UNIVERSITÉ DE STRASBOURG

ECOLE DOCTORALE MSII

ICUBE – UMR7357

THÈSE présentée par :

Sven RINK

soutenue le : **14 Décembre 2023**

pour obtenir le grade de : **Docteur de l'Université de Strasbourg**

Discipline/ Spécialité : Science de l'ingénieur / Micro- et Nanoélectronique

Modelization of Single-Photon Avalanche Diode (SPAD) for the Design of Integrated Imager Pixels in advanced CMOS technology

THÈSE dirigée par :

M. LALLEMENT Christophe	Professeur, Université de Strasbourg
Et co-dirigée par	
M. UHRING Wilfried	Professeur, Université de Strasbourg

RAPPORTEURS :

M. CHARBON Edoardo	Professeur, établissement École Polytechnique Fédérale de Lausanne (EPFL), Suisse
M. SALLESE Jean-Michel	Maître d'enseignement et de recherche, École Polytechnique Fédérale de Lausanne (EPFL), Suisse

AUTRES MEMBRES DU JURY :

Mme. DESGREYS Patricia	Professeur, Télécom Paris – Institut polytechnique de Paris
M. GOUGET Gilles	Ingénieur, STMicroelectronics Crolles
M. KAMMERER Jean-Baptiste	Maître de conférences, Université de Strasbourg
M. HEBRARD Luc	Professeur, Université de Strasbourg

INVITÉS : M. QUENETTE Vincent	Docteur Ingénieur, STMicroelectronics Crolles

Acknowledgements

I would like to express my sincere gratitude to Christophe Lallement for directing this research work and for his patience, support, and encouragement. Without his supervision this research work would not have been possible.

The interesting ideas presented in this research work would not have been thinkable without the brilliant and instructive suggestions of my supervisors Jean-Baptiste Kammerer and Vincent Quenette and my co-director Wilfried Uhring. I thank them very much for their support, discussions and unique pedagogical guidance aswell as their constant advice and instructions that immensely helped in developing the model in the best way.

I would also like to thank Gilles Gouget and André Juge who greatly contributed to this thesis notably through our combined effort in the IPCEI NANO2022 reports. I also want to express my gratitude to Jean-Robert Manouvrier who provided the model framework in which the developed model inserts itself, and the corresponding guidance associated with it.

Additionally, I want to thank Denis Rideau, Raul-Andres Bianchi, Alexandre Dartigues, Elsa Lacombe, Raphael Neri, Florent Brun and Dominique Golanski for their effort in collaboration with our team of the ICube laboratory to develop and provide insights on the ‘‘KNACK” test circuit.

Most importantly, I would like to thank my family, friends and particularly my beloved partner for their permanent support, patience and encouragement without whom this research work would not have been possible.

Summary

Table of figures

Notations

Fundamental constants

Designation	Name and Value
Elementary charge	$q = 1.602 \cdot 10^{-19}$ C
Boltzmann's constant	$k = 1.380658 \cdot 10^{-23}$ J. K^{-1}
Planck's constant	$h = 6.62607004 \cdot 10^{-34}$ m^2. kg. s^{-1}
Bandgap energy at 300K for Silicon	$E_g = 1.12\ eV$
Dielectric permittivity of void	$\varepsilon_0 = 8.854 \cdot 10^{-12}$ F. m^{-1}
Dielectric permittivity of Silicon	$\varepsilon_{Si} = 11.8 \cdot \varepsilon_0$
Dielectric permittivity of Silicon oxyde	$\varepsilon_{ox} = 3.9 \cdot \varepsilon_0$

Notations

Designation	Name	Unit
Intrinsic concentration for Silicon	n_i	m^{-3}
Electron mobility for Silicon	μ_e	m. V^{-1}. s^{-1}
Carrier saturation velocity for Silicon	v_s	m. s^{-1}
Space charge region width	W	m
Semiconductor temperature	T	K
P-type doping concentration	N_d	m^{-3}
N-type doping concentration	N_a	m^{-3}
Breakdown voltage	V_{br}	V
Excess voltage	V_{ex}	V

Reverse biasing voltage	V_{bias}	V
Effective electron mass	m_e^*	kg
Effective hole mass	m_h^*	kg
Electron density	n	m^{-3}
Hole density	p	m^{-3}
Life time of minority carriers	τ_0	s
Fermi energy level for P-type Silicon	E_{F_p}	eV
Fermi energy level for N-type Silicon	E_{F_n}	eV
Conduction band level	E_C	eV
Valence band level	E_V	eV
Impact ionization threshold energy	E_{th}	eV
Temperature coefficient for breakdown voltage	β	K^{-1}
Ionization coefficient for electrons	α_n	m^{-1}
Ionization coefficient for holes	α_p	m^{-1}
Electrostatic potential	$\Psi(z)$	V
Dead space	d	m
Thermal and Trap Assisted Tunneling genration rate	$G_{SRH-TAT}$	m^{-3}
Band To Band Tunneling genration rate	G_{BTBT}	m^{-3}
Conduction Band carrier density	N_c	m^{-3}
Valence Band carrier density	N_v	m^{-3}
Band gap trap density on level i	N_{tni}	m^{-3}
Electron capture cross section	σ_{ni}	m^2
Electron capture rate	$U_{capt,ni}$	$\mathrm{m}^{-3}.\mathrm{s}^{-1}$

Number of electrons generated during avalanche	N_{av}	/
SPAD surface	S_{SPAD}	m^2
Number of electrons captured by traps on level i	Λ_{ni}	/
Trap released electron concentration	R_{ni}	m^{-3}
Hurkx field enhancement parameter	$\Gamma_{TAT}(F)$	/
Trap lifetime	τ_{ni}	s
Trap Energy	E_{tni}	eV
Trap release probability of electrons	η_i	/

Abbreviations

Designation	Name
SPAD	Single Photon Avalanche Diode
CMOS	Complementary Metal Oxyde Semiconductor
FF	Fill factor
PDP	Photon Detection Probability
DCR	Dark Count Rate
SPTR	Single Photon Timing Resolution
PDE	Photon Detection Efficiency
DT	Dead Time
AP	Avalanche Probability
TCAD	Technology Computer–Aided Design
SRH-TAT	Schockley-Read-Hall Trap Assisted Tunneling
BTBT	Band To Band Tunneling

SCR	Space Charge Region
QE	Quantum Efficiency
STI	Shallow Trench Isolation
DTI	Deep Trench Isolation
VGR	Virtual Guard Ring
ToF	Time of Flight
TDC	Time to Digital Converter
TCSPC	Time-Correlated Single Photon Counting

Summary

After a brief introduction on integrated imagers in CMOS technology used in today's semiconductor industry and a description of the studied device, the Single Photon Avalanche Diode (SPAD), in **Chapter I**, we first go through the stakes of the thesis. Following up the motivations and objectives in terms of characterization and compact modelling, the principal figures of merit of the SPAD device are enumerated and described. Then the core phenomenon of the avalanche diodes, the avalanche breakdown induced by impact ionization of carriers inside the Space Charge Region (SCR) of the device, is studied in detail. After that, different modelling approaches of avalanche breakdown and figures of merit are studied to provide insights on how those figures can already be designed. Finally, different SPAD architectures, quenching circuits and measuring techniques are rounding up this bibliographic chapter to provide a baseline of the state of the art in terms of SPAD modelling and putting an accent on the need for a new, different, and more elegant approach on avalanche breakdown modelling.

Chapter II then delves into the first step in the developed modelling approach, first describing the studied device's reach through type architecture and opening on the impact ionization modelling approach used. After this crucial mechanism has been clearly defined and the design approach of this phenomenon outlined, the translation of impact ionization as it is implemented in our model is then described in detail. This description goes from the definition of the loop-based model structure to the definition of the underlying physical equations and parameters. To close this chapter, the first simulation results are confronted to characterization data on several key figures of merit including Charge per Pulse (CPP) and Pulse Width, providing validation to the base model as proof of concept.

Building upon the results of the base single loop model, **Chapter III** goes in further detail on the distributed modelling approach that will take full advantage of the transient measurements carried out on the dedicated test structure that was developed. As modelling the transient response of the SPAD was the main drive of this thesis, the distributed approach described in this chapter serves as a major milestone in the direction of a precise and reliable transient SPAD model. In the first subchapter, the "KNACK" test chip and the corresponding measurement setup and method are detailed. First observations on characterization results are also made to emphasize the stochastic nature of the SPAD and thus also the need for a precise transient model. After that, the absorption depth modelling, focusing on the spatialized description of the model, is detailed. This comprises the explanation of the different areas inside the device, the translation of those regions to a multiple loop distributed model and also the details of the quenching criteria based on the loop gain of the device. Following this section, the model calibration and tuning is developed and functional additions to the model, such as a charge based quenching criterion and electric field degradation, are portrayed. Lastly, the parameter extraction and adjustment methods are illustrated to close the chapter and put forward the accuracy of the model in comparison with transient measurements conducted on the device.

Chapter IV brings this thesis to a close and starts with the different perspectives this distributed model can offer in the future, emphasizing on the ease of implementation of key phenomena observed in all the transient measurements on the device. From the delayed absorption of avalanche generated holes at the anode, to the statistical jitter modeling and finally the statistical amplitude modelling, this chapter is closed by a global conclusion, recounting and concluding all the different milestones of this thesis work and manuscript.

Chapter I Bibliography

1 Introduction

1.1 Integrated imagers in CMOS technology

Over the last decade, there has been a considerable development in pixel array imager applications using SPAD avalanche diodes embedded in CMOS technology. This technology combines pixel arrays, each consisting of a SPAD photodiode and a control and readout circuit, with an integrated data processing circuit in CMOS technology. This has led to the realization of multiple imagers for imaging and sensing applications in various fields, including biomedical, environmental, industrial, and automotive [1][2][3].

One of the main applications of SPAD imaging is Time-of-Flight (ToF) sensors. These sensors are used for their ranging and proximity detection capabilities in smartphones, gaming applications, and robotics. The ToF principle is based on the measurement of the time it takes for a light pulse to travel from a source to a target and back to a detector. By measuring the time-of-flight of the light pulse, the distance to the target can be calculated.

In recent years, there has been a significant increase in the use of SPAD-based ToF sensors in various applications. For example, in smartphones, ToF sensors are used for facial recognition and augmented reality applications. In gaming applications, they are used for motion tracking and gesture recognition. In robotics, they are used for obstacle detection and localization.

SPAD-based ToF sensors have several advantages over other ToF technologies. They offer high sensitivity, high temporal resolution, and low power consumption. Additionally, they can operate in a wide range of ambient light conditions, making them suitable for use in various applications.

The Time-of-Flight (ToF) principle is a method for measuring the distance to an object using light. The sensor combines an infrared (IR) emitter and a Single-Photon

Avalanche Diode (SPAD) based receiver. An IR laser is emitted, reflecting on the object that needs to be detected. It then reflects part of the emitted photons back to the photodetector. The delay between emission and reception is directly yielding the distance to the object. This principle is illustrated in ***Figure I-1*** [4].

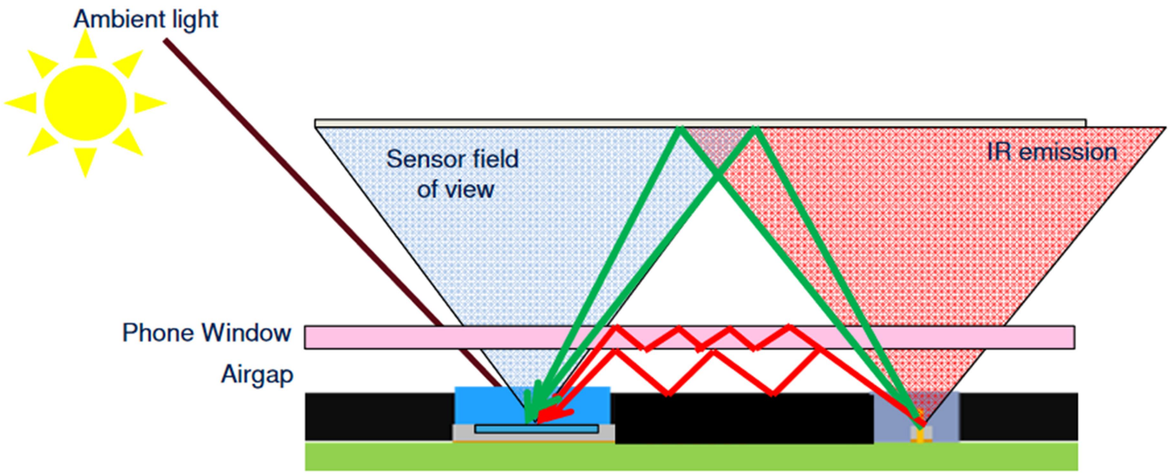

Figure I-1 ToF principle schematic, STMicroelectronics ToF sensor

Typical ToF sensor architectures, as shown in ***Figure I-2*** [5], are composed of a SPAD array with its front-end circuitry, including amplifiers and filters, and address decoded recharge triggering thanks to sensor control, Multiplexer (MUX), and Time-to-Digital Converter (TDC) circuitry. The SPAD array is responsible for detecting the reflected light, while the front-end circuitry is responsible for amplifying and filtering the signal. The MUX is used to select the appropriate SPAD element, while the TDC is used to measure the time-of-flight of the light pulse.

STMicroelectronics' technology used in this research combines a pixel matrix chip and an image processing chip in a 3D integration scheme. This technology enables the integration of the SPAD array, front-end circuitry, MUX, and TDC on a single chip, resulting in a compact and efficient ToF sensor. The pixel matrix chip is responsible for detecting the reflected light, while the image processing chip is responsible for processing the signal and generating the distance measurement.

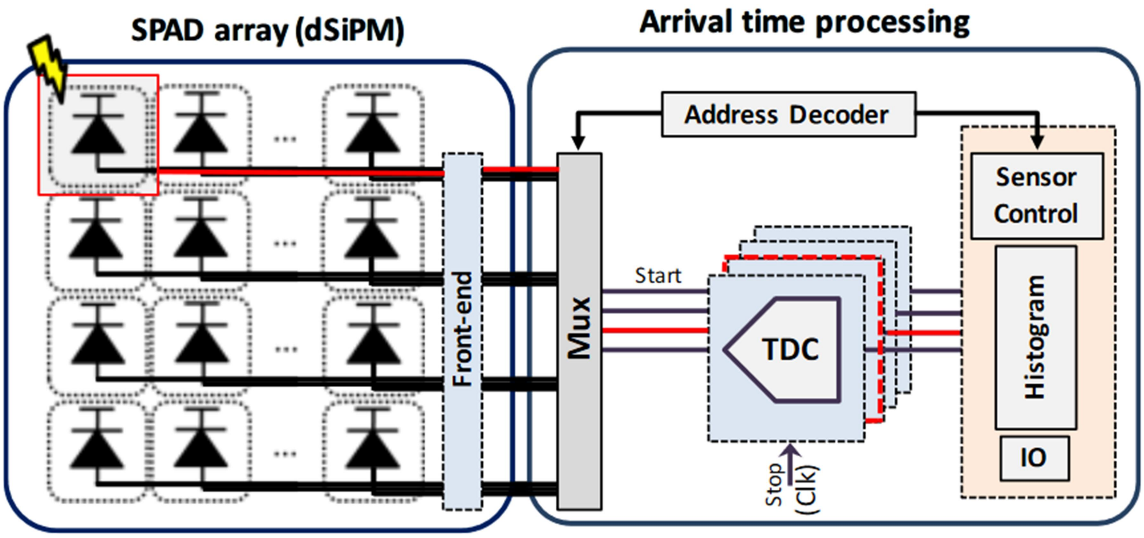

Figure I-2 Direct ToF TCSPC system block diagram with TDC sharing scheme: the assigned TDCs extract the timing of the detected photon arrivals

Overall, the ToF principle is a powerful method for measuring the distance to an object using light. Typical ToF sensor architectures are composed of a SPAD array with its front-end circuitry, MUX, and TDC circuitry. STMicroelectronics' technology used in this research combines a pixel matrix chip and an image processing chip in a 3D integration scheme, resulting in a compact and efficient ToF sensor.

1.2 Single-photon avalanche photodiodes: SPAD

The Single-Photon Avalanche Diode (SPAD) is a type of photodiode that operates in an atypical operating domain for most semiconductor devices. Unlike traditional photodiodes, a reverse bias is applied beyond the breakdown voltage of the junction, generating a metastable state in the presence of free carriers generated either by the absorption of an incident photon or by other mechanisms such as generation-recombination or tunnelling. This operating domain is called the Geiger Mode.

In Geiger Mode, the SPAD diode is highly sensitive to individual photons, making it an ideal device for applications such as single-photon detection and time-resolved photon counting. When a photon is absorbed by the SPAD diode, it generates a large number of electron-hole pairs, which are then amplified through an avalanche process, resulting in a detectable electrical signal.

Figure I-3 [4] provides a detailed SPAD characteristic with its different modes. The figure shows the typical behavior of a SPAD diode in different operating modes, including the linear mode, the breakdown mode, and the Geiger Mode. In the linear mode, the SPAD diode behaves like a traditional photodiode, generating a linear response to incident light. In the breakdown mode, the diode is operated at a voltage beyond the breakdown voltage, resulting in a large reverse current. In the Geiger Mode, the diode is operated at a voltage beyond the breakdown voltage, but with a quenching circuit to limit the duration of the avalanche process.

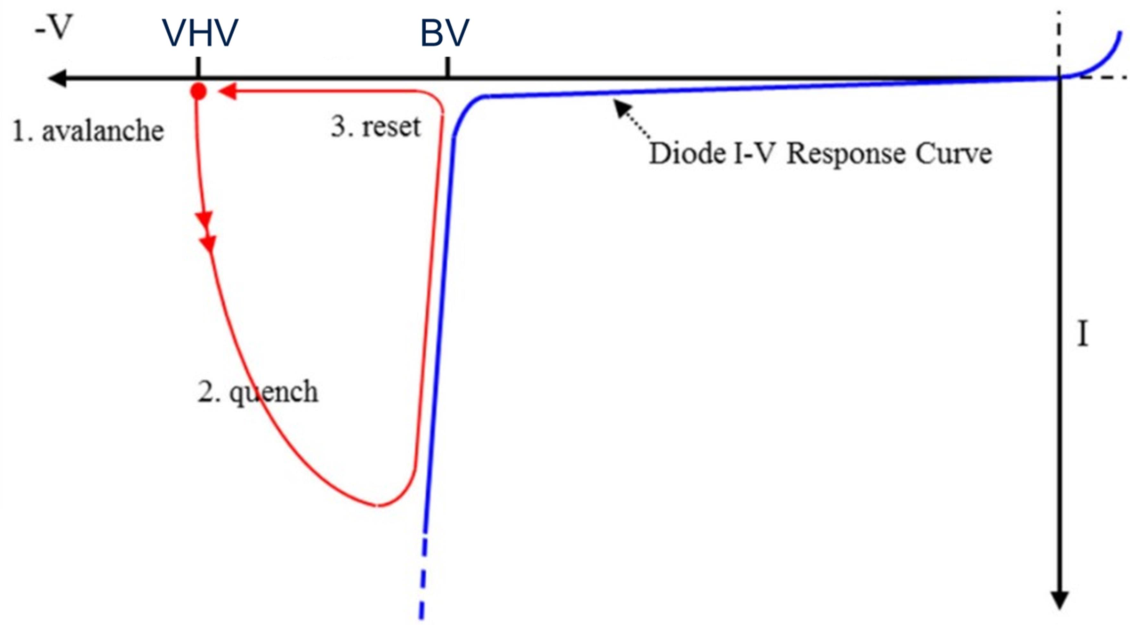

Figure I-3 SPAD reverse bias I-V characteristic with the different avalanche breakdown phases

The Geiger mode operation of the device can be broken down into three main steps, each of which plays a crucial role in the functioning of the device:

- The first step is the avalanche, which is the active part of the device. During this step, the single-photon avalanche diode (SPAD) conducts current and avalanche breakdown occurs when the biasing voltage exceeds the breakdown voltage of the diode. This results in an exponential increase in the number of carriers, which in turn leads to an increase in current. The carriers are multiplied through impact ionization, which is a process that occurs when high-energy electrons collide with atoms in the material, causing them to ionize and create additional electrons and holes.

- The second step of the Geiger mode operation is the quench, which is the first step in the recharge of the SPAD. The main mechanism behind this step is the extinction of the avalanche, which is generally caused by an increase in current through a quenching resistor. This increase in current effectively increases the voltage drop across the resistor, which in turn decreases the voltage applied to the SPAD. This quenching process allows the voltage across the SPAD to be reduced below its breakdown voltage, thus stopping the avalanche current.
- The third and final step of the Geiger mode operation is the reset, which follows the quenching step. This step is caused by the reset of the voltage across the SPAD to its initial biasing value VHV (High Voltage node), which is caused by the extinction of the diode. Following this, the SPAD is able to be retriggered again to detect incoming photons. This reset step is crucial in ensuring that the device is able to function properly and accurately detect incoming photons.

The process of photon absorption in a SPAD is a complex one, involving the creation of an electron-hole pair in just a few picoseconds. However, the polarization at the breakdown limit imposes an electric field that is favorable to the multiplication of electron-hole pairs via the phenomenon of ionization by impact. This leads to an exponential increase in the current, which becomes observable at the macroscopic scale for each absorbed photon, after a few tens to hundreds of picoseconds.

The amplification factor of the current depends on the applied electric field and can vary over several decades. This makes it necessary to limit the current through the use of control electronics. The purpose of these electronics is not only to avoid degradation of the SPAD, but also to turn off the avalanche as soon as the first light pulse is detected by the reading electronics. This is important to prevent the avalanche from continuing and potentially causing damage to the device.

In addition to turning off the avalanche, the control electronics must also reset the polarization of the SPAD before the next light pulse. This is a crucial step that must be completed in just a few nanoseconds to ensure that the device is able to accurately detect the next incoming photon. The combination of these processes allows the SPAD to function as a highly sensitive detector of single photons, with the ability to detect even the weakest of signals.

The impact ionization regime is a crucial factor in increasing the sensitivity of the SPAD as a photon detector. However, it is the control electronics that allows for the precise control of the avalanche transient current, including the amplification, extinction, and reset of the device. The speed of the trigger combined with the response time of the control electronics determines the temporal resolution of each imager pixel. This is a critical factor in ensuring that the device is able to accurately detect and record the arrival time of each photon.

It is important to note that the physical processes of photon absorption and impact ionization are stochastic in nature. As a result, the temporal response of the SPAD to a photon absorption not only has a deterministic component that depends on structural parameters, such as the choice of material and the size of the absorption area, but is also subject to a fluctuation known as jitter. This jitter must be taken into account in the definition of the minimum temporal resolution of each pixel.

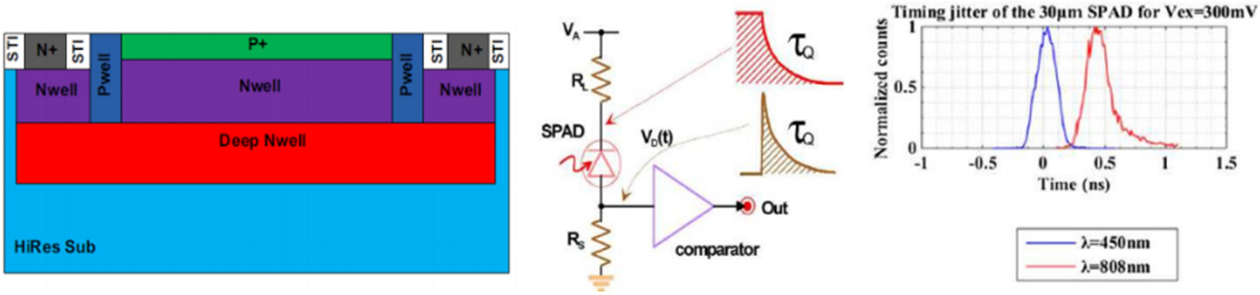

Figure I-4 Example of SPAD cross section (left), Impulse response principle of passively quenched SPAD by two avalanche current limiting resistors (middle), Example of timing response fluctuation of SPAD at two different

Figure I-4 [1] illustrates the importance of taking jitter into account when defining the minimum temporal resolution of each pixel. These figures demonstrate the effect of jitter on the temporal response of the SPAD, showing that even small fluctuations can have a significant impact on the accuracy of the device. By carefully considering the impact of jitter and other factors on the temporal resolution of the SPAD, it is possible to design highly sensitive and accurate photon detectors that are capable of detecting even the weakest of signals.

Figure I-5 [2] provides valuable insight to two key characteristics of SPAD diodes. First, the DC current I(V) characteristic which is the basis of the design in order to select the fitted components around the device. Then, as mentioned in previous paragraphs, SPAD's are sensitive to a large spectrum of wavelengths, but at various degrees. This is when the Photon Detection Probability (PDP) comes into the picture, which for silicon has its maximum at the lower spectrum of visible light in the blue range (400-500nm in the case of Figure I-5).

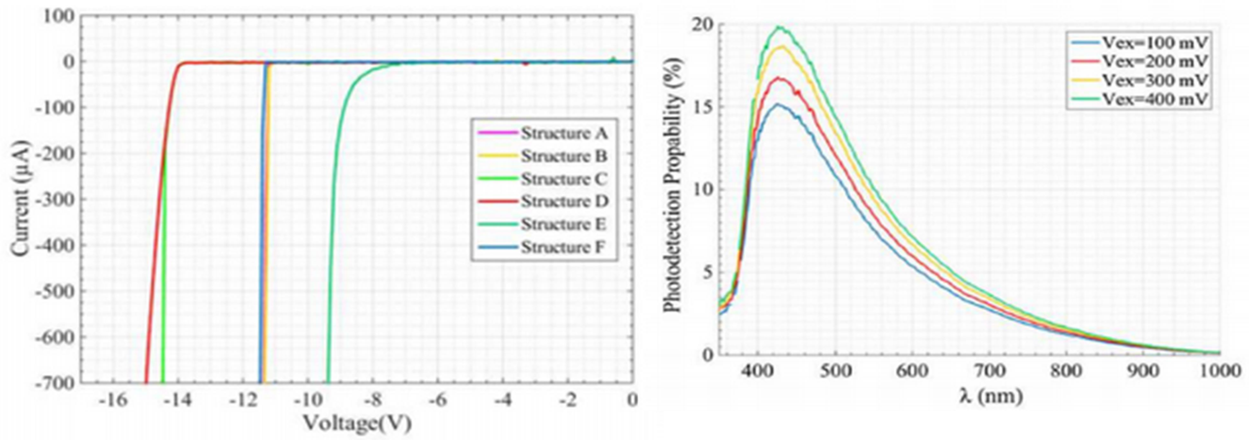

Figure I-5 DC caracteristics of different SPAD structures (left), Photon detection efficiency PDP for a same SPAD structure at different reverse biasing voltages [$V_{ex} = V_{bias} - V_{br}$] (right)

2 The stakes of the thesis

2.1 Objective

The primary objective of this modeling project is to address the needs of designers who are working on imaging circuits based on SPADs. The impulse response of a given pixel is determined by a range of factors, including the static (DC), dynamic (AC), and statistical temporal behavior of the SPAD, as well as the response time of the control and readout electronics. All of these effects must be accurately modeled and implemented in the circuit design environment to achieve the following goals:

- Optimization of the implementation of individual SPADs: This involves optimizing the merit figures of the SPAD, including photon detection probability (PDP), dark count rate (DCR), single-photon timing resolution (SPTR), jitter, dead time (DT), and afterpulsing probability (AP). Additionally, the optimization of pixel electronics, such as fill factor (FF), photon detection efficiency (PDE), input capacity, bandwidth, and other parameters, is also critical.
- Evaluation of the performance of a pixel matrix: Once the individual SPADs and pixel electronics have been optimized, it is necessary to evaluate the performance of the pixel matrix as a whole. This involves assessing the sensitivity of the pixel matrix and identifying any limitations that may be present, such as optical coupling issues.

By accurately modeling and implementing all of these factors in the circuit design environment, it is possible to develop highly sensitive and accurate imaging circuits based on SPAD diodes. This is critical in a range of applications, including medical imaging, remote sensing, and scientific research.

The statistical modeling approach that will be used in this project is a complex and multi-faceted process that involves a range of techniques and methods. The first step in the process will be to conduct a thorough analysis of the physical effects that are involved in the operation of SPADs. This will involve a review of the relevant literature, as well as the use of advanced simulation tools such as TCAD to model the behavior of SPADs under different conditions.

Once the physical effects have been analyzed and modeled, the next step will be to develop an analytical model that can be applied in a design environment. This will involve using the results of the simulation to derive a mathematical model that accurately describes the behavior of SPADs under different conditions. The model will be validated using experimental data, which will be obtained using a range of ultra-fast optoelectronic characterization techniques that are based on the use of femtosecond lasers.

The experimental validation will be based on a range of techniques, including time-resolved photoluminescence, time-correlated single photon counting, and other advanced methods. These techniques have been developed and implemented in the laboratory as part of the Nano2022 project, which is funded by the French government and aims to consolidate the national electronic components industry.

To ensure that the modeling approach is applicable in an environment of imaging circuit design, two levels of modeling will be considered. The first level will focus on modeling an elementary SPAD, which is a priority for this project. The second level will involve modeling a pixel matrix, which is a more complex task that requires the integration of multiple SPADs into a single imaging system.

The Verilog-A behavioral language will be used for modeling, as it is compatible with circuit simulators that are commonly used in analog design. This will allow the analytical model to be integrated into a range of imaging circuit designs, including those used in medical imaging, remote sensing, and scientific research.

Overall, the statistical modeling approach that will be used in this project is a critical step in the development of highly accurate and reliable imaging circuits based on SPAD diodes. By consolidating the national electronic components industry, this project will also help to promote innovation and economic growth in France.

2.2 Characterization

The development of optoelectronic characterization methods with high temporal resolution is a critical challenge in this project. The behavior of SPADs is complex and dynamic, and it is important to accurately capture the behavior of the device in order to optimize its performance as a photon detector. One of the key challenges in developing these characterization methods is ensuring that the reading electronics do not significantly disturb the observation of the SPAD transient response in nominal value and dispersion. This requires careful design and optimization of the reading electronics, as well as the development of advanced experimental techniques that can accurately capture the behavior of the SPADs.

To address this challenge, the project will focus on developing optoelectronic characterization methods with high temporal resolution that will allow for the experimental observation of the dynamics of avalanche triggering in SPADs. The avalanche current in SPADs reaches its nominal value at the scale of a few hundred picoseconds, making it important to develop characterization methods that can accurately capture this behavior.

The study will evaluate the limits of this new metrology and critically analyze the results of completely depleted diode-type 40nm SPAD devices. This analysis is required to interpret the temporal waveforms accessible via the SPAD reading circuit and to determine the temporal behavior of the SPAD itself, which is not directly observable but remains necessary for the development and experimental validation of the model.

The development of these optoelectronic characterization methods is a critical step in the optimization of SPAD performance. By accurately characterizing the behavior of SPADs, it is possible to develop highly accurate and reliable models that can be used to optimize the performance of imaging circuits based on SPAD diodes.

The project will also help to promote innovation and economic growth in France by consolidating the national electronic components industry. This will involve collaboration with industry partners to develop and commercialize new technologies based on the results of the project. In addition, the project will provide training and education opportunities for students and researchers, helping to build the next generation of experts in the field of optoelectronics and photonics. Overall, this project represents a significant step forward in the development of advanced imaging technologies based on SPAD diodes.

2.3 Compact Modeling

The current state of diode modeling in design tools is inadequate in capturing the full range of dynamic components and stochastic aspects of the avalanche regime that are specific to the operation of Single Photon Avalanche Diodes (SPADs). The avalanche process in SPADs is characterized by a rapid multiplication of charge carriers, which is triggered by the impact ionization of a single carrier. This process is often modeled using a simple empirical approach that assumes quasi-static equilibrium in reverse polarization mode. However, this approach fails to account for the carrier generation mechanisms, such as direct or indirect tunnel current, that can trigger the avalanche in a dark situation. Moreover, the dynamics and temporal variability of these mechanisms are not considered, which can lead to inaccuracies in the model.

In the case of SPADs, the absorption of photons must also be taken into account, including its dynamics, temporal variability, and the dimensions of the absorption zone, which can contribute to Jitter. The absorption of photons in SPADs can be modeled using a variety of approaches, including the use of Monte Carlo simulations or analytical models based on the diffusion equation. However, these models often make simplifying assumptions that can limit their accuracy, such as assuming a uniform absorption profile or neglecting the effects of surface recombination.

While some existing models do incorporate different figures of merit and physical effects, such as the dark count rate and photon detection efficiency, there is still a need for more comprehensive and accurate models that can fully capture the complexities of SPAD operation. Such models could help improve the design and optimization of SPAD-based systems.

Several analytical and CAD models have been developed to analyze the experimental results of SPADs and estimate their performance metrics. One such model is the Analytical Avalanche Model developed by L. Pancheri in 2014 [3], which is used to analyze the avalanche process in SPADs and estimate their performance metrics. Another model is the Avalanche Establishment Time Analytical Model developed by F. Sun in 2019 [4], which is used to estimate the Jitter in SPADs based on the time it takes for the avalanche process to establish.

In addition to analytical models, several deterministic CAD models have been developed to simulate the behavior of individual or matrix SPADs. One such model is the Deterministic CAD Model based on equivalent circuitry developed by A. Dalla Mora in 2007 [1], which is used to simulate the behavior of individual SPADs. Another model is the High-Level Model of the Statistical Distributions of SPAD Merit Figures developed by R. Kazma in 2015 [5] or also the model developed by F.Villa in 2015 [6], which are used to simulate the statistical distributions of SPAD performance metrics.

Several static behavioral CAD models have also been developed to simulate the behavior of the SPAD avalanche. One such model is the Static Behavioral CAD Model (DC) of the SPAD Avalanche developed by A. Dieguez in 2015 [7], which is based on the Verilog-A language and also the model developed by L. Zheng in 2017 [8]. Another model is the Stochastic Avalanche CAD Model for Individual SPADs developed by G. Giustolisi in 2011-2012 [9], which is also based on the Verilog-A language.

Finally, several stochastic avalanche CAD models have been developed to simulate the behavior of individual SPADs and estimate their performance metrics. One such model is again the Stochastic Avalanche CAD Model for Individual SPADs developed by G. Giustolisi in 2011-2012 [9], which is based on the Verilog-A language. Another model is the Stochastic Avalanche CAD Model for Individual SPADs Extended to

Afterpulse Triggering developed by Z. Cheng in 2016 [10], which is also based on the Verilog-A language. These models are useful for simulating the behavior of SPADs under different operating conditions and optimizing their performance for specific applications.

3 SPAD figures of merit

The figures of merit of a SPADs [2] are critical in characterizing its temporal response and are governed by stochastic phenomena. These performance metrics include PDP, DCR, SPTR, Jitter, DT, and AP. The design of the control electronics for each pixel in a SPAD matrix requires the consideration of these stochastic phenomena in the SPAD model, in addition to the precise electrical behavior of each component of the pixel.

To optimize the performance of the pixel and the pixel matrices, a two-level modeling approach is required. The first level involves a fine modeling of the mechanisms involved in the SPAD avalanche, including their transient dynamics and the associated statistics. This modeling approach requires a detailed understanding of the physical processes that govern the SPAD avalanche, such as carrier generation, impact ionization, and multiplication. It also involves the development of accurate models for the various stochastic phenomena that affect the SPAD performance metrics, such as the statistical distribution of the avalanche gain and the probability of afterpulsing.

