Doctoral Thesis

Behavioral Modeling of CMOS SPADs Based on TCAD Simulations

Author:

Juan Manuel López Martínez

Advisors:

Ángel Rodríguez Vázquez
Ricardo Carmona Galán

Sevilla, 2022

Behavioral Modeling of CMOS SPADs Based on TCAD Simulations

By:

Juan Manuel López Martínez

PROPUESTA DE TESIS DOCTORAL

PARA LA OBTENCIÓN DEL GRADO DE

DOCTOR EN CIENCIAS Y TECNOLOGÍAS FÍSICAS

Advisors:

Ángel Rodríguez Vázquez

Ricardo Carmona Galán

Sevilla, 2022

Behavioral Modeling of CMOS SPADs Based on TCAD Simulations

Doctoral Thesis
Universidad de Sevilla

Juan Manuel López Martínez
Licentiate in Physics

Advisor: Ángel Rodríguez Vázquez
Full Professor
Electronics and Electromagnetism Department
Universidad de Sevilla

Advisor: Ricardo Carmona Galán
Tenured Scientist
Instituto de Microelectrónica de Sevilla
Consejo Superior de Investigaciones

Universidad de Sevilla
Instituto deMicroelectrónica de Sevilla
Centro Nacional de Microelectrónica
Consejo Superior de Investigaciones Científicas

Calle Américo Vespucio, 28.
Parque Científico y Tecnológico Cartuja.
41092 Sevilla

Author e-mail: jmlopezma@gmail.com

This work is licensed under a Creative Commons Attribution-Non Commercial-No Derivatives 4.0 International License.

Para Ixell,
y para Freya, Èric y Ender,
que me han enseñado que, realmente,
no toda la gente errante anda perdida.

CONTENTS

Chapter 1 Overview of the Single-Photon Avalanche Diode 1

1.1 Introduction, 1
1.2 Modern planar Single-Photon Avalanche Diode structures, 7
1.3 Single-Photon Avalanche Diodes in the literature, 10
1.4 Thesis Overview: Establishing a new model workflow, 11
1.5 References, 12

Chapter 2 SPAD models workflow & development 17

2.1 Introduction, 17
2.2 TCAD Model: Fabrication Simulation with ATHENA, 17
2.3 TCAD Model: Device simulation with ATLAS, 23
2.4 SPAD selection and model overview, 27
2.5 References, 32

Chapter 3 Physical foundations of the basic Verilog-A model 35

3.1 Introduction, 35
3.2 SPAD physical parameters, 35
3.3 SPAD analytical model, 38
3.4 Avalanche triggering events, 40
3.5 Temperature effects modeling, 49
3.6 References, 52

Chapter 4 Building the VERILOG-A model 55

4.1 Introduction, 55
4.2 VERILOG-A basic model description, 56
4.3 Model's state diagram, 64
4.4 References, 65

Chapter 5 Empirical validation of the model workflow 67

5.1 Introduction, 67
5.2 TCAD extracted parameters, 67
5.3 VERILOG-A results, 75
5.4 References, 79

Chapter 6 Theory and modeling of response time and photon-timing jitter in SPADs **81**

6.1 Introduction, 81

6.2 Photon-timing jitter mechanics, 81

6.3 Simulation and experimental validation, 88

6.4 References, 91

Chapter 7 Regarding crosstalk in photodiodes **93**

7.1 Introduction, 93

7.2 TCAD model configuration and simulation, 93

7.3 Restricted quantum efficiency: detection region, 93

7.4 Crosstalk in SPADs, 98

7.5 Crosstalk in Pinned Photodiodes, 102

7.6 References, 108

Conclusions **109**

Appendixes **111**

I Basic G12 model code, 111

II P-gated/Virtual guard ring dual-junction SPAD design, 131

III Acronyms, 135

Acknowledgements **137**

List of publications **139**

CHAPTER 1
OVERVIEW OF THE SINGLE-PHOTON AVALANCHE DIODE

1.1 INTRODUCTION

1.1.1 CMOS IMAGE AND VISION SENSORS

CMOS image and vision sensors are semiconductor devices that employ CMOS technological processes to generate *electrically-coded images* when excited by *visible light* [1] [1]. As compared to other image sensor technologies based on CCDs (*charge-coupled devices*) [2], the main asset of using CMOS resides in the possibility of merging sensing circuitry, signal conditioning circuitry, data conversion circuitry, and processing circuitry on a common substrate. Exploiting this possibility has clear benefits, including:

- Sensors capable of delivering *digital images* are available as **stand-alone chips** that embed *Analog-to-Digital Converter* (ADC) circuits together with the light-sensing *photodiodes*. In other words, the outputs of CMOS imaging chips are ready to interface with *digital processors* without further bridge chips needed. This feature significantly differs from CCD sensors, whose primary outputs are *analog* signals. Moreover, CMOS image sensors may embed *error-correction* digital circuitry to correct artifacts caused by circuital errors on-chip, thus delivering corrected, ready-to-use digital images to the outside. Modern CMOS sensors include one or two ADCs *per column* to achieve high speed sensor-to-processor communication [3]. All-in-all, these capabilities of CMOS sensors allow system engineers to construct sensory systems with reduced *SWaP* [2] and enhanced performance and enable the incorporation of imaging capabilities in applications where CCDs are unfeasible due to cost, size, or performance limitations.
- Pixels can be slightly modified to **adapt** their response to the light stimuli's nature without needing external processing resources. For instance, pixels can incorporate *analog memories* for local error correction with global shutter acquisition and thus preclude motion artifacts [4]. Also, they can include circuitry for adaptive voltage to time encoding, thus allowing the acquisition of scenes with vast *dynamic ranges* [5]. The catalog of CMOS pixels is broad and diverse.
- Pixels can be radically modified to incorporate **early processing capabilities** close to the sensors, thus filtering redundant information and reducing the amount of data right where images are acquired. This concept is behind the so-called *event-driven* vision sensors, which operation resembles that of human retinas [6]. Near-to-sensor circuits also fuel so-called *multi-functional pixel micro-processors* that embed microprocessors' processing and storage features, although operating on analog data [7].

---

[1] Electrical images are sets of analog or digital signal values that encode optical scene information. Electrical images can be pointwise (encoding a single light spot) or consist of linear or matrix (area) arrays. Area sensors, the most prevalent ones, are composed of pixels (single picture elements) arranged typically on a rectangular, uniform set of rows and columns. Each pixel senses a spatial sample of optical scenes (projected on the array and focused through lenses) and produces an analog signal that is the result of transducing the incoming photons first into electrical charge and then into current or, most usually, voltage. Digital images consist of arrays of digital numbers obtained from converting pixel voltages, or currents, onto digital codes by analog-to-digital converter circuits.

[2] SWaP is the acronym for Size, Weight, and Power. Reducing the SWaP of systems is one of the significant benefits of microelectronics technologies. Increasing the number and modalities of components that are embedded per chip unit area enables the construction of miniaturized systems capable of sensing, processing, and actuating.

FIGURE 1.1 illustrates the concepts of CMOS image and vision sensors by displaying, at the top, a typical CMOS imaging chip architecture. The bottom-left sub-figure is a microprocessor pixel block diagram employed for a commercial ultra-compact and ultra-fast area vision sensor intended for machine vision applications – shown at the bottom-right.

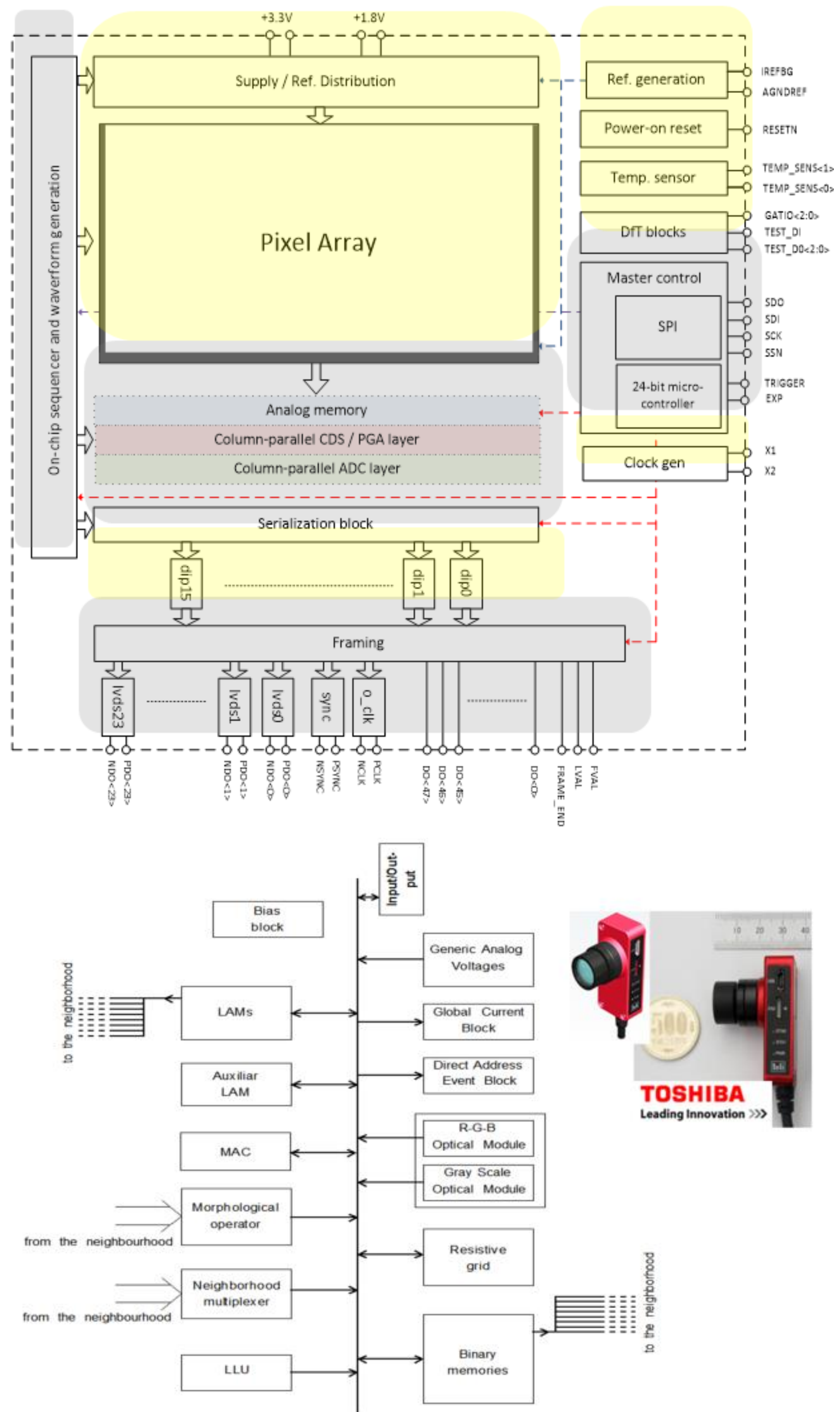


FIGURE 1.1 (top) CMOS imager chip architecture illustrating the functional embedding feature typical of CMOS imagers; (bottom) block diagram of multi-functional micro-processor pixel used for industrial high-speed vision cameras on-a-chip.

The earliest CMOS image architectures and chips are from the late ‘60s of the XXth century [8][9]. The potential of CMOS for function embedding, reduced cost, and improved SWaP was clear from the beginning. However, some 40 years later, many camera companies still hesitated to replace the CCD sensors they used in their products with CMOS counterparts. Moreover,

they were utterly reluctant to use today´s customary solutions, such as distributed ADCs, because they thought these strategies would impact image quality [10]. The proposal of improved *pinned photodiodes* and the associated *4-T pixel* architecture was crucial to remove these obstacles and system engineer doubts [11]-[13]. These strategies enabled pairing the imaging quality of CCDs, thus paving the way for exploiting all potential benefits of CMOS. Indeed, CMOS imager production has been growing significantly during the last decade, as FIGURE 1.2 from Yole development illustrates [14]. While a significant percentage of this growth is rooted in mobile phones and other personal equipment, applications of CMOS image and vision sensors have spanned a myriad of sectors, from scientific instrumentation to the automotive.

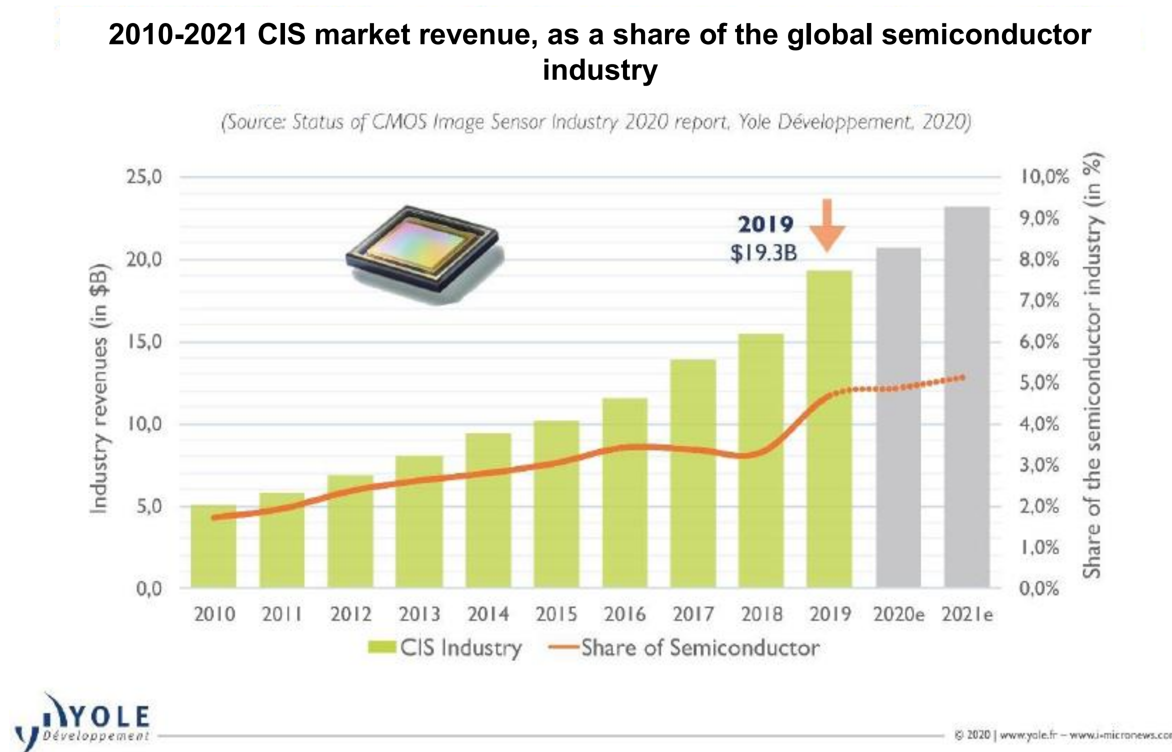


FIGURE 1.2 Growing rate of the revenue share of CMOS image chips over the last decade [14].

Regarding sensing function, most CMOS pixels employ *photodiodes* consisting of *reverse-biased* P-N junctions. Under this biasing, a large electric field exists across a depleted area around the junction. Without any stimuli, a tiny reverse current circulates through the photodiode, called *dark or saturation current*. This dark current is spurious – the smaller the dark current, the better the quality of the photodiode. When light impinges on the photodiode, photons produce electron-hole pairs that may either recombine or be swept to the diode terminals by the depleted region electric field. In this latter case, a photocurrent is generated proportional to the power of the incoming light [15]:

$$I_{ph} = \pm q \cdot \frac{QE(\lambda) \cdot \lambda \cdot A \cdot P_{opt}}{hc} \tag{1.1}$$

where q is the elementary charge, $QE(\lambda)$ the quantum efficiency, λ the wavelength of the absorbed photons, A the area where these photons produce charge carriers that can be collected, also called the active area region, P_{opt} the optical power, h the Planck's constant, and c the speed of light.

Depending on the impedance loading of the photodiode, either this current (low impedance loading) or the voltage resulting after integrating the photocurrent on a capacitor is measured to get pixel electrical data.

1.1.2 GENERALITIES ABOUT SPADS

The conventional photodiode usage described in the previous section implements a *gainless* photo-transduction function; i.e., the power of the pixel electrical data is only a fraction of the incoming light power. Unlike conventional photodiodes, SPADs, the sensing structures addressed in this Thesis have *intrinsic gain*.

SPAD stands for Single Photon Avalanche Detectors. SPADs are photodiodes structurally similar to those used in conventional image sensors. However, while the conventional ones are biased in the low voltage zone within the inverse region, where there is no gain, the SPADs are biased in the avalanche region [16]. In this region, specifically, when operating in **Geiger mode** [17], a single photon is enough to start a *chain reaction* that manifests as a current pulse. This pulse's sharp, active edge encodes the time instant when the photon was detected, which supports the use of SPADs to measure Time-Of-Flight and, from these, *distances* to the objects focused on the sensor. SPADs are therefore suitable for capturing 3D images, that is, for estimating the *depth of objects* in an image, without resorting to binocular vision or interferometry techniques. They can also estimate *light intensities* and therefore capture 2D images based on *counting pulses* [16].

SPAD avalanches must be extinguished (through dedicated *quenching circuitry*) to preclude device damage and prepare the sensors for new photon event measurements. FIGURE 1.3 illustrates some relevant aspects of the operation of SPADs, in particular:

- the strategy to bias them and quench the avalanche (upper left);
- the waveforms of the avalanche current pulses (bottom middle);
- circuits used to quench and extinguish the avalanches and control the SPADs (bottom left and right) [18]; and
- the complexity of the physical structure (top right) is determined by the need to control the electric fields to preclude unwanted avalanches [19].

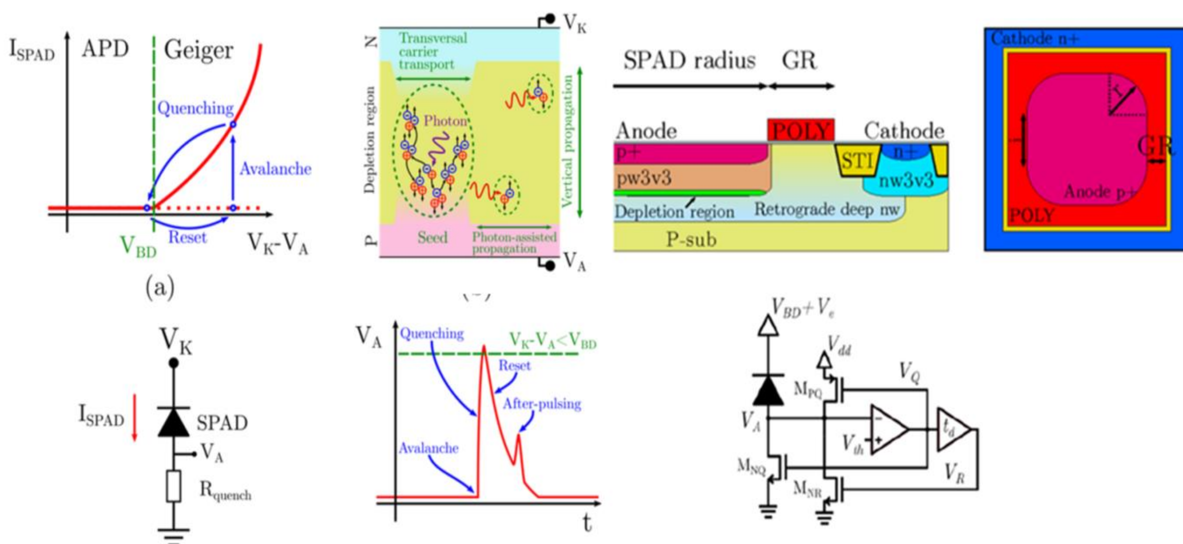


FIGURE 1.3 **(top-left): Illustration of the phenomenology and operation of SPADs; (top-right): SPAD structure (top) and avalanche extension circuitry by means of active circuits.**

The biasing and quenching circuitry must be embedded together with the photodiodes in the pixels.

Getting SPADs to work correctly and qualify for practical applications poses scientific-technical challenges at different levels. At the device level, technologies and physical structures (geometries and sizes) optimized to respond only to target stimuli should be used, reducing as much as possible spurious avalanches, parameterized by the Dark Count Rate (DCR), and increasing the probability of photon detection, parameterized by the Photon Detection Probability (PDP) and the Photon Detection Efficiency (PDE). On the one hand, at the pixel level, it is necessary to minimize the overhead of the circuitry used to extinguish the avalanches and control the waveforms of the signals that polarize the photodiodes. This impacts both the pitch (pixel size) and the Fill-Factor (FF, percentage of the pixel area sensitive to light), as well as spurious phenomena like *after-pulsing* [20] and *jitter* [21] (uncertainty in the measurement of time). On the other hand, at the reading circuitry level, it is necessary to design channels capable of encoding events and encoding time in digital format by using time-to-digital converters) with resolutions appropriate to the application. Among other challenges, at the post-processing level, it is necessary to filter out spurious avalanches and, if possible, extract features or identify events, for example, spatio-temporal contrasts, in the captured images.

Significant methodological challenges are also related to designing complete systems, including SPADs, quenching circuitry, reading circuitry, etc. Simulations must verify these systems before fabrication, and electrical models capturing SPAD physical operation must be devised for that purpose. Building these models defines this dissertation's primary challenge, namely predicting SPAD behavior through accurate models to facilitate the analysis of the impact of the SPAD design decisions on the whole sensing device.

1.1.3 BRIEF REVIEW OF THE DEVELOPMENT OF SPADS

As already mentioned, a SPAD is a P-N junction biased above its *breakdown voltage*. In these conditions, the electric field in the junction is so high that a single carrier, created by, for example, the absorption of a photon in the multiplication region, may provoke a large avalanche current by being accelerated to enough kinetic energy to cause exponential growth of electron-hole pairs, a process known as Impact Ionization [15]. Following an avalanche and the subsequent current pulse, the quenching circuitry extinguish the valance and restores the diode to the initial bias voltage so that it can detect another photon arrival and avoid the eventual thermal burn. The quenching circuitry can be either *passive* (see bottom left schematics in FIGURE 1.3) or *active* (bottom right) [18][19][22][23].

The first developed SPADs were the Reach-Through SPADs (RT-SPAD) developed in the 1970s by McIntyre [24] (FIGURE 1.4 – left). These devices consist of several microns thick structures that are *backside* illuminated. These devices have a broad multiplication region that makes their breakdown voltage in the hundreds of volts, which causes them to have an enormous power consumption. Also, the large multiplication region affected another critical parameter of the SPADs. The statistical variation of the response time, the photo-timing jitter, which is typically expected to be below 150ps in most applications, and in the first early RT-SPADs, could be greater than 1ns.

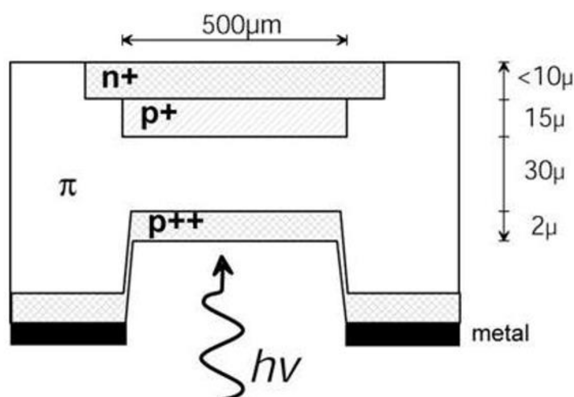
In that early decade, research on these new structures kept on, and although Haitz had already invented it in 1963 [25], it was not until 1981 when S. Cova [26] introduced the first SPAD developed in planar technology, which allowed for a significant decrease in size to a few microns (FIGURE 1.4 – right). Planar implementation implied the possibility of higher doping concentrations in a smaller multiplication region, which increased the junction's electric field

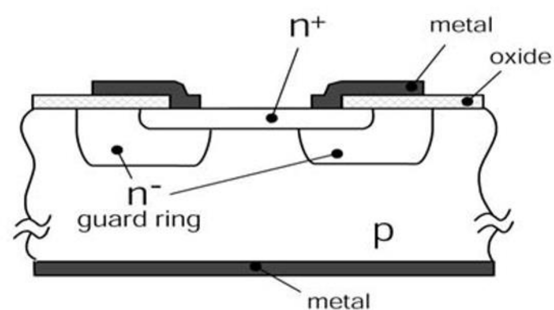
and reduced the breakdown voltage between terminals and, thus, the power consumption. Also, another benefit of a reduced multiplication region is the reduced jitter.

Several researchers kept improving the designs of the SPADs through the 80s and 90s. Still, the giant leap was when Rochas, in 2003, managed to integrate the SPADs along the circuitry needed for its operation in a standard CMOS process. Thus, he developed the technology that allowed the fabrication of *SPAD arrays* in massive numbers [27].

Miniaturization was boosted further with the fabrication of SPAD arrays in *submicron* [28] and *deep-submicron* CMOS technologies [29]. Miniaturization in SPADs is a double-edged sword. On the one hand, it allows larger arrays and improved capabilities, such as reduced power consumption and photo-timing jitter. On the other hand, as the size of the SPADs decreases, the doping concentration at the junction becomes higher, making the junction's electric field higher, which increases the noise associated with tunneling processes [30].

McIntyre's reach-through diode **Haitz's planar diode**





"Thick" SPAD **"Thin" SPAD**

FIGURE 1.4 Basic structure of a reach-trough SPAD and a planar SPAD [31].

1.1.4 ON SPAD APPLICATIONS

Due to their capabilities for signaling events in time and counting photons, CMOS SPAD-based sensors have many applications, particularly when the sensor has to excel in time response capabilities and high counts of photon fluxes.

The Time-Correlated Photon Counting (TCSPC) technique [32] is used in Fluorescence Lifetime Imaging Microscopy (FLIM). This technique enables the identification of some molecules through the mean time required for the molecule to go, from exit to ground level, also called lifetime. This transition release photons a few nanoseconds after the initial excitation, which the sensor can measure. FLIM is often used in biological research [33].

Another application is TOF imaging [34]. A camera with TOF capabilities can recreate objects in three dimensions and determine their distance when they are in motion using an active, pulsed illumination system (FIGURE 1.5). In this system, the actual position of the peak response is used to determine the TOF and, therefore, the distance of the reflection under the TCSPC technique. TOF is becoming essential in technological fields like robotics, gaming, and assisted driving.

Significant contributions have also been made in the *biomedical imaging* field through *time-resolved* single-photon sensors. The best examples of this are Positron Emission Tomography (PET) and Single-Photon Emission Computed Tomography (SPECT), in which most modern equipment uses Silicon PhotoMultipliers (SiPMs) [35]. As SiPMs approach the timing

resolution limit of 100 ps, they are being replaced by TOF PETs. FIGURE 1.6 shows an example of PET imaging.

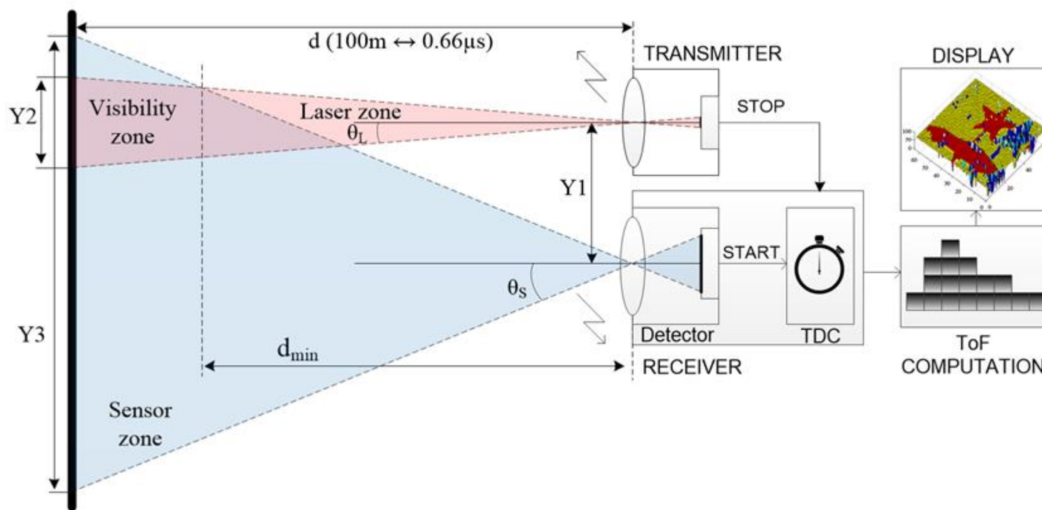


FIGURE 1.5 Conceptual scheme of a Light Detection and Ranging system based on a sensor capable of TOF measurement [36].

SPAD imagers are also used in *Raman spectroscopy*, which provides data on the chemical composition and molecular structure non-destructive manner [37]. This technique relies on light scattering with vibrating molecules, whereby Stokes-scattered Raman photons present a redshift in their spectrum. SPADs allow the design of compact, solid-state detectors for Raman spectroscopy operating in time-resolved mode, whether via very short gates, ideally in the 10–100ps range, given the nature of the Raman signal. This mode of operation also reduces the DCR contribution and thus enhances the overall Signal-to-Noise Ratio (SNR). Several linear SPAD-based systems have been recently developed, usually comprising one or a few lines, initially targeting applications such as mineralogy and biophotonics [38].

1.2 MODERN PLANAR SINGLE-PHOTON AVALANCHE DIODES STRUCTURES

Modern planar SPADs are implemented by implantation and annealing on an epitaxial-growth substrate to create a **P-N junction** near the SPAD surface. This junction is surrounded by the **depletion region** or **multiplication region**, the region where avalanches occur through Impact Ionization processes. This depletion region is encompassed by regions where charge carriers can diffuse to the depletion region and still trigger an avalanche, called **neutral regions**. Also, the electric field maximizes when the junction is reverse-biased above its breakdown voltage. However, this electric field is not homogenous. This is because the dopant implantation gradient is more significant in the perpendicular direction of the doping beam, making the electric field in that part of the junction greater, and thus the probability of avalanche, making the SPAD sensitive only in a portion of its surface.

Researchers have proposed different solutions to overcome this problem. FIGURE 1.7 shows some most used structures. The P-N junction is supposed *round-shaped*, but other shapes, like squares, rounded squares, hexagons, and octagons, are often used [29][39][40].

- Cova used an N lightly doping implant at the edge of the more heavily doped N+ implant [22] (FIGURE 1.7a). In this configuration, several of these structures share the same substrate or the well, so they suffer from increased noise and crosstalk from neighboring devices.

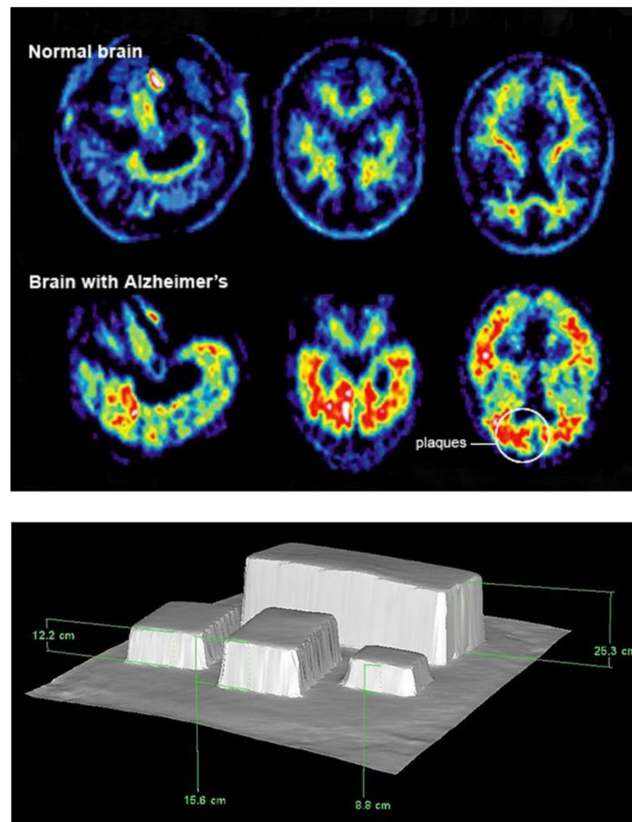


FIGURE 1.6 (top) Example of tomography shows a patient with a positive Alzheimer's diagnostic; (bottom) A range image with measurement results of a TOF-camera.

- FIGURE 1.7b shows Rochas's design [27] which encapsulates Cova's structure in an individual well to reduce crosstalk and protect the multiplication region from several noise sources. Also, he flipped the doping to meet standard CMOS technologies specifications.
- After Rochas's structure, researchers kept looking for other ways to improve SPADs capabilities while trying to stick to a standard CMOS technology with mixed success. These changes involved new concepts of guard rings. Finkelstein attempted to substitute the wells implanted at the edge of the junction with a Shallow Trench Isolation (STI) [41], intending to reduce the lower limit to the junction size of the depletion region generated by the wells that formed the classic guard ring (FIGURE 1.7c). This strategy yielded reduced pixel sizes and enhanced the fill factors. However, the STIs in contact with the junction increased the dark counts to the order of MHz due to the traps present on the surface of the silicon crystal, which made the design unusable. Gerbash proposed coating the STIs with several layers of semiconductor materials to force carriers generated by the surface traps to recombine before arriving at the multiplication region [42]. Nevertheless, this process, although compatible with CMOS processes, it is not available in standard CMOS technologies.
- Richardson [43] proposed other innovations in the guard ring area with up to three new structures, which took advantage of the availability of a buried N-type well formed by a high-energy ion implantation step before forming the N-Well. This strategy resulted in a buried deep N-Well with a retrograde doping profile. If used to encapsulate a P-Well completely, the edge of this well is in contact with an increasingly less doped N-type well as it gets closer to the surface, which will weaken the electric field. Thus, the SPAD enters

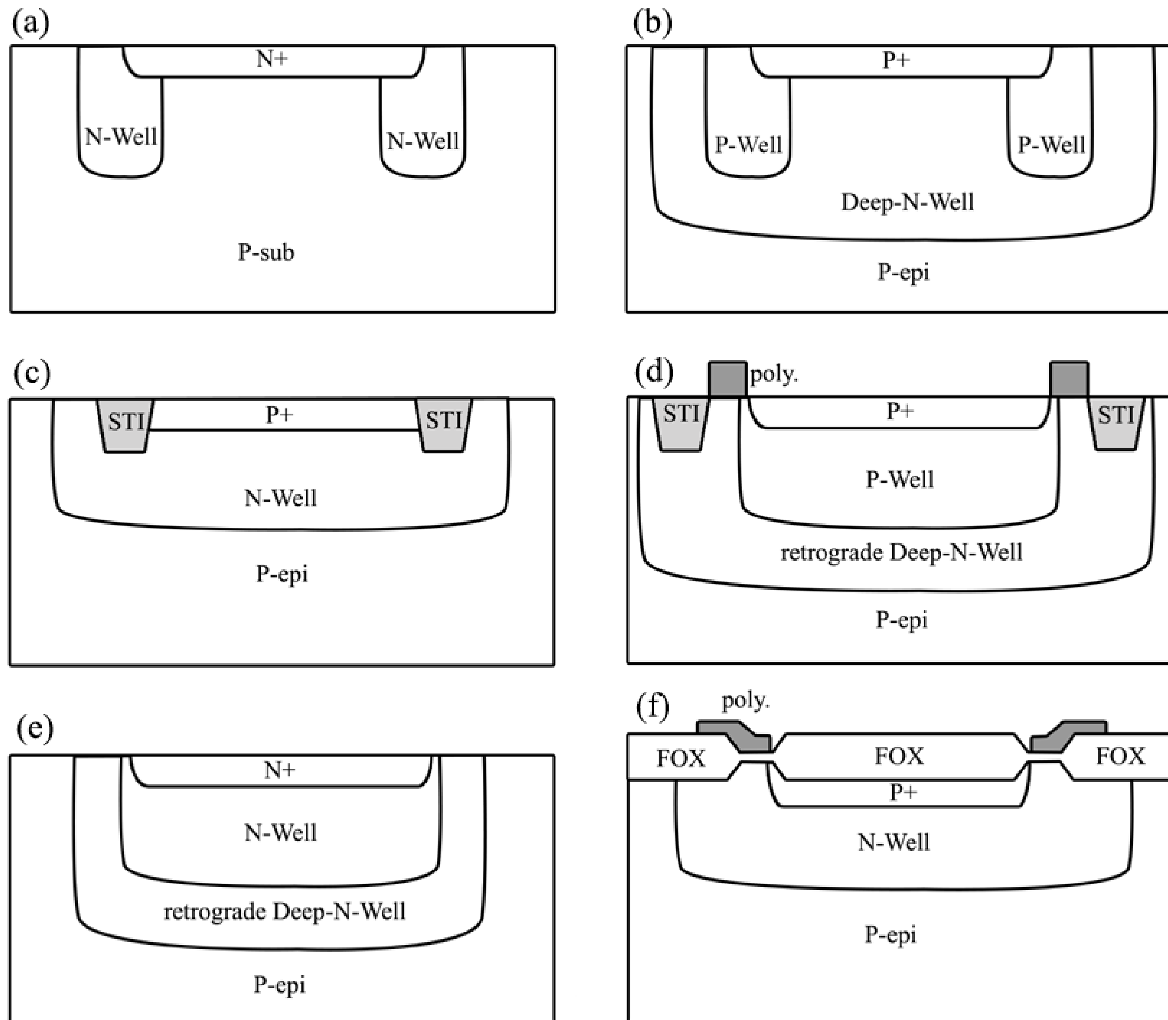


FIGURE 1.7 Modern SPADs simplified structures. They are defined by the way they avoid premature edge breakdown. (a) was the first structure which had a guard ring to prevent edge breakdown, and was proposed by Cova in 1981 [22]. An upgraded version of this structure was used by Rochas [27], whereas structure (c) was proposed by Finkelstein [41]. Structure (d) was proposed by Richardson [43] and structure (e) by Webster [44]. Finally, a novel method for preventing edge breakdown in structure (f) was proposed by Dandin [45].

the breakdown in the deepest part of the junction, creating a de-facto guard ring. This guard ring is called a virtual guard ring. FIGURE 1.7d depicts the most used out of the three proposals, which can be fabricated with standard CMOS technologies. This design presents a lower doped junction than FIGURE 1.7b, which lowers the electric field enough to make the noise due to tunneling processes negligible. Besides, as the junction is more profound in bulk, it also improves the PDE at longer optical wavelengths; however, it does not perform well in the shortest wavelength of the visual spectrum.

- Shortly after, Webster proposed to take advantage of the retrograde nature of the epitaxial growth where the wells are implanted to create a virtual guard ring around an N-type well and form a junction at the most profound boundary of a Deep-N-Well (FIGURE 1.7e) [44]. Therefore, this device offers excellent detection capabilities on the longer end of the spectrum.
- Finally, Dandin proposed a new approach to control premature edge breakdown. He applied a gate to the edge of the junction to prevent edge breakdown (FIGURE 1.7f) [45]. This device is called Perimeter Gated Single-Photon Avalanche Diode (PGSPAD). The

main issue with this SPAD is that it does not meet the requirements to be fabricated in standard CMOS technology. Custom technologies raise the price and make unpractical mass-produce these diodes.

1.3 SINGLE-PHOTON AVALANCHE DIODE MODELS IN THE LITERATURE

The simplest analytic model was proposed by Oldham (FIGURE 1.8a) in 1972 [46]. It consists of a DC source to act as the diode breakdown voltage (V_B), a series resistor as the internal diode resistance (R_s), a capacitor as the junction capacitance (C_d), and a switch to simulate the impact of a photon.

Mora introduced some parasitic components to account for real-like SPADs' recovery and quenching times [47]. FIGURE 1.8b shows the capacitor acting as junction capacitance as in FIGURE 1.8ba (C_{AC}). Still, also we can observe the stray capacitances from anode and cathode to substrate (C_{AS} and C_{CS}, respectively). Also, the diode resistance is given by the sum of the resistance of the multiplication region and the resistance of the neutral regions crossed by the avalanche current, and the switch of FIGURE 1.8ba that mimics the avalanche has been replaced by an nMOS transistor. This work also included a piecewise linear voltage source to describe better the current-voltage characteristics of the SPAD. This analytic model was designed to be quenched by a passive quenching circuit, which offers some limitations. It was developed in SPICE with standard cell libraries, making it accessible for other researchers. Later Zappa developed an active quenching circuit for this model that solved some of the issues [48]. However, it still presented convergence problems in the piecewise linear function that spurious artifacts appeared when the device operated in extreme conditions in the presence of other complex electronics.

The analytic description of the core SPAD behavior was not modified in subsequent models. The only thing that changed is the capacity to describe the behavior of the diode with more precision through a Hardware Description Language (HDL). The first model using the VERILOG-A HDL [49], an analog extension of the common VERILOG HDL and Cadence SPECTRE Simulator, was proposed by Mita [50]. As all the essential simulation tools can perform mixed-mode simulations, the models written in Verilog-A can be easily used in any design or simulation context. This first model solved the convergence problems mentioned

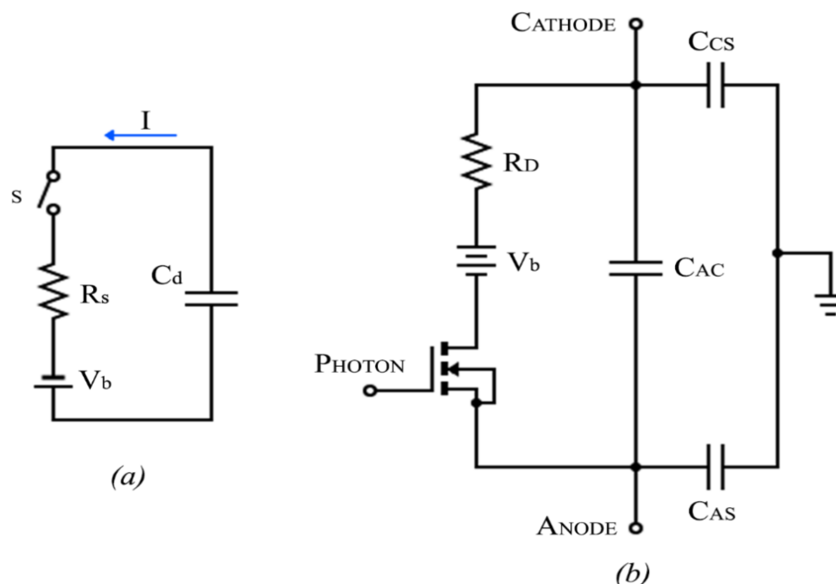


FIGURE 1.8 Oldham's (a) and Mora's (b) analytic single-photon avalanche diodes respectively.

above, as more complex mathematical functions could be used in Verilog-A. Besides, it included the voltage dependence of the diode junction capacitance, which boosted the model precision. However, this model did not describe any behavior associated with statistical phenomena, like dark counts, after-pulsing or response time, and photon-timing jitter simulation.

Shortcomings of Mita's model were partially solved by Giustolici [51], who successfully provided a VERILOG-A model capable of calculating the Carrier Generation Rate (CGR), albeit only due to thermal origin, thus giving an approximation of the dark count's rate. Not only that, but his model also provided the first simulation of the after-pulsing phenomena. Giustolici's model did not include some critical noise contributions associated with tunneling processes despite its improvements. Also, the after-pulsing modeling had two defects. The first one is that the activation energy of the deep-level traps used in the model is plausible but lacks data backing them. Secondly, the after-pulsing depends on the time interval since the previous avalanche event was not addressed.

The more recent VERILOG-A model proposed by Cheng [52] addresses almost all Giustolici's model issues. It provides dark counts from band-to-band tunneling and considers the after-pulsing dependence of the time interval from the last avalanche. Most importantly, it contemplates the extraction of the physical parameters. However, Cheng's model still lacks coverage of important phenomena, namely:

- the Trap-Assisted Tunneling (TAT) contribution to the dark count rate is not included;
- there is no validation of the Shockley-Read-Hall (SRH) theory of recombination through defects based on an actual existing defect;
- the coverage of temperature dependencies is incomplete;
- the model does not include the possible effects on the results of other internal SPAD structures like a guard ring or a second junction.

1.4 THESIS OVERVIEW: ESTABLISHING A NEW MODELING WORKFLOW

When trying to apply Cheng's model to the typical SPAD structures of FIGURE 1.7, we realize there is no clue regarding what kind of single-photon avalanche is actually being modeled. Remember that all previous models, including Cheng´s one, use the same analytic description based on FIGURE 1.8b. There is no mention of a guard ring or a second junction contributing to their capacitance or other complex SPAD structures. In other words, all previous models only consider a simple SPAD composed of a sole junction. Our methodology overcomes this drawback by pursuing physically-consistent models, i.e., models that take into account the SPAR inner structure and map the underlying physics on a VERILOG-A description. In the quest of this challenge, this Thesis proposes a workflow including the following points:

- Select an actual SPAD structure, described at physical level, which model is targeted.
- Simulate its fabrication process with Athena from the Silvaco tools suite to feed a TCAD simulator with an accurate model structure that can include fabrication defects.
- Extract its key parameters with TCAD simulation of the device, performed with ATLAS, from the Silvaco TCAD tools suite.
- Combine physical parameters and analytical descriptions and the data extracted from the TCAD simulations to build an accurate VERILOG-A model.

Besides using this physically-consistent methodology, our models embed the following new features:

- Inclusion of the contributions to the dark count rate from the TAT processes.

- New approximation to the inclusion of the Band-To-Band Tunneling (BTBT) contribution to dark counts.
- Inclusion of Webster's spectroscopy results about traps and deep-level traps [53].
- Simulation of the SPAD self-heating and the SPAD dynamic behavior with the temperature.
- Simulation of the time response and photon-timing jitter.
- Crosstalk analysis of the models.

The Thesis Chapter explains the proposed method and the approach followed to include the above-mentioned features.

1.5 REFERENCES

[1] M. Kriss, *Handbook of Digital Imaging – Vol 1: Image Capture and Storage*. Wiley 2015.

[2] J. Nakamura (Editor), *Image Sensors and Signal Processing for Digital and Still Cameras*. Taylor & Francis 2006.

[3] J.A. Leñero-Bardallo and A. Rodríguez-Vázquez, "ADCs for Image Sensors: Review and Performance Analysis." Chapter 3 in Analog Electronics for Radiation Detection (edited by R. Turchetta), CRC Press 2016.

[4] S. Velichko, "Overview of CMOS Global Shutter Pixels." *IEEE Transactions on Electron Devices*, vol. 69, no. 6, pp 2806-2814, June 2022.

[5] S. Vargas-Sierra, G. Liñán-Cembrano, and A. Rodríguez-Vázquez, "A 151dB High Dynamic Range CMOS Image Sensor Chip Architecture with Tone Mapping Compression Embedded in-Pixel." *IEEE Sensors Journal,* vol. 15, pp. 180-195, January 2015.

[6] J.A. Leñero-Bardallo, R. Carmona-Galán, and A. Rodríguez-Vázquez, "Applications of Event-Based Image Sensors—Review and Analysis." *International J. of Circuit Theory and Applications.* vol. 46, pp. 1620-1630, September 2018.

[7] A. Rodríguez-Vázquez, J. Fernández-Berni, J.A. Leñero-Bardallo, I. Vornicu, and R. Carmona-Galán, "CMOS Vision Sensors: Embedding Computer Vision at Imaging Front-Ends." *IEEE CIrcuIts and Systems MagazIne*, vol. 2ndQuarter 2018, pp. 90-107, April 2018.

[8] G. Weckler, "Operation of p-n Junction Photodectors in a Photon Flux Integration Mode." *IEEE Journal of Solid State Circuits*, vol. 2, no. 3, pp. 65–73, 1967.

[9] P. J. W. Noble, "Self-Scanned Silicon Image Detector Arrays." *IEEE Transactions on Electron Devices*, vol. 15, pp. 202–209, April 1968.

[10] A. Rodríguez-Vázquez, private communication from AnaFocus Ltd. Business exploration activities in Japan.

[11] E. Fossum, "CMOS Image Sensors: Electronic Camera-on-a-Chip." *IEEE Transactions on Electron Devices,* vol. 44, no.10, pp. 1689-1698, October 1997.

[12] T. Lule, S. Benthien, H. Keller, F. Mutze, P. Rieve, K. Seibel, M. Sommer, M. Bohm, Sensitivity of Cmos Based Imagers and Scaling Perspectives." *IEEE Transactions on Electron Devices*, vol. 47, pp. 2110–2122, 2000.

[13] E. R. Fossum and D. B. Hondongwa, "A Review of the Pinned Photodiode for CCD and CMOS Image Sensors." *IEEE J. Electron Devices*, vol. 2, no. 3, pp. 33–43, May 2014.

[14] Yole Development, *Press Release CIS Market Update*, November 2020. www.yole.fr

[15] S. M. Sze and K. K. Ng, *Physics of Semiconductor Devices*, 3rd ed. New York, NY, USA: Wiley, 2007.

[16] M. Perenzoni, "Single-Photon Avalanche Diode-Based Detection and Imaging." *IEEE Solid-State Circuits Magazine*, pp. 26-34, Summer 2018.

[17] N. Britun and A. Nikiforov, *Photon Counting - Fundamentals and applications*. Rijeka, HRV: IntechOpen, 2018.

[18] I. Vornicu, R. Carmona-Galán, B. Pérez-Verdú, and A. Rodríguez-Vázquez, "Compact CMOS Active Quenching/Recharge Circuit for SPAD Arrays." *International J. of Circuit Theory and Applications*, vol. 44, pp. 917-928, April 2016.

[19] I. Vornicu, J. M. López-Martínez, F. Bandi, R. Carmona-Galán, and A. Rodríguez-Vázquez, "Design of High-Efficiency SPADs for LiDAR Applications in 110nm CIS Technology." *IEEE Sensors Journal*, vol. 21, no. 04, pp. 4776-4785, February 2021.

[20] M. Moreno-García, L. Pancheri, M. Perenzoni, R. del Río, O. Guerra-Vinuesa, and A. Rodríguez-Vázquez, "Characterization-Based Modeling of Retriggering and After-pulsing for Passively Quenched CMOS SPADs." *IEEE Sensors Journal*, vol. 19, no. 14, pp. 5700-5709, July 2019.

[21] I. Vornicu, R. Carmona-Galán, and A. Rodríguez-Vázquez, "Arrayable Voltage-Controlled Ring-Oscillator for Direct Time-of-Flight Image Sensors." *IEEE Transactions on Circuits and Systems-I: Regular Papers.* vol. 64, no. 11, pp. 2821-2834, November *2017.*

[22] S. Cova, M. Ghioni, A. Lacaita, C. Samori, and F. Zappa, "Avalanche Photodiodes and Quenching Circuits for Single-Photon Detection." *Applied Optics*, vol. 35, no. 12, pp. 1956–1976, 1996.

[23] F. Zappa, M. Ghioni, S. Cova, C. Samori, and A. C. Giudice, "An Integrated Active-Quenching Circuit for Single-Photon Avalanche Diodes." *IEEE Transactions on Instrumentation and Measurement,* vol. 49, no. 6, pp. 1167–1175, Dec. 2000.

