. Dark count rate (DCR) as a function of excess bias voltage at four temperatures. Temperatures are set with an uncertainty of +/- 2°C.
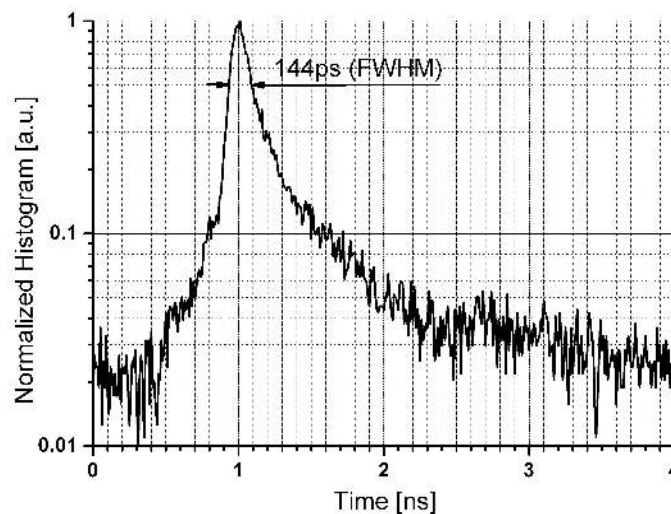

Fig. 10. Histogram of the response of SPADs to a low jitter light impulse.

Timing jitter was characterized in this work also based on the TCSPC technique. A fast laser source with pulse width of 40ps and repetition rate of 40MHz emitting a beam with a wavelength of 637nm was used to illuminate the SPAD. The

time interval between the laser output trigger and the leading edge of the SPAD signal, operated with the active recharge circuit, was measured via a high performance oscilloscope operating as a TDC. The oscilloscope, a LeCroy 8600A, features 20GS/s and 3ps of uncertainty. A histogram was built as the time interval measurements were repeated over very large number of times. In order to prevent the typical pile-up effect, optical neutral density filters were used to reduce the SPAD firing rate to a few tens of kHz. The resulting jitter is reported in the normalized histogram shown in Fig. 10. The FWHM value of the resulting pulse was 144ps. This timing uncertainty includes the uncertainty of the hybrid active recharge setup as well as the laser pulse width. We believe that lower timing uncertainties can be expected when the SPAD will be monolithically integrated with its front-end circuit.

5. CONCLUSIONS

The first single photon avalanche diode implemented in 130nm CMOS technology is reported. Techniques to fabricate the device using available layers within standard design rules are described in detail. The characterization of the device yielded photon detection probabilities similar to other CMOS single photon detectors found in the literature. The dark count rate and timing jitter of this device have also been measured at various operating conditions. In the future, arrays of this device will be monolithically integrated with front-end and application circuits, to be used in a number of applications requiring high dynamic range and timing resolution.

ACKNOWLEDGEMENTS

REFERENCES

[1] S. Cova, M. Ghioni, F. Zappa, I. Rech and A. Gulinatti, "A View on Progress of Silicon Single Photon Avalanche Diodes", *Advanced Photon Counting Techniques*, Proc. of SPIE Vol. 6372 63720I-1, Oct. 2006.

[2] H. Finkelstein, M. J. Hsu, and S. C. Esener, "STI-Bounded Single-Photon Avalanche Diode in a Deep-Submicrometer CMOS Technology", *IEEE Electron Device Letters*, Vol. 27, No. 11, November 2006.

[3] T. Mamamoto, "Sidewall damage in a silicon substrate caused by trench etching", Applied Physics Letters, vol. 58, no. 25, 24, pp 2942-4 June 1991.

[4] H. I. Kwon, M. I. Kang, B.-G. Park, J. D. Lee, S. S. Park, "The analysis of dark signals in the CMOS APS imagers from the characterization of test structures", *IEEE Transactions on Electron Devices*, vol. 51, no. 2, pp. 178-184, February, 2004.

[5] A. Rochas, M. Gani, B. Furrer, P. A. Besse, R. S. Popovic, G. Ribordy, and N. Gisin, "Single photon detector fabricated in a complementary metal–oxide–semiconductor high-voltage technology", *Review of Scientific Instruments*, vol. 74, no. 7, pp. 3263-3270, July 2003.

[6] C. Niclass A. Rochas, P.A. Besse, and E. Charbon, "Design and Characterization of a CMOS 3-D Image Sensor Based on Single Photon Avalanche Diodes", *IEEE J. of Solid-State Circuits*, Vol. 40, N. 9, pp. 1847-1854, Sep. 2005.

[7] IdQuantique, www.idquantique.ch.

[8] A. Lacaita, F. Zappa, S. Cova, and P. Lovati, "Single-photon detection beyond 1μm: performance of commercially available InGaAs@InP detectors", Applied Optics, vol. 35, no. 16, June 1996.

[9] A. Rochas, "Single Photon Avalanche Diodes in CMOS technology", Ph.D. Thesis, EPF-Lausanne, 2003.