The second level of modeling involves taking into account the optical coupling between SPADs, which is also of a stochastic nature. This modeling approach requires a detailed understanding of the optical properties of the SPAD matrix, including the absorption and scattering of photons within the pixel and the surrounding materials. It also involves the development of accurate models for the optical crosstalk between adjacent pixels, which can significantly impact the performance of the SPAD matrix.

Overall, the optimization of SPAD matrices requires a comprehensive modeling approach that takes into account both the electrical and optical properties of the device, as well as the stochastic phenomena that govern its performance metrics. Such models can help guide the design of control electronics for SPAD matrices and optimize their performance.

3.1 Dark Count Rate DCR

The DCR is a phenomenon that occurs in Single Photon Avalanche Diodes (SPADs) when an avalanche is triggered without the arrival of a photon that would have generated an electron-hole pair. This can be caused by various physical phenomena that occur in the device, including the spontaneous thermal generation of carriers due to the Schockley Read Hall (SRH) effect or the Trap Assisted Tunneling (TAT) effect. As a result, DCR is often referred to as SRH-TAT DCR.

In addition to SRH-TAT DCR, Afterpulsing is another source of dark counts in SPADs. Afterpulsing occurs when a carrier is trapped in a silicon defect state during the avalanche process and is later released, triggering another avalanche. This can result in false detections and reduce the overall performance and accuracy of the SPAD.

Figure I-6 provides a summary of all the contributions to DCR, including both thermally impacted and thermally decorrelated contributions. The thermally impacted contributions are shown in red and include the SRH-TAT DCR as well as other sources of dark counts that are influenced by temperature, such as surface leakage and tunneling. The thermally decorrelated contribution is shown in green and includes dark counts that are not influenced by temperature, such as those caused by cosmic rays or electronic noise.

Accurately characterizing and modeling DCR is critical for optimizing the performance of SPADs and SPAD-based systems. This requires a detailed understanding of the physical mechanisms that contribute to DCR, as well as accurate models for the various sources of dark counts.

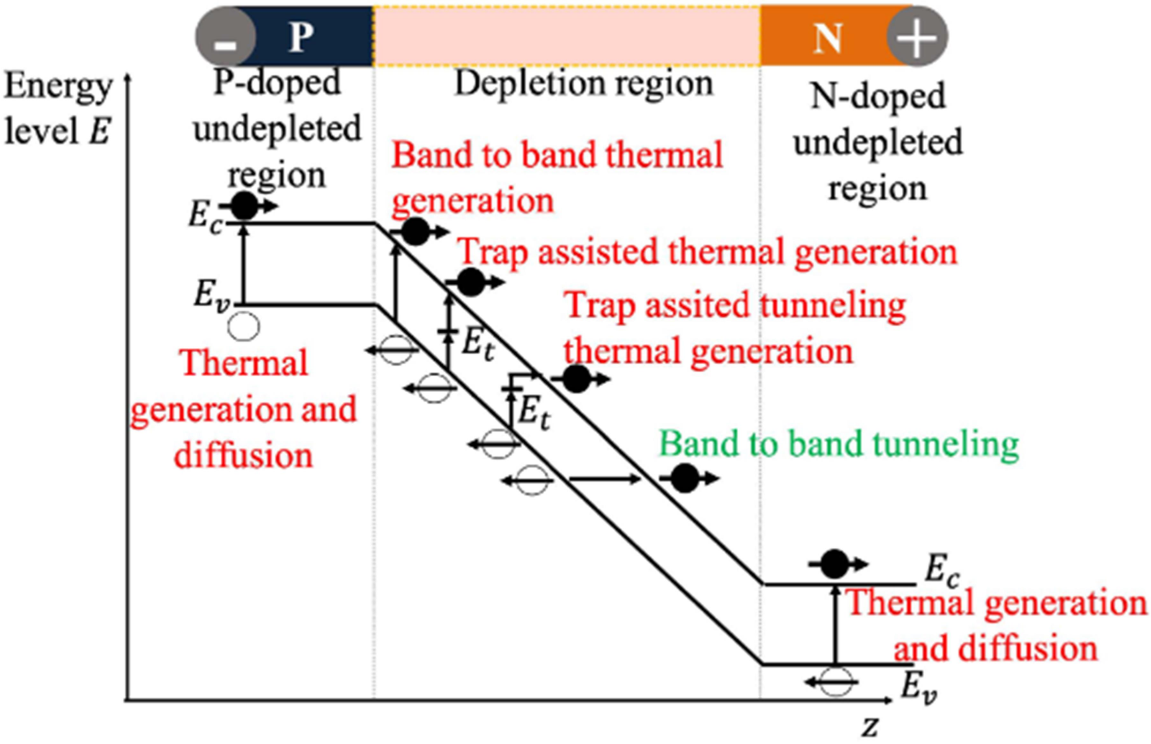

Figure I-6 DCR contributions illustration for a reverse biased p-n junction, setting aside the afterpulsing contribution

3.1.1 Schockley Read Hall (SRH) spontaneous thermal generation

In an equilibrium semiconductor, the electron density n multiplied with the hole density p is assumed to be constant. These densities follow the follow ***Eq. I-1***:

$$n \cdot p = {n_i}^2 \qquad \textbf{\textit{Eq. I-1}}$$

Where n_i is the intrinsic carrier concentration of the material.

When this balance is disturbed, the generation and recombination phenomena try to restore the balance. The generation and recombination rate is caused by traps whose energy level is situated in the band gap of the semiconductor. These traps can be created by impurities, defects, or other sources of lattice distortion.

The generation and recombination rate in a non-equilibrium semiconductor is governed by the Schockley-Read-Hall (SRH) equation, which takes into account the effects of traps on the carrier lifetime. The SRH generation rate is given by $G_{SRH-TAT}$ and is defined by ***Eq. I-2*** [10]:

$$G_{SRH-TAT} = \frac{np - {n_{i_{\text{eff}}}}^2}{\tau_h(n + n_1) + \tau_e(p + p_1)} \qquad \textbf{\textit{Eq. I-2}}$$

Where $n_{i_{\text{eff}}}$ is the effective intrinsic carrier concentration, τ_h and τ_e are the hole and electron lifetimes, and n_1 and p_1 are the excess carrier concentrations due to doping of the silicon.

The rate of generation and recombination depends on several parameters, including the energy levels of the generation/recombination centers (traps), the intrinsic concentration of the material n_i, the electron and hole concentration, defined by the doping of the material N_{d} and N_{a}, the temperature T, and the lifetime of minority carriers τ_0. These parameters can significantly impact the performance of semiconductor devices and must be carefully considered in their design and optimization.

If the electric field becomes very large, the generation is amplified by a phenomenon called the TAT effect. This effect occurs when an electron tunnels from a trap state to the conduction band, creating a hole in the valence band. The TAT effect can significantly increase the rate of carrier generation and recombination and must be carefully considered in the design of high-field devices such as avalanche photodiodes and tunnel diodes.

Overall, understanding the generation and recombination phenomena in semiconductors is critical for optimizing the performance of semiconductor devices and improving their reliability.

3.1.2 Band to Band Tunneling BTBT

When the electric field in a semiconductor is very strong, electrons can pass from the valence band to the conduction band by tunneling, a phenomenon known as Band-to-

Band Tunneling (BTBT). This process occurs when the electric field is strong enough to overcome the energy barrier between the valence and conduction bands, allowing electrons to tunnel through the barrier. The BTBT process generates an electron-hole

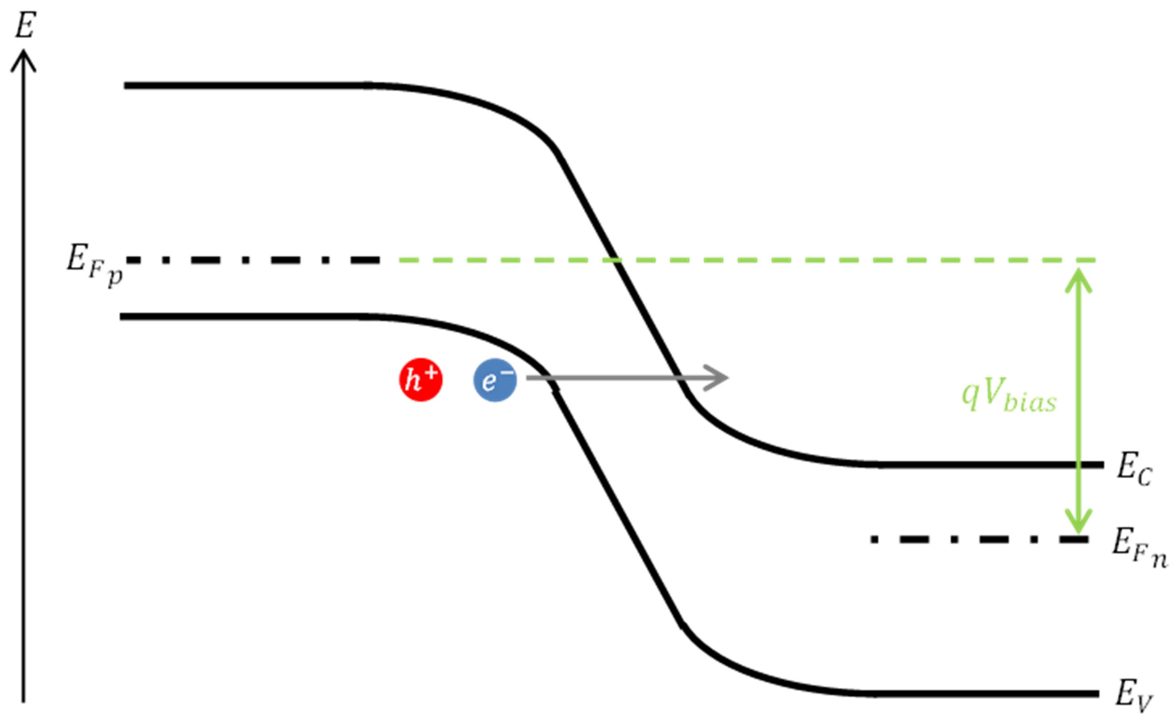

Figure I-7 Band diagram of P/N junction representing Band-to-Band Tunneling

pair, which can eventually trigger an undesired avalanche and participate in the dark count rate. ***Figure I-7*** shows the corresponding band diagram to illustrate the BTBT mechanism.

Dark counts in SPADs can be initiated by carriers generated by the BTBT effect. These carriers can be generated in the depletion region of the SPAD, where the electric field is strongest, and can contribute to the overall dark count rate of the device. The BTBT effect is particularly problematic in SPADs, where the detection of single photons requires a high level of sensitivity and low levels of noise.

The Band-to-Band Tunneling effect generated Dark Counts fix the minimal DCR threshold in SPADs. The DCR is a critical performance metric for SPADs, as it determines the minimum level of background noise that the device can tolerate without compromising its performance. The BTBT effect can significantly impact the DCR of SPADs and must be carefully considered in their design and optimization.

3.2 Afterpulsing

Traps created by impurities or defects in a semiconductor can become generation or recombination centers, which can significantly impact the performance of semiconductor devices such as SPADs. Avalanche breakdown that occurs within the space charge region (SCR) of a SPAD generates a current peak, increasing the likelihood that a trap will be filled by a carrier generated from the avalanche. This trap can in turn release this carrier and thus generate a second avalanche that can be detected as a photon, which distorts the count and thus participates in the dark count rate.

The afterpulsing effect in SPADs is closely related to the avalanche extinction or hold-off time, which is the time required for the device to recover from an avalanche and return to its quiescent state. The afterpulsing effect occurs when a carrier is trapped in a defect state during the avalanche process and is later released, triggering another avalanche. This can result in false detections and reduce the overall performance of the SPAD. The afterpulsing effect increases with the avalanche extinction/hold-off time, mainly depending on the quenching circuit and the current I, which depends on the applied biasing voltage V_{bias}. Reverse biasing increases the width of the SCR and thus the number of traps that can capture a carrier, which can significantly impact the afterpulsing effect.

The dependencies of afterpulsing on the avalanche extinction/hold-off time, quenching circuit, current, and applied biasing voltage are shown in ***Figure I-8*** [11]. Accurately characterizing and modeling the afterpulsing effect is critical for optimizing the performance of SPADs and improving their reliability. This requires a detailed understanding of the physical mechanisms that contribute to afterpulsing, as well as accurate models for the various sources of noise and distortion in the device.

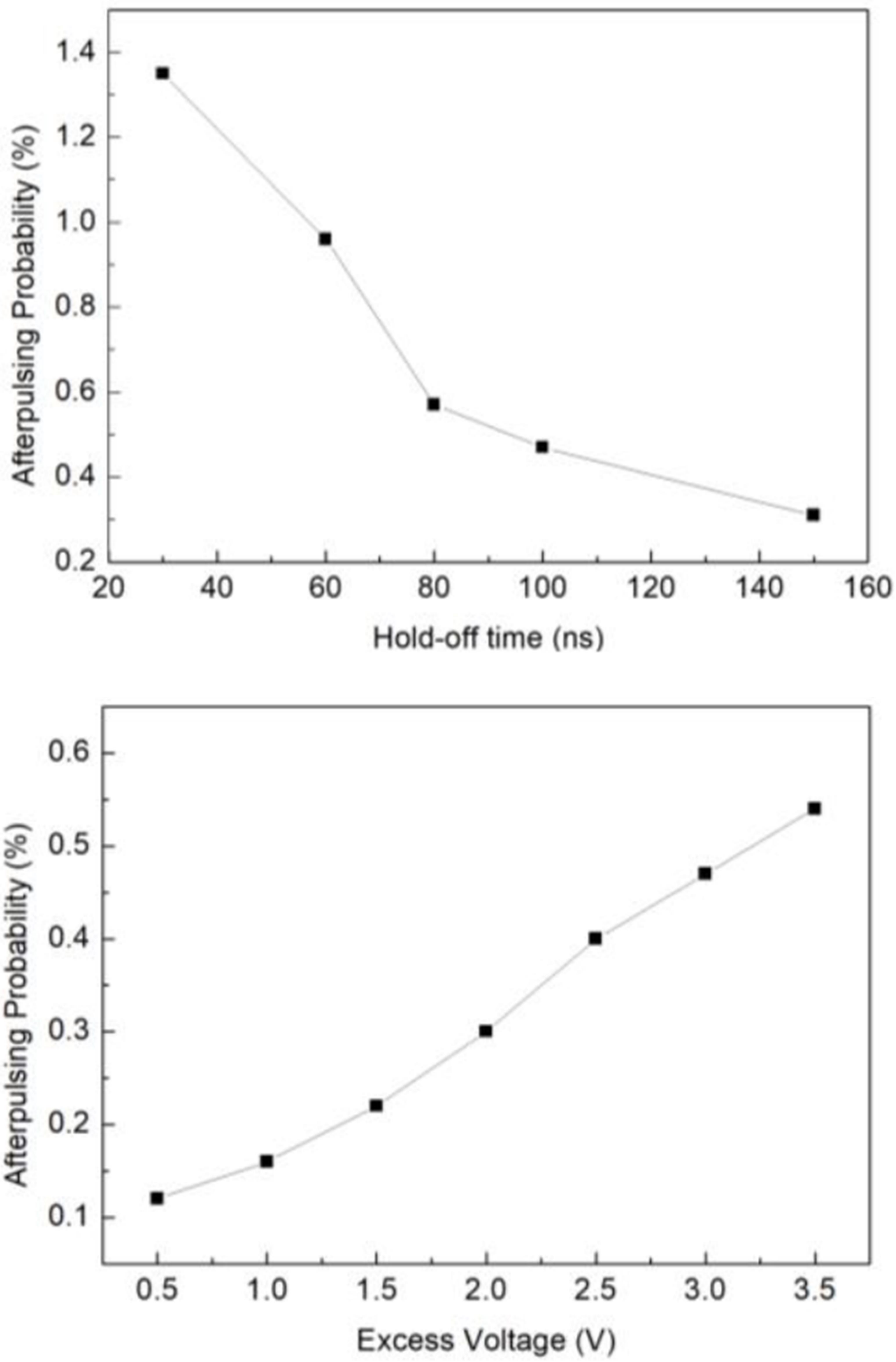

Figure I-8 Example of Afterpulsing probability as a function of hold-off time (top) and excess voltage (bottom)

4.3 Photon Detection Probability

The absorption probability of a photon depends on several factors, including the wavelength of the photon, the material used in the photodetector, and the dimensions of the absorption region. For example, the absorption probability of a photon in a silicon-based photodetector is highest for wavelengths around 800 nm, which corresponds to the peak sensitivity of silicon. However, other materials such as SiGe or InGaAs can have different absorption characteristics that are better suited for different wavelengths. ***Figure I-9*** [3] perfectly illustrates the PDP as a function of wavelength in silicon devices.

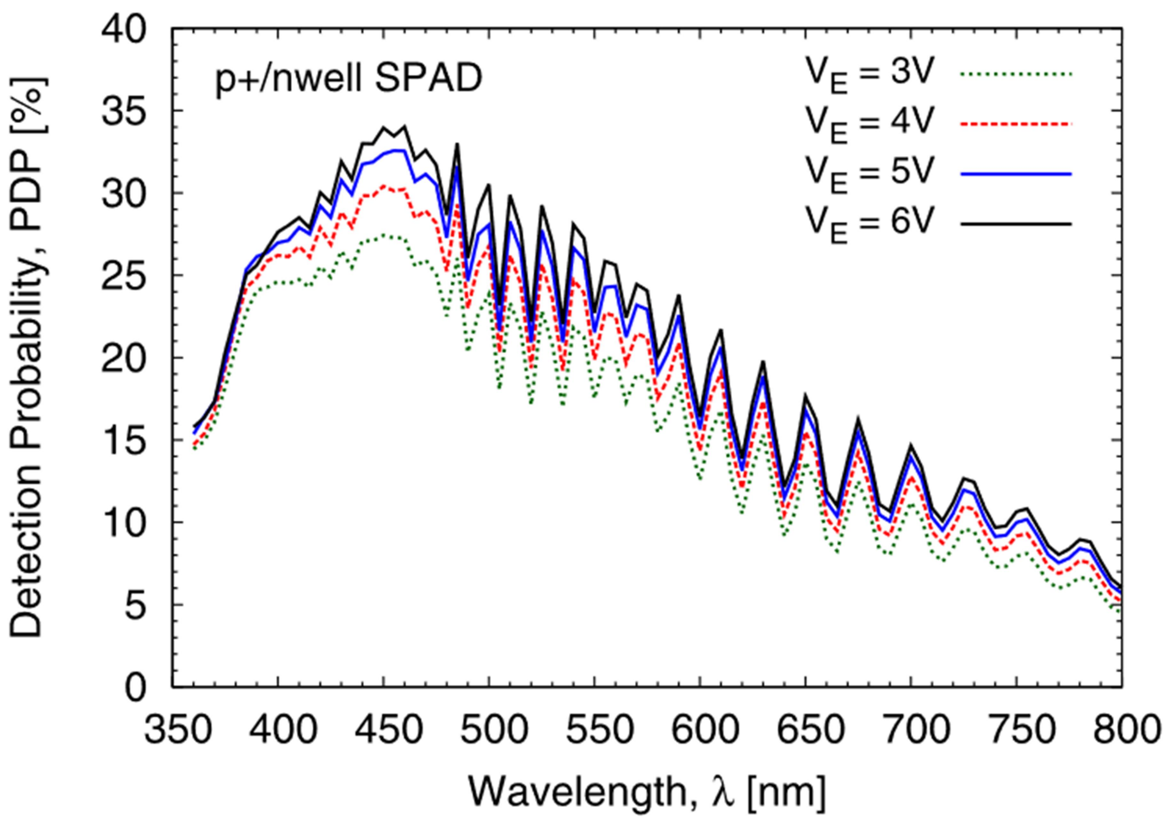

Figure I-9 PDP as a function of wavelength at different excess voltages [3]

The architecture of the photodetector also plays a role in the absorption probability and PDP. For example, SPADs are designed to operate in the Geiger mode, where a single photon can trigger an avalanche of electron-hole pairs that produces a detectable signal. This architecture allows SPADs to achieve high PDP values, typically above 20-30%. However, other types of photodetectors, such as photomultiplier tubes or avalanche photodiodes, may have different architectures that affect their PDP. Even in the domain of SPAD architectures, the PDP can vary depending on the chosen architecture of the device (reach-through, PIN, etc.).

The fill factor (FF) is a parameter that characterizes the active area of the photodetector relative to the total area of the pixel that includes the control electronics. In other words, it represents the fraction of the pixel area that is actually sensitive to light. A higher fill factor means that more of the pixel area is devoted to the active region, which can increase the PDP and the overall efficiency of the photodetector.

Finally, the photon detection efficiency (PDE) is a measure of the probability that a photon will be detected by the photodetector, taking into account all the losses and inefficiencies in the detection process. This includes factors such as the absorption probability, the efficiency of the avalanche process, the quantum efficiency of the detector, and any losses due to reflection or transmission at the interfaces. The PDE is a critical parameter for many applications that require high sensitivity and low noise, such as quantum communication or fluorescence microscopy. By optimizing the design of the photodetector to maximize the PDE, researchers can improve the performance of these applications and enable new ones.

4.4 Jitter

The temporal resolution of a photodetector is a critical parameter that determines its ability to accurately measure the arrival time of a photon and is closely linked to what is called "jitter". However, the temporal resolution is limited by various stochastic variations that occur during the detection process. These variations are specific to the physical phenomena involved in the detection process, such as photon absorption, avalanche amplification and extinction, and current pulse reading.

One of the main sources of stochastic variation is the impact ionization process [12], which is responsible for triggering the avalanche amplification. The probability of impact ionization depends on the external electric field applied to the photodetector, which can vary during the amplification and extinction transient. As a result, the pulse detected by the control and readout electronics can exhibit a jitter, which is a random variation in the pulse arrival time. The magnitude of the jitter depends on various factors, such as the external electric field, the material properties of the photodetector, and the design of the control and readout electronics.

Another factor that affects the temporal resolution is the diffusion of carriers in the photodetector. When a photon is absorbed by the material, it creates an electron-hole pair that can diffuse through the material before reaching the region where the avalanche amplification occurs. This diffusion process can introduce a delay between the arrival of the photon and the detection of the avalanche by the processing circuit. The magnitude of this delay depends on various factors, such as the material properties of the photodetector, the dimensions of the active region, and the intensity of the incident photon.

To quantify the temporal resolution of a photodetector, researchers often use the distribution of counts over time, which represents the probability of detecting a photon at a certain time after its incidence. This distribution can be affected by various factors, such as the jitter, the diffusion delay, and the noise characteristics of the detector. ***Figure I-10*** [4] shows an example of the distribution of counts over time for a standard jitter SPAD, where the normalized distribution exhibits a characteristic shape that reflects the temporal resolution of the detector.

To improve the temporal resolution of photodetectors, researchers have developed various techniques such as time-correlated single photon counting (TCSPC) [13], which allows for precise measurement of the arrival time of photons with sub-nanosecond resolution. Other techniques such as gating or time-stamping [14] can also be used to reduce the impact of jitter and diffusion delays on the temporal resolution.

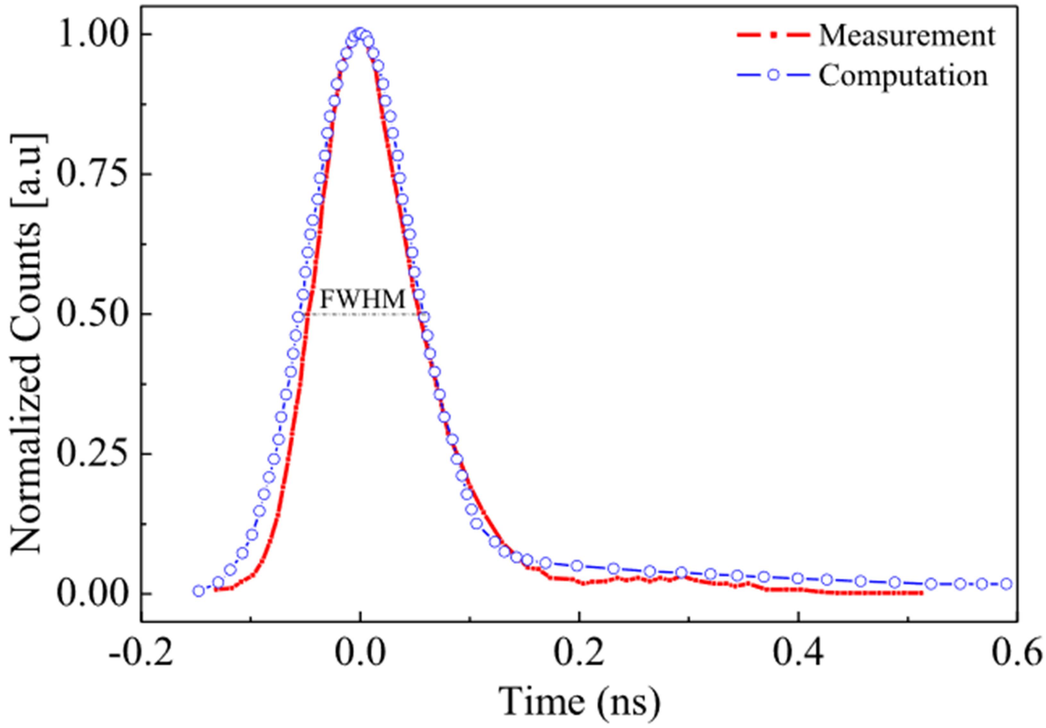

Figure I-10 Normalized timing jitter curve representing the counts over time [4]

4.5 Dead Time DT

Dead Time (DT) is a critical parameter that characterizes the ability of a SPAD to detect multiple photons in a short time interval. When a photon is detected by the SPAD, it triggers an avalanche of electron-hole pairs that produces a detectable signal. However, this avalanche process also creates a large number of moving charges that need to be evacuated before the SPAD can detect another photon, which can be seen in ***Figure I-11***. The time required for this evacuation process and for the SPAD to become sensitive again is known as the Dead Time.

It is also a crucial parameter that limits the maximum count rate of the SPAD. If the count rate exceeds the inverse of the Dead Time, the SPAD will miss some of the photons, leading to a reduction in the overall sensitivity and accuracy of the detector. Therefore, it is essential to minimize the Dead Time and optimize the design of the SPAD and its control electronics to achieve high count rates and low dead times.

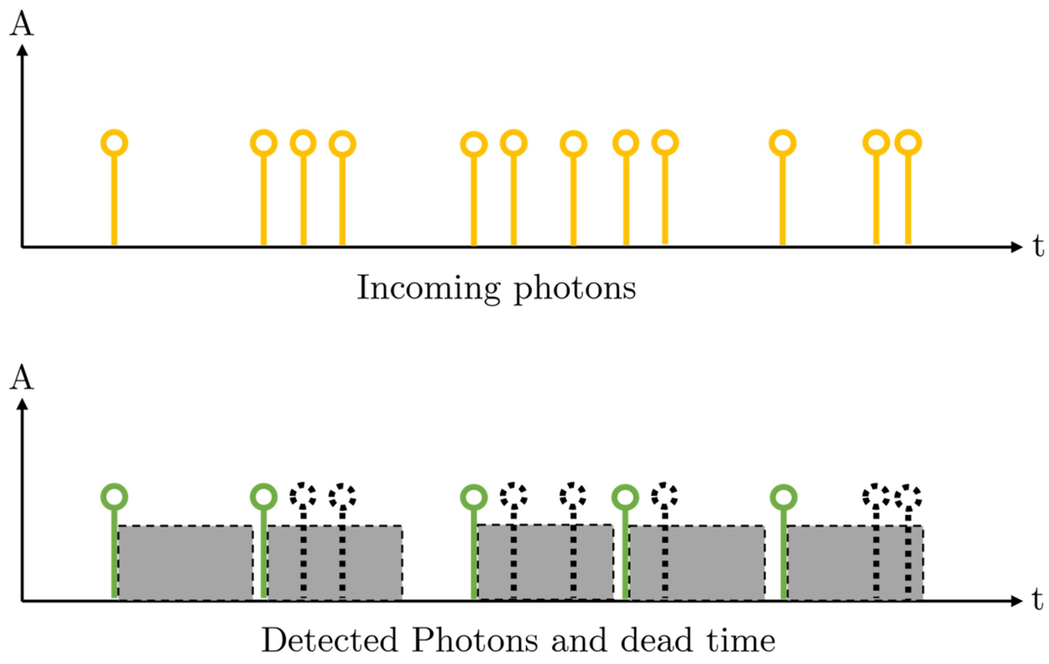

Figure I-11 Visual representation of dead time.

Dead Time depends on various factors, such as the material properties of the SPAD, the dimensions of the active region, and the design of the control electronics. For example, the Dead Time can be affected by the capacitance of the SPAD, which determines the amount of charge that needs to be evacuated before the SPAD can detect another photon. Dead Time can also be influenced by the design of the readout electronics, such as the timing and synchronization of the control signals.

To measure the Dead Time of a SPAD, researchers often use a pulsed laser or a light source with a known repetition rate [15]. By varying the repetition rate and measuring the count rate of the SPAD, we can determine the Dead Time and optimize the design of the detector and its control electronics.

To improve the Dead Time of SPADs, researchers have developed various techniques such as gating or time-stamping [14], which allow for precise control of the detection window and reduction of the Dead Time. Other techniques such as active quenching or fast reset can also be used to speed up the evacuation of moving charges and reduce the Dead Time [16].

4.6 Crosstalk

The avalanche process of a SPAD generates a large number of hot carriers that can interact with the material and produce additional photons by decelerating the carriers. This process, known as carrier multiplication or impact ionization, can lead to the emission of secondary photons that can trigger the avalanche of neighboring unlit SPADs. This phenomenon is known as optical crosstalk and can occur in arrays of SPADs where the active regions are in close proximity to each other.

Optical crosstalk can be a significant issue in SPAD arrays, as it can lead to false detections and reduce the overall sensitivity and accuracy of the detector. To minimize optical crosstalk, researchers have developed various techniques such as lateral isolation or optical filters that can block the transmission of secondary photons between neighboring SPADs. These techniques can improve the performance of SPAD arrays and enable new applications such as time-resolved imaging or fluorescence correlation spectroscopy.

In addition to optical crosstalk, SPADs can also be affected by electrical crosstalk [17], which occurs when a photon generated by a neighboring SPAD creates a carrier beyond the active zone of the detector. This carrier can diffuse to the surface of another SPAD and trigger an avalanche that is not photogenerated. Electrical crosstalk can be a significant issue in SPAD arrays, especially when the active regions are closely spaced or when the control electronics are not properly designed. To reduce electrical crosstalk, researchers have developed various techniques such as guard rings or floating wells that can isolate the active regions and prevent the diffusion of carriers between neighboring SPADs.

Figure I-12 depicts the mechanisms of optical and electrical crosstalk in SPAD arrays. The figure shows how the emission of secondary photons by one SPAD can trigger the avalanche of neighboring SPADs through optical coupling, and how the diffusion of carriers between SPADs can lead to false detections and reduce the overall performance of the detector.

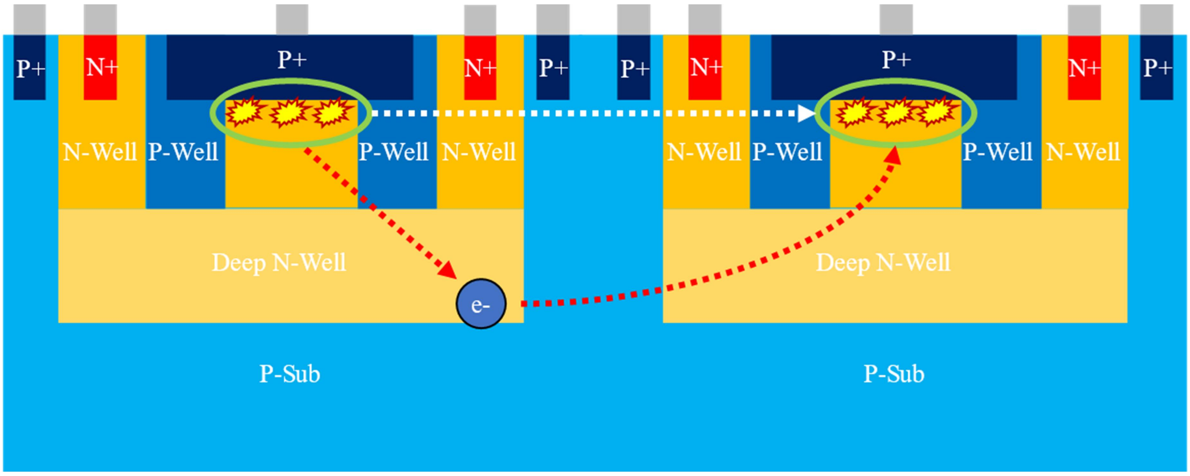

Figure I-12 Diagram of optical (white dashed arrow) and electrical (red dashed arrow) crosstalk mechanism

4 The avalanche phenomenon

The avalanche phenomenon is a critical mechanism that underlies the operation of various types of photodetectors, including Avalanche Photodiodes (APDs) and SPADs. When a photon is absorbed by the material of the detector, it creates an electron-hole pair that can be accelerated by the electric field present in the device. In the high field zone, also known as the multiplication region or the space charge region (SCR), the electric field is strong enough to accelerate the carriers to high energies, which can lead to the emission of additional carriers through impact ionization.

Impact ionization is a critical process that enables the multiplication of carriers and the amplification of the signal. When a high-energy carrier collides with an atom in the material, it can transfer enough energy to ionize the atom and create additional electron-hole pairs. These secondary carriers can then be accelerated by the electric field and create even more carriers through impact ionization, leading to an avalanche of carriers that produces a detectable signal.

The avalanche process is a highly nonlinear and stochastic phenomenon that can be affected by various factors, such as the material properties of the detector, the dimensions of the active region, and the intensity of the incident photon. For example, the avalanche process can be influenced by the electric field distribution in the SCR, which can affect the probability of impact ionization and the overall efficiency of the detector. The avalanche process can also be affected by noise sources such as thermal noise or dark current, which can introduce additional carriers and reduce the signal-to-noise ratio of the detector.

4.1 Impact ionization threshold energy

Impact ionization is a phenomenon that occurs when a carrier reaches a specific energy level known as the ionization threshold energy E_{th}. The ionization threshold energy is dependent on the gap energy E_{g} of the semiconductor, which is the energy required to remove a charge from the material. The smaller the gap energy, the easier it is to remove a charge from the semiconductor, as the valence band is closer to the conduction band, and the required electric field $\vec{F}$ is lower. The gap energies also depend on the material with the following examples:

- Si: $E_{\mathrm{g}} = 1.12\ eV$
- InAs: $E_{\mathrm{g}} = 0.33\ eV$
- GaP: $E_{\mathrm{g}} = 2.24\ eV$

In 1960, Keldysh established the relationship between impact ionization and the ionization threshold energy E_{th} [18]. Later, in 1972, Anderson and Cravell developed an algorithm to calculate the threshold energy for real band structures [19]. This algorithm was further refined by Ballinger, who developed a more precise method.

The threshold energy E_{th} is expressed as a function of the gap energy E_{g} of the semiconductor and the isotropic effective mass of electrons m_{e}^* and holes m_{h}^* with the ***Eq. I-3*** and ***Eq. I-4*** [20] :

$$E_{\mathrm{th_e}} = E_{\mathrm{g}}\left(1 + \frac{m_{\mathrm{e}}^*}{m_{\mathrm{e}}^* + m_{\mathrm{h}}^*}\right) \qquad \textbf{\textit{Eq. I-3}}$$

$$E_{\mathrm{th_h}} = E_{\mathrm{g}}\left(1 + \frac{m_{\mathrm{h}}^*}{m_{\mathrm{e}}^* + m_{\mathrm{h}}^*}\right) \qquad \textbf{\textit{Eq. I-4}}$$

⇨ If $m_{\mathrm{e}}^* = m_{\mathrm{h}}^*$ then $E_{\mathrm{th_e}} = E_{\mathrm{th_h}} = \frac{3}{2}E_{\mathrm{g}}$

In reality, band structures of semiconductors are never as perfectly parabolic as in ***Figure I-13***, which means that the ionization threshold energy E_{th} is strongly dependent on the crystallographic orientation of the semiconductor. This is because crystallographic orientation affects the effective mass of electrons and holes, which in turn affects the ionization threshold energy. Therefore, the ionization threshold energy is not a fixed value for a given semiconductor, but rather varies depending on the crystallographic orientation of the material.

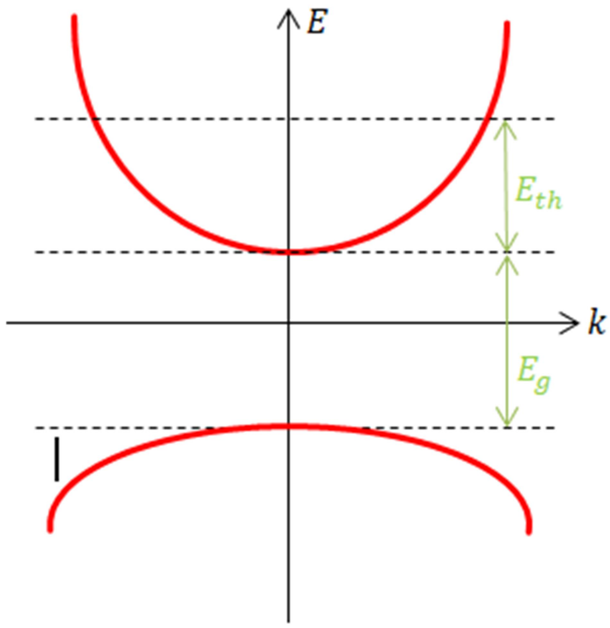

Figure I-13 Semiconductor band structure

The impact ionization probability $\bar{p}$ is directly proportional to the ionization threshold energy E_{th} [20] according to the ***Eq. I-5***:

$$\bar{p} = a(E - E_{\mathrm{th}})^b \qquad \textbf{\textit{Eq. I-5}}$$

- a and b proportionality coefficients depending on the material, $b = 2$ [21]
    - For Si in equilibrium at 300K, $\bar{p} \approx 6 \cdot 10^{-13}$ [21]

This means that as the ionization threshold energy increases, so does the probability of impact ionization occurring. The impact ionization probability is a measure of the likelihood that an electron or hole will undergo impact ionization when it reaches the ionization threshold energy. This probability increases with the average carrier energy and therefore for very high electric fields $\vec{E}$ and is also influenced by a variety of factors, including the energy of the carrier, the crystallographic orientation of the semiconductor, and the presence of impurities or defects in the material.

4.2 Carrier energy

Carriers inside semiconductor junctions derive most of their energy from the electric field by ***Eq. I-6***:

$$E = E_0 - qv\tau_e F \qquad \textbf{Eq. I-6}$$

- E_0 the energy at equilibrium $E_0 = \frac{3}{2}kT$
- q the elementary charge
- v the carrier velocity
- τ_e the relaxation time $\approx 10^{-12}, 10^{-13}\ s$
- F the electric field

This energy has a limit due to the saturation of carrier velocity when the field is too strong. ***Eq. I-7*** and ***Eq. I-8*** sum up the electric field equations at both low and high electric fields.

Low Electric fields:

$$E_{\mathrm{LF}} = E_0 - q\mu_{\mathrm{e}}F^2\tau_{\mathrm{e}} \qquad \textbf{Eq. I-7}$$

High electric fields:

$$E_{\mathrm{HF}} = E_0 - qv_{\mathrm{s}}\tau_{\mathrm{e}}F \qquad \textbf{Eq. I-8}$$

Here, v_{s} represents the saturation velocity, which is approximately 10^7 cm.s^{-1} in Silicon (Si), and μ_{e} represents the electron mobility, which is approximately 1500 cm.V^{-1}.s^{-1}. The mobility of carriers, and therefore their speed, is influenced by temperature and doping levels. This carrier mobility can be visualized using ***Figure I-14*** [22], which shows the effect of the electric field for different doping levels on it.

The carrier mobility and consequently the energy of carriers, is therefore also dependent on temperature, doping concentration, and the strength of the electric field. It is important to note that for maximum average energy, the temperature should be low, and the doping should not be too high.

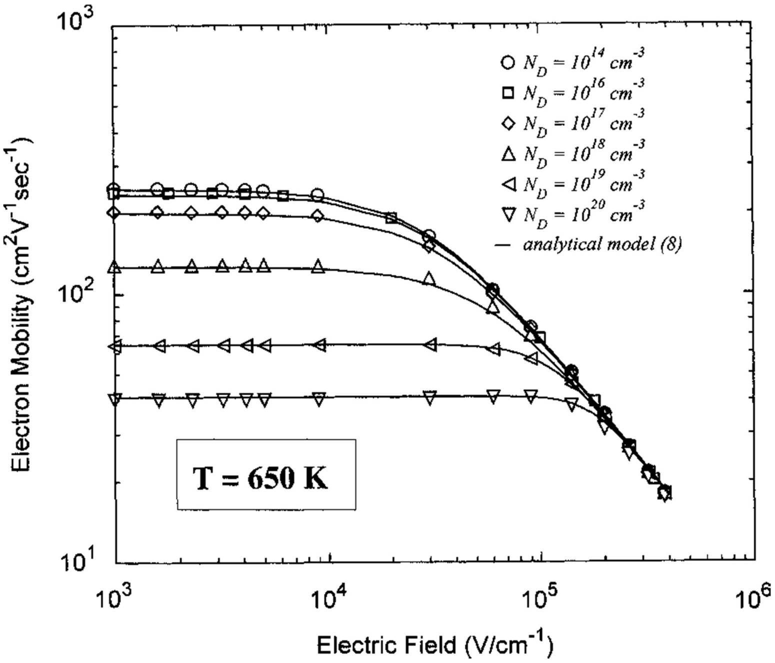

Figure I-14 Graph of electron mobility as a function of both electric field and doping concentration

4.3 Impact ionization phenomena

When studying the phenomenon of impact ionization in semiconductors, there are two main approaches that can be used: approximation by characteristic breakdown field and approximation by ionization rates.

The first approach, approximation by characteristic breakdown field, involves calculating the electric field at which impact ionization occurs and using this value as a characteristic breakdown field for the material. This approach is based on the assumption that impact ionization occurs when the electric field reaches a critical value, which is determined by the material properties. The characteristic breakdown field is an important parameter for determining the performance and reliability of semiconductor devices, as it provides a measure of the maximum electric field that the material can withstand before experiencing breakdown.

The second approach, approximation by ionization rates, involves calculating the rate at which impact ionization occurs as a function of the electric field strength. This approach is based on the assumption that the probability of impact ionization occurring increases with increasing electric field strength. By calculating the ionization rate as a function of the electric field, it is possible to predict the behavior of carriers in the material and optimize the performance of semiconductor devices.

Both of these approaches are widely used in the field of semiconductor physics and engineering, and each has its own advantages and limitations. The choice of approach will depend on the specific application and the desired level of accuracy.

4.4 Approximation by characteristic breakdown field

The approach of using the characteristic breakdown field to predict avalanche breakdown in semiconductors is based on the principle that if the maximum electric field F_{max} exceeds the characteristic breakdown field F_{br}, then impact ionization occurs and leads to an avalanche breakdown. To better understand this approach, it is necessary to introduce the concept of breakdown voltage and its relationship to avalanche breakdown.

The breakdown voltage V_{br} is the voltage at which a diode experiences breakdown and conducts current. The breakdown voltage is dependent on temperature and can be expressed by ***Eq. I-9*** [23]:

$$V_{\mathrm{br}} = V_{\mathrm{br}_0}[1 + \beta(T - T_0)] \qquad \textbf{\textit{Eq. I-9}}$$

where V_{br_0} is the breakdown voltage at room temperature $T_0 = 300K$, β is the temperature coefficient, and T is the junction temperature. The temperature coefficient β varies depending on the structure of the semiconductor, with values of $\beta = 6.82 \cdot 10^{-4}\, K^{-1}$ for a circular single-photon avalanche diode (SPAD) structure and $\beta = 6.79 \cdot 10^{-4}\, K^{-1}$ for an elliptical structure [23].

In a diode, the electric field F reaches its maximum value F_{max} at the boundary between the P-doped and N-doped zones, which is located in the Space Charge Region (SCR). This occurs when the reverse bias voltage V_pol of the diode is higher than its breakdown voltage V_{br}. The SCR is a region where the majority carriers are depleted, and the remaining carriers are minority carriers that have diffused across

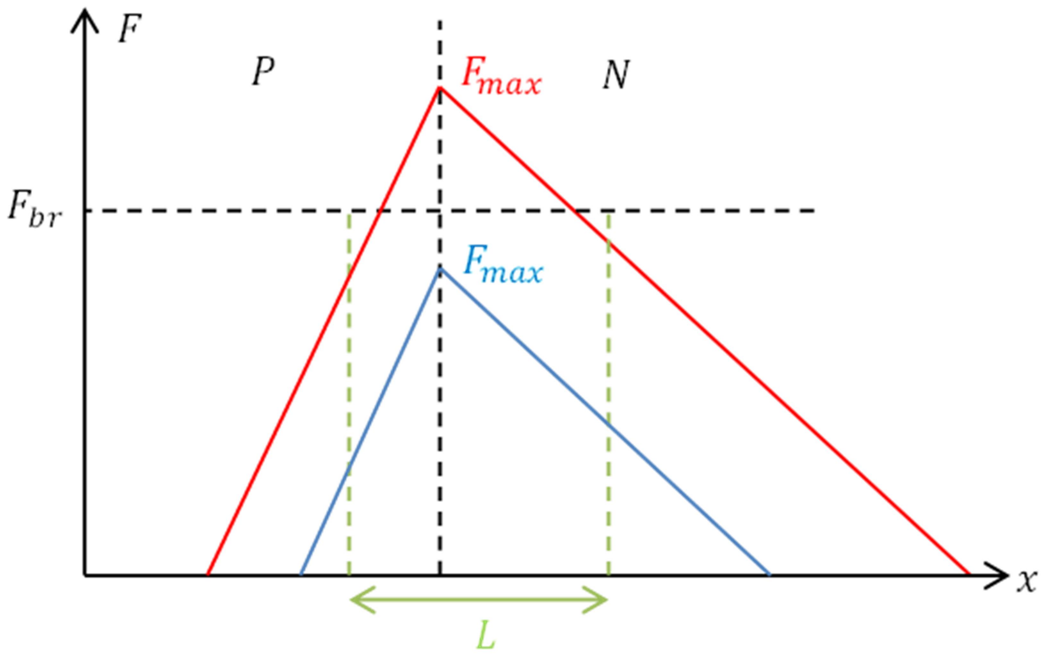

Figure I-15 Profile of the electric field F in a p/n junction

the junction.

As shown in ***Figure I-15***, the behavior of the diode depends on the relationship between the maximum electric field F_{max} and the characteristic breakdown field F_{br}.

- If F_{max} is greater than F_{br}, an avalanche breakdown occurs, as shown by the red plot. In this case, the electric field is strong enough to cause impact ionization, which leads to the generation of additional carriers and a rapid increase in current. This breakdown can cause permanent damage to the diode and is typically undesirable in most applications.
- On the other hand, if F_{max} is less than F_{br}, no breakdown occurs, as shown by the blue plot. In this case, the electric field is not strong enough to cause impact ionization, and the diode operates normally.

4.5 Approximation by ionization rates

The approximation of impact ionization in semiconductors is based on two models: the local model and the non-local model. These models are used to describe the behavior of carriers in regions of strong electric field, where impact ionization is likely to occur.

The concept of dead space is central to this approximation. The dead space is defined as the minimum distance that a carrier must travel after being injected into a region of strong electric field $\vec{F}$ or after undergoing impact ionization, in order to reach the threshold energy E_{th} required for ionization to occur. During its transit, the carrier may collide with phonons, which are vibration particles in the material.

The dead space can be calculated using ***Eq. I-10*** [24]:

$$d = \frac{E_{\mathrm{th}} - N_{\mathrm{ph}} E_{\mathrm{ph}}}{q|F|} \qquad \textbf{\textit{Eq. I-10}}$$

Where E_{th} is the threshold energy for ionization, N_{ph} is the number of phonons in the material, E_{ph} is the energy of a phonon, q is the charge of the carrier and F is the electric field strength.

According to Anderson and Cravell [19], the minimum threshold energy $E_{\mathrm{th}_{\mathrm{min}}}$ for impact ionization is dependent on the crystallographic orientation of the material. For a $\langle 100 \rangle$ orientation, the minimum threshold energy is 1.1eV for electrons and 1.8eV for holes.

4.5.1 Local model

The local model is a commonly used approach for approximating impact ionization in semiconductors. This model is based on several assumptions:

- The electric field $\vec{F}$ is time independent and uniformly distributed in space.
- The avalanche breakdown area's width W is large compared to the dead space.

- The stationary state must be quickly reached by the carriers under the effect of the electric field $\vec{F}$ and phonons.

However, this model has limitations: for structures with a strong electric field gradient $\vec{F}$, the local model cannot accurately predict the reality, since the ionization coefficients depend directly on $\vec{F}$.

The ionization coefficients for silicon are defined by ***Eq. I-11*** and ***Eq. I-12*** [25]:

$$\alpha_\mathrm{n} = A_\mathrm{n} \cdot \exp\left(-\frac{B_\mathrm{n}}{F}\right) \qquad \textbf{\textit{Eq. I-11}}$$

$$\alpha_\mathrm{p} = A_\mathrm{p} \cdot \exp\left(-\frac{B_\mathrm{p}}{F}\right) \qquad \textbf{\textit{Eq. I-12}}$$

Those equations are also known as the Chynoweth expressions and accurately represent the evolution of the ionization coefficients. The coefficients $B_{\mathrm{n/p}}$ can be extended according to ***Eq. I-13*** [30]:

$$B_{\mathrm{n/p}} = C_{\mathrm{n/p}} + D_{\mathrm{n/p}}T \qquad \textbf{\textit{Eq. I-13}}$$

The coefficients $D_{\mathrm{n/p}}$ introduce the temperature dependancy of the ionization phenomenon, crucial to accurately model the SPAD's behavior over its entire temperature range.

Parameter	n	p
$A_{n/p}$	$4.43 \cdot 10^5\ cm^{-1}$	$1.13 \cdot 10^6\ cm^{-1}$
$C_{n/p}$	$9.66 \cdot 10^5\ V.cm^{-1}$	$1.71 \cdot 10^6\ V.cm^{-1}$
$D_{n/p}$	$4.99 \cdot 10^2\ V.cm^{-1}.K^{-1}$	$1.09 \cdot 10^3\ V.cm^{-1}.K^{-1}$

Figure I-16 Impact ionization parameter table

In his study, Massey determined the values of factors $A_{\mathrm{n/p}}$, $C_{\mathrm{n/p}}$ and $D_{\mathrm{n/p}}$ giving a good fit for ionization coefficients, given in ***Figure I-16***. The valid conditions for the best fit are the following:

- Electric field range : 200-800 $\mathrm{kV \cdot cm^{-1}}$
- Temperature range : 15-420 K

Under those conditions, the values for the different parameters of the impact ionization coefficients are summarized in ***Figure I-16***.

4.5.2 Non-local model

In this model, the parameters dependent on the ionization coefficients are multiple and no longer depend only on the local field $\vec{F}$. For this model, we take into account the "history" of the carrier, in other words the energy that has already been absorbed before.

There are different interpretations concerning the consideration of the dead space after the generation of the carrier and after each of the ionizing shocks in which it has participated or only after the generation.

Several known models that have been developed using the non-local approach are those of McIntyre [27] or Hayat and Pancheri [3].

The two main models that take into account dead space are the Hard dead space and Soft dead space models. In both of these models, a probability density function is defined as follows [29]:

$$p_{\mathrm{n/p}} = \begin{cases} 0, x_0 < d_{\mathrm{n/p}} - x \\ \alpha_{\mathrm{n/p}}(F) \cdot \exp\left(\alpha_{\mathrm{n/p}}(F) \cdot \left(x_0 + x - d_{\frac{\mathrm{n}}{\mathrm{p}}}\right)\right), x_0 \geq d_{\mathrm{n/p}} - x \end{cases} \qquad \textbf{\textit{Eq. I-14}}$$

4.5.2.1 Hard dead space

The hard dead space model excludes the possibility of ionization if the ionizing carrier hasn't traveled its corresponding dead space.

Until a carrier has traveled its dead space $d_{\mathrm{n/p}}$, its ionization coefficient is zero. Once it has reached $d_{\mathrm{n/p}}$, it spontaneously takes a value $\alpha^*_{\mathrm{n/p}}$ called microscopic ionization coefficient that are directly correlated to their local counterpart by ***Eq. I-15*** [24]:

$$\frac{1}{\alpha_{\mathrm{n/p}}} = \frac{1}{\alpha^*_{\mathrm{n/p}}} + d_{\mathrm{n/p}} \qquad \textbf{\textit{Eq. I-15}}$$

By rearranging this equation to ***Eq. I-16***, the microscopic ionization coefficients can be expressed using the local coefficients to greatly simplify implementation inside a model.

$$\alpha^*_{\mathrm{n/p}} = \frac{1}{\frac{1}{\alpha_{\mathrm{n/p}}} - d_{\mathrm{n/p}}} \qquad \textbf{\textit{Eq. I-16}}$$

Considering Okuto and Crowell's model for non-local impact ionization, the ionization coefficient $\alpha_{\mathrm{OC}}(x)$ depending on the distance z traveled by the ionizing carrier created at $z = 0$ is expressed by ***Eq. I-17*** [24]:

$$\alpha_{\mathrm{OC}}(x) = \alpha^*_{\mathrm{n/p}} \cdot U(x - d_{\mathrm{n/p}}) \qquad \textbf{\textit{Eq. I-17}}$$

With $U(z - d_{\mathrm{n/p}})$ the unit step function:

$$U(x - d_{\mathrm{n/p}}) = \begin{cases} 0 \; if \; x - d_{\mathrm{n/p}} < 0 \\ 1 \; if \; x - d_{\mathrm{n/p}} \geq 0 \end{cases} \qquad \textbf{\textit{Eq. I-18}}$$

The associated probability density function is defined as follows [24]:

$$h_{\mathrm{OC}}(x) = \alpha^*_{\mathrm{n/p}} \cdot \exp\left(-\alpha^*_{\mathrm{n/p}}(x - d_{\mathrm{n/p}})\right) \cdot U(x - d_{\mathrm{n/p}}) \qquad \textbf{\textit{Eq. I-19}}$$

The main issue with this model is its step behavior as shown on ***Figure I-17*** [28] below. Such a behavior is physically impossible, thus the need for a smoother model to accurately describe the ionization phenomena which lead to the soft dead space model.

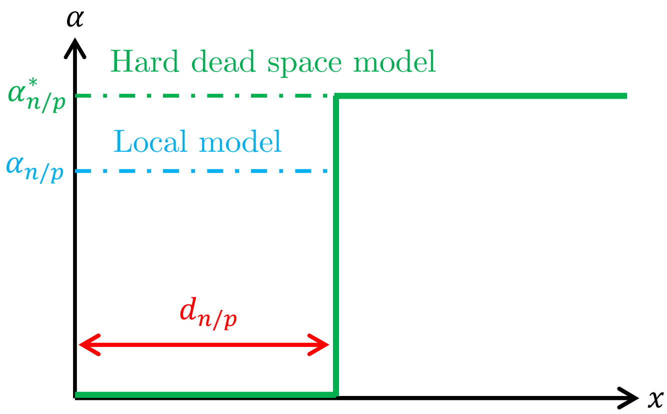

Figure I-17 Macro- and microscopic ionization coefficients depending on distance from carrier generation

4.5.2.2 Soft dead space

The soft dead space model is an extension of the hard dead space model, as it takes into account electric field fluctuations, resulting in carriers being able to ionize without having travelled their dead space. Those fluctuations are described by an effective electric field taking into account the dead space, according to Eq. ***I***-20 [29]:

$$F_{\mathrm{eff}_{\mathrm{n/p}}}(x) = F(x) \cdot \mathrm{erf}\left(\frac{x}{d_{\mathrm{n/p}}}\right) \qquad \textbf{\textit{Eq. I-20}}$$

The field is therefore modulated by the Gauss error function **erf**, considering that the electric field $F(x)$ is constant along the dead space. This effective electric field

$F_{\mathrm{eff}_{\mathrm{n/p}}}(x)$ replaces the electric field in the local model's ionization coefficient formulas.

For the soft dead space model, there are 3 other approaches [30]: The Kwon [31], Sheffield [32] and Liverpool [33] models. It has been proven through Monte Carlo

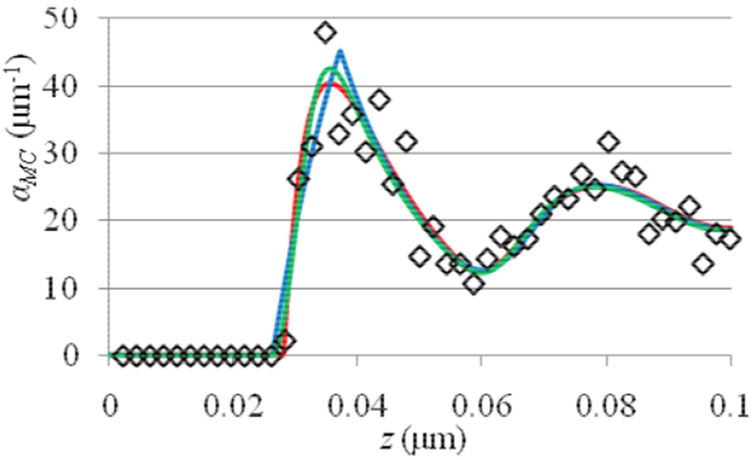

Figure I-18 Kwon, Sheffield and Liverpool models coefficients compared to Monte Carlo data from Plimmer

simulations that they are equivalently accurate, the differences between them being smaller than the random error due to Monte Carlo simulation. ***Figure I-18*** [30] below shows the comparison of the ionization factors of the different models and Monte Carlo simulation data from S.A. Plimmer [34]. Therefore, only one model is detailed in this paper: The Sheffield model developed by C.H. Tan and his team of Sheffield University.

This model is based on the same principle as the effective electric field model above: The Okuto Crowell ionization coefficient $\alpha_{\mathrm{OC}}(x)$ is taken as a base and can be described as half a Gaussian curve which introduces a spreading factor s and keeps an asymptotic value $\alpha^*_{\mathrm{n/p}}$ at $x \to +\infty$ representing the value of the microscopic ionization coefficient. Following equation ***Eq. I-21*** was developed to express a new ionization coefficient $\alpha_{\mathrm{Sheffield}}(x)$ [30]:

$$\alpha_{\mathrm{Sheffield}}(x) = \alpha^*_{n/p}\left(1 - \exp\left(\frac{-(x-d)^2}{s^2}\right)\right)U(x - d_{n/p}) \qquad \textbf{\textit{Eq. I-21}}$$

This new coefficient leads to the probability density function $h_{\text{Sheffield}}(x)$ [30]:

$$h_{\text{Sheffield}}(x) = \alpha_{\text{Sheffield}}(x) \exp\left(-\alpha^*_{\text{n/p}}\left(x - d_{\text{n/p}} - \frac{s\sqrt{\pi}}{2}\text{erf}\left(\frac{x - d_{\text{n/p}}}{s}\right)\right)\right) \qquad \textbf{\textit{Eq. I-22}}$$

5 Approaches of Physical and Statistical Modelling of Parameters of SPADs

5.1 Avalanche breakdown probability

As for ionization rates, there are two models to describe the avalanche breakdown phenomena of SPADs: a localized approach, considering only the local ionization coefficients that are considered, and non-localized approach that takes into account the past history of the carrier.

5.1.1 Local model

Let $P_{\text{p}}(z)$ be the probability that a carrier (or electron/hole pair) generated in z can generate an avalanche. We have ***Eq. I-23*** [29] :

$$P_{\text{p}}(z) = P(P_{\text{p}_\text{e}}(z) \cup P_{\text{p}_\text{h}}(z)) \qquad \textbf{\textit{Eq. I-23}}$$

$P_{\text{p}_\text{e}}(z)$ and $P_{\text{p}_\text{h}}(z)$ are independent events so the following equality, ***Eq. I-24*** can be derived by the properties of the union $\cup$ operator [35] :

$$\begin{aligned} P_{\text{p}}(z) &= P_{\text{p}_\text{e}}(z) + P_{\text{p}_\text{h}}(z) - \left(P_{\text{p}_\text{e}}(z) \cap P_{\text{p}_\text{h}}(z)\right) \\ &= P_{\text{p}_\text{e}}(z) + P_{\text{p}_\text{h}}(z) - P_{\text{p}_\text{e}}(z)P_{\text{p}_\text{h}}(z) \end{aligned} \qquad \textbf{\textit{Eq. I-24}}$$

Let an electron (or hole) be generated at $z + dz$, it may trigger an avalanche breakdown at z in two ways:

- A: it triggers an avalanche breakdown arriving at z, $P(A) = \begin{Bmatrix} P_{\mathrm{p_e}}(z) \\ P_{\mathrm{p_h}}(z) \end{Bmatrix}$
- B: it generates an electron-hole pair which will in turn initiate an avalanche breakdown traveling dz.

Let $\alpha_{\mathrm{n}} dz$ be the probability that the electron generates a secondary pair, we have ***Eq. I-25 to I-27***:

$$P(B) = \begin{Bmatrix} \alpha_{\mathrm{n}} dz P_{\mathrm{p}}(z) \\ \alpha_{\mathrm{p}} dz P_{\mathrm{p}}(z) \end{Bmatrix} \qquad \textbf{\textit{Eq. I-25}}$$

$$\begin{Bmatrix} P_{p_e}(z+dz) = P_{p_e}(z) + \alpha_n dz P_p(z) - \alpha_n dz P_p(z) P_{p_e}(z) \\ P_{p_h}(z+dz) = P_{p_h}(z) + \alpha_p dz P_p(z) - \alpha_h dz P_p(z) P_{p_h}(z) \end{Bmatrix} \qquad \textbf{\textit{Eq. I-26}}$$

$$\begin{Bmatrix} \dfrac{dP_{p_e}}{dz} = (1 - P_{p_e}) \alpha_n (P_{p_e} + P_{p_h} - P_{p_e} P_{p_h}) \\ \dfrac{dP_{p_h}}{dz} = (1 - P_{p_h}) \alpha_p (P_{p_e} + P_{p_h} - P_{p_e} P_{p_h}) \end{Bmatrix} \rightarrow \begin{Bmatrix} P_{p_e}(z_n) = 0 \\ P_{p_h}(z_p) = 0 \end{Bmatrix} \qquad \textbf{\textit{Eq. I-27}}$$

Those equations are Oldham's equations, constituting the framework for the local model [36].

5.1.2 Non-local model

The principle is as follows: a particle must have enough energy before it can ionize by travelling a certain distance called dead space.

Let d_e and d_h be the dead spaces for electrons and holes and E_{th_e} and E_{th_h} the threshold energies of the latter. In the SCR, the dead space is related to the electrostatic potential $\Psi(z) = \frac{E_{\mathrm{pot}}}{q}$ by the following relationships of ***Eq. I-28*** and ***Eq. I-29*** [3]:

$$\Psi(z + d_{\mathrm{e}}(z)) - \Psi(z) = \frac{E_{\mathrm{th_e}}}{q} \qquad \textbf{\textit{Eq. I-28}}$$

$$\Psi(z) - \Psi(z - d_{\mathrm{h}}(z)) = \frac{E_{\mathrm{th_h}}}{q} \qquad \textbf{\textit{Eq. I-29}}$$

The potential energy reached by travelling a distance of $d_{\mathrm{e}}/d_{\mathrm{h}}$ is characterized by $E_{\mathrm{th_e}}/E_{\mathrm{th_h}}$, the threshold energies of the considered carriers. The opposite signs in the positions of the electrostatic potentials are caused by the fact that holes move in the opposite direction of electrons.

The profile of the electrostatic potential can be derived from the electric field profile ($\vec{F} = -\mathrm{grad}(\Psi)$) and then solve digitally.

Let $h(\varepsilon|z)$ be the probability density functions of the distance ε to the first ionizing impact.

For electrons, we have [3]:

$$\left\{\begin{array}{c} \varepsilon < d_{\mathrm{e}} : h_{\mathrm{e}}(\varepsilon|z) = 0 \\ \varepsilon \geq d_{\mathrm{e}} : \; h_{\mathrm{e}}(\varepsilon|z) = \alpha_{\mathrm{n}}(z+\varepsilon) \cdot exp\left(-\int_{d_{\mathrm{e}}(z)}^{\varepsilon} \alpha_{\mathrm{n}}(z+y)dy\right) \end{array}\right\} \qquad \textbf{\textit{Eq. I-30}}$$

And analogously for holes [3]:

$$\left\{\begin{array}{c} \varepsilon < d_{\mathrm{h}} : h_{\mathrm{e}}(\varepsilon|z) = 0 \\ \varepsilon \geq d_{\mathrm{h}} : \; h_{\mathrm{h}}(\varepsilon|z) = \alpha_{\mathrm{p}}(z-\varepsilon) \cdot exp\left(-\int_{d_{\mathrm{h}}(z)}^{\varepsilon} \alpha_{\mathrm{p}}(z-y)dy\right) \end{array}\right\} \qquad \textbf{\textit{Eq. I-31}}$$

α_{n} and α_{p} are the adapted impact ionization coefficients of the non-local model, as seen in the hard dead space description in **Chapter I**, **Subhapter 4**.

Let $P_{\mathrm{Z}}(z)$ be the probability that an electron in z generates a finite number of electrons and holes, related to $P_{\mathrm{p_e}}(z)$ by the relation: $P_{\mathrm{Z}}(z) = 1 - P_{\mathrm{p_e}}(z)$

Let $P_{\mathrm{Y}}(z)$ be the probability that a hole in z generates a finite number of electrons and holes, related to $P_{\mathrm{p_h}}(z)$ by the relation: $P_{\mathrm{Y}}(z) = 1 - P_{\mathrm{p_h}}(z)$

From those probabilities, the following equations ***Eq. I-32*** and ***Eq. I-33*** [37] :

$$P_{\mathrm{Y}}(z) = \int_{z}^{\infty} h_{\mathrm{h}}(\xi|z)d\xi + \int_{0}^{z} P_{\mathrm{Y}}^{2}(z+\xi)P_{\mathrm{Z}}(z+\xi)\, h_{\mathrm{e}}(\xi|z)d\xi \qquad \textbf{\textit{Eq. I-32}}$$

$$P_{\mathrm{Z}}(z) = \int_{L-z}^{\infty} h_{\mathrm{e}}(\xi|z)d\xi + \int_{0}^{L-z} P_{\mathrm{Z}}^{2}(z+\xi)P_{\mathrm{Y}}(z+\xi)\, h_{\mathrm{e}}(\xi|z)d\xi \qquad \textbf{\textit{Eq. I-33}}$$

- With L the width of the space charge region (SCR)

The avalanche triggering probability, when combining ***Eq. I-32*** and ***Eq. I-33***, caused by an electron-hole pair generated in z, $P_{\mathrm{p}}(z)$ equals to ***Eq. I-34***:

$$P_{\mathrm{p}}(z) = 1 - P_{\mathrm{Z}}(z)P_{\mathrm{Y}}(z) \qquad \textbf{\textit{Eq. I-34}}$$

From this equation, we can deduce the probabilities of an avalanche being triggered by an electron, $P_{\mathrm{p_e}}(z)$, and by a hole, $P_{\mathrm{p_h}}(z)$.

5.2 Photon detection probability PDP

As discussed and introduced in **Chapter I**, **Subchapter 3**, the photon detection probability is defined by the probability that a photon is detected by the device and generated an output signal. ***Eq. I-35*** is the formal definition of the PDP.

$$PDP = P_T \cdot QE \qquad \textbf{\textit{Eq. I-35}}$$

The PDP is a function of two components:

- the photon transmission probability P_T, which is the probability that a photon is transmitted across the device and reaches the active region

- the quantum efficiency QE which is the probability that the photon that arrives at the active region generates an electron, hole or electron-hole pair.

5.2.1 Modeling P_T

For the modeling of the photon transmission probability, a matrix method can be used [38]. ***Figure I-19*** shows a propagation matrix of i layers with the corresponding reflection and transmission vectors.

The pass relation, ***Eq. I-36*** between layer i and $i+1$ is:

$$\begin{bmatrix} F_i^+ \\ F_i^- \end{bmatrix} = \frac{1}{t_{i\rightarrow i+1}} M_{i+1} \begin{bmatrix} F_{i+1}^+ \\ F_{i+1}^- \end{bmatrix} \qquad \textbf{Eq. I-36}$$

- $t_{i\rightarrow i+1}$ transmission coefficient between layer i and $i+1$
- F_i^+ incident electric field
- F_i^- reflected electric field

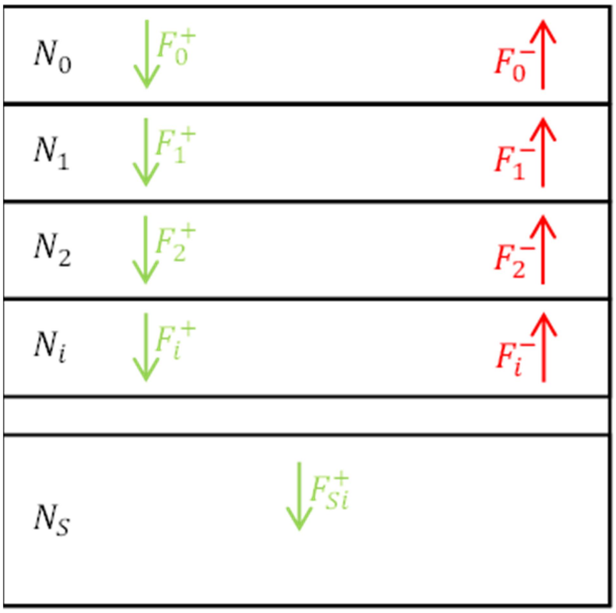

Figure I-19 Diagram of reflection and transmission of the electric field

M_{i+1} , as in ***Eq. I-37***, is the propagation matrix between layer i and $i+1$:

$$M_{i+1} = \begin{bmatrix} exp(j\delta_{i+1}) & r_{i\to i+1} \cdot exp(-j\delta_{i+1}) \\ r_{i\to i+1} \cdot exp(j\delta_{i+1}) & exp(-j\delta_{i+1}) \end{bmatrix} \qquad \textbf{\textit{Eq. I-37}}$$

- $r_{i\to i+1}$ reflexion coefficient between layer i and $i+1$
- δ_{i+1} the phase shift in layer i, δ_i following ***Eq. I-38***:

$$\delta_i = \frac{2\pi N_i d_i \cos\theta_i}{\lambda} \qquad \textbf{\textit{Eq. I-38}}$$

- N_i complex refraction index of layer i, following ***Eq. I-39***:

$$N_i(\lambda) = n_i(\lambda) - jk_i(\lambda) \qquad \textbf{\textit{Eq. I-39}}$$

- n_i refraction index of layer i
- k_i extinction coefficient of the material, $k = 0$ for a transparent material
- d_i thickness of layer i
- θ_i incidence angle on layer i
- λ wavelength of incident photon

Added to those equations, according to the Snell-Descartes relationship, we have ***Eq. I-40:***

$$n_i \sin\theta_i = n_{i+1} \sin\theta_{i+1} \qquad \textbf{\textit{Eq. I-40}}$$

Knowing that the reflection of the Si layer is null, $F_{Si}^- = 0$ we establish the following relation in ***Eq. I-41***:

$$\begin{bmatrix} F_0^+ \\ F_0^- \end{bmatrix} = \left[\prod_1^{Si} \frac{1}{t_{i\to i+1}}\right]\left[\prod_1^{Si} M_i\right]\begin{bmatrix} F_{Si}^+ \\ 0 \end{bmatrix} \qquad \textbf{\textit{Eq. I-41}}$$

By solving numerically ***Eq. I-41***, we find the transmission probability of the photon P_T and its expression as in ***Eq. I-42***:

$$P_T = \frac{n_{Si}}{n_0}\left|\frac{F_{Si}^{+}}{F_0^{+}}\right|^2 \qquad \textbf{\textit{Eq. I-42}}$$

The sine and cosine terms, which are caused by the exponential terms in the transmission matrix M_i, represent oscillations at the transmission level between layers. Their frequency depends on phase shift δ_i : it increases when λ decreases and vice versa.

5.2.2 Modeling *QE*

The first step seen in the modelling approaches of the quantum efficiency (QE), the junction has to be first represented graphically and properly annotated. In this case, let the P/N junction be the one represented in ***Figure I-20***.

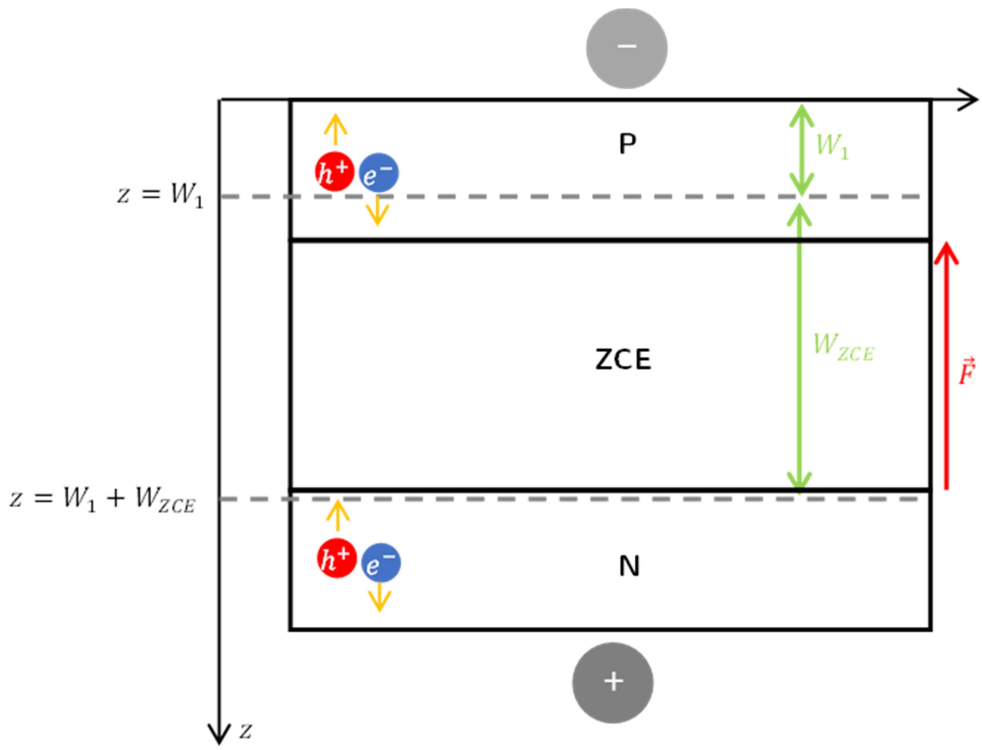

Figure I-20 Diagram of a P/N junction with visualization of the carriers movements and areas of interest

If we come back to the definition of Quantum efficiency (QE), it is described as the probability that a photon reaching the active area of the device (photosensitive silicon) is absorbed and generates an avalanche. This figure of merit has a strong dependance on the wavelength λ of the incident and absorbed photon.

We can represent quantum efficiency as the sum of three contributions, as in ***Eq. I-43***:

$$QE(\lambda) = QE_P(\lambda) + QE_{ZCE}(\lambda) + QE_N(\lambda) \qquad \textbf{\textit{Eq. I-43}}$$

- QE_P quantum efficiency in the P-type region outside the SCR
- QE_{ZCE} quantum efficiency inside the SCR
- QE_N quantum efficiency in the N-type region outside the SCR

The equations used in the following parts, describing an existing modelling method in the different sections of the device, are also based on the schematic of ***Figure I-20***.

5.2.2.1 QE In the non-depleted regions

When a photon is absorbed in zone P and creates an electron-hole pair, there is carrier diffusion in this zone. The electron can diffuse up to the limit of the SCR W_1 and thus be carried along by the strong electric field present in the SCR.

The probability of absorption of a photon at depth z follows the following absorption law in ***Eq. I-44*** [39]:

$$P_{\text{abs}} = \alpha_{\text{abs}}(\lambda) \cdot \exp(-\alpha_{\text{abs}}(\lambda) z) \qquad \textbf{\textit{Eq. I-44}}$$

- α_{abs} absorption coefficient of a photon at wavelength λ

After absorption of the photon and generation of an electron-hole pair, the electron must diffuse up to W_1 without recombining by Auger [40] or Schockley-Read-Hall effect [41]. The probability is following ***Eq. I-45*** [42]:

$$P_{\mathrm{e_{diff}}} = \exp\left(-\frac{W_1 - z}{L_\mathrm{e}}\right) \qquad \textbf{\textit{Eq. I-45}}$$

- L_e the electron scattering length, following ***Eq. I-46***:

$$L_\mathrm{e} = \sqrt{D_\mathrm{e}\tau_\mathrm{e}} \qquad \textbf{\textit{Eq. I-46}}$$

- D_e the diffusion coefficient (Fick's Law), Einstein's relation gives ***Eq. I-47***:

$$\frac{D_\mathrm{e}}{\mu_\mathrm{e}} = \frac{D_\mathrm{h}}{\mu_\mathrm{h}} = \frac{kT}{q} \Rightarrow D_\mathrm{e} = \mu_\mathrm{e}\frac{kT}{q} \qquad \textbf{\textit{Eq. I-47}}$$

When the two probability equations are put together and integrated from 0 to W_1, in order to obtain the probability that a photogenerated electron diffuses up to the limit of the SCR W_1 [43], we get ***Eq. I-48***:

$$P_1 = \int_0^{W_1} P_{\mathrm{abs}}P_{\mathrm{e_{diff}}} = \int_0^{W_1} \alpha_{\mathrm{abs}}(\lambda)\cdot\exp(-\alpha_{\mathrm{abs}}(\lambda)z)\cdot\exp\left(-\frac{W_1 - z}{L_\mathrm{e}}\right)dz$$

Eq. I-48

$$= \frac{L_\mathrm{e}\alpha_{\mathrm{abs}}}{L_\mathrm{e}\alpha_{\mathrm{abs}} - 1}\left[\exp\left(-\frac{W_1}{L_\mathrm{e}}\right) - \exp(-\alpha_{\mathrm{abs}}(\lambda)W_1)\right]$$

This electron that has reached the SCR must be driven by the field, multiply by impact ionization and trigger an avalanche to be detected by the device. Therefore, using the probability of triggering an avalanche by an electron generated in z, $P_{\mathrm{p_e}}(z)$, in this case at W_1, we get ***Eq. I-48***:

$$QE_\mathrm{P} = P_1P_{\mathrm{p_e}}(W_1) \qquad \textbf{\textit{Eq. I-49}}$$

When applying the same reasoning for a hole generated in the N-type region of the device, which diffuses up to the other limit of the SCR, $W_1 + W_{\mathrm{ZCE}}$, ***Eq. I-50*** is obtained [43]:

$$\begin{aligned} P_2 &= \int_{W_1+W_{\mathrm{ZCE}}}^{z_{\mathrm{bot}}} P_{\mathrm{abs}} P_{\mathrm{p_{diff}}} \\ &= \int_{W_1+W_{\mathrm{ZCE}}}^{z_{\mathrm{bot}}} \alpha_{\mathrm{abs}}(\lambda) \cdot \exp(-\alpha_{\mathrm{abs}}(\lambda) z) \cdot \exp\left(-\frac{z-(W_1+W_{\mathrm{ZCE}})}{L_{\mathrm{h}}}\right) dz \\ &= \frac{L_{\mathrm{h}}\alpha_{\mathrm{abs}}}{L_{\mathrm{h}}\alpha_{\mathrm{abs}}+1} \cdot \exp(-\alpha_{\mathrm{abs}}(\lambda)(W_1+W_{\mathrm{ZCE}})) \end{aligned} \qquad \textbf{\textit{Eq. I-50}}$$

From this, it is derived that the quantum efficiency in the N-type region is defined by ***Eq. I-51***:

$$QE_{\mathrm{N}} = P_2 P_{\mathrm{p_h}}(W_1 + W_{\mathrm{ZCE}}) \qquad \textbf{\textit{Eq. I-51}}$$

5.2.2.2 QE In the depleted region (SCR)

For quantum efficiency in the SCR, the calculation is simpler since the photogenerated carriers do not need to diffuse and are directly driven by the electric field. The two carriers of the photogenerated electron-hole pair can trigger an avalanche, so using $P_{\mathrm{p}}(z)$ we derive the relation ***Eq. I-52*** for QE_{ZCE} [43]:

$$QE_{\mathrm{ZCE}} = \int_{W_1}^{W_1+W_{\mathrm{ZCE}}} \alpha_{\mathrm{abs}}(\lambda) \cdot \exp(-\alpha_{\mathrm{abs}}(\lambda) z)\, P_{\mathrm{p}}(z) dz \qquad \textbf{\textit{Eq. I-52}}$$

Here, the probability of an avalanche triggered by an electron-hole pair $P_p(z)$ is integrated over the whole width of the SCR W_{ZCE}, mainly because avalanche triggering carriers can be generated in the whole SCR.

5.2.3 Total contribution and Photon Detection Efficiency (PDE)

This entire parameter calculation with the various expressions finally leads to the value of the photon detection probability PDP, which combines the transmission probability of a photon P_{T} and quantum efficiency QE. It is the probability that a photon arrives on the active surface of the device, creates an electron-hole pair and triggers an avalanche which can then be detected by the photon counting circuit, it was previously established by ***Eq. I-35*** that we have:

$$PDP = QE \cdot P_{\mathrm{T}}$$

An important parameter in SPAD modeling, directly linked to the PDP is the Photon Detection Efficiency PDE [44], taking into account the geometry of the studied device. This PDE is expressed by ***Eq. I-53***:

$$PDE = PDP \cdot F_{\mathrm{G}} = QE \cdot P_{\mathrm{T}} \cdot F_{\mathrm{G}} \qquad \textbf{\textit{Eq. I-53}}$$

The correction (or geometric) factor F_{G}, also called "Fill Factor" is defined as the ratio between the active area of the SPAD and the total area of the pixel [45].

Optimization of the PDE requires a good geometry of the SPAD, which should ideally have a uniform avalanche triggering probability over the entire multiplication area

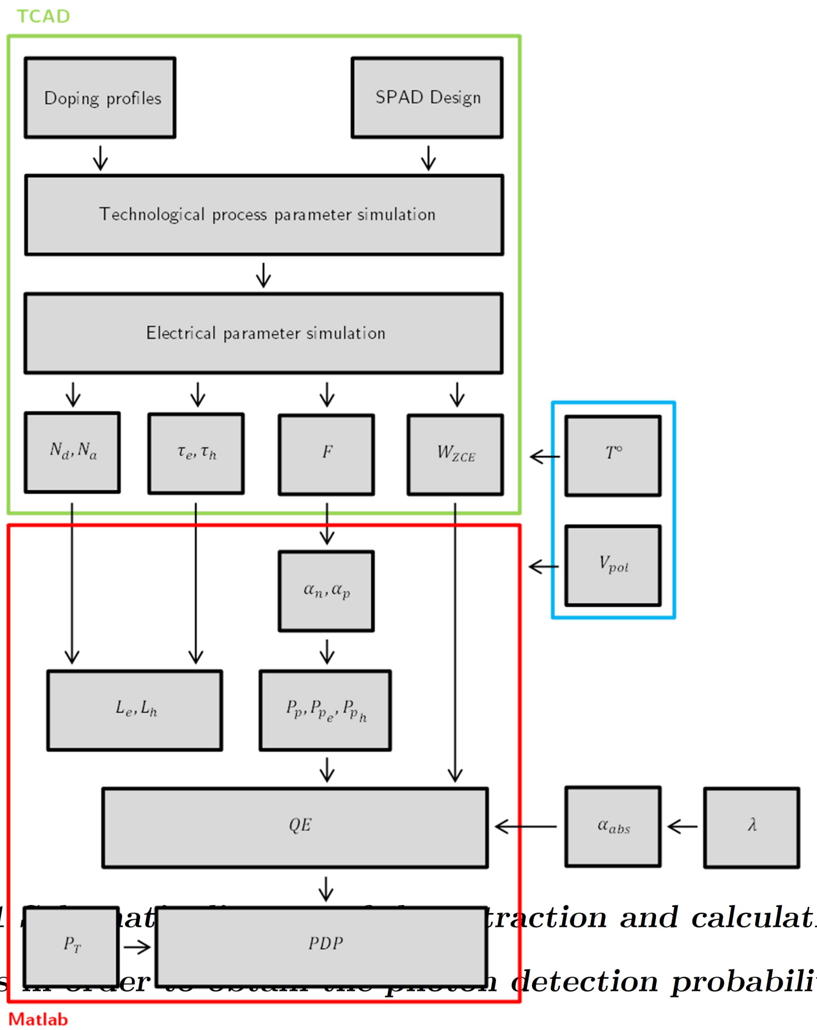

Figure I-21 Schematic diagram of the extraction and calculation of the parameters in order to obtain the photon detection probability (PDP)

[46]. This is why the optimal geometry for a SPAD is circular, to limit the strong fields at the edges of the device. Extraction and calculation of the PDP in the particular method presented in this manuscript is described in ***Figure II-21*** [29].

5.3 Modeling Dark Count Rate DCR

As seen previously, the Dark Count Rate is the main source of noise in SPAD devices. It is mainly due to spontaneous generation of electrons, holes or electron-hole pairs and is composed of three contributions according to ***Eq. I-54***:

$$DCR = DCR_{\mathrm{SRH-TAT}} + DCR_{\mathrm{BTBT}} + DCR_{\mathrm{Afterpulsing}} \qquad \textit{\textbf{Eq. I-54}}$$

- $DCR_{\mathrm{SRH-TAT}}$ SRH generation contribution, alone or enhanced by trap assisted tunneling (TAT)
- DCR_{BTBT} Band-to-Band tunneling contribution
- $DCR_{\mathrm{Afterpulsing}}$ Afterpulsing contribution

It is possible to decorrelate the Afterpulsing contribution $DCR_{\mathrm{Afterpulsing}}$ by simulations and/or measurements, so the Afterpulsing effect will be covered separately in the following section **5.4**.

The DCR can be evaluated for each region of the junction by following the method detailed by A. Panglosse [47] and ***Eq. I-55*** can thus be deduced:

$$DCR_{\mathrm{SRH-TAT}} + DCR_{\mathrm{BTBT}} = DCR_{\mathrm{P}} + DCR_{\mathrm{SCR}} + DCR_{\mathrm{N}} \qquad \textit{\textbf{Eq. I-55}}$$

5.3.1 DCR in non-depleted regions

$P_{\mathrm{p_e}}(W_1)$ being the probability that an avalanche is triggered by an electron generated in the P-type region of the junction, we have by Fick's law the diffusion currents:

$$\begin{cases} J_{\mathrm{e}} = qD_{\mathrm{e}} \cdot \mathrm{grad}(\vec{n}) = qD_{\mathrm{e}} \cdot \dfrac{dn}{dx} \\ J_{\mathrm{h}} = -qD_{\mathrm{h}} \cdot \mathrm{grad}(\vec{p}) = -qD_{\mathrm{h}} \cdot \dfrac{dp}{dx} \end{cases} \qquad \textbf{\textit{Eq. I-56}}$$

The surface DCR represents the number of thermally generated electrons that have scattered to the edge of the SCR, W_1, and triggered an avalanche. (Counts per unit area). Derived from ***Eq. I-56*** are the resulting contributions DCR_{P} and DCR_{N}, as detailed in ***Eq. I-57*** and ***Eq. I-58***:

$$DCR_{\mathrm{P}} = P_{\mathrm{p_e}}(W_1)\frac{J_{\mathrm{e}}(W_1)}{q} \qquad \textbf{\textit{Eq. I-57}}$$

$$DCR_{\mathrm{N}} = P_{\mathrm{p_h}}(W_1)\frac{J_{\mathrm{p}}(W_1 + W_{ZCE})}{q} \qquad \textbf{\textit{Eq. I-58}}$$

In order to model the diffusion current densities J_{p} and J_{n}, the continuity equations of ***Eq. I-59*** below are used, resulting in the current densities' equations in ***Eq. I-60***:

$$\begin{cases} \dfrac{dn}{dt} = D_{\mathrm{e}}\dfrac{d^2n}{dx^2} - \dfrac{n - n_0}{\tau_0} \\ \dfrac{dp}{dt} = -D_{\mathrm{h}}\dfrac{d^2p}{dx^2} - \dfrac{p - p_0}{\tau_0} \end{cases} \qquad \textbf{\textit{Eq. I-59}}$$

- $\frac{n-n_0}{\tau_0}$ the electron generation rate
- $\frac{p-p_0}{\tau_0}$ the hole generation rate

$$\begin{cases} J_{\mathrm{e}} = qD_{\mathrm{e}}\dfrac{dn}{dx} \Rightarrow \dfrac{dJ_{\mathrm{e}}}{dx}\dfrac{1}{q} = D_{\mathrm{e}}\dfrac{d^2n}{dx^2} = \dfrac{\nabla J_{\mathrm{e}}}{q} \\ J_{\mathrm{h}} = -qD_{\mathrm{h}}\dfrac{dp}{dx} \Rightarrow \dfrac{dJ_{\mathrm{h}}}{dx}\dfrac{1}{q} = -D_{\mathrm{h}}\dfrac{d^2p}{dx^2} = \dfrac{\nabla J_{\mathrm{h}}}{q} \end{cases} \qquad \textbf{\textit{Eq. I-60}}$$

From ***Eq. 60***, the divergence of the current densities can be isolated and can be expressed as a function of the electron-hole generation rates in ***Eq. I-61***:

$$\begin{Bmatrix} \nabla J_{\mathrm{e}} = q\dfrac{dn}{dt} + qG_{\mathrm{TOT}} \\ \nabla J_{\mathrm{h}} = q\dfrac{dp}{dt} + qG_{\mathrm{TOT}} \end{Bmatrix}$$ ***Eq. I-61***

- G_{TOT} the electron-hole generation rate, sum of all contributions participating in the generation mechanism, following Eq. ***I-***62:

$$G_{\mathrm{TOT}} = G_{\mathrm{SRH-TAT}} + G_{\mathrm{BTBT}} + G_{\mathrm{ionisation}}$$ ***Eq. I-62***

- $G_{\mathrm{ionisation}}$ the impact ionization generation rate, which essentially represents the avalanche breakdown carrier generation rate and is expressed by ***Eq. I-63***:

$$G_{\mathrm{ionisation}} = \alpha_{\mathrm{n}} n v_{\mathrm{e}} + \alpha_{\mathrm{p}} p v_{\mathrm{h}}$$ ***Eq. I-63***

- v_e and v_h the recombination velocities of electrons and holes respectively

In the P and N zones, without the presence of an electric field, the major contribution is the thermal one so the diffusion current is low. DCR_{N} and DCR_{P} are therefore rarely taken into account due to their values being negligeable compared to DCR_{SCR} which will be developed in the following paragraph.

5.3.2 *DCR* in the SCR

For the DCR that takes place in the depletion zone, thermally or band-to-band generated electron/hole pairs can generate an avalanche, of probability $P_p(z)$. It is expressed as follows:

$$DCR_{\mathrm{SCR}} = \int_{W_1}^{W_1+W_{\mathrm{ZCE}}} P_{\mathrm{p}}(z) G_{\mathrm{SRH-TAT}}(z) dz + \int_{W_1}^{W_1+W_{\mathrm{ZCE}}} P_{\mathrm{p}}(z) G_{\mathrm{BTBT}}(z) dz$$ ***Eq. I-64***

- G_{SRH-T} (z) being the SRH-TAT thermal generation rate following Eq. ***I-***65 [10], [29]:

$$G_{\mathrm{SRH-TAT}} = \frac{np - {n_{\mathrm{i_{eff}}}}^2}{\tau_{\mathrm{h}}(n + n_1) + \tau_{\mathrm{e}}(p + p_1)} \qquad \textbf{\textit{Eq. I-65}}$$

By considering $n_{\mathrm{i_{eff}}}$, the model takes into account the Band-gap narrowing (BGN) phenomena.

  - $n_{\mathrm{i_{eff}}} = n_{\mathrm{i}} \cdot \exp\left(\frac{E_{\mathrm{BGN}}}{2kT}\right)$ the effective intrinsic carrier density, E_{BGN} the band-gap-narrowing energy [48], with n_{i} defined by ***Eq. I-66***:

$$n_{\mathrm{i}} = \sqrt{N_{\mathrm{c}}(T)N_{\mathrm{v}}(T)} \cdot \exp\left(-\frac{E_{\mathrm{g}}(T)}{2kT}\right) \qquad \textbf{\textit{Eq. I-66}}$$

  - τ_{e} and τ_{h} minority carrier lifetimes of electrons and holes respectively

The lifetime of minority carriers $\tau_{\mathrm{n,p}}$ depends both on the physical parameters of the recombination centers and the electric field F [44] :

$$\tau_{\mathrm{n,p}} = \frac{1}{v_{\mathrm{th_{e,h}}} N_{\mathrm{T}} \sigma_{\mathrm{T}} (1 + \Gamma_{\mathrm{TAT}}(F))} \qquad \textbf{\textit{Eq. I-67}}$$

    - $\Gamma_{\mathrm{TAT}}(F)$ the Field enhancement parameter of Hurkx's model [49]
    - $v_{\mathrm{th_{e,h}}}$ thermal velocity of carriers [9]

$$v_{\mathrm{th_{e,h}}} = \sqrt{\frac{3kT}{m_{\mathrm{e,h}}}} \qquad \textbf{\textit{Eq. I-68}}$$

    - N_{T} recombination center/trap density
    - σ_{T} the cross section of recombination centers [50]

$$\sigma_{\mathrm{T}} = \sigma_{300} \cdot \exp\left(\frac{E_{\mathrm{A}}}{k}\left(\frac{1}{300} - \frac{1}{T}\right)\right) \qquad \textbf{\textit{Eq. I-69}}$$

  - With $n_1 = n_{\mathrm{i_{eff}}} \cdot \exp\left(\frac{E_{\mathrm{trap}}}{kT}\right)$ and $p_1 = n_{\mathrm{i_{eff}}} \cdot \exp\left(-\frac{E_{\mathrm{trap}}}{kT}\right)$

- $G_{\mathrm{BTBT}}(z)$ Band-to-band tunnelling generation rate

The presence of a very strong electric field F can pull electrons out of the valence band and into the conduction band by the tunnel effect.

This phenomenon is described by Hurkx's model [51] and expresses the band-to-band tunneling generation rate G_{BTBT} , following ***Eq. I-70***:

$$G_{\mathrm{BTBT}} = B|F|^{\frac{5}{2}} \cdot \exp\left(\frac{F_0}{F}\right) \qquad \textbf{\textit{Eq. I-70}}$$

- B a constant depending on electron-phonon interactions and
- With $F_0 = F_{corr}\left(-\frac{E_g(T)}{E_{g_{300}}}\right)^{\frac{3}{2}}$

It is possible to decorrelate the $DCR_{\mathrm{SRH-TAT}}$ and DCR_{BTBT} contributions depending on the temperature at which the device is situated, according to the following graph on ***Figure I-22*** [47]:

- At low temperatures (up to about 30K from room temperature), the DCR_{BTBT} is predominantly higher than $DCR_{\mathrm{SRH-T}}$, which thus represents the dark count floor as the DCR will never go below that DCR_{BTBT} baseline value.

- At high temperatures, (up above 300K, room temperature) the $DCR_{\mathrm{SRH-T}}$ contribution becomes dominant when compared to the DCR_{BTBT} noise floor.

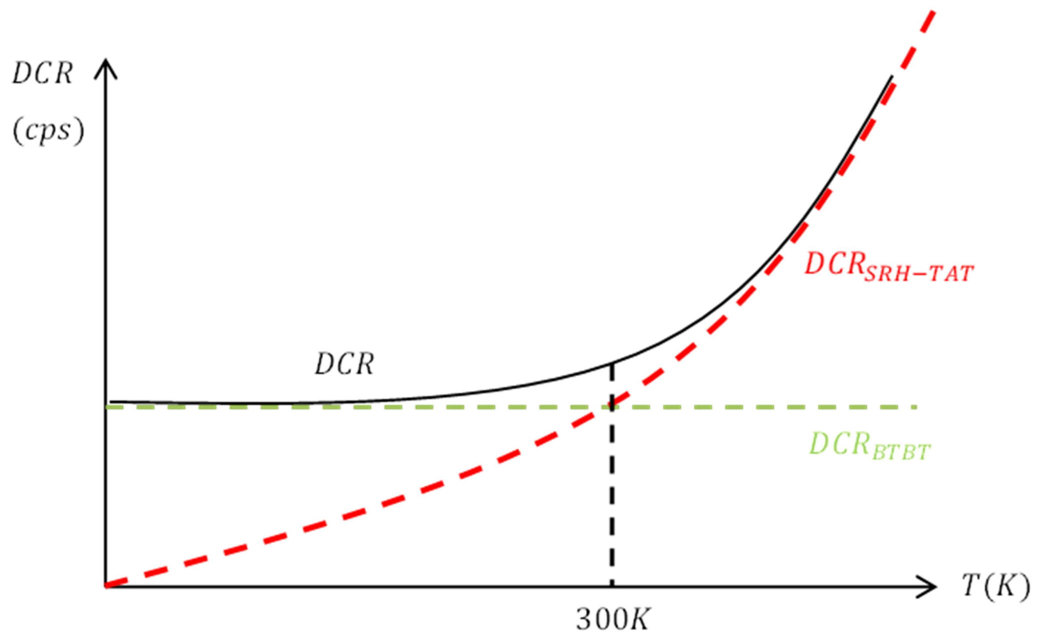

Figure I-22 Representation of the contributions DCR_{BTBT} and $DCR_{SRH-TAT}$ as a function of device temperature

5.4 Modeling Afterpulsing

5.4.1 DCR Afterpulsing contribution

A photogenerated avalanche can, because of the strong current peak it generates, cause afterpulsing by capturing a carrier in a trap that can be released afterwards. This phenomenon is highly dependent on dead time. On ***Figure I-23***, it is shown that when a carrier is released from its trap during the dead time of the device, afterpulsing is not detected. However, if the carrier is released outside the dead time, the latter will be, if it generates an avalanche, detected at the exit and will

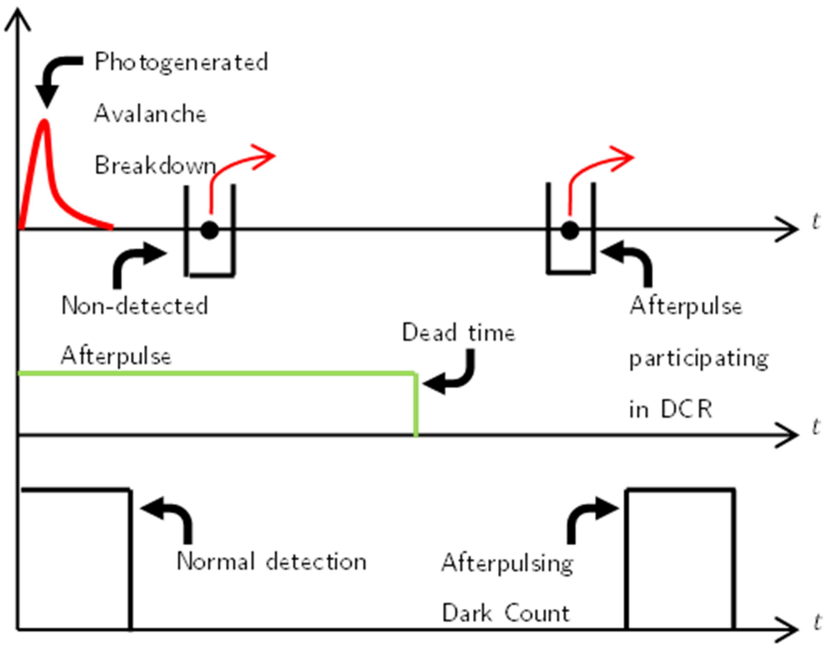

Figure I-23 Afterpulsing phenomenon (top axis) and its consequences on the digital output (bottom axis) as a function of the dead time (middle axis)

participate to the DCR.

We can easily deduce that the lower the dead time is, the higher the probability of afterpulsing. For a dead time $\Delta t_{\mathrm{m}} = 1\,\mu s$, the probability of afterpulsing is very low or almost null [52], yielding ***Eq. I-71***:

$$\Delta t_{1\mathrm{s}} = DCR \cdot \Delta t_{\mathrm{m}} \qquad \textbf{\textit{Eq. I-71}}$$

- $\Delta t_{1\mathrm{s}}$ the total duration of dead time in one second

- DCR the measurable Dark Count Rate ($DCR = DCR_{\mathrm{SRH-T}} + DCR_{\mathrm{BTBT}} + DCR_{\mathrm{Afterpulsing}}$)
- Δt_{m} the dead time caused by the extinction circuitry

Since the measurable Dark Count Rate DCR does not take into account the events hidden by the dead time and thus not detected, therefore, the DCR has to be expressed as a function of the total dead time in one second $\Delta t_{\mathrm{m}_{1\mathrm{s}}}$ and DCR :

$$DCR_{\Delta t_{1\mathrm{s}}} = DCR \cdot \Delta t_{1\mathrm{s}} = DCR^2 \cdot \Delta t_{\mathrm{m}} \qquad \textbf{\textit{Eq. I-72}}$$

The real Dark Count Rate DCR_{real} taking into account the counts hidden by the dead time, $DCR_{\Delta t_{1\mathrm{s}}}$, follows ***Eq. I-73***:

$$DCR_{\mathrm{real}} = DCR + DCR_{\Delta t_{1\mathrm{s}}} = DCR(1 + DCR \cdot \Delta t_{\mathrm{m}}) \qquad \textbf{\textit{Eq. I-73}}$$

It is the real Dark Count Rate DCR_{real} which defines the Dark Count Rate Afterpulsing contribution $DCR_{\mathrm{afterpulsing}}$ as a function of the Afterpulsing probability P_{a} with the following equation:

$$DCR_{\mathrm{afterpulsing}} = DCR_{\mathrm{real}} \cdot P_{\mathrm{a}} \qquad \textbf{\textit{Eq. I-74}}$$

The afterpulsing phenomenon can be broken down into four distinct stages:

- In the first place, there must be an impact ionization caused avalanche breakdown, generated by a photon or a dark count.
- An avalanche carrier must be trapped by a recombination site.
- This carrier is then released with a certain delay depending on the characteristics of the trap and driven by the strong electric field.
- Finally, if the SPAD has been recharged before the carrier is released and the carrier generates an avalanche, this avalanche will be detected by the device and will contribute to the DCR.

5.4.2 Modeling of carrier trapping by recombination centers

As previously explained, following a first avalanche, correlated secondary avalanches can occur. These can have different activation energies E_{A} and time constants τ depending on the physical characteristics of the traps. The electron capture rate $U_{\mathrm{capt,ni}}$ can be expressed by ***Eq. I-75*** [11]:

$$U_{\mathrm{capt,ni}}(t) = N_{\mathrm{tni}}(1 - f_{\mathrm{tni}})\frac{j_{\mathrm{e}}(t)}{q}\sigma_{\mathrm{ni}} \qquad \textbf{\textit{Eq. I-75}}$$

- N_{tni} the density of traps in the band-gap on level i
- $(1 - f_{\mathrm{tni}})$ the fraction of unoccupied electron traps on level i at the beginning of the avalanche
- $j_{\mathrm{e}}(t)$ electron current density in the depleted region SCR
- σ_{ni} electron capture cross section

After a time t, at the end of the avalanche, the number of electrons captured by the traps of level i is defined with ***Eq. I-76*** [11]:

$$\Lambda_{\mathrm{ni}} = \int_0^t U_{\mathrm{capt,ni}}(t)dt = N_{\mathrm{tni}}(1 - f_{\mathrm{tni}})\frac{N_{\mathrm{av}}}{S_{\mathrm{SPAD}}}\sigma_{\mathrm{ni}} \qquad \textbf{\textit{Eq. I-76}}$$

- N_{av} the number of electrons generated during an avalanche, follows ***Eq. I-77***:

$$N_{av} = \frac{C \cdot V_{ex}}{q} \qquad \textbf{\textit{Eq. I-77}}$$

  - V_e the excess bias voltage, depending on the reverse bias voltage V_{bias} of the device:

$$V_{ex} = V_{bias} - V_{br} \qquad \textbf{\textit{Eq. I-78}}$$

- C the capacity, sum of the parasitic capacities of the extinction and treatment circuit and the internal capacity of the SPAD C_{SPAD}.

$$C_{\mathrm{SPAD}} = \varepsilon_0 \varepsilon_{\mathrm{r}} \frac{S_{\mathrm{SPAD}}}{W_{\mathrm{SCR}}} \qquad \textbf{\textit{Eq. I-79}}$$

- S_{SPAD} the considered area of the SPAD
- σ_{ni} electron capture cross section

Following the same reasoning, similar formulas can be obtained for holes.

To be able to decrease the probability of afterpulsing, by analyzing the previous equations, it appears that the capacity of the SPAD has to be decreased, thus decreasing the surface S_{SPAD} and increasing the SCR width W_{SCR} of the device. It is also necessary to reduce as much as possible the parasitic capacities of the circuitry surrounding the SPAD, which is why more and more SPADs are co-integrated with CMOS circuits.

5.4.3 Modeling of trapped carrier release

Let R_{ni} be the concentration of electrons released by a trap of energy E_{tni} after the current peak caused by the avalanche breakdown:

$$R_{\mathrm{ni}} = \Lambda_{\mathrm{ni}} \cdot \eta_{\mathrm{i}} \qquad \textbf{\textit{Eq. I-80}}$$

- η_i the release probability of electron by a trap of energy E_{tni}

$$\eta_{\mathrm{i}} = \langle e_{\mathrm{ni}} \rangle_{\mathrm{HF}} \cdot \tau_{\mathrm{ni}} \qquad \textbf{\textit{Eq. I-81}}$$

  - $\langle e_{\mathrm{ni}} \rangle_{\mathrm{HF}}$, as described in ***Eq. I-82*** is the average probability for an electron inside the bandgap to be released inside the high electric field region (SCR) with $\Gamma_{\mathrm{TAT}}(F)$ [51] the Hurkx field, ***Eq. I-83*** enhancement parameter which can also be used in the TAT DCR modeling.

$$\langle e_{\mathrm{ni}}\rangle_{\mathrm{HF}} = n_{\mathrm{i}}(1+\Gamma_{\mathrm{TAT}}(F))\cdot\exp\left(\frac{E_{\mathrm{i}}-E_{\mathrm{tni}}}{kT}\right)\langle c_{\mathrm{ni}}\rangle \qquad \textbf{\textit{Eq. I-82}}$$

$$\Gamma_{\mathrm{TAT}}(F) = 2\sqrt{3\pi}\frac{|F|}{F_\Gamma}\cdot exp\left(\frac{|F|^2}{{F_\Gamma}^2}\right), \qquad F_\Gamma = \frac{\sqrt{24m_{\mathrm{t}}(kT)^3}}{qh} \qquad \textbf{\textit{Eq. I-83}}$$

- τ_{ni} the life time of a trap of energy E_{tni}, the average time between 2 successive captures, as expressed by ***Eq. I-84*** [53]. τ_{ni} increases when the temperature T decreases and decreases the more E_{tni} is near E_C or E_V.

$$\tau_{\mathrm{ni}} = \tau_0\cdot\exp\left(\frac{E_{\mathrm{tni}}}{kT}\right), \qquad \tau_0 = 2\cdot 10^{-7}\cdot\exp\left(-\frac{0.267}{kT}\right) \qquad \textbf{\textit{Eq. I-84}}$$

6 SPAD in CMOS technology

Since the conception of the first SPAD, there have been many different SPADs that have been developed and manufactured. Two main architectures stand out:

- Diffusion guard ring and isolation
- Virtual guard ring and isolation

6.1 Diffusion guard ring and isolation

- P-Well diffusion guard ring for the P+ version of the diode
- N-Well type guard ring for the N+ version of the diode

This architecture, seen on ***Figure I-24*** [29] is adopted by many SPAD manufacturers and scientific researchers.

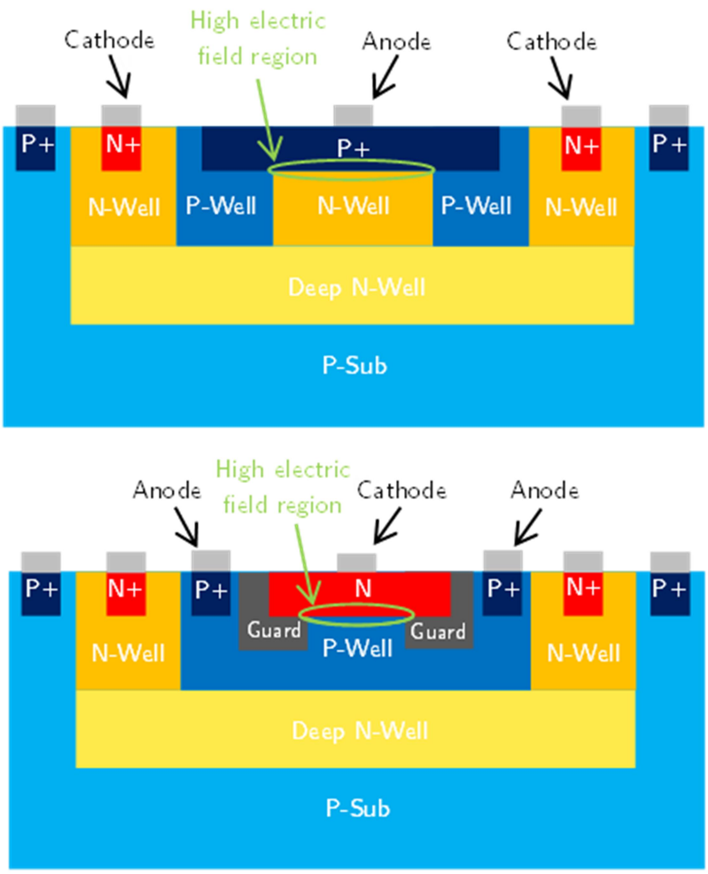

Figure I-24 SPAD architecture with diffusion guard ring (P-type guard ring on the left and N-type guard ring on the right)

In this type of architecture, the active region circled in green is situated buried in the structure beneath the highly dope upper layer (either P- or N-doped) with a W-shaped carrier flow through the N-Well and Deep N-Well. Isolation of this active region is either assured by guard rings or additional P-Wells between the N-Well of the active region and the outer ones.

6.2 Virtual guard ring and isolation

This type of alternative architecture, ***Figure I-25***, uses a guard ring which is nothing more than the P-Substrate with no additional P- or N-Well isolation implants around trhe central active region implant. The SCR is deeper than with a diffusion guard ring, as now the active region is situated between the P-Well and the Deep N-Well.

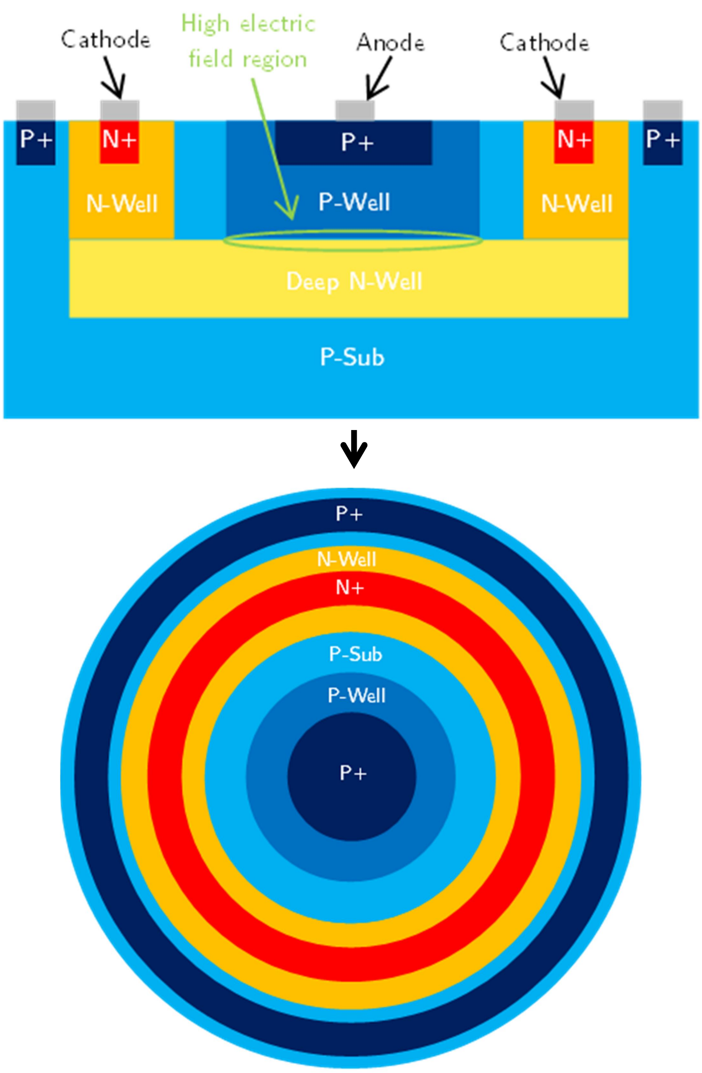

Figure I-25 SPAD architecture with virtual guard ring (left) and top view (right)

Due to the deep location of the virtual guard ring as pictured on ***Figure I-25*** [29], it can surround the multiplication region in order to prevent leakage of the electric field while also maintaining a low doping concentration of the P-Substrate compared to the conventional diffusion type guard ring. [54]

6.3 Architecture improvements

Some adjustments can be made to basic architectures in order to improve effects such as electric field leakage and crosstalk.

6.3.1 Shallow Trench Isolation (STI) and Deep Trench Isolation (DTI)

The use of Shallow Trench Isolation [29], or STI, is easily implementable in the fabrication process and provides both advantages and drawbacks.

STI consists of trenches, as seen on ***Figure I-26***, generally filled with a thick layer of undoped polysilicon or silicon dioxide (SiO_2). This causes a lot of defects because the interface between the STI and monocrystal silicon is very poor. Therefore, it often leads to high DCR and Afterpulsing probability which can be devastating for certain application with low noise tolerances.

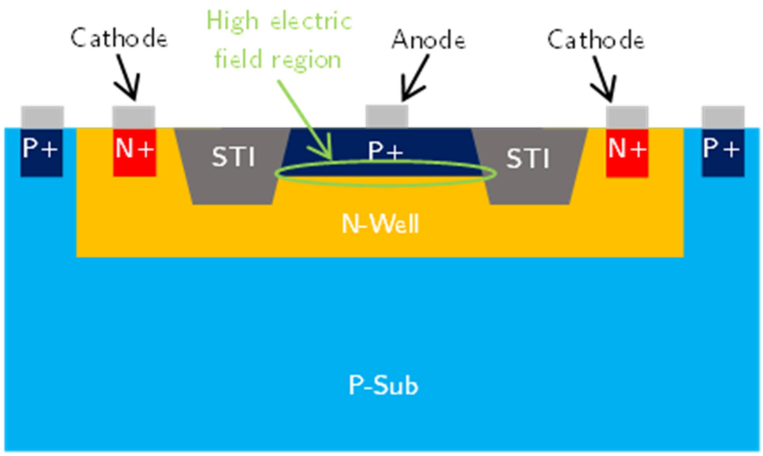

Figure I-26 SPAD Architecture with Diffusion Guard Ring and Shallow Trench Isolation (STI)

On the flipside, thanks to STI, the electric field between anode and cathode is prevented from being too high, preventing electrical current leakage between those terminals.

The SPAD studied in this manuscript, designed in STMicroelectronics' architecture, uses a derivative of STI which is a deeper alternative hence its name: Deep Trench Isolation, or DTI.

DTI is mainly used to fully isolate each pixel of the array from the surrounding ones, as the deep isolation wells create a physical and electrical barrier between each SPAD pixel. ***Figure I-27*** [55] below shows the general DTI placement in the architecture, on the outer ends of the pixel next to the outer electrode's doping regions. This greatly reduces electrical and optical crosstalk, as the isolation is impervious to both photons and carriers (electrons or holes).

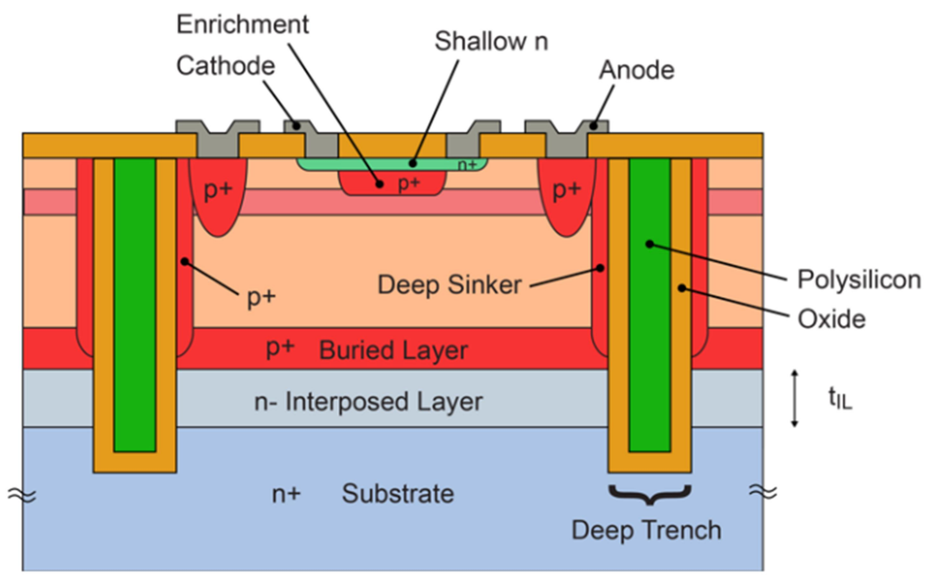

Figure I-27 Deep Trench Isolation (DTI) in a SPAD architecture

6.3.2 Optimized implants

Another improvement frequently integrated in SPAD designs is optimized implants, which consist in additional doping regions aiming to improve the device's behavior as seen in ***Figure I-28***.

In this architecture improvement, the doping of different areas is essential. The doping of the N/P-Well VGR (virtual guard ring) zone must be higher than that of the standard N/P-Well zone. This allows to concentrate the carrier multiplication at the interface (N/P-Well VGR) // (N/P+). The downside is the high doping concentration which can enhance Dark counts and Afterpulsing probability.

Therefore, the N/P-Well zones should not be too heavily doped, as this may create a strong electric field outside the active zone and a weaker field inside the active zone.

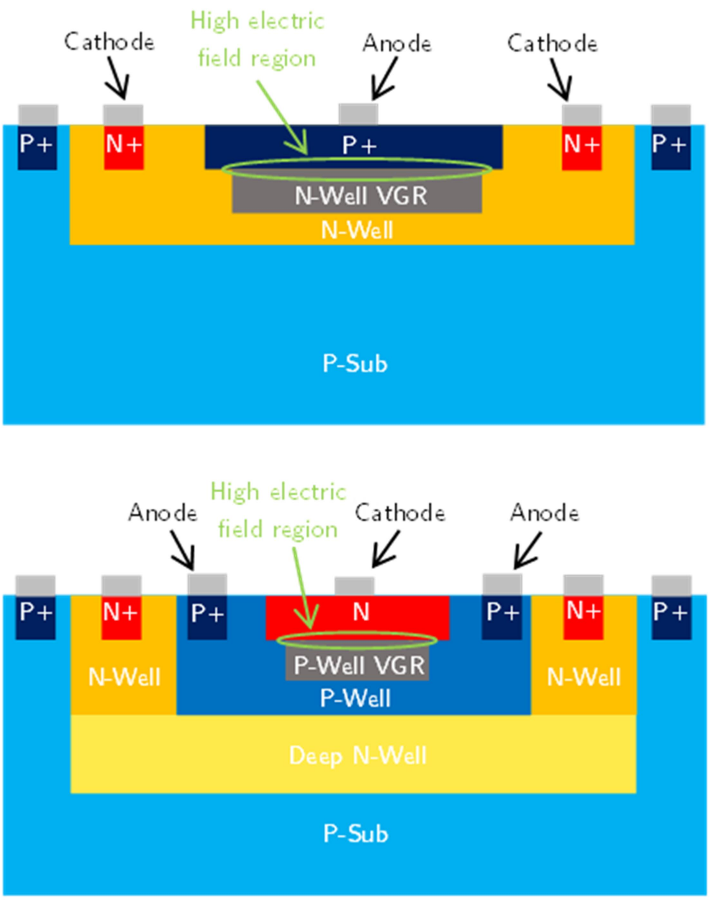

Figure I-28 SPAD architecture with virtual guard ring (P-type virtual guard ring on the left and N-type on the right)

7 Quenching circuits

Once the SPAD has triggered an avalanche after the detection of a photon or a dark count, it must be recharged in order to detect again. An electric circuit, also called a "quenching circuit" is responsible in insuring the recharge of the SPAD and to enable retriggering. It can be of two types [45] :

- Passive quenching
- Active quenching

7.1 Passive quenching

Passive quenching is achieved with passive components (mainly resistors). When an avalanche occurs, the avalanche current flows through the resistor, thus increasing the voltage drop across it. Therefore, voltage drop across the SPAD decreases to a point where the current drops low enough to decrease the ionization probability until it becomes negligible. At this point, the avalanche extincts and the device is recharged after an RC time constant depending on the quenching circuit used. If the circuits on ***Figure I-29*** are considered [56], the time constant is mainly dependent on R_{L} the ballast biasing resistor, C_{s} the stray capacitance (capacitance to ground of SPAD terminal connected to R_{L}) and both R_{d} and C_{d} the diode resistance (depending on the junction and well resistances) and the junction capacitance.

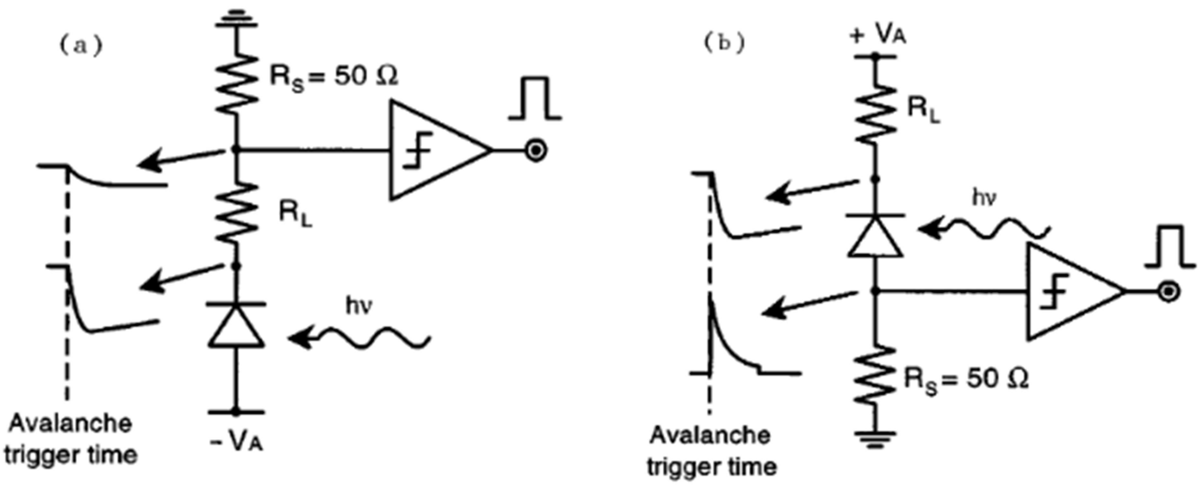

Figure I-29 Passive quenching circuits, (a) Cathode quenched, (b) Anode quenched

The quenching time constant, which will determine the quenching dynamics of the circuit, T_{q} is expressed by the following equation ***Eq. I-85***:

$$T_{\mathrm{q}} = (C_{\mathrm{d}} + C_{\mathrm{s}})\frac{R_{\mathrm{d}}R_{\mathrm{L}}}{R_{\mathrm{d}} + R_{\mathrm{L}}} \qquad \textbf{\textit{Eq. I-85}}$$

7.2 Active quenching

As opposed to passive quenching, active quenching is based on the use of passive and active components to control the avalanche current of SPADs. On the left [57] of ***Figure I-30*** below is a simple quenching circuit with a PMOS current mirror and an external resistor to set the current. Q_P operates in saturation mode as a constant current source (no voltage drop across). When an avalanche occurs, the voltage drop across the SPAD drops significantly. Q_P then operates in linear mode and if its internal resistance is high enough, the voltage across the SPAD will drop below its breakdown voltage, extincting the avalanche and resetting the device. This is what is called active quenching.

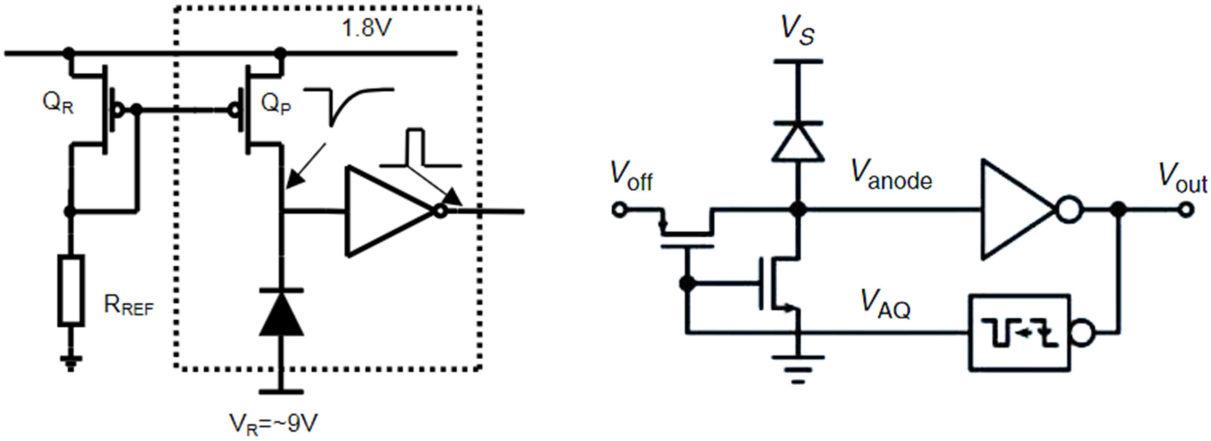

Figure I-30 Active quenching circuits, (left) current mirror quenching, (right) retroactive feedback quenching

Other types of active quenching circuits exist such the one as on the right [45] of ***Figure I-30***, where a retroaction triggers the SPAD quenching (in this case on falling edge) in order to reduce its recharge time to enable faster photon redetection.

More complex active quenching circuits can be designed to further reduce the recharge time of the SPAD and optimize the overall responsiveness, consumption and performance of imagers using avalanche diodes.

8 Measuring techniques

8.1 Photon detection efficiency

As previously explained in Part I, Chapter 5, Photon Detection Efficiency is a crucial figure of merit of SPADs. Therefore, being able to establish measuring techniques is mandatory to determine the effectiveness and performance of the device. Additionally, those measurements help tuning developed SPAD models to improve their accuracy.

In order to be able to measure the Photon Detection Probability and thus Efficiency, or PDP and PDE, the following measurement setup on ***Figure I-31*** [58] below can be used.

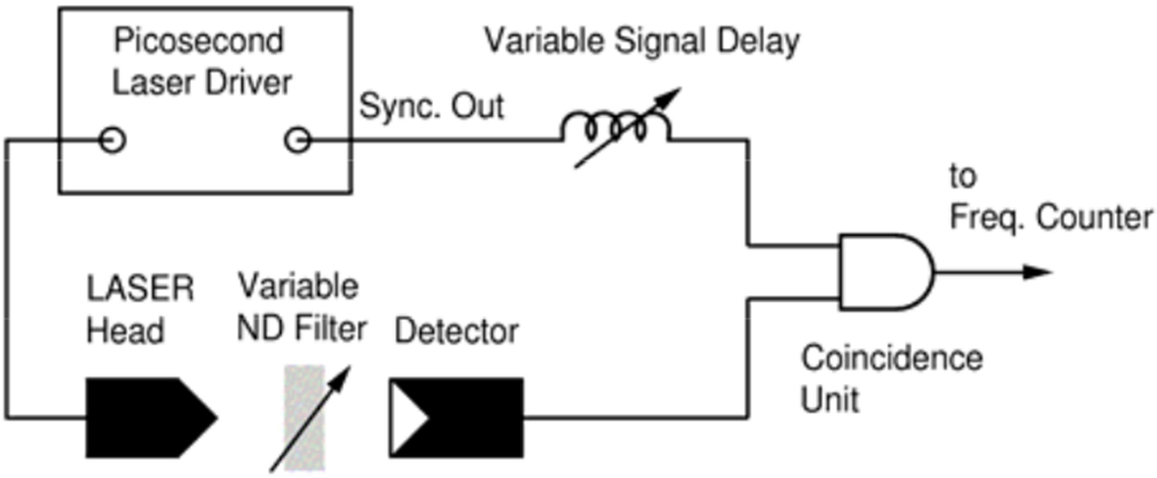

Figure I-31 Test setup of PDP measurement

A picosecond Laser is used to generate precise laser pulses in order to detect photons in a tight coincidence window (approximately 5ns), therefore detecting only incoming photons suppressing both Afterpulsing and noise. Photon detection efficiency is defined as the ratio between the average frequency of detected photons and the average frequency of incoming photons. The latter is obtained with a photodetector of known PDE and is kept constant for the following measurements.

8.2 Afterpulsing

As it is the case with the PDE measurement, measuring and quantifying Afterpulsing is critical to provide valuable information on the device and help adjusting SPAD models accordingly.

The Afterpulsing measurement setup, as seen in ***Figure I-32*** [58], uses exactly the same base setup as for PDE measurement, but with the only difference being the output stage:

- Output B is the same output as in the PDE setup and is counting the amount of incoming photons detected by the SPAD.
- Output A uses photon detection as a trigger for a 500ns interval generated by a pulse generator, interval during which the events happening between two "real" photon detections (afterpulses and dark counts) are counted.

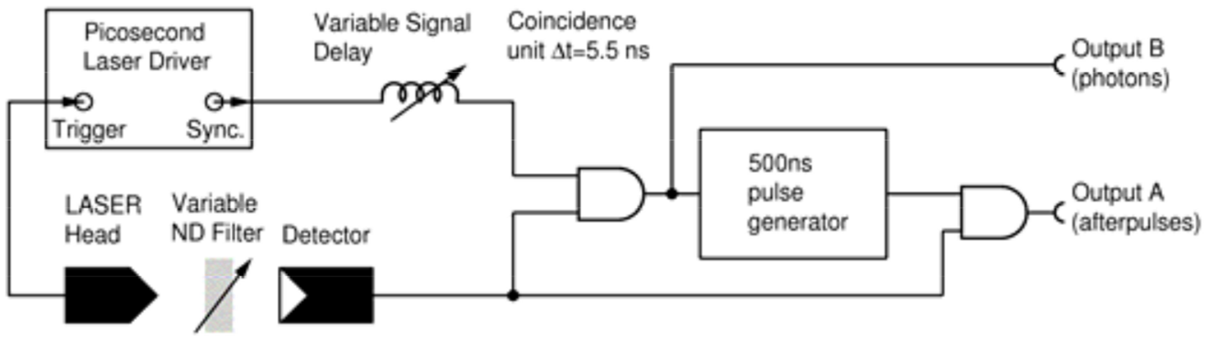

Figure I-32 Setup for Afterpulsing measurement

8.3 Jitter

A final key figure of merit that requires specific measurement techniques is jitter. The timing resolution of SPAD devices and its responsiveness are both responsible for the difficulty in measuring jitter, as the time constants can be as low as a few hundreds of picoseconds. ***Figure I-33*** [59] below shows a setup for Jitter measurements.

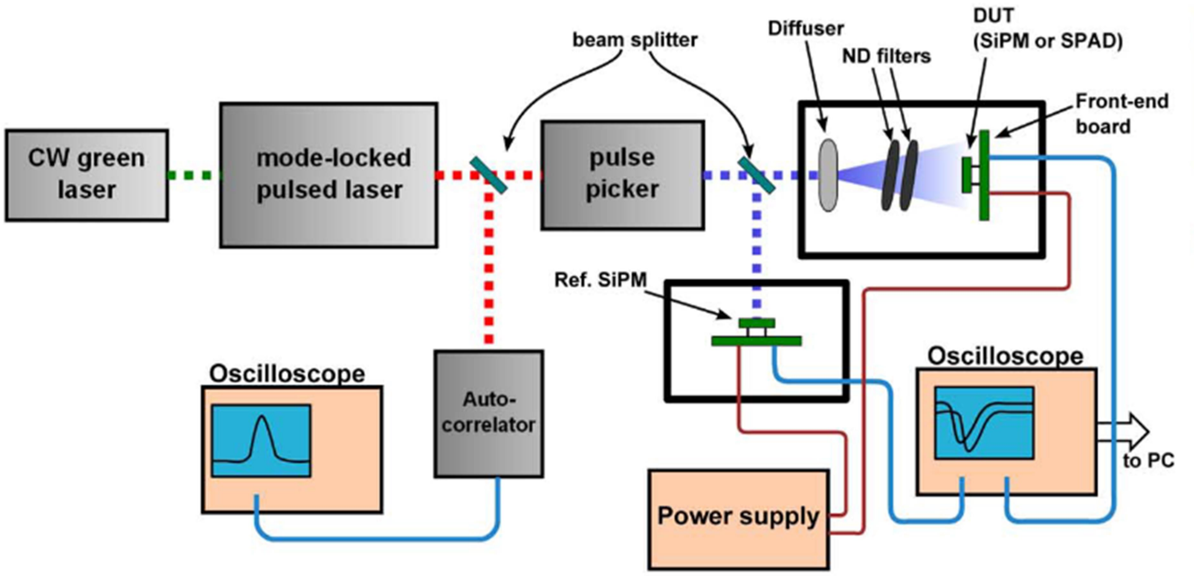

Figure I-33 Setup for Jitter measurement

The continuous wave laser pumps the mode locked laser to provide extremely narrow laser pulses (picoseconds). The autocorrelator is there to control the pulse shape. As the frequency of this laser setup is higher than the typical recharge time of SPADs, a pulse picker synchronized with the laser is used to only let 1 in 100 pulses pass. This way, the dead time between two pulses is high enough to measure the time between the photon emission and its detection by the SPAD.

Synthesis

During the bibliography, we saw that the various modelling approaches used to describe the different phenomena of the SPAD were often linked to complex and behavioral functions and exhaustive methods. We saw that the impact ionization mechanism is central to the device's behavior, but the descriptions of this phenomenon fail to capture the mechanism's workings easily and accurately. As the modelling descriptions are quite far away from a physical description of the SPAD behavior, the aim of this thesis was to develop a model that is as close to the physical mechanisms as possible.

From what was seen in the bibliography, certain aspects were kept as they corresponded to some extent to the modelling approach developed during this thesis. For example, the idea of the transit time of carriers and its link to the dead space was used to quantify the transition between the nodes of the model that will be described in the next chapters. Added to that, the knowledge of the causes and consequences of different figures of merit such as DCR, Jitter and Dead time is crucial to understand and better model the SPAD's behavior. Some notions, however were sidelined. Even though the modelling methods presented in **Subchapter 5** are an interesting approach on building a SPAD model, they presented evident flaws, mainly linked to a mathematically heavy approach which can be very abstract. Furthermore, as mentioned in the previous paragraph, those approaches do not aim to simulate the transient analog behavior, which is precisely where our model inserts itself in the state of the art of SPAD models.

The most important thing to note in the state of the art of SPAD modelling, as introduced in the previous paragraph, is the lack of comparison with transient measurements of the device. In the context of both this PhD and the associated IPCEI project, those transient measurements presented in **Chapter II** and in [60] are central, as they permitted to get invaluable insights on the transient behavior of the SPAD. Thanks to the characterization data that is acquired during this project, we were able to design a compact model that can accurately reproduce the behavior of the device.

In the next chapter, the notions seen in this bibliography are implemented in a modelling approach that aims to achieve a spatially distributed model. This model is capable of physically modelling the avalanche breakdown mechanism with a loop-based approach, be adjusted on both digital and transient analog characterization data and yields good fits to the device's responses.

Chapter II Single Loop SPAD Model

Introduction

The model developed in this manuscript was based on the idea of a spatialized SPAD model. This spatialization is translated into a combination of elementary loops and each single loop element bases itself on the recurring nature of the impact ionization phenomenon. In this first chapter, the outline of the base model consisting of a single loop, will be described. The resulting model is implemented into an existing model structure from STMicroelectronics and sets itself in the role of the avalanche current I_{avl}.

The SPAD's current is composed of four different current contributions that are each following their own equations. ***Eq. II-1*** is the expression of the total current as the sum off all four individual contributions.

$$I_{\mathrm{tot}} = I_{\mathrm{for}} + I_{\mathrm{rec}} + I_{\mathrm{rev}} + I_{\mathrm{avl}}$$

Eq. II-86

The first current contribution is the direct current source that is derived from Schokleys equations as ***Eq. II-2*** states.

$$I_{\mathrm{for}} = I_{\mathrm{s_{for}}} \cdot \left(\exp\left(\frac{V_{\mathrm{D}}}{n.\,U_{\mathrm{T}}}\right) - 1\right)$$

Eq. II-2

- n the diode's ideality factor
- $I_{\mathrm{s_{for}}}$ the forward saturation current
- U_{T} the thermodynamic potential at equilibrium

The second contribution, ***Eq. II-3***, is the recombination current source which as the name states is caused by the recombination of the carriers inside the silicon.

$$I_{\mathrm{rec}} = I_{\mathrm{s_{rec}}} \cdot \left(\exp\left(\frac{V_{\mathrm{D}}}{n_{\mathrm{r}}.U_{\mathrm{T}}}\right) - 1\right) \qquad \textbf{\textit{Eq. II-3}}$$

- n_{r} the recombination ideality factor
- $I_{\mathrm{s_{rec}}}$ the recombination saturation current

The third contribution is the reverse current source and uses a simplified model as ***Eq. II-4*** shows.

$$I_{\mathrm{rev}} = I_{\mathrm{s_{rev}}} \cdot \left(\exp\left(\frac{-V_{\mathrm{D}}}{n_{\mathrm{rev}}}\right) - 1\right) \qquad \textbf{\textit{Eq. II-4}}$$

- n_{rev} the reverse current voltage dependancy factor
- $I_{\mathrm{s_{rev}}}$ the reverse current prefactor

The last contribution I_{avl}, as previously mentioned, is the contribution of the avalanche current. This current is the object of this modelling approach as it is the most important contribution in the total SPAD current, due to the complex nature of the impact ionization phenomenon. The model therefore integrates itself in the role of the avalanche current I_{avl} in the total SPAD current of ***Eq. II-1***.

The main goal of this modeling approach is to exploit the transient measurements made on the device and use this information to calibrate the transient response of the model for an accurate description of the device's behavior.

1 Translation of Impact Ionization Phenomena to a Looped Model Architecture

1.1 Device architecture

The spatialized compact modeling approach employed for the device is intricately tied to its intrinsic architecture, which plays a pivotal role in its overall functionality and performance. An initial crucial step in the modeling process involved extracting essential information pertaining to the structural configuration of the Single-Photon Avalanche Diode (SPAD), as depicted in ***Figure II-1.***

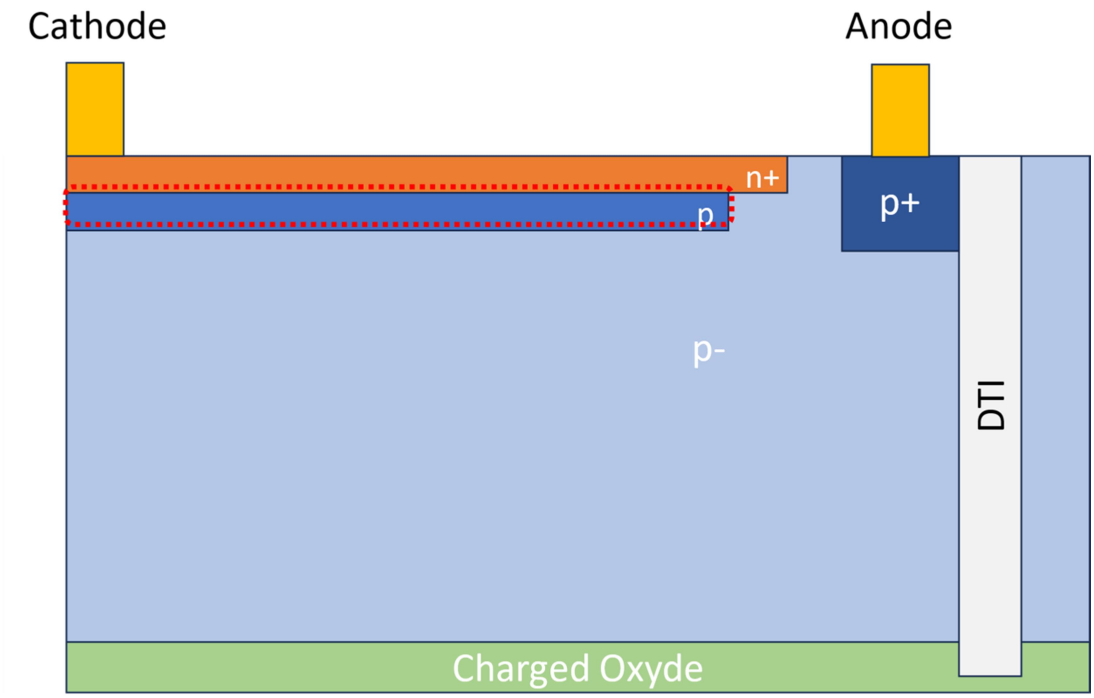

Figure II-1 Simplified cross-section of SPAD device with all different doping regions, isolations (DTI). Multiplication region is highlighted in the red dotted line.

The SPAD structure comprises distinct doping regions that are strategically implemented to achieve desired characteristics and operational efficiency. The cathode contact, denoted as the n+ region, is electrically connected to the device, while the anode contact is similarly connected to the p+ region. Sandwiched between these two regions is a weakly doped p- region, which is assimilated to a quasi-intrinsic layer. This unique arrangement of the n+/p-/p+ stacking within the SPAD structure is a defining characteristic of a PIN diode architecture.

Furthermore, an interesting feature within the SPAD structure is the presence of the n+/pwell junction, which serves as a transition point before transitioning into the classic PIN layer stacking. This particular junction configuration is indicative of a reach-through type architecture. The incorporation of this reach-through architecture contributes to enhanced device performance and enables specific operational advantages in terms of charge carrier collection and avalanche breakdown.

By carefully examining and characterizing the SPAD's intrinsic architecture, valuable insights into its underlying principles and operational behavior can be obtained. This comprehensive understanding of the structural components and their interplay facilitates the development of accurate and predictive compact models. These models enable efficient analysis and design optimization of SPAD devices for a wide range of applications, including photon counting, time-resolved imaging, and quantum communication.

This specific type of diode exhibits several key distinctions when compared to its PIN counterpart, demonstrating unique characteristics and behavior. Firstly, the electric field distribution within the Space Charge Region (SCR) is notably more localized compared to a conventional PIN diode. Instead of a uniform electric field across the entire SCR, the electric field experiences a significant variation. Particularly, the electric field attains exceptionally high values at the n+/pwell junction, where the junction is narrowest, before gradually decreasing to lower values within the remaining depletion area. This concentrated high electric field at the narrow junction region allows for the occurrence of carrier multiplication through impact ionization, thus earning this region the name of the multiplication region.

Conversely, in the considerably wider low electric field region of the SCR, the electric field strength is not sufficient for the carriers to reach their impact ionization energy threshold. Carriers injected or generated within this region predominantly undergo drift either directly towards the cathode (in the case of holes) or towards the multiplication region. This expansive region, characterized by its larger dimensions and relatively weaker electric field, serves as the primary site for absorbing incident photons, consequently earning it the designation of the absorption region. Its substantial width enables effective absorption of a greater number of photons, enhancing the device's sensitivity to light. ***Figure II-2*** represents the cross section that was established to develop the corresponding SPAD model.

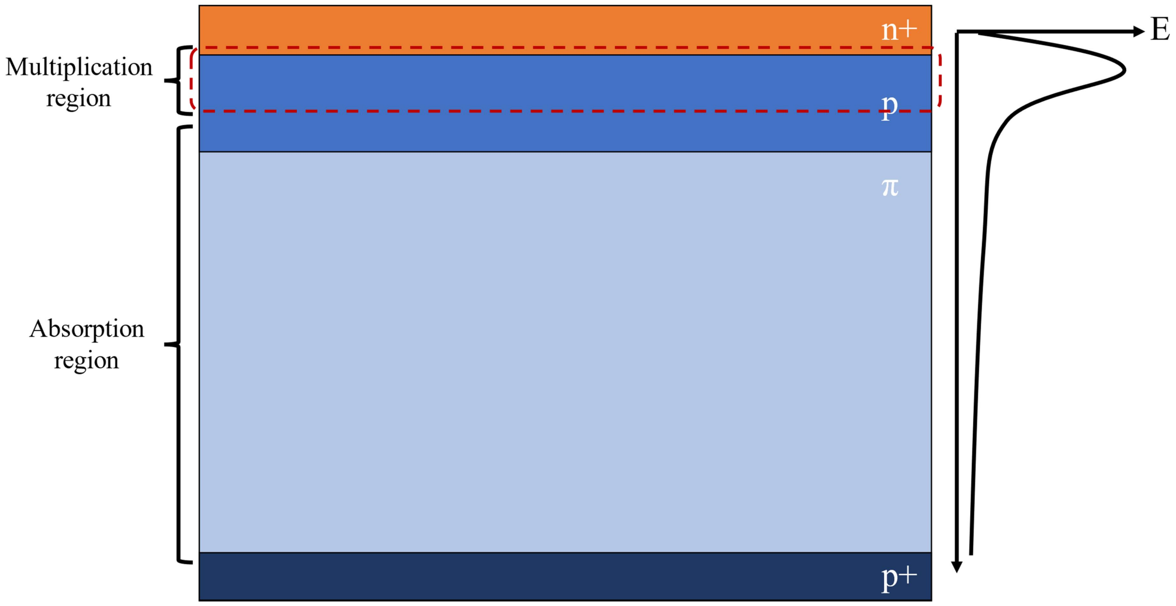

Figure II-2 Simplified cross-section of a n+/p/π/p+ reach-through SPAD and the corresponding electric field distribution (right) inside the SCR

These distinct electric field characteristics and the resulting localized electric field distribution across the SCR significantly differentiate this diode from its PIN counterpart. The presence of the multiplication region, with its high electric field concentration, enables efficient carrier multiplication through impact ionization, contributing to the device's avalanche breakdown behavior and its ability to detect and amplify weak optical signals. Conversely, the absorption region, encompassing a wider low electric field region, facilitates efficient photon absorption and promotes the conversion of incident photons into measurable electrical signals.

By comprehensively understanding and analyzing the electric field distribution within the SCR, we can gain valuable insights into the underlying mechanisms and performance attributes of this unique diode structure.

In summary, the specific characteristics of this diode, including the localized electric field distribution, the presence of a multiplication region with a high electric field concentration, and the wider absorption region, contribute to its distinct behavior and functionality. These features enable efficient carrier multiplication, avalanche breakdown, and enhanced photon absorption, making it a valuable component for various optoelectronic applications demanding high sensitivity and performance.

1.2 Impact ionization modeling approach

The foundation of the model primarily revolves around the fundamental physical phenomenon of impact ionization. By leveraging the base mechanism of carrier multiplication through impact ionization, the model offers a physics-centric approach to describing the behavior of the SPAD. This deliberate modeling choice aims to minimize reliance on arbitrary fitting equations, enabling a more accurate representation of the real phenomenon with minimal adjustment parameters.

- The photo-generation of an electron/hole pair within the SCR of the SPAD serves as the initial step in the carrier multiplication process. This photo-generation occurs when incident photons impact the device, resulting in the generation of an electron and a hole pair within the SCR.
- Subsequently, the carriers are propelled by the electric field present within the SCR, progressively gaining energy as they accelerate. This acceleration facilitates their movement towards the surrounding silicon (Si) atoms.
- Upon colliding with the Si atoms, if the electron or hole possesses an energy equal to or greater than the ionization threshold energy, it can dislodge a valence band electron, liberating it from its atomic bond. The liberated electron is then accelerated and has the potential to generate additional

electron/hole pairs upon subsequent impacts, leading to an exponential multiplication of electrons (and holes) within the device.

To encapsulate this ionization mechanism and the cyclic nature of the physical phenomenon, the model employs a loop architecture. This design choice mirrors the inherent loop-like behavior observed in the impact ionization process.

The ionization coefficients, denoted as $\alpha_{\mathrm{n,p}}$, which are inversely related to a characteristic distance known as the dead space $\alpha_{\mathrm{n,p}}{}^{-1}$, are critical parameters within the model. These coefficients depend on the local electric field F_{m} and are mathematically defined by ***Eq. II-5*** and ***Eq. II-6***:

$$\alpha_{\mathrm{n,p}} = A_{\mathrm{n,p}} \cdot exp\left(\frac{-B_{\mathrm{n,p}}}{F_{\mathrm{m}}}\right) \qquad \textbf{\textit{Eq. II-5}}$$

$$B_{\mathrm{n,p}} = C_{\mathrm{n,p}} + D_{\mathrm{n,p}}T \qquad \textbf{\textit{Eq. II-6}}$$

Where A_{n} = 4.43·10^{5} cm^{-1}, A_{p} = 1.13·10^{6} cm^{-1}, C_{n} = 9.66·105 V·cm^{-1}, C_{p} = 1.71·10^{6} V·cm^{-1}, D_{n} = 4.99·10^{2} V·cm^{-1}·K^{-1}, D_{p} = 1.09·10^{3} V·cm^{-1}·K^{-1} and T (in °K) the absolute temperature and F_{m} is the electric field inside the multiplication region of the SCR [25].

By incorporating the local field information and the corresponding ionization coefficients, the model captures the nuanced relationship between the electric field and carrier multiplication, enabling a more precise and comprehensive representation of the SPAD's performance. By adopting this physics-centric approach and embracing the impact ionization phenomenon, the model achieves a more realistic and accurate depiction of the SPAD's behavior.

In conclusion, the model builds upon the physical principles underlying impact ionization, utilizing carrier multiplication as the core mechanism to describe the SPAD's behavior. By faithfully capturing the steps of the ionization process within a loop architecture and incorporating the ionization coefficients dependent on the local field, the model offers a physics-centric and data-driven approach to modeling the

SPAD's performance. This advanced modeling framework facilitates more accurate simulations, enabling the development of high-performance SPAD devices tailored to specific application requirements.

2 Single-loop model principle and concept

2.1 Model structure

Considering the cyclic nature of the underlying physical phenomenon, as detailed in **Chapter II**, **Subchapter 1.2**, the compact model is structured as a loop, providing a comprehensive representation of the SPAD behavior, as depicted in ***Figure II-3***. This loop-based architecture comprises two distinct paths, representing the flows of electrons (indicated in green) and holes (depicted in red) within the device. Importantly, this internal structure is solely connected to the main model through two output nodes, which serve as the endpoints of the Space Charge Region (SCR).

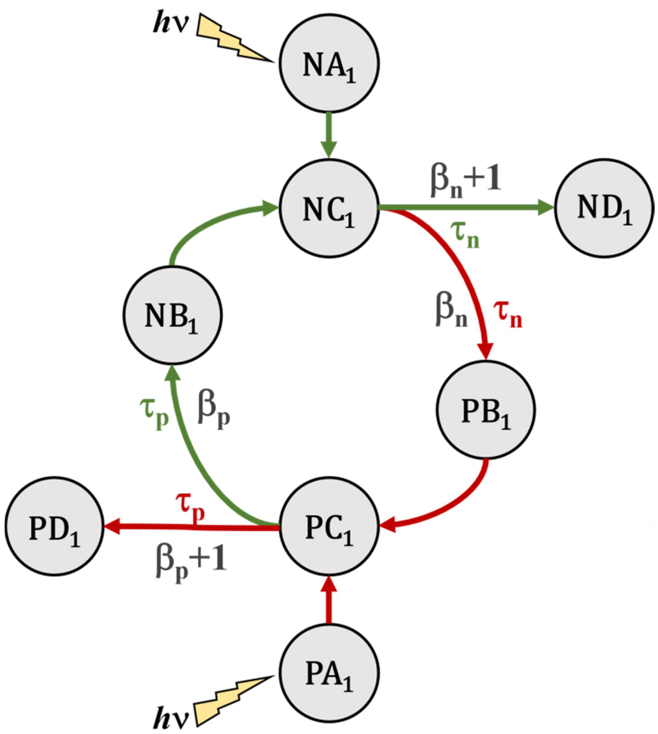

Figure II-3 Loop model schematic principle

The model encompasses various essential nodes, each serving a specific purpose and contributing to the overall functionality of the SPAD representation. Let's examine each of these nodes in detail:

- NA and PA: These nodes correspond to the initial injection points for electron-hole pairs and are utilized to model the photoelectric effect. Notably, these entry points can also accommodate the injection of charges induced by other phenomena, such as the release of trapped carriers (Afterpulsing) or thermally induced generation of electron-hole pairs. This versatility allows for a comprehensive exploration of the SPAD's behavior under diverse operating conditions.

- NC and PC: These nodes function as summing nodes, responsible for aggregating the currents originating from NA/NB and PA/PB, respectively. At each iteration within the loop, these nodes accumulate and combine the incoming currents, enabling the comprehensive analysis of the carrier flow dynamics and interactions within the device.
- NB and PB: As denoted by their names, these nodes serve as multiplication nodes. Their primary purpose is to multiply the current from the previous iteration using the extended ionization coefficients $\beta_{\mathrm{n,p}}$ (refer to ***Eq. II-7*** in the subsequent subsection). Through this multiplication process, the model accounts for the impact ionization phenomenon, simulating the exponential multiplication of carriers as they traverse the SPAD.
- ND and PD: Serving as the output nodes, ND and PD play a crucial role in the model's functionality. These nodes facilitate the extraction of the current at the ongoing iteration and inject it into the output terminals of the model, providing a representation of the SPAD's overall current behavior.

By incorporating these various nodes within the loop structure, the compact model encapsulates the key aspects of carrier dynamics, multiplication, and current flow within the SPAD. The interconnected nature of these nodes enables the simulation of intricate charge interactions, facilitating a deeper understanding of the device's performance under diverse operating conditions.

In summary, the compact model's loop-based architecture, closely aligned with the cyclic nature of the underlying physical phenomenon, allows for a comprehensive and accurate representation of the SPAD's behavior. The diverse nodes within the model, including the injection, summing, multiplication, and output nodes, collectively capture the intricacies of carrier dynamics, impact ionization, and current flow within the device.

2.2 Extended ionization coefficients

As highlighted in the preceding section, the carrier multiplication process within the model relies on the utilization of extended ionization coefficients. To gain a comprehensive understanding of these coefficients, it is necessary to revisit the concept of impact ionization, based on its following definition:

When an electron or a hole is subjected to acceleration over an average distance of ${\alpha_{\mathrm{n,p}}}^{-1}$ (also referred to as the dead space) as illustrated in ***Figure II-4***, it undergoes a collision, leading to the release of excess energy that generates an additional electron-hole pair. Subsequent collisions involving both electrons and holes can trigger an avalanche breakdown of the junction, further amplifying the carrier multiplication effect.

In the context of the model, a carrier accelerated by an electric field generates a single electron-hole pair after traversing a distance of ${\alpha_{\mathrm{n,p}}}^{-1}$.

However, if this same carrier continues to be accelerated across a distance corresponding to the width of the space charge region W_{SCR}, it can lead to consecutive and cumulative electron-hole multiplications. The number of successive multiplications is determined by the ratio of W_{SCR} to ${\alpha_{\mathrm{n,p}}}^{-1}$, yielding $W_{\mathrm{SCR}}/{\alpha_{\mathrm{n,p}}}^{-1}$.

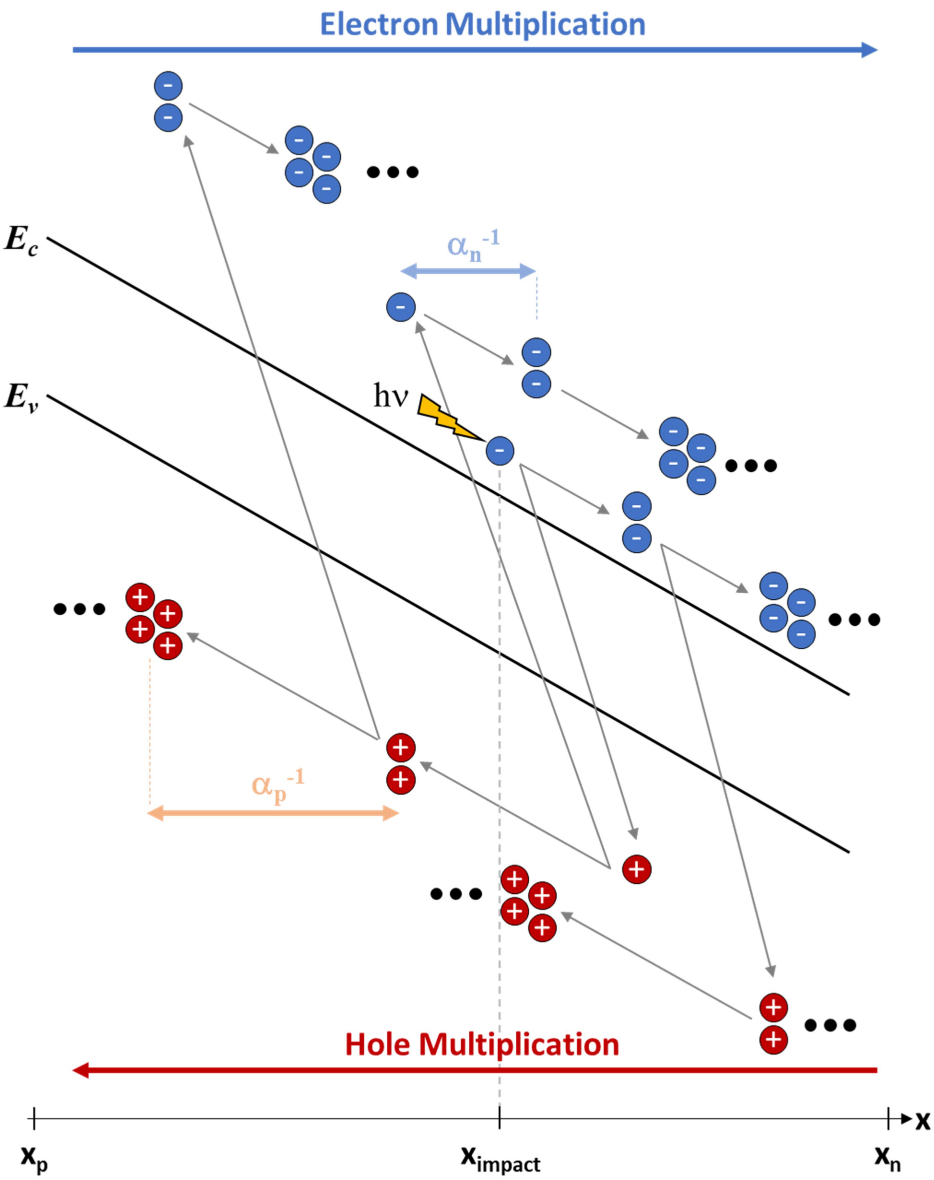

Figure II-4 Impact ionization mechanism, starting with electron generation

To express this relationship mathematically, the extended ionization coefficients $\beta_{\mathrm{n,p}}$ can be defined through the following equation:

$$\beta_{\mathrm{n,p}} = \alpha_{\mathrm{n,p}} \cdot W_{\mathrm{SCR}} \qquad \textbf{\textit{Eq. II-7}}$$

This equation illustrates how the extended ionization coefficients $\beta_{\mathrm{n,p}}$ are derived from the product of the original ionization coefficients $\alpha_{\mathrm{n,p}}$ and the width of the space charge region W_{SCR}. By incorporating these extended coefficients, the model effectively captures the cumulative impact ionization effect, offering a more accurate representation of carrier multiplication within the SPAD.

It is worth noting that a similar approach utilizing effective multiplication widths was employed in a model developed by Oussaiti et al. [61]. This demonstrates the wider adoption of such methodologies in the modeling and characterization of avalanche breakdown phenomena in semiconductor devices.

By incorporating the concept of extended ionization coefficients and their relationship to the space charge region width, the model enhances the accuracy and realism of the carrier multiplication process within the SPAD.

2.3 Transit time

The transit time $\tau_{\mathrm{n,p}}$ across the width of the space charge region W_{SCR} corresponds to the duration it takes for carriers to traverse this region. In the context of the SPAD device, the carriers are generated within the high electric field region. Consequently, they experience acceleration due to the Lorentz force, with no magnetic field present in our specific case. It is important to note that the electric field within the multiplication region is exceptionally high as a result of the applied high reverse bias to the device.

Due to this intense electric field, the carriers acquire velocities equal to their respective saturation velocities $v_{\mathrm{n,p}_{\mathrm{sat}}}$. The saturation velocity represents the maximum velocity that carriers can attain in a given material under high electric field conditions. The carriers within the SPAD device travel at this saturation velocity as they are propelled through the multiplication region. The resulting carrier transit times $\tau_{\mathrm{n,p}}$ are calculated with ***Eq. II-8***:

$$\tau_{\mathrm{n,p}} = \frac{W_{\mathrm{SCR}}}{v_{\mathrm{n,p}_{\mathrm{sat}}}} \qquad \textbf{\textit{Eq. II-8}}$$

By considering the transit time and the carriers' motion at their saturation velocities, the model takes into account the dynamics of carrier movement within the space charge region. This detailed understanding of carrier behavior is crucial for accurately capturing the temporal characteristics of the SPAD's response to incident photons and ensuring the model's fidelity in reproducing experimental observations.

Moreover, it is worth mentioning that the transit time and the carriers' velocity profiles are influenced by various factors, including the device's material properties, doping concentrations, and applied bias conditions. By incorporating these considerations, the model provides a comprehensive description of carrier transport within the SPAD, all while staying simple as the compact model. The accurate representation of carrier transit time and velocity profiles is pivotal in understanding the temporal response of the SPAD.

3 Model equations and parameters

3.1 Loop description and nodes equations

The description of the loop model involves breaking it down into distinct parts, each serving a specific purpose. The first part focuses on the generation of carriers through the photoelectric effect and the initiation of the avalanche process. When a trapped electron is released or when an absorbed photon creates an electron/hole pair, a trigger signal is generated and detected by the NA node for electron injection. Additionally, the PA node, representing hole injection, may also detect this signal, leading to the generation of either an electron or an electron/hole pair. This event, illustrated in ***Figure II-5***, gives rise to a current step with an amplitude of $q/\tau_{\mathrm{n,p}}$ and a duration of $\tau_{\mathrm{n,p}}$, ensuring that the total injected charge corresponds to the elementary charge q.

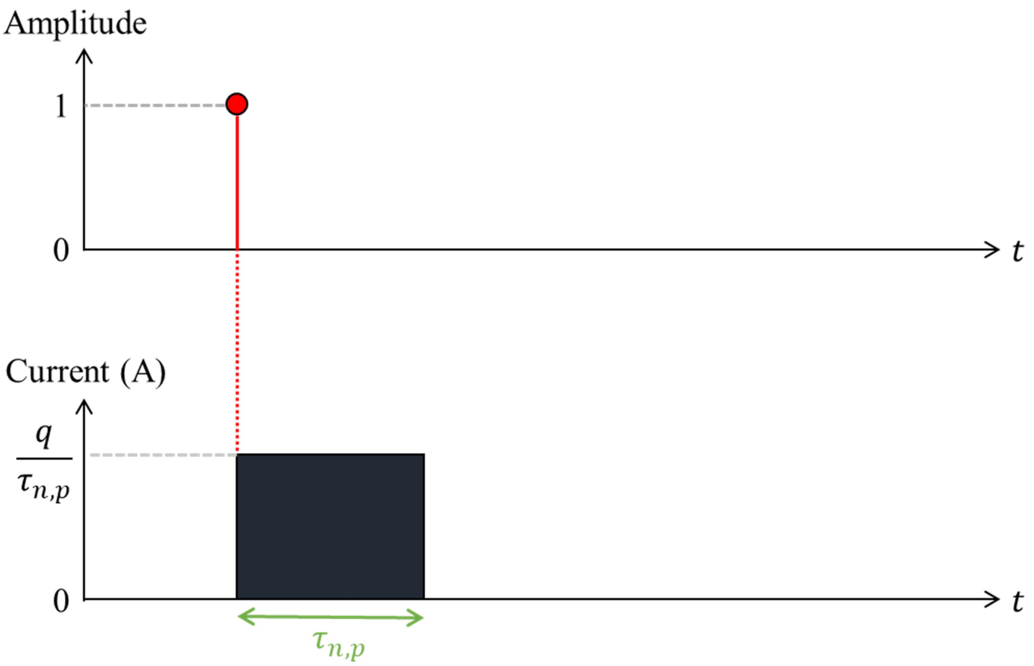

Figure II-5 NA/PA node signal generation explanation (bottom), trigger signal (top) and time correlation

Moving forward in the loop model, we encounter the summing nodes NC and PC. These nodes play a crucial role in collecting the currents generated in the NA and PA nodes, respectively. As the loop is repeated, these summing nodes not only have the ability to gather additional photogenerated carriers but also accumulate the contributions from the NB and PB nodes according to the following equations **Eq. *II*-8** and **Eq. *II*-9**:

$$\boldsymbol{NC = NA + NB} \qquad \textbf{\textit{Eq. II-9}}$$

$$\boldsymbol{PC = PA + PB} \qquad \textbf{\textit{Eq. II-9}}$$

By incorporating the currents from NA, NB, PA, and PB, the summing nodes provide an aggregated current value that reflects the combined effects of carrier generation and multiplication within the SPAD device. The inclusion of the NB and PB nodes allows for the amplification of the collected currents and accounts for the iterative nature of carrier multiplication.

The most critical aspect of the loop model lies in the multiplication of the currents gathered in the NC and PC nodes. This step involves two essential components: the multiplication nodes NB and PB, and the output nodes ND and PD.

The NB and PB multiplication nodes are responsible for representing the electrons generated by holes and the holes generated by electrons, respectively. By multiplying the currents present in the NC and PC nodes with the respective extended ionization coefficients $\beta_{\mathrm{n,p}}$, the following equations are obtained:

$$\boldsymbol{NB} = \beta_{\mathrm{p}} \cdot \boldsymbol{PC} \qquad \textbf{\textit{Eq. II-10}}$$

$$\boldsymbol{PB} = \beta_{\mathrm{n}} \cdot \boldsymbol{NC} \qquad \textbf{\textit{Eq. II-11}}$$

These equations highlight the fact that the currents in NB and PB are determined by the multiplication of the currents in the NC and PC nodes with the appropriate ionization coefficients. This multiplication process accounts for the cascading effect of carrier generation and multiplication within the SPAD device.

Furthermore, the ND and PD nodes are also considered as multiplication nodes. These nodes represent the electrons generated by electrons and the holes generated by holes, respectively. Similar to the NB and PB nodes, the currents induced by the multiplication process are added to the currents already present in the loop. However, the equations associated with these nodes differ slightly due to their role as output nodes:

$$\boldsymbol{ND} = (\beta_{\mathrm{n}} + 1) \cdot \boldsymbol{NC} \qquad \textbf{\textit{Eq. II-12}}$$

$$\boldsymbol{PB} = \left(\beta_{\mathrm{p}} + 1\right) \cdot \boldsymbol{PC} \qquad \textbf{\textit{Eq. II-13}}$$

These equations demonstrate that the currents in the ND and PD nodes are obtained by adding the multiplication currents from the NB and PB nodes to the currents already present in the ND and PD nodes, respectively.

By incorporating these multiplication nodes and considering their impact on the overall current flow, the loop model provides a comprehensive understanding of the carrier multiplication process.

It is important to note that the behavior of the compact model changes depending on the multiplication factor, especially when it is low or negligible. In such cases, when the reverse biasing voltage is lower than the breakdown voltage, the compact model exhibits behavior similar to that of a conventional photodiode. In this scenario, the photogenerated carriers from NA/PA contribute to the photocurrent without significant multiplication effects.

However, as the reverse biasing voltage approaches the breakdown voltage, the multiplication of the photocurrent becomes more pronounced, and the compact model accurately captures this behavior. Biasing the photodiode just below the breakdown voltage allows the compact model to properly describe the multiplication process, taking into account the impact ionization phenomenon and the resulting avalanche breakdown.

Figure II-6 provides a comprehensive visual representation of the impact ionization mechanism through a series of successive iterations as executed by the model over time. The figure showcases the intricate dynamics of carrier multiplication within the device, capturing both hole and electron multiplications originating from the initial photogeneration of the first electron-hole pair.

As the avalanche process unfolds, the compact model takes into account the successive generations of carriers and their significant contributions to the overall avalanche current.

The iterations, depicted in ***Figure II-6*** on the following page, serve as a visual testament to the robustness and accuracy of the compact model in simulating the intricate dynamics of impact ionization. By considering each successive generation of carriers during the avalanche process, the model provides valuable insights into the behavior and characteristics of the SPAD device under varying operational conditions and multiple generations.

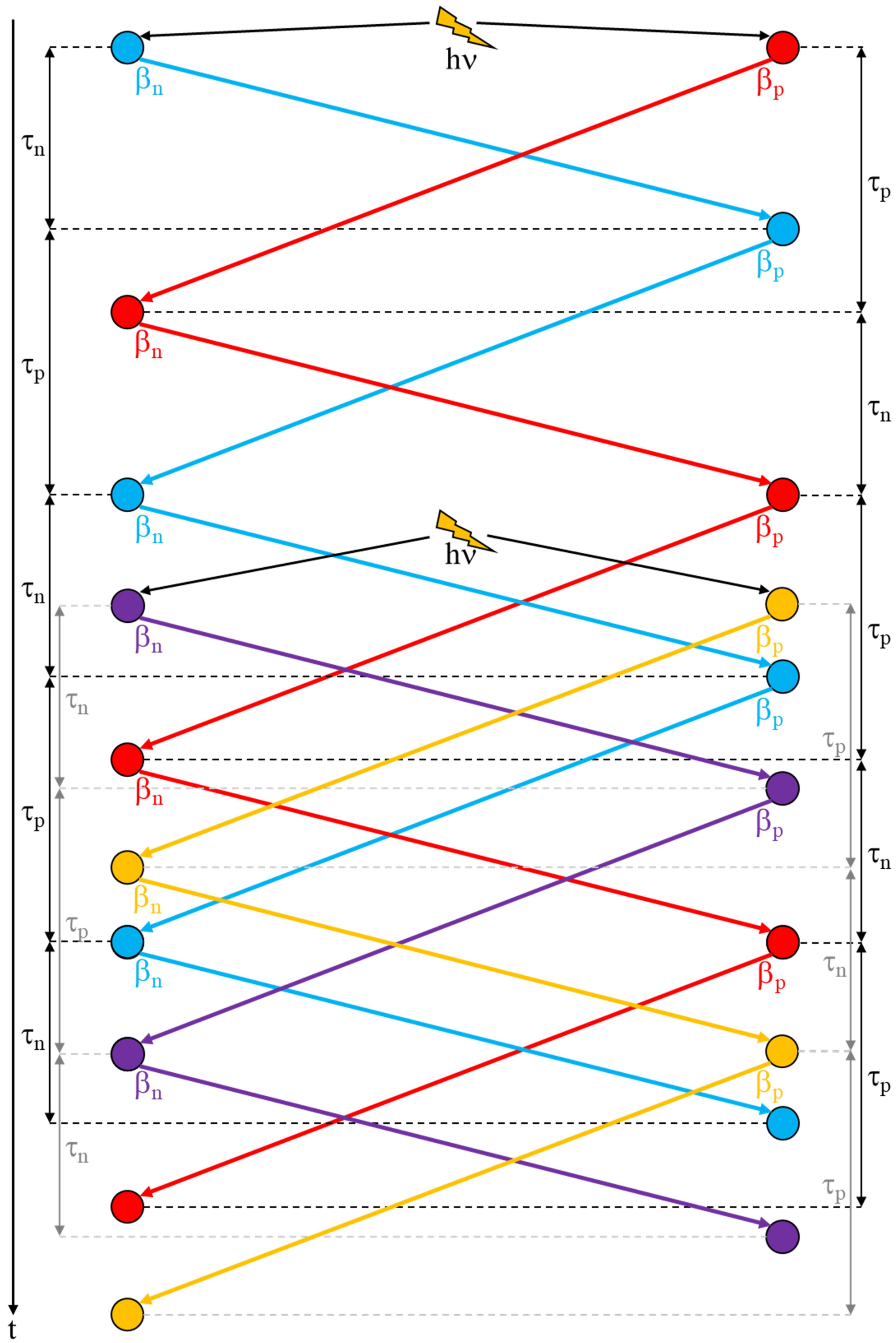

Figure II-6 : Avalanche breakdown sequence and subsequent impact ionization events, starting with a photogenerated electron hole pair.

3.2 Loop Gain

The looped structure of the model provides a foundation for comprehensively describing the avalanche breakdown process, as showcased in the control theory schematic depicted in ***Figure II-7***. This schematic serves as a powerful tool for analyzing the DC operating point of the model and delving into the intricate dynamics of the system.

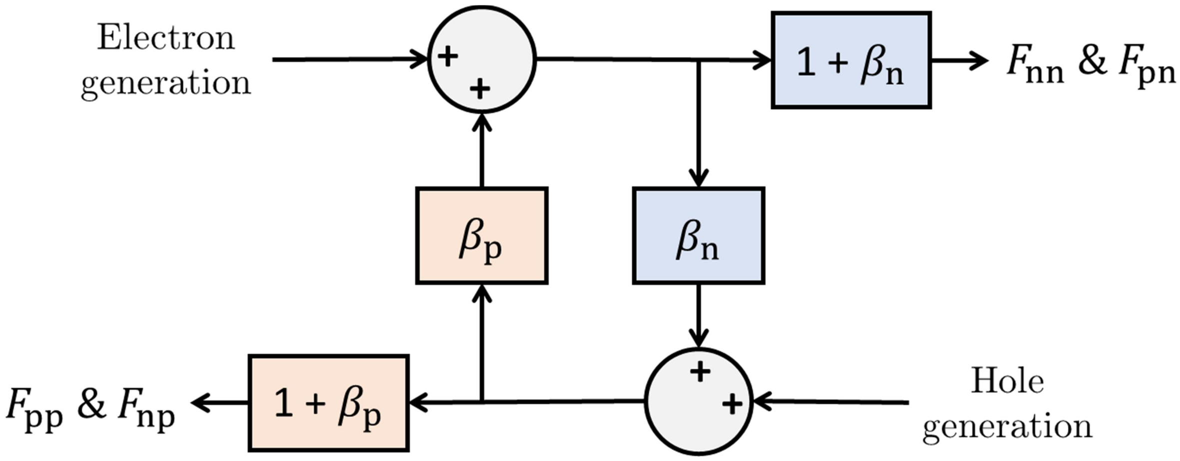

Figure II-7 Control theory description of loop model, DC operating point

Drawing upon principles from control theory, the system depicted in the schematic operates with two inputs and two outputs, resulting in four distinct transfer functions: F_{nn}, F_{pn}, F_{pp} and F_{np}. Each transfer function captures the relationship between the inputs and outputs, offering valuable insights into the interplay of various parameters and their impact on the behavior of the system.

The indexing convention employed in the transfer functions provides a clear distinction between the inputs and outputs. The first index denotes the input, whereas the second index signifies the output, allowing for a systematic and comprehensive analysis of the model's behavior under different conditions.

By examining these transfer functions, we gained a deeper understanding of the underlying dynamics governing the avalanche breakdown process in SPAD devices.

This comprehensive analysis aids in the optimization of device performance, enabling the design to be closer to reality.

Moreover, the control theory schematic provides a framework for studying the stability and controllability of the system, which in our case indicates and confirms the unstable diverging nature of the avalanche phenomena.

The analysis of the control system reveals that all its closed-loop transfer functions exhibit positive feedback. By considering the equation for a closed-loop positive feedback transfer function in ***Eq. II-14***, we can determine the transfer functions for the entire control system, providing valuable insights into the system's behavior.

$$TF_{\text{closed loop}} = \frac{TF_{\text{feedforward}}}{1 - TF_{\text{open loop}}} \qquad \textbf{\textit{Eq. II-14}}$$

The closed-loop transfer function $TF_{\text{closed loop}}$ is expressed as the division of the feedforward transfer function $TF_{\text{feedforward}}$ consisting of the direct path from input to output by the difference between 1 and the open-loop transfer function $TF_{\text{open loop}}$. This formulation allows us to evaluate the overall system response in terms of electric signal.

In the context of the avalanche breakdown process, where electron and hole generation play crucial roles, the closed-loop gains resulting from both an initial electron and hole generation on the electron output can be calculated. The closed-loop gain F_{nn}, representing electron-induced gain on the electron output, is determined by ***Eq. II-15***. Similarly, the closed-loop gain F_{pn}, which represents the effect of initial hole generation on the electron output, can be derived using ***Eq. II-16***. Both gains take into account both the electron and hole multiplication factors β_{n} and β_{p} and their interplay within the control system.

$$F_{\text{nn}} = \frac{(1 + \beta_{\text{n}})}{1 - \beta_{\text{n}} \cdot \beta_{\text{p}}} \qquad \textbf{\textit{Eq. II-15}}$$

$$F_{\text{pn}} = \frac{(1 + \beta_{\text{n}}) \cdot \beta_{\text{p}}}{1 - \beta_{\text{n}} \cdot \beta_{\text{p}}} \qquad \textbf{\textit{Eq. II-16}}$$

On the other hand, when the hole output is considered, the closed-loop gain F_{pp} reflects the hole-induced gain on the hole output and ***Eq. II-17*** provides the corresponding expression. Furthermore, the closed-loop gain F_{np}, representing the effect of initial electron generation on the hole output, can be determined using ***Eq. II-18***.

$$F_{\mathrm{pp}} = \frac{(1+\beta_{\mathrm{p}})}{1-\beta_{\mathrm{n}}\cdot\beta_{\mathrm{p}}} \qquad \textbf{\textit{Eq. II-17}}$$

$$F_{\mathrm{np}} = \frac{(1+\beta_{\mathrm{p}})\cdot\beta_{\mathrm{n}}}{1-\beta_{\mathrm{n}}\cdot\beta_{\mathrm{p}}} \qquad \textbf{\textit{Eq. II-18}}$$

As previously explained, the avalanche breakdown is defined by the generation of an infinite number of carriers by impact ionization. Therefore, the breakdown criterion of the model model is defined by the divergence of the transfer functions. This directly translates to the denominator becoming zero. The loop-gain G_{loop} of the model is defined as the multiplication of the $\beta_{\mathrm{n}}\cdot\beta_{\mathrm{p}}$ and plays a crucial role in determining its behavior.

- If $G_{\mathrm{loop}} < 1$, the model operates in the normal photodiode mode. In this mode, the generated electron/hole pairs are transported to the output nodes without undergoing multiplication. If the loop gain drops below 1, indicating that the voltage across the Space Charge Region (SCR) falls below the breakdown voltage, the avalanche current decreases, ultimately leading to the quenching of the avalanche process. It is important to note that the discharge of the SCR's capacitance by the avalanche current can cause the voltage across the SCR to automatically drop below the breakdown voltage, influenced by the electrical environment of the SPAD. Factors in the electrical environment such as the presence of a quenching resistor or an active quenching circuit play a role in this process.
- On the other hand, when the loop gain $G_{\mathrm{loop}} \geq 1$, the breakdown criterion is satisfied, and the model enters Geiger mode. In this mode, a generated

electron/hole pair undergoes exponential multiplication through impact ionization, resulting in the self-sustaining buildup of avalanche carriers. This Geiger mode operation signifies the occurrence of the avalanche breakdown and the amplification of carriers within the SPAD.

Understanding the behavior and transition between the normal photodiode mode and the Geiger mode is of significant importance in the design and optimization of SPAD devices. By accurately characterizing the loop gain and the associated breakdown criterion, the model is able to accurately simulate the breakdown voltage of the device as it is directly linked to the gain criterion established in this compact SPAD model.

3.3 Loop Gain correction

The complex and intricate nature of the SPAD device has posed challenges in accurately describing its loop gain, leading to the development of more refined models. As highlighted in the preceding paragraph, the breakdown voltage of the device is defined as the value at which the loop gain equals 1, signifying the onset of avalanche breakdown within the device.

The inaccuracy in describing the loop gain and the resulting breakdown voltage arises primarily from the unique architecture of SPADs, where the avalanche breakdown occurs in a specific region known as the multiplication region. Unlike the entire width of the SCR, the multiplication region represents only a fraction of the total SCR width. This narrow and critical region is where the impact ionization phenomena leading to avalanche breakdown predominantly take place.

Due to the confinement of the avalanche breakdown process within the multiplication region, the conventional description of the loop gain falls short in capturing the true dynamics and intricacies involved. To address this limitation, we had to adjust the W_{SCR} parameter, which represents the width of the space charge region, in order to more accurately account for the impact ionization occurring within the narrow multiplication region. This parameter was extracted from V_{br}. To extract the value of the corrective factor, needed to ajust the multiplication width, the DC simulation is compared to I(V) measurements to align the simulation with the measured value of the breakdown voltage V_{br}.

To comprehensively describe the process, the device is initially subjected to simulation, allowing for the characterization of its loop gain. Subsequently, the loop gain value, denoted as $G_{\mathrm{loop}_{\mathrm{extracted}}}$, is extracted at the theoretical breakdown voltage, V_{br}. In order to refine the description, the loop gain, G_{loop}, is equated to 1 by varying the parameter W_{SCR}. This adjustment ensures that the loop gain accurately corresponds to the condition at which avalanche breakdown occurs.

Taking into account the impact of this refinement, a corrected loop gain, denoted as $G_{\mathrm{loop}_{\mathrm{corr}}}$, is introduced. It incorporates an effective width parameter, W_{eff}, to account for the modifications made to W_{SCR} in ***Eq. II-19***. The corrected loop gain, $G_{\mathrm{loop}_{\mathrm{corr}}}$, is given by ***Eq. II-20,*** which incorporates the product of the ionization coefficients α_{n}, α_{p}, W_{SCR}, and a scaling factor γ.

$$G_{\mathrm{loop}} = \beta_{\mathrm{n}} \cdot \beta_{\mathrm{p}} = \alpha_{\mathrm{n}} \cdot \alpha_{\mathrm{p}} \cdot {W_{\mathrm{SCR}}}^2 \qquad \textbf{\textit{Eq. II-19}}$$

$$G_{\mathrm{loop}_{\mathrm{corr}}} = \alpha_{\mathrm{n}} \cdot \alpha_{\mathrm{p}} \cdot {W_{\mathrm{eff}}}^2 = \alpha_{\mathrm{n}} \cdot \alpha_{\mathrm{p}} \cdot {W_{\mathrm{SCR}}}^2 \cdot \gamma^2 = 1 \qquad \textbf{\textit{Eq. II-20}}$$

$$\gamma = \frac{1}{\sqrt{G_{\mathrm{loop}_{\mathrm{extracted}}}}} \qquad \textbf{\textit{Eq. II-21}}$$

To determine the appropriate scaling factor, γ, ***Eq. II-21*** is employed, relating it to the inverse square root of the extracted loop gain, $G_{\mathrm{loop}_{\mathrm{extracted}}}$ when identifying it thanks to ***Eq. II-19***. This scaling factor, γ, effectively compensates for any deviations or inaccuracies in the original loop gain extraction process.

Following this correction, the behavior of the model aligns more accurately with experimental results, providing a closer match to the observed phenomena. In the subsequent section, the results of these refinements will be further discussed, highlighting the improved agreement between the corrected model and experimental data. This comprehensive analysis underscores the importance of accurately characterizing and refining the loop gain in order to achieve a reliable and realistic representation of the SPAD device's behavior.

4 Simulations and experimental results

The first step in validating the single loop model as a proof of concept was to compare its simulated response to measurements. This comparison was made between a few key figures of merit that were available at the time. As a point of reference, measurements on the device were made mainly using its digital response `VSPADOUT` obtained with the circuit on ***Figure II-8***, as those were the only measurements

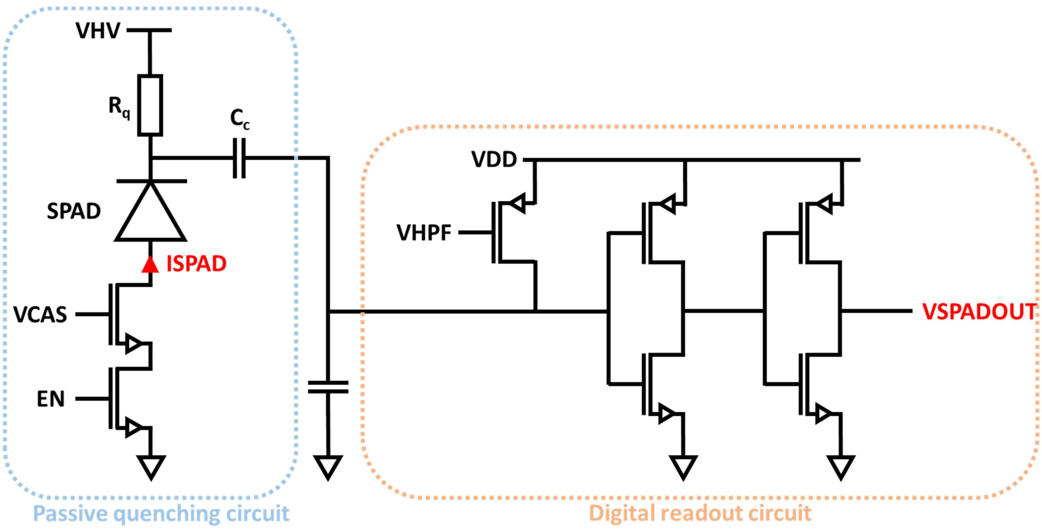

Figure II-8 Test circuit for digital measurements to obtain CPP and PW

available at that time.

The key objective was to challenge the baseline of the model's concept, which is represented by the Single Loop Model and serves as proof of concept. The key figures of merit in this comparison, evaluated in the following sections, were the charge per pulse (CPP), the pulse width (PW), the VHV0, and the electric field (EF).

4.1 Charge Per Pulse (CPP)

The first figure of merit used to compare the model's simulation results to measurements is Charge Per Pulse (CPP). The definition of CPP is the amount of charge that is contained in one current pulse generated by the SPAD diode. This

CPP value is obtained by integrating the measured current and dividing it by the number of pulses that have been detected or simulated, as in ***Eq. II-22:***

$$\mathrm{CPP} = \frac{\int_{t_0}^{t_f} I_{\mathrm{SPAD}}(t)dt}{n_{\mathrm{pulses}}} \qquad \textbf{\textit{Eq. II-22}}$$

The importance of this figure of merit is consequential as it is directly correlated with the power consumption of the circuit, enabling designers to take into account the accurate consumption linked to the SPAD.

This figure of merit has been measured and simulated at three temperatures: -20°C, 60°C, and 80°C. Comparing the model at different temperatures is important to evaluate the accuracy of the model when thermal conditions vary, which can have a major impact on the output of the device.

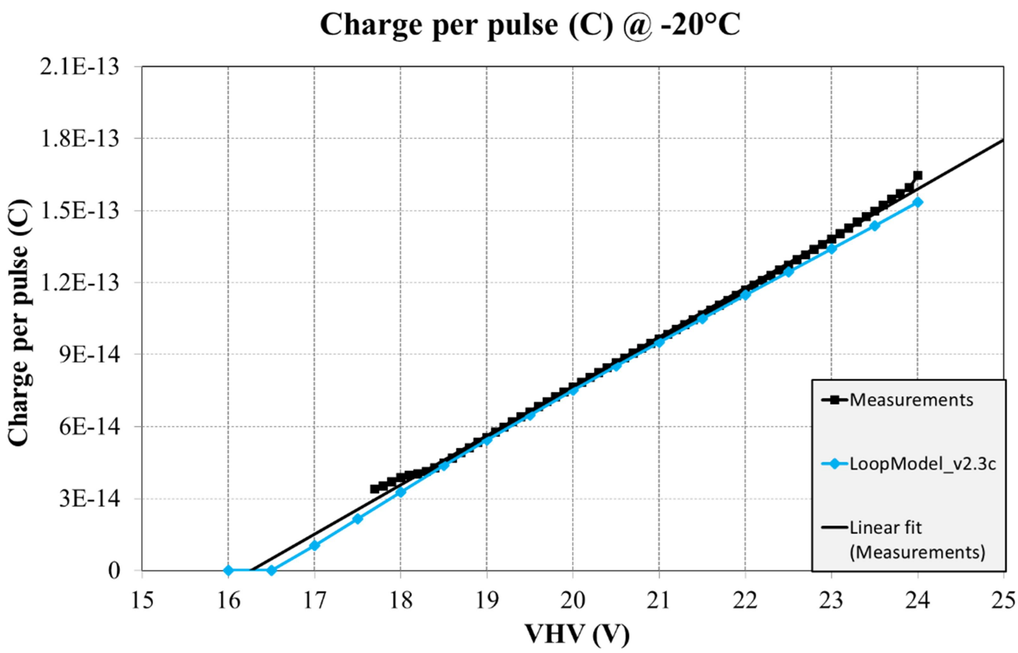

Figure II-9: Charge per pulse measurement versus model comparison at a temperature of -20°C

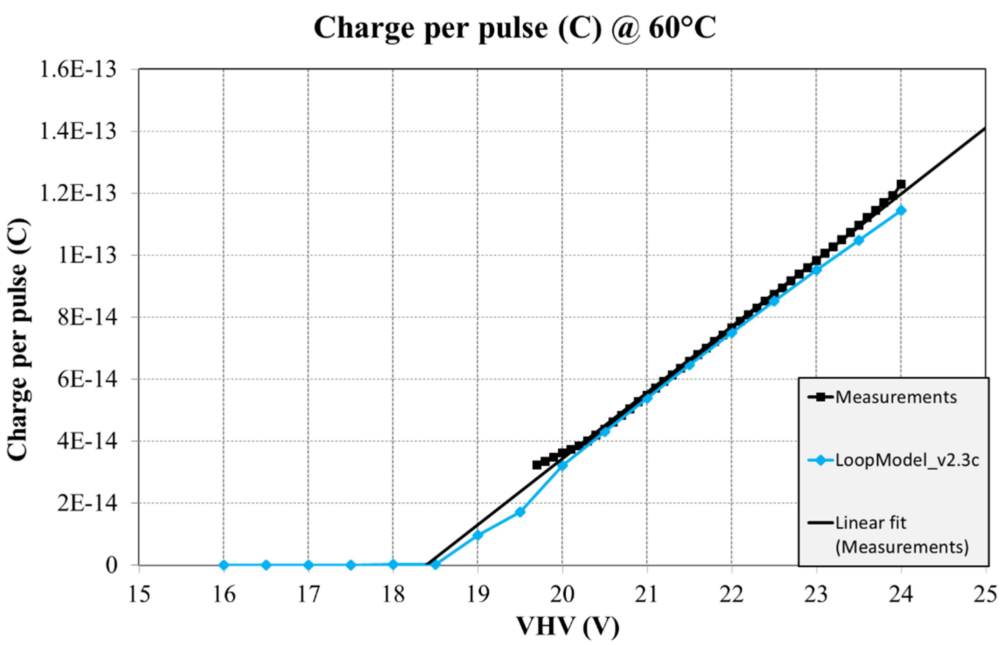

Figure II-11 : Charge per pulse measurement versus model comparison at a temperature of 60°C

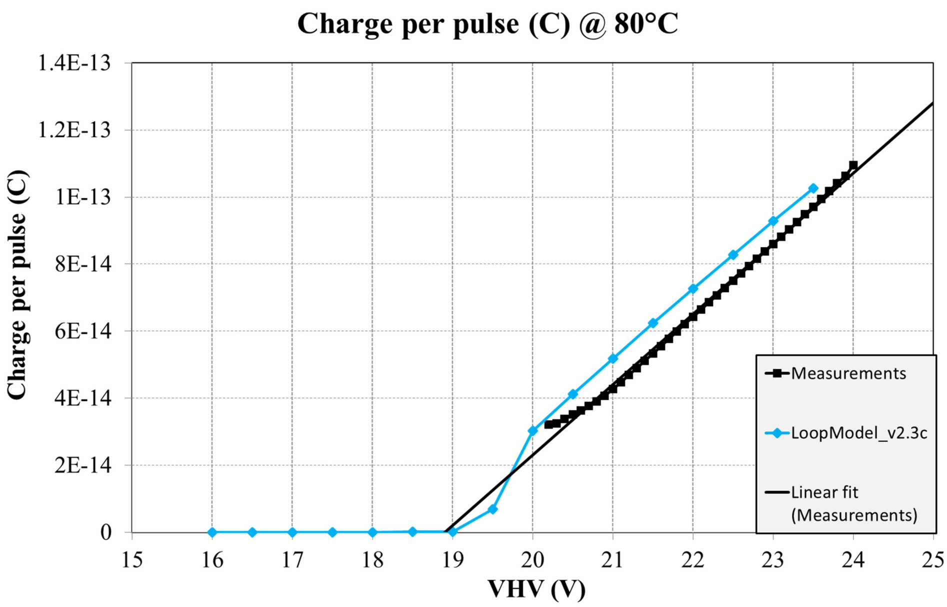

Figure II-10 : Charge per pulse measurement versus model comparison at a temperature of 80°C

As shown in ***Figures II-9***, ***Figures II-10*** and ***Figures II-11***, the model simulations provide a good fit to the experimental curve over most of the linear response of the CPP curves. This proves the ability of the model to predict power consumption in a complex circuit and its accuracy in regards to the physical device's response. It is worth noting that when the biasing voltage VHV is low, leading to a low excess voltage, the measurement points in this region are not exploitable and are therefore not represented due to the measurement methods used.

4.2 Pulse Width

The second Figure of Merit that was evaluated by confronting the simulated model responses to the device measurements was the Pulse Width. This value is measured in time and is defined by the duration of the digital output pulse of the SPAD when an input signal and biasing is applied, as represented in ***Figure II-12***.

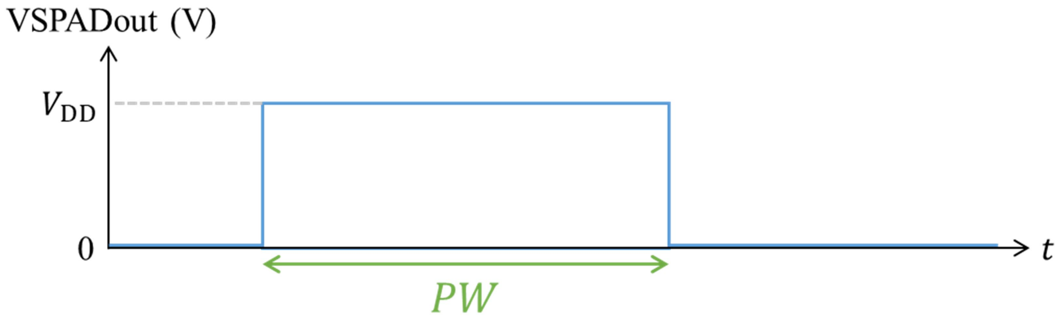

Figure II-12 Graphical representation of the pulse width figure of merit

The Pulse Width has been measured at 60°C for the entire polarization voltage range. To obtain the mean result represented in ***Figure II-13***, 1000 pulses of output have been averaged in order to be coherent with the model's response which, at first, is simulating the mean response of the device. The resulting comparison was promising, as the simulated response curve is close to the measurements. The factor of error can be attributed to factors such as electric field degradation or carrier recombination.

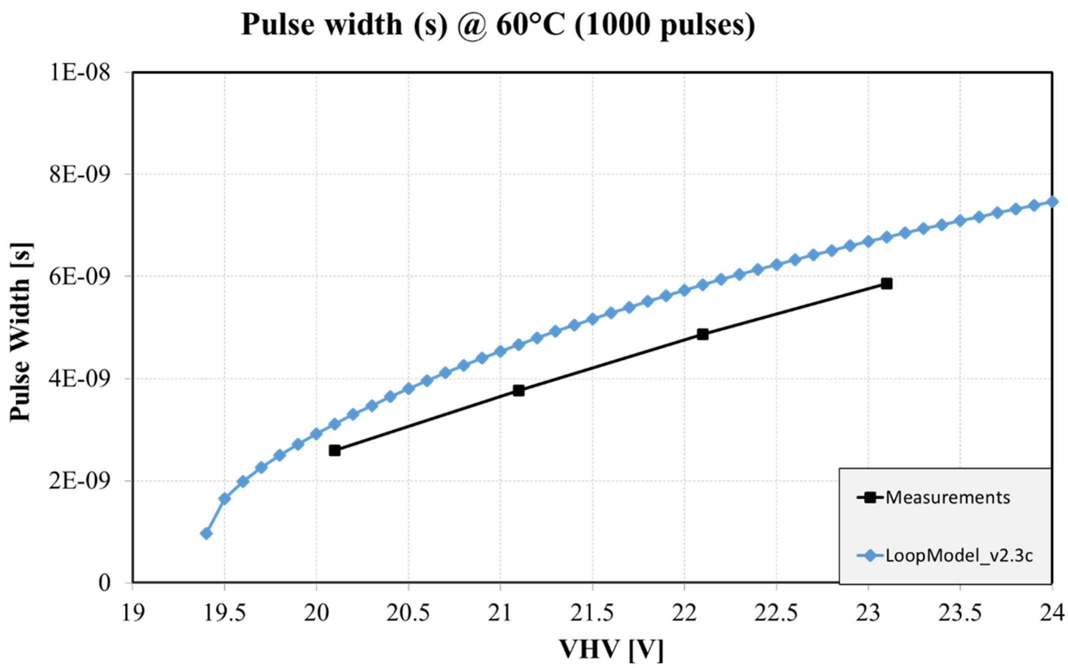

Figure II-13 Pulse Width measurement curve (black) compared to the model's response (blue) at 60°C

4.3 VHV0

The last of the Figures of merit that have been evaluated and compared to the measurements made on the device is VHV0. This figure is defined as the voltage at which 10% of the applied trigger pulses are generating an avalanche breakdown of the device. It is near the breakdown voltage of the device and provides a well-defined reference frame to compare the model's response to the characterization data.

Figure II-14 shows this comparison, with a dependency on the temperature that is considered, as the SPAD's avalanche breakdown phenomenon is heavily dependent on temperature.

The result of this confrontation yields a satisfactory fit, as the simulated model response is following the measurement curve with little deviation across the entire temperature range. The information collected can provide us with the confirmed accuracy of the model in terms of VHV0, no matter the temperature the model is considered and used at.

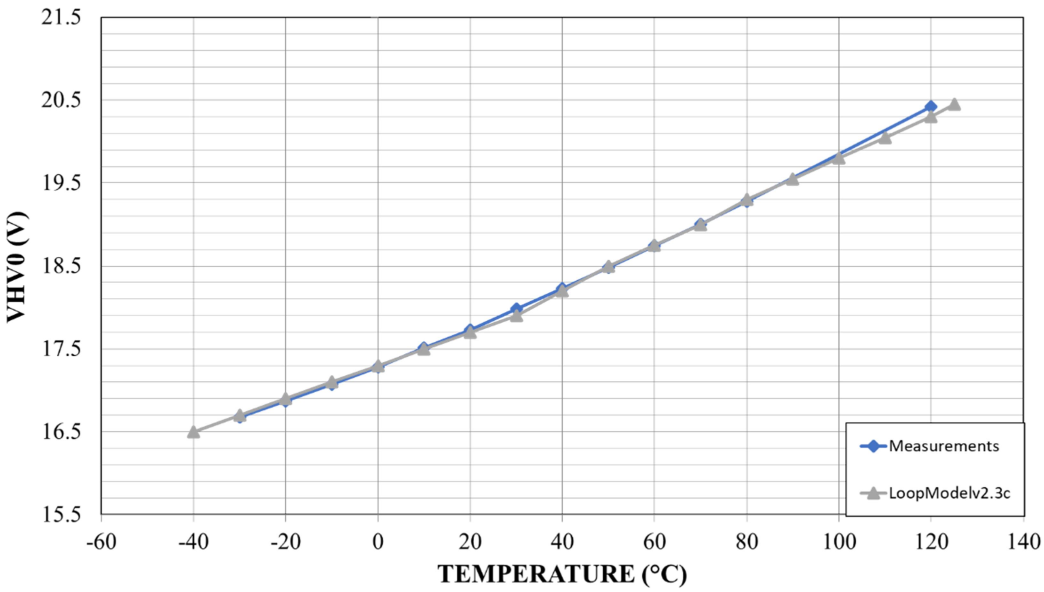

Figure II-14 VHV0 measurement curve (blue) compared to the model's response (grey) over the entire temperature range

4.4 Electric field

As a final challenge to the model's effectiveness to emulate the SPAD behavior, it has been compared to TCAD simulations of the electric field inside the SCR. This electric field is taken at its peak value, as this peak largely dictated the avalanche behavior of the device.

Given the comparison shown in ***Figure II-15***, the model is more than capable of accurately simulating the electric field inside the SCR of the SPAD. Once the model was correctly tuned, the simulated and extracted electric field coincided with the TCAD simulated curve, with little error and deviation. This overall fit over the entire biasing voltage range both provides a valuable information on the model's capabilities, and further proof of concept, giving us a solid starting point towards a distributed modelling approach.

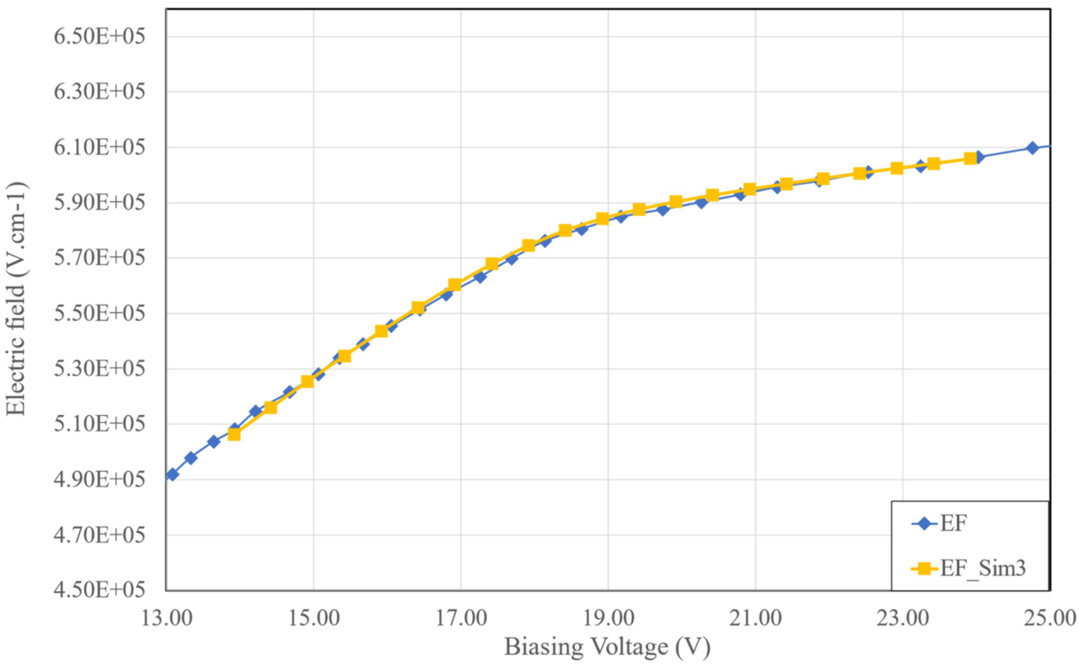

Figure II-15 Electric field TCAD simulated curve (blue) compared to the model's response (yellow) over the entire biasing voltage range

Conclusion

The Single-loop compact SPAD model, although highly functional, serves as a crucial proof of concept and has garnered validation and recognition through its publication at the IEEE international ESSDERC/ESSCIRC conference [62]. From a functional standpoint, the model demonstrates its capability to deliver accurate results when compared to experimental measurements of charge per pulse, pulse width, and VHV0, highlighting its effectiveness in capturing the fundamental behavior of SPAD devices. Additionally, the model is able to fit the electric field simulated by TCAD, further confirming its capability.

However, it is important to acknowledge the limitations that arise due to the simplified architecture of the Single-loop model, which solely focuses on describing the behavior of the SPAD within the high electric field multiplication region. Consequently, this simplified model fails to account for certain stochastic effects inherent to the device. The two primary components of these stochastic effects are timing jitter of the avalanche breakdown and variations in the peak avalanche current.

Timing jitter primarily manifests as a delay between the absorption of an incoming photon and the occurrence of the avalanche breakdown phenomenon. It arises from various factors, including the transportation of carriers within the device, carrier trapping and recombination processes, and the complex interplay of carrier dynamics. The comprehensive understanding and characterization of timing jitter play a crucial role in accurately assessing the performance and reliability of SPAD devices. Detailed explanations regarding the breakdown of timing jitter will be expounded upon in the subsequent chapter, unraveling its underlying mechanisms and shedding light on its implications for device behavior.

Peak avalanche current variations also stem from the intricate geometry and architecture of SPAD devices. These variations arise due to non-uniform carrier multiplication processes and the spatial distribution of carriers within the multiplication region. Understanding the causes and effects of peak avalanche current variations is vital for optimizing device performance and reliability. A thorough explanation of the rationale behind the development of the compact modeling solution can be found in the corresponding chapter, providing insights into how the model addresses peak avalanche current variations and accounts for their impact.

To effectively capture the stochastic effects and embrace the complete complexity of SPAD devices, the Single-loop model necessitates expansion to encompass multiple loops. This extension to multiple loops enables the accurate modeling of distributed effects, carrier transport phenomena, and spatial distribution, facilitating a multidimensional description of the device. By incorporating these additional loops, the model can adequately account for the statistical effects associated with timing jitter and peak avalanche current variations. The subsequent chapter will delve into the details of this extension to multiple loops, presenting a comprehensive analysis of SPAD devices within a broader modeling framework and offering a more profound understanding of their behavior.

This next step in the model's design apprehends one of the principal goals and motivations of this approach's development: the transient response of the device. Thanks to advanced circuitry and measurement techniques with a dedicated high precision test setup, the voltage transients of both the device's electrodes can be clearly visualized and exploited, providing crucial input for the SPAD model that was developed.

Chapter III Distributed 1D compact model

Introduction

As mentioned earlier, the limitations of the Single-loop model primarily arise from its lack of spatialization, as it fails to consider the distributed effects inherent in SPAD devices. It has been observed in **Chapter II**, **Subchapter 1.1** that the structure of the space charge region (SCR) within the SPAD is not spatially uniform; instead, it can be divided into two distinct regions: the absorption region and the multiplication region. To address this limitation, the initial step in developing a more comprehensive model involves introducing spatial distribution along the depth of the junction.

The purpose of incorporating this spatial distribution is to simulate the stochastic nature of photon absorption within the device. Depending on the wavelength of the incident photons, which can vary across different applications, the generation of initial electron-hole pairs can occur throughout the entire SCR. Consequently, the location of electron absorption within the device has a significant impact on the timing of the avalanche phenomenon, giving rise to timing jitter in the SPAD response. This spatial distribution allows for the consideration of delayed avalanche events based on the position of electron absorption, contributing to the overall timing jitter experienced by the device.

Furthermore, the 1D distribution along the width of the device also introduces an additional effect known as the arrival delay of the hole current. This delay in the hole current manifests as a "snail-shaped" total avalanche current when combined with the electron current contribution. The non-uniform distribution of carriers along the width of the device introduces a spatial variation in the multiplication process, resulting in a distinctive shape for the avalanche current waveform. By incorporating this width-based distribution, the model captures the delayed arrival of the hole current and its consequential influence on the overall avalanche process.

By considering both the depth and width distributions in the model, the spatialization of the device provides a more comprehensive understanding of the stochastic effects exhibited by the SPAD. This improved model allows for a detailed analysis of the timing jitter and the distinct shape of the avalanche current waveform, shedding light on the intricate interplay between carrier distributions and the resulting device behavior.

First, the measurement test chip and setup will be explained, followed by a first insight into the characterization data that can be obtained. Then, the modelling approach considering the depth structure of the device will be detailed using the properties of the two main areas of the space charge region: the absorption and multiplication regions. Following this, the different calibrations of the model to fit the mean transient measurements will be described in detail, finishing with the parameter extraction scheme used to adjust the model to produce an accurate output.

1 Measurements

1.1 “KNACK” Test circuit

As part of the IPCEI (Important Project of Common European Interest) project under the Nano2022 initiative led by the French government, a collaborative effort between STMicroelectronics and ICube Laboratory has resulted in the development of a sophisticated test circuit. The primary objective of this test circuit is to facilitate direct analog measurements of the transient anode and cathode voltages in Single-Photon Avalanche Diodes (SPADs). This collaborative endeavor represents a significant advancement in the field of nanotechnology research and development.

The test circuit, illustrated in ***Figure III-1***, comprises three distinct components, each serving a specific purpose in the measurement process:

- The main SPAD pixel, which forms the core of the circuit, is intricately designed to incorporate a passive quenching circuit and pixel selection cascoded transistors. These components play a crucial role in ensuring precise and controlled operation of the SPAD, enabling accurate measurement of the transient anode and cathode voltages. The passive quenching circuit serves to rapidly extinguish the avalanche current, allowing the SPAD to recover quickly and prepare for subsequent measurements. The pixel selection cascoded transistors enable efficient and selective activation of individual pixels within the SPAD array, facilitating targeted voltage measurements.
- RF amplifiers specifically tailored for transient measurements of the anode and cathode voltages are integrated into the test circuit. These dedicated RF amplifiers are meticulously designed to optimize the amplification of the voltage signals while minimizing noise and distortion. Their high-performance characteristics ensure reliable and accurate measurement of the transient voltage variations, providing valuable insights into the dynamic behavior of the SPAD device.

- A digital readout circuitry is incorporated into the test circuit, comprising essential elements such as a high-pass filter transistor and a buffer stage consisting of two successive inverter stages. This digital readout circuitry serves as an interface between the analog voltage signals captured from the SPAD and the subsequent digital processing stages. The high-pass filter transistor enables the removal of low-frequency noise and interference, enhancing the fidelity of the measured voltage signals. The buffer stage, consisting of multiple inverter stages, ensures signal integrity and amplification, preparing the voltage signals for subsequent digital analysis and interpretation.

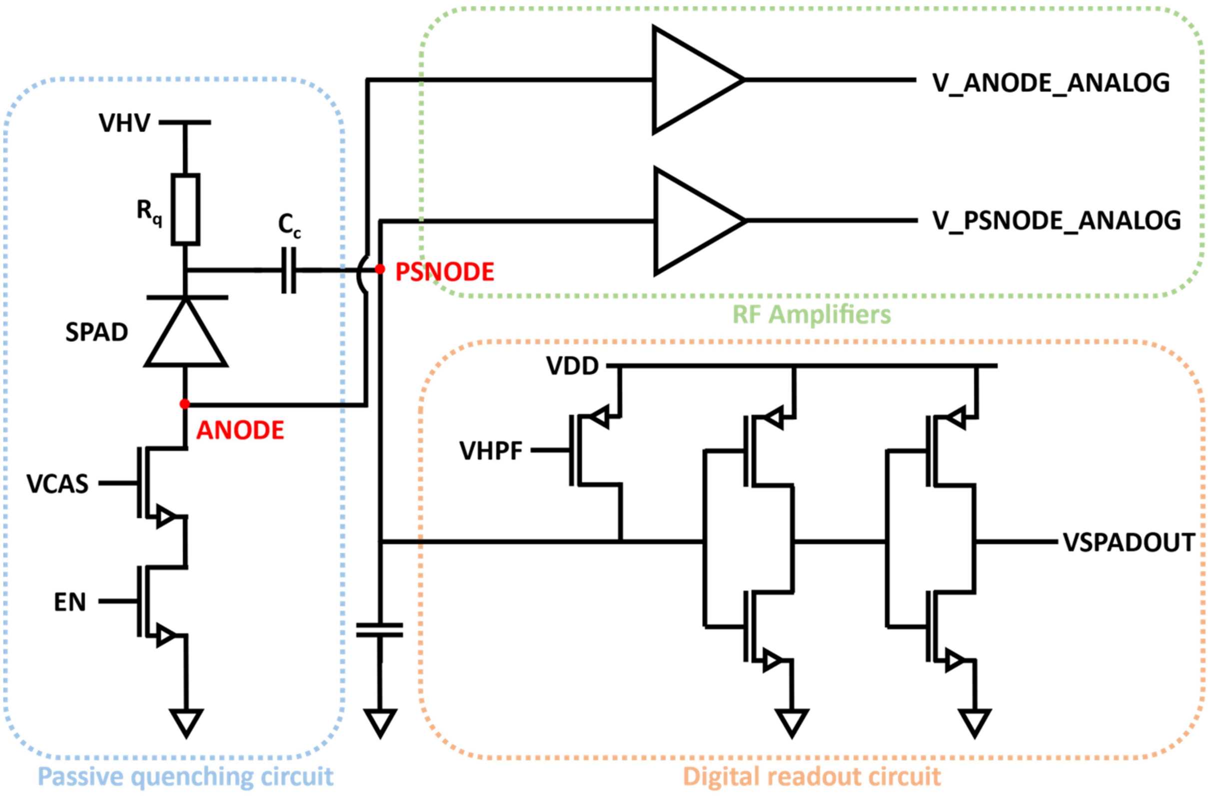

Figure III-1: "KNACK" test circuit diagram. Reference design used for both measurements and simulation

The collaborative development of this sophisticated test circuit within the IPCEI project of the Nano2022 initiative represents a significant milestone in the advancement of SPAD technology. By enabling direct analog measurements of the transient anode and cathode voltages, this circuitry enhances our understanding of the intricate dynamics and performance characteristics of SPAD devices. It paves the way for further advancements in the field of nanotechnology and establishes a solid foundation for future research and development endeavors.

In the optoelectronic characterization method employed for measuring ultrafast Single-Photon Avalanche Diode (SPAD) voltage transients, a comprehensive voltage measurement strategy is implemented. The method involves the direct measurement of the `V_ANODE` voltage, while the measurement of the `V_CATHODE` voltage requires several steps due to the high applied voltage VHV, ranging from 22 to 25 volts.

To accurately measure the `V_CATHODE` voltage, a carefully designed capacitive divider is employed. This capacitive divider incorporates a Metal-Oxide-Metal (MOM) capacitance, which is directly integrated into the SPAD pixel. Additionally, the parasitic capacitance to ground from both the digital output circuit and the RF amplifier is considered in the measurement setup. By utilizing this capacitive divider configuration, the `V_CATHODE` voltage is effectively stepped down to a new node referred to as `V_PSNODE`. `V_PSNODE` represents an attenuated image of the `V_CATHODE` voltage, allowing for precise voltage measurement while considering the voltage limitations imposed by the high applied voltage VHV.

Once `V_PSNODE` and `V_ANODE` are obtained, both voltages undergo amplification using dedicated RF amplifiers that are specifically designed for this particular application. These RF amplifiers are optimized to amplify the voltage signals accurately and with high fidelity. The amplification stage enhances the signal-to-noise ratio and allows for more precise and reliable measurements. The amplified voltages, namely `V_PSNODE_ANALOG` and `V_ANODE_ANALOG`, provide an enhanced representation of the respective voltage signals and enable detailed characterization and analysis of the SPAD device.

By incorporating these specialized RF amplifiers in the measurement setup, the optoelectronic characterization method ensures comprehensive and accurate measurement of both `V_PSNODE` and `V_ANODE` voltages. This sophisticated approach facilitates a detailed examination of the voltage characteristics of the SPAD during ultrafast voltage transients, enabling a thorough understanding of its behavior, performance, and response dynamics in various operating conditions.

1.2 Test setup

The optoelectronic characterization method developed to measure ultrafast SPAD voltage transients encompasses several essential steps and additional components for an exhaustive analysis. This comprehensive method involves illuminating the SPAD device with a Titanium-doped Aluminum Oxide (Ti:Al2O3) femtosecond laser and meticulously monitoring the analog voltage variations at the SPAD electrodes. Let's delve into the details of this method and its supplementary components.

To ensure precise temporal referencing, a low jitter trigger signal with a remarkable temporal stability of less than 1 picosecond (ps) is employed. This trigger signal, synchronized with the femtosecond laser pulse, serves as an excellent temporal reference for the measurements. The femtosecond laser itself emits pulses with an incredibly short duration of 100 femtoseconds (fs), allowing for the capture of ultrafast SPAD voltage transients.

The synchronization between the trigger signal and the femtosecond laser pulse is achieved by utilizing a beam splitter, which diverts a portion of the laser's signal to enable the synchronization process. This synchronization ensures accurate alignment between the trigger signal and the laser pulse, enhancing the precision of the measurements.

In addition to the synchronization setup, a Pockels cell is incorporated into the method to address the relatively high jitter associated with the SPAD device when operating in single photon counting mode. The Pockels cell plays a crucial role in modulating the laser beam passing through it. By adjusting the cell's modulation parameters, the number of photons reaching the SPAD device is significantly reduced to approximately 10. This controlled reduction in photon flux, coupled with a low repetition frequency of 800 hertz (Hz), enables a controlled and precise illumination of the SPAD for accurate measurements.

The combined use of the Ti:Al2O3 femtosecond laser, the low jitter trigger signal, and the Pockels cell offers several advantages in the optoelectronic characterization process. The femtosecond laser's ultrashort pulse duration enables the capture of ultrafast SPAD voltage transients occurring within picosecond timescales. The low jitter trigger signal provides a stable temporal reference, ensuring accurate temporal alignment of the measurements. The inclusion of the Pockels cell allows for controlled illumination of the SPAD device, mitigating the effects of jitter and providing precise measurements. ***Figure III-2*** depicts the entire test setup for the measurements of `V_PSNODE and V_ANODE`.

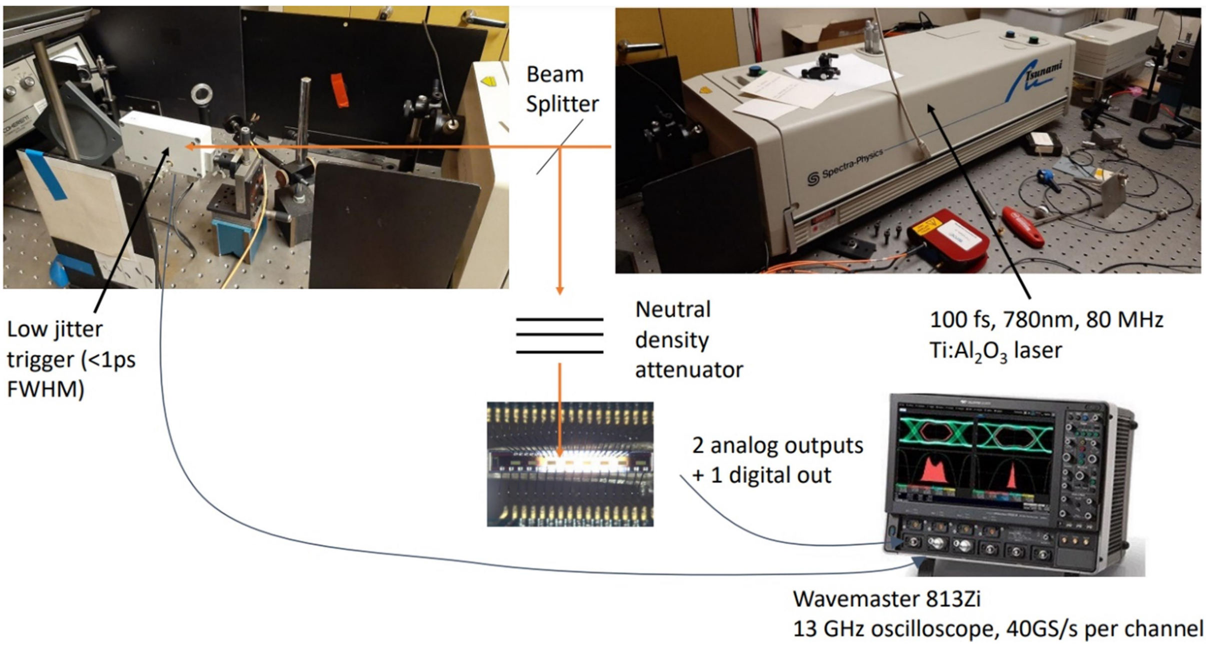

Figure III-2 Knack test setup for transient measurements of `V_PSNODE` and `V_ANODE`

To summarize, the optoelectronic characterization method developed for measuring ultrafast SPAD voltage transients involves illuminating the device with a Ti:Al2O3 femtosecond laser while simultaneously monitoring the analog voltage at the SPAD electrodes. This method incorporates a low jitter trigger signal synchronized with the femtosecond laser pulse for precise temporal referencing. A beam splitter is utilized to facilitate synchronization, and a Pockels cell is employed to control the illumination of the SPAD device, reducing photon flux and ensuring accurate measurements. This comprehensive approach maximizes the accuracy and reliability of the optoelectronic characterization process.

1.3 Transient Measurements

With the developed 'KNACK" test circuit and the corresponding measurement method and its test setup, highly detailed and precise measurements of the transient response of SPAD diodes can be carried out. Using the power of the RF-Amplifiers to provide a signal that is easily exploitable in terms of resolution and amplitude, the voltages at the two electrodes of the device were measured and visualized. On ***Figures III-3*** and ***III-4***, the voltage swings of the PSNode and the anode are

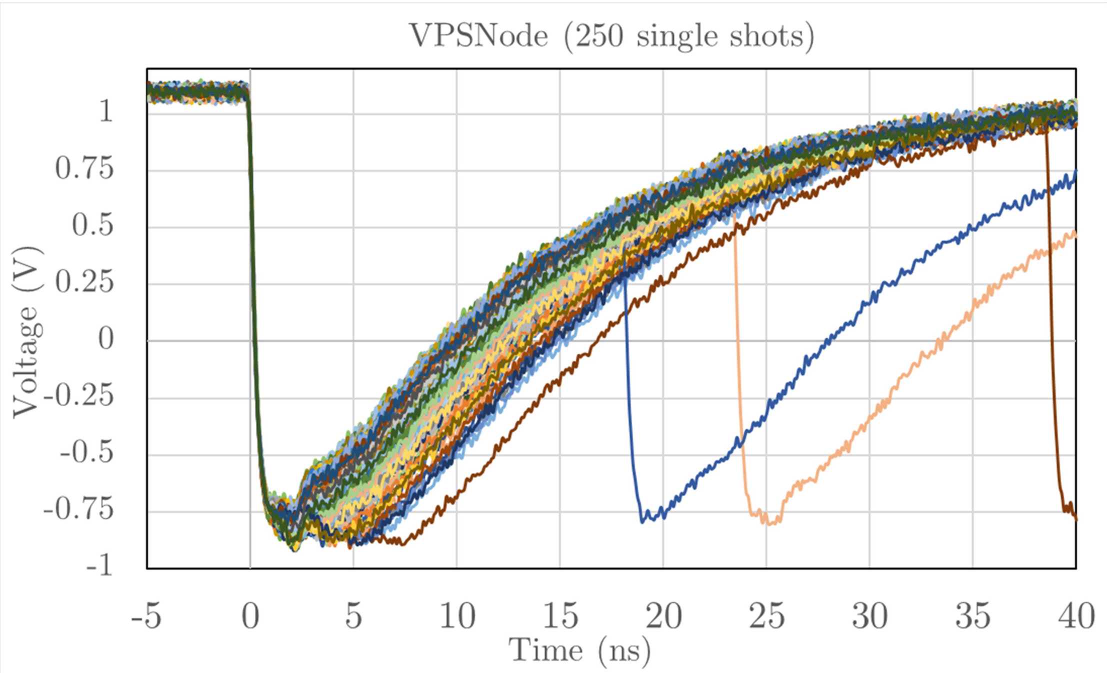

Figure III-3 Voltage transient measurements of the `PSNode`***, 250 single shots at a biasing voltage*** `VHV` ***of 25V***

represented respectively.

Starting with the voltage transients on ***Figure III-3***, representing the voltage drop on the `VPSNode` of the circuit. This node is the direct image of the SPAD's cathode voltage through a capacitive divider in order to achieve a lower output voltage that can in turn be processed by the digital output circuitry. It is clearly visible that the SPAD's response is of a stochastic nature. Fluctuations can be observed in terms of voltage drop amplitude, duration and recharge curve. Furthermore, as visible on ***Figure III-3*** in the blue, yellow and brown transients, there is involuntary retriggering and secondary avalanche. This highlights the incidence of stochastic phenomena like DCR or Afterpulsing.

If the Anode voltage of the `V_Anode` node is visualized, as on ***Figure III-4***, another phenomenon can be observed. Indeed, the amplitude fluctuations as seen on ***Figure III-3*** are still present. In addition, the `V_Anode` voltage presents a secondary peak after approximately 2.5ns for a 25V `VHV` polarization. The origins and possible causes resulting in this behavior will be further explored in **Chapter IV**.

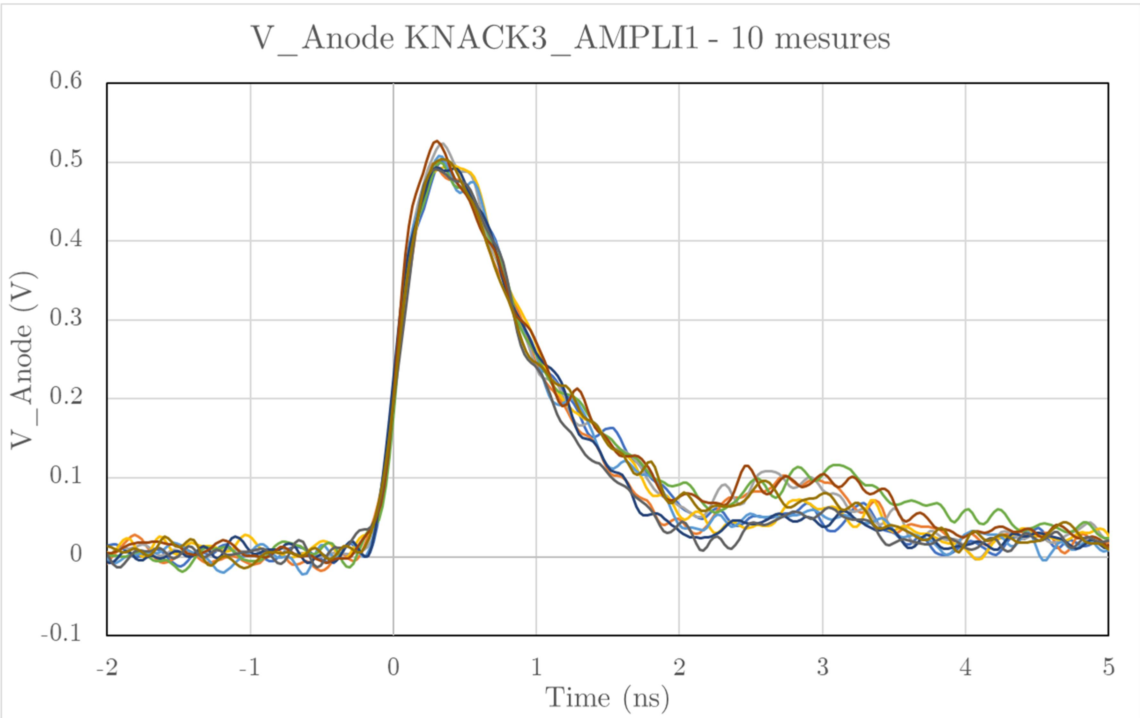

Figure III-4 Voltage transient measurements of the `V_Anode`***, 10 single shots at a biasing voltage*** `VHV` ***of 25V***

Based on the transient measurements provided by the developed "KNACK" test structure Several preliminary observations can be made, based on the characterization results:

- The output voltage drop across both the `VPSNode` and `V_Anode` circuit nodes are presenting statistical fluctuations in their respective amplitude values. The cause of those fluctuations is mainly due to the difference in the local electric field if the multiplication area in which the impact ionization is occurring. Variations in this electric field can cause small variations in impact ionization coefficients, resulting in fluctuations of avalanche dynamics.

- For the `VPSNode` transients, the plateau area on the bottom of the voltage drop has variations in its duration. This may be caused by a sustained avalanche at its peak, resulting in carriers remaining in the multiplication region that will delay the quenching of the device.
- For the `V_Anode` transients, a secondary peak has been observed through the high precision measurements.

The model developed in this manuscript is aimed to represent the mean transient behavior, so the reference measurement, ***Figure II-5***, used was obtained by averaging 100 transient responses that do not contain afterpulses. The absence of those afterpulses insures a clean mean transient response without distortion that may be caused by those events.

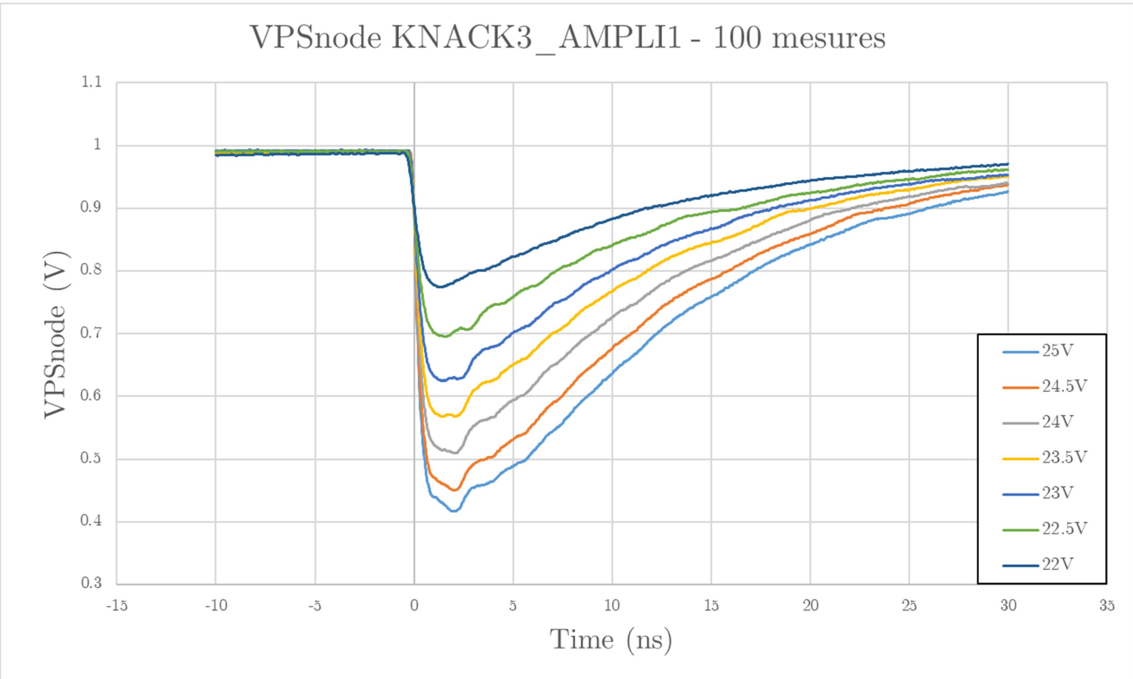

Figure III-5 Mean transient responses of the `VPSnode` ***obtained by averaging 100 measurements***

2 Photon absorption depth modelling approach

The architecture of the SPAD device, specifically the two distinct parts of SCR, played a pivotal role in our modeling approach, enabling a comprehensive representation of the device's behavior. To capture the spatial distribution within the device, we adopted a systematic approach by dividing the SCR into smaller portions. Each of these portions was represented by an individual loop within the model, incorporating its unique characteristics such as ionization coefficients, widths, and transit times. This level of granularity allowed for a detailed and accurate simulation of the SPAD's operation.

Figure III-6 provides a comprehensive visualization of the spatial structure employed in our model, showcasing the integration of these individual loops into the overall compact model framework.

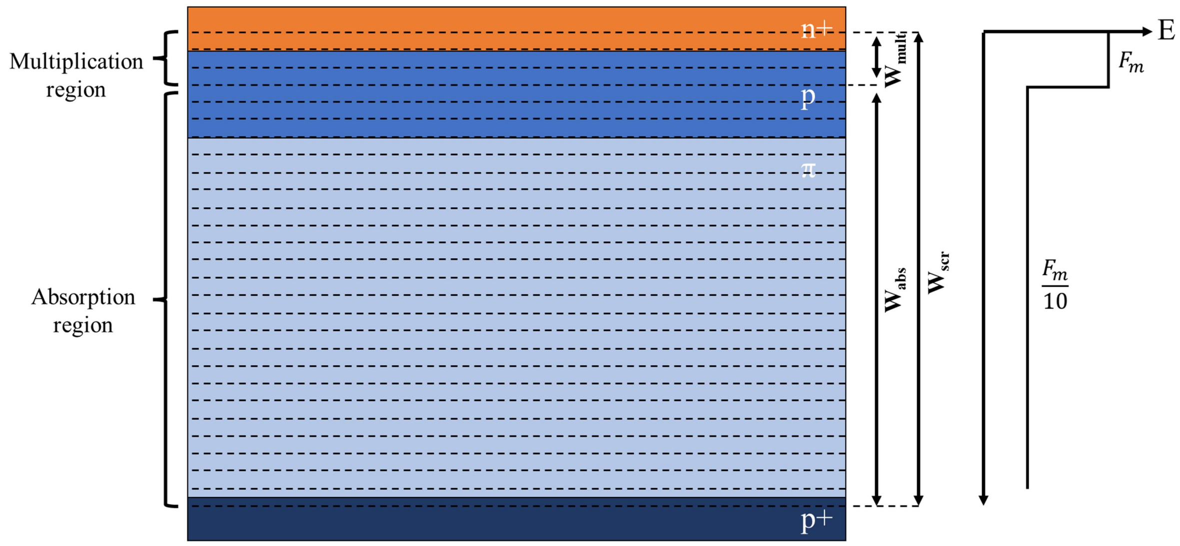

Figure III-6 Spatial distribution of device chosen for the compact model

To begin, the total width of the SCR, denoted as W_{SCR}, was further divided into two distinct regions: the absorption region W_{abs} and the multiplication region W_{mult}. This division was not only applied to the physical width of the SCR but also extended to the electric field distribution within the device.

In the wider portion of the SCR, characterized by a significantly lower electric field strength compared to the multiplication region (approximately 10 times weaker), carriers do not possess sufficient energy to surpass the impact ionization threshold. As a result, carriers injected or generated in this low electric field region predominantly drift towards the anode (in the case of holes) or move towards the multiplication region (in the case of electrons). This wider carrier drift region, due to its larger size, functions primarily as the absorption region, effectively capturing and absorbing a significant portion of the incident photons. The chosen value for the electric field within this region, set at 1/10th of the electric field present in the multiplication region, was determined through extensive TCAD simulations [63] conducted on devices with similar technological characteristics.

Furthermore, for finer modeling and increased precision, both the absorption and multiplication regions can be subdivided into additional subsections. This approach allows for a more detailed and localized representation of the device's behavior, capturing variations within specific regions.

Subsequent sections of the study will delve into a more comprehensive and detailed description of both the absorption and multiplication regions within the SPAD device. Through an exhaustive analysis, we aim to elucidate the intricate mechanisms and characteristics of these regions, considering factors such as carrier transport, impact ionization phenomena, and spatial distribution. This comprehensive understanding of the spatial distribution within the SPAD device is essential for accurately modeling its performance and behavior, facilitating further advancements in the field of single-photon detection.

2.1 Multiplication region

Within the multiplication region of the SCR, the electric field attains a sufficiently high magnitude to facilitate impact ionization, thereby enabling the avalanche breakdown process. This multiplication region, in comparison to the overall SCR, is relatively small, with dimensions approximately one-tenth of the total SCR width. To simplify the modeling process and capture the essential characteristics of this narrow region, we initially assumed that the electric field inside the multiplication region could be approximated as a constant field. This constant field corresponds to the maximum electric field strength, denoted as F_{m}, present within the SPAD device.

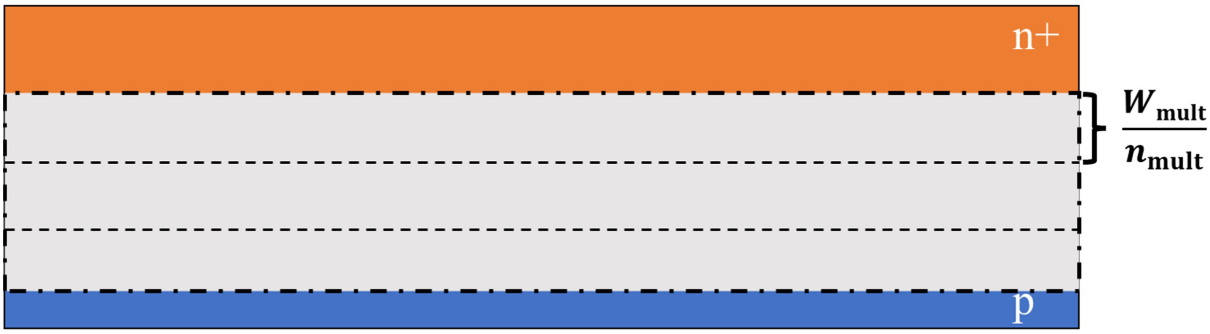

Figure III-7 Multiplication region discretization and distance of each element

To achieve a more detailed representation of the multiplication region, as seen in ***Figure III-7***, we further divided the width of this region, denoted as W_{mult}, into n_{mult} equal parts. Each of these parts corresponds to an individual loop within the model, enabling the capture of the localized behavior and characteristics of that specific portion of the multiplication region. The value of W_{mult} is dependent on the biasing voltage applied to the SPAD device, as per the relationship defined by ***Eq. III-1***. To simplify the expression, a highly asymmetric junction is considered due to the very high doping concentration of the n+ region, resulting in a width of depletion region spanning mainly in the low doped Pwell zone.

$$W_{\mathrm{mult}} = \sqrt{\frac{2 \cdot \varepsilon_0 \varepsilon_{\mathrm{Si}} \cdot V_{\mathrm{d}}}{q \cdot \mathrm{N}_{\mathrm{d}_{\mathrm{Pwell}}}}} \qquad \textbf{\textit{Eq. III-87}}$$

By subdividing the multiplication region into multiple loops and assigning specific widths to each loop, we aim to be able to achieve a more refined and accurate modeling of the impact ionization process. This approach allows us to account for variations in electric field strength and carrier dynamics across different sections of the multiplication region, offering a comprehensive understanding of the SPAD device's behavior under various operating conditions.

The selection of the number of loops and their corresponding widths within the multiplication region is based on careful consideration of the device's design parameters, performance requirements, and fabrication technology. Through systematic analysis and experimentation, we can then optimize the model to accurately represent the physical and electrical characteristics of the multiplication region, facilitating a more precise simulation of the avalanche breakdown process and its impact on the SPAD's overall performance.

2.2 Absorption region

Within the absorption region of the SCR, the electric field does not reach the necessary intensity to facilitate impact ionization and the subsequent avalanche breakdown. As a result, the precise characterization of the electric field within this region is deemed unnecessary in our initial modeling approach. Instead, we adopt an approximation where the electric field strength inside the absorption region is assigned a value of $F_{\mathrm{m}}/10$. This choice is made to ensure that the electrons and holes traversing through the SCR maintain a velocity close to their saturation velocity, v_{sat}.

By selecting an electric field magnitude of $F_{\mathrm{m}}/10$, we aim to accurately capture the carrier transport behavior within the absorption region. This choice is motivated by the desire to maintain carrier velocities at a level consistent with the maximum achievable under the given bias conditions. The absorption region width, denoted as W_{abs}, can be calculated by subtracting the width of the multiplication region, W_{mult}, from the total SCR width, W_{SCR},.

To ensure a comprehensive representation of the absorption region, as in ***Figure III-8,*** we further divide the width of this region into a suitable number of subsections, specifically n_{loop}- n_{mult}. Here, n_{loop} represents the total number of loops constituting the model, while n_{mult}. corresponds to the number of loops assigned to the multiplication region. The division of the absorption region into these subsections allows once more for a more refined modeling approach, accounting for the localized characteristics and carrier transport behavior within specific portions of the absorption region.

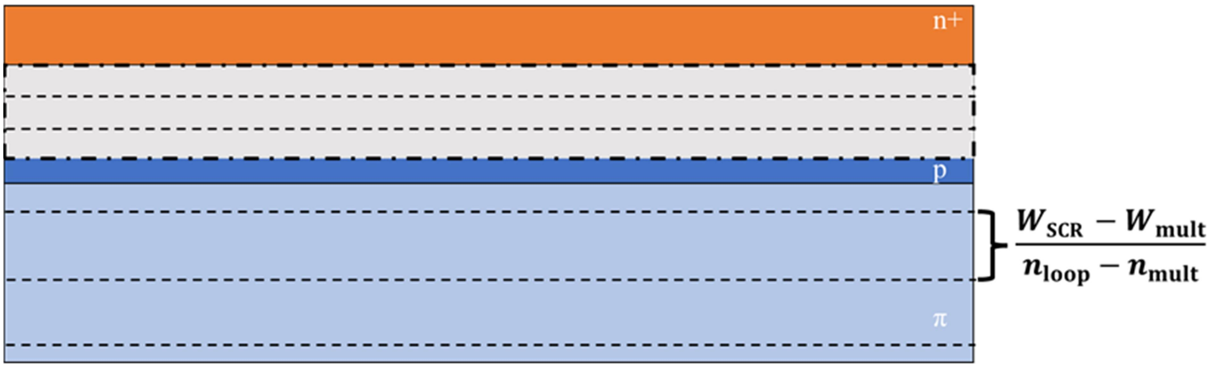

Figure III-8 Absorption region discretization and distance represented by each individual element

2.3 Multiple loop Structure

To effectively implement the distributed structure of the SPAD device in our code, it was necessary to establish the overall schematic principle of the model. Building upon previous work conducted on the single loop model, we utilized its structure as a starting point and extended it to incorporate multiple loops. In this specific example, as depicted in ***Figure III-9***, we allocated two loops for each distinct region of the SPAD, namely the absorption region and the multiplication region. These loops are interconnected through their respective input and output nodes, denoted as PD and ND.

By adopting this design approach, we achieved the flexibility to implement various combinations of regions with a dedicated number of loops assigned to each one. This design choice empowers users to customize the model according to their specific diode topology requirements and preferences. It allows for the creation of models tailored to different SPAD configurations, ensuring a more accurate representation of the device's behavior

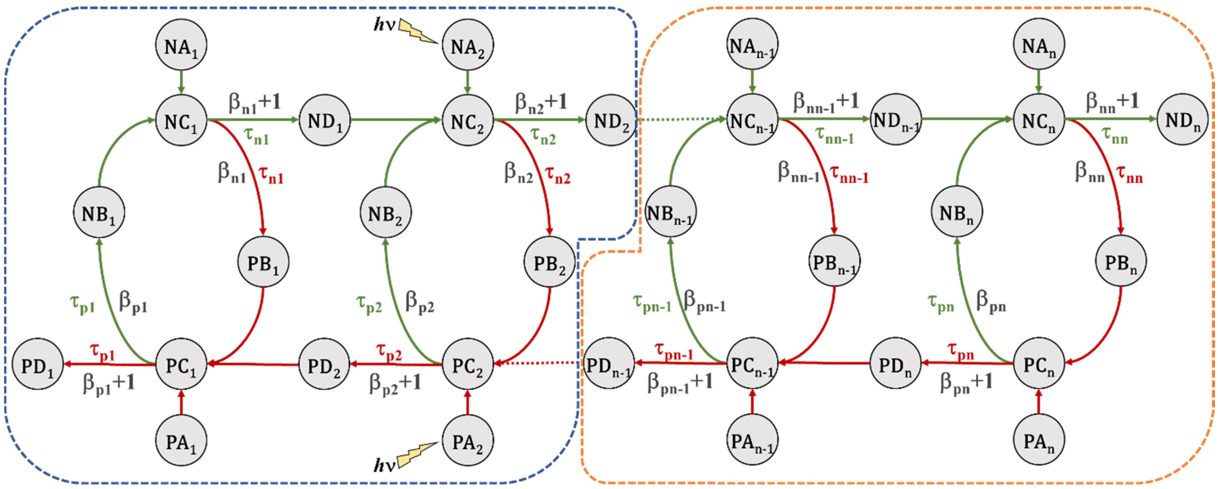

Figure III-9 Distributed SPAD model at n-th order with two loops represented per SPAD region

During the course of this PhD research, both the absorption region and the multiplication region will be represented as a single loop each. This decision was made considering the escalating complexity that arises with an increased number of loops, especially while modelling the narrow multiplication region. Moreover, the physical characteristics of the device we focused on modeling featured a reach-through structure with an extremely narrow multiplication region, measuring only several hundreds of nanometers at most. Consequently, employing a single loop for each region provides a sufficiently accurate representation of the device's physical properties and behavior.

It is important to note that while we have chosen to describe the absorption and multiplication regions as single loops in this particular case, the modeling framework we have developed remains highly adaptable. Future researchers or designers can easily modify the number of loops allocated to each region, or even introduce additional loops to account for more intricate device architectures and specific requirements. This adaptability ensures that the model can, in time, be applied to a wide range of SPAD designs and enables researchers to explore various scenarios and optimize the model for different applications.

2.4 Gain quenching criteria

To ensure the accuracy and stability of the model, it was essential to implement a reliable quenching criterion that effectively addresses convergