

Limitation of SPADs quantum efficiency due to the dopants concentration gradient.

Juan Manuel López-Martínez, Ricardo Carmona-Galán and Ángel Rodríguez-Vázquez

Instituto de Microelectrónica de Sevilla-CNM
CSIC-Universidad de Sevilla
Sevilla, Spain
jmlopez@imse-cnm.csic.es

Abstract— Single-photon avalanche diodes are highly sensitive devices capable of registering the arrival of an individual photon. They are biased beyond breakdown voltage, and thus the electron-hole pairs generated by any incident photon is accelerated by the strong electric field triggering an avalanche current. This paper examines the role of the dopants concentration gradient in the gathering of photons in these devices, and how it can be engineered to maximize quantum efficiency and explain his role to minimize undesirable effects like crosstalk.

Keywords—single-photon avalanche diode (SPAD); device simulation; TCAD simulation; quantum efficiency.

I. INTRODUCTION

Single-photon avalanche diodes (SPADs) are semiconductor devices with the ability of registering the incidence of individual photons. They offer higher photon detection efficiency, especially in the red and near infrared spectrum than photomultiplier tubes, which SPADs were thought to replace. SPADs can be employed either to sense low-intensity luminous signals, by photon counting, or to timestamp the arrival these photons, in order to estimate their time-of-flight (ToF). Nowadays they can be built in CMOS technology, what allows incorporating functionality and opens the door to many more applications. Apart from depth and proximity sensors, SPADs are finding suitable application scenarios in fluorescence lifetime microscopy (FLIM) [1] and fluorescence correlation spectroscopy [2].

When operational, SPADs are reverse biased above its breakdown voltage. In this condition they can maintain a delicate state of equilibrium without triggering an avalanche, that is Geiger mode. In this state, any charge carrier that reach to or is created within the depletion region, would potentially trigger an avalanche, which would be registered by the circuitry. In the last two decades, much effort has been made by investigators to refine these devices, in order to improving their characteristics, for instance, their quantum efficiency, dark counts and fill factor. To that end they have been reshaped over the years. One the main drives for this have been the search of a suitable guard ring that can either avoid early edge-breakdown, that would limit greatly the active area size of the SPAD, and scale down these devices to maximize fill factor. One of the first solutions to this problem was to implant a low doped well in the periphery of the active area region to avoid its periphery to prematurely enter breakdown (Fig. 1). However, the problem

with this configuration is that there is limit for scaling it down, as the depletion regions around this low doped wells would merge, depleting of carriers the active area and disabling the possibility of triggering avalanches [3]. Another solution to this problem has been the use of shallow trench isolation (STI). However, in this configuration the junction would reach the silicon-silicon oxide boundary, making surface traps skyrocket dark counts to megahertz and thus making it inviable for most applications [4]. The most modern form of guard ring is that of the virtual guard ring. This configuration involves avoiding a deep n-well to fully form near the surface, making the well less doped as the surface gets closer (retrograde deep n-well). Then a P-Well is implanted in the center of the device making room at the periphery so that a P-Well/Retrograde Deep-N Well junction can be formed [5]. This junction would have then a strong electric field in the active area that will weaken as its get closer to the surface, increasing the lateral breakdown voltage gradually and forming a de-facto guard ring (Fig. 2). This concept also has been used in [6] with the growth of p-epitaxial layer over a p-substrate to form a guard ring between the p-epitaxial layer and a deep N-well (Fig. 3). Both solutions offer the possibility of scaling down further the SPADs and the possibility of enhancing the detection of the longer wavelengths, as the junctions are located deeper in the bulk. Then, the gain in quantum efficiency in the last two configurations at those

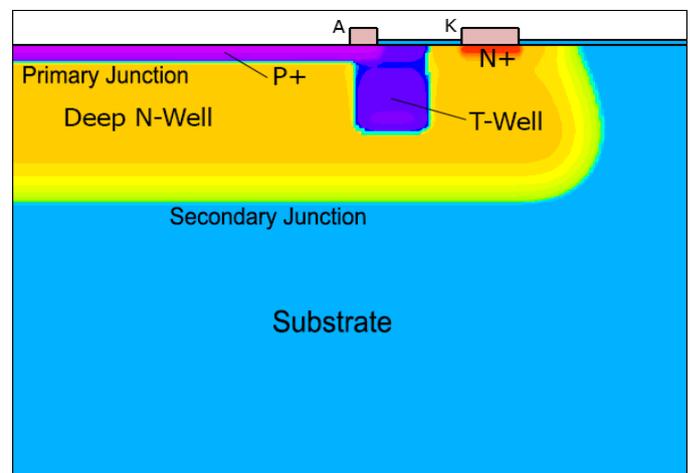


Fig. 1. Doping profile of a P+/DNW SPAD done by TCAD simulations. The left border acts as axis of revolution.