[24] R. J. McIntyre, "The Distribution of Gains in Uniformly Multiplying Avalanche Photodiodes: Theory." *IEEE Transactions on Electron Devices*, vol. 19, no. 6, pp. 703-713, 1972.

[25] A. Goetzberger, B. McDonald, R. H. Haitz, and R. M. Scarlett, "Avalanche Effects in Silicon p-n Junctions. II. Structurally Perfect Junctions." *Journal of Applied Physics*, vol. 34, no. 6, pp. 1591-1600, June 1963.

[26] S. Cova, A. Longoni, and A. Andreoni, "Towards picosecond resolution with single-photon avalanche diodes." *Review of Scientific Instruments*, vol. 52, no. 3, pp. 408-412, March 1981.

[27] A. Rochas, A. R. Pauchard, P. A. Besse, D. Pantic, Z. Prijic, and R. S. Popovic, "Low-noise Silicon Avalanche Photodiode Fabricated in Conventional CMOS Technologies." *IEEE Transactions on Electron Devices*, vol. 49, no. 3, pp. 387-394, March 2002.

[28] C. Niclass, M. Sergio, and E. Charbon, "A single photon avalanche diode array fabricated in 0.35-μm CMOS and based on an event-driven readout for TCSPC experiments." *Proc. SPIE 6372, Advanced Photon Counting Techniques*, 63720S, November 2006.

[29] C. Niclass, M. Gersbach, R. Henderson, L. Grant, and E. Charbon, "A Single Photon Avalanche Diode Implemented in 130-nm CMOS Technology." *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 13, no. 4, pp. 863-869, July-Aug. 2007.

[30] S. M. Sze, M-K. Lee, *Semiconductor Devices: Physics and Technology*. Wiley, 2012.

[31] S. Cova, M. Ghioni, A. Lotito, I. Rech, and F. Zappa, "Evolution and prospects for single-photon avalanche diodes and quenching circuits." *Journal of Modern Optics*, vol. 51, no. 9-10, pp. 1267-1288, 2004.

[32] W. Becker, *Advanced Time-Correlated Single-Photon Counting Techniques*. New York, NY: Springer, 2005.

[33] W. Becker, "Fluorescence Lifetime Imaging—Techniques and Applications." *Journal of Microscopy*, vol. 247, pp. 119–136, 2012.

[34] F. Piron, D. Morrison, M. R. Yuce, and J. M. Redouté, "A Review of Single-Photon Avalanche Diode Time-of-Flight Imaging Sensor Arrays," *IEEE Sensors Journal*, vol. 21, no. 11, pp. 12654-12666, Jun. 2021.

[35] A. Muntean, E. Venialgo, S. Gnecchi, C. Jackson and E. Charbon, “Towards a Fully Digital State-of-The-Art Analog SiPM.” *2017 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, 2017, pp. 1-4.

[36] I. Vornicu, R. Carmona-Galán, and A. Rodríguez-Vázquez, “CMOS SPADs with Embedded Smartness.” *Proc. 2002 International SPAD Sensors Workshop*, Edinburg 2020.

[37] C. Krafft, B. Dietzek, J. Popp, and M. Schmitt, “Raman and Coherent Anti-Stokes Raman Scattering Microspectroscopy for Biomedical Applications.” *J. Biomed. Opt.* 17(4) 040801, April 2012.

[38] Y. Maruyama, J. Blacksberg, and E. Charbon, “A 1024× 8, 700-ps Time-Gated SPAD Line Sensor for Planetary Surface Exploration With Laser Raman Spectroscopy and LIBS.” *IEEE Journal of Solid-State Circuits*, vol. 49, no. 1, pp. 179-189, Jan. 2014.

[39] 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 Transactions on Electron Devices*, vol. 55, no. 3, pp. 760-767, Mar. 2008.

[40] D. Palubiak, M. M. El-Desouki, O. Marinov, M. J. Deen, and Q. Fang, “High-Speed, Single-Photon Avalanche-Photodiode Imager for Biomedical Applications.” *IEEE Sensors Journal*, vol. 11, no. 10, pp. 2401-2412, Oct. 2011.

[41] 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, pp. 887-889, Nov. 2006.

[42] M. Gersbach, J. Richardson, E. Mazaleyrat, S. Hardillier, C. Niclass, R. Henderson, L. Grant, and E. Charbon, “A Low-Noise Single-Photon Detector Implemented in A 130nm CMOS Imaging Process,” *Solid-State Electron*, vol. 53, pp. 803–808, 2009.

[43] 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 Transactions on Electron Devices*, vol. 58, no. 7, pp. 2028-2035, July 2011.

[44] E. A. G. Webster, L. A. Grant, and R. K. Henderson, “A High-Performance Single-Photon Avalanche Diode in 130-nm CMOS Imaging Technology.” *IEEE Electron Device Letters*, vol. 33, no. 11, pp. 1589-1591, Nov. 2012.

[45] M. Dandin, A. Akturk, B. Nouri, N. Goldsman, and P. Abshire, “Characterization of Single-Photon Avalanche Diodes in a 0.5um Standard CMOS Process—Part 1: Perimeter Breakdown Suppression.” *IEEE Sensors Journal*, vol. 10, no. 11, pp. 1682-1690, Nov. 2010.

[46] W. G. Oldham, R. R. Samuelson, and P. Antognetti, “Triggering phenomena in avalanche diodes.” *IEEE Transactions on Electron Devices*, vol. 19, no. 9, pp. 1056-1060, Sept. 1972.

[47] A. Dalla Mora, A. Tosi, S. Tisa, and F. Zappa, “Single-Photon Avalanche Diode Model for Circuit Simulations.” *IEEE Photonics Technology Letters*, vol. 19, no. 23, pp. 1922-1924, Dec.1, 2007.

[48] F. Zappa, A. Tosi, A. Dalla Mora, and S. Tisa, “SPICE Modeling of Single Photon Avalanche Diodes.” *Sensors and Actuators A: Physical*, vol 153, no. 2, pp. 197-204, 2009.

[49] *Verilog-A Language Reference Manual*. Los Gatos, CA: Open Verilog International, 1996

[50] R. Mita, G. Palumbo, and P. G. Fallica, “Accurate Model for Single-Photon Avalanche Diodes.” *IET Circuits, Devices Syst.*, vol. 2, no. 2, pp. 207–212, Apr. 2008.

[51] G. Giustolisi, R. Mita, and G. Palumbo, “Verilog-A modeling of SPAD statistical phenomena.” *Proc. 2011 IEEE International Symposium of Circuits and Systems (ISCAS)*, pp. 773-776, 2011.

[52] Z. Cheng, X. Zheng, D. Palubiak, M. J. Deen, and H. Peng, “A Comprehensive and Accurate Analytical SPAD Model for Circuit Simulation.” *IEEE Transactions on Electron Devices*, vol. 63, no. 5, pp. 1940-1948, May 2016.

[53] E. A. G. Webster and R. K. Henderson, "A TCAD and Spectroscopy Study of Dark Count Mechanisms in Single-Photon Avalanche Diodes." *IEEE Transactions on Electron Devices*, vol. 60, no. 12, pp. 4014-4019, Dec. 2013.

CHAPTER 2
SPAD MODELS WORKFLOW & DEVELOPMENT

2.1 INTRODUCTION

As stated in Chapter 1, some previously reported SPAD models make assumptions that may not be consistent with the device's physical operation and may result in quantitative and even qualitative deviations from real-like SPADs. Because SPAD sensors exploit fundamental device behavior, being closer to the underlying physical reality is mandatory to reduce the risk of system-level artifacts and faults. This dissertation pursues this challenge by using a *bottom-up* modeling methodology that relies on TCAD simulations of practical SPAD devices.

At the bottom level, the methodology employs a process simulator based on device physics called ATHENA [1]. This simulator analyzes structures resulting from process sequences by using systems of equations that describe the physics of semiconductor operation and helps map them onto SPAD external behavioral features. Potential advantages of using this simulator include:

- Accurate enough, physically-consistent prediction of SPAD performance;
- Insightful model description;
- Quicker response than empirical models with even simpler structures;

At the next level, the methodology employs a physically-consistent device simulator, ATLAS [2], to characterize the ATHENA physical models. ATLAS predicts the electrical characteristics that are associated with specified physical structures and bias conditions. The tool solves a set of differential equations derived from Maxwell's Laws to describe carrier transport within SPAD structures, thus allowing for elucidating behavior that remained reluctant for other approaches, such as the electric field or the parasitic capacitances between terminals.

These physically-realistic models provide the support for describing VERILOG-A HDL models that are fully compatible with the main circuit simulators, such as Cadence SPECTRE or Synopsys HSPICE.

This chapter firstly covers the simulation of the SPAD fabrication process with ATHENA. Then, the device structure capabilities derived from it are tested and analyzed with ATLAS, which will be integrated into a VERILOG-A model. Finally, we will review the different models resulting from this process.

2.2 TCAD MODEL: FABRICATION SIMULATION WITH ATHENA

ATHENA uses a sequence of commands to capture the physical process that result in a final device structure. It is convenient to illustrate the simulation of SPAD fabrication processes by using the structure of FIGURE 2.1. Rochas proposed this structure at the early stages of the SPAD sensor roadmap [3]; complementary to this work, the reference [4] provides a model-oriented description of its empirical performance.

FIGURE 2.1 shows the doping profile of the final structure, where the yellowish-reddish parts are doped with N-type dopants, whereas bluish-purplish parts are doped with P-type dopants. The process steps leading to this final structure are listed below:

- Mesh definition.
- Substrate definition. The substrate is the physical material that holds the device's components.
- Channeling effect.

- Wells Implantation.
- Diffusion processes.
- Declaring electrodes and saving the structure.

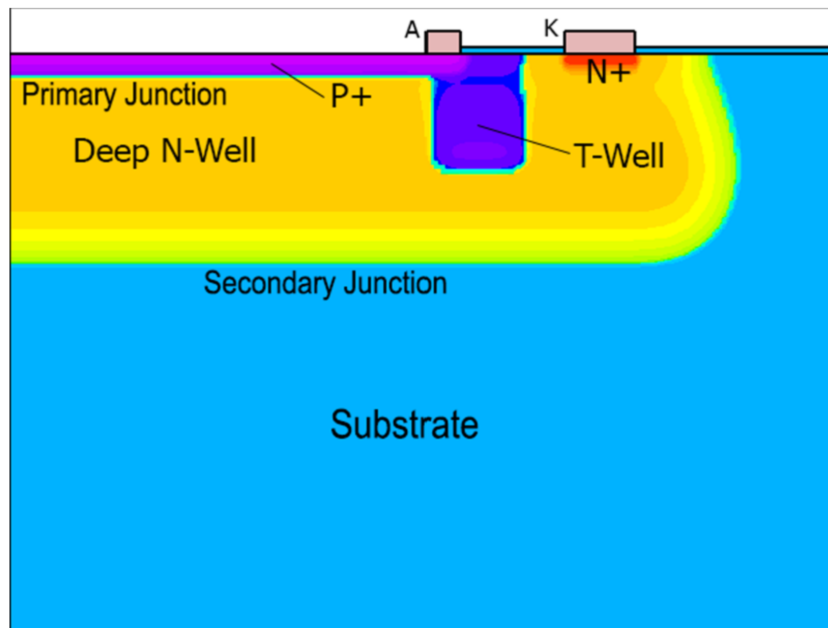


FIGURE 2.1 Final device structure based on the Rochas's SPAD developed with Athena. The left border acts as axis of revolution.

2.2.1 MESH DEFINITION

Both ATHENA and ATLAS are initialized through an open text file with DECKBUILD, an interactive, graphic runtime environment for developing process and device simulation input, from the same tool family. To call ATHENA, we write the following statement:

```
go athena
```

All the subsequent statements will be thereby considered ATHENA's statements.

The first step to specifying devices in ATLAS and ATHENA is to define the *mesh*, which will be used as a basis for the calculations of simulations. TCAD tools partition the space through vertical and horizontal lines. These lines divide the space into *squares*, and these, in turn, are divided into two *triangles*. These triangles are the basic unit of simulation; the smaller the triangle sizes, the more accurate the simulation. Non-uniform zones of the device, such as depletion regions or the regions near the contact surfaces between materials, should be defined by small enough triangles since the variables in these regions change rapidly with distance. However, too small triangles consume system resources and result in long simulation times, thereby raising a trade-off during the device specification phase.

For this reason, it is important to choose the proper density of triangles for every structure within the device, specifying the desired density of vertical and horizontal lines that form them. This is done by declaring a minimum of two points in any given location with the *line* statement:

```
line [dimension] location=[µm] spacing=[µm]
```

where *dimension* can be either *x* for horizontal lines or *y* for vertical lines, and *spacing*, the lines per micrometer in that point. This form of declaring the mesh creates a density gradient of lines between those two points.

Before defining the mesh, it is essential to be aware of the final size of the device and its components since the definition of the mesh already has implicit data. For example, the basic SPAD *guard ring* model has 10µm in diameter and is 1.8µm deep, with the primary junction at 0.2µm deep. Hence the mesh must be thinner to secure accuracy in those places (see FIGURE 2.2).

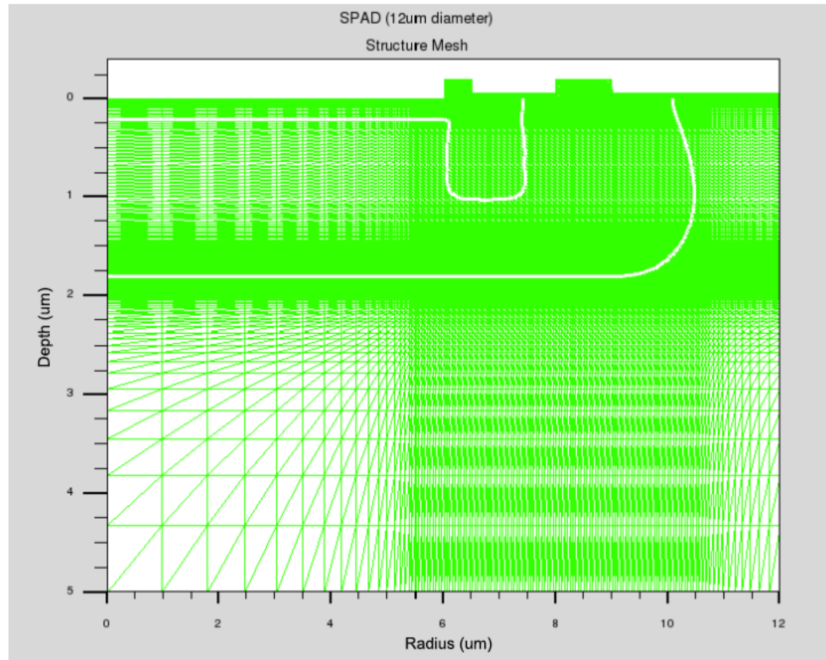


FIGURE 2.2 G12 SPAD model mesh. White lines represent the primary and secondary junctions.

2.2.2 SUBSTRATE DEFINITION

Following mesh definition, the initial *substrate* is established. All the SPADs addressed in this Thesis have a common silicon substrate doped with boron at a concentration of 6×10^{14} impurities/cm^3 with cylindrical symmetry and a crystal lattice orientation of {100}. This orientation for the silicon wafer is preferred in integrated circuits over others like {111} or {110} because the state density in the Si/SiO_2 surface is minimum, which results in higher carrier mobility [5]. The syntax of the statement that would define the initial substrate is as follows:

```
init silicon boron c.boron=6e14 orientation=100 cylindrical
```

The *cylindrical* parameter allows extrapolating 3D results from a 2D model, making it quite convenient to simplify calculations.

In the case of Rocha's SPAD, the different layers of the SPAD will lie on an *epitaxial layer*. Normally, the epitaxial layer is grown from a phosphorus-implanted substrate, but for this SPAD, it is not necessary. Thus, a relatively low doped material is defined instead as the initial material, serving as the epitaxial layer. Bear in mind that a normal phosphorus-implanted substrate (P-substrate) has a dopant density of at least two orders of magnitude higher than that of the typical epitaxial layer.

2.2.3 CHANNELING EFFECT

The next step is to implant the different *wells* that shape the device, but before that, we have to prevent an undesirable effect: the *channeling effect*. When crystals are irradiated with *ions*, the

ions' implantation direction may coincide with that of the silicon crystal planes, thus resulting in exponential patterns rather than the desired *Gaussian* patterns. There are two ways to avoid this effect, and both of them are used:

- The first is to incline the incident angle of the ions implantation beam between 5 and 10 degrees versus the silicon crystal plane. Optionally the plane containing the beam can also be rotated between 20 and 30 degrees.
- The second way is to add a nanometers thick screen silicon dioxide layer, so when the ions collide with that layer molecules, they are scattered and implanted randomly. In the several models developed, an 8nm oxide layer will be placed above the device to avoid the channeling effect of the subsequent implantations. That layer uses the following code:

```
deposit oxide thick=0.008 divisions=4
```

It is interesting to add that these processes are already related to the real fabrication process as total alignment with the crystal plane never occurs in reality, and, in any fabrication process, a small oxide layer is always created in the substrate due to the presence of oxygen.

2.2.4 WELLS IMPLANTATION

The next step is to begin the procedure of implanting the different wells within the substrate. Ion implantation distributions are calculated through analytical models and the use of *spatial moments*. This calculation method is based on *range concepts* [6] in which an ion-implantation profile is constructed from a previously prepared (calculated or measured) set of spatial moments. For example, the simplest approximation to ion implantation is a Gaussian distribution, which has two spatial moments than can be obtained by theory or experimentally, the penetration depth or range of the ions and its projected straggle or standard deviation. However, it is well established today that a realistic approximation needs at least 4 spatial moments so that it can take into account the lateral spread and its standard deviation. For that purpose, ATHENA uses, as a basis, the Pearson IV distribution introduced by Hofker in 1975 [7]. However, the Gaussian distribution is still useful for easily providing ATHENA with the needed data for implanting the wells adequately, as will be shown later on.

The order of well implantation goes from more to less depth. Taking as a reference FIGURE 2.1, that order would be:

- Deep-N-Well
- T-Well,
- P+ Well
- N+ Well.

The process of well implantation is described in the following paragraphs.

First, a mask is deposited on our device to limit the area in which the ion implantation will occur. So we will begin depositing the material that will serve as a mask:

```
deposit barrier thick=0.03 divisions=3
```

Then it is modified by cutting it into the desired shape. In a real fabrication process, this is done by projecting the mask patterns in a process called *photolithography* or by electron-beam lithography (EBL), which is a maskless process and is able to actually draw much smaller (sub 10nm) details in the resist. The *etch* statement can simulate these processes. The uses of *etch* present in the model are listed below.

i) Cut a polygon of a given material indicating the coordinates of its vertices. For example, a quadrilateral:

```
etch [material] start x=[x1] y=[y1]
```

```
etch cont x=[x2] y=[y2]
etch cont x=[x3] y=[y3]
etch done x=[x4] y=[y4]
```

ii) To lower the air-exposed surface of a material:

```
etch [material] dry thick=[µm]
```

iii) Remove all material from one position in one direction:

```
etch [material] right/left p1.x=[µm]
etch [material] below/above p1.y=[µm]
```

iv) Delete a material from the entire device:

```
etch [material] all
```

v) Then, several ion implantations with different doses will shape the well. The syntax to implant impurities is the following:

```
implant [impurity] energy=[KeV] dose=[particles/cm2] tilt=[degrees]
rotation=[degrees]
```

All of these steps correspond to the real fabrication process itself, which result in realistic structures that can be fabricated.

As said previously, the *implant* statement establishes a Pearson IV implantation profile by default, as it includes lateral spread. However, a Gaussian profile can still be used to calculate the penetration range easily. Then, choosing the dose and energy carefully, it is possible to calculate the peak concentration as well as the range of said peak in silicon. The peak concentration is given by [8]:

$$n(R_P) = \frac{S}{\sqrt{2\pi}\,\sigma_P} \qquad (2.1)$$

where σ_P is the projected *straggle* or standard deviation, R_P the projected range (the depth where the concentration will be maximum), and S the dose. Both the silicon projected range and the projected straggle are tabulated for each impurity, as shown in FIGURE 2.3.

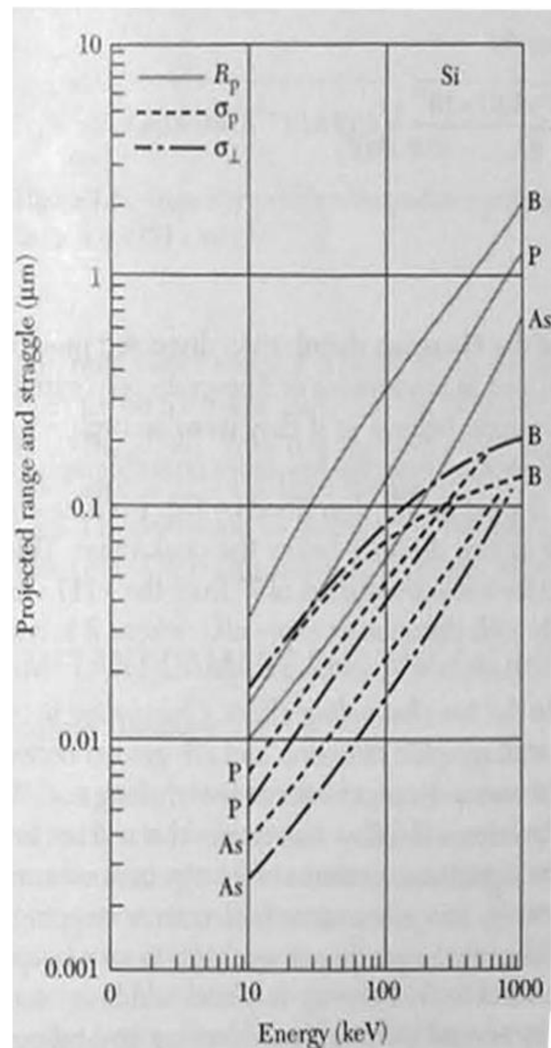


FIGURE 2.3 Boron, phosphorous and arsenic projected range, projected straggle and lateral straggle in silicon [8].

For instance, let us say it is desired to implant a boron P+ Well with a peak concentration of 4×10^{20} impurities/cm³ with a projected range of 0.1µm. Taking as reference FIGURE 2.3, we determine that the impurity beam needs an energy of 30KeV, which results in a projected straggle of 0.032µm. Then, isolating the dose at equation (2.1) results in 3.21×10^{15}

impurities/cm^2. The statement to implant such a well would be:

```
implant boron energy=30KeV dose=3.21e15 tilt=7 rotation=27 s.oxide=0.008
```

where the *tilt*, *rotation* and *s.oxide* parameters are related to the channeling effect mentioned before.

2.2.5 DIFFUSION PROCESSES

Several diffusion processes are performed during the simulation to, respectively: adjust the position of the junctions and the depth reached by the wells and de-homogenize the depletion regions. These processes are needed for more realistic simulations.

A diffusion process smooths the Gaussian dopant profile according to the diffusion time, extending the impurities through the device.

The syntax is as follows:

```
diffuse time=[minutes] temp=[ºC] inert
```

where the *inert* parameter is to specify that diffusion is done in the presence of nitrogen, that is, dry air. The time units can be either minutes, or seconds, or hours.

For instance, let us say that we have a P+ Well/N-Well junction, and we want to displace that junction by 0.1 μm. As the P+ Well has a higher impurity concentration by orders of magnitude, this well would 'invade' the N-Well, displacing the junction in the diffusion process by a certain length. This length is the diffusion length and is governed by the next equation:

$$L = \sqrt{D \cdot t} \tag{2.2}$$

where t is the diffusion time and D the diffusion coefficient. As the p+ well is composed of boron impurities, to know the time it would require for a diffusion process of 1050 ºC to displace the junction 0.1μm, we need to know the diffusion coefficient of Boron at that temperature. FIGURE 2.4 shows the diffusion coefficient versus the temperature for several materials in silicon. The diffusion coefficient, in this case, would be approximately $8 \times 10^{-14}\text{cm}^2/\text{s}$. The expected diffusion time would be 80 ns.

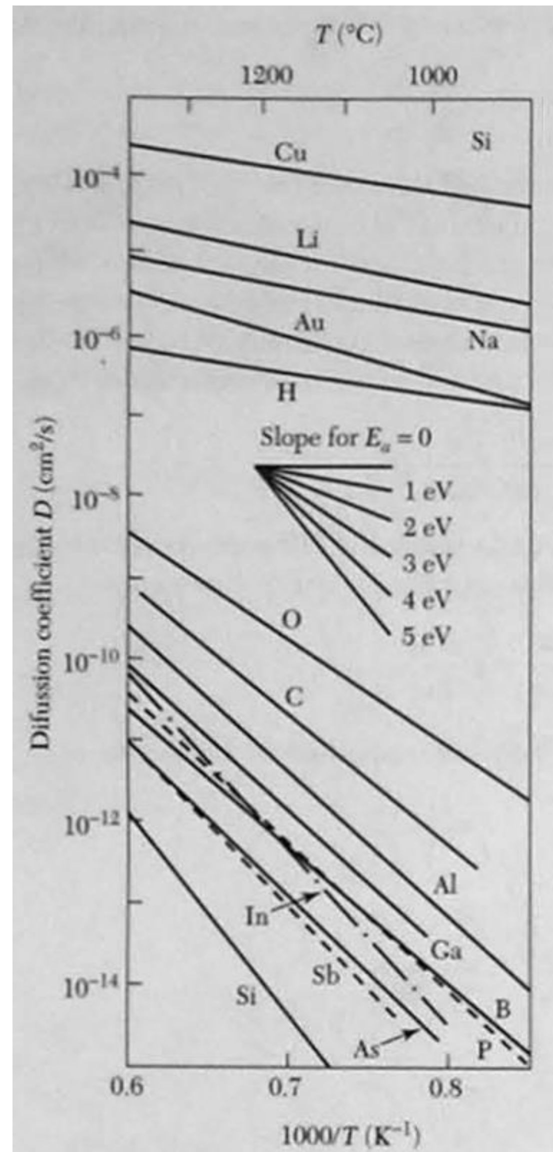


FIGURE 2.4 Diffusion coefficient vs temperature in silicon [8].

2.2.6 DECLARING ELECTRODES AND SAVING THE STRUCTURE

Before declaring the electrodes, the *debris* from the exposed oxide layer is removed. Then, a metal layer is deposited above the device (typically aluminum). Using a mask, we can remove the undesired part of this metal layer, and the remnants will be the device's electrodes.

Electrodes are declared using the following simple syntax:

```
electrode name=[name] x=[X coordinate of a point inside the electrode] y=[Y coordinate of a point inside the electrode]
```

where the *name* parameter may be some key values, like, for example, *gate*, *drain* or *cathode*.

To finalize, we save the structure in an *str* file with the next code:

```
structure outfile=my_structure.str
```

To summarize, the process to create a SPAD structure in ATHENA is as follows:

- First, the mesh is declared using at least two instances of the line statement:

```
line [dimension] location=[µm] spacing=[µm]
```

- Then, the substrate is initialized:

```
init silicon boron c.boron=6e14 orientation=100 cylindrical
```

- A shallow layer is deposited to avoid the channeling effect:

```
deposit oxide thick=0.008 divisions=4
```

- The different wells are implanted using a mask:

```
deposit barrier thick=0.03 divisions=3
etch [material] start x=[x1] y=[y1]
etch cont x=[x2] y=[y2]
etch cont x=[x3] y=[y3]
etch done x=[x4] y=[y4]
implant [impurity] energy=[KeV] dose=[particles/cm2] tilt=[degrees] rotation=[degrees]
```

- Several diffusion processes are simulated to adjust the junctions:

```
diffuse time=[minutes] temp=[ºC] inert
```

- And finally, the different electrodes are declared before saving the structure:

```
electrode name=[name] x=[X coordinate of a point inside the electrode] y=[Y coordinate of a point inside the electrode]
```

2.3 TCAD MODEL: DEVICE SIMULATION WITH ATLAS

After defining the device structure, the device capabilities must be tested using the device simulator ATLAS. Just like ATHENA, it is called from DECKBUILD like this:

```
go atlas
```

Next is to tell ATLAS which *structure file* (*.str) will be used. The structure file contains the physical device information previously defined with ATHENA. The following command loads the structure file:

```
mesh infile=my_structure.str cylindrical
```

where the *cylindrical* parameter will tell ATLAS to consider that the input device has cylindrical geometry, with its left edge acting as an axis of revolution. Now, ATLAS will have a physically plausible SPAD with the layer information of FIGURE 2.1. As they have already been declared in ATHENA, the electrode information is already embedded in the structure file; this device has three keywords associated with them that may be used to sweep their voltages: *anode*, *cathode*, and *substrate*.

Once the structure file is loaded, the next decision is the choice of models. ATLAS's physical models are grouped into five classes, namely:

- *carrier statistics,*
- *mobility,*
- *recombination,*
- *impact ionization*, and
- *tunneling.*

Bear in mind that Atlas only will statically test the device structure. That is, avalanche triggering or other dynamics behaviors like the dark count rate are not assessed. For that reason, out of the previous models, the tunneling ones will not be needed, as they are related to Band-To-Band tunneling processes that only can be observed under dynamic conditions.

2.3.1 CARRIER STATISTICS MODEL

Electrons in thermal equilibrium within a semiconductor lattice obey *Fermi-Dirac statistics* [5]. The probability that an available state with Energy E is occupied by an electron is:

$$f(E) = \frac{1}{1 + e^{\frac{E - E_F}{k_B T}}} \tag{2.3}$$

where E_F is Fermi energy, and k_B is the Boltzmann's constant. When $E - E_F$ “ $k_B T$, equation (2.3) can be approximated by:

$$f(E) = e^{\frac{E_F - E}{k_B T}} \tag{2.4}$$

Statistics based on equation (2.4) are known as *Boltzmann statistics* [5] and can be justified in semiconductor device theory if we assume non-degeneracy, as certain properties of very high doped material can be overseen in our case. Thus, the first model we will use would be *boltzmann*, which uses equation (2.4) and simplifies subsequent calculations [2].

As the doping level increases, bandgap separation decreases, as conduction and valence bands approximate each other in virtually the same amount. To emulate this effect, the bandgap narrowing model, *BGN,* based upon the work of Slotboom and De Graaf, is used [9]:

$$\Delta E_g = 9 \times 10^3 \ln \frac{N}{10^{17}\mathrm{cm}^{-3}} + \left[\left(\ln \frac{N}{10^{17}\mathrm{cm}^{-3}} \right)^2 + 0.5 \right]^{0.5} \quad [\mathrm{eV}] \tag{2.5}$$

where N is the Net doping in $\mathrm{dopants/cm^3}$.

2.3.2 MOBILITY MODELS

For mobility models, analytical expressions based upon the work of Caughey and Thomas [10] are used to specify doping and temperature-dependent low-field mobilities for electrons (μ_{n0}) and holes (μ_{p0}), respectively:

$$\mu_{n0} = \mu_{minn} \left(\frac{T}{T_0} \right)^{\alpha_n} + \frac{\mu_{maxn} \left(\frac{T}{T_0} \right)^{\beta_n} - \mu_{minn} \left(\frac{T}{T_0} \right)^{\alpha_n}}{1 + \left(\frac{T}{T_0} \right)^{\gamma_n} \left(\frac{N}{N_{crit}} \right)^{\delta_n}} \tag{2.6}$$

$$\mu_{p0} = \mu_{minp}\left(\frac{T}{T_0}\right)^{\alpha_p} + \frac{\mu_{maxp}\left(\frac{T}{T_0}\right)^{\beta_p} - \mu_{minp}\left(\frac{T}{T_0}\right)^{\alpha_p}}{1 + \left(\frac{T}{T_0}\right)^{\gamma_p}\left(\frac{N}{N_{crit}}\right)^{\delta_p}} \quad (2.7)$$

being μ_{min}, μ_{min}, α, β, γ, δ and N_{crit}, parameters obtained empirically, $T_0 = 300K$, the reference temperature, and T the temperature.

The simulation also includes the *fldmob* model to emulate the mobility dependence on the parallel electric field to model velocity saturation effects [10]:

$$\mu_n(E) = \mu_{n0}\left[\frac{1}{1 + \left(\frac{\mu_{n0}E}{V_{sn}}\right)^{\beta_n}}\right]^{\frac{1}{\beta_n}} \quad (2.8)$$

$$\mu_p(E) = \mu_{p0}\left[\frac{1}{1 + \left(\frac{\mu_{p0}E}{V_{sp}}\right)^{\beta_p}}\right]^{\frac{1}{\beta_p}} \quad (2.9)$$

Going deeper into this topic is unnecessary because of objectives of this Thesis. However, if the reader is interested in broadening his/her knowledge, it is recommended to carefully read the works in [2] and [10].

2.3.3 RECOMBINATION MODELS

Two complementary models are available here, namely the *consrh* and the *auger* models. The first one considers the *Schokley-Read-Hall theory* (SRH) of recombination through defects and the direct transitions of three carriers at high current densities or Auger recombination. Section 3.4.2, in Chapter 3, includes explanations regarding this model.

The *auger* model takes into account *Auger's effect*, where an electron may be expelled from the atom when another electron recombines in the valence band [11].

2.3.4 IMPACT IONIZATION MODEL

Finally, Selberherr's impact ionization model (*selb*) is used to calculate the avalanche's current. This model is a variation of the classical Chynoweth model to calculate the ionization rates [12]:

$$\alpha_n(E) = \alpha_{n\infty} \exp\left(\frac{-b_n}{|E|}\right)^{\beta_n} \quad (2.10)$$

$$\alpha_p(E) = \alpha_{p\infty} \exp\left(\frac{-b_p}{|E|}\right)^{\beta_p} \quad (2.11)$$

2.3.5 MODEL DECLARATION AND DEVICE TESTS

So to define our models, we write the next statements:

```
models boltzmann bgn analytic fldmob consrh auger
impact selb
```

The next step is to specify the method of calculation of ATLAS. The next statement makes the subsequent calculation likely to converge. This statement is standard in ATLAS and will not be further discussed. Refer to [2] for further reading.

```
method newton autonr carr=2 climit=1e-3 maxtraps=20
```

Once the models are chosen, the device's main characteristics must be checked to extract data needed for calculations. These checks include:

- Measuring the electron and hole concentration at the primary junction to be able to determine the electron injection during avalanches
- The electric field in the primary junction, the secondary junction, and different sections of the guard ring
- Electron and hole avalanche probability when measuring in *Geiger mode* (a model designed solely for this purpose, *geiger*, would be used instead of *consrh*).

The statement *probe*:

```
probe name=[my_measurement] x=[x coord.] y=[y coord] [keyword]
```

performs these checkings, where *keyword* is the parameter that is needed to be measured (for example, *field* for electric field). Now that we have our probes defined, it is necessary to command ATLAS to store our probes data and terminals parameters in a text file:

```
log outfile=my_data.log
```

Now the *simulation run* can be defined. For example, measuring quantum efficiency in a transient simulation may be interesting. To be able to do so, it is needed first to illuminate our device by declaring a light beam statement:

```
beam num=1 x.origin=[x beam origin] y.origin=[y beam origin] xmin=[beam left
border] xmax=[beam right border] angle=[beam angle in degrees]
wavelength=[um].
```

Then, the device is given initial conditions, and the cathode is ramped. To measure quantum efficiency, the device does not need to be ramped over breakdown. A small inverse polarization would do:

```
solve init
solve vcathode=0.1 vfinal=1.0 vstep=0.1 name=cathode
```

Finally, we ramp our light beam:

```
solve b1=[W/cm2] ramp.lit ramptime=[seconds] dt=[temporal steps]
tstop=[temporal stop] vcathode=1.0
```

With this, we can measure the external quantum efficiency, *EQE*, by dividing the available photocurrent by the source photocurrent. The source photocurrent is the rate of photons incident on the device expressed as current density [2]:

$$I_S = q\frac{B_n\lambda}{hc}W_t \qquad (2.12)$$

where W_t is the beam width and B_n the intensity of the beam number n. The available photocurrent is [2]:

$$I_S = q\frac{B_n \lambda}{hc}\sum_{i=1}^{N_R} W_R \int_0^{Y_i} P_i \alpha_i e^{-\alpha_i y} dy \tag{2.13}$$

being P_i the attenuation before the start of the ray due to non-unity transmission coefficients and absorption prior to ray start, α_i the absorption coefficient of the material the ray is traversing and Y_i is the ray path in that material.

Another parameter, the *internal quantum efficiency*, *IQE*, for a terminal, can be calculated by dividing that terminal current by the available photocurrent. The *IQE* represents how much of the photocurrent generated by the photons is finally collected in a given terminal. This parameter does not make much sense in a SPAD, which is ready to collect and amplify the current generated by a single photon in theory, but it is of much importance in other devices, such as solar cells. To measure the efficiency of a SPAD, we use the *EQE*, which represents the total current generated by the incident photons on the device.

As work is done with single-photon avalanche diodes, finding the breakdown voltage or the parasitic capacities between terminals may be interesting too. In this case, it is needed to ramp the cathode in alternating current and find at which point the electric field becomes constant:

```
solve vcathode=0.1 vfinal=15.0 vstep=0.1 name=cathode ac frequency=1e6
direction
```

Results can be checked with TONYPLOT, a tool to visualize structures and log files when the simulations are done.

2.4 SPAD SELECTION AND MODEL OVERVIEW

SPAD structures addressed for behavioral modeling in this Thesis are selected based on relevance considerations. First, the so-called G-family (FIGURE 2.5a), based on Rocha's work [3], is simple yet suitable for applications. Second, the V12 SPAD (FIGURE 2.5b), based on Richardson's work [13], is rooted in industrial deployments recently made by ST. We will see that the latter is an improvement over the former. Third, the A12 structure (FIGURE 2.5c), based on Webster's SPAD [14], is an example of a SPAD that, due to its layer configuration and despite outstanding capabilities, presents some unexpected problems that make its usability challenging. Finally, The P12 structure in FIGURE 2.5d, a new SPAD structure proposed in this work, uses the concept of perimeter guarding to prevent edge breakdown. The different layer configurations will be described in the next section.

2.4.1 SPAD MODEL OVERVIEW: G-FAMILY MODELS

G-family SPADs are based on Rocha's SPAD (FIGURE 2.5a) [3] and consist of a single cylindrical SPAD model with four different diameters of active area: G04, G08, G12 (FIGURE 2.1), and G16, which have diameters of 4µm, 8µm, 12µm, and 16µm respectively. Said models consist of a Deep-N-Well implanted in a P-doped epitaxial layer. Within that Deep-N-Well, a shallow P+ layer is surrounded by a P-Well, that acts as a guard ring to avoid premature edge breakdown. Therefore, this model has two junctions: a P+/Deep-N-Well junction, also called primary junction, and a Deep-N-Well/P-Substrate junction, also called secondary junction. This junction prevents avalanche carriers from being scattered across the P-Substrate, where they could drift away to other nearby SPAD, generating avalanche pulses as well, thus generating undesired dark counts.

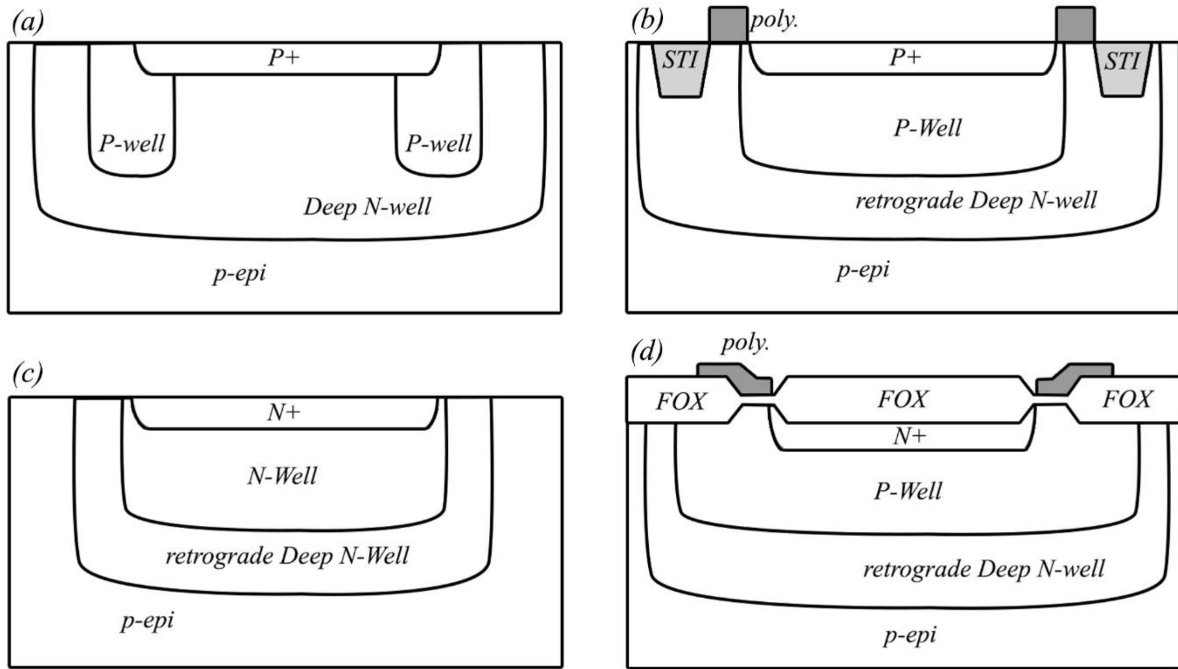


FIGURE 2.5 Models selected for this work: (a) G12 model. (b) V12 model. (c) A12 model. (d) P12 model.

When the SPAD is reverse biased over the breakdown voltage of the primary junction, V_{B1}, generated carriers within the depletion region can ignite avalanches. It is considered that the breakdown voltage of the primary junction is that of the active area region, as the guard ring has a higher breakdown voltage. To put the SPAD primary junction on reverse, usually, the cathode is positively biased. This put the secondary junction on reverse as well, but due to the lower doping profile of this junction, the electric field is lower than that of the primary junction, which leads to a higher breaking voltage, V_{B2}. So normally, the SPAD operation limit, when reverse biased, is within a range of $V_{B2} - V_{B1}$. FIGURE 2.6 shows the electric field across the device at 1.25 of excess bias voltage. White lines show where the junctions are located, while black ones delimit the extent of depletion regions for both junctions. The electric field is

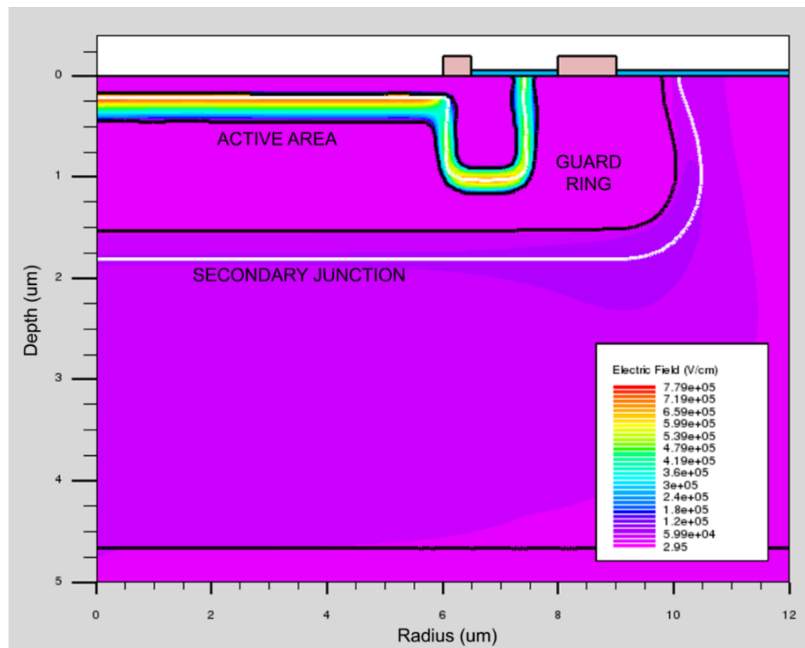


FIGURE 2.6 Electric field on the G12 model at 1.25V of excess bias voltage.

confined within the limits of the depletion region, which widens with the increasing reverse bias.

FIGURE 2.7 compares the electric field across the active area and guard ring of the primary and secondary junctions. The more abrupt the slope of the electric field, the more doped is that side of the junction.

G-family SPADs are illustrative of the capabilities of TCAD regarding exploring the impact of size *scaling*. Other simple changes are also possible, like a wider guard area ring or changes in the doping profiles to alter the primary junction depth, and only require small changes in the VERILOG-A model. Complex changes like new structures within the device that requires profound changes in the VERILOG-A model are also possible.

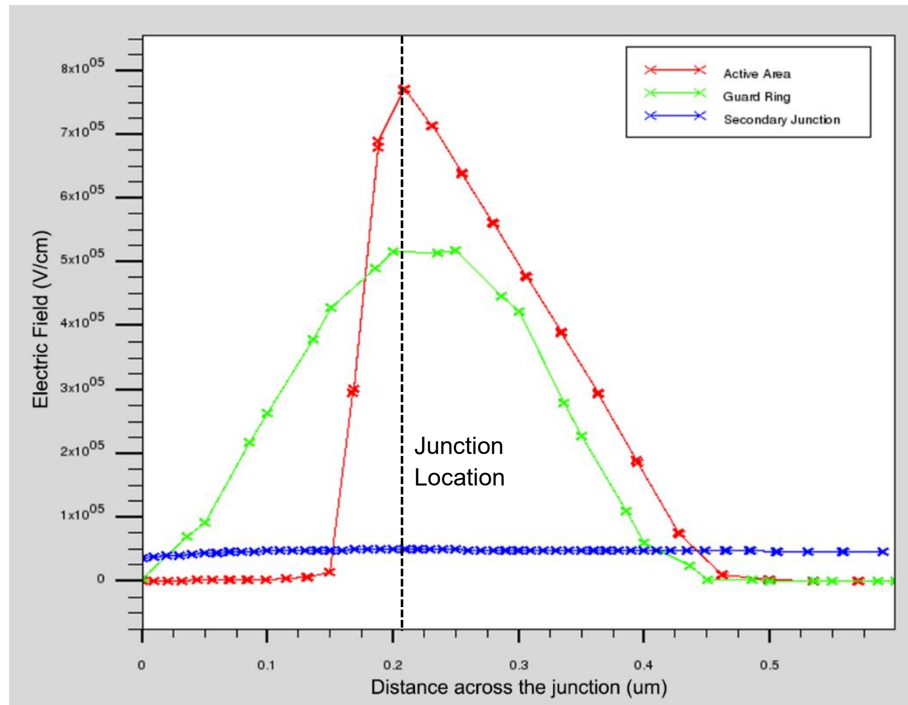


FIGURE 2.7 Electric field comparison of several junction parts.

2.4.2 SPAD MODEL OVERVIEW: THE V12 MODEL

The V12 model also has an active area region diameter of 12µm, as shown in FIGURE 2.8. This model is based on Richardson's SPAD (FIGURE 2.5b) [13]. This model also consists of a Deep-N-Well implanted in a P-Substrate, but unlike the G family models, it has a central P-Well, so that the Deep-N-Well and the P-Well form the primary junction. The main advantage of this structure is the lower electric field of the primary junction, as the doping profile of the P-Well is several orders of magnitude lower than that of the P+ Well. It is so low that it manages to eliminate the dark counts due to the band-to-band tunneling and the effects of trap-assisted tunneling that also have a strong electric field dependence. It also diminishes the avalanche current, making the SPAD less prone to self-heating.

Using the type of central well does not eliminate the need for a guard ring to avoid premature edge breakdown. To that end, a virtual guard ring has been developed. A virtual guard ring is formed by partially blocking the ion implantation of the Deep-N-Well in the desired zone, where we will have a retrograde Deep-N-Well. The retrograde Deep-N-Well has a decreasing doping profile as it approaches the surface, which is about two orders of magnitude lower. Thus, the breakdown voltage will increase and be maximum near the surface, forming an effective guard ring. Then, only the deep plateau of the junction will get over breakdown when the diode

is reverse biased. FIGURE 2.9 shows an electric field comparison with the G12 model, where it can be appreciated how the junction electric field is about 15% lower. A detail of the electric field profile is shown in FIGURE 2.10, where the decreasing electric field of the virtual guard ring can be appreciated.

As shown in Chapter 5, another of the advantages of this model is that the junction is deeper than that of the G12 model, which makes it better to sense longer wavelengths.

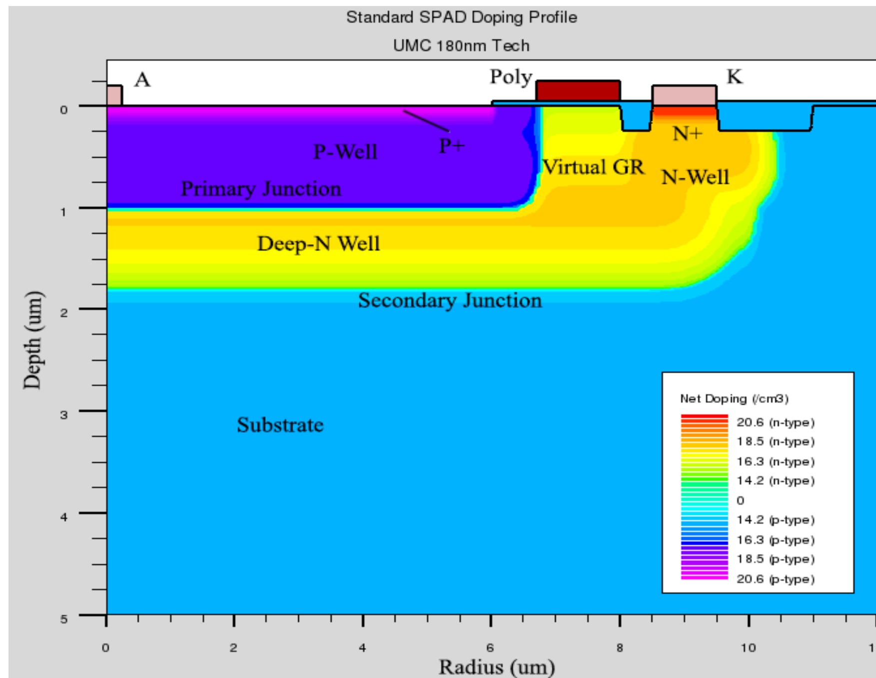


FIGURE 2.8 Doping profile of the V12 model. The left border acts as axis of revolution.

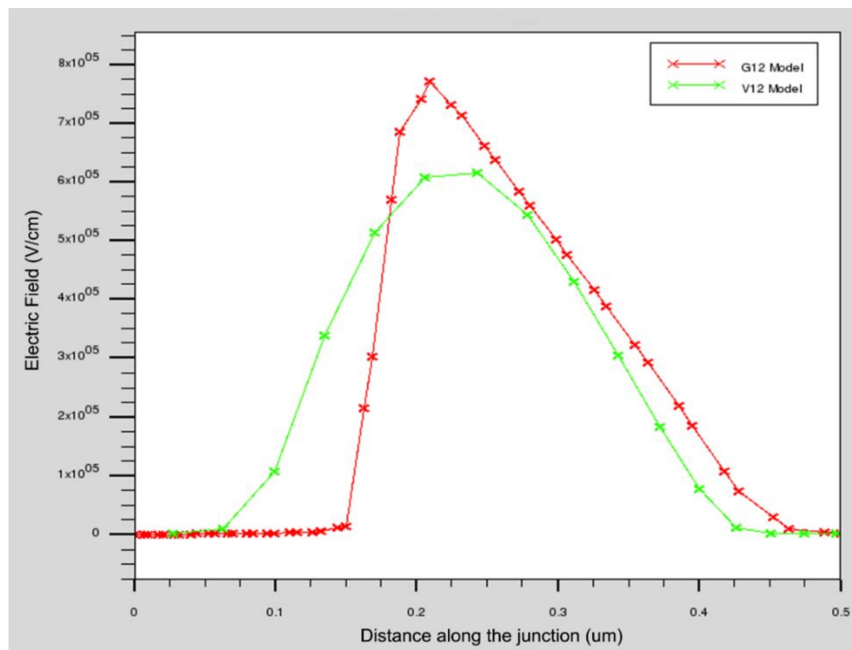


FIGURE 2.9 Electric field at the primary junction for the G12 and the V12 model.

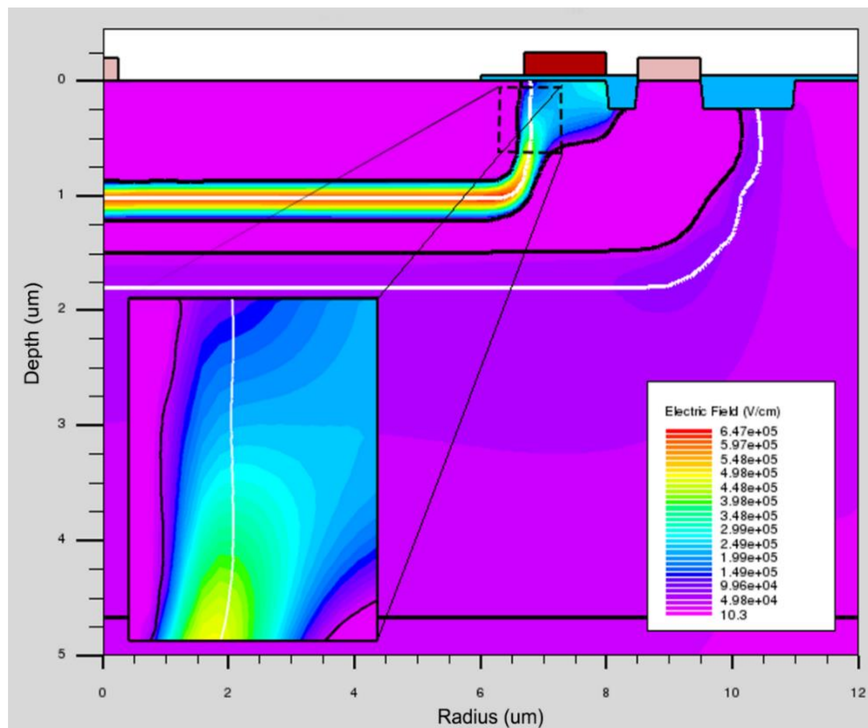


FIGURE 2.10 Detail of the electric field at the edge of the central P-Well region of the V12 model.

2.4.3 SPAD MODEL OVERVIEW: THE A12 MODEL

Webster's structure has also used the virtual guard ring concept [14] (FIGURE 2.5c). This device takes advantage of the decreasing nature of the doping profile of the epitaxial layer over the substrate. That way, in a similar fashion to the V12 model, a virtual guard ring can be formed. This structure requires the growing of an epitaxial layer that is simulated by invoking the command *epitaxy* after the substrate is defined, feeding Athena with a time to form and the temperature of the process, which will determine the gradient of the epitaxial layer, and the thickness of this layer:

```
epitaxy time=[minutes] temperature=[ºC] thickness=[µm]
```

The doping profile of the A12 model can be observed in FIGURE 2.11. The V12 and the A12 models offer the possibility of further scaling down the SPADs and the possibility of enhancing the detection of longer wavelengths, as the junctions are located deeper in the bulk. Thus, the gain in quantum efficiency in the last two configurations at those wavelengths should be considerable, given the difference in junction depth. Later in this Thesis, Chapter 7 overviews the capabilities of the A12 model.

2.4.4 SPAD MODEL OVERVIEW: THE P12 MODEL

The P12 model is based on a new structure developed in this work (FIGURE 2.5d). It consists of a stacked triple sensing junction SPAD (two acting as sensing junctions, one as an insulator) modeled by TCAD simulations in standard 180-nm technology. The main design novelty is that the primary junction is prevented from edge breakdown in a perimeter-gating fashion [15]. The device is similar to the P-Well device developed in [13], which used a virtual guard ring to avoid a peripheral breakdown in the primary junction. However, in our device, we added an N+ Well within the P-Well, thus creating a third junction that is prevented from edge junction breakdown with the perimeter-gating mentioned above.

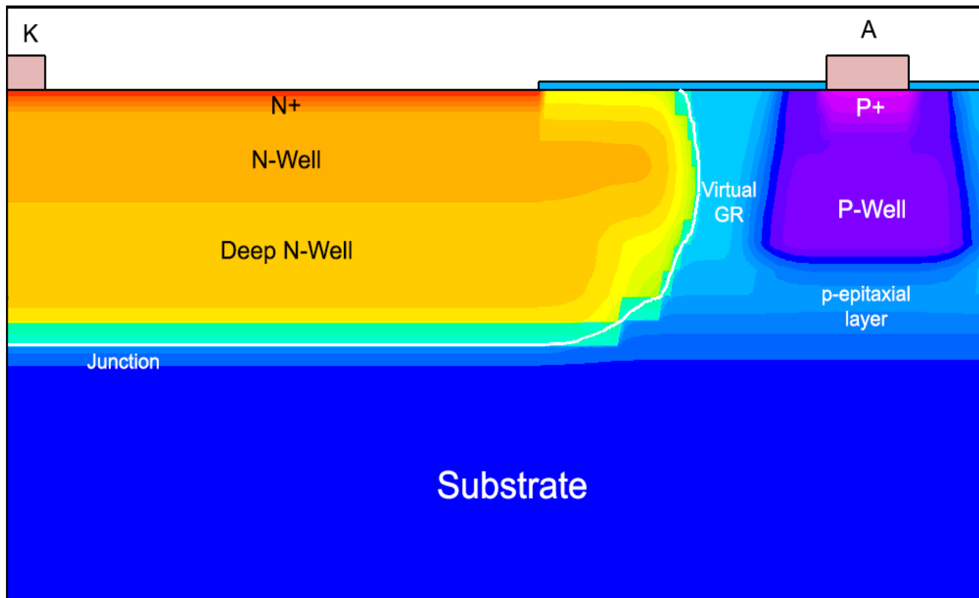


FIGURE 2.11 Doping profile of the A12 model. The white line represents the junction. The left border acts as axis of revolution.

FIGURE 2.12 shows The SPAD structure. The primary junction is formed by the N+ Well and the P-Well, where a positive potential at the Gate electrode avoids edge breakdown. The P-Well and the Deep-N-Well form the secondary junction. The virtual guard ring formed by blocking the peripheral N-Well formation ensures a retrograde doping profile, making the electric field in the junction lower so that it needs a greater potential to enter breakdown. That way, only the deepest part will trigger avalanches. The tertiary junction, formed by the Deep-N-Well and the substrate, would serve as an insulator when a positive potential is applied to the second cathode, being, therefore, reverse biased. The capabilities of this model will be discussed in Appendix II.

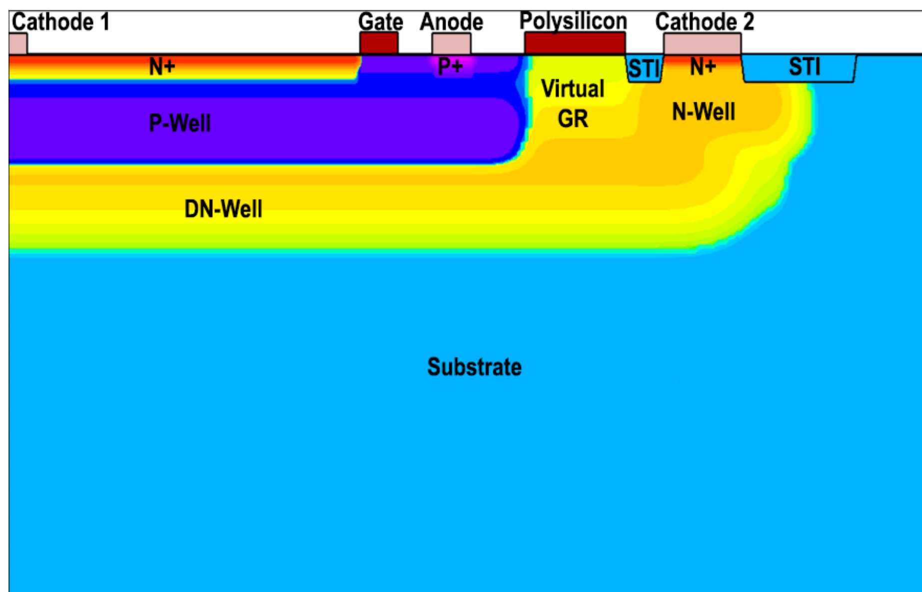


FIGURE 2.12 Doping pofile of the P12 model. Left border acts as an axis of revolution.

2.5 REFERENCES

[1] *Athena User's Manual*. Silvaco Inc., Santa Clara, CA, 2015.

[2] *Atlas User's Manual: Device Simulation Software*. Silvaco Inc., Santa Clara, CA, 2016.

[3] A. Rochas, A. R. Pauchard, P. A. Besse, D. Pantic, Z. Prijic, and R. S. Popovic, "Low-noise Silicon Avalanche Photodiode Fabricated in Conventional CMOS Technologies." *IEEE Transactions on Electron Devices*, vol. 49, no. 3, pp. 387-394, March 2002.

[4] I. Vornicu, R. Carmona-Galán, B. Pérez-Verdú and Á. Rodríguez-Vázquez, "Compact CMOS Active Quenching/Recharge Circuits for SPAD Arrays." *Int. J. of Circuit Theory and Application*, Vol. 44, No. 4, pp. 917-928, Apr. 2015.

[5] S. M. Sze and K. K. Ng, Physics of Semiconductor Devices, 3rd ed. New York, NY, USA: Wiley, 2007.

[6] J. Lindhard, M. Scharff, and H.E. Schiott. "Range Concepts and Heavy Ion Ranges." *Kgl. Dan. Vid. Selsk. Mat.-fys. Medd.*, v. 33, 1963.

[7] W. K. Hofker, D. P. Oosthoek, N. J. Koeman and H. A. M. de grefte, "Concentration Profiles of Boron Implantations in Amorphous and Polycrystalline Silicon." *Radiation Effects*, Vol. 24, No.4, pp. 223-231, 1975.

[8] G.S. May and S.M. Sze, *Fundamentals of Semiconductor Fabrication*. Wiley, 2004. Pages 108 and 129.

[9] J. W. Slotboom and H. C. De Graaf, "Measurements of Bandgap Narrowing in Silicon Bipolar Transistors." *Solid State Electronics*, vol. 19, pp. 857-862, 1976.

[10] D. M. Caughey and R. E. Thomas. "Carrier Mobilities in Silicon Empirically Related to Doping and Field." *Proceedings of the IEEE*, Vol. 55, pp 2192-2193, 1967.

[11] S. Selberherr, *Analysis and Simulation of Semiconductor Devices*. Wien, New York: Springer-Verlag, 1984.

[12] R. Van Overstraeten and H. Deman, "Measurement of the Ionization Rates in Diffused Silicon p-n Junctions." *Solid-State Electronics* 13 (1970): 583-608.

[13] 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 Transactions on Electron Devices*, Vol. 58, no. 7, pp. 2028-2035, July 2011.

[14] E. A. G. Webster, L. A. Grant, and R. K. Henderson, "A High-Performance Single-Photon Avalanche Diode in 130-nm CMOS Imaging Technology." *IEEE Electron Device Letters*, Vol. 33, no. 11, pp. 1589-1591, Nov. 2012.

[15] J. Gu, M. Habib Ullah Habib, and N. McFarlane, "Perimeter-Gated Single-Photon Avalanche Diodes: An Information Theoretic Assessment." *IEEE Photonics Technology Letters*, Vol. 28, no. 6, Mar. 2016.

CHAPTER 3
PHYSICAL FOUNDATIONS OF A BASIC VERILOG-A MODEL

3.1 INTRODUCTION

Chapter 2 overviews the use of TCAD tools to create SPAD models from physical data. The outcome of these analyses supports the development of VERILOG-A models that contain all the data provided by TCAD simulations and the physical conditions that describe the operation of a SPAD. This model-building procedure relies on descriptions of device physics. However, this Thesis does not aim to explain the foundations of physical processes like, for instance, *Impact Ionization*, but rather how the macroscopic variables and structure within the device affect its performance.

Descriptions in this chapter mostly use the G-family (see Section 2.4.1 in Chapter 2) to describe the physics of single-photon avalanches. The rationale for using this family includes the following: i) its layer configuration and capabilities are widely known; ii) there is abundant literature about it; iii) the base layer configuration has been used for 30 years now [1].

Other models addressed in the Thesis, namely V12, A12, and P12, will be covered in Chapters 5, 7, and Appendix II, respectively. The V12 model [2] illustrates how layers of a standard technology can be engineered to improve the performance of SPADs. Working with standard technologies is appealing because of cost and affordability considerations.

The A12 model, based on the Webster SPAD structure [3], also employs standard technology layers and arranges them to achieve larger sensitivity in the near-infrared region of the spectrum. This arrangement increases the noise from neighboring devices and produces crosstalk. Thus, it requires non-standard technology to isolate every SPAD, which raises serious doubts about its usability in large arrays.

Finally, the P12 model, a new SPAD structure devised in this Thesis, is a representative example of how to apply this methodology for SPAD improvement.

3.2 SPAD PHYSICAL PARAMETERS

FIGURE 3.1 shows the G12 SPAD structure, the workbench used throughout this chapter to describe the physical properties included in the VERILOG-A model. This device is a 12µm diameter active area device with a guard ring of low-doped P-well material, or T-Well, around the central P+/Deep-N-Well breakdown region. Likewise, the Deep-N-well is implanted in an epitaxial layer of a very low doped p material. Therefore, this device has two junctions:

- the P+/Deep-N-Well junction or primary junction;
- and the Deep-N-Well/substrate junction or secondary junction.

Let us first explore the behavior of these junctions with the bias voltage and the electric field.

The first step to studying a SPAD is biasing it. The primary and secondary junctions are both reverse-biased when the *cathode* (K) is biased with a positive potential and the *anode* (A) is connected to the ground. They may reach the breakdown voltage if the positive potential is high enough. Different cases arise depending on the nature of the junctions′ doping profiles – addressed in the following sections.

3.2.1 ABRUPT JUNCTIONS

The following general equation describes the breakdown voltage, V_B, of both junctions for *one-sided abrupt* junctions [4]:

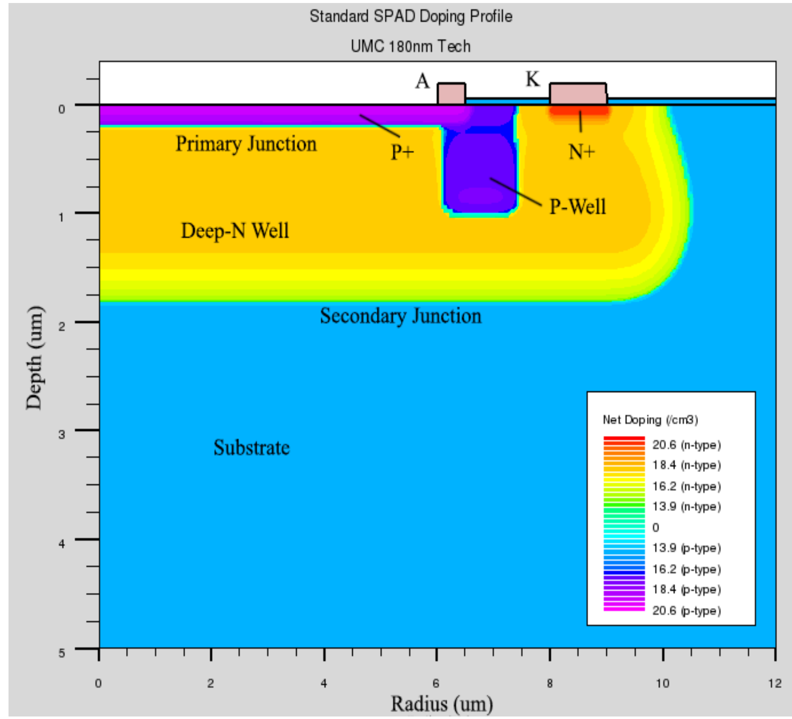


FIGURE 3.1 Final device structure of the G12 SPAD model developed with Athena.

$$V_B = \frac{E_c W_{Dm}}{2} = \frac{\varepsilon_s E_c^2}{2qN} \tag{3.1}$$

being E_c the critical electric field in the junction, W_{Dm} the width of the depletion region, and N the ionized background impurity concentration on the lightly doped side.

According to this equation, the primary junction breakdown voltage should be smaller than that of the secondary junction due to the former having an impurity concentration higher than the latter ($2 \times 10^{17}\text{cm}^{-3}$ and $6 \times 10^{14}\text{cm}^{-3}$ respectively). To be precise, the G12 model has a breakdown voltage of $V_{B1} = 10.3\text{V}$ in the active area region of the primary junction and a breakdown voltage of $V_{B2} \sim 25 - 30\text{V}$ in the secondary junction. Then, the operation within this range is mandatory for the device to work correctly.

The junction's electric field strongly depends on impurity concentration and is almost constant when the SPAD is reverse-biased beyond the breakdown voltage; this quasi-constant field is known as the *critical* electric field. The higher the impurity concentration, the higher the electric field. For abrupt junctions [4]:

$$E_c = \frac{4 \times 10^5}{1 - \frac{1}{3}\log_{10}\left(\frac{N}{10^{16}\text{cm}^{-3}}\right)} \ [\text{Vcm}^{-1}] \tag{3.2}$$

where N, the impurity concentration, is given in cm^{-3}.

3.2.2 NON-ABRUPT JUNCTIONS

Actual junctions are neither abrupt nor even *linearly* graded. Indeed, linearly graded junctions are only a suitable approximation when the extension of the depletion layer is large enough [5], as FIGURE 3.2 illustrates. Consequently, the electric field within actual, small junctions is not constant.

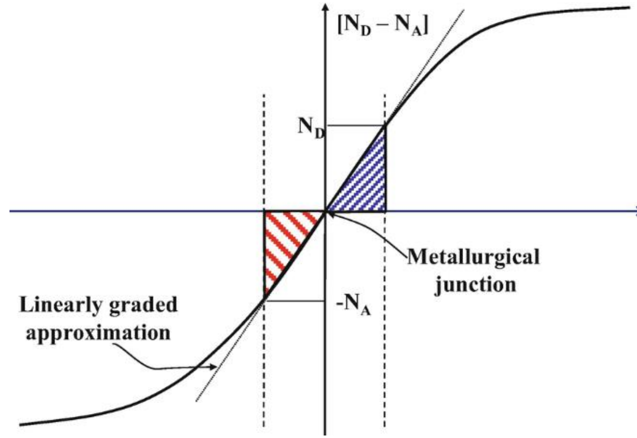


FIGURE 3.2 Linearly graded approximation to a real like junction [7].

This work approximates the value of the effective depletion layer width, W_e, by using the maximum electric field across the junction, E_m, extracted from the TCAD model of the device [6]:

$$W_e = \frac{2\psi}{E_m} \tag{3.3}$$

where ψ is the *built-in potential* [4]:

$$\psi = \frac{k_B T}{q} \ln\left(\frac{N_D N_A}{n_i^2}\right) \tag{3.4}$$

being N_D and N_A the donor and acceptor impurities and n_i the silicon intrinsic concentration. The impurities concentrations are taken from the TCAD model, and the intrinsic concentration will be regarded later in this chapter.

The intensity of the electric field is higher at the edges of the active area region, thus causing the depletion layer to narrow due to a higher curvature radius. As a result, the *breakdown voltage* decreases in these regions, triggering *Premature Edge Breakdowns* (PEB). These undesired edge effects, which make the SPAD only sensitive on the edge of the junction, are overcome in the G12 model with the guard ring implantation (FIGURE 3.3). The guard ring junction has a lower impurity concentration than the p+ layer, thus increasing the breakdown voltage in this area and precluding PEBs.

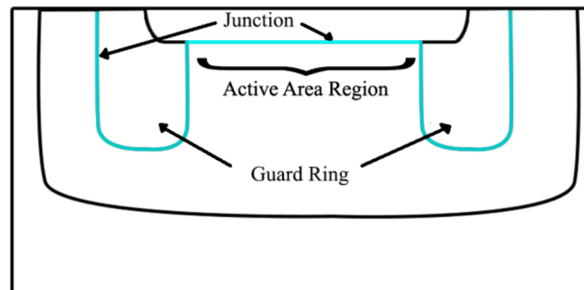


FIGURE 3.3 SPAD diagram showing the guard ring and the junction. The portion of the junction that is sensitive to illumination and enters breakdown is highlighted, and it is called the active area region. The guard ring protects the edge of this area from entering in a breakdown before the rest of the junction.

FIGURE 3.4 shows the variation of the primary junction electric field and the SPAD current with the reverse bias voltage. This graph shows how TCAD simulation gives a detailed electric field profile, unlike other work that just offers constant or average electric field values.

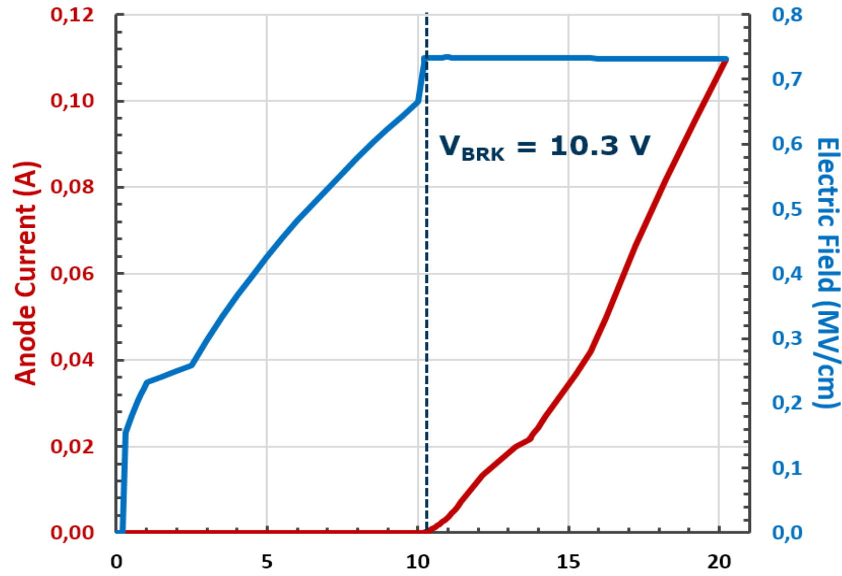


FIGURE 3.4 Anode current and primary junction electric field against reverse bias.

3.3 SPAD ANALYTICAL MODEL

The SPAD analytical model in this Thesis (see FIGURE 3.5) displays the following major differences versus the previously proposed Giustolici's [8] model:

- It takes into account the second junction between the deep-n-well and the substrate; hence the saturation current from that junction and the junction capacitance are added to the new model.
- Unlike previous works, both primary and secondary junctions are considered in order to describe the DC current-to-voltage relation.

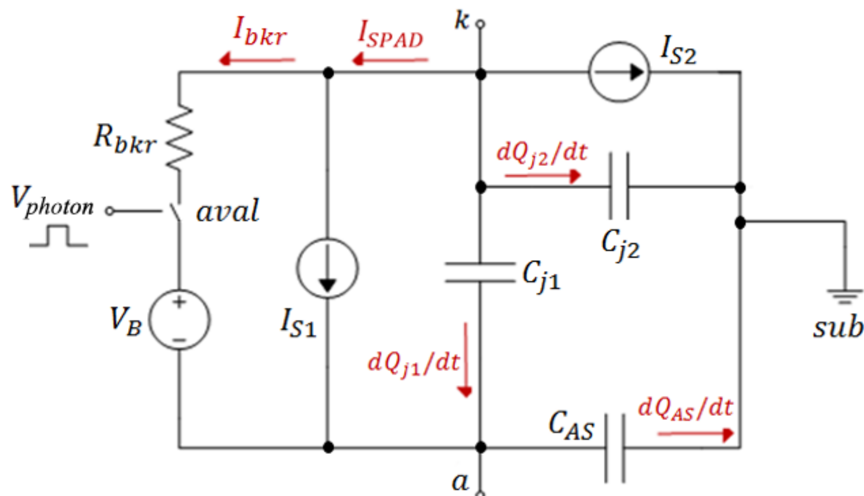


FIGURE 3.5 SPAD analytical model.

The current in a SPAD, I_{SPAD}, is dominated by the saturation current through the primary junction when there is no avalanche, I_{S1}, and for the avalanche current when the avalanche is being triggered, I_{bkr}. To solve the convergence problems in the boundary between them, a *pseudo-max* function is used to avoid calculations problems in VERILOG-A [8]:

$$\max(x, y) \approx K \cdot \ln\left(e^{\frac{x}{K}} + e^{\frac{y}{K}}\right) \tag{3.5}$$

A pseudo-max function is a composite of two functions approximately equal to the greater function. If $x \approx y$, then $\max(x, y)$ strongly depends on both functions. The advantage of using this function is that it is fully differentiable and solves any calculation issue that it may happen. Through some calculations, we can reach the following formula for I_{SPAD}, where the pseudo-max function is used in the avalanche term ($x = 0, y = V_E, K = V_n$):

$$I_{SPAD} = \begin{cases} I_{S1} & \text{no avalanche} \\ I_{S1} + \dfrac{V_n}{R_{brk}} \ln\left(1 + e^{\frac{V_E}{V_n}}\right) & \text{avalanche} \end{cases} \tag{3.6}$$

being R_{bkr} the breakdown series resistance, $V_E = V_{ka} - V_B$ the excess bias voltage, and $V_n \approx 50\text{mV}$ a normalization voltage that allows the resulting current to adjust to the TCAD simulation results. As $V_E > 0$ while the SPAD is in avalanche, $\max(0, V_E) \approx V_E$. Notice, that is not enough that $V_E > 0$ for the SPAD to trigger an avalanche. The SPAD also needs that a charge carrier arrives to its depletion region.

TCAD simulations show that the breakdown series resistance is related to the active area region radius (r):

$$R_{brk}(r) = R_0\, r^{-R_1} \tag{3.7}$$

being R_0 ($\approx 313.464\Omega \cdot \mu\text{m}^{R1}$) and R_1 (≈ 0.818) device-dependent constants, and r given in µm.

The dynamic (AC) behavior of the SPAD can be determined by three contributions: the charges being stored in the primary and secondary depletion regions (Q_{j1} and Q_{j2} respectively) and the anode-to-substrate (Q_{AS}) stray capacitor. They are determined by the following equations [9]:

$$Q_{j1} = A_1 \frac{\psi_1\, C_{01}}{1 - m_j}\left(1 + \frac{V_d}{\psi}\right)^{1-m_j} + A_g \frac{\psi_g\, C_{0g}}{1 - m_j}\left(1 + \frac{V_d}{\psi}\right)^{1-m_j} \tag{3.8}$$

$$Q_{j2} = A_2 \frac{\psi_2\, C_{02}}{1 - m_j}\left(1 + \frac{V_d}{\psi}\right)^{1-m_j} \tag{3.9}$$

$$Q_{AS} = C_{AS}\, V_A \tag{3.10}$$

being A the area, ψ the built-in potential (equation (3.4)), C_0 the zero-bias capacitance per unit area, $m_j = 0.5$ the junction grading coefficient, V_D the voltage across the diode and V_A the voltage in the anode terminal. The '1', 'g' and '2' subscripts refer to the primary junction, the guard ring, and the secondary junction, respectively. Equation (3.8) considers the contributions

of the active area region and guard ring to the primary junction. The zero-bias capacitance, C_0 (TABLE 3.1), is given by:

$$C_0 = \frac{\varepsilon_s}{W_0} \tag{3.11}$$

where W_0 the effective depletion layer width at zero bias.

Therefore, the current in the diode terminals is:

$$I(K) = I_{SPAD} + I_{S2} + \frac{dQ_{j1}}{dt} + \frac{dQ_{j2}}{dt} \tag{3.12}$$

$$I(A) = -I_{SPAD} - \frac{dQ_{j1}}{dt} + \frac{dQ_{AS}}{dt} \tag{3.13}$$

where I_{S2} is the saturation current of the secondary junction.

TABLE 3.1 Zero-bias capacitance extracted from TCAD simulations.

	$[\mathbf{F \cdot cm^{-2}}]$
C_{01}	7.983×10^{-8}
C_{0g}	3.681×10^{-8}
C_{02}	6.051×10^{-9}

3.4 AVALANCHE TRIGGERING EVENTS

The events that may trigger SPAD avalanches are separated into two groups:

- the first is the arrival of photon-induced carriers to depletion regions, caused either by direct impact or diffusion, which are intended to be detected;
- the second group includes all the sources of non-desirable avalanches of *dark counts* generated by the device.

FIGURE 3.6 illustrates the dark count sources, subdivided into *primary* and *secondary* dark counts. **Primary dark** counts include:

(a) *Bulk thermal generation* and diffusion towards the depletion region. They are insignificant due to high recombination rates for the minority carriers outside the depletion region, thus having a shallow generation rate.

(b) *Band-to-band thermal generation* is also unlikely due to the large silicon bandgap.

(c) *Trap-Assisted Thermal Generation* (TATG) is a major source of noise in SPADs. Defects in the silicon crystal create new energy levels between the conduction and valence band enabling carriers to get excited.

(d) *Trap-Assisted Tunnelling* (TAT) happens when carriers get excited thermally to the new energy levels created by the defects. Then the strong electric field in the depletion region makes them tunnel the rest of the potential barrier.

(e) *Band-To-Band Tunnelling* (BTBT), which is the mechanism that allows the charge carriers to tunnel the potential barrier between the valence and conduction band without thermal assistance due to a strong enough electric field.

Secondary dark counts have two contributions:

- The first one is the *after-pulsing* effect (f), in which traps near the valence and conduction band, also called deep-level traps, hold carriers trapped after an avalanche and, after a variable time, are released and have the possibility of triggering a new avalanche.
- The second one is the *crosstalk*, in which charge carriers or photons generated in neighbouring SPADs trigger spurious avalanches.

Triggering events are *statistical* phenomena, the most relevant among which are explained in the following subsections. Regarding crosstalk, and because it depends upon the behavior of other SPADs, it is not included in this chapter but addressed in Chapter 7.

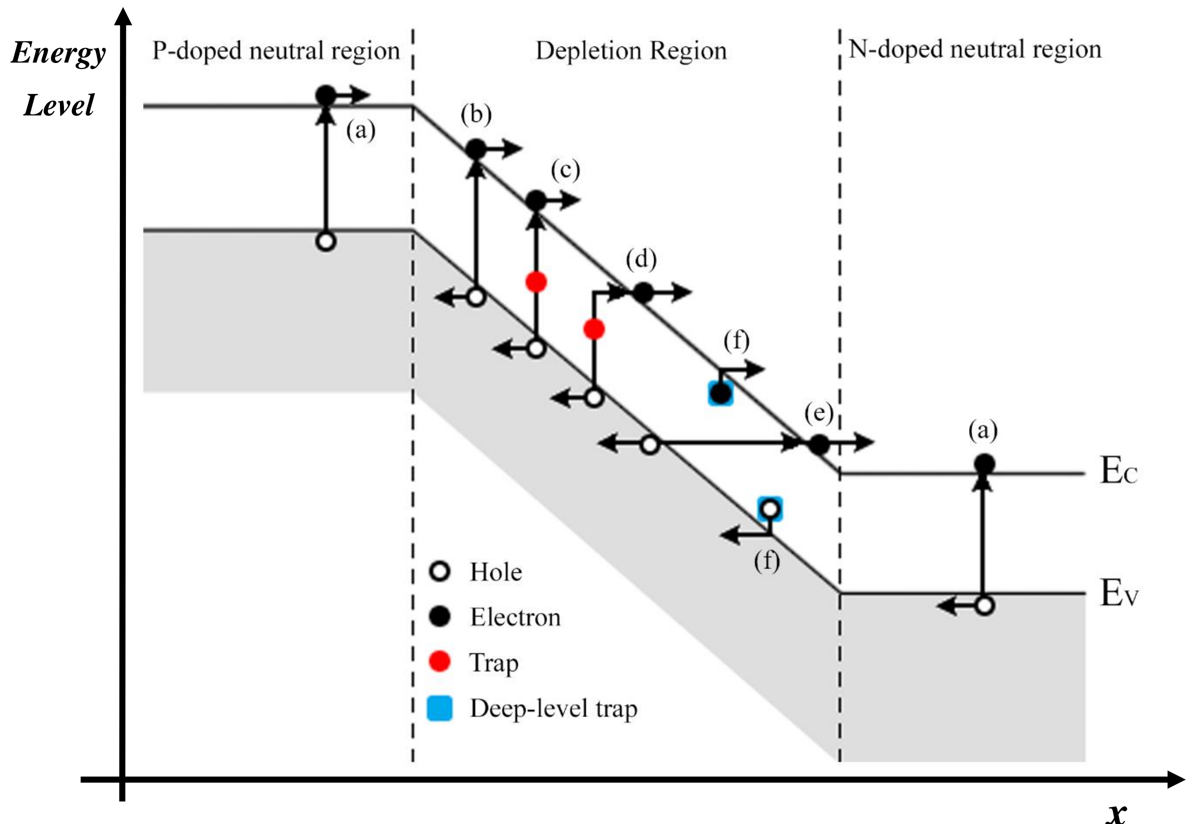


FIGURE 3.6 Schematic of several sources of dark counts; (a) Bulk thermal generation and diffusion; (b) Band-to-band thermal generation; (c) Trap-assisted thermal generation; (d) Trap-assisted tunneling; (e) Band-to-band tunneling; (f) After-pulsing. (This figure was handcrafted by this Thesis' authors from theory present in [4])

3.4.1 PHOTON ARRIVAL

The probability of a photon arriving at the space charge region and triggering an avalanche is quantified by the *Photon Detection Efficiency* (PDE):

$$PDE = QE(\lambda) \cdot FF \cdot P_{tr} \tag{3.14}$$

being $QE(\lambda)$ the *quantum efficiency*, which can be determined by TCAD simulations and can be seen in FIGURE 3.7, FF the fill factor and P_{tr} the avalanche triggering probability. P_{tr} can be then approximated experimentally with the PDE data provided in [10] using equation (3.14), as it proved a better approach than a theoretical approximation used by [8]. The triggering probability is shown in FIGURE 3.8.

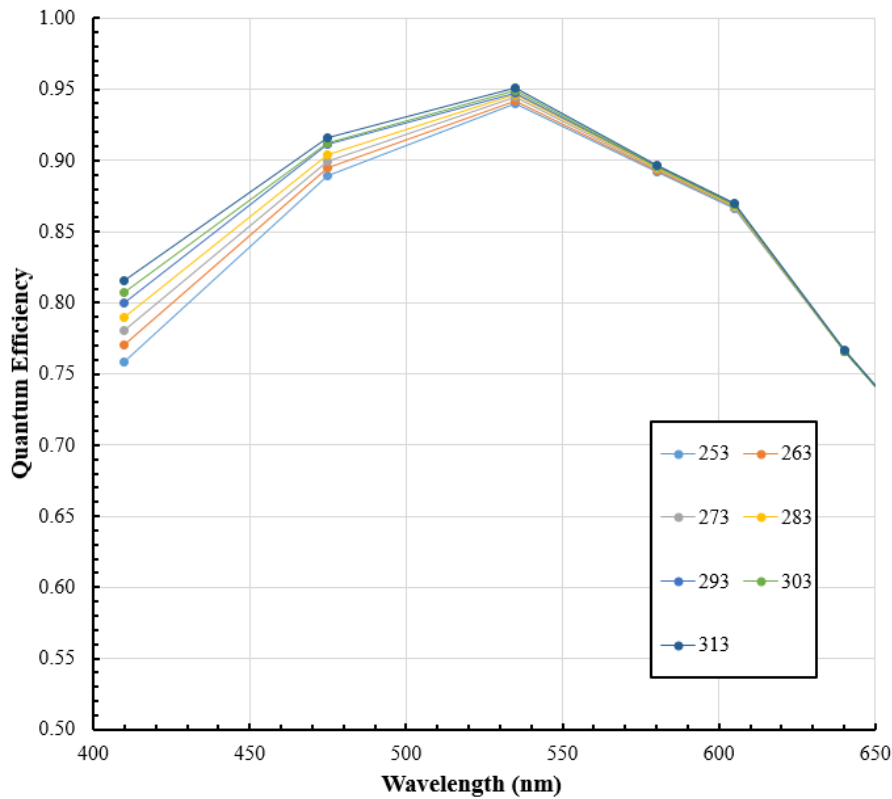


FIGURE 3.7 G12 Model: Quantum Efficiency against wavelength for several temperatures (in *Kelvin*).

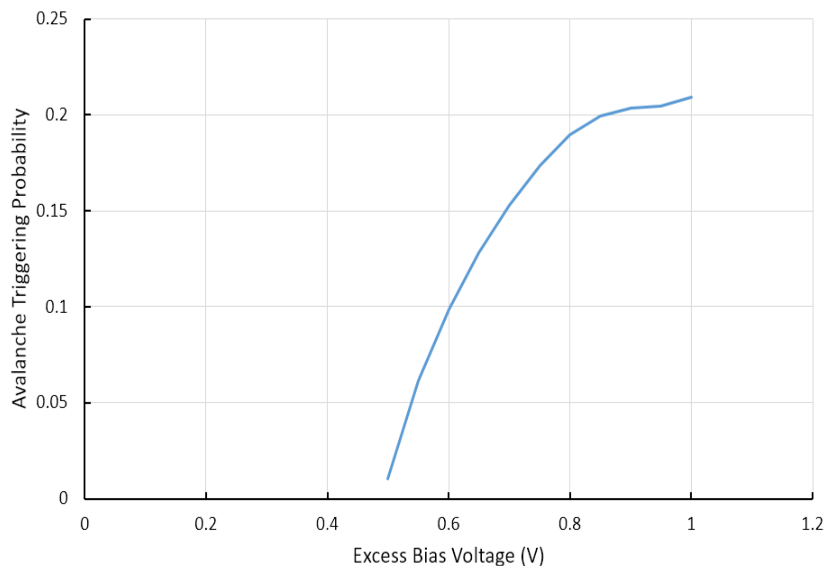


FIGURE 3.8 Avalanche triggering probability against excess bias voltage. Extracted from [10].

The basic G12 VERILOG-A model cannot consider the SPAD response time and its statistical deviation, the photon-timing jitter. However, an expansion of the basic model which can simulate these parameters will be treated separately in Chapter 6.

3.4.2 PRIMARY DARK COUNTS: TRAP-ASSISTED THERMAL GENERATION

It corresponds to carriers generated within the depletion region due to the presence of defects in the silicon crystal lattice; a process that shows a large temperature dependence and that can

be formulated by the *S̲hockley-R̲ead-H̲all* (SRH) theory of recombination through defects. This formulation makes the following assumptions [11]:

1. the semiconductor is *non-degenerate*;
2. the defect levels involved all have the same energy E_t (single-level trap);
3. the energy position of the defect level is stable, i.e., especially independent of the charge state of the impurity;
4. the impurity concentration is negligibly small compared to the background doping;
5. the *relaxation times* of the captured carriers are much shorter than their *reemission times*.

The assumption of non-degeneracy is valid in the wafer bulk in almost all cases, provided a non-diffused bulk wafer. The recombination dominance of a stable single-level trap is also assumed to be valid based on the work on *dark count spectroscopy* [12], suggesting that the major DCR source in modern SPADs is the *phosphorus-vacancy defect* or *E-center*. This defect has an activation energy of $E_t \approx 0.44$ eV below the conduction band and a cross-section of $\sigma_T = 1.1 \times 10^{-13}\text{cm}^2$ [13]. The common problem of the coexistence of a deep and a shallow level, dominating carrier lifetime under different conditions, is compatible with the standard SRH model as both types of defects are treated independently in this work.

The last two points in the assumption list above ensure that the defect center acts mainly as a recombination center and that eventual trapping effects due to the defect center are negligible for the validity of the SRH model. For this to happen, the recombination center density, N_T, must be at least one order of magnitude lower than a critical recombination center density, N_{CRIT}, to ensure the accuracy of the standard SRH model below 10 % under low-injection conditions [14]. To be able to use this criterion in this Thesis, first we have to calibrate N_T around our experimental results (see Chapter 5) and then compare its results to N_{CRIT}.

The critical recombination center density depends not only on the doping concentration N_D and the excess carrier concentration or injection density Δn, but also on the energy level E_t and the capture cross-sections σ_n and σ_p of the given recombination center, thus representing a unification of the prerequisites 4 and 5. That is, it represents a limit beyond which those prerequisites are not met. In parallel with the increasing dominance of trapping effects, N_{CRIT} decreases with decreasing N_D and Δn, respectively. Thus, for a given defect center and sample, the low-injection approximation N_{CRIT}^{LLI} represents a lower bound for N_{CRIT}. For *p*-type doping, it takes the form:

$$N_{CRIT}^{LLI} = \frac{(p_0 + p_1)\left[n_1 + \left(\frac{\tau_{n_0}}{\tau_{p_0}}\right)(p_0 + p_1)\right]}{\left|p_0 - \left(\frac{\tau_{n_0}}{\tau_{p_0}}\right)p_1\right|} \tag{3.15}$$

being p_0 the hole concentration at thermal equilibrium, τ_{n_0} and τ_{p_0} the recombination lifetime for electrons and holes, which will be mentioned later, and n_1 and p_1 the SRH densities:

$$n_1 = N_C e^{\frac{-\Delta E_n}{k_B T}} \tag{3.16}$$

$$p_1 = N_V e^{\frac{-\Delta E_p}{k_B T}} \tag{3.17}$$

$$\Delta E_n = E_C - E_t \tag{3.18}$$

$$\Delta E_p = E_t - E_V \tag{3.19}$$

where N_C and N_V are the effective density of states in the conduction and valence band, E_t the defect energy level, and E_C and E_V the lowest energy of the conduction band and the highest energy of the valence band, respectively.

For the G-models (despite the active area radius) $N_{CRIT}^{LLI} \approx 2 \times 10^{12}\,\mathrm{traps/cm^3}$ which is almost two orders of magnitude higher than the final recombination center density, $N_t = 6.37 \times 10^{10}\,\mathrm{traps/cm^3}$, which ensures the validity of the SRH model. The N_{CRIT} criterion thus constitutes an upper limit of tolerable contamination. Also, since the N_{CRIT} criterion only accounts for trapping effects arising from the dominating recombination center itself (in this case, the phosphorus vacancy), and this work only considers a unique recombination center, this assures the validity of the SRH model. As we will see later, this work considers those trapping effects.

Once the validity of the SRH model has been established, the carrier generation rate due to defects in the crystal lattice or thermal generation can be defined by [15]:

$$CGR_{TH} = \frac{n_i \cdot A \cdot W_e}{\tau_{eff}} \left(1 - e^{\frac{-qV}{2k_bT}}\right) \tag{3.20}$$

being A the active area of the primary junction, W_e the effective depletion layer width, n_i, the silicon intrinsic concentration, V the voltage across the junction, and τ_{eff} the effective carrier lifetime. The exponential term is only important when the voltage across the junction is very small, well below the breakdown voltage. So for the sake of simplicity, we can approximate equation (3.20) to:

$$CGR_{TH} \approx \frac{n_i \cdot A \cdot W_e}{\tau_{eff}} \tag{3.21}$$

where n_i, the silicon intrinsic concentration, is given by:

$$n_i = (N_C N_V)^{\frac{1}{2}}\, e^{\frac{-E_g}{2k_BT}} \tag{3.22}$$

being E_g the silicon forbidden bang gap energy. Bear in mind that the forbidden bandgap is affected by narrowing due to the device's high doping levels ($\approx 10^{17}\,\mathrm{traps/cm^3}$). This is highly significant as it affects energy gaps between the conduction and valence bands and the phosphorus-vacancy energy activation, thus affecting all the model calculations. Besides, the temperature also affects the forbidden bandgap, although the temperature dependence of the model will be treated in a later section. Therefore, the silicon forbidden bang gap is determined in this work as [4]:

$$E_g = E_{g0} - \frac{\alpha_g\, T^2}{T + \beta_g} - E_{gn} \tag{3.23}$$

being $E_{g0} = 1.166\ \mathrm{eV}$, $\alpha_g = 4.73 \times 10^4\ \mathrm{eV/K}$, $\beta_g = 636\ \mathrm{K}$, and E_{gn} the silicon bandgap narrowing, which is [16]:

$$E_{gn} = 9 \cdot 10^{-3} \left\{ \ln\left(\frac{N_D}{10^{17}}\right) + \left[\ln\left(\frac{N_D}{10^{17}}\right)^2 + 0.5\right]^{\frac{1}{2}} \right\} \tag{3.24}$$

with N_D being the net doping in atoms/cm^3.

The main difficulty in determining the carrier generation rate (equation (3.20)) is calculating the effective carrier lifetime, τ_{eff}. Provided that the assumptions of the SRH model meet, and using the injection density, Δn, in an avalanche event provided by the TCAD model, the effective carrier lifetime can be calculated by [11]:

$$\tau_{eff} \approx \frac{\tau_{n_o}(p_0 + p_1 + \Delta n) + \tau_{p_o}(n_0 + n_1 + \Delta n)}{n_0 + p_0 + \Delta n} \tag{3.25}$$

where n_0 is the electron concentration at thermal equilibrium and τ_{n_o} and τ_{p_0} the recombination lifetimes for electrons and holes, as mentioned above. These recombination lifetimes are given by [17]:

$$\tau_{n_0} = \frac{1}{v_{de} N_T \sigma_T} \tag{3.26}$$

$$\tau_{p_0} = \frac{1}{v_{dh} N_T \sigma_T} \tag{3.27}$$

being v_{de} and v_{dh} the drift velocity for electrons and holes, respectively, σ_T the defect cross-section and N_T the recombination center density. The latter is a technology parameter that we cannot possibly know, so all the models will be calibrated around this parameter to match the experimental results. See Chapter 5 for more information.

3.4.3 PRIMARY DARK COUNTS: TRAP-ASSISTED TUNNELING

When the electric field is large enough, trap-to-band *phonon-assisted tunneling* can enhance the emission of electrons and holes. In this process, carriers are trapped first in the defect energy level and then tunnel the remaining potential barrier due to a high electric field. According to [17], equations (3.26) and (3.27) must be modified to include this enhancement effect as follows:

$$\tau_{n_0} = \frac{1}{v_{de} N_T \sigma_T (1 + \Gamma_n)} \tag{3.28}$$

$$\tau_{p_0} = \frac{1}{v_{dh} N_T \sigma_T (1 + \Gamma_p)} \tag{3.29}$$

where the added parameters Γ_n and Γ_p are the *field-enhancement factors* for electrons and holes, respectively, which expressions are [17]:

$$\Gamma_{n,p} = \frac{\Delta E_{n,p}}{k_B T} \int_0^1 exp\left[\frac{\Delta E_{n,p}}{k_B T}\, u - K_{n,p}\, u^{\frac{3}{2}}\right] du \tag{3.30}$$

with:

$$K_{n,p} = \frac{4}{3} \frac{\sqrt{2m^* \Delta E_{n,p}}}{q\hbar |E_m|} \tag{3.31}$$

being E_m the electric field in the primary junction, and m^* the effective mass for conductivity calculations.

Numerical methods are mandatory to solve equation (3.30). However, based on the numerical outcomes, the following exponential fitting functions can be used for approximate solutions:

$$\Gamma_n \approx G_1(T)\, e^{G_2(T)|E_m|} \tag{3.32}$$

$$\Gamma_p \approx G_3(T)\, e^{G_4(T)|E_m|} \tag{3.33}$$

where the G variables show a quadratic dependence with the temperature that fits the numerical solution of equation (3.30), with a regression coefficient of $R^2 \approx 1$:

$$G_1(T) = -1.33 \times 10^{-5}\, T^2 + 9.985 \times 10^{-3}\, T - 1.255 \tag{3.34}$$

$$G_2(T) = 3.83 \times 10^{-10}\, T^2 - 2.863 \times 10^{-7}\, T + 6.048 \times 10^{-5} \tag{3.35}$$

$$G_3(T) = -2.3 \times 10^{-5}\, T^2 + 1.702 \times 10^{-2}\, T - 2.556 \tag{3.36}$$

$$G_4(T) = 5.486 \times 10^{-10}\, T^2 - 3.962 \times 10^{-7}\, T + 7.899 \times 10^{-5} \tag{3.37}$$

FIGURE 3.9 shows the variation of the field-enhancement factor with the electric field. When the SPAD is reverse-biased beyond the breakdown voltage, this factor can be as high as 300 at ambient temperature, effectively enhancing the carrier generation by this factor. As the recombination center density, N_T, remains unknown at this point, not including this effect can be the cause of overestimating it by at least two orders of magnitude in devices with high impurity concentration.

It is important to mention that, under a low enough electric field, equations (3.28) and (3.29) are reduced to equations (3.26) and (3.27), respectively. Note also that trap-assisted tunneling is noticeably greater at lower temperatures. However, the next chapter shows that the carrier generation rate drops greatly with temperature, so TAT has almost no impact at low temperatures.

3.4.4 PRIMARY DARK COUNTS: BAND-TO-BAND TUNNELING

As CMOS technologies scale down, the depletion layer becomes increasingly thinner, and the electric field increases (equation (3.3)). When the electric field approaches $7 \times 10^5\ \mathrm{V/cm}$, the probability of an avalanche being triggered by a carrier generated by a BTBT process becomes significant.

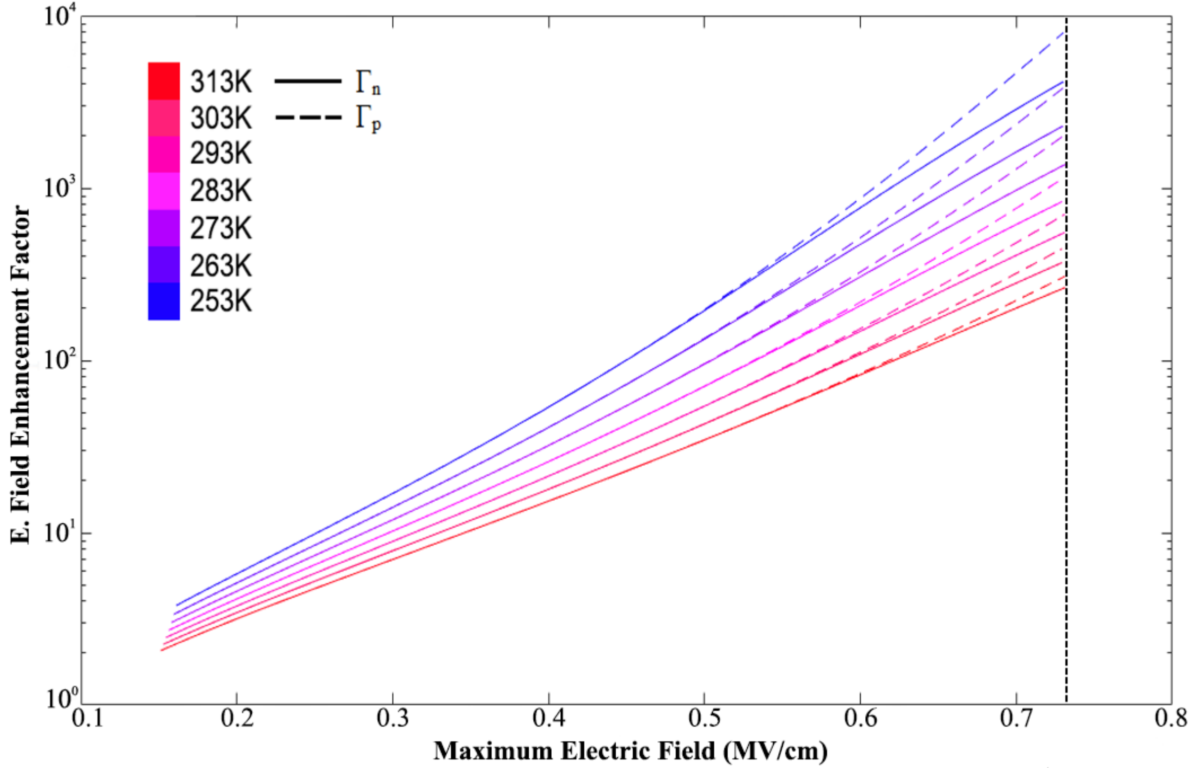


FIGURE 3.9 Field enhancement factors for electrons and holes for several temperatures against the primary junction electric field for the phosphorus vacancy defect. The dashed vertical line represent the primary junction's electric field when the SPAD is reverse biased beyond the breakdown voltage, and thus, Γ_n and Γ_p remain practically constant from there on.

This Thesis relies on Kao's model [18], which validity has been checked among others in [12][19], to determine the carrier generation rate due to BTBT. The carrier generation rate due to BTBT is given by:

$$CGR_{BTBT} = \left[A_K \left(\frac{E_m}{E_0} \right)^P e^{\left(\frac{-B_K}{E_m} \right)} \right] \cdot A \cdot W_e \tag{3.38}$$

where $A_K = 3.29 \times 10^{15} \text{cm}^{-3}\text{s}^{-1}$ and $B_K = 2.38 \times 10^{6} \text{Vcm}^{-1}$ are *Kane's parameters*, $E_0 = 1\text{Vcm}^{-1}$, and $P = 2.5$ as transitions in silicon are indirect [18].

There is no consensus on Kane's parameters values. The issue is of paramount importance since a small change in their values significantly impacts the final dark count rate. TABLE 3.2 shows values used by different researchers. This Thesis founds that Kao's parameters solution fits its results well and proved successfully by [19].

Once the primary sources to dark counts have been defined, the dark count rate due to primary sources can be expressed as [9]:

$$DCR_{PRI} = (CGR_{TH} + CGR_{BTBT})\, P_{tr} \tag{3.39}$$

Each carrier generation process follows a *Poissonian distribution* whose average value is its own *CGR*. Then, the time interval between two subsequent carrier generation events of each carrier generation process has an exponential distribution; whose *expected mean value* is:

$$\tau_{TH,BTBT} = \frac{1}{CGR_{TH,BTBT}} \tag{3.40}$$

TABLE 3.2 Kane parameters values by several authors for silicon indirect transitions.

References	$A_K(cm^{-3}s^{-1})$	$B_K(Vcm^{-1})$
Hurxk 1992 [17]	4.00×10^{14}	19.0×10^{6}
Kao 2012 [18]	3.29×10^{15}	23.8×10^{6}
Webster 2013 [12]	2.00×10^{15}	23.9×10^{6}

3.4.5 SECONDARY DARK COUNTS: AFTER-PULSING

Deep-level traps are defects in the semiconductor lattice that cause valid energy levels in the forbidden bandgap. When holes or electrons are trapped in them, they require a larger energy than the characteristic thermal energy $k_B T$ to be released. During the triggering of an avalanche, some carriers may be captured by these deep-level traps. After a statistical time called the *trap lifetime*, the carrier is released and may trigger an additional avalanche, which eventually may provoke the capture of additional carriers. This phenomenon is called *after-pulsing*.

There are typically several deep level traps that can be modeled by different *trap lifetimes* given by [9]:

$$\tau_i = \tau_{0i} e^{\frac{E_{A_i}}{k_B T}} \tag{3.41}$$

being E_A the activation energy and τ_0 a pre-exponential factor for the *ith* trap.

The probability of after-pulsing is time-dependent as stated by [9] and is given by [20][21]:

$$P_a(t) = \sum_{i=1}^{N} \frac{A_i}{\tau_i} e^{\frac{-t}{\tau_i}} \tag{3.42}$$

where A_i is a pre-exponential factor, N the number or deep-level traps, i the *ith* deep-level trap and, τ_i the trap-lifetime of the *ith* trap.

The pre-exponential factor represents the probability for a carrier to be captured by a unique deep-level *ith* trap during an avalanche and, if released, the probability for that carrier to trigger an avalanche. This pre-exponential factor can be approximated by [4]:

$$A_i = \frac{H_i t_a}{V} P_{tr} \tag{3.43}$$

$$H_i = \sigma_{Ti}\, v_{de}\, n_e \tag{3.44}$$

$$n_e = \int_{t_0}^{t_f} \frac{I_{spad}}{q} dt \tag{3.45}$$

where H_i represents the electron capture rate per unit volume for the ith deep-level trap, σ_{Ti} the cross-section of the ith deep-level trap, v_{de} the electron thermal velocity, n_e the electrons generated during the last avalanche, V the depletion region volume, t_a the duration time of the last avalanche, and P_{tr} the avalanche triggering probability.

The values of the pre-exponential factors τ_0 and the activation energies E_A are determined empirically. This Thesis uses the data provided by [9], which models after-pulsing with 3 deep-level traps (1 slow trap of 156 ns of trap-lifetime and 2 fast traps of 23 and 28 ns trap-lifetime at a temperature of 300K).

The traps cross-sections are approximated from the data provided by the work in [8]. TABLE 3.3 displays trap values chosen to be physically reasonable from the data present in the works [8][9].

TABLE 3.3 Deep level traps lifetimes, activation energies and cross-sections.

ith trap	$\tau_0(s)$	$E_A(eV)$	$\sigma_T(cm^2)$
1	8.742×10^{-12}	0.253	5×10^{-16}
2	9.947×10^{-13}	0.265	5×10^{-15}
3	5.296×10^{-10}	0.097	5×10^{-15}

3.5 TEMPERATURE EFFECTS MODELING

All previously reported models assume a constant temperature. This Thesis considers temperature changes happening during operation due to *self-heating* effects. Temperature dependences of some relevant parameters have already been presented in previous sections, namely the intrinsic concentration (equation (3.22)) and the SRH densities (Equations (3.16)(3.17)). These variables depend on the **density of states in the conduction and the valence bands**, which in their turn depend on temperature [4]:

$$N_c = 2 \times 10^{-6} M_c \left(\frac{2\pi m_{de}^* k_B T}{h^2}\right)^{\frac{3}{2}} \quad [\text{cm}^{-3}] \tag{3.46}$$

$$N_V = 2 \times 10^{-6} \left(\frac{2\pi m_{dh}^* k_B T}{h^2}\right)^{\frac{3}{2}} \quad [\text{cm}^{-3}] \tag{3.47}$$

where M_C = 6 is the number of equivalent minima for Si, and m_{de}^* and m_{dh}^* are the hole and electron effective mass for density of states purposes. According to [4], these masses are given by

$$m_{de}^* = (m_l m_t m_t)^{\frac{1}{3}} \tag{3.48}$$

$$m_{dh}^* = \left(m_{lh}^{\frac{3}{2}} + m_{hh}^{\frac{3}{2}}\right)^{\frac{2}{3}} \tag{3.49}$$

where m_l and m_t are the silicon longitudinal and the transverse electron effective masses, and $m_{lh}$and m_{hh} accounts for the silicon light and heavy hole effective mass.

The **effective masses** of electrons and holes for silicon are also relevant regarding temperature dependence. According to [22]:

$$m_e^* = (1.045 + 4.5 \times 10^{-4}T)m_0 \tag{3.50}$$

$$m_h^* = (0.523 + 0.0014T - 1.48 \times 10^{-6}T^2)m_0 \tag{3.51}$$

being m_0 the electron mass. These masses impact the electron and hole drift velocities [4]:

$$v_{de} = \sqrt{\frac{3k_bT}{m_e^*}} \tag{3.52}$$

$$v_{dh} = \sqrt{\frac{3k_bT}{m_h^*}} \tag{3.53}$$

which are used to determine the recombination lifetimes (equations (3.26)-(3.29)).

The **breakdown voltage** also depends on temperature. According to [8], such dependence can be captured into:

$$V_B = V_{B0}(1 + \beta \cdot \Delta T) \tag{3.54}$$

where β [K^{-1}] is the temperature constant for the breakdown voltage, V_{B0} is the breakdown voltage at a reference temperature, $T_0 = 253K$, and ΔT is the temperature increment with respect to T_0. TCAD simulations results in FIGURE 3.10 show that the breakdown voltage for the different models does not seem to have a linear dependence. In fact, data shows that for three different structures of the G-models, the temperature constant has a linear dependence on temperature:

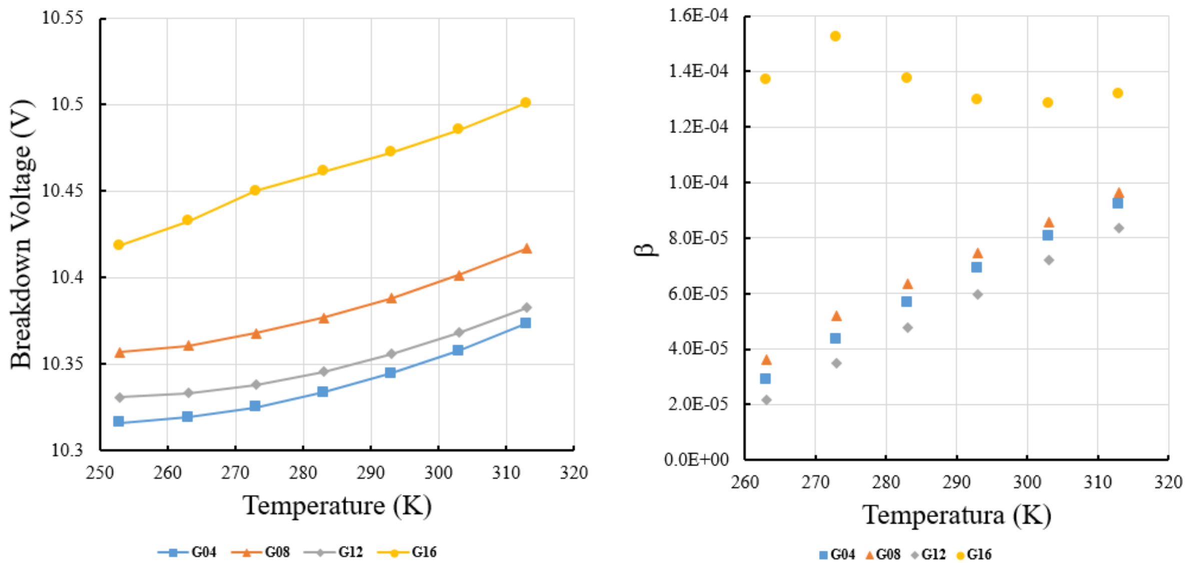


FIGURE 3.10 Breakdown voltage (left side) and β (right side) variation with temperature.

$$\beta = B_1 T + B_0 \tag{3.55}$$

being B_0 [K^{-1}] and B_1[K^{-2}] the constants that fit this relation. TABLE 3.4 shows extracted values of these parameters for different G-models.

Another major player, the **electric field**, also depends on temperature:
being V_{ka} the SPAD bias voltage, E_0 the electric field dependent on the excess bias voltage at a temperature of reference $T_0 = 253$K, and ΔT the temperature increment with respect to T_0. The electric field is measured in Vcm^{-1}. More information about $E_0(V_E)$ will be given in Chapter 4 (equation (4.1)). Equation (3.56) has been calculated from numerical approximation with a coefficient regression of $R^2 = 0.9907$. As we only need to know the electric field when the SPAD is biased over the breakdown voltage, a constant value greater than 0 is assigned to the electric field when the SPAD is biased below the breakdown voltage for model stability purposes.

$$E = \begin{cases} 10^5 & V_{ka} < V_B \\ E_0(V_E)[1 + (6.748 \times 10^{-7}T - 2.247 \times 10^{-4})\Delta T] & V_{ka} \geq V_B \end{cases} \tag{3.56}$$

Temperature also affects the **traps cross-section**, and it is related to Equation (3.41) of the trap's lifetime since equation (3.26) relates them to each other. Interestingly, previous works only apply this temperature dependence to the trap lifetime of deep-level traps. This Thesis modifies the mid-gap level traps cross-section to reflect this temperature dependence according to [23]:

$$\sigma_T = \sigma_0 e^{\frac{-E_t}{k_B T}} \tag{3.57}$$

being σ_0 a pre-exponential factor and E_t the activation energy of the phosphorus vacancy. Taking the trap cross-section at 300K as a reference:

$$\sigma_T = \sigma_{300} e^{\frac{E_t}{k_B}\left(\frac{1}{300} - \frac{1}{T}\right)} \tag{3.58}$$

Regarding **self-heating**, the source of this heat is not clear, as it could be generated by the current through the device or the peripheral circuitry. Considering the simple model of joule heating, self-heating can be modeled as:

$$\Delta T = P_w \cdot R_T \tag{3.59}$$

TABLE 3.4 *B* Parameters for Several SPAD Models

model	$\boldsymbol{B_1}$[K^{-2}]	$\boldsymbol{B_0}$[K^{-1}]
G04	1.252×10^{-6}	-2.987×10^{-4}
G08	1.186×10^{-6}	-2.734×10^{-4}
G12	1.234×10^{-6}	-3.021×10^{-4}
G16	-2.95×10^{-7}	2.211×10^{-4}

$$R_T = \frac{L}{k\,(T)} \tag{3.60}$$

where P_w is the dissipated power, R_T the thermal resistance of the device volume, L the device thickness, and $k(T)$, the silicon thermal conductivity.

3.6 REFERENCES

[1] S. Cova, A. Longoni, and A. Andreoni, "Towards Picosecond Resolution with Single-Photon Avalanche Diodes." *Review of Scientific Instruments*, vol. 52, no. 3, pp. 408-412, March 1981.

[2] 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 Transactions on Electron Devices*, vol. 58, no. 7, pp. 2028-2035, July 2011.

[3] E. A. G. Webster, L. A. Grant, and R. K. Henderson, "A High-Performance Single-Photon Avalanche Diode in 130-nm CMOS Imaging Technology." *IEEE Electron Device Letters*, vol. 33, no. 11, pp. 1589-1591, Nov. 2012.

[4] S. M. Sze and K. K. Ng, *Physics of Semiconductor Devices*, 3rd ed. New York, NY, USA: Wiley, 2007.

[5] N. Rinaldi, "Analysis of The Depletion Layer of Exponentially Graded P-N Junctions with Non-Uniformly Doped Substrates." *IEEE Transactions on Electron Devices*, vol. 47, no. 12, pp. 2340-2346, Dec. 2000.

[6] S. Khorasani, "Abrupt PN Junctions: Analytical Solutions Under Equilibrium and Non-Equilibrium." *Solid State Electronics*, vol. 122, pp. 37-44, Oct. 2016.

[7] B. El-Kareh and L.N. Hutter, *Silicon Analog Components*, 2nd ed. Germany: Springer, 2020.

[8] G. Giustolisi, R. Mita, and G. Pallumbo, "Behavioral modeling of statistical phenomena of single-photon avalanche diodes." *Int. J. Circuit Theory and Applications*, vol. 40, no. 7, pp. 661–679, 2012.

[9] Z. Cheng, X. Cheng, D. Palubiak, M. J. Deen, and H. Peng, "A Comprehensive and Accurate Analytical SPAD Model for Circuit Simulation." *IEEE Trans. Electron Devices*, vol. 63, no. 5, pp. 1940–1948, May 2016.

[10] I. Vornicu, R. Carmona-Galán, B. Pérez-Verdú, and Á. Rodríguez-Vázquez, "Compact CMOS Active Quenching/Recharge Circuits for SPAD Arrays." *Int. J. Circuit Theory and Applications*, vol. 44, No. 4, pp. 917-928, Apr. 2015.

[11] S. Rein, *Lifetime Spectroscopy. A Method of Defect Characterization in Silicon for Photo-voltaic Applications*. 1st Ed. Berlin Heidelberg, Germany, Springer-Verlag: 2005.

[12] E. A. G. Webster and R. K. Henderson, "A TCAD and Spectroscopy Study of Dark Count Mechanisms in Single-Photon Avalanche Diodes." *IEEE Trans. Electron Devices*, vol. 60, no. 12, pp. 4014–4019, Dec. 2013.

[13] M. M. Sokoloski, "Structure and Kinectics of Defects in Silicon." NASA, Washington D.C., Nov. 1967.

[14] D. Macdonald and A. Cuevas, "Validity of Simplified Shockley-Read-Hall Statistics for Modeling Carrier Lifetimes in Crystalline Silicon." *Phys. Rev. B (Condensed Matter and Materials Physics*), vol. 67 (7), pp. 75203–1–7, 2003.

[15] D. A. Ramirez, M. M. Hayat, and M. A. Itzler, "Dependence of the Performance of Single Photon Avalanche Diodes on the Multiplication Region Width." *IEEE Journal of Quantum Electronics*, vol. 44, no. 12, Dec. 2008.

[16] J. W. Slotboom and H. C. De Graaf, "Measurements of Bandgap Narrowing in Silicon Bipolar Transistors." *Solid State Electronics*, vol. 19, pp. 857-862, 1976.

[17] G. A. M. Hurkx, D. B. M. Klaasen, and M. P. G. Knuvers, “A New Recombination Model for Device Simulation Including Tunneling.” *IEEE Trans. Electron Devices*, vol. 39, no. 2, pp. 331–338, Feb. 1992.

[18] K. H. Kao, A. N. Verhulst, W. G. Vandenberghe, B. Sorée, G. Groeseneken, and K. De Meyer, “Direct and Indirect Band-to-Band Tunneling in Germanium-Based TFETs.” *IEEE Trans. Electron Devices*, vol. 59, no. 2, pp. 292–301, Feb. 2012.

[19] J. Yipeng, K. Wei, W. Tahiuan, and G. Du, “Comparison of Band-to-Band Tunneling Models in Si And Si–Ge Junctions.” *J. of Semiconductors*, vol. 34, no. 9, Sep. 2013.

[20] D. B. Horoshko, V. N. Chizhevsky, and S. Ya, Kilin, “Full Response Characterization of Afterpulsing in Single-Photon Detectors.” *J. of Modern Optics*, vol. 64, no. 2, pp. 191–195, Mar. 2017.

[21] B. Korzh, T. Lunghi, K. Kuzmenko, G. Boso, and H. Zbinden, “Afterpulsing Studies of Low-Noise Ingaas/Inp Single-Photon Negative-Feedback Avalanche Diodes.” *J. Mod. Opt.*, vol. 62, no. 14, pp. 1151–1157, Nov. 2015.

[22] *Atlas User’s Manual: Device Simulation Software*. Silvaco Inc., Santa Clara, CA, 2016.

[23] A. Mitonneau, A. Mircea, G. M. Martin, and D. Pons, “Electron and Hole Capture Cross-Sections at Deep Centers in Gallium Arsenide.” *Revue de Physique Appliquée*, Tome 14, pp. 853-861, Oct. 1979.

CHAPTER 4
BUILDING THE VERILOG-A MODEL

4.1 INTRODUCTION

As in previous chapters, the G12 SPAD will serve as a workbench to explain the VERILOG-A models devised in this Thesis. Anyhow, despite the specific SPAD structure, VERILOG-A model building involves the following steps:

- electrodes and branches declaration;
- constants declaration;
- variables declaration;
- device parameters setting;
- static and dynamic behaviour;
- carrier generation rates:
- initial conditions:
- macroscopic parameters definition;
- photon arrival;
- thermal dark counts generation;
- BTBT dark counts generation;
- after-pulsing generation;
- turn-off;
- setting of the current contribution;
- self-heating temperature contribution.

Appendix I contains the code of the G12 model. The text in this chapter includes references to specific code lines of this appendix. FIGURE 4.1, repeated here for easier reading, shows the circuital structure of the model.

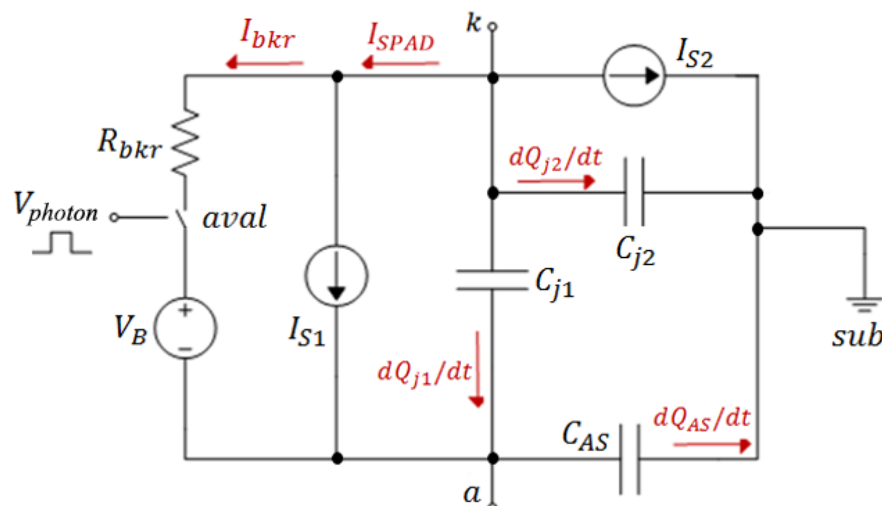


FIGURE 4.1 SPAD analytical model.

4.2 VERILOG-A BASIC MODEL DESCRIPTION

4.2.1 ELECTRODE AND BRANCH DECLARATION

The VERILOG-A model has the same terminal arrangement as Giustolici's model [1], with three terminals, namely, *a*, *k*, and *photon*, out of which the latter is for the light stimuli and the former, together with an additional ground terminal, *gnd* (*sub* in FIGURE 4.1), are for circuit connections. The declaration is as follows:

```
11. module spad_gr_12(a,k,photon);
12. inout a,k,photon;
13. electrical a, k, photon, gnd;
14. ground gnd;
15. branch (k,a) SPAD, JUNC_1;
16. branch (k,gnd) JUNC_2, C_SUB;
17. branch (a,gnd) A_SUB;
```

4.2.2 CONSTANTS AND VARIABLES

Declaration of suitable **Constant** values is of paramount relevance because small changes may significantly impact device running conditions and characteristics, according to the equations presented in Chapter 3. Constants are divided into three groups:

- *Device constants*: constants that are specific to the device (i.e., SPAD radius)
- *Physical constants*: fundamental physical constants (i.e., silicon bandgap voltage).
- *Trap-related constants*: constants that control the crystal defects type.

TABLE 4.1 details the values of the constants employed in this model. Some of these, like the light beam's wavelength that illuminates the SPAD, are often modified to change run conditions.

TABLE 4.1 Constants used in the G12 Verilog-A model.

Symbol	Quantity	Value
DEVICE CONSTANTS		
V_{b0}	Breakdown voltage at room temperature	10.331 V
T_0	Reference temperature	253 K
r	SPAD's radius	6×10^{-4} cm
η_{T1}	Upper triggering probability parameter	0.415
η_{T2}	Lower triggering probability parameter	4.36
m_{Ilat}	Avalanche turn-off mean current	10^{-4} A
σ_{Ilat}	Avalanche turn-off mean current standard deviation	5×10^{-6} A
V_n	Normalization voltage	5×10^{-2} V
R_{brk}	Breakdown resistance	72.555 Ω
N_1	DN-Well ion density (active area)	2.1×10^{17} cm^{-3}
N_2	DN-Well ion density (guard ring)	2.3×10^{17} cm^{-3}
N_3	DN-Well ion density (secondary junction)	6.0×10^{17} cm^{-3}
P_1	P+-Well ion density	4.8×10^{19} cm^{-3}

Symbol	Quantity	Value
P_2	T-Well ion density	1.1×10^{18} cm^{-3}
P_3	Epitaxial layer ion density	6.0×10^{14} cm^{-3}
C_{AS}	Anode-to-substrate stray capacitance	8.64×10^{-14} F
C_{CS}	Cathode-to-substrate stray capacitance	2.83×10^{-14} F
m_j	Junction grading coefficient	0.5
C_{01}	Primary junction zero-bias capacitance (active area)	7.98×10^{-8} Fcm^{-2}
C_{0g}	Primary junction zero-bias capacitance (guard ring)	3.68×10^{-8} Fcm^{-2}
C_{02}	Secondary junction zero-bias capacitance	6.05×10^{-9} Fcm^{-2}
A_{TOT}	Pixel area	1.024×10^{-5} cm^2
A_H	Kane A parameter	3.29×10^{15}cm^{-3}s^{-1}
B_H	Kane B parameter	2.38×10^{7} Vcm^{-1}
p	Indirect/Direct transition parameter	2.5
T_D	Total doping density at the primary junction	3.8×10^{17}cm^{-3}
R_E	Device electrical resistance	105.46 Ω
PHYSICAL CONSTANTS		
E_{g0}	Silicon bandgap	1.166 eV
m_0	Electron mass	9.11×10^{-31} Kg
m_{lh}	Silicon light hole mass	0.16 m_0
m_{hh}	Silicon heavy hole mass	0.49 m_0
m_l	Silicon longitudinal electron mass	0.98 m_0
m_t	Silicon transverse electron mass	0.19 m_0
M_C	Number of equivalent conduction band	6.0
λ	Wavelength	447 nm
TRAP-RELATED CONSTANTS		
E_t	Phosphorus vacancy-defect activation energy	0.45 eV
σ_T	Phosphorus vacancy-defect cross-section	1.1×10^{-13} cm^2
E_{A1}	1st deep-level trap activation energy	4.0535×10^{-20} J
E_{A2}	2nd deep-level trap activation energy	4.2382×10^{-20} J
E_{A3}	3rd deep-level trap activation energy	1.5585×10^{-20} J
τ_{01}	1st deep-level trap lifetime at 300K	8.7417×10^{-12} s
τ_{02}	2nd deep-level trap lifetime at 300K	9.9472×10^{-12} s
τ_{03}	3rd deep-level trap lifetime at 300K	5.296×10^{-12} s
σ_{T1}	1st deep-level trap cross-section at 300K	5×10^{-16} cm^2
σ_{T2}	2nd deep-level trap cross-section at 300K	5×10^{-15} cm^2
σ_{T3}	3rd deep-level trap cross-section at 300K	5×10^{-15} cm^2

4.2.3 DEVICE PARAMETERS SETTING

Parameters within this group are updated in every iteration of a simulation run, as temperature may cause them to change over time. The only exception is the **recombination center density** N_T, whose value changes with the type of technology or layer and which is needed to calculate recombination lifetimes (see equations (3.26)-(3.29)). The modeling methodology in this Thesis employs iterative procedures to select this parameter value such that simulated dark counts fit experimental measurements. Monitorization relies on three built-in model features:

- self-heating (SH);

- mid-gap trap cross-section dependence on temperature (TRAP.T); and
- trap assisted tunneling (TAT).

TABLE 4.2 contains calculated N_T for each feature combination.

Following N_T estimation, the **self-heating** feature updates the temperature by adding the heating (equation (3.59)) from the last iteration (line 267). This update will affect the entire model until the next iteration, and the following temperature-dependent parameters are affected:

- The **breakdown voltage** (V_B) and the temperature constant for the breakdown voltage (β), see equations (3.54) and (3.55), are modeled in line 272.
- The **bandgap narrowing** (E_{gn}) from equation (3.24) is also calculated in this section.
- The **energy activation levels** of the different trap families (E_{Ai}) are updated accordingly from line 288 onwards.
- The **conduction and valence bands density of states**, N_C (equation (3.46)) and N_V (equation (3.47)).
- The **effective masses**, m_e^* (equation (3.50)) and m_h^* (equation (3.51)).

These parameter values are then employed to:

- Calculate **built-in voltages** (ψ) for both junctions according to equation (3.4). This enables obtaining the voltage across both junctions and the guard ring providing the assumption that doping density is constant along the entire junction.

The next calculations refer to the **electric field**. As Chapter 3 explains (see section 3.2.2), the linear approximation to both primary and secondary junctions is not appropriate given the doping profile in the surroundings of the primary junction – see FIGURE 4.2. Instead, our method employs TCAD simulations (FIGURE 4.3) to express the primary junction electric field as a polynomial piecewise function of the excess bias voltage at a reference temperature of $T_0 = 253\text{K}$. This calculus is unnecessary for the second junction or the guard ring, as only the primary junction's electric field is used for dark count estimation. This electric field approximation is modeled in line 340 and is as follows:

$$E_0(V_E) = \begin{cases} 10^5 & V_E < 0 \\ -10241\,V_E^4 + 23868\,V_E^3 - 21056\,V_e^2 + 10844\,V_e + 728340 & 0 \leq V_E < V_1 \\ -1.33\,V_E^4 + 31.55\,V_E^3 - 250.07\,V_e^2 + 544.84\,V_e + 731358 & V_E \geq V_1 \end{cases} \quad (4.1)$$

TABLE 4.2 **Recombination Center Density according to the choice of models.**

~~SH~~	~~TRAP.T~~	~~TAT~~	$\mathbf{5.49 \times 10^{13}\ traps \cdot cm^{-3}}$
~~SH~~	~~TRAP.T~~	TAT	$1.17 \times 10^{11}\ \mathbf{traps \cdot cm^{-3}}$
~~SH~~	TRAP.T	~~TAT~~	$7.16 \times 10^{10}\ \mathbf{traps \cdot cm^{-3}}$
~~SH~~	TRAP.T	TAT	$1.17 \times 10^{11}\ \mathbf{traps \cdot cm^{-3}}$
SH	~~TRAP.T~~	~~TAT~~	$3.26 \times 10^{13}\ \mathbf{traps \cdot cm^{-3}}$
SH	~~TRAP.T~~	TAT	$7.35 \times 10^{10}\ \mathbf{traps \cdot cm^{-3}}$
SH	TRAP.T	~~TAT~~	$2.82 \times 10^{13}\ \mathbf{traps \cdot cm^{-3}}$
SH	TRAP.T	TAT	$6.37 \times 10^{10}\ \mathbf{traps \cdot cm^{-3}}$

being V_E the excess bias voltage and $V_1 = 0.856\text{V}$, and E_0 measured in Vcm^{-1}. FIGURE 4.4 also shows the variation of the electric field and the non-linearity breakdown voltage variation with temperature.

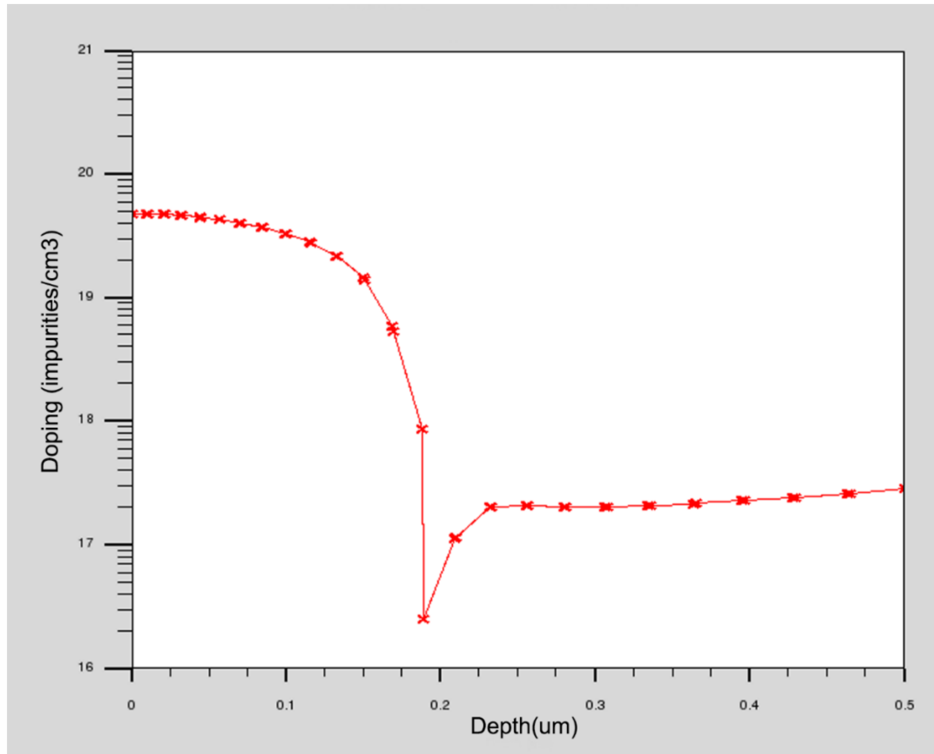


FIGURE 4.2 G12 SPAD doping profile of the primary junction at a radius of 3 µm.

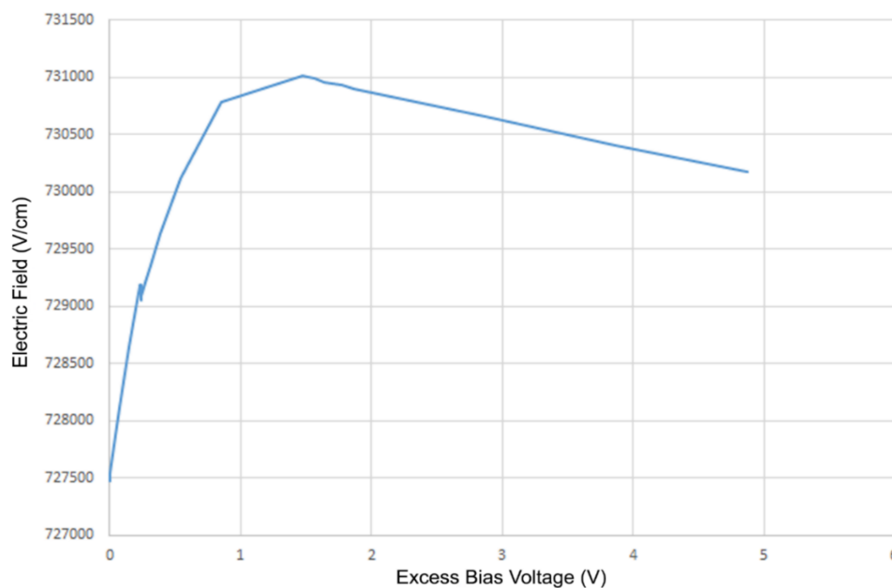


FIGURE 4.3 Primary Junction Electric Field Vs Excess Bias Voltage at 293K for the G12 SPAD model.

The list of bullet points below summarizes the next steps:

- The calculation of the maximum electric field, E_m, supports the determination of the **depletion region effective width** (W_e) in line 367 through equation (3.3).
- Next, the **electron and hole thermal velocities**, v_{de}, and v_{dh}, are determined. (equations (3.52) y (3.53)).

- Previous steps and the updating of the **deep-level traps cross-section**, σ_{Ti} (equation (3.58)) make it possible to update the **electron capture rate** of the deep-level trap

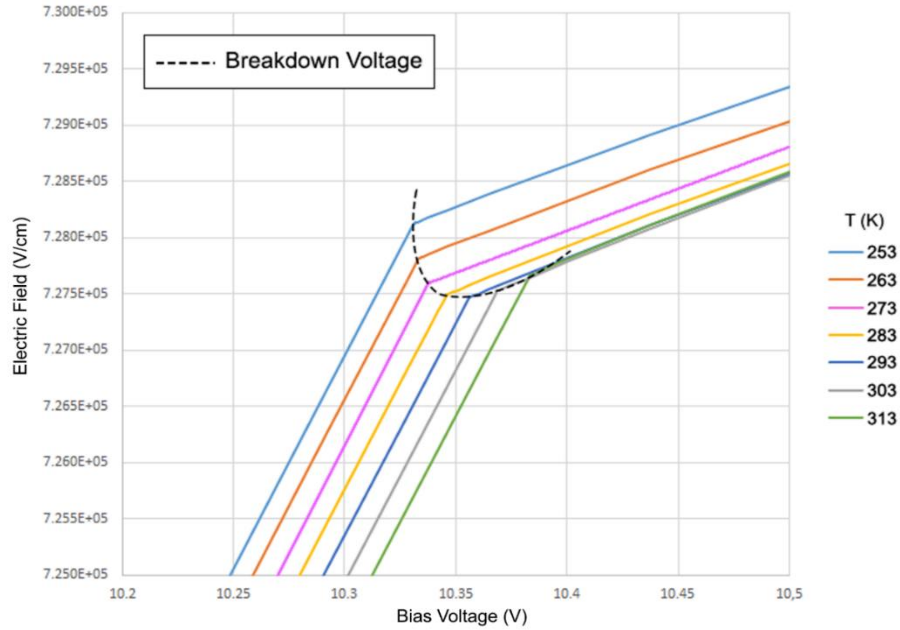


FIGURE 4.4 Electric Field against the bias voltage in the G12 SPAD model. The stripped line represents the breakdown voltage as a function of temperature.

(equation (3.44)) [1]:

$$\theta_{ei} = v_{de} \cdot \sigma_{Ti} \tag{4.2}$$

- The **saturation current** for both junctions is calculated as a linear function of the voltage across the junction, whose coefficients are exponential functions of temperature. Those coefficients are determined in line 371.
- Regarding the **photon-induced avalanche triggering probability**, it employs equation (3.14), where quantum efficiency is extracted from TCAD simulations, and the fill factor and the photon detection efficiency employ experimental results from [2]. These results are valid within the range of excess bias contemplated in reference [2], which is [0.5,1.0V]. Outside of that, the next model has to be applied [3]:

$$P_{tr} = 1 - \exp\left(\frac{-V_e}{\eta V_B}\right) \tag{4.3}$$

where η is a normalization constant that allows us to ensure P_{tr} to be a continuous function. The triggering probability has to be updated every iteration due to the changes to the breakdown voltage caused by the temperature.

- Finally, the **silicon thermal conductivity** is updated, k, since it also depends on the temperature. Its new value is used then to update the silicon thermal resistance, R_T – see equation (3.60).

4.2.4 STATIC AND DYNAMIC BEHAVIOR

Setting this behavior defines the next procedural step.

- First, the **saturation current**s for the primary and the secondary junction, respectively, are calculated using TCAD simulation outcomes:

$$I_{Si} = A_i(T) \cdot V_i + B_i(T) \tag{4.4}$$

where V_i is the voltage across the ith junction, and $A_i(T)$ and $B_i(T)$ parameters that fit I_{Si} a linear regression dependent on temperature:

$$A_1(T) = 2.174 \times 10^{-24} e^{(9.186 \cdot 10^{-2} T)} \tag{4.5}$$

$$B_1(T) = -2.233 \times 10^{-16}\, T^2 + 1.174 \times 10^{-13} T - 1.543 \times 10^{-11} \tag{4.6}$$

$$A_2(T) = 5.674 \times 10^{-26} e^{(1.04 \times 10^{-1}\, T)} \tag{4.7}$$

$$B_2(T) = 2.184 \times 10^{-24} e^{(9.187 \times 10^{-2}\, T)} \tag{4.8}$$

Then the avalanche part of the current that crosses the junction (I_{bkr}, the second term of equation 3.6) is modeled in line 408.

- Then, the **stored charge** in both junctions (equations (3.8) and (3.9)) are determined in line 411 as they require the built-in voltages.
- Finally, the **charge stored in the anode-to-substrate stray capacitor** (equation (3.10)) is estimated in line 415.

4.2.5 CARRIER GENERATION RATES

After setting the basic device's physical and electrical parameters, the turn is for the parameters that control the carrier generation rates. Expressions in Chapter 3 assume a unique type of recombination center, the *E-Centre*, whose activation energy is $E_A \approx 0.44eV$, below the conduction band energy level. However, as already stated, the activation energy of this defect is affected by the silicon bandgap narrowing due to high doping and temperature. Thus, the gap to the conduction band will be somewhat smaller. The issue of bandgap narrowing is far from trivial and has been addressed by many researchers (see FIGURE 4.5). Small variations in the activation energy of the recombination centers can lead to great changes in the carrier generation rate and must be treated with great precision. Therefore, the first task is calculating the energy gap of this defect to the conduction band (ΔE_n) and the valence band (ΔE_p). As E_A is defined below the conduction band, this is exactly the difference between the lowest energy level of the conduction band and that of the E-Centre:

$$\Delta E_n = E_A \tag{4.9}$$

and therefore, as $E_g = \Delta E_n + \Delta E_p$, being E_g the forbidden band gap energy with the narrowing included (equation (3.23)):

$$\Delta E_p = E_g - E_A \tag{4.10}$$

It is convenient expressing both energy gaps as a function of E_g, introducing the parameter α, whose expression is shown below beside the gap expressions using this parameter (line 418):

$$\alpha = \left(\frac{E_g - E_A}{E_g}\right) \tag{4.11}$$

$$\Delta E_p = \alpha E_g \tag{4.12}$$

$$\Delta E_n = (1-\alpha)E_g \tag{4.13}$$

- **Electron and hole concentrations** at thermal equilibrium in line 421 are extracted from TCAD simulations.
- Also, **SRH densities**, N_1 and P_1, from equations (3.16) and (3.17) can now be determined as we know the energy gaps from equations (3.18) and (3.19) as well as the **densities of states for de conduction and valence band**.
- Now we set the **injection density of charge carriers** (Δn) as a piecewise function of the current through the primary junction (I_{SPAD}) extracted from TCAD simulations, minus the electron concentration at thermal equilibrium in line 445:

$$\Delta n = n(I_{SPAD}) - n_0 \tag{4.14}$$

- We also include here the calculations of **gamma factors**, equations (3.32) and (3.33), that constitute the contribution of the trap-assisted tunneling to the thermal carrier generation rate (line 453).
- Now we can determine the **effective minority carrier lifetime** from SRH theory from equation (3.25), calculating first the electron and hole capture times from equations (3.28) and (3.29) in line 467, as other requirements are met.
- All previous steps enable the calculation of the **thermal carrier generation rate** in equation (3.21) (line 473), as well as the **carrier generation rate due to BTBT** from equation (3.38). The inverse of both values gives us the respective time between two carrier generation rates, allowing us to schedule the carrier generation events.
- The **mean time between two carrier release events from deep-level traps** is also updated due to temperature variations, and finally, we calculate the dark count rate in line 509 using equation (3.39).

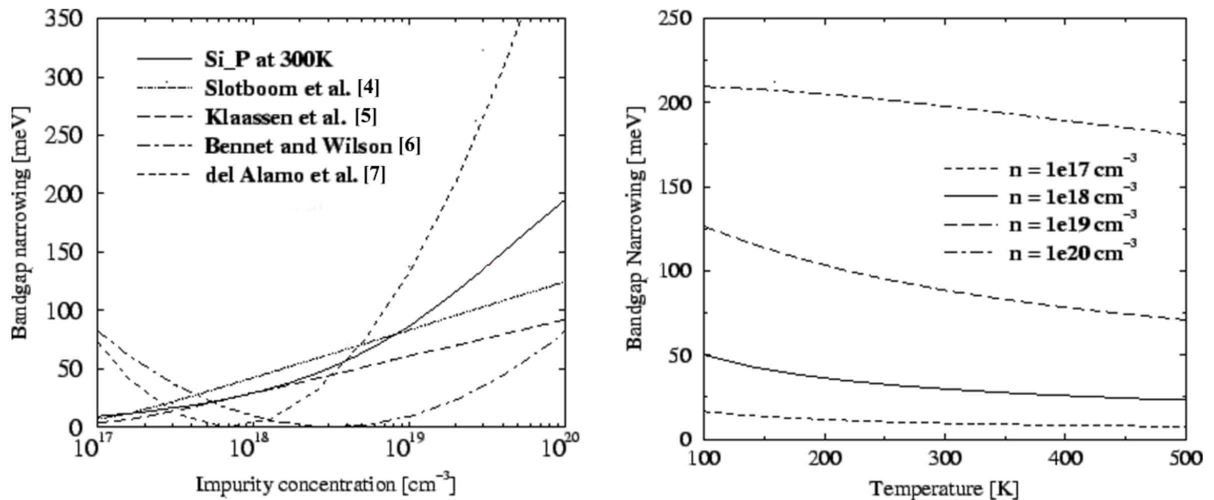


FIGURE 4.5 How impurity concentration and temperature affect band-gap narrowing.

4.2.6 INITIAL CONDITIONS AND MACROSCOPIC PARAMETER DEFINITION

This section is dedicated to initializing parameters and scheduling the first events for carrier generation or carrier release from deep-level traps. Additionally, it also sets the fill factor.

In line 593, the quantum efficiency is updated every iteration, as lower wavelengths significantly vary with temperature. Accordingly, the photon detection efficiency is updated too from equation (3.14).

4.2.7 PHOTON ARRIVAL

As in Giustolici's work, photon arrival is emulated by forcing the voltage of the photon terminal above a threshold value and allowing an interval of a few picoseconds, which is enough to start an avalanche so that the PDE requisite is met. By changing the value of the constant *aval* from 0 to 1, the avalanche is ignited:

$$I_{SPAD} = I_{s1} + aval \cdot I_{bkr} \tag{4.15}$$

During an avalanche, the model cannot ignite new avalanches. Also, the constant *reset* is pulled up to 1 from 0, making the model integrate the avalanche's current over time to measure the number of electrons. This permits, through equation (3.42), to determine the after-pulsing probability. In Giustolici's model [1], the probability of a carrier release was checked every time an avalanche was extinguished with the VERILOG-A function @(timer(start_time)), which made the simulation very slow. In fact, there was no probability check at all. Every interval of time that passed in which it was considered that there had to be at least one charge carrier freed from deep-level traps, one after-pulsing avalanche was triggered. In addition to not being random, this system had the problem of not considering the time that had passed since the triggering of the last avalanche. With the system developed in this work, every time an avalanche is turned off, the after-pulsing probability is checked without the need to call the *timer* function, improving speed simulation greatly. Also, it considers the time that has passed since the last avalanche and the current size to calculate the after-pulsing probability. To that end, in the case of an avalanche event, the model always records the initial and final absolute times of an avalanche, as well as the time between avalanches, so that the after-pulsing can be properly simulated.

4.2.8 DARK COUNT GENERATION, TURN-OFF, AND UPDATING CURRENT CONTRIBUTIONS

When scheduled, a generated carrier within the depletion region, either for thermal or tunneling causes or released carrier from deep-level traps, can generate an avalanche, provided the event passes the triggering probability check. Whether that be or not the case, the next event will be scheduled following an exponential distribution if the event is caused by generated charge carriers, as stated in Chapter 3.

While the avalanche is on, electrons are counted (line 735). While the avalanche current is higher than a certain limit value I_{lat}, the avalanche is self-sustained, and when it gets less than that limit is self-quenched (line 740). I_{lat} is the latching current, which exhibits a statistical fluctuation, where m_{Ilat} is the mean value and σ_{Ilat} the statistical fluctuation of a normal distribution.

After that, the probability of after-pulsing of each deep level is checked. In case of after-pulsing, an event is scheduled (line 746 onwards).

Finally, we establish the current across the SPAD in the current iteration in line 778. Then we calculate the current contributions from the stored charge in both junctions, as well as in the parasitic capacitances. To do that, the function *ddt* is used to integrate over the simulation time.

4.2.9 SELF-HEATING TEMPERATURE CONTRIBUTION

Depletion zones subjected to avalanches experience temperature increases owing to the circulating current and the corresponding power dissipation. Enabling the self-heating option allows the model to update the temperature. For that purpose, the dissipated power is firstly calculated in those SPAD regions where the avalanche current flows (line 800) using:

$$P_w = R_E \cdot I_{SPAD}^2 \tag{4.16}$$

where R_E stands for the SPAD electrical resistance.

To calculate the temperature increase due to self-heating, a comparison is made with the power dissipated in the last iteration. If the current is larger, self-heating is applied according to equation (3.59) to create a self-heating continuum. Indeed, it is assumed that SPAD cannot dissipate heat efficiently.

In the event of anomalously large avalanches, the avalanche is cut in half to maintain the heating continuum at a stable level. For stability, the temperature increase is also limited by self-heating to 100K since at higher temperatures and higher bias voltage, self-heating may escalate exponentially and cause the simulation run to halt.

4.3 MODEL'S STATE DIAGRAM

The flow diagram of FIGURE 4.6 provides a pictorial description of the model operation. First, the SPAD is biased beyond the breakdown voltage to enable valance behaviors. With this biasing, the SPAD enters a meta-state of equilibrium called **Geiger mode**. In this state, any carrier generated within the depletion region, or reaching this region by a diffusion process, may trigger an avalanche and get the SPAD out of that equilibrium. Until that happens, the device remains *idle*. However, since the beginning of the operation, the first band-to-band and thermal carriers are scheduled to be generated within the depletion region.

The device's illumination is performed by a periodic function where incident photons are simulated. When one of these photon strikes the SPAD, the PDE is checked (Equation (3.14)). If it happens to pass, an avalanche is triggered. In this state of avalanche, the SPAD cannot detect more incoming photons, and for that reason, this state is called **dead time**. When the quenching circuit extinguishes the avalanche, the device gets out of this state and is effectively reset into the Geiger mode state.

While the device is idle and no photon is absorbed, band-to-band and thermally generated carriers may be generated within the depletion region. If such a case, the photon triggering probability, P_{tr}, is checked, and if positive, a dark count avalanche is triggered. Either way, the next carrier due to band-to-band tunneling or thermal generation, respectively, is scheduled to be generated according to an exponential distribution.

Additionally, any avalanche can generate an **afterpulsing**, that is, a new avalanche after a random time after an avalanche due to the release of trapped carriers in deep-level traps during the previous avalanche. So every time an avalanche is extinguished, the probability of afterpulsing, $P_a(t)$ (equation (3.42)), for every one of the three types of deep-level trap is checked. The probability of afterpulsing depends on the photon triggering probability, but they can be separated so the model can be more efficient. From equations (3.42) and (3.43), the probability of a carrier from the *ith* deep-level trap to be released would be:

$$P_{ci} = \frac{H_i t_a}{V \tau_i} e^{\frac{-t}{\tau_i}} \tag{4.17}$$

Operating this way precludes checking the triggering probability if the carrier is not even released, which shortens the computer time required to run the model. So when the carrier release is positive, a charge carrier generation is scheduled to happen after a short random time after an avalanche. Notice that this avalanche can generate subsequent avalanches due to the release of charge carriers from deep-level traps.

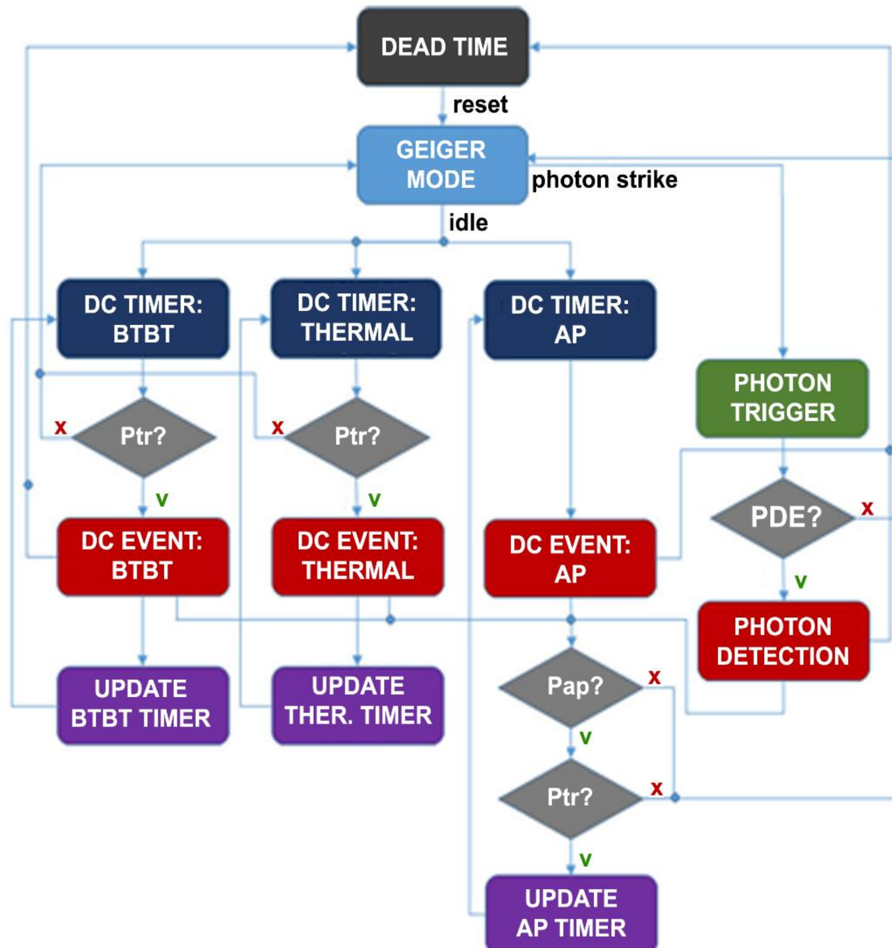


FIGURE 4.6 Generic state diagram of a SPAD model.

4.4 REFERENCES

[1] G. Giustolisi, R. Mita, and G. Pallumbo, "Behavioral Modeling of Statistical Phenomena of Single-Photon Avalanche Diodes." *Int. J. Circuit Theory and Applications*, vol. 40, no. 7, pp. 661–679, 2012.

[2] I. Vornicu, R. Carmona-Galán, B. Pérez-Verdú, and Á. Rodríguez-Vázquez, "Compact CMOS Active Quenching/Recharge Circuits for SPAD Arrays." *Int. J. Circuit Theory and Applications*, vol. 44, no. 4, pp. 917-928, Apr. 2015.

[3] S. M. Sze and K. K. Ng, *Physics of Semiconductor Devices*. 3rd ed. New York, NY, USA: Wiley, 2007.

[4] J. W. Slotboom and H. C. De Graaf, "Measurements of Bandgap Narrowing in Silicon Bipolar Transistors." *Solid State Electronics* Vol. 19 (1976): 857-862.

[5] D. B. M. Klaassen, J. W. Slotboom, and H.C. De Graaff, "Unified Apparent Bandgap Narrowing in n- and P-type Silicon." *Solid State Electronics,* vol. 35, no. 2, pp. 125-129, 1992.

[6] H. S. Bennett and C. L. Wilson, “Statistical Comparisons of Data on Band-Gap Narrowing in Heavily Doped Silicon: Electrical and Optical Measurements.” *J. Appl. Phys.* vol 55, no. 10, pp: 3582-3587, 1984.

[7] J. del Alamo, S. Swirhun, and R. M. Swanson, “Simultaneous Measuring of Hole Lifetime, Hole Mobility and Bandgap Narrowing in Heavily Doped N-Type Silicon.” *IEDM Technical Digest*, pp. 290-293, December 1985.

CHAPTER 5
EMPIRICAL VALIDATION OF THE MODEL WORKFLOW

5.1 INTRODUCTION

Previous chapters focused on defining and establishing the bases of a workflow to produce reliable SPAD VERILOG-A models. This process encompasses the following steps:

- Identify the device structure.
- Simulate the structure fabrication with ATHENA so that the resultant TCAD models are sufficiently for reliable simulation.
- Device simulation of the TCAD models to extract static parameters, which will help to develop the VERILOG-A model and gain a more profound knowledge of the SPAD operation.
- Establish the physical foundations of the single-photon avalanche diodes and their statistical phenomena.
- Make use of all resulting data to build the different SPAD VERILOG-A models.

This process has been applied to TCAD and VERILOG-A models for several SPAD physical structures. This chapter compiles electrical behavior results for G structures (Section 2.4.1), based on the P+/Deep-N-Well SPAD (from now on, P+ SPAD), and the V12 model (Section 2.4.2), based on the P-Well/Deep-N-Well SPAD (PW SPAD). These SPAD structures are relevant because of their maturity and physical plausibility. Also, the V12 represents an improvement that makes it worth considering. First, the results from the static parameters extracted from the TCAD models, the G models, and the V12 model will be discussed. Then, the statistical phenomena will be discussed before making a comparison with the literature.

5.2 TCAD EXTRACTED PARAMETERS

5.2.1 I-V CHARACTERISTICS

The first thing is to test the created TCAD models to check the I-V characteristics, so it resembles the behavior of an avalanche diode. When the SPADs are biased under the breakdown voltage, we can measure the current of the primary and secondary junctions due to the saturation current. This current is in the order of picoamperes, as shown in the logarithmic graph at the right in FIGURE 5.1. This figure also shows how the avalanche current increases linearly with the bias voltage when the devices are biased above the breakdown voltage. This is true as long as the guard ring breaking voltage is not reached. The linear graph on the left in FIGURE 5.1 shows that the G models follow these results, with the slope being inversely proportional to the device breakdown resistance (TABLE 5.1), as shown in equation (3.6). This resistance is higher the smaller the active area region of the SPAD (equation (3.7)). However, the V12 model exhibits an abrupt increase in the generated current for a bias voltage of around 14.5V due to premature edge avalanche triggering. That happens because the lateral breakdown voltage of the central P-Well of the V12 model is slowly being surpassed beyond 14.5V. These results show how the PW SPAD has a narrower range of operation than a P+ SPAD.

Regarding breakdown voltage values, the logarithmic scale in FIGURE 5.1 highlights the sudden current increase, from picoamperes to milliamperes, corresponding to the breakdown voltage limit. As expected, the breakdown voltage of the differently sized P+ SPADs is quite similar because this parameter does not depend on the device's size but on the layer's defect

TABLE 5.1 Breakdown resistance of the developed models.

Model	$R_{brk}[\Omega]$
G04	182.75
G08	103.1
G12	72.56
G16	54.94
V12	218.43

concentration that forms the junction. The breakdown voltage of a PW SPAD is higher because the primary junction's maximum electric field is lower, as explained in the next section

5.2.2 PRIMARY JUNCTION'S ELECTRIC FIELD

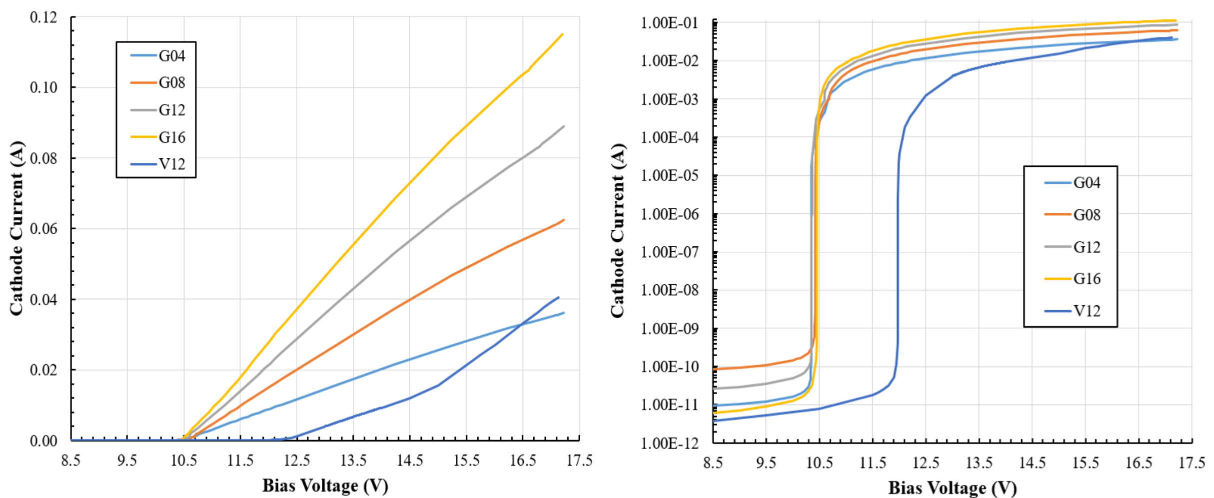


FIGURE 5.1 I-V characteristics from the G models and the V12 model in linear (left) and logarithmic (right) scale.

FIGURE 5.2 shows how the primary junction's electric field behaves in the active area region. Below the breakdown voltage, the electric field increases with the bias voltage. However, when the bias exceeds the breakdown voltage, the electric field plateaus, and the SPAD enters the Geiger operation region. As explained above, the V12 model experiences premature edge breakdown at 14.5V. When this happens, the current generated in the active area region of the SPAD drops significantly from that point, and the SPAD stops being sensitive to incoming photons. This range of operation can be improved, making, for example, the depletion region of the SPAD thicker. Chapters 6 and 7 further discuss the detection limits of the SPADs.

The PW SPAD, therefore, has a maximum electric field significantly lower than the P+ SPAD. This feature results in a decrease in noise, as it will be shown later.

When it comes to taking the electric field results to a VERILOG-A model, it is easily done through a piecewise function, which can be seen in Appendix I (Line 322 onwards). Notice that the G12 function for the electric field is the same for all the G models, as the layer configuration does not change.

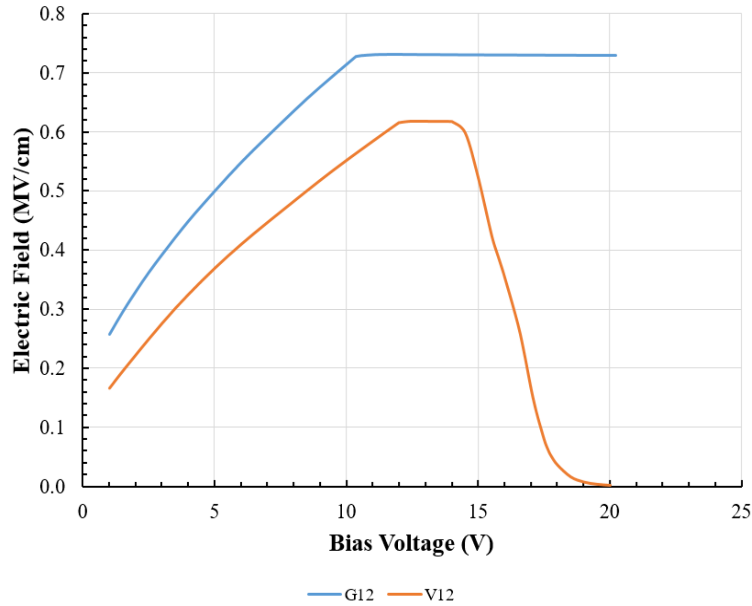


FIGURE 5.2 Primary junction's electric field vs bias voltage for the G12 and the V12 models.

5.2.3 BREAKDOWN VOLTAGE

As was shown in Chapter 3, the junction's electric field depends on the dopant density (equation (3.2)). A stronger electric field narrows the depletion region (equation (3.3)), thus decreasing the breakdown voltage. The G models have a breakdown voltage of approximately 10.3V at room temperature, similar to other works in the literature built under the same 180 nm standard technology as the layer configuration used for the models developed in this work [1]-[3].

The V12 model was also built following the layer configuration of a standard 180nm technology. However, the results in [4] correspond to a 130nm technology instead of a 180nm one, so empirical values are not comparable. Regardless, the breakdown voltage of the V12 models was found to be 11.92 V, with is a reasonable value given the lower electric field at the primary junction.

5.2.4 PHOTON DETECTION PROBABILITY

The photon detection efficiency, PDE, is a difficult parameter to compare with those of the literature. As shown in Chapter 3, PDE depends upon the fill factor, FF, and the avalanche triggering probability, P_{tr} (equation (3.14)). The fill factor varies according to the pixel's circuitry, and often its value is not disclosed in publications. On the other hand, the P_{tr} is a very complex parameter, as we will see, and depends on the excess bias voltage and the layer configuration of the foundry, so different devices will have different photon triggering probabilities. Nevertheless, a fair comparison can be made by taking only the **photon detection probability**, PDP, to examine the differences in responsiveness between the models and real-like SPADs:

$$PDP = QE(\lambda) \cdot P_{tr} \tag{5.1}$$

FIGURE 5.3 shows the photon detection probability versus the wavelength for the G12, the V12 model and several real SPADs from the literature. The first thing to highlight is that Niclass´s [5] and Richarson´s [4] SPADs (a P+ SPAD and a PW SPAD, respectively) use both a 130nm technology, whose layer configuration boosts quantum efficiency making both

SPADs' PDP higher. Secondly, the models and Niclass´s and Richardson´s SPADs all exhibit small valleys at 540nm and 570nm, which are probably related to the absorption characteristics of silicon. The fact that these valleys are present both in actual SPADs and in the models devised in the Thesis is one qualitative evidence of the model´s validity. Also, as with real SPADs, the V12 PDP is displaced towards the longer wavelengths with respect to the G12 model. It can also be seen that neither Faramazpour´s SPAD [2] nor Vornicu´s SPAD [1] present these valleys. Besides, their PDP are noticeably smaller than those of the models, and even their general shape is different. According to Faramazpour, the reason is the presence of a passivation layer over the SPADs, which may also be the case for Vornicu's SPAD.

5.2.5 AVALANCHE TRIGGERING PROBABILITY

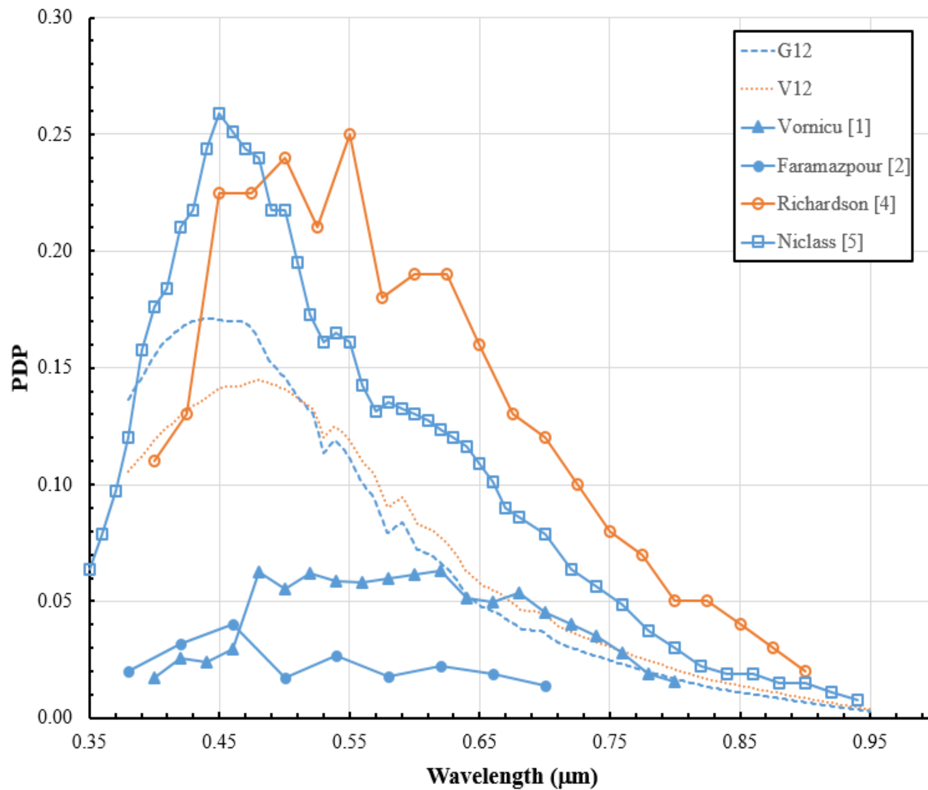


FIGURE 5.3 Photon detection probability vs wavelength at 1V of excess bias for the G12 and V12 models, and for several real SPADs from the literature. Those based in a P+ SPAD are depicted in blue, whereas those base in a PW SPAD are depicted in orange.

As explained in Chapter 3, the avalanche triggering probability used in this work follows the proposal of [1]. This reference provides a PDE for several excess biases at 447nm, thus allowing to extract the avalanche triggering probability, P_{tr}, for these biases using equation (3.14) and the TCAD-extracted quantum efficiency for a 447nm wavelength. However, problems arose when comparing those results to that of the existing theoretical model for the P_{tr} [6]:

$$P_{tr} = 1 - e^{\left(-\frac{V_E}{\eta V_B}\right)} \tag{5.2}$$

being V_E the excess bias over the breakdown voltage, V_B, and η a device-dependent fitting parameter.

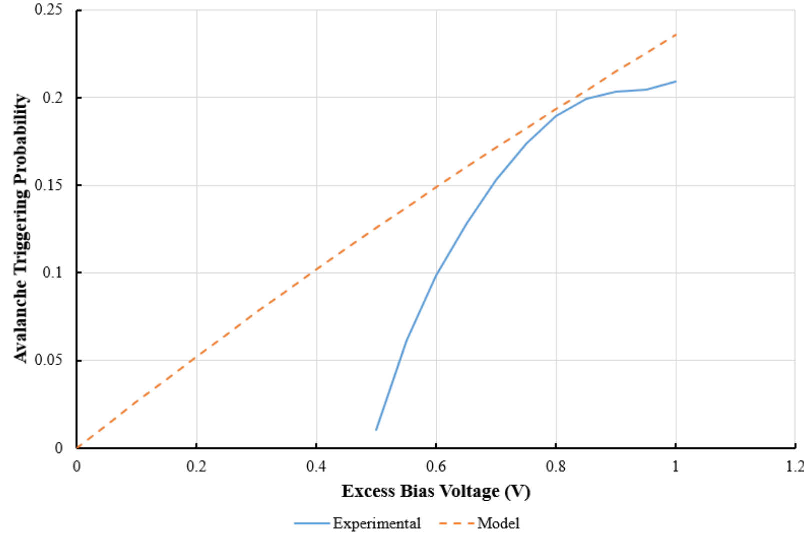


FIGURE 5.4 Avalanche Triggering Probability vs Excess Bias Voltage for experimental data extracted from Vornicu´s work [1] and from the model.

FIGURE 5.4 shows that the theoretical model hardly fits the experimental results extracted from [1]. It is interesting to notice that the experimental P_{tr} drops to zero at 0.5V excess bias voltage and, after, rises sharply to a moderate increase near 1V excess bias. As explained in Chapter 6, this dropping is likely to be associated with the pixel readout circuit, and we have not found any device-related feature producing it. The moderate increase near 1V excess bias observed in FIGURE 5.4 can be related to the breakdown voltage increase due to the self-heating effect (equation (3.54)) that makes the P_{tr} moderate its increase. Self-heating contribution is addressed in Section 5.3.3.

The avalanche triggering probability can be assessed using TCAD tools. When an electron-hole pair is created in silicon due to a photon strike, either the electron or the hole can trigger an avalanche and thus contribute to the avalanche triggering probability. For that reason, the P_{tr} has two components: the electron avalanche triggering probability, P_e, and the hole avalanche triggering probability, P_h:

$$P_{tr} = P_e + P_h \tag{5.3}$$

FIGURE 5.5 shows how these components, and the composed P_{tr}, change with the depth where the electron-hole pair is created within the G12 structure. As the SPAD is reverse biased, the cathode has a higher bias than the anode, so when an electron-hole pair is created near the surface within the P+ region of the device, the hole reaches the anode or recombines, while the electron drifts to the depletion region. For that reason, P_e dominates near the surface. Consequently, when the pair is created below the depletion region, the hole drifts towards the anode, and in its path, it has to cross the depletion region, making it the dominant component in that region. As the layer’s thickness and doping are different on the surface and below the depletion region, and the electron and holes have different drift velocities, their respective avalanche triggering probabilities also differ.

As can be deduced from FIGURE 5.5, the avalanche triggering probability for this G12-model is higher for photons absorbed near the surface. Most of these photons have a shorter

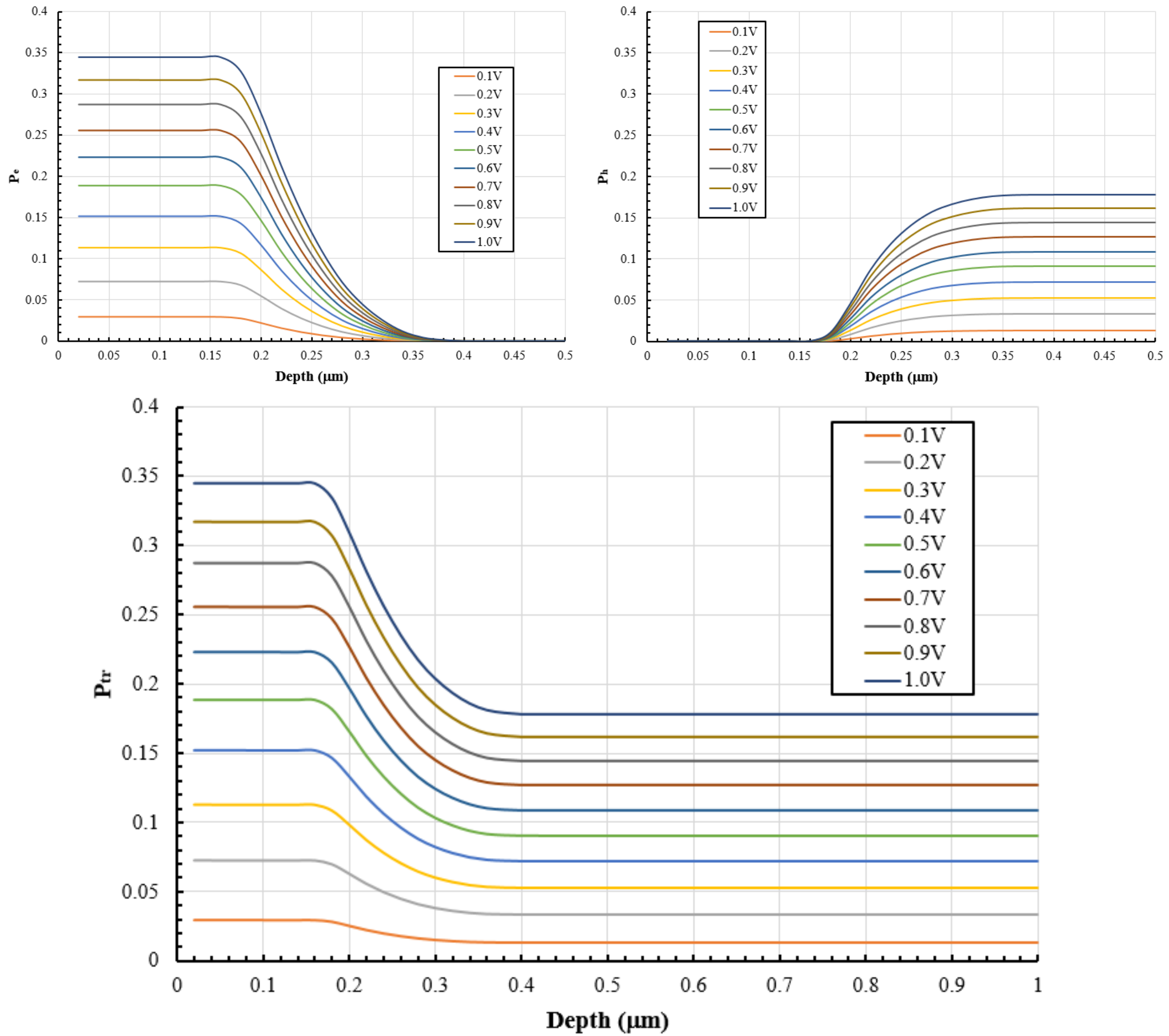


FIGURE 5.5 Avalanche triggering probability (bottom) and its components (top) vs depth for several excess bias voltages for the G12 model.

wavelength than those absorbed deeper in bulk. Thus, a relationship exists linking the avalanche triggering probability and the wavelength of these photons through the *absorption coefficient*, α. This Thesis demonstrates that photons absorbed near the surface have not only a higher quantum efficiency but also a higher probability of triggering an avalanche.

The absorption of photons in silicon follows an exponential distribution which depends on the absorption coefficient. This is calculated through the generation rate formula in a given device [7]:

$$G = \eta_0 \frac{P\lambda}{hc} \alpha e^{-\alpha y} \tag{5.4}$$

being α the absorption coefficient and y the absorption depth. Equation (5.4) will be explained further in Chapter 6. For now, it suffices to know the form of the distribution that follows the absorption of photons – see FIGURE 5.6. In this figure, the generation rate corresponding to every wavelength is normalized by the maximum value of each one, given as the absorption profiles against depth.

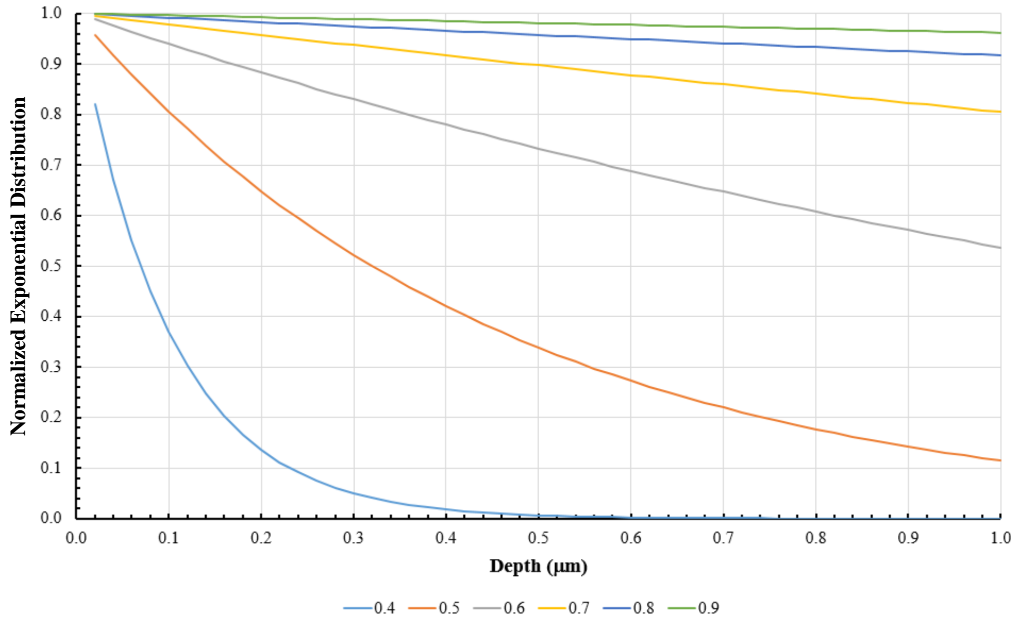


FIGURE 5.6 Normalized exponential distributions of photon silicon absorption depths for several wavelengths (µm).

The avalanche triggering probability can be obtained as a function of wavelength and excess bias voltage by using the normalized data of (FIGURE 5.6) and the weighted mean with the data provided in FIGURE 5.5:

$$P_{tr}(V_e,\lambda) = \frac{\sum_{i=0}^{n=y} \bar{\alpha}_i(\lambda) \cdot P_i(V_e)}{\sum_{i=0}^{n=y} \bar{\alpha}_i(\lambda)} \qquad (5.5)$$

being $\bar{\alpha}_i(\lambda)$ the normalized exponential distribution for a certain depth i, $P_i(V_e)$ the avalanche triggering probability for a certain depth i, and y the maximum depth at which photons can be absorbed.

The continuous drawings in Figure 5.7 show the results of equation (5.5). On the other hand, the dotted line represents the predictions made using equation (5.2), where the wavelength has no impact. Note that this old modeling technique fits long wavelengths beyond approximately 550nm. However, for shorter wavelengths, the equation (5.5) modeling technique anticipates significant variations that are indeed observed in practical SPADs.

The junction of the G12 model is 0.21µm below the surface. If the same junction were placed at, for example, the same depth as that of the V12 model, 1µm, the variation of the P_{tr} with wavelength would be lesser because, at that depth, the normalized exponential distributions dependence on the absorption coefficient shows fewer differences (FIGURE 5.8).

To include these results in the model of equation (5.2), the parameter η, which adjusts the exponential curve, must be modified according to the wavelength. This Thesis has found that a cubic polynomial equation is the best fit for this parameter:

$$\eta(\lambda) = (2.2541\,\lambda^3 - 5.3962\,\lambda^2 + 4.2961\,\lambda - 0.7772) \cdot \eta_0 \qquad (5.6)$$

being λ the wavelength in µm and η_0 a calibration parameter. This fit has a regression coefficient of $R^2 = 0.9963$. Then equation (5.2) can be generalized for the G12 model as:

$$P_{tr}(\lambda) = 1 - e^{\left(\frac{-V_E}{\eta(\lambda)V_B}\right)} \tag{5.7}$$

$$V_E = V_{ka} - V_B \tag{5.8}$$

FIGURE 5.9 shows the avalanche triggering probability dependence on wavelength for 1V of excess bias voltage for the G12 model. As the data from the G12 model were extracted from Vornicu's [1] data at 1V of excess bias voltage and a wavelength of 447nm, we calibrated the η_0 parameter, so that PDP would be the same at that wavelength.

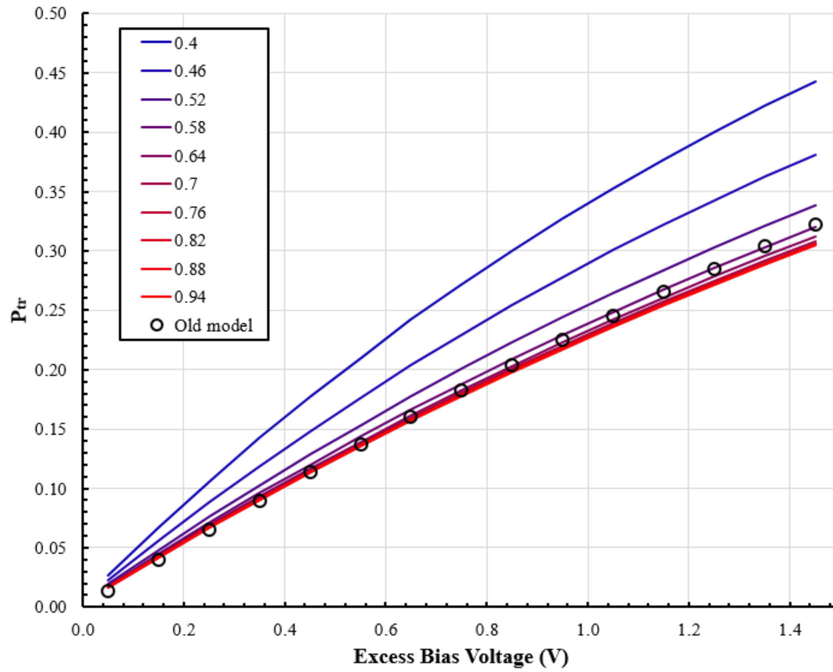


FIGURE 5.7 Avalanche triggering probability vs excess bias voltage for several wavelengths (μm). The dashed blue line represents the model from Figure 4. The junction is 0.21μm below the surface.

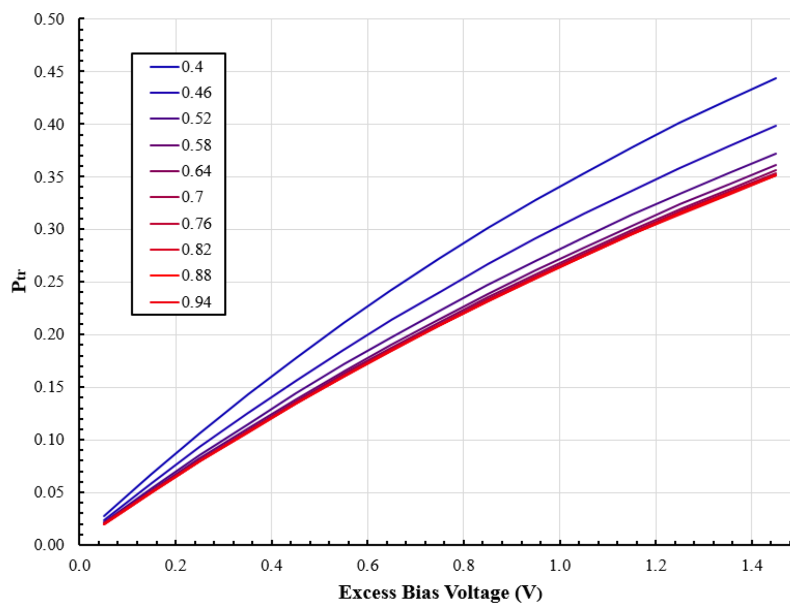


FIGURE 5.8 Avalanche triggering probability vs excess bias voltage for several wavelengths (μm). The junction is 1 μm below the surface.

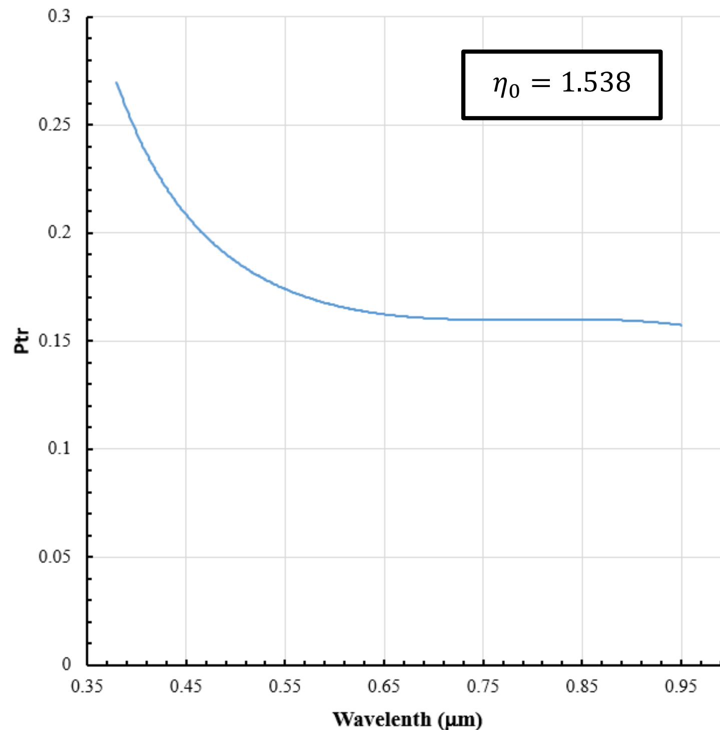


FIGURE 5.9 Avalanche triggering probability versus wavelength for the G12 model at 1V of excess bias voltage.

5.3 VERILOG-A RESULTS

The VERILOG-A model results have been obtained operating the model with a 100-kΩ passive quenching resistor (FIGURE 5.10).

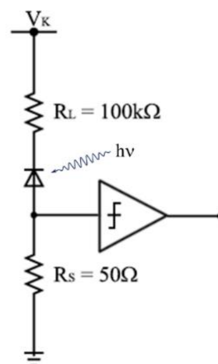


FIGURE 5.10 Basic passive quenching circuit

5.3.1 DARK COUNT RATE AND RECOMBINATION CENTER DENSITY

When making the VERILOG-A model, the most difficult parameter to determine is the **recombination center density**, N_T. Because this is a technology parameter that is not often available (see Chapter 3), a process of calibration is needed. The calibration made in this Thesis employs the experimental DCR results obtained by [1] and takes the DCR value at 0.65V of excess voltage as a reference.

Bear in mind that the VERILOG-A model in this Thesis has novel new features as compared to previous ones; among them:

- Temperature-related effects and self-heating.
- Mid-gap trap cross section temperature dependence.
- Trap-assisted tunneling (TAT) contribution to the dark count rate.
- Inclusion of Webster's spectroscopy results about traps and deep-level traps [8].
- New approximation to the inclusion of the Band-to-band tunneling (BTBT) contribution to dark counts.

Including the three first effects above required several iterative assessment cycles. Indeed, we noticed that the first runs of the model underestimate dark counts with respect to experimental results – see the green line in FIGURE 5.11. Besides, the density of recombination centers (original value not represented here) happened to be excessively large, according to other studies. This latter issue was solved with the inclusion of TAT calculations. However, the problem of dark counts was a more difficult problem to address. The inclusion of the mid-gap trap cross-section temperature dependence (yellow line) slightly improved the results, but not conclusively. Out of the two main *DCR* components, the tunneling was discarded to be responsible for the deviations as it only depends on the electric field, which remains relatively constant beyond the breakdown voltage. That only left the thermal component, as it is essentially determined by the effective carrier lifetime from equation (3.20), and through it, the recombination lifetimes for electrons and holes from equations (3.28) and (3.29), respectively, which we reproduce again here:

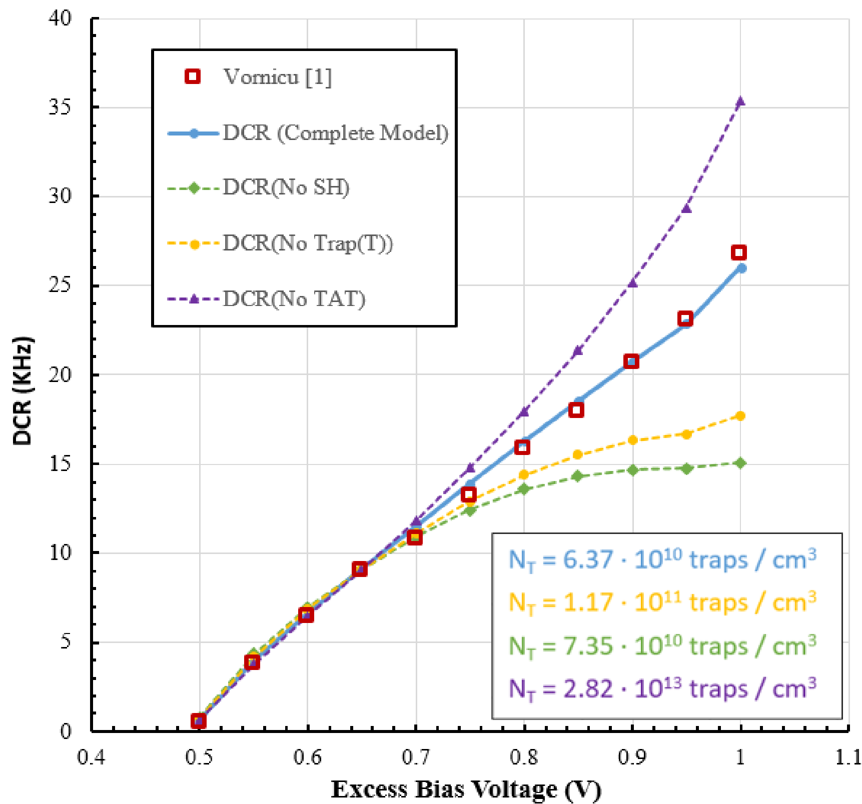


FIGURE 5.11 Dark count rate vs excess bias voltage for several instances of the G12 model compared to Vornicu's work [1]. Recombination center density values for every instance are showed.

$$\tau_{n_0} = \frac{1}{v_{de} N_T \sigma_T (1 + \Gamma_n)} \tag{5.9}$$

$$\tau_{p_0} = \frac{1}{v_{dh} N_T \sigma_T \left(1 + \Gamma_p\right)} \tag{5.10}$$

If these equations are examined, it can be seen that leaving aside Γ and σ_T, being both well documented, the only two other suspects are the **drift velocity**, v_d, of the charge carriers and the **recombination center density**, N_T. As changing the latter implies losing the adjustment below 0.7V in excess, only the former is left. The drift velocity only depends on the square root of the temperature, so it was concluded that the actual device was suffering from an increase in the internal temperature that was boosting the charge carriers' drift velocity. Effectively, the device was undergoing **self-heating**.

The greater the excess bias voltage, the larger the self-heating. When including the necessary calculations, the final results were those that correspond to the blue line. These match the experimental results quite well, getting a relative error no greater than 7% from 0.55V excess bias voltage onwards. With these results, it can be said that:

- The SPAD is indeed heavily affected by self-heating.
- Not taking the TAT contribution to *DCR* can lead to an overestimation of dark counts as a result of overestimating N_T by more than two orders of magnitude.

TABLE 5.2 compares the value estimated for N_T in this work and others in the literature corresponding to other SPAD models in VERILOG-A. The work of Giustolisi [6] lacked several essential features, so its assumption for N_T may not be realistic. While He's and Cheng's works [9][10] were more complete, our belief is that they overestimated it by not including some features that were included in our model. For instance, He's work did neither include the BTBT nor the TAT in his dark counts' calculations, while Cheng's work lacked the inclusion of the TAT.

TABLE 5.2 Recombination center density

References	$N_T [cm^{-3}]$
Giustolisi 2012 [6]	9×10^{9}
He 2013 80[9]	8×10^{11}
Cheng 2016 [10]	$8.7 \sim 10.7 \times 10^{11}$
This Thesis	6.37×10^{10}

FIGURE 5.12 (right) shows how *DCR* and its components vary with temperature. The thermal component of the *DCR* clearly grows exponentially with the ambient temperature, while the BTBT slightly decreases with temperature, as expected and reported by other works [8]. The after-pulsing effect shows a mild increase, but these results will be discussed in the next section.

For a fair comparison of device models, an equal defect density must be assumed. This can be done now that the recombination center density has been calibrated. FIGURE 5.12 (left) shows a *DCR* comparison between the G family and the V12 structure. As expected, the *DCR* grows with the active area region radius, and the V12 model is a huge improvement over the G12 model, decreasing the *DCR* by 95%. However, these values are still higher a few times than those obtained by Richardson [4]. It has to be emphasized that Richardson's devices employ a 130nm non-standard technology, which is newer and will supposedly have a reduced recombination center density, making the thermal noise much lower.

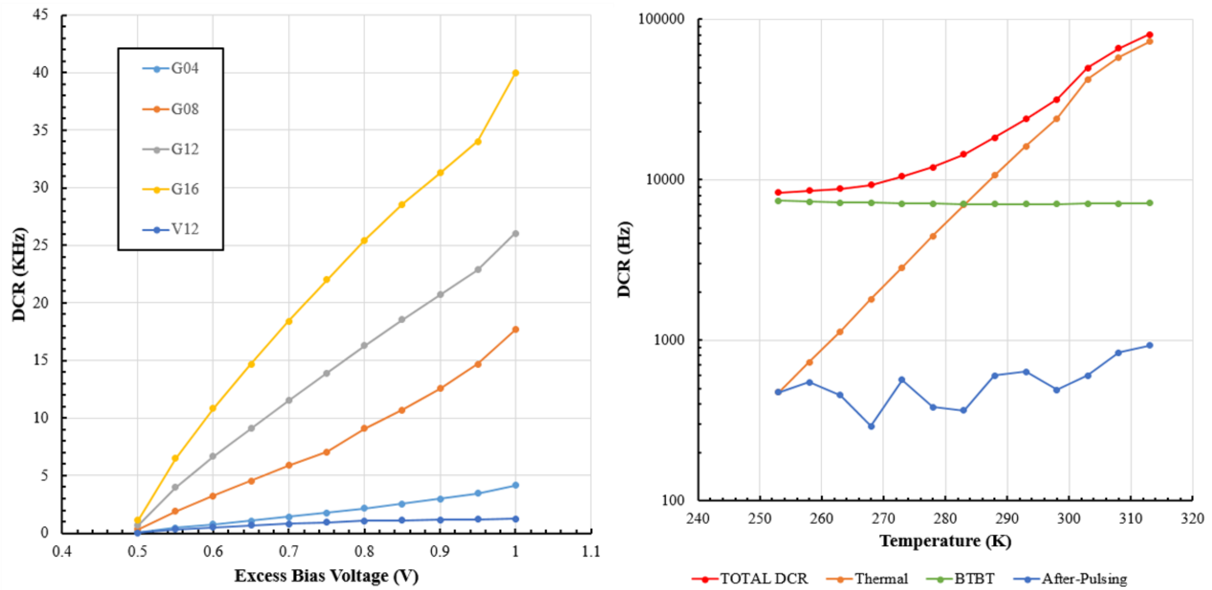


FIGURE 5.12 (Left) Dark count rate vs excess bias voltage for several Verilog-A models. (Right) Dark count rate vs temperature for several components of the dark count rate for the G12 model at 1V of excess bias voltage.

5.3.2 AFTER-PULSING

After-pulsing results, shown in FIGURE 5.12 (right), present a somewhat irregular increasing trend with temperature. This is due to the simulation times of VERILOG-A, which typically are in the order of a few dozen of microseconds, while experimental results are obtained over the extent of times several magnitudes larger. Nonetheless, FIGURE 5.13 shows how after-pulsing probability (equation (3.42)) presents a downward trend, albeit also irregular, with the temperature. This is because the after-pulsing dependence on temperature is weaker than that of the primary dark counts due to thermal generation. That is, dark counts due to after-pulsing

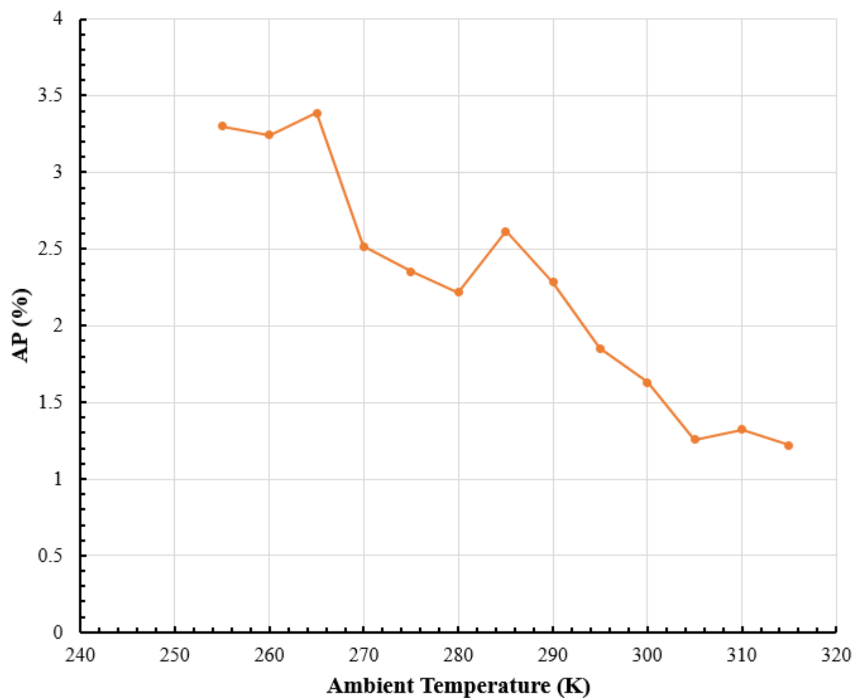


FIGURE 5.13 After-pulsing probability vs ambient temperature for the G12 model.

do rise with temperature, but the percentage is lower. After-pulsing results show a clear dependence with the excess bias voltage; however, at very low excess bias at those times of simulation, results are insufficient to be reliable. It has to be taken into account that FIGURE 5.12 (right) after-pulsing results are an extrapolation of shorter times results, and the researchers of this work believe that the original after-pulsing count variation with temperature is not reliable enough to pick out a trend.

5.3.3 SELF-HEATING

Results from simulations suggest that the experimental device suffers from self-heating. The causes of this self-heating seem to be related to the avalanche current. However, it is not crystal clear if the main cause of self-heating is the SPAD itself or the pixel circuitry. Regardless, this effect has been simulated according to equations (3.59) and (3.60).

FIGURE 5.14 shows that self-heating mainly affects smaller devices, which makes sense as it should be easier to heat a smaller volume. However, given the G04 and G08 device results, it appears to exist an upper limit depending on the excess bias. On the other hand, the V12 device would suffer less from this problem because of the lesser avalanche current.

Another interesting result is that the self-heating depends linearly on the ambient temperature (FIGURE 5.15), making some cooling device desirable. In this case, the V12 model seems almost immune to this contribution, which indicates another advantage of the PW SPAD device over the P+ SPAD.

5.4 REFERENCES

[1] I. Vornicu, R. Carmona-Galán, B. Pérez-Verdú, and Á. Rodríguez-Vázquez, “Compact CMOS Active Quenching/Recharge Circuits for SPAD Arrays.” *Int. J. of Circuit Theory and Applications*, vol. 44, no. 4, pp. 917-928, Apr. 2015.

[2] N. Faramarzpour, M. J. Deen, S. Shirani, and Q. Fang, “Fully Integrated Single Photon Avalanche Diode Iin Standard CMOS 0.18µm Technology.” *IEEE Transaction on Electron Devices*, vol. 55, no. 3, pp. 760–767, Mar. 2008.

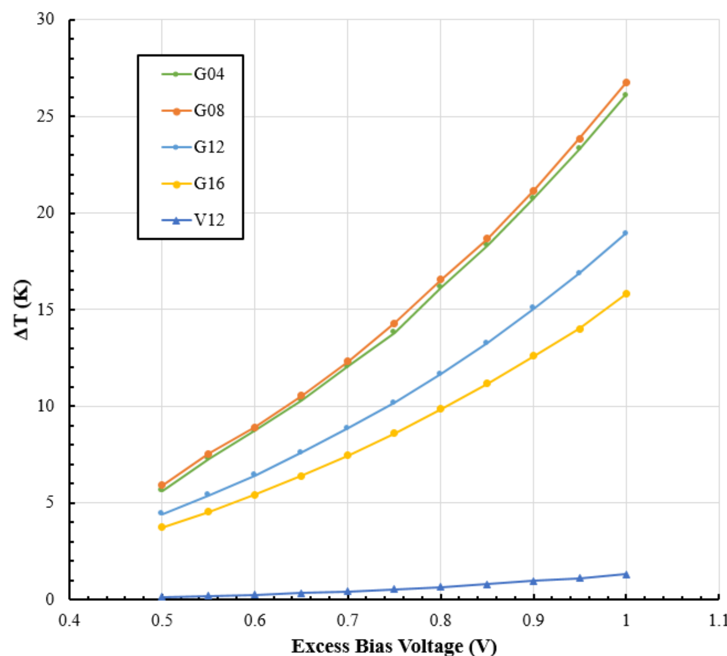


FIGURE 5.14 Self-heating vs excess bias voltage for several models

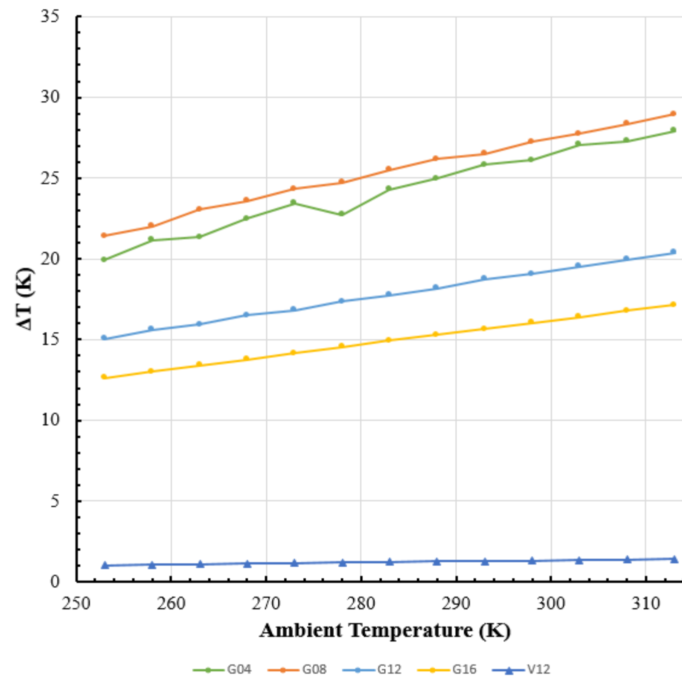


FIGURE 5.15 Self-heating vs ambient temperature for several models.

[3] J. Cao, L. Wang, C. Shen, L. Liu, and N. Wu, "A Single Photon Avalanche Diode in Standard CMOS Technology for Range Finding." *2016 13th IEEE International Conference on Solid-State and Integrated Circuit Technology (ICSICT)*, 2016, pp. 674-676, doi: 10.1109/ICSICT.2016.7999008.G.

[4] 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 Transactions on Electron Devices*, vol. 58, no. 7, pp. 2028-2035, July 2011.

[5] C. Niclass, M. Gersbach, R. Henderson, L. Grant, and E. Charbon, "A Single Photon Avalanche Diode Implemented in 130-nm CMOS Technology." *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 13, no. 4, pp. 863-869, July-aug. 2007.

[6] G. Giustolisi, R. Mita, and G. Pallumbo, "Behavioral Modeling of Statistical Phenomena of Single-Photon Avalanche Diodes." *Int. J. of Circuit Theory and Applications*, vol. 40, no. 7, pp. 661–679, 2012.

[7] *Atlas User's Manual: Device Simulation Software*. Silvaco Inc., Santa Clara, CA, 2016.

[8] E. A. G. Webster and R. K. Henderson, "A TCAD and Spectroscopy Study of Dark Count Mechanisms in Single-Photon Avalanche Diodes." *IEEE Transactions on Electron Devices*, vol. 60, no. 12, pp. 4014-4019, Dec. 2013.

[9] Q. He, Y. Xu, and F. Zhao, "An Accurate Simulation Model for Single-Photon Avalanche Diodes Including Important Statistical Effects." *J. Semicond.*, vol. 34, no. 10, pp. 104007-1–104007-6, Oct. 2013.

[10] Z. Cheng, X. Zheng, D. Palubiak, M. J. Deen, and H. Peng, "A Comprehensive and Accurate Analytical SPAD Model for Circuit Simulation." *IEEE Trans. Electron Devices*, vol. 63, no. 5, pp. 1940–1948, May 2016.

CHAPTER 6

THEORY AND MODELING OF RESPONSE TIME AND PHOTON-TIMING JITTER IN SPADS

6.1 INTRODUCTION

Deviations from the nominal, ideal time instances of electrical waveforms constitute one of the primary error sources in circuits intended for timing. In electronic oscillators, these deviations are related to *phase errors* and produce random changes in the oscillation frequency along the operation cycle [1]. In SPAD circuits, random changes in the time event stamps result in errors in the TOF evaluation and hence in the depth of objects. Jitter assessment is hence crucial for SPADs and their circuit models.

SPAD models in previous chapters consider that SPADs are dimensionless, point objects. Hence, these models have almost instantaneous response time, depending only on the time constants associated with the junctions and stray capacitances in the range of fentoFarads. Thus, the place where photons are absorbed and the time it takes for them to reach the depletion region are omitted from the analysis. Indeed, previously reported SPAD models do share the drawback of not including do not include photon-timing jitter. One reason for this lack is the complicated relationships between technological data, which are not complete in many cases, and the absorption coefficients and the electric field profile on the device when it is operated in Geiger mode. However, including this feature in a VERILOG-A description based on ***analytical models*** would provide clear benefits, as it would not require time-costly Monte Carlo computations. This chapter shows that analytical modeling of the photo-timing jitter provides insight into the SPAD device and how the quenching circuit affects the results and how this offers clues to make improvements.

6.2 PHOTON-TIMING JITTER MECHANICS

When a photon successfully strikes the active area region of SPAD, it can potentially create an electron-hole pair along its way inside the device. The probability of this photon being detected by the SPAD is governed by the photon detection efficiency, which expression and components are described in previous chapters and repeated here for easier reading:

$$PDE = QE(\lambda) \cdot FF \cdot P_{tr} \tag{6.1}$$

where $QE(\lambda)$ is the quantum efficiency (evaluable by TCAD simulations), FF is the fill factor, and P_{tr} the avalanche triggering probability.

Once the electron-hole pair is created within the device, both carriers will accelerate diffusion according to the electric field profile present in the SPAD until they either trigger the first impact ionization or recombine; the former option may lead to an avalanche. The time it takes for an electric charge carrier to diffuse and provoke that first ionization is called *diffusion time*. After the first impact ionization occurs, subsequent impact ionizations occur due to the accelerated charge carriers created in the first impact. At this point, the avalanche ignition may still fail. The time it takes for the avalanche to stabilize finally is called *build-up time*. Finally, after the avalanche is stabilized, it spreads throughout the active area of the SPAD. The time it takes for the avalanche to spread and reach a minimum current threshold detectable by the circuitry is the spread time. Altogether, they add to make the avalanche response time, whose statistical fluctuation is the photon-detection timing jitter (FIGURE 6.1).

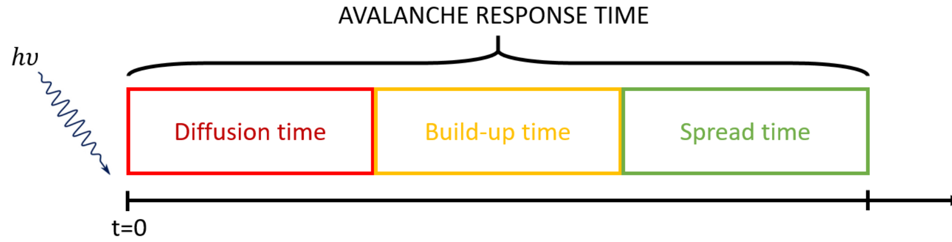


FIGURE 6.1 Avalanche response time components.

6.2.1 DIFFUSION TIME

This study models the different components of the timing jitter under the assumption that a photon has been detected and has triggered an avalanche Electron-hole pairs, able to originate this avalanche, can only be created in a determined zone of the SPAD, as shown in FIGURE 6.2. Here the module and direction of the electric field can be seen within a portion of the active area region of the SPAD. Notice that at, approximately 1µm-deep, the vertical component of the electric field changes direction, and carriers generated below that point cannot reach the depletion region and trigger an avalanche, establishing a sensing limit. Besides, any carrier created in any of the neutral regions may have diffused to the depletion region (electrons for the shallow region and holes for the deep region), accelerated until they reach the saturated drift velocity, and triggered an impact ionization which might result in an avalanche current. The time it takes for these carriers to reach the depletion zone and trigger the first impact ionization, which may or may not trigger an avalanche, is called diffusion time. Notice that even carriers created in the depletion region, where the electric field is maximum, have to diffuse and gain momentum to impact-ionize for the first time. Spinelli states that the minimum time for a carrier to gain this needed momentum is around 10ps [2].

The first step in calculating this diffusion time is to estimate the **depth at which the photon has been absorbed**. As we have assumed that the avalanche has taken place, the photon had to

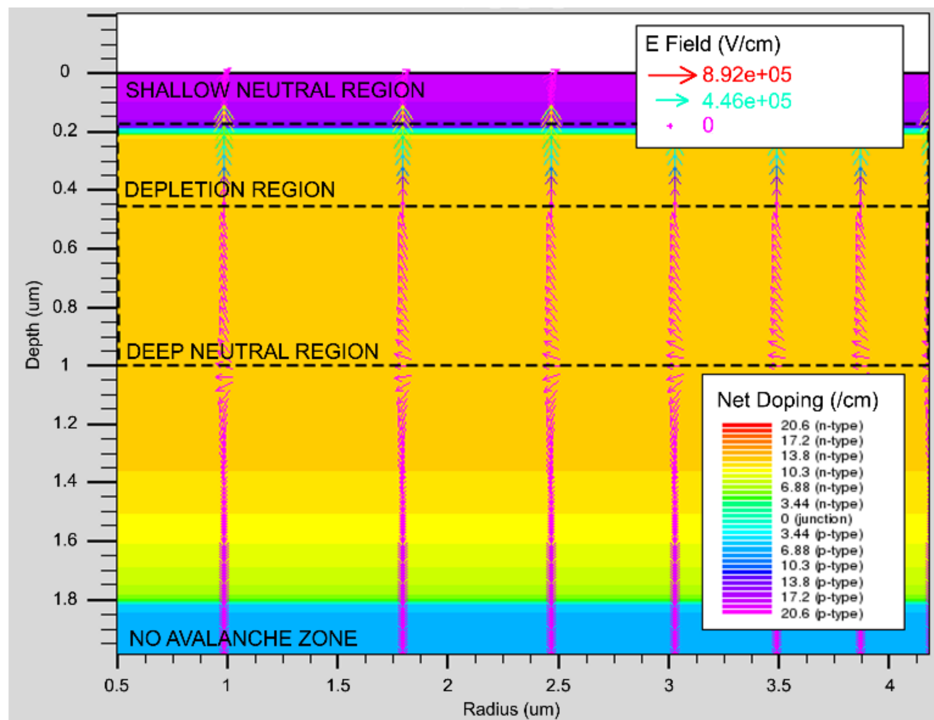


FIGURE 6.2 Electric field profile in the SPAD active area region.

be absorbed no further than one micrometer-deep in this case. To make a statistical law that generates a random depth to be absorbed, we define the charge carriers' generation rate of a given wavelength λ as [3]:

$$G = \eta_0 \frac{P\lambda}{hc} \alpha e^{-\alpha y} \tag{6.2}$$

$$\alpha = \frac{4\pi}{\lambda} k \tag{6.3}$$

where: α is 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 is the depth of absorption, and k is the imaginary part of the optical index of refraction or extinction coefficient.

One main problem here is that the **extinction coefficient**, and therefore the **absorption coefficient**, is dependent on temperature and, more importantly, on the doping concentration. The extinction coefficient in low-doped silicon is well known and offers no significant variations when the doping concentration is less than $10^{17}\,\text{cm}^{-3}$. In this range, the coefficient follows a quasi-exponential law in the optical spectrum [4]. However, our model contains regions where the doping concentration is far greater, and therefore it has to be taken into account. Through TCAD simulation, we have determined that the absorption coefficient in the active area region of our device can be adjusted to the following power law:

$$\alpha(\lambda) = 8.254 \times 10^{-26} \cdot \lambda^{-6.84}\ [\text{cm}^{-1}] \tag{6.4}$$

Note that equation (6.2) follows an exponential law. As we assume that the photon is actually detected, we must limit the range of this equation to the region of possible detection. However, such a constraint can hardly be implemented in VERILOG-A**.** To overcome this limitation, we rely on a **uniform distribution** that can be easily constrained to generate the constrained exponential distribution:

$$P(\lambda) = \frac{-\ln(1 - a \cdot U)}{\alpha(\lambda)} \tag{6.5}$$

$$a = 1 - e^{-\alpha(\lambda)\cdot d} \tag{6.6}$$

where: U is the uniform distribution; a is a normalization value that establishes the maximum probability of the uniform distribution (due to the fact that the photon absorption range does not cover all the space as we assumed that it was indeed absorbed in the 0-1μm range and therefore the probability must be limited); and d is the absorption width of the sensitive area in centimeters, in this case 10^{-4}cm.

The inclusion of d makes the result of equation (6.5) directly the statistical fluctuation of the photon absorption depth. Equation (6.6) is the result of clearing a in equation (6.5) assuming $P(\lambda) = d$ and $U = 1$. For example, we wish to generate a random photon absorption of a given wavelength λ. As we assume this photon has been detected, the range of absorption depth is $[0, d]$, following an exponential distribution (equation 6.2). Equation 6.6 calculates a limit, $a \in (0,1)$, to the uniform distribution, U (which range is $[0,1]$), following the $\alpha(\lambda)$ shape modulated by d. By multiplying $a \cdot U$ in equation (6.5), the range would be limited to $[0, a]$.

Then, the result of equation (6.5) is a random depth of photon absorption in the $[0, d]$ range, following an exponential distribution that depends upon the form of $\alpha(\lambda)$.

FIGURE 6.3 shows the normalized exponential distributions of the probability of photon absorption for several photon wavelengths. Photons of shorter wavelength distributions are more condensed around the depletion region; thus, it is more probable for those photons to be absorbed there and detected.

FIGURE 6.4 shows in what region of the SPAD is more probable that a photon of a given wavelength strikes and creates an electron-hole pair. As can be seen, shorter wavelengths tend to be absorbed near the surface of the device and longer wavelengths deep in bulk. These results are in accordance with the performance of real-like SPADs.

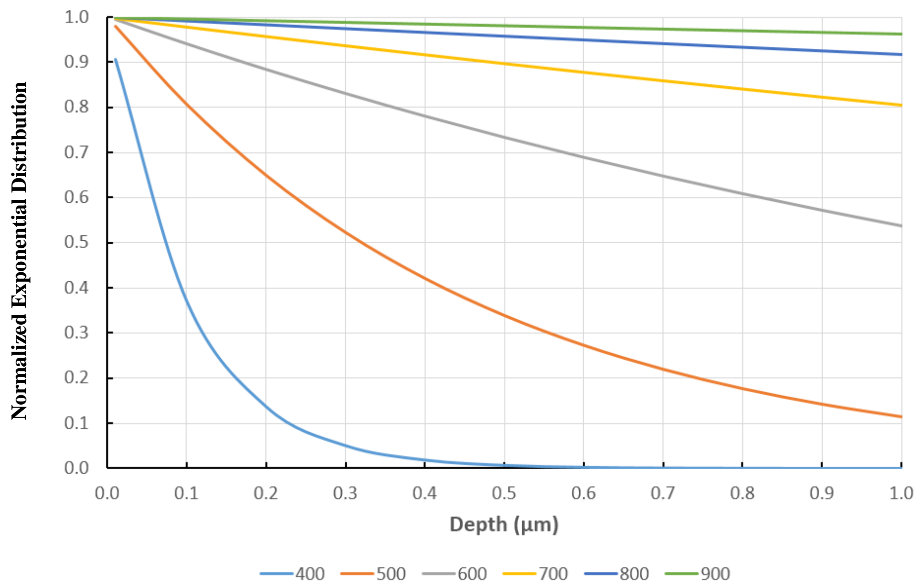


FIGURE 6.3 Normalized exponential distributions of photon absorption depths for several wavelengths (nm).

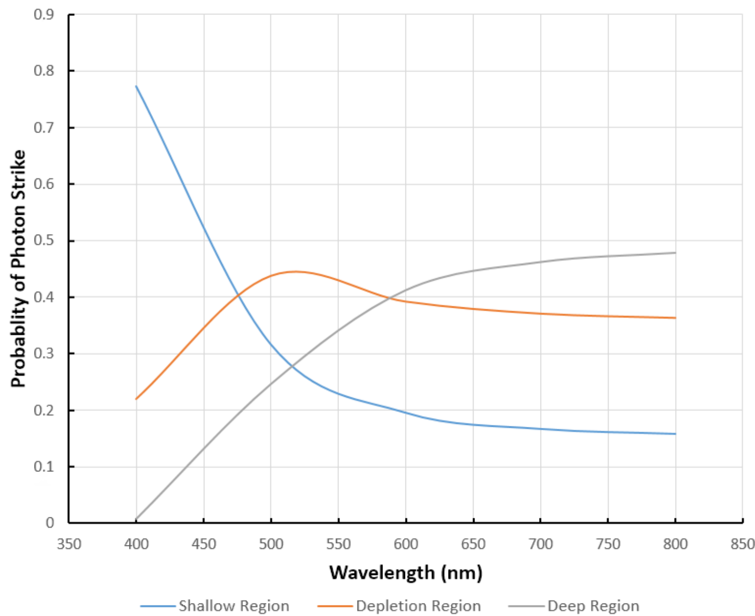


FIGURE 6.4 Probability of photon strike vs wavelengths for the different regions of the SPAD.

In the SPAD model described in Chapter 4, we emulated generating photons by raising a voltage signal over a determined threshold. Now, those photons are given a **random absorption depth** based on the equations (6.5) and (6.6).

Once the depth at which the electron-hole pair was created is determined, the time the charge carrier takes to diffuse to the depletion zone needs to be estimated. This can be done by extracting the **carrier velocity profile** from the TCAD model and then integrating these velocity curves to arrive at the following expressions for the diffusion time for electrons in the shallow region and holes in the deep region, respectively:

$$t_{dn} = \frac{e^{-v_{n2} \cdot Y} - e^{-v_{n2} \cdot y_0}}{v_{n1} \cdot v_{n2}} \quad (6.7)$$

$$t_{dp} = \frac{\ln|v_{p1} \cdot Y + v_{p2}| - ln|v_{p2}|}{v_{p1}} \quad (6.8)$$

being Y the absorption depth, y_0 the shallow limit depth of the depletion region, and v_{n1}, v_{n2}, v_{p1} and v_{p2}, velocity parameters that depend linearly on the excess bias voltage:

$$v_{n1} = 2136\, V_E + 6952.1 \quad (6.9)$$

$$v_{n2} = -2.487\, V_E + 26.636 \quad (6.10)$$

$$v_{p1} = -1.89 \text{ x } 10^6\, V_E - 1.9 \text{ x } 10^6 \quad (6.11)$$

$$v_{p2} = 2.13 \text{ x } 10^6\, V_E - 2.04 \text{ x } 10^6 \quad (6.12)$$

FIGURE 6.5 shows the electron velocity profile versus depth for SPAD active region; a similar figure is observed for holes. In the shallow neutral region (left of the depletion region), the profile follows an exponential fit, whereas the fit is lineal in the deep neutral region,

6.2.2 BUILD-UP TIME

Since it is known that there is **noise** associated with the avalanche multiplication process [2][5], a contribution to the timing jitter is expected to arise from this source. The current resulting from the first carrier, which has an estimated initial value of $2 \cdot 10^{-8}$A [2][5] follows an irregular time pattern due to the ongoing multiplication process, and it is not stabilized until it reaches 10^{-6}A. Hence, this contribution to the timing jitter cannot be reduced by lowering the discriminator threshold, which is as low as 10^{-4}A in the literature. The time that takes the current to reach that level has a normal distribution and depends on the excess bias voltage and the electric field profile. As earlier works failed to match experimental results of the build-up time at low excess bias voltages [2][5], we decided to use the experimental results obtained by Ingiargola. (FIGURE 6.6).

Ingiargola developed two SPAD models [5], namely **S44** and **S62**; the only difference between them is in the electric field profile, the first having a larger maximum electric field.. Using the results from FIGURE 6.6, we developed a numerical fit consisting in piecewise relation with a power and a linear equation that fits Ingiargola's measurements and is consistent with Spinelli's work [2]:

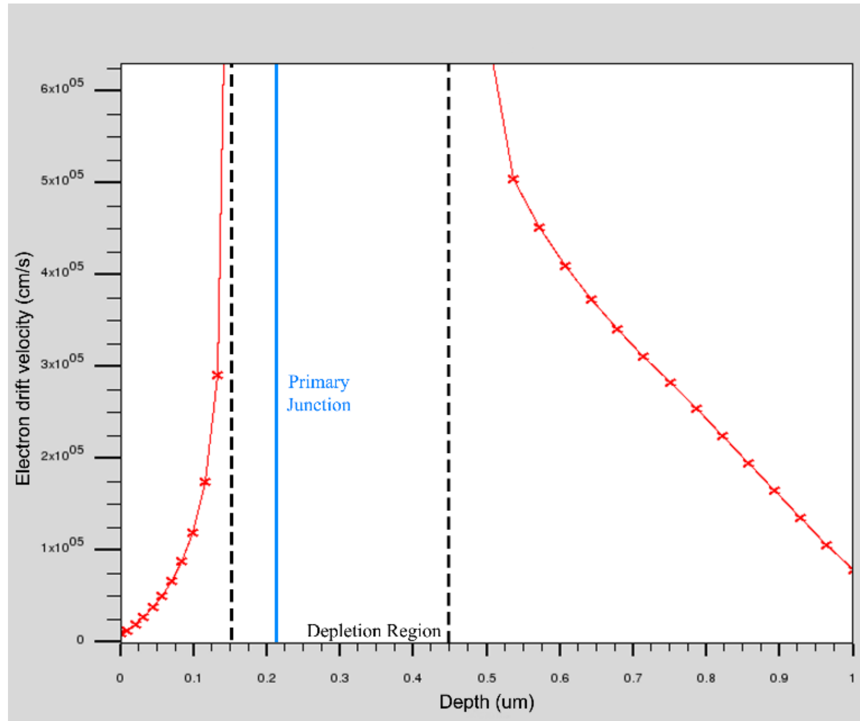


FIGURE 6.5 Electron drift velocity (in red) vs depth.

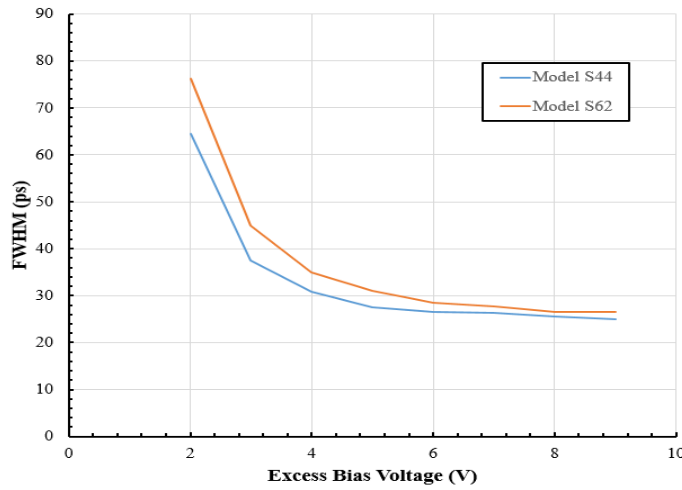


Figure 6.6 Ingiargola's build-up times for his developed models [5].

$$t_{b,FWHM} = \begin{cases} B_1(E_m) \cdot V_E^{B_2(E_m)} & V_E < V(E_m) \\ B_3(E_m) \cdot V_E + B_4(E_m) & V_E \geq V(E_m) \end{cases} \text{[ps]} \tag{6.13}$$

where: E_m is the primary junction maximum electric field of the non-uniform electric field profile that characterizes real like SPADs (MV/cm); V_E is the excess bias voltage; $V(E_m)$ is a potential value representing the knee point between the power fitting and the linear fitting (first and second terms of equation (6.13)); and $B_x\ (E_m)$ are parameters that depend linearly on the electric field and fit equation (6.13):

$$B_1(E_m) = -449.85 \cdot E_m + 369.5 \tag{6.14}$$

$$B_2(E_m) = 1.01 \cdot E_m - 1.501 \tag{6.15}$$

$$B_3(E_m) = 2.092 \cdot E_m - 1.717 \tag{6.16}$$

$$B_4(E_m) = -37.14 \cdot E_m + 50.877 \tag{6.17}$$

Analysis shows that the knee point has a quadratic dependence on the electric field:

$$V(E_m) = -50.15\, E_m^2 + 49.2\, E_m\text{-}7.096 \tag{6.18}$$

Our results show that the primary electric field contribution dominates over the excess bias contribution to the build-up time. Also, a low excess bias or electric field increases the build-up time. According to the device applications, these results may lead to important decisions, such as engineering the primary junction to avoid an excessive high electric field junction, and the subsequent noise due to band-to-band tunneling may lead to higher jitter due to the build-up time contribution.

6.2.3 SPREAD TIME

Once the avalanche is stable, it spreads laterally fast through the active area region. As already described in [2], the lateral spreading of the avalanche's free carriers decreases the electric field just before the upcoming avalanche front. As the avalanche spreads, the front widens, increasing the density of free carriers, and thus the electric field decreases evermore as the avalanche keeps spreading. The consequence of this is that the lateral spreading velocity of the avalanche falls gradually. The model of the spreading velocity of the avalanche front for our SPAD is similar to that of [2] and is given by [6]:

$$v_p(r) = \begin{cases} v_{p0} & r < r_0 \\ v_{p0}\, r^{-k_1} & r \geq r_0 \end{cases} \tag{6.19}$$

$$v_{p0} = 2\sqrt{\frac{D \cdot V_E}{\beta}} \tag{6.20}$$

where D is the diffusion coefficient of the charge carriers, r is the avalanche front radius in micrometers, $r_0 \approx 1\mu\text{m}$ four our devices, V_E is the excess bias, β is a constant that in our devices is 5.71×10^{-12}ps V, and $k_1 = 0.359$ is a constant that is related to the spread of free carriers from the circular form of the avalanche spreading front. FIGURE 6.7 illustrates equation (6.19).

As the avalanche spreads through the SPAD area, the avalanche current builds up until the avalanche reaches the whole area. At this point, the avalanche current is maximum. Supposing the avalanche front spreading velocity is known, the percentage of the SPAD active area region undergoing avalanche can be calculated at any given time, and thus the percentage of the maximum avalanche current at that moment. When this current reaches the minimum detectable current or **threshold current**, I_{min}, imposed by the quenching circuitry, the avalanche is detected, and a value can be assigned to the spreading time and, of course, to the timing response of the avalanche. The problem here is determining that threshold current in actual pixels, but this can be circumvented by taking into account that the maximum avalanche current, I_{SPAD} (equation 3.6), depends on the excess bias voltage. If this maximum current is lower than the minimum detectable current, avalanche triggering may be occurring, but our circuitry should

not be able to detect it. Then, a minimum excess bias of avalanche detection can be established. At that limit, the threshold current is the maximum avalanche current and can be deduced from equation (3.6).

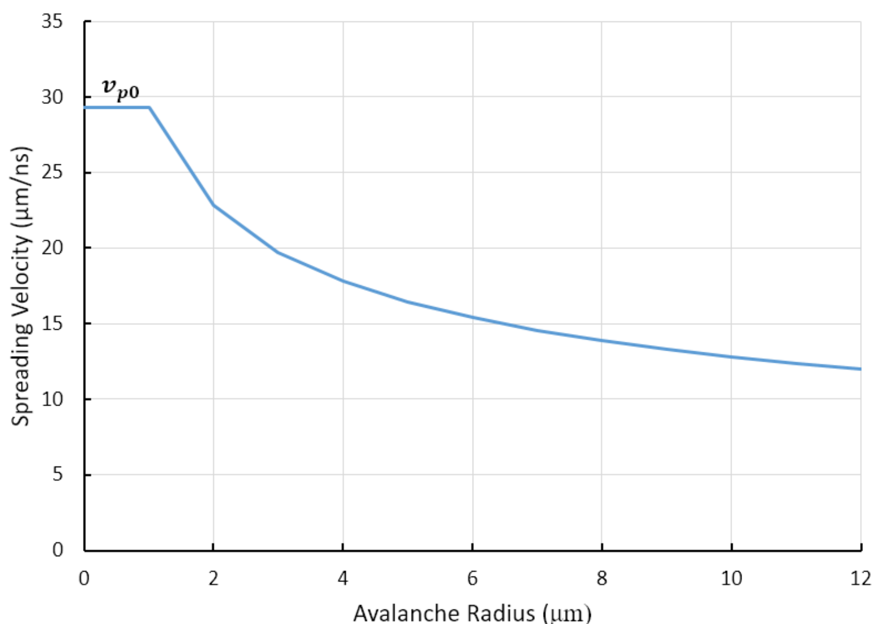


FIGURE 6.7 Spreading velocity of the avalanche front vs the avalanche radius.

6.3 SIMULATION AND EXPERIMENTAL VALIDATION

The VERILOG-A model analytic SPAD model described in Chapter 3 (Figure 3.5) cannot directly embed features like the response time and the photon-timing jitter. Hence, it must be modified to enable the emulation of photon strike, its absorption and diffusion, and the modeling of the ignition of the subsequent spreading of avalanches. These additions are far from trivial due to the difficulties of performing analytical estimations of the triggering probability. This parameter encompasses:

- the probability of the charge carriers successfully making the first impact ionization and not recombining while drifting; and
- the probability of the avalanche stabilizing successfully.

Instead of resorting to analytical procedures, it is simpler to determine it experimentally and then do the calculations once it is known that the avalanche is happening after checking the triggering probability.

Calculation of the response time involves the following steps:

- First of all, a random **photon absorption depth** is assigned through equation (6.5), and a diffusion time is established with equations (6.7) and (6.8).
- We assume that the avalanche has happened and is stabilized; hence, we discard the probability of failing stable avalanches, thus enabling the calculation of a **build-up time** using equation (6.13).
- We compute the time needed for the avalanche to spread to the percentage of the SPAD area that results in the minimum detectable current or **threshold current**. As, in principle, this parameter is unknown, the threshold current is used as incognita to validate our model, assigning it several values in an iterative procedure.

FIGURE 6.8 shows an example of a run simulation. Every change of height in both graphs represents the strike of a new photon in the simulated device. When a photon is absorbed out of the depletion zone, the response is higher and rises more the further away it is from it.

Similar to previous works [5][7], we found that the timing jitter of a SPAD is heavily influenced by the threshold current of the quenching circuit, namely:

- if that threshold is low enough (below 2.5mA in the G12 device structure), then the timing jitter will be dominated by the **build-up time**, as the spread time would be negligible;
- otherwise, the timing jitter is dominated by the **spread time**.

We have tried to find how the G12 model behaves, supposed the quenching circuitry imposes a minimum threshold current. To do that, we have limited the spreading area of the avalanche when the current has reached that threshold. FIGURE 6.9 shows the timing jitter for various excess voltages versus the anode voltage, V_A. If we take a look to FIGURE 5.10, we can see that the avalanche current at that anode terminal depends upon the R_S resistance. By measuring V_A instead, our measures do not depend upon the circuit elements that may be between the anode terminal and the *ground*, and thus, these results can be compared to others works that use another circuit configuration.

We can see in FIGURE 6.9 that the higher the threshold, the higher the contribution of the spread time to the timing jitter. Also, a higher excess voltage improves the avalanche response, as both build-up and spreading develop faster. FIGURE 6.10 shows an example of photon-timing

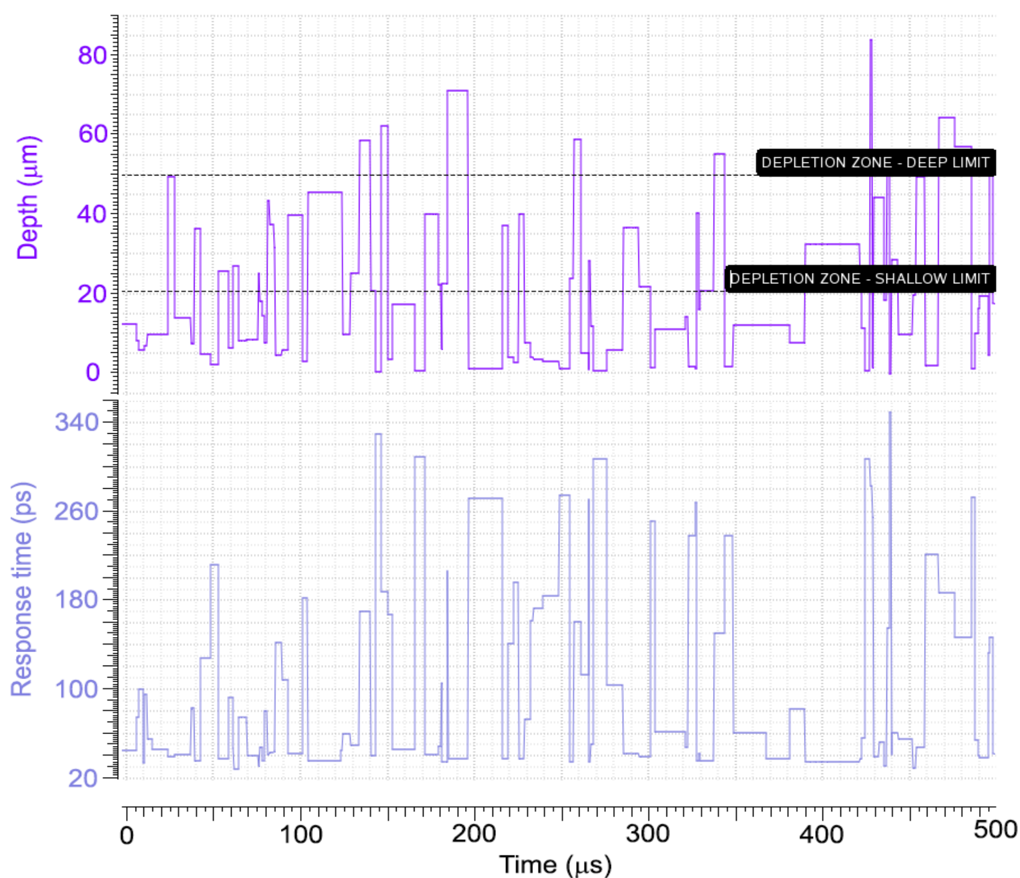


FIGURE 6.8 Photon strike depth and response time in a simulation run. The excess bias voltage is 1V.

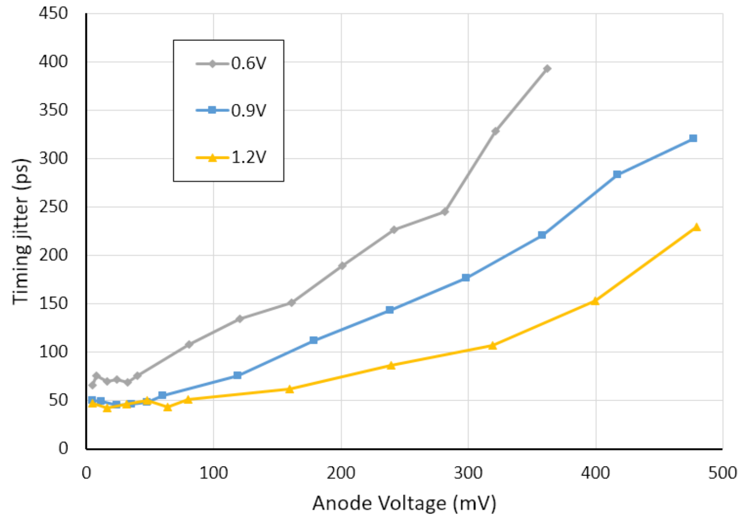


FIGURE 6.9 Timing jitter vs anode voltage for several excess bias voltages.

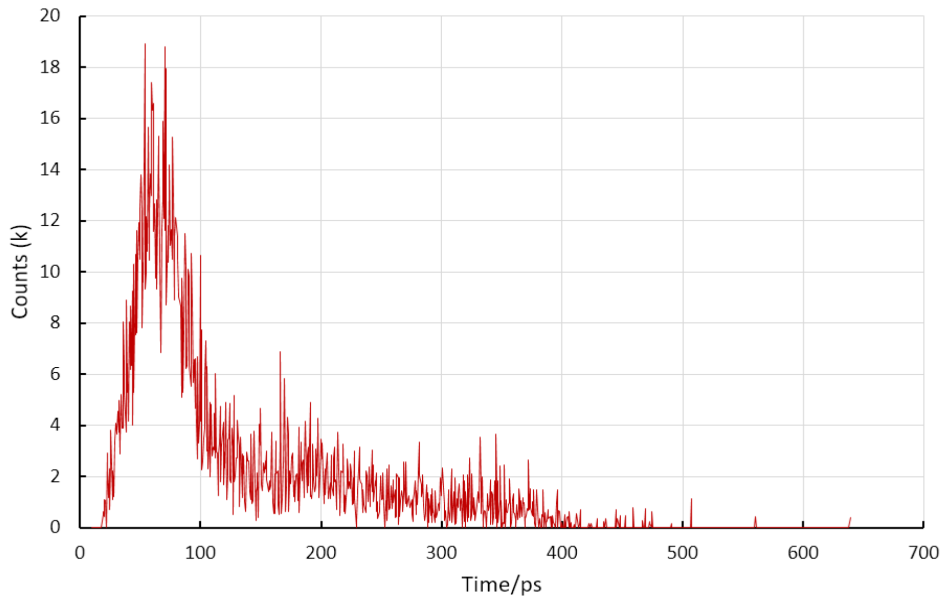


FIGURE 6.10 Example of timing jitter results for a 20 ms run simulation, under 0.9V of excess bias when the SPAD is illuminated with pulses of 80 ps long, with a 447 nm wavelength beam with a frequency of 2.5 MHz.

jitter obtained for a low threshold current, where the build-up time (~42ps) dominates over the spread time. In this case, the threshold current is 0.1mA with a photon-timing jitter represented in the form of the FWHM of 48ps. If we observe FIGURE 6.9, this value represents the lowest photon-timing jitter the SPAD can achieve if we manage to lower the threshold current imposed for the quenching circuit to that value.

The results also showed that under an excess bias of 0.9V, the 187ps of timing jitter obtained experimentally could also be obtained for an anode voltage of 313mV corresponding to a threshold current of 6.26mA with the avalanche covering the 52.5% of the SPAD active area. Interestingly, this implies that any avalanche whose maximum current, when expanded to the entire active area, is less than that of the threshold current of 6.26mA will not be detected. As the maximum avalanche current depends on the excess bias, the lower limit to the excess bias mentioned before can be obtained. Therefore, the excess bias corresponding to this threshold current is 0.47V. Indeed, it can be seen how the experimental photon detection efficiency (PDE) results obtained in [8] plummets close to 0.5V excess bias as the circuitry becomes unable to

detect avalanches in the SPAD. The similarity of the model and experimental results provides a qualitative argumentation for model validation.

These results are compatible with other SPADs in the literature. Reference [9] reports several 8μm-diameter devices, all of which have greater timing jitters than the one described in our model, both running with an excess bias of 1.2V. The difference resides in the engineered primary junction of these SPADs, which presents a lower electric field (~0.6MV/cm) than the one employed by this work (0.729MV/cm).

6.4 REFERENCES

[1] A. Rodríguez-Vázquez and I. Vornicu, “Harmonic Oscillator: Basic Concepts and Circuits.” *Wiley Encyclopedia on RF Circuits*, 2023.

[2] A. Spinelli and A. L. Lacaita, “Physics and Numerical Simulation of Single Photon Avalanche Diodes.” *IEEE Trans. Electron Devices*, vol. 44, no. 11, pp. 1931-1943, Nov 1997.

[3] *Atlas User's Manual: Device Simulation Software*. Silvaco Inc., Santa Clara, CA, 2016

[4] D. E. Aspnes, and A. Studna, “Dielectric Functions and Optical Parameters of Si, Ge, GaP, GaAs, GaSb, InP, InAs, and InSb from 1.5 to 6.0 eV.” *Phys. Rev. B.*, vol. 27, 10.1103/PhysRevB.27.985.

[5] A. Ingargiola, M. Assanelli, A. Gallivanoni, I. Rech, M. Ghioni, and S. Cova, “Avalanche Buildup and Propagation Effects on Photon-Timing Jitter In Si-SPAD with Non-Uniform Electric Field.” *Proc. SPIE 7320, Advanced Photon Counting Techniques III*, 73200K, April 2009.

[6] A. Lacaita, M. Mastrapasqua, and S. Vanoli, “Observation of Avalanche Propagation by Multiplication Assisted Diffusion in P-N Junctions.” *Appl. Phys. Lett.,* vol. 57, pp. 489–491, 1990.

[7] A. Gulinatti, P. Maccagnani, I. Rech, M. Ghioni, and S. Cova, “35ps Time Resolution at Room Temperature with Large Area Single Photon Avalanche Diodes.” *Electron. Lett.*, vol. 41, pp. 272–274, 2005.

[8] I. Vornicu, R. Carmona-Galán, B. Pérez-Verdú, and Á. Rodríguez-Vázquez, “Compact CMOS Active Quenching/Recharge Circuits for SPAD Arrays.” *Int. J. of Circuit Theory and Applications*, vol. 44, no. 4, pp. 917-928, Apr. 2015.

[9] 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.

CHAPTER 7
REGARDING CROSSTALK IN PHOTODIODES

7.1 INTRODUCTION

Crosstalk (CTK) occurs when charges or photons generated on a pixel are finally sensed by a neighboring pixel, thus degrading the sensor's performance. It also cuts down the spatial resolution by blurring the sharp edges. Other detrimental effects are the reduction of the overall sensitivity, the degradation of the color separation, and increased image noise. Crosstalk is defined as the percentage of the total charge generated by incident light that is diverted to not-illuminated pixels in the neighborhood.

There are two components for CTK. The *optical crosstalk* is due to radiative recombination when an electron drops to its equilibrium energy band and radiates a photon [1], which may reach depletion regions of neighboring photodiodes. The second component is *electrical*, and it involves the diffusion of photo-generated carriers between adjacent devices. This Chapter addresses the crosstalk characterization of two types of photodiodes, namely:

- first, a crosstalk study of 3 kinds of SPADs that suffer from optical crosstalk will be done;
- then, to compare operability, a crosstalk study of an entirely different photodiode, the *pinned photodiode*, will be shown.

We analyze both types of devices to gain a global understanding of the crosstalk impact by using TCAD simulations to assess the impact of crosstalk on the quantum efficiency QE. We will see that the doping concentration gradient limits QE and show that this limitation has a collateral benefit, as it also helps keep the crosstalk noise at a low level.

7.2 DEVICE SELECTION AND TCAD MODEL CONFIGURATION

The devices selected for studying quantum efficiency and crosstalk have been the G12, V12, and A12 models (See Chapter 2 for these device structures). The process simulation was carried out with ATLAS as part of the SILVACO TCAD suite. This simulation includes 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 [2]. Simulations are intended to characterize QE for different wavelengths, for a window of illumination that matches the active area, and to examine as well the physical properties of the bulk in Geiger mode.

7.3 RESTRICTED QUANTUM EFFICIENCY: DETECTION REGION

In the last two decades, investigators have put much effort into refining SPADs to improve their characteristics, for instance, their *QE, DCR,* and *FF*. To that end, SPADs have been reshaped over the years. A main improvement line has been the use of a suitable *guard ring* that can either avoid early edge breakdown, which would significantly limit the active area size of the SPAD, and scale down these devices to maximize *FF*. One of the first solutions to this problem was to implant a low doped well in the active area region's periphery to prevent its periphery from prematurely entering breakdown (P+ SPAD, see Chapter 5). However, the problem with this configuration is that there is a limit for scaling it down, as the depletion regions around these low doped wells would merge, depleting the active area carriers 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 non-viable for most applications [4]. The most modern form of guard ring is that of the *virtual guard ring*. This configuration prevents a Deep-N-Well (DNW) from fully forming 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] (PW SPAD, see Chapter 5). This junction would then have a strong electric field in the active area that will weaken as it gets closer to the surface, gradually increasing the lateral breakdown voltage and forming a de-facto guard ring. This concept also has been used in [6] with the growth of a P-epitaxial layer over a P-substrate to form a guard ring between the P-epitaxial layer and a Deep-N-well. Both solutions offer the possibility of further scaling down the SPADs and enhancing the detection of longer wavelengths, as the junctions are located deeper in bulk. Thus, the gain in quantum efficiency in the last two configurations at those wavelengths should be considerable, given the difference in junction depth; however, this is not the case.

To see why, the QE of the selected devices will be compared. To measure the wavelength profile of QE all devices are illuminated with a uniform beam of a fixed intensity with a variable wavelength. The current collected in the anode (I_A) is then divided by the source photocurrent (SPC), which is the total current that could generate an illumination source if every photon generates an 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 [2]:

$$EQE(\lambda) = \frac{|I_A|}{SPC(\lambda)} \tag{7.1}$$

FIGURE 7.1 displays the EQE for different SPAD configurations. As can be seen, the G12 model presents a maximum EQE at a wavelength of 0.44µm, whereas the V12 model has its maximum at 0.48µm, and from there, the quantum efficiency is only marginally higher than that of the G12 model. At first sight, the primary junction depth of the V12 model, which is around 1µm deep, should have enabled a more differentiable maximum EQE shift over the former, in which the primary junction is only 0.2µm deep. However, this difference is minimal. That shift is effectively observed with the A12 model, which presents a maximum at 0.64µm with its junction at 1.5µm deep.

The reason for these behaviors lies in the electric field profile of the different SPADs when they are ready to absorb a photon and trigger an avalanche. Electron-hole pairs able to originate an avalanche can only be created by illuminating a determined zone of the SPAD, the active

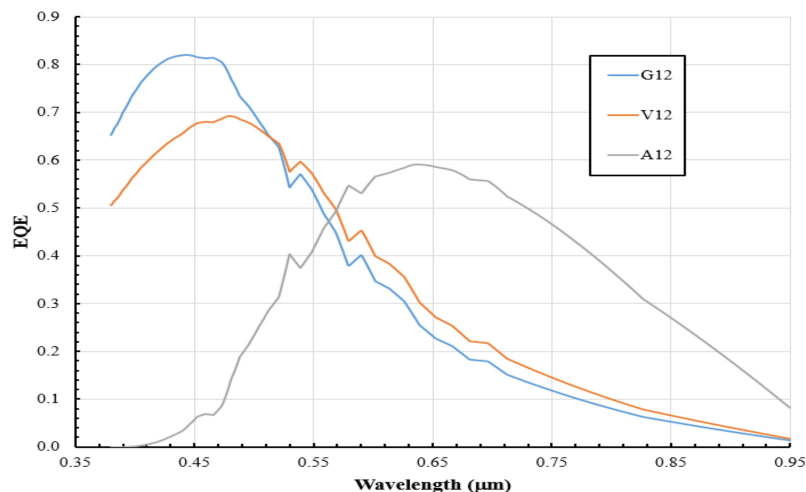


FIGURE 7.1 External quantum efficiency obtained for several SPAD models

area region. In the FIGURE 7.2 the module and direction of the electric field can be seen within a portion of the active area region of the SPAD. The deeper part of the Deep-N-Well presents a weak electric field. That electric field is caused by the Deep-N-Well (in yellow in FIGURE 7.2) presents a gradient of doping concentration instead of a flat uniform doping. Thus, this gradient provokes that any pair of points within the DNW presents a difference in charge that will enable an electric field between them. Instead of random diffusion, this electric field will route charge carriers to or away from the depletion zone. 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**, as is shown in FIGURE 7.2. On the other hand, any carrier created between the surface and this sensing limit may diffuse to the depletion region, which is around the junction and within those limits. These carriers will accelerate until they reach the saturated drift velocity and trigger impact ionization which would result in an avalanche current.

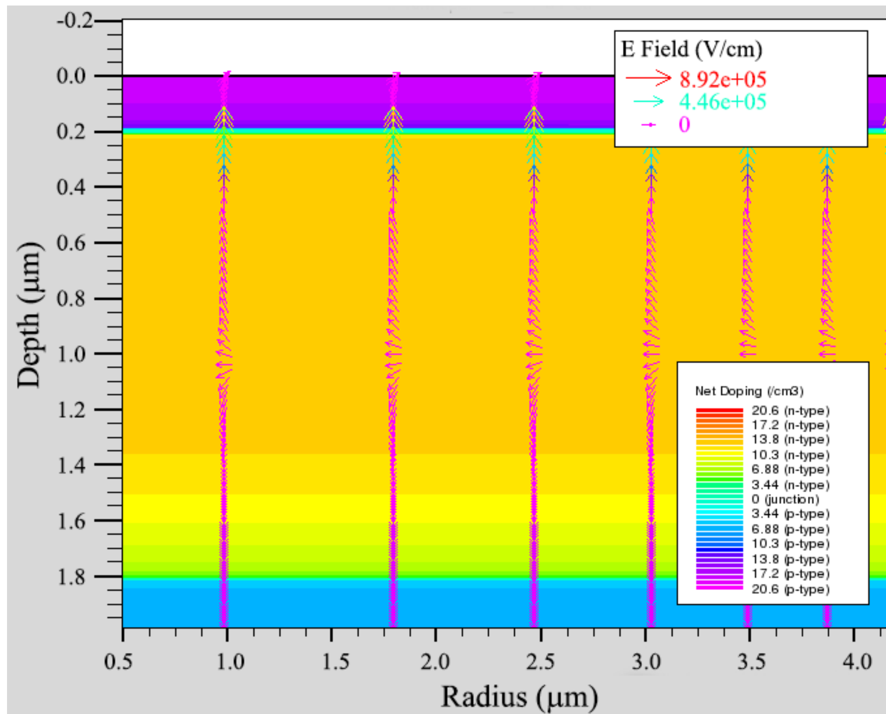


Figure 7.2 Electric field represented in vector arrows in the active area region of the G12 model.

This spatial sensing limit restricts *EQE* since the farther it is from the junction, the greater the volume capable of gathering photo-generated charge carriers. Naturally, it can be used then to define a region within the SPAD that encompasses 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 trigger avalanches and are hence undetectable for the electronics. FIGURE 7.3 shows the detection region for the three selected devices. Through it, we can explain the quantum efficiency from FIGURE 7.1, but first, we will define the current generated for incident photons. The generation rate is given by [2]:

$$G = \eta_0 \frac{P\lambda}{hc} \alpha e^{-\alpha y} \tag{7.2}$$

$$\alpha = \frac{4\pi}{\lambda} k \tag{7.3}$$

where: α is 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;

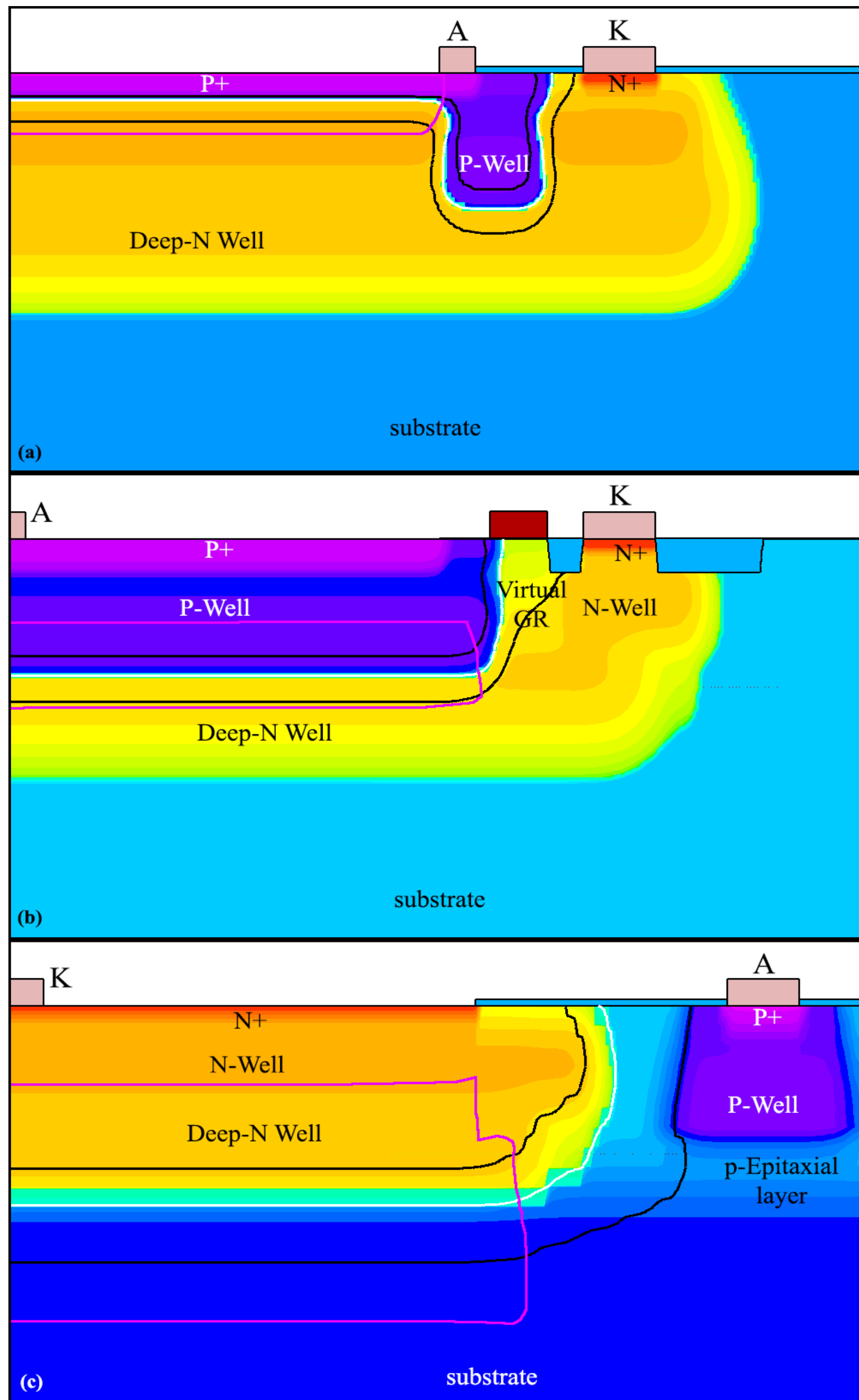


FIGURE 7.3 **Layer configuration of the different SPAD models: (a) G12, (b) V12 and (c) A12 models. The left border acts as an axis of revolution. The depletion region limits are depicted by black lines, whereas a pink line delimits the detection region.**

η_0 is the internal quantum efficiency, which represents the number of carrier pairs generated per photon observed; y is the depth of absorption, and k is the imaginary part of the optical index of refraction or extinction coefficient.

Equation (7.2) explains that a photon's absorption probability decays exponentially with depth, and the longer the wavelength, the slower the decay. Hence, the shortest wavelengths will be absorbed almost entirely near the surface, whereas the longest ones will be more uniformly absorbed through the bulk as the absorption probability decay less with depth.

The G12 model has a detection region ranging from the surface to 1µm deep (FIGURE 7.3(a)), so it is very good at detecting short wavelengths. Then the probability of detecting photons decays as the wavelength of the photons gets larger, as the detection region becomes too small compared to the region of possible absorption. The detection region for the V12 model range from 0.65µm to 1.3µm (FIGURE 7.3(b)). As expected, this configuration performs worse when gathering photons of short wavelengths, but the smaller detection volume makes it worse at detecting too medium wavelengths. Finally, it is slightly better at detecting longer wavelengths because of the deeper limits of the detection region. The last SPAD model has a detection region that ranges from 0.7µm to 2.5µm (FIGURE 7.3(c)). In this last case, the distance the charge carriers have to travel to the depletion region is so high that the probability of recombining again with the bulk is high enough. The deeper boundary for this region, which is well inside the substrate, is defined by this *recombination length*, or the distance a charge carrier travels inside the bulk before recombining again. As the substrate has a uniform doping concentration that starts to vary as it gets closer to the SPAD, the movement of charge carriers generated near the deeper boundary is governed by random diffusion. This high recombination probability also explains why the probability of detecting short wavelengths is so low. Instead, this device is excellent at detecting long-wavelength photons due to its thick 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 A12 model arrangement do these electrons reach the depletion region, triggering avalanches. As described before, in the case of the G12 model and V12 model, this detection volume is within the SPADs themselves. The electrons generated in the DNW and the substrate are 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 makes them almost immune to electrical crosstalk, as charge carriers generated in other SPADs 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 an avalanche if they are absorbed in the detection region. This makes the photodiode based in the V12 model especially well protected due to the low cross-section of its detection region. The A12 model, however, would suffer this undesirable effect due to the larger cross-section of its detection region.

A solution to improve quantum efficiency in this device would be to smooth the doping concentration gradient so that the vertical component of the weak electric field present in the neutral regions would be less likely to change direction. In the V12 model, for example, the P-Well impurities concentration could maintain a downward trend until it reaches the depletion region, which would generally improve quantum efficiency, especially in short wavelengths, as the SPAD would be able to detect photons absorbed near the surface. 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 applies to the other SPAD configurations.

7.4 CROSSTALK IN SPADS

The photon detection probability is defined in terms of the quantum efficiency:

$$PDP = QE(\lambda) \cdot P_{tr} \tag{7.4}$$

Bear in mind that the avalanche triggering probability P_{tr} varies with the device. Hence, to compare different devices on equal basis, this is forced to its maximum value ($P_{tr} = 1$) in all of them, a situation that happens when the excess bias is large enough. As QE only depends on wavelength, QE defines the maximum PDP in these conditions. We will see that this assumption does not significantly alter the final results in terms of crosstalk. As in the previous section, we mean EQE when referring to quantum efficiency.

As the G12 and V12 models have a secondary junction, they are protected from charge carriers generated below this junction, whether from the illumination of the SPAD itself or the illumination of neighboring pixels, also called electrical crosstalk. On the other hand, the A12 model is sensitive to this kind of spurious signal, and it will be studied in the next section. All of them suffer from optical crosstalk.

7.4.1 ELECTRICAL CROSSTALK IN SPADS

The electrical crosstalk in the A12 model has two sources, described in the bullet points below.

- The first one is due to the charge carriers generated by the illumination. When a SPAD is illuminated, some of the generated electrons are collected in the cathodes of neighboring pixels, thus triggering spurious avalanches. FIGURE 7.4 shows the crosstalk of up to three neighboring pixels for Deep-Ntubs separated by 4.5μm and 9μm, respectively. This figure shows how crosstalk increases with wavelength. The longer the wavelength, the deeper the pairs created within the substrate, and the weaker the electric fields, which allows the electrons to diffuse to adjacent pixels easier.

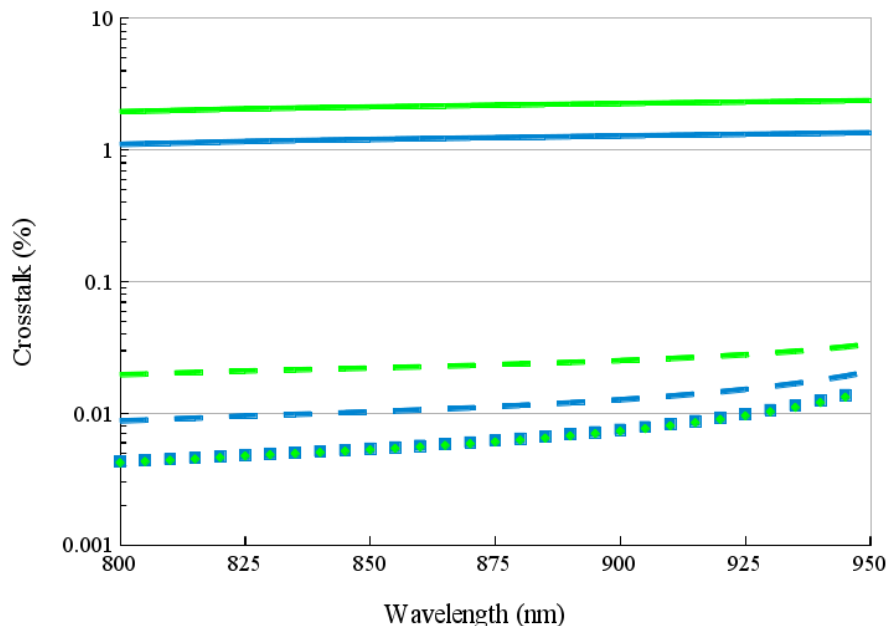


FIGURE 7.4 Crosstalk due to illumination for Deep-Ntubs separated 4.5 μm (green) and 9 μm (blue), where solid lines refer to first neighbors, stripped lines to second neighbors and dots to third neighbors.

- The second source of electrical crosstalk is due to the charge carriers generated by an avalanche that can diffuse to neighboring pixels. Although present, this source is negligible when compared to the previous one.

7.4.2 OPTICAL CROSSTALK IN SPADS

As already explained, the optical crosstalk is due to radiative recombination, which involves the emission of photons when electrons drop to their equilibrium energy band. In SPADs, this photon emission reaches its peak during an avalanche, affects all configurations, and presents an emission spectrum that can only be obtained experimentally. In this work, the emission spectrum used was that of the work of Rech (FIGURE 7.5) [7]. This spectrum corresponds to the emission of an avalanche in Silicon, so all Si-based SPADs will present this continuum of wavelengths. As can be seen, this spectrum presents a maximum of 960nm. An omnidirectional illumination of this wavelength will be absorbed by Silicon almost entirely within a radius of around 30μm. To quantify the optical crosstalk, first, the emission power of a SPAD during an avalanche is measured through TCAD simulations. An example of this can be seen in FIGURE 7.6, where the maximum radiative recombination rate corresponds to the depletion region. This recombination rate follows the next equation:

$$RR = B \cdot (np - n_i^2) \qquad (7.5)$$

where $B = 4.73 \times 10^{-15} \mathrm{cm^3 s^{-1}}$ is the *radiative recombination coefficient* at 300K for crystalline Si [8], and the rest represents the deviation of np, that is the product of charge carrier concentration, from the Si intrinsic concentration at thermal equilibrium, due to an avalanche.

Then, the **luminous power** is obtained by integrating the radiative recombination over the entire device. Therefore, the power of this emission would be linearly dependent on the avalanche current, which depends upon the excess bias voltage. In this work, the optical crosstalk was also measured as external quantum efficiency from the adjacent SPADs for this

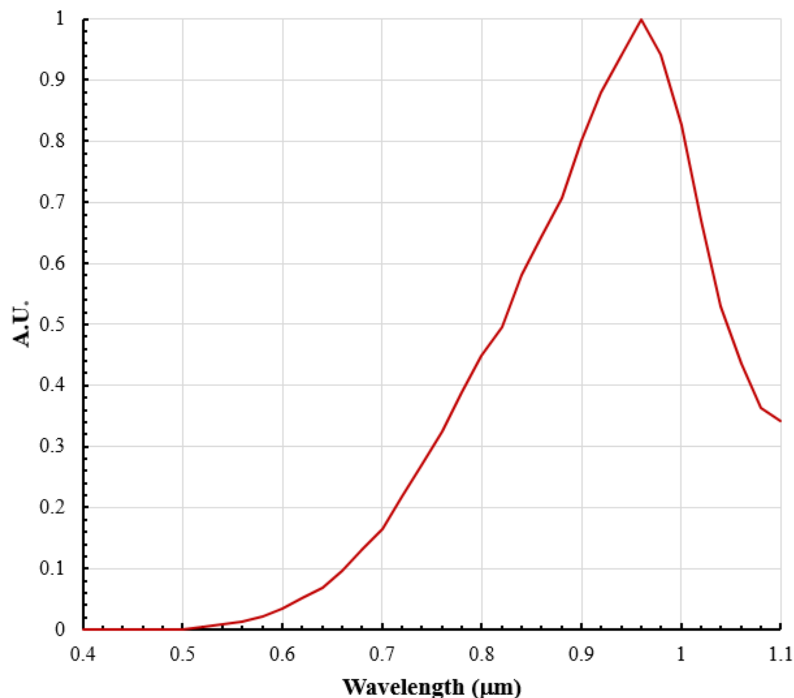


FIGURE 7.5 Emission spectrum due to the emission of an avalanche in Si [7].

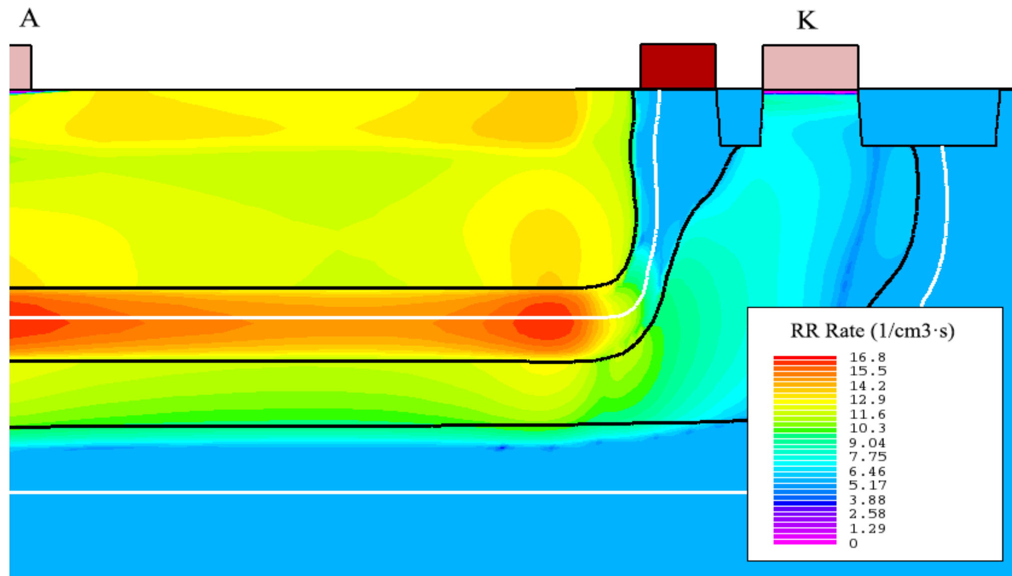


FIGURE 7.6 Radiative Recombination rate in a V12 model during an avalanche at 1V of excess bias voltage. White lines represent the junctions, and black lines the limits of the depletion regions. Scale is logarithmic.

kind of illumination. As quantum efficiency does not depend on the source's intensity, measures are taken at a fixed excess bias of 1V.

Once the emission power of every SPAD configuration is obtained, an avalanche triggering SPAD was considered a point source of illumination with the emission spectrum mentioned earlier to simulate the illumination on SPADs separated at a certain distance from the source. This quantum efficiency is considered the maximum photon detection probability ($P_{tr} = 1$) like normal illumination. This simplification can raise concerns because of the long wavelengths of the spectrum emission. Indeed, FIGURE 7.1 shows that in the near-infrared region of the QE graph, the G12 and V12 models' quantum efficiency are very similar. Like that of the avalanche triggering probability (FIGURE 7.7), a simple difference may do crosstalk results variate one respect to the other greatly. However, this difference is not enough to be significant, as shown in FIGURE 7.8. In this figure, it can be seen how the difference between both devices

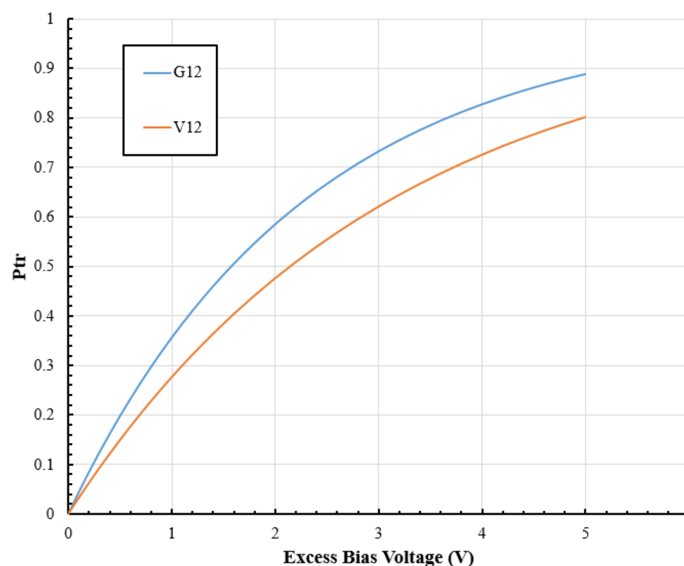


FIGURE 7.7 Avalanche triggering probability vs excess bias for the P+/DNW device and the PW/DNW device extracted from TCAD simulations. Different doping concentration at the junction results in a variation of the electric field affecting the avalanche triggering probability.

is kept at several excess bias voltages, being the V12 model quantum efficiency always greater than that of the G12 model.

FIGURE 7.9 shows this contribution to crosstalk for the SPADs model considered in logarithmic scale for a detector located at a certain distance from the emitter. The A12 model presents a very high quantum efficiency for this kind of emission, which decays fast with distance following roughly a Gaussian profile. Although the avalanche current is greater than that of the V12 model, the high crosstalk from this source in these devices does not depend upon the intensity of the signal from the avalanching diode. This is due to the capacity of these SPADs to gather charge carriers from the substrate into the depletion region due to a higher

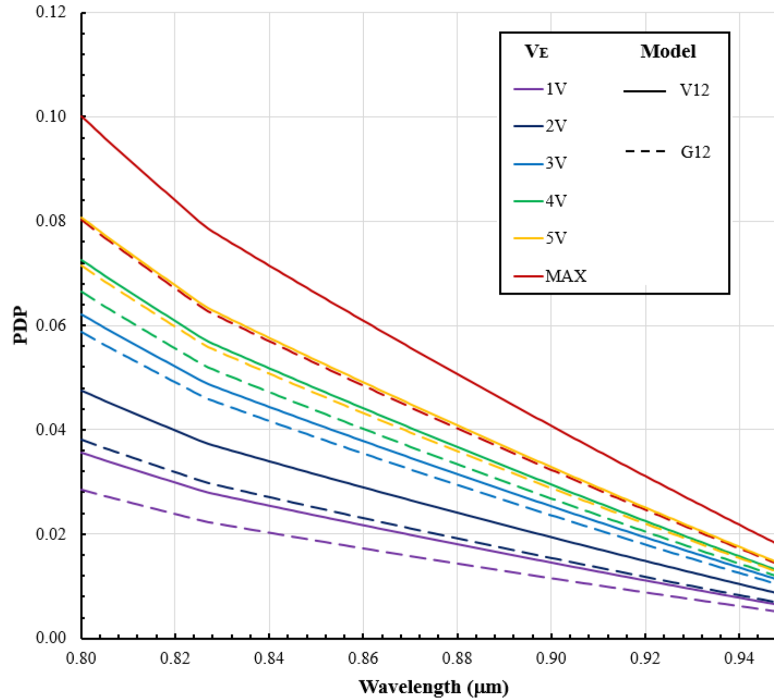


FIGURE 7.8 Quantum efficiency in the near infrared for several excess bias voltages for the PW/DNW and P+/DNW devices. Photon detection probability or maximum quantum efficiency is shown in red.

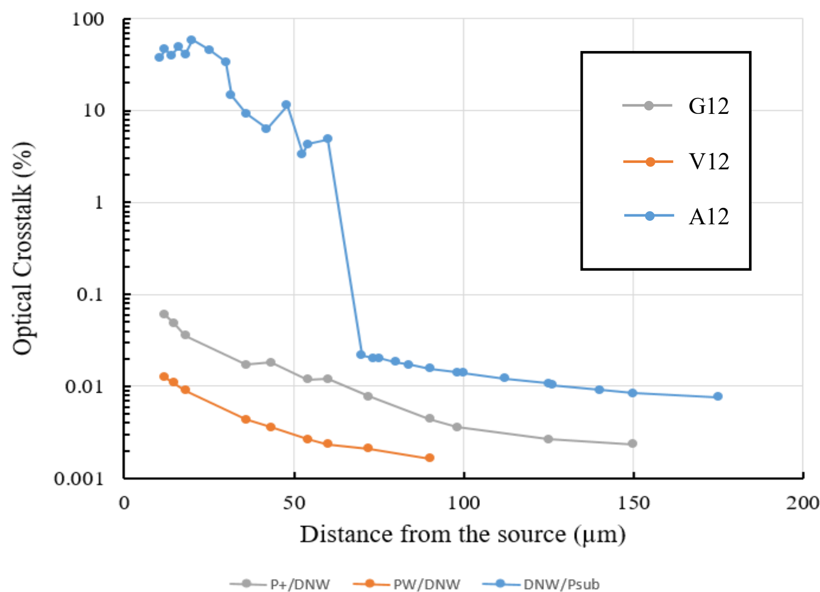


FIGURE 7.9 Optical crosstalk vs distance from the emitter SPAD for several devices.

volume of detection. There have been attempts to minimize the optical crosstalk in SPADs in the literature, and all of them proposed the location of an optical barrier in the direct path of the emission. In [7], a high doping implant is used between devices ($10^{20}\mathrm{cm}^{-3}$) to act as an optical barrier. Calandri [9] proposes using a metallic trench in an InGaAs/InP SPAD array. Either way, they both report the detection of optical crosstalk due to reflections on the device's backside. Despite the high sensitivity in the NIR spectrum of the A12 model [6], the crosstalk impedes the integration of these SPADs in array format. If done, it would be at the expense of the fill factor, which would greatly diminish the photon detection efficiency or implants not present in standards technologies.

7.5 CROSSTALK IN PINNED PHOTODIODES

This section summarizes a study of electrical crosstalk for the Pinned PhotoDiode (PPD) [10]. To that end, the TCAD simulation consisted of the creation of a partial 4-transistor pixel (4T-APS). Particularly, the study explores the CTK relation with quantum efficiency and the impact of the thickness, and hence the depth, of the epitaxial layer – a key parameter of the structure that affects both CTK and QE [11]. This exploration is relevant for SPADs because:

- PPDs are easier to simulate than SPADs, so several of them can be simultaneously simulated without requiring long computer calculation times. Making a single TCAD simulation of a SPAD can take hours because the simulator has to converge solutions for the breakdown point when ramping the SPAD bias voltage to the desired. The TCAD simulation of a PPD just takes minutes.
- Electrical crosstalk is larger for PPDs than for SPADs, and the PPD currents can be visualized easily. This particular PPD feature illustrates in a very clear way the important role that the second junction plays in SPADs and why it cannot be dispensed.
- The visualization of crosstalk provides tips about how charge carriers drift through the epitaxial layer in a similar fashion to SPADs.
- PPDs do not suffer from optical crosstalk due to radiative recombination, so they do not need to be protected against this source.

For simplicity, all simulations employ a 2D description of the structure – see FIGURE 7.10. Also, we consider an array of five pixels (FIGURE 7.11), each including a PPD with its Transfer Gate (TG), a Floating Diffusion (FD), and a ReSet Transistor (RST) to restore the floating diffusion voltage to a reset value. The PPD includes the usual layers:

- The buried N-Well and the P+ pinning implantations.
- The P-Well isolates the circuitry from the bulk.
- N+ boxes used for floating diffusion and the drain of the reset transistor.
- A TG threshold adjustment layer.

The device fabrication process is simulated with ATHENA, SSUPREM4, and ELITE tools, belonging to the set of SILVACO TCAD tools [12]. We rely on the dose information provided in [10] to specify the *doping profile*s. Custom doping profiles for the P-Well or PW and the p-substrate compatible with a standard 180nm CMOS technology were used.

The resulting device has about 75,000 nodes [2], and a single simulation run with ATLAS takes about 85 hours of simulation time with a 4-core processor, which gives an idea of its complexity.

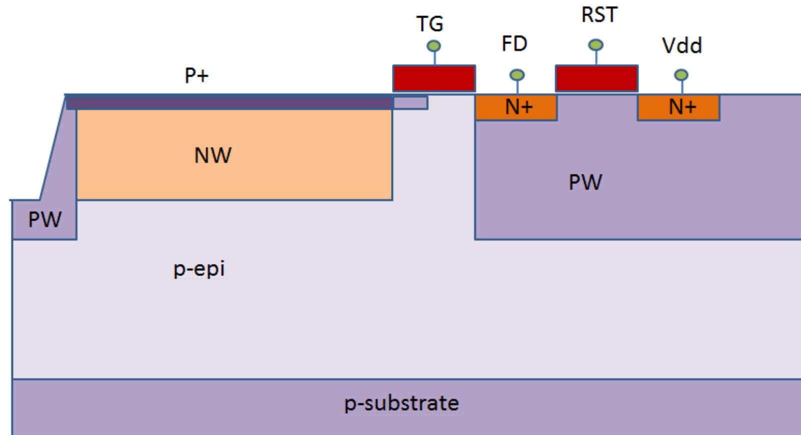


FIGURE 7.10 Pinned photodiode basic structure.

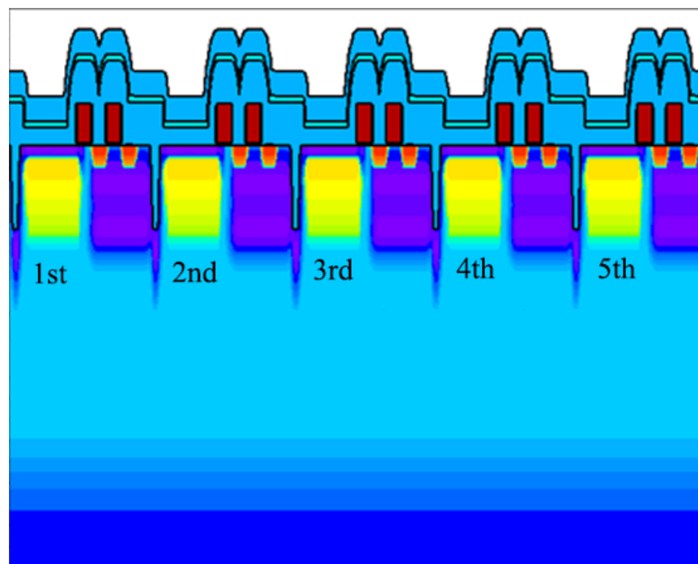


FIGURE 7.11 Structure of a five 4T-APS pixel array visualized from a doping profile image.

7.5.1 SIMULATED TESTS

Simulations of the device process employ ATLAS (as part of SILVACO TCAD tools) and the following models [2]:

- ***Consrh***: a recombination model which uses concentration-dependent carrier lifetimes using the Shockley-Read-Hall theory of recombination through defects, similarly to SPADs (Section 3.4.2, Chapter 3).
- ***Cvt*: lombardi** (CVT) model. A complete mobility model that takes into account doping, temperature, and electric field effects. The mobility model, *fldmob*, used in Chapter 2, was required to emulate the saturation of drift velocity for SPADs, which is not needed for PPDs.

  Fermi: Fermi-Dirac statistical model. SPADs complicated calculations required a simpler model so the solution could converge, but that is no longer necessary for PPDs, which can make use of a more complex and precise carrier statistics model.

Crosstalk characterization employs a spot-scanning technique. It consists of:

- first, illuminating an individual pixel inside an array and;

- then measuring its effects on adjacent pixels.

The crosstalk in PPDs is defined as:

$$CTK(\%) = \frac{I_0}{I_T} 100 \quad \textbf{(7.6)}$$

where I_0 is the total current induced by stray carriers in the adjacent pixels when the central pixel is illuminated, and I_T is the total photo-generated current. Both are measured as the slopes of the generated charges against a fixed transient time [10].

Regarding the **optical transmission**, we consider that the whole array is masked, except the central pixel active area, as we want to evaluate the electrical component of the crosstalk separately. Hence, the optical transmission is zero. The illumination intensity of the central pixel in FIGURE 7.11 is $1.4 \cdot 10^{-2}\mathrm{W/cm^2}$ so that the photodiode does not saturate during a fixed integration time, considering an illumination window of $1.45\mu\mathrm{m}$ that match that of the active area. We have measured the charge stored in the buried NW of the five pixels at a fixed luminous intensity and exposition time when they are reverse biased so that an evaluation can be made of the photocurrent available on the non-illuminated pixels against the central one.

7.5.2 QUANTUM EFFICIENCY AND EPITAXIAL LAYER THICKNESS

A key parameter to consider is the thickness of the p-epitaxial layer (ELT). As stated in [11], there is a quantifiable *QE* degradation when the p-epitaxial layer thickness decreases. This is because minority carriers generated in this layer are efficiently collected in the buried NW of the pixel due to the high electric field established between the substrate and the well when the pixel is reverse-biased. Also, the low concentration of impurities reduces the recombination rate of photo-generated carriers, thus increasing the diffusion length. When this layer is thinner, the substrate is closer to the surface. Minority carriers in the substrate are more likely to recombine due to a high impurity concentration, not reaching the high-field area. Therefore, *QE* degrades at longer wavelengths because these photons are absorbed in greater depths. This can be seen in FIGURE 7.12. Shorter wavelengths, like 410 and 474nm in the graph, do not seem to suffer from this effect. Therefore, the longer the wavelength, the larger the loss of *QE*.

7.5.3 CROSSTALK AND EPITAXIAL LAYER THICKNESS

ELT also affects CTK. To understand this effect, we have to study the electric fields within the bulk to be able to know how carrier motion is established. TCAD simulations allow us to measure both the module and the direction of the electric field. Also, CTK can be visualized as electron current density between the second and third pixels.

FIGURE 7.13 shows three different regions between the depletion regions of two adjacent pixels that can be clearly identified when the epitaxial layer is thinner. The first one is at the edge of the substrate, and it will be labeled as **border region** (B). The second is characterized by opposing electric fields that create a potential valley where minority carriers can only diffuse between depletion regions and thus will be labeled as **valley region** (V). Finally, there is a strip of a relatively strong electric field between V and the PW, which will be the **buffer region** (BF). In B, the electric field is not very strong but is high enough to make minority carriers drift into V. As in B, the BF electric field conducts minority carriers towards the substrate and into V. The opposing electric fields allow the carriers to move freely in this region. Some of these carriers can get through and reach the second-pixel depletion region. Then, the minority carriers are collected in the buried NW, and thus the CTK occurs. As in the previous case, minority carriers that travel through V and reach the 4th-pixel depletion region will be collected and measured as CTK. On this occasion, however, there is a small potential barrier at the end of the

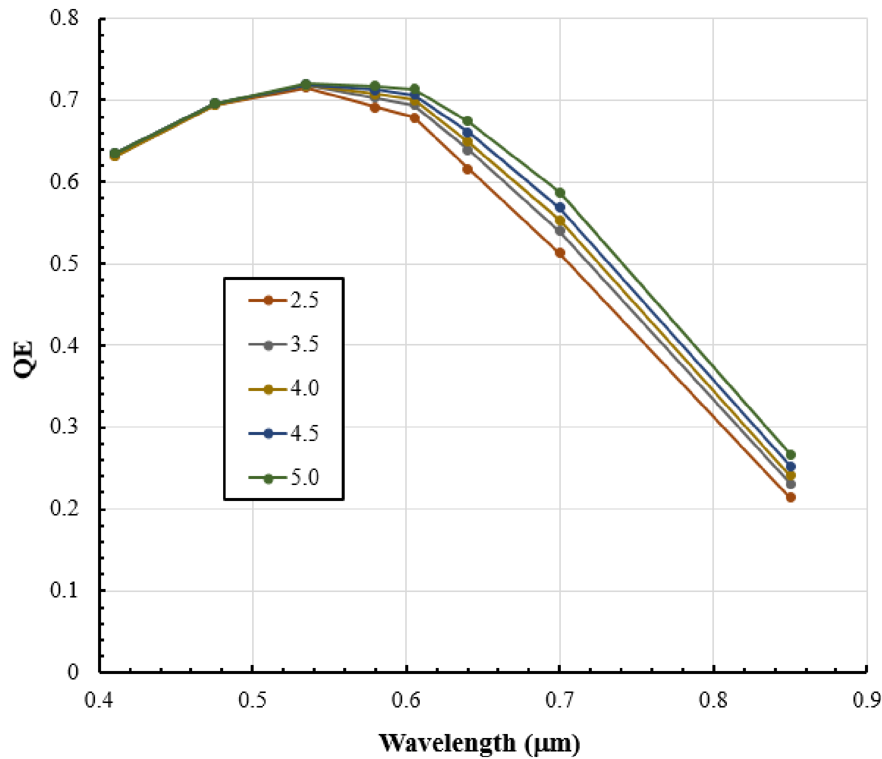


FIGURE 7.12 Quantum efficiency vs wavelength for several ELT in µm.

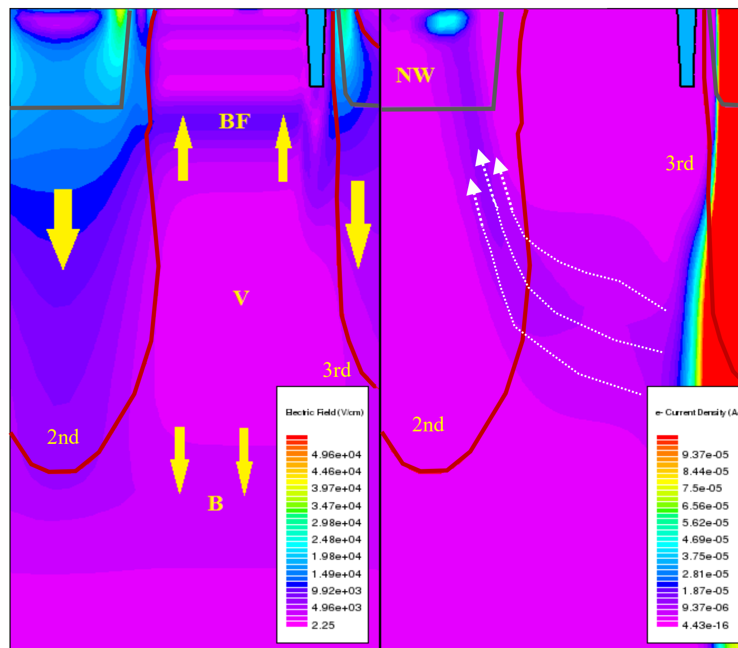


FIGURE 7.13 Crosstalk between 2nd and 3rd pixels with ELT at 1.5µm. Electric field is represented in the left graph. General electric field vector arrows are depicted. Electron current density is shown in the right graph. The red lines represent the depletion region edge and the grey ones the junctions.

V region due to the trench, preventing many minority carriers in free diffusion from reaching the depletion region, and thus CTK is considerably lower. Notice that 1st and 5th pixels registered meaningful CTK values only at shorts ELTs and longer wavelengths in our simulations, but they were under 0.5% even then.

On the one hand, when ELT decreases, the V region becomes thinner, therefore reducing the volume where minority carriers may diffuse, thus effectively increasing current density in this

zone. This caused minority carrier velocity and diffusion length to rise in V. Now carriers, which would have been recombined before, can reach the adjacent pixels leading to CTK increase, as FIGURE 7.14 shows.

On the other hand, when ELT increases from a certain point on, B regions between pixels disappear. This happens when ELT is greater than the depletion region edge's reach, allowing the photo-generated carriers to diffuse freely across the p-epitaxial layer into any of the pixels. If the rise of photo-generated carriers from the *QE* boost is also considered, it is easy to see the CTK increase with ELT. This effect can be seen in FIGURE 7.15. Total CTK, defined as the arithmetic sum of all adjacent pixels CTK against wavelength for several ELTs, is shown in

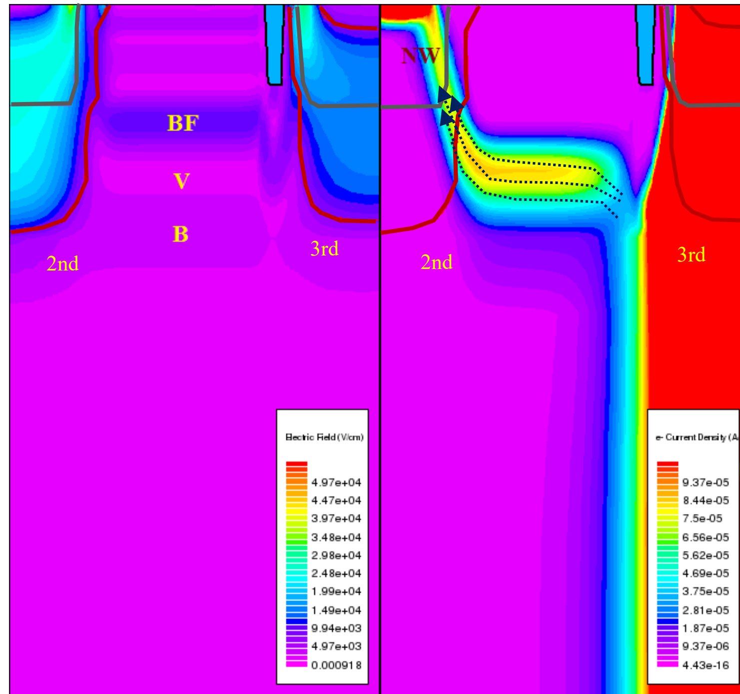


FIGURE 7.14 Crosstalk between 2nd and 3rd pixels with ELT at 0.5 μm.

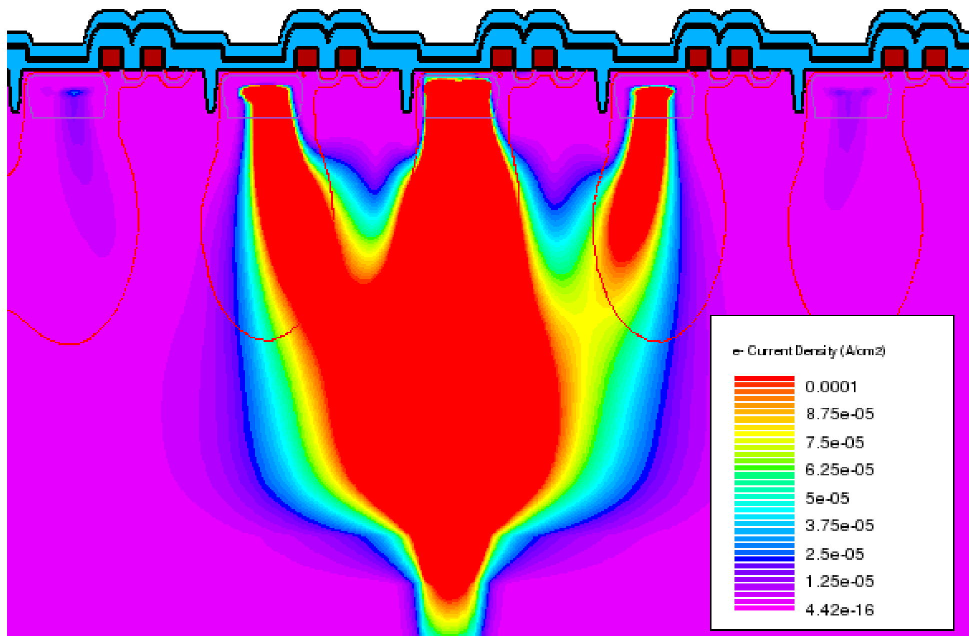


FIGURE 7.15 Crosstalk representation as electron current density with ELT at 5 μm. In this graph, maximum electron current density is cut by 2 orders of magnitude so crosstalk in 1st and 5th pixels can be visualized.

FIGURE 7.16. As can be seen, CTK increases with wavelength, as shown in other works [13][14]. Nevertheless, the total CTK clearly expresses the dependence of the ELT.

The CTK values obtained are at their lowest in a range of 2.5 and 3µm of ELT when they reach 4% for a wavelength of 850nm. These values are low and in line with that of current literature [13]. However, estimated CTK values can exceed 10% if ELT is too large or even 15% if too reduced.

Therefore, with these results, we can draw conclusions about the differences in how crosstalk affects PPDs and SPADs.

- P+ SPADs and PW SPADs are protected from electrical crosstalk by their second junction, which acts as a barrier.
- Both PPDs and SPAD based on the A12 model present electrical crosstalk that increases with wavelength, and the reasons for both of them are the same: longer wavelength photons create more carriers at the depth where the charges carriers responsible for the electrical crosstalk diffuse.
- Although the A12 model depletion region is not protected by a secondary junction, it is protected laterally by the virtual guard ring. This guard ring prevents charge carriers from reaching a sensitive area, whereas the PPDs do not have any of those protections, which explains the greater crosstalk effect.

These conclusions give a reasonable explanation of the crosstalk results in both SPADs and PPDs and can help researchers in future SPADs and pixel designs to minimize the impact of this source of dark counts.

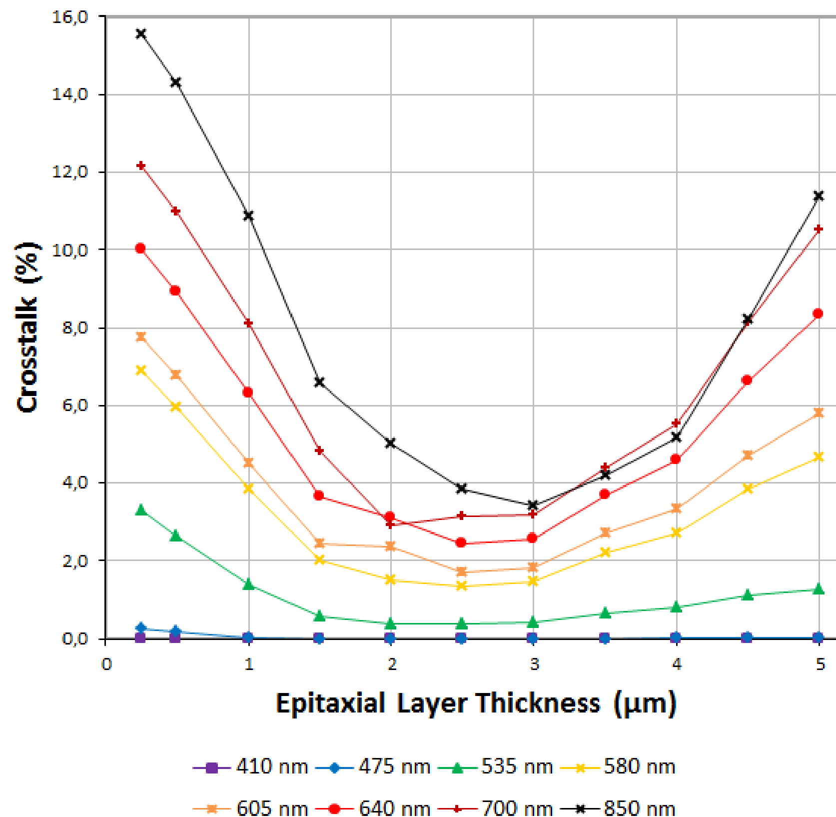


FIGURE 7.16 Total crosstalk vs. ELT for several wavelengths.

7.6 REFERENCES

[1] J. Bude, N. Sano, and A. Yoshii, "Hot-Carrier Luminescence in Si." *Physical Review. B, Condensed Matter*, vol. 45, pp. 5848–5856, Apr. 1992.

[2] *Atlas User's Manual: Device Simulation Software*. SILVACO INC., Santa Clara, CA, 2016

[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." *IEEE Electron Device Letters,* vol. 33, no. 11, pp. 1589-1591, Nov. 2012.

[7] I. Rech, A. Ingiargola, R. Spinelli, I. Labanca, S. Marangoni, M. Ghioni, and S. Cova, "Optical Crosstalk in Single Photon Avalanche Diode Arrays: A New Complete Model." *Opt. Express,* vol. 16, pp. 8381–8394, Jul. 2008.

[8] T. Trupke and M. A. Green, "Temperature Dependence of the Radiative Recombination Coefficient of Intrinsic Crystalline Silicon." *J. Appl. Phys.*, vol. 94, no. 8, Oct. 2003.

[9] N. Calandri, M. Sanzaro, L. Motta, C. Savoia, and A. Tosi, "Optical Crosstalk in InGaAs/InP SPAD Array: Analysis and Reduction With FIB-Etched Trenches." *IEEE Photonics Technology Letters*, vol. 28, no. 16, pp. 1767-1770, Aug. 2016.

[10] E. R. Fossum and D. B. Hondongwa, "A Review of the Pinned Photodiode for CCD and CMOS Image Sensors." *IEEE J. Electron Devices*, vol. 2, no. 3, pp. 33–43, May 2014.

[11] G. Agranov, R. Mauritzson, S. Barna, J. Jiang, A.Dokoutchaev, X. Fan, and X. Li, "Super Small, Sub 2µm Pixels for Novel CMOS Image Sensors." *Proc. Int. Image Sensor Workshop*, pp. 307–310, 2007.

[12] *Athena User's Manual*. Silvaco Inc., Santa Clara, CA, 2015.

[13] M. Estribeau and P. Magnan, "Pixel Crosstalk and Correlation with Modulation Transfer Function of CMOS Image Sensor." *Proc. SPIE* vol. 5677, pp. 98-108, Mar. 2005.

[14] B. C. Burkey, W. C. Chang, J. Littlehale, T. H. Lee, T. J. Tredwell, J. P. Lavine, and E. A. Trabka, "The Pinned Photodiode for an Interline-Transfer CCD Image Sensor." *IEDM Tech. Dig.*, vol. 30, pp. 28–31, Dec. 1984.

CONCLUSIONS

This Thesis addresses the construction of physically-consistent VERILOG-A models for mixed-signal electrical simulation of SPAD-based sensors. The methodology consists of a combination of analytical modeling of SPAD behaviors and empirical fitting based on TCAD simulations of actual SPAD structures. Compared to previous VERILOG-A SPAD models, this We consider avalanches in primary, nominal sensitive junctions, and secondary junctions like those associated with guard rings. Besides this, proposed models contemplate features that were not incorporated to previous ones, namely:

- the TAT contribution to the dark count rate is not included;
- there is no validation of the SRH theory of recombination through defects based on an actual existing defect;
- the coverage of temperature dependencies is incomplete;
- the model does not include the possible effects on the results of other internal SPAD structures like a guard ring or a second junction.

Thesis conclusions are summarized below:

1. Using TCAD simulations of actual SPAD structures to fit analytic models allows for surmounting the difficulties of directly mapping TCAD results onto circuit model structures and providing model outcomes that fit experimental qualitative and quantitative experimental SPAD behavioral data.
2. Several basic SPAD structures are identified out from the state-o-the-art and selected for analysis and modeling based on their intrinsic value and their application scope.
3. A modeling workflow has been defined, starting with fabrication process simulation of SPAD structures and ending with VERILOG-A models. This has involved procedures for translating doping profiles into data required for TCAD tools. The procedure has been applied to a 180nm CMOS technology.
4. Besides analyzing different state-of-the-art SPADs structures, the methodology enabled the proposal of a new device structure, a SPAD with three junctions (two sensitive to incoming photons), and a P-gated guard ring to avoid edge breakdown described in Appendix II.
5. Previous VERILOG-A models, basically based on the analytical descriptions made by Dalla Mora[1] have been improved by including TAT and making the model rely on a solid validation of the SRH model for the carrier generation based on the E-center defect.
6. Self-heating has been included to the model thus allowing, by the first time according to our knowledge, to change dynamically model parameters and behaviors as a function of temperature.
7. VERILOG-A models have been devised for the family of G SPAD structures of different sizes, based in the P+/Deep-N-Well SPAD (often abbreviated P+ SPAD in this Thesis), and for the V12 model (based on the P-Well/Deep-N-Well or PW SPAD). Other structures, such as the A12 and P12, are addressed only at the TCAD level. Comparison to experimental results show deviations below 7%, notwithstanding that doping profile information was incomplete due to foundry confidentiality issues.

---

[1] A. Dalla Mora, A. Tosi, S. Tisa, and F. Zappa, "Single-Photon Avalanche Diode Model for Circuit Simulations." *IEEE Photonics Technology Letters*, vol. 19, no. 23, pp. 1922-1924, Dec.1, 2007.

8. TCAD simulations of the G12 structure showed that the avalanche photon triggering probability could be described through the wavelength of the incident photons, thus improving the classic model used to describe that behavior.
9. Improved descriptions of the response time and the photon-timing have been made and checked with the G12 structure to show that the SPAD photon-timing jitter is limited by the components of the SPAD quenching circuit through the threshold current or minimum detectable current.
10. Crosstalk has been analyzed with the following outcomes:
    - P+ SPADs and PW SPADs are protected from electrical crosstalk by their second junction that acts as a barrier, whereas the DNW SPAD is affected a little by electrical crosstalk and heavily by optical crosstalk due to radiative recombination.
    - Both PPDs and DNW SPAD present electrical crosstalk that increases with wavelength. The reasons for both are the same: longer wavelength photons create more carriers at the depth where the charge carriers responsible for the electrical crosstalk diffuse into the depletion region of the A12 model.
    - Although a secondary junction does not protect the DNW SPAD depletion region, it is protected laterally by the virtual guard ring. This guard ring prevents charge carriers from reaching a sensitive area, whereas the PPDs do not have any protections, which explains the greater crosstalk effect.

APPENDIX I
BASIC G12 MODEL CODE

1. MODEL SETUP

```
1   `include     "constants.vams"
2   `include     "disciplines.vams"
3
4   module spad_gr_12(a,k,photon);
5   inout        a,k,photon;
6   electrical   a, k, photon, gnd;
7   ground       gnd;
8   branch (k,a)        SPAD, JUNC_1;
9   branch (k,gnd)      JUNC_2, C_SUB;
10  branch (a,gnd)      A_SUB;
```

2.CONSTANT DECLARATION

```
11  parameter real   V_b0=10.331;
12  parameter real   PhThereshold=1;
13  parameter real   T_0=253.0;
14  parameter real   r=6e-4;
15  parameter real   eta_T_1=0.415;
16  parameter real   eta_T_2=4.36;
17  parameter real   mIlat=1e-4;
18  parameter real   sigIlat=5e-6;
19  parameter real   V_n=0.05;
20  parameter real   Rbkr=72.555;
21  parameter real   N_1=2.0845e17;
22  parameter real   N_2=2.3e17;
23  parameter real   N_3=6.0255e17;
24  parameter real   P_1=4.8084e19;
25  parameter real   P_2=1.0789e18;
26  parameter real   P_3=6.0e14;
27  parameter real   Cas=8.64e-14;
28  parameter real   Cks=2.83e-14;
29  parameter real   mj=1.0/2.0;
30  parameter real   Cj0=7.9827e-8;
31  parameter real   Cj0_g=3.681e-8;
32  parameter real   Cj0_dw=6.051e-9;
33  parameter real   A_tot=1024e-8;
34  parameter real   A_H=3.29e15;
35  parameter real   B_H=2.38e7;
36  parameter real   p=2.5;
37  parameter real   TD=3.828e17;
```

```
38  parameter real   SPADRes=105.46;
39  parameter real   Eg0=1.166;
40  parameter real   m_0=9.1093897e-31;
41  parameter real   m_lh=0.16;
42  parameter real   m_hh=0.49;
43  parameter real   m_l=0.98;
44  parameter real   m_t=0.19;
45  parameter real   Mc=6.0;
46  parameter real   lambda = 447;
47  parameter real   Et0=0.45;
48  parameter real   sig_00=1.1e-13;
49  parameter real   Ea0_1=4.0535e-20;
50  parameter real   Ea0_2=4.2382e-20;
51  parameter real   Ea0_3=1.5585e-20;
52  parameter real   t0_1=8.7417e-12;
53  parameter real   t0_2=9.9472e-13;
54  parameter real   t0_3=5.296e-10;
55  parameter real   sigmat_1 = 5e-16;
56  parameter real   sigmat_2 = 5e-15;
57  parameter real   sigmat_3 = 5e-15;
```

3. VARIABLE DECLARATION

```
58  real Nt;
59  real T;
60  real B;
61  real V_b;
62  real delta_V_b;
63  real V_e;
64  real cgr_th;
65  real cgr_btbt;
66  real cgr_tot;
67  real tau_cgth;
68  real tau_cgbtbt;
69  real deltatcg_th;
70  real deltatcg_btbt;
71  real tcg_th;
72  real tcg_btbt;
73  real tau_cr1;
74  real tau_cr2;
75  real tau_cr3;
76  real tcr1;
77  real tcr2;
78  real tcr3;
79  real deltatcr1;
```

```
80   real deltatcr2;
81   real deltatcr3;
82   real Ptr;
83   real Ispad;
84   real Ibkr;
85   real Is;
86   real Is_2;
87   real Ilat;
88   real Qj_1;
89   real Qj_2;
90   real Qas;
91   real Qks;
92   real ni;
93   real vthh;
94   real vthe;
95   real tau_0n;
96   real tau_0p;
97   real Is_a;
98   real Is_b;
99   real Is_2a;
100  real Is_2b;
101  real c_1;
102  real c_2;
103  real c_3;
104  real A;
105  real A_g;
106  real A_dw;
107  real V_d;
108  real V_dg;
109  real V_ddw;
110  real FF;
111  real QE;
112  real QE_table[1:8];
113  real l_table[1:8]={410,475,535,580,605,640,700,850};
114  real r1;
115  real pde;
116  real delta_T;
117  real E_field;
118  real E_field_0;
119  real W;
120  real phi;
121  real phi_g;
122  real phi_dw;
123  real n;
124  real delta_n;
```

```
125     real n0;
126     real p0;
127     real n1;
128     real p1;
129     real Nv;
130     real Nc;
131     real Eg;
132     real BGN;
133     real m_h;
134     real m_e;
135     real m_de;
136     real m_dh;
137     real alpha;
138     real Eb_p;
139     real Eb_n;
140     real gA;
141     real gB;
142     real gamma_n;
143     real gamma_p;
144     real g_1;
145     real g_2;
146     real g_3;
147     real g_4;
148     real tau_eff;
149     real Pa_1;
150     real Pa_2;
151     real Pa_3;
152     real theta_1;
153     real theta_2;
154     real theta_3;
155     real tn_notrap;
156     real tp_notrap;
157     real st1;
158     real st2;
159     real st3;
160     real te_nt;
161     real cgr_nt;
162     real av_i_time;
163     real av_time;
164     real b_av_time;
165     real n_aval;
166     real pc_1;
167     real pc_2;
168     real pc_3;
169     real sig_0;
```

```
170    real dcr;
171    real dcr_th;
172    real dcr_btbt;
173    real power;
174    real th_cond;
175    real th_res;
176    real heat;
177    real Ij1;
178    real Ij2;
179    real Ias;
180    real ncrit;
181    real i_1;
182    real Ntrap;
183    real delta_Eg;
184    real Et;
185    real Ea_1;
186    real Ea_2;
187    real Ea_3;
188    real Iks;
189    real pow_i;
190    real aux1;
191    real pow_r;
192    integer aval;
193    integer sd_cgth;
194    integer sd_cgbtbt;
195    integer sd_cr1;
196    integer sd_cr2;
197    integer sd_cr3;
198    integer sd_tr;
199    integer sd_pa1;
200    integer sd_pa2;
201    integer sd_pa3;
202    integer sd_ct1;
203    integer sd_ct2;
204    integer sd_ct3;
205    integer kcg_th;
206    integer kcg_btbt;
207    integer kdc_th;
208    integer kdc_btbt;
209    integer kdc;
210    integer kdet;
211    integer kndet;
212    integer kap;
213    integer kap_1;
214    integer kap_2;
```

```
215     integer kap_3;
216     integer sdIlat;
217     integer i;
218     integer sd_det;
219     integer reset;
220     integer SH_enable;
221     integer TRAP_TEMP_enable;
222     integer TAT_enable;
```

4.DEVICE PARAMETERS SETTING

```
223     analog begin
224
225     //Setting recombination center density
226     //according to model features
227     if (TAT_enable==1) begin
228         if (SH_enable==0) begin
229             if (TRAP_TEMP_enable==0) begin
230                 Nt=6.07e11;
231             end
232             else begin
233                 Nt=227.2e10;
234             end
235         end
236         else begin
237             if (TRAP_TEMP_enable==0) begin
238                 Nt=137.8e10;
239             end
240             else begin
241                 Nt=73.9e10;
242             end
243         end
244     end
245     else begin
246         if (SH_enable==0) begin
247             if (TRAP_TEMP_enable==0) begin
248                 Nt=1.184e13;
249             end
250             else begin
251                 Nt=1.798e13;
252             end
253         end
254         else begin
255             if (TRAP_TEMP_enable==0) begin
256                 Nt=5.25e12;
```

```
257                 end
258                 else begin
259                         Nt=1.291e13;
260                 end
261         end
262     end
263 
264     Ntrap=Nt;
265 
266     //Getting system temperature and update it with the heating
267     T=$temperature+heat;
268     delta_T=T-T_0;
269 
270     //Setting Breakdown voltage. B has a small variation with T.
271     B=1.234e-6*T-3.021e-4;
272     V_b=V_b0*(1.0+B*delta_T);
273     delta_V_b=V_b-V_b0*(1.0+(1.234e-6*$temperature-3.021e-4)*($temperature-
274 T_0));
275 
276     //SPAD active area, guard ring area and second junction area.
277     A = `M_PI*r**2;
278     A_g=2*`M_PI*(r+0.06e-4)*0.808e-4+`M_PI*((1.34e-4+r)**2-(r+0.06e-
279 4)**2)+2*`M_PI*(r+1.34e-4)*1.01e-4;
280     A_dw=2*`M_PI*(r+4.04e-4)*1.79e-4+`M_PI*(r+4.04e-4)**2;
281 
282     //Set Band-gap narrowing and Band-Gap Energy
283     BGN=9e-3*(ln(TD/1e17)+sqrt((ln(TD/1E+017))**2+0.5));
284     Eg=(Eg0-((4.73e-4*T**2)/(T+636))-BGN)*`P_Q;
285     delta_Eg = (Eg0*`P_Q-Eg)/2;
286 
287     //Updating Trap Activation Energies due to forbidden band-gap narrowing.
288     Et = (Et0 - delta_Eg/`P_Q);
289     Ea_1 = (Ea0_1 - delta_Eg);
290     Ea_2 = (Ea0_2 - delta_Eg);
291     Ea_3 = (Ea0_3 - delta_Eg);
292 
293     //Effective masses
294     m_de=pow(m_l*m_t*m_t,1.0/3.0);
295     m_dh=pow(pow(m_lh,1.5)+pow(m_hh,1.5),2.0/3.0);
296     m_e=1.045+4.5e-4*T;
297     m_h=0.523+0.0014*T-1.48e-6*T**2;
298 
299     //Density of states
300     Nc=2.0e-6*Mc*pow(((2*`M_PI*m_de*m_0*`P_K*T)/(`P_H)**2),1.5);
301     Nv=2.0e-6*pow(((2*`M_PI*m_dh*m_0*`P_K*T)/(`P_H)**2),1.5);
```

```
302
303     //Intrinsic concentration
304     ni=sqrt(Nc*Nv)*exp(-Eg/(2*`P_K*T)); //(1/cm3)
305
306     //Set Built-In voltage for junctions
307     phi=(`P_K*T)/(`P_Q)*ln((N_1*P_1)/ni**2);
308     phi_g=(`P_K*T)/(`P_Q)*ln((N_2*P_2)/ni**2);
309     phi_dw=(`P_K*T)/(`P_Q)*ln((N_3*P_3)/ni**2);
310
311     //Updating several potential parameters
312     V_e = (V(JUNC_1)+delta_V_b) - V_b;
313     V_d = (V(JUNC_1)+delta_V_b) + phi;
314     V_dg =  (V(JUNC_1)+delta_V_b) + phi_g;
315     V_ddw = V(JUNC_2) + phi_dw;
316
317     //Setting the junction maximun electric field
318     //Electric field is involved with BTBT and TAT dark counts so
319     //no point of modelling it below the breakdown voltage.
320
321     if (V_e < 0) begin
322         E_field_0=1.0e5;
323     end
324
325     if (V_e >= 0.856) begin
326         E_field_0 = -1.3299*V_e**4+31.55*V_e**3-
327 250.07*V_e**2+544.84*V_e+730578+780;
328     end
329
330     if (V_e >= 0 && V_e < 0.856 ) begin
331         E_field_0 = -10241*V_e**4+23868*V_e**3-
332 21056*V_e**2+10844*V_e+727460+780;
333     end
334
335     //Electric field temperature dependence
336     if (E_field_0 <= 0) begin
337         E_field=1.0e5;
338     end
339     else begin
340         E_field=E_field_0*(1+(6.748e-7*T-2.247e-4)*delta_T);
341     end
342
343     //Electron and hole thermal velocity
344     vthh=sqrt(abs((3*`P_K*T)/(m_h*m_0)));       //(m/s)
345     vthe=sqrt((3*`P_K*T)/(m_e*m_0));            //(m/s)
346
```

```
347     //Setting trap and deep level-trap cross sections //(cm2)
348     if (TRAP_TEMP_enable==0) begin
349         sig_0=sig_00;
350         st1=sigmat_1;
351         st2=sigmat_2;
352         st3=sigmat_3;
353     end
354     else begin
355         sig_0=sig_00*exp((Et*`P_Q/`P_K)*((1.0/300.0)-(1.0/T)));
356         st1=sigmat_1*exp((Ea_1/`P_K)*((1.0/300.0)-(1.0/T)));
357         st2=sigmat_2*exp((Ea_2/`P_K)*((1.0/300.0)-(1.0/T)));
358         st3=sigmat_3*exp((Ea_3/`P_K)*((1.0/300.0)-(1.0/T)));
359     end
360 
361     //Setting Electron capture rate for the ith trap (cm3/s)
362     theta_1=vthe*st1*100;
363     theta_2=vthe*st2*100;
364     theta_3=vthe*st3*100;
365 
366     //Junction effective width (m)
367     W=2*phi/E_field;
368 
369     //Setting saturation current for primary and secondary junctions (from TCAD
370 simulations) (A)
371     Is_a=2.174e-24*limexp(9.186e-2*T);
372     Is_b=-2.233e-16*T**2+1.174e-13*T-1.543e-11;
373     Is_2a=5.674e-26*limexp(1.04e-1*T);
374     Is_2b=2.184e-24*limexp(9.187e-2*T);
375 
376     if (Is_a > 1e-3 || Is_2a > 1e-3) begin
377         Is_a = 0;
378         Is_2a= 0;
379     end
380 
381     //Updating avalanche triggering probability for every run as V_e depends upon
382 Temperature
383     //For  0 <= V_e <= 1 we take Vornicu 2015 PDE results
384     //For V_e > 1 and V_e < 0.5 we follow the model.
385     if (V_e > 0) begin
386         if (V_e >= 0.5 && V_e <= 1.0) begin
387                 Ptr=32.4509*V_e**5-122.0404*V_e**4+182.0232*V_e**3-
388 135.571*V_e**2+51.1655*V_e-7.8192;
389         end
390         if (V_e > 1.0) begin
391                 Ptr = 1.0 - exp(-V_e/(eta_T_1*V_b));
```

```
392            end
393            if (V_e < 0.5) begin
394                    Ptr = 1.0 - exp(-V_e/(eta_T_2*V_b));
395            end
396        end
397        else begin
398            Ptr = 0.0;                        //If V_e <= 0 then Ptr = 0. For stability purposes.
399        end
400
401        if (Ptr < 0) begin            //Prevents Ptr from becoming negative. For stability
402    purposes
403            Ptr = 1e-8;
404        end
```

5. STATIC AND DYNAMIC CHARACTERISTICS

```
405        //Updating Junctions Current
406        Is = Is_a*(V(JUNC_1)+delta_V_b)+Is_b;
407        Is_2 = Is_2a*V(JUNC_2)+Is_2b;
408        Ibkr=(V_n/Rbkr)*ln(1+limexp(V_e/V_n));
409
410        //Updating Junctions and Stray Capacitors Stored Charge
411        Qj_1=(A*phi*Cj0/(1-mj))*pow(abs(1+V_d/phi),(1-mj))+(A_g*phi_g*Cj0_g/(1-
412    mj))*pow(abs(1+V_dg/phi_g),(1-mj));
413        Qj_2=(A_dw*phi_dw*Cj0_dw/(1-mj))*pow(abs(1+V_ddw/phi_dw),(1-mj));
414
415        Qas=Cas*V(A_SUB);
416        Qks=Cks*V(C_SUB);
```

6. PRIMARY JUNCTION'S CARRIER GENERATION RATE

```
417        //Carrier concentration and Injection density (cm-3)
418        alpha=((Eg/`P_Q)-Et)/(Eg/`P_Q);
419        Eb_p=alpha*Eg;
420        Eb_n=(1.0-alpha)*Eg;
421        n0=4.808e-5*exp(1.046e-1*T);
422        p0=ni**2/n0;
423        n1=Nc*exp(-Eb_n/(`P_K*T));
424        p1=Nv*exp(-Eb_p/(`P_K*T));
425
426        //Updating Injection density. From TCAD simulations.
427        if (abs(Ispad)<=1e-12) begin
428            n=7.878e3*exp(4.38e12*abs(Ispad));
429        end
430        else begin
431            if (abs(Ispad) <= 1e-4) begin
```

```
432                     n=1.192e18*abs(Ispad)+6.659e9;
433             end
434             else begin
435                     if (abs(Ispad) <= 5e-3) begin
436                             n=abs(-
437     1.256e20*(abs(Ispad))**2+1.192e18*(abs(Ispad))+3.736e11);
438                     end
439                     else begin
440                             n=3e15;
441                     end
442             end
443         end
444
445         delta_n=n-n0;
446
447         //Trap Assisted Tunneling: electric field enhancement factor (EFEF). From TCAD
448     simulations
449         g_1=((-1.388e-3)*T)+0.9303;
450         g_2=((-2.026e-10)*T)+1.04e-7;
451         g_3=((-8.721e-4)*T)+0.8198;
452         g_4=((-2.918e-10)*T)+1.377e-7;
453         gamma_n=g_1*limexp(E_field*100*g_2);
454         gamma_p=g_3*limexp(E_field*100*g_4);
455
456         //Carrier capture time by traps (s)
457         tau_0n=1/((vthe*100*Nt*sig_0)*(1+gamma_n));
458         tau_0p=1/((vthh*100*Nt*sig_0)*(1+gamma_p));
459         tn_notrap=1/((vthe*100*Nt*sig_0));    //Electron capture time (TAT disabled)
460         tp_notrap=1/((vthh*100*Nt*sig_0));    //Hole capture time (TAT disabled)
461
462         //Effective minority carrier lifetime (SRH + Auger + Radiative + Trap assisted
463     tunneling) (s)
464         //t_SRH << t_Auger and t_Radiative so t_Auger and t_Radiative can be
465     disregarded.
466
467         tau_eff=(tau_0n*(p0+p1+delta_n)+tau_0p*(n0+n1+delta_n))/(p0+n0+delta_n);
468                     //TAT enabled
469         te_nt=(tn_notrap*(p0+p1+delta_n)+tp_notrap*(n0+n1+delta_n))/(p0+n0+delta_n);
470                 //TAT disabled
471
472         //Thermal carrier generation rate (s-1)
473         cgr_th=((ni*W*A)/(tau_eff));     //TAT enabled
474         cgr_nt=((ni*W*A)/(te_nt));       //TAT disabled
475
476         //Band-to-band tunneling carrier generation rate (Hurkx Model for indirect
477     semiconductors)
```

```
478     cgr_btbt=A_H*pow(E_field,p)*limexp(-B_H/E_field)*A*W; //(s-1)
479
480     //CGR cannot be lesser than 1 for stability
481     if (cgr_th<=1) begin
482         cgr_th=1.0;
483     end
484     if (cgr_nt<=1) begin
485         cgr_nt=1.0;
486     end
487     if (cgr_btbt<=1) begin
488         cgr_btbt=1.0;
489     end
490
491     //Total carrier generation rate and time between 2 CG events for thermal
492 generation
493     if (TAT_enable==1) begin
494         cgr_tot = cgr_th + cgr_btbt;        //(s-1)
495         tau_cgth = 1/(cgr_th);                      //(s)
496         dcr_th=Ptr*cgr_th;
497     end
498     else begin
499         cgr_tot = cgr_btbt + cgr_nt;        //(s-1)
500         tau_cgth = 1/(cgr_nt);                      //(s)
501         dcr_th=Ptr*cgr_nt;
502     end
503
504     //Setting time between 2 CG events for BTBT generation
505     tau_cgbtbt = 1/(cgr_btbt);
506
507     //Setting Dark Count Rate for thermal and BTBT generation
508     dcr_btbt=Ptr*cgr_btbt;
509     dcr=cgr_tot*Ptr;
510
511     //Updating Ilat
512     Ilat = $rdist_normal(sdIlat,mIlat,sigIlat);
513
514     //Setting thermal conductivity (W/cmK) and thermal resistivity (K/W)
515     th_cond=(-1.4067e-5*T**3+1.6083e-2*T**2-6.5365*T+1.0404e3)/100;
516     th_res=1299.63/(th_cond);
517
518     //Updating mean time between two 1st RC events because of T variations
519     tau_cr1 = t0_1*exp(Ea_1/(`P_K*T));
520
521     //Updating mean time between two 2nd RC events because of T variations
522     tau_cr2 = t0_2*exp(Ea_2/(`P_K*T));
```

```
523
524     //Updating mean time between two 3rd RC events because of T variations
525     tau_cr3 = t0_3*exp(Ea_3/(`P_K*T));
```

7. INITIAL CONDITIONS

```
526     @(initial_step) begin
527
528     //Schedule the first CG event for thermal generation
529     deltatcg_th = $rdist_exponential(sd_cgth,tau_cgth);
530     tcg_th = $abstime + deltatcg_th;
531
532
533     //Schedule the first CG event for btbt generation with an initial mean for
534 convergence.
535     tau_cgbtbt=1e-5;
536     deltatcg_btbt = $rdist_exponential(sd_cgbtbt,tau_cgbtbt);
537     tcg_btbt = $abstime + deltatcg_btbt;
538
539     //mean time between two 1st RC events
540     tau_cr1 = t0_1*exp(Ea_1/(`P_K*T));
541
542     //mean time between two 2nd RC events
543     tau_cr2 = t0_2*exp(Ea_2/(`P_K*T));
544
545     //mean time between two 3rd RC events
546     tau_cr3 = t0_3*exp(Ea_3/(`P_K*T));
547
548     //Set Fill Factor (Based on Vornicu 2015)
549     FF=`M_PI*r**2/A_tot;
550
551     //Parameters initialization
552     reset=1;
553     sd_cgth=5;
554     sd_cgbtbt=15;
555     sd_cr1=20;
556     sd_cr2=25;
557     sd_cr3=30;
558     sd_tr=35;
559     sd_pa1=40;
560     sd_pa2=45;
561     sd_pa3=50;
562     sd_ct1=55;
563     sd_ct2=60;
564     sd_ct3=65;
565     heat=0;
```

```
566     kcg_th=0;
567     kcg_btbt=0;
568     kdc_th=0;
569     kdc_btbt=0;
570     kdc=0;
571     kdet=0;
572     kndet=0;
573     kap=0;
574     kap_1=0;
575     kap_2=0;
576     kap_3=0;
577     pow_r=0;
578     aux1=0;
579
580     // Model Options (0=NOT ENABLED,1=ENABLED)
581         //Self-Heating
582         SH_enable=1;                    //Self-Heating
583
584         // Trap and deep-level trap cross-section temperature depence
585         TRAP_TEMP_enable=1;
586
587         //Carrier generation by Trap-assisted tunneling
588         TAT_enable=1;
589
590     end
591
592     //Update Quantum Efficiency (From TCAD simulations)
593     QE_table[1]=-3.501e-6*T**2+2.923e-3*T+0.2434;
594     QE_table[2]=-2.887e-6*T**2+2.083e-3*T+0.5466;
595     QE_table[3]=-7.319e-7*T**2+6.063e-4*T+0.8331;
596     QE_table[4]=-4.501e-7*T**2+3.344e-4*T+0.8361;
597     QE_table[5]=0.87;
598     QE_table[6]=0.77;
599     QE_table[7]=0.62;
600     QE_table[8]=0.24;
601
602     i = 1;
603     for (i=1;i<8;i=i+1) begin
604         if ((lambda >= l_table[i]) && (lambda <= l_table[i+1])) begin
605                 QE=((QE_table[i+1]-QE_table[i])/(l_table[i+1]-l_table[i]))*(lambda-
606 l_table[i])+QE_table[i];
607
608         end
609
610     end
```

```
611
612     //Set Photon Detection Efficiency
613     pde = Ptr*QE*FF;
614
615     ncrit = ((p0+p1)*(n1+n0+delta_n+(tau_0n/tau_0p)*(p0+p1+delta_n)))/abs(p0-
616  (tau_0n/tau_0p)*p1);
617
618
```

8. PHOTON ARRIVAL

```
619     if ((V(photon) > PhThereshold) && (aval==0) && V_e>=0)
620     begin
621         //Cheking Photon Detection Efficiency
622         r1= $rdist_uniform(sd_det,0.0,1.0);
623         if ( r1 < pde ) begin
624             aval = 1;
625             reset=1;
626             b_av_time=$abstime - av_i_time;
627             av_i_time=$abstime;
628             kdet = kdet + 1;
629         end
630     end
```

9. DARK COUNTS: THERMAL CARRIER GENERATION

```
631     @(timer(tcg_th))
632     begin
633         // Increasing CG counter
634         kcg_th = kcg_th + 1;
635
636         // Setting bias excess voltage
637         if (aval == 0 && V_e >= 0)
638         begin
639             //DC event
640             if ($rdist_uniform(sd_tr,0.0,1.0) < Ptr)
641             begin
642                 //Start the avalanche
643                 aval=1;
644                 reset=1;
645                 b_av_time=$abstime - av_i_time;
646                 av_i_time=$abstime;
647                 kdc_th = kdc_th +1;
648                 kdc = kdc+1;
649             end
650         end
```

```
651
652         //Schedule next CG event
653         deltatcg_th = $rdist_exponential(sd_cgth,tau_cgth);
654         tcg_th = $abstime + deltatcg_th;
655     end
```

10. DARK COUNTS: BAND-TO-BAND TUNNELING

```
656     @(timer(tcg_btbt))
657     begin
658         // Increasing CG counter
659         kcg_btbt = kcg_btbt + 1;
660
661         // Setting bias excess voltage
662         if (aval == 0 && V_e >= 0)
663         begin
664                 //DC event
665                 if ($rdist_uniform(sd_tr,0.0,1.0) < Ptr)
666                 begin
667                         //Starting the avalanche
668                         aval=1;
669                         b_av_time=$abstime - av_i_time;
670                         reset=1;
671                         av_i_time=$abstime;
672                         kdc_btbt = kdc_btbt +1;
673                         kdc=kdc+1;
674                 end
675         end
676         //Schedule next CG event
677         deltatcg_btbt = $rdist_exponential(sd_cgbtbt,tau_cgbtbt);
678         tcg_btbt = $abstime + deltatcg_btbt;
679
680     end
```

11.DARK COUNTS: AFTER-PULSING

```
681     @(timer(tcr1))
682     begin
683         if  (aval==0 && V_e>=0) begin
684
685                 //There is a chance of after-pulsing. Checking triggering probability...
686                 if ($rdist_uniform(sd_tr,0.0,1.0) < Ptr) begin
687                         //Starting the avalanche
688                         aval=1;
689                         b_av_time=$abstime - av_i_time;
690                         reset=1;
```

```
691                     av_i_time=$abstime;
692                     kap = kap +1;
693                     kap_1=kap_1+1;
694                 end
695         end
696     end
697 
698     @(timer(tcr2))
699     begin
700         if  (aval==0 && V_e>=0) begin
701 
702                 //There is a chance of after-pulsing. Checking triggering probability...
703                 if ($rdist_uniform(sd_tr,0.0,1.0) < Ptr) begin
704                 //Starting the avalanche
705                     aval=1;
706                     b_av_time=$abstime - av_i_time;
707                     reset=1;
708                     av_i_time=$abstime;
709                     kap = kap +1;
710                     kap_2=kap_2+1;
711                 end
712         end
713     end
714 
715     @(timer(tcr3))
716     begin
717         if  (aval==0 && V_e>=0) begin
718 
719                 //There is a chance of after-pulsing. Checking triggering probability...
720                 if ($rdist_uniform(sd_tr,0.0,1.0) < Ptr) begin
721                     //Start the avalanche
722                     aval=1;
723                     b_av_time=$abstime - av_i_time;
724                     reset=1;
725                     av_i_time=$abstime;
726                     kap = kap +1;
727                     kap_3=kap_3+1;
728                 end
729         end
730     end
```

12. TURN-OFF

```
731     if (aval==1 && reset==1) begin
732         reset=0;
```

```
733     end
734 
735     n_aval=idt(Ispad,0,reset)/(`P_Q);
736 
737     if ((aval==1) && (Ibkr < Ilat))
738     begin
739         //Quenching the avalanche
740         aval=0;
741         av_time=$abstime-av_i_time;
742 
743         //After an avalanche turn-off, checking deep-level traps
744         //for afterpulsing probabilities.
745 
746         //Updating pre-exponential factors
747         pc_1=theta_1*n_aval*av_time*1e-7/(A*W*tau_cr1);
748         pc_2=theta_2*n_aval*av_time*1e-7/(A*W*tau_cr2);
749         pc_3=theta_3*n_aval*av_time*1e-7/(A*W*tau_cr3);
750 
751         //Checking after-pulsing probability for 1st deep-level traps (Ptr check not
752 included here)
753         deltatcr1 = $rdist_exponential(sd_cr1,tau_cr1);
754         Pa_1= ((pc_1))*exp(-(deltatcr1)/tau_cr1);
755         if ($rdist_uniform(sd_pa1,0.0,1.0)<Pa_1) begin
756             tcr1 = av_i_time + deltatcr1;
757         end
758 
759         //Checking after-pulsing probability for 2nd deep-level traps (Ptr check not
760 included here)
761         deltatcr2 = $rdist_exponential(sd_cr2,tau_cr2);
762         Pa_2= ((pc_2))*exp(-(deltatcr2)/tau_cr2);
763         if ($rdist_uniform(sd_pa2,0.0,1.0)<Pa_2) begin
764             tcr2 = av_i_time + deltatcr2;
765         end
766 
767         //Check after-pulsing probability for 3rd deep-level traps (Ptr check not
768 included here)
769         deltatcr3 = $rdist_exponential(sd_cr3,tau_cr3);
770         Pa_3= ((pc_3))*exp(-(deltatcr3)/tau_cr3);
771         if ($rdist_uniform(sd_pa3,0.0,1.0)<Pa_3) begin
772             tcr3 = av_i_time + deltatcr3;
773         end
774 
775         reset=1;        //Reset the electron counter or the avalanches.
776     end
```

13. CURRENT CONTRIBUTIONS SETTING

```
//Set current across the SPAD
Ispad = Is + aval*Ibkr;

//Updating device currents
Ij1 = ddt(Qj_1);
Ij2 = ddt(Qj_2);
Ias = ddt(Qas);
Iks = ddt(Qks);

I(SPAD)<+ Ispad;
I(JUNC_1)<+ ddt(Qj_1);
I(JUNC_2)<+ ddt(Qj_2);
I(JUNC_2)<+Is_2;
I(A_SUB)<+ ddt(Qas);
I(C_SUB)<+ ddt(Qks);

//Setting Power (W) and Self-Heating contribution (K)
//The following code calibrate the heat due to power dissipation
//so it is not affected by initial conditions or anomalous
//current variations which may make power arise beyond 1.06 times
//of its initial value.

i_1 = abs(I(SPAD));
pow_r = SPADRes*i_1**2;

if (pow_r > power) begin
    power = SPADRes*i_1**2;
    heat = th_res*power*SH_enable;

    if (aux1 < 1 && power > 1e-4) begin
            pow_i = power;
            aux1 = aux1 + 1;
    end

end

if (power > (1.06*pow_i)) begin
    power = power*0.5;
end
end

endmodule
```

APPENDIX II

P-GATED/VIRTUAL GUARD RING DUAL-JUNCTION SPAD DESIGN

1. INTRODUCTION

Photodiode structures typically embed *stacked pn junctions*, some of which are secondary for the nominal detection function. These secondary functions may be considered parasitics or exploited for enhanced detection efficiency and spectral discrimination [1][2][3]. Similar to conventional photodiodes, SPADs structures include several junctions [4][5], and different authors have addressed their possible usage. The bullet points below describe two major lines in this direction.

- Finkelstein et al [6] employ a CMOS SPAD with two sensing junctions for biological imaging of separated fluorescents probes. However, the shallowest junction is protected from edge breakdown by a Shallow-Trench-Isolation or STI, which makes the dark counts grow enormously due to surface traps. Besides, the deeper junction does not have a guard ring to avoid peripheral breakdown.
- Henderson et al. [7] propose a technique to guard both junctions appropriately. However, the deeper junction would trigger avalanches directly into the substrate, making the SPAD unsuitable for large arrays because measures to preclude crosstalk result in reduced fill factors. Still, the pulse duration for both junctions varied with the wavelength, and in theory, this feature enables the discrimination of absorbed photons by wavelength.

This Appendix focuses on the P12 structure, which consists of a stacked triple junction SPAD (two acting as sensing junctions, one as an insulator) modeled by TCAD simulations in standard 180nm technology – see device cross-section in FIGURE 1. It employs a perimeter-gating approach that prevents the primary junction from edge breakdown. The device is similar to the P-Well device developed in [4], which uses a virtual guard ring to avoid a peripheral breakdown in the primary junction. However, the proposed device includes a N+ well within the P-Well, thus creating a third junction, N+/P-Well, that is prevented from premature edge breakdown with the perimeter gating approach, instead of the classic guard ring or virtual guard ring.

As in [7], this device inherits the capabilities of its parent devices [5][8]: low DCR, high sensitivity, and high temporal resolution. However, quantum efficiency proves to be lower than that of other SPADs in the literature.

2. DEVICE STRUCTURE

The left edge in FIGURE 1 acts as an revolution axis. The primary junction is formed by the N+ well and the P-Well, where a positive potential at the Gate electrode avoids edge breakdown. The P-Well and the Deep-N-Well form the secondary junction. The virtual guard ring formed by blocking the peripheral N-Well formation ensures a retrograde doping profile, lowering the electric field in the junction so that it needs a larger potential to enter breakdown. That way, only the deepest part will trigger avalanches. The tertiary junction, formed by the Deep-N-Well and the substrate, would serve as an insulator when a positive potential is applied to the second cathode, being reverse biased.

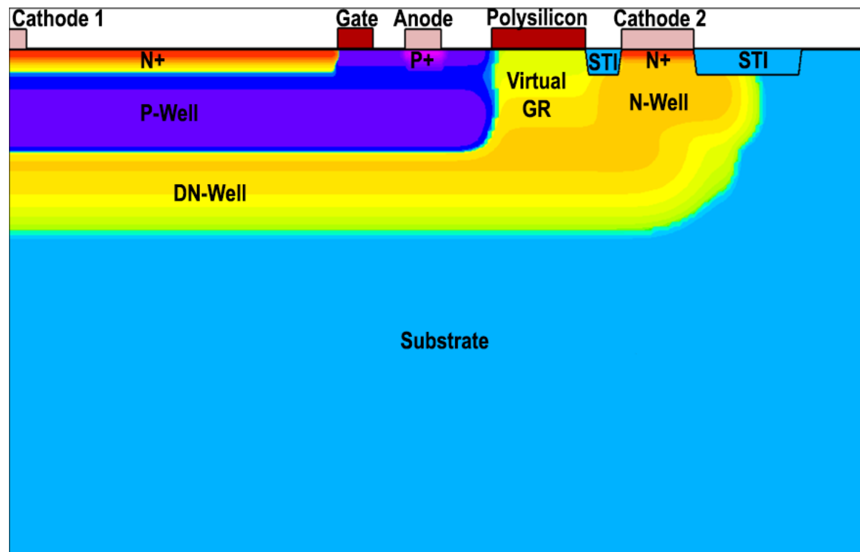


FIGURE 1 Device cross section. Left border serves as axis of revolution.

3. DEVICE CHARACTERISTICS

The breakdown voltage of the secondary junction is $V_{B2} = 12.292$V, while the breakdown voltage of the primary junction is regulated by the gate voltage. When the gate is not biased, the primary junction breakdown region is localized at the edge instead of the active area region. This is due to the fact that the doping profile at the junction edge is more abrupt, which causes the electric field in this region to be higher. As a consequence, the primary junction active area is unable to trigger avalanches from photon strikes at low bias voltages. If the gate electrode is positively biased, electrons accumulate below the gate, lowering the potential on that side of the primary junction edge and making the electric field across the junction lower. This cause the breakdown region of the junction to actually move towards a zone with a now higher electric field when the junction is reverse biased over the breakdown voltage, ejecting it from the edge. As a result, the breakdown voltage rises to that of the active area region. FIGURE 2 shows how as the Gate voltage rises whereas the electric field at the edge of the junction drops. As a result, the breakdown voltage rises. This effect can also be seen in FIGURE 3, where the anode current is represented against the Cathode 1 potential for several Gate voltages.

Both cathodes must have a positive bias over the breakdown voltage to operate both junctions in Geiger mode. In this mode, photons can generate electron-hole pairs capable of triggering avalanches in either junction. The detection region (see Chapter 7) of a normal P-Well/Deep-N-Well SPAD would then be simply divided into two detection regions, one for each junction. Also, the dispersion time of this carrier to reach any of the depletion zones will be shorter, positively affecting the timing jitter.

FIGURE 4 shows the *QE* obtained from TCAD simulations when illuminating the primary junction active area. The first junction has its maximum quantum efficiency at 450nm, while the second junction has a maximum of 600nm. This difference should make the device able to differentiate colors. Besides, as was shown in Chapter 5, the avalanche triggering probability can be related to wavelength too. Because ATLAS cannot work with two junctions simultaneously, an approximation has been made based on the *QE* of the two junctions.

Finally, FIGURE 5 shows the anode current against the excess bias voltage over breakdown voltage for the primary and the secondary junctions.

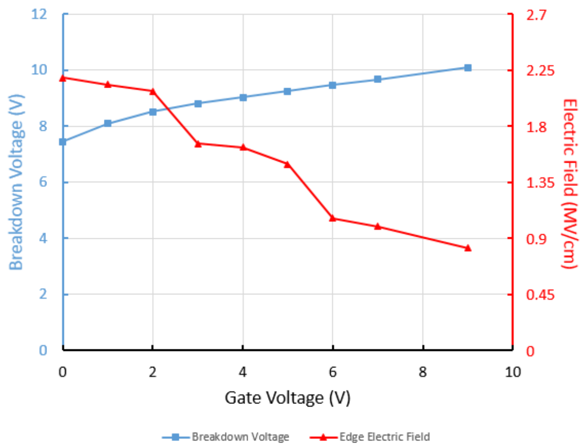


FIGURE 2 Primary junction breakdown voltage and electric field at the edge of the primary junction vs Gate voltage.

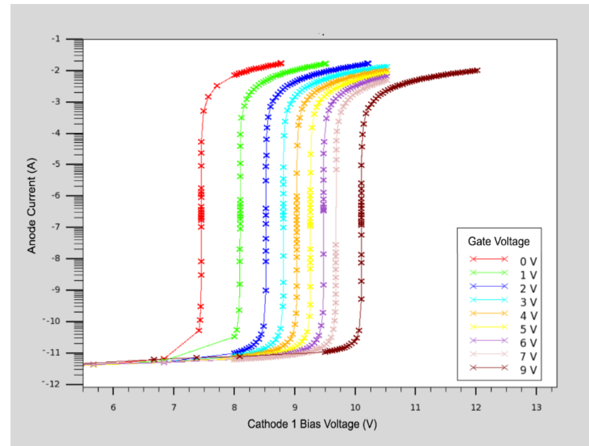


FIGURE 3 Anode current (logarithmic scale) vs Cathode 1 Bias voltage while keeping the secondary junction below the breakdown voltage for several gate voltages.

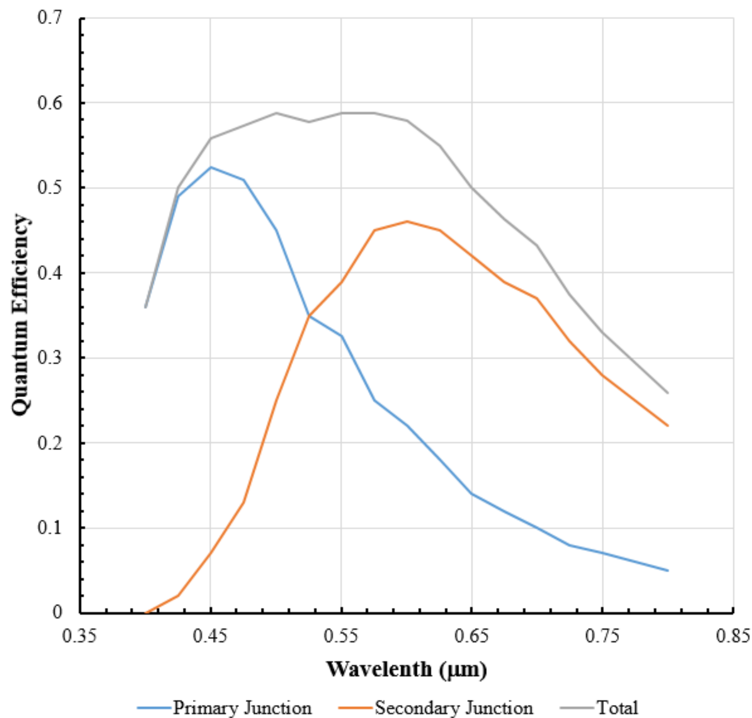


FIGURE 4 QE vs Wavelength for the primary and secondary junction and an approximation of the total combination.

In summary, this Appendix overviews a new device based on the stacked junctions' technology. This device, with color sensing capabilities, can also be used efficiently in pixel arrays unlike previous stacked junctions SPADs in the literature, which opens its possible use in many applications.

4. REFERENCES

[1] R. B. Merrill, "Color separation in an active pixel cell imaging array using a triple-well structure." U.S. Patent 5 965 875, Oct. 12, 1999.

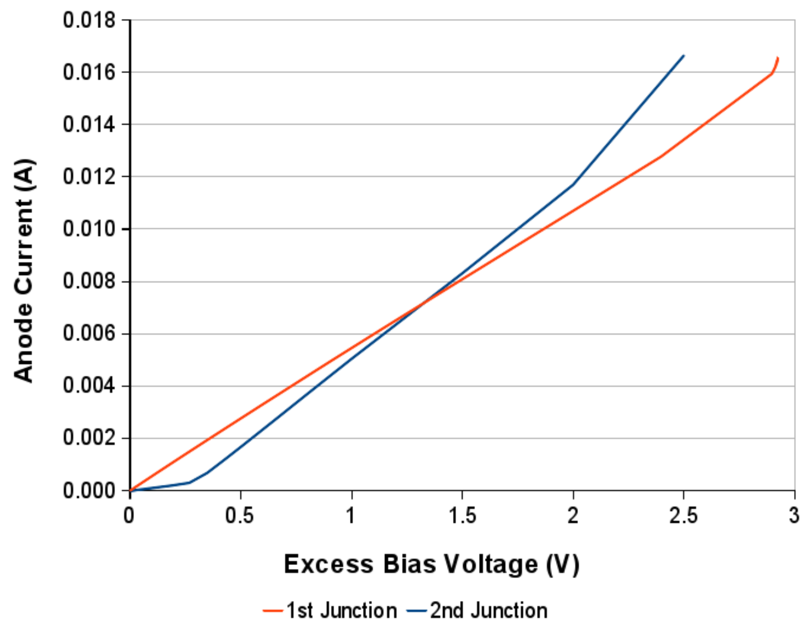


FIGURE 5 Anode Current vs Excess Bias Voltage for both, primary and secondary junctions.

[2] S. Feruglio, G.-N. Lu, P. Garda, and G. Vasilescu, “A Review of the CMOS Buried Double Junction (BDJ) Photodetector and its Applications.” *Sensors*, vol. 8, no. 10, pp. 6566–6594, Oct. 2008.

[3] R. Gómez-Merchán, D. Palomeque-Mangut, J.A. Leñero-Bardallo, M. Delgado-Restituto, and A. Rodríguez-Vázquez, "A Comparative Study of Stacked-Diode Configurations Operating in the Photovoltaic Region.” *IEEE Sensors Journal,* vol. 20, no. 16, pp. 9105-9113, August 2020.

[4] M. Ghioni, A. Guilinatti, I. Rech, F. Zappa, and S. Cova, “Progress in Silicon Single-Photon Avalanche Diodes.” *IEEE J. Sel. Topics Quantum Electron.*, vol. 13, no. 4, pp. 852–862, Jul./Aug. 2007.

[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 Transactions on Electron Devices*, Vol. 58, no. 7, pp. 2028-2035, July 2011.

[6] H. Finkelstein, M. J. Hsu, and S. C. Esener, “Dual-junction single-photon avalanche diode.” *Electron. Lett.*, vol. 43, no. 22, pp. 1228–1229, Oct. 2007.

[7] R. K. Henderson, E. A. G. Webster and L. A. Grant, “A Dual-Junction Single-Photon Avalanche Diode in 130-nm CMOS Technology.” *IEEE Electron Device Letters*, vol. 34, no. 3, Mar. 2013.

[8] J. Gu, M. Habib Ullah Habib, and N. McFarlane, “Perimeter-Gated Single-Photon Avalanche Diodes: An Information Theoretic Assessment.” *IEEE Photonics Technology Letters*, vol. 28, no. 6, Mar. 2016.

APPENDIX III
ACRONYMS

The following table describes the meaning of various acronyms and special terms used throughout the dissertation.

Acronym	Meaning
4T-APS	4-Transistor Active Pixel Sensor
AQC	Active Quenching Circuit
ATHENA	Simulator for numerical, physically-based, two-dimensional simulation of semiconductor processing owned by Silvaco
ATLAS	Physically-based two and three dimensional device simulator owned by Silvaco
BTBT	Band-To-Band Tunneling
CGR	Carrier Generation Rate
CMOS	Complementary Metal-Oxide-Semiconductor
CTK	Crosstalk
DCR	Dark Counts Rate
DECKBUILD	Interactive, graphic runtime environment for developing process and device simulation input decks owned by Silvaco
DNW	Deep-N-Well
ELT	Epitaxial Layer Thickness
EQE	External Quantum Efficiency
FD	Floating Diffusion
FF	Fill factor
FLIM	Fluorescence Lifetime Imaging Microscopy
HDL	Hardware Description Language
IQE	Internal Quantum Efficiency
PDE	Photon Detection Efficiency
PDP	Photon Detection Probability
PEB	Premature Edge Breakdown
PET	Positron Emission Tomography
PGSPAD	Perimeter Gated Single-Photon Avalanche Diode
PPD	Pinned Photodiode
PQC	Passive Quenching Circuit
QE	Quantum Efficiency
RST	Reset Transistor
RT-SPAD	Reach-Through Single-Photon Avalanche Diode
SH	Self-Heating
SiPM	Silicon Photomultiplier
SNR	Signal-to-Noise Ratio
SPAD	Single-Photon Avalanche Diode
SPC	Source Photo-Current
SPECT	Single-Photon Emission Computed Tomography
SPECTRE	SPICE-class circuit simulator owned by Cadence Design Systems
SPICE	General-purpose, open-source analog electronic circuit simulator
SRH	Shockley-Read-Hall Theory

STI	Shallow Trench Isolation
TAT	Trap-Assisted Tunneling
TATG	Trap-Assisted Thermal Generation
TCAD	Technology Computer-Aided Design. Branch of electronic design automation that models semiconductor fabrication and semiconductor device operation.
TG	Transfer Gate
TOF	Time Of Flight
TONYPLOT	Graphical post-processing tool for use with all Silvaco simulators.
TRAP.T	Mid-gap trap cross-section dependence on temperature
TSCPC	Time Correlated Single Photon Counting
VERILOG-A	Industry standard modeling language for analog circuits

ACKNOWLEDGEMENTS

Este es el fin de mi viaje para obtener el doctorado. Un viaje que ha estado lleno de alegrías, de dolores y de experiencias únicas que te ponen delante de ese espejo que es la vida y hacen relativices tus miedos e inseguridades. Aún recuerdo, hace seis años, cuando iba a mi andadura por el Instituto de Microelectrónica de Sevilla, las palabras del Dr Ricardo Carmona Galán, el que sería mi tutor, en su primera llamada para ofrecerme la posibilidad de convertirme en un doctorando bajo su tutela, diciéndome que no sería un camino fácil. Que necesitaría mucha sacrificio y dedicación en el futuro. Sinceramente, pensaba que no sería fácil, ¡pero tampoco pensaba que ese futuro incluiría una pandemia y tres hijos!

A pesar de las extraordinarias adversidades que me salieron al paso, me encuentro ahora, escribiendo estas líneas, pensando en todas esas fantásticas personas y en todas esas fantásticas experiencias que he tenido el privilegio de vivir.

En primer lugar, gracias a mi tutor, Dr. Ricardo Carmona Galán, cuya hábil tutela y profundo conocimiento de los sensores de imagen y los fotodiodos, me han servido de inspiración y han hecho posible que esta Tesis llegue adonde está ahora. Gracias también por todo ese apoyo que no pocas veces me ha ayudado a seguir adelante. Gracias.

Gracias también a mi director Dr. Ángel Rodríguez Vázquez, por ser esa figura que sabes que siempre estará ahí, contigo, frente a los problemas. Por ser esa persona que, con su cercanía y su abnegada dedicación, y su increíble capacidad para este trabajo, hace que los que le rodean podamos alcanzar metas más altas de las que creíamos posible, tanto como personas, como profesionales. Gracias.

Me gustaría también agradecer al Dr. Fernando Vidal Verdú de todo corazón, la posibilidad de hacer la estancia virtual con su grupo, dadas las circunstancias excepcionales que se derivaron de la pandemia, y espero poder tener la oportunidad en el futuro de colaborar de nuevo con él.

Gracias también a todas esas personas que tenían parte de su hogar en el Instituto de Microelectrónica de Sevilla y que tuve la fortuna de encontrarme en su camino, y que juntas formaban seguramente el entorno más profesional que he conocido. Gracias Ion, Jorge, Marco, Pablo, Antonio, Valentín, Norberto, David, Franco. Gracias también a todos esos profesionales que permiten que el IMSE siga funcionando como un reloj.

Muchísimas gracias también a la Dra. Gloria Huertas y a la Dra. Rocío del Rio, que me brindaron la maravillosa oportunidad de impartir docencia en la Universidad. Muchas gracias.

Por último, me gustaría dar especialmente las gracias a esas personas, sin cuyo sacrificio, amor y apoyo incondicional, que necesitaría seguramente más de una vida para poder corresponder, no habría sido posible esta Tesis. Gracias Txell, por elegirme como compañero para este breve lapso en el tiempo del universo al que llamamos vida. Gracias Freya. Gracias Èric. Gracias Ender. Gracias por hacerme reír y sacar al niño interior que llevo dentro. Gracias por darle sentido a todo.

LIST OF PUBLICATIONS

JOURNAL PAPERS

I. Vornicu, **J. M. López-Martínez**, F. N. Bandi, R. C. Galán and Á. Rodríguez-Vázquez, "Design of High-Efficiency SPADs for LiDAR Applications in 110nm CIS Technology," in *IEEE Sensors Journal*, vol. 21, no. 4, pp. 4776-4785, 15 Feb.15, 2021, doi: 10.1109/JSEN.2020.3032106.

CONFERENCE PROCEEDINGS

J. M. López-Martínez, R. Carmona-Galán and A. Rodríguez-Vázquez, "Limitation of SPADs quantum efficiency due to the dopants concentration gradient," *2020 27th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, 2020, pp. 1-4, doi: 10.1109/ICECS49266.2020.9294798.

J. M. López-Martínez, R. Carmona-Galán and Á. Rodríguez-Vázquez, "Photon-Detection Timing-Jitter Model in Verilog-A," *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2020, pp. 1-5, doi: 10.1109/ISCAS45731.2020.9181222.

J. M. López-Martínez, I. Vornicu, R. Carmona-Galán and Á. Rodríguez-Vázquez, "An Experimentally-Validated Verilog-A SPAD Model Extracted from TCAD Simulation," *2018 25th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, 2018, pp. 137-140, doi: 10.1109/ICECS.2018.8617962.

J. M. López-Martínez, R. Carmona-Galán and Á. Rodríguez-Vázquez, "Characterization of electrical crosstalk in 4T-APS arrays using TCAD simulations," *2017 13th Conference on Ph.D. Research in Microelectronics and Electronics (PRIME)*, 2017, pp. 237-240, doi: 10.1109/PRIME.2017.7974151.

WORKSHOPS

J. M. López-Martínez, I. Vornicu, R. Carmona-Galán, J. Fernández-Berni, and Á. Rodríguez-Vázquez, "TCAD Simulation of Electrical Crosstalk in 4T-Active Pixel Sensors," in *6th Workshop on Architecture of Smart Cameras (WASC)*, Jun. 5-6, 2017, Córdoba.