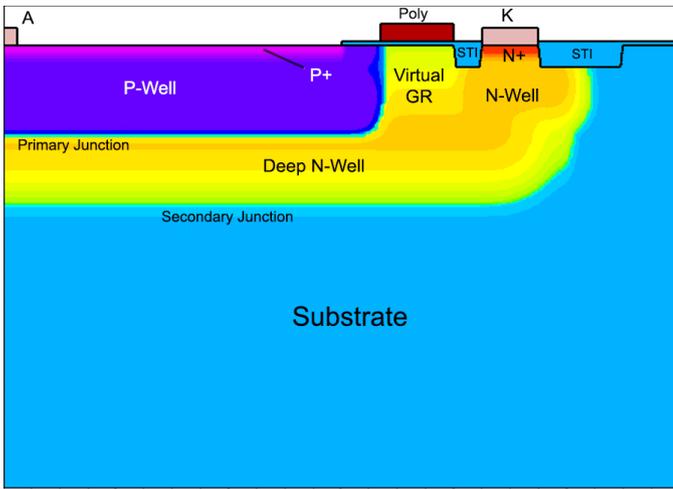


Fig. 2. Doping profile of a P-Well/DNW SPAD by TCAD simulations. The left border acts as axis of revolution.

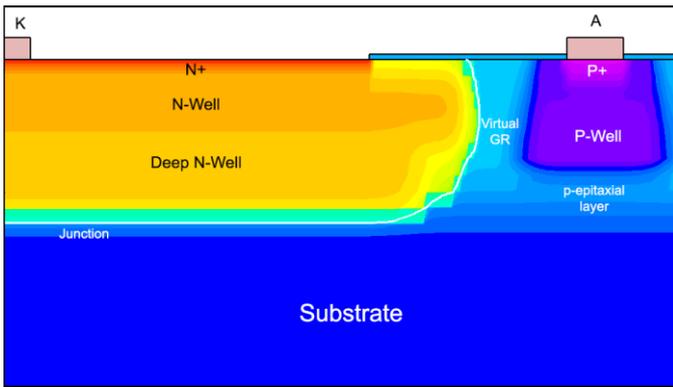


Fig. 3. Doping profile of a DNW/PSub SPAD. The white line represents the junction. The left border acts as axis of revolution.

wavelengths should be considerable, given the difference in junction depth, however it is not.

This paper will examine the quantum efficiency (QE) of these devices through TCAD simulations, and will show how the doping concentration gradient limits quantum efficiency and protect devices from crosstalk, as well as possible solutions.

II. TCAD SIMULATION

The devices selected for this work were the P+/DNW (Fig. 1), the P-Well/DNW (Fig. 2) and the DNW/PSub (Fig. 3) SPADs. To study these devices, a simulation fabrication process flow was carried out with Athena, SSuprem4 and Elite tools from Silvaco TCAD suit, using doping profiles of a standard CMOS 0.18 technology. Then process simulation has been carried out with Atlas as part of Silvaco TCAD suite. This simulation include carrier mobility models (ANALYTIC, FLD.MOB), recombination models (CONSRH, AUGER), carrier statistics models (FERMI, BGN) and impact ionization models (SELB), all necessary to describe the physical phenomena present in the operation of a SPAD. The simulations were intended to characterize quantum efficiency for different wavelengths, for a window of illumination that matched the active area, and to examine as well the physical properties of the bulk in Geiger mode.

III. EXPERIMENTAL RESULTS

First quantum efficiency was determined. As QE only depends on wavelength, it can be seen as the maximum photon detection probability (PDP):

$$PDP = QE(\lambda) \cdot P_{tr} \quad (1)$$

where P_{tr} is the avalanche triggering probability. To measure QE , all devices are illuminated with a uniform beam of a fixed intensity with variable wavelength. The current collected in the anode (I_A) is then divided by the source photo current (SPC), which is the total current that could generate an illumination source if every photon would generate and electron-hole pair. The result then would be the external quantum efficiency (EQE) which will be the one we are referring to from now on [7]:

$$EQE(\lambda) = \frac{|I_A|}{SPC(\lambda)} \quad (2)$$

The quantum efficiency for different SPAD configurations is presented in Fig. 1. As it can be seen the P+/DNW configuration presents a EQE maximum at a wavelength of $0.44 \mu\text{m}$, whereas the P-Well/DNW has its maximum at $0.48 \mu\text{m}$ and from there the quantum efficiency is only marginally higher than that of P+/DNW. At first sight the primary junction depth of the latter, around $1 \mu\text{m}$ deep, should have enabled a more differentiable maximum EQE shift over that of the former, which primary junction is only $0.2 \mu\text{m}$ deep. That shift is effectively observed with the DNW/PSub configuration, which presents a maximum at $0.64 \mu\text{m}$ with its junction in $1.5 \mu\text{m}$ deep. The reason for this results lays in the electric field profile of the SPAD when it is reverse biased above its breakdown region but has not trigger an avalanche yet.

Electron-hole pairs, able to originate an avalanche, can only be created by illumination in a determined zone of the SPAD as shown in Fig. 5. Here the module and direction of the electric field can be seen within a portion of active area region of the SPAD. The deep neutral region presents a weak electric field. That electric field is caused by the fact that the Deep N-Well

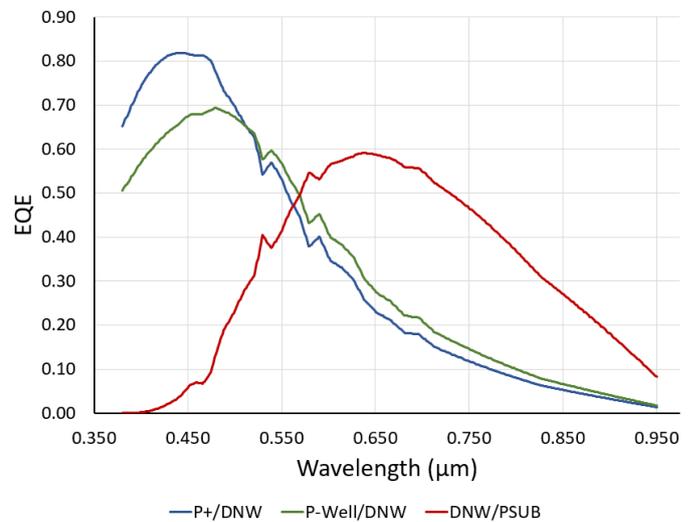


Fig. 4. External quantum efficiency obtained for the devices of this work.

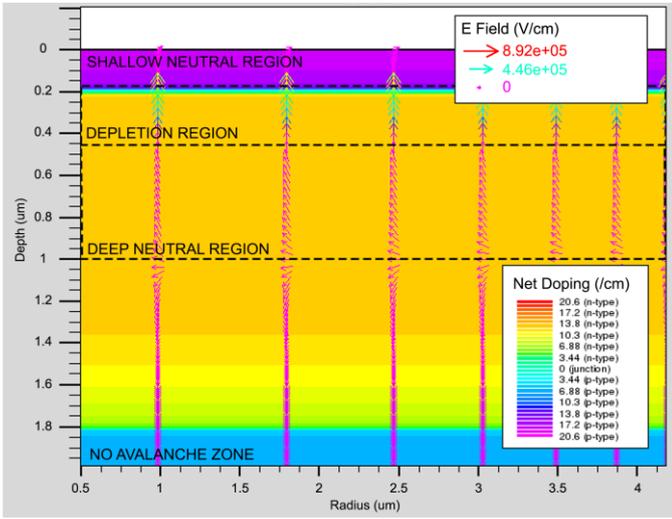


Fig. 5. Electric field profile of the active area region of the P+/DNW SPAD.

presents a gradient of doping concentration instead of a flat uniform doping. Thus, any pair of points within the Deep N-Well will present a difference in charge that will enable an electric field between them. This electric field will route charge carriers to or away from the depletion zone instead of random diffusion. Approximately at 1-micrometer-deep, the doping concentration becomes maximum, and consequently the vertical component of the electric field changes direction and carriers generated below that point are unable to reach the depletion region and trigger an avalanche, establishing in fact a sensing limit. On the other hand, any carrier created in any of the neutral regions may diffuse to the depletion region (electrons for the shallow region and holes for the deep region), accelerate until they reach the saturated drift velocity and trigger impact ionization which would result in an avalanche current.

This sensing limit restrict quantum efficiency, since the farther it is from the junction the greater the volume capable of gathering photogenerated charge carriers. Naturally, it can be used then to define a region within the SPAD that encompass the depletion region, and that is characterized by being the region where photon absorption may lead to avalanche triggering. That is, the detection region. Electron-hole pairs generated by photons absorbed outside this region cannot possibly trigger avalanches. Fig. 6 show the detection region for the three selected devices. Through it, we can explain the quantum efficiency from Fig. 4, but first we are going to define the current generated for incident photons. This current is given by [7]:

$$G = \eta_0 \frac{P\lambda}{hc} \alpha e^{-\alpha y} \quad (3)$$

$$\alpha = \frac{4\pi}{\lambda} k \quad (4)$$

being α the absorption coefficient, P is the ray intensity factor, which contains the cumulative effects of reflections, transmissions, and loss due to absorption over the ray path, η_0 is the internal quantum efficiency, which represents the number of carrier pairs generated per photon observed, y the depth of

absorption, and k is the imaginary part of the optical index of refraction or extinction coefficient. Equation (3) explains that the probability of a photon of being absorbed, decays exponentially with depth and the longer the wavelength the slower the decay. So shortest wavelengths will be absorbed almost entirely near the surface, whereas longest ones will be more uniformly absorbed through the bulk as the absorption probability decay less with depth.

The P+/DNW SPAD has a detection region ranging from the surface to 1 μm deep (Fig. 6-a), so is very good detecting short wavelengths. Then the probability to detect photons decays as the photons wavelength get larger, as the detection window become too small compared to the region of possible absorption. The detection region for the P-Well/DNW SPAD range from 0.65 μm to 1.3 μm (Fig. 6-b). As expected this configuration performs worse in gathering photons of short wavelengths, but the smaller window of detection makes it worse at detecting too medium wavelengths. Finally, its slightly better at detecting

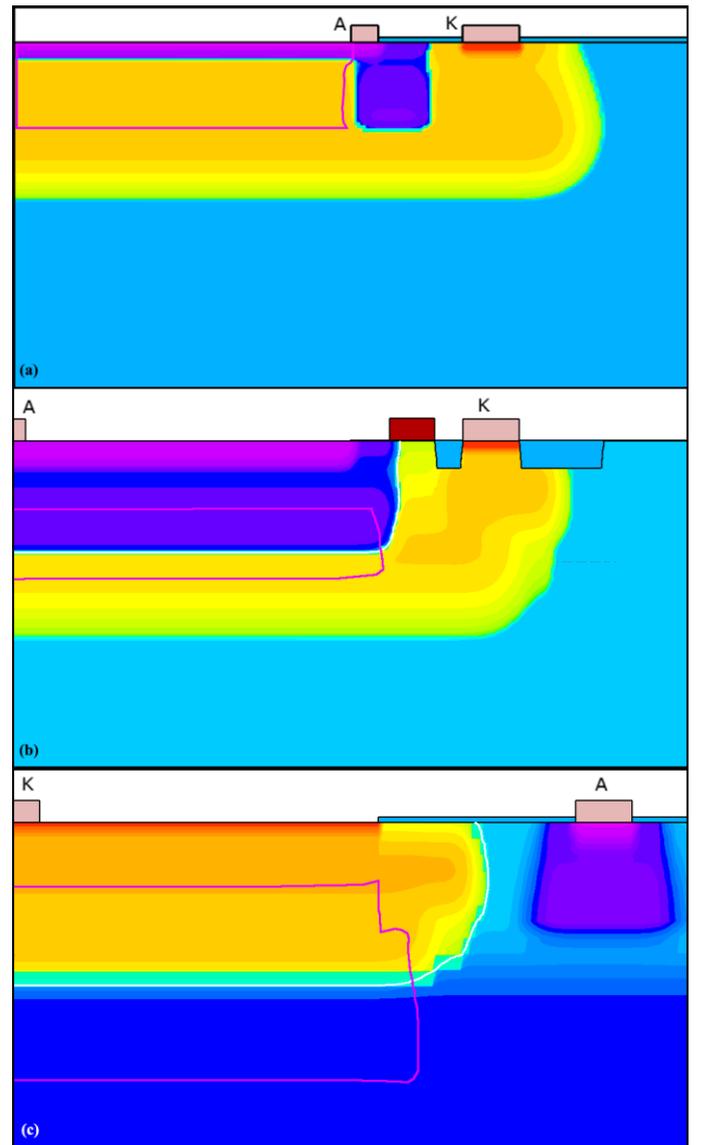


Fig. 6. Detection region for the P+/DNW (a), the P-Well/DNW (b) and the DNW/PSub SPADs.

longer wavelengths, because the deeper detection region. The last SPAD configuration has a detection region that range from 0.7 μm to 2.5 μm (Fig. 6-c). In this last case, the distance the charge carriers have to travel to the depletion region is high enough that the probability of recombining with the bulk is high enough. In fact, the deeper boundary for this region, that is well inside the substrate, is defined by this recombination length. As the substrate has a uniform doping concentration that starts to vary as it gets closer to the SPAD, the movement of charge carriers in generated near the deeper boundary is governed by random diffusion. This high probability of recombination explains too why is so low the probability of detecting short wavelengths. Instead, this device is excellent at detecting long wavelength photons due to its great detection region.

As said before, when a positive potential is applied on the cathode of any SPAD, they are reversed biased. Thus, electrons generated in the substrate by photons of longer wavelengths can diffuse from the substrate to the DNW to be collected by the cathode. Only in the DNW/PSub arrangement, these electrons reach the depletion region, triggering avalanches. As described before, in the case of the P+/DNW and P-Well/DNW SPADs, this detection volume is within the SPADs themselves. The electrons generated in the DNW and the substrate are simply attracted to the biased cathode, while only some holes generated within the detection region could be attracted by the anode to the depletion region and thus generate an avalanche. In other words, the electrons and the holes generated in the substrate by the illumination can never generate an avalanche in these devices, hence their low detection in long wavelengths. This also make them almost immune to electrical crosstalk, as charge carriers from other pixels cannot reach either the depletion region. This is also true for any side effect that can generate charge pairs, like illumination by radiative recombination of neighboring SPADs that are undergoing avalanche, also known as optical crosstalk. Photons generated in this way only could trigger avalanche if they are absorbed in the detection region. This make the P-Well/DNW photodiode especially well protected due to low cross-section of its detection region. The DNW/PSub SPAD however would suffer considerably this undesirable effect due to the larger cross-section of its detection region.

A solution to improve quantum efficiency in this devices would be to smooth the doping concentration gradient. In the P-Well/DNW SPAD for example, the P-Well impurities concentration could maintain a downward trend until it reaches the depletion region, generally improving quantum efficiency, especially in short wavelengths. This however would imply a

custom P-Well that not all the foundries offer and would make them prone to crosstalk, as the detection region grows larger. This solution is applicable to the other SPAD configurations.

IV. CONCLUSION

In this work the fabrication and processes of three different types of SPADs have been simulated by TCADs simulation. Thank to that a region of detection only in which the device can detect incoming photons have been defined. This region is determined through opposing electric fields which are originated by the difference of dopant concentration through the bulk. Thank to that have been possible to explain the differences in quantum efficiency of the three devices and improve the understanding of these devices.

ACKNOWLEDGEMENT

This work was funded by Junta de Andalucía's through grant TIC 2012-2338, the Spanish MINECO and European Region Development Fund (ERDF/FEDER) through Project RTI2018-097088-B-C31, the US Office of Naval Research through Grant No. N00014-19-1-2156 and by EU H2020 MSCA through Project ACHIEVE-ITN (Grant No 765866).

REFERENCES

- [1] M. Bohmer and F. Pampaloni, "Time-resolved confocal scanning device for ultrasensitive fluorescence detection," *Rev. Sci. Instr.* vol. 11, no. 72, pp. 4145-4152, 2001.
- [2] C. Niclass, C. Favi, T. Kluter, F. Monnier, and E. Charbon, "Single-synchronous detection," *IEEE J. Solid-State Circuits*, vol. 44, no. 77, pp. 1977-1989, Sep. 2009.
- [3] N. Faramarzpour, M. J. Deen, S. Shirani, and Q. Fang, "Fully Integrated Single Photon Avalanche Diode Detector in Standard CMOS 0.18- μm Technology," *IEEE Trans. Electron Devices*, vol. 55, no. 3, pp. 760-767, Mar. 2008.
- [4] H. Finkelstein, M. J. Hsu, and S. C. Esener, "STI-bounded single photon avalanche diode in a deep-submicrometer CMOS technology," *IEEE Electron Device Lett.*, vol. 27, no. 11, pp. 887-889, Nov. 2006.
- [5] J. A. Richardson, E. A. G. Webster, L. A. Grant, and R. K. Henderson, "Scaleable Single-Photon Avalanche Diode Structures in Nanometer CMOS Technology," *IEEE Trans. Electron Devices*, vol. 58, no. 7, pp. 2028-2035, Jul 2011.
- [6] E. A. G. Webster, L. A. Grant and R. K. Henderson, "A High-Performance Single-Photon Avalanche Diode in 130-nm CMOS Imaging Technology," in *IEEE Electron Device Letters*, vol. 33, no. 11, pp. 1589-1591, Nov. 2012.
- [7] *Atlas User's Manual: Device Simulation Software*. Silvaco Inc., Santa Clara, CA, 2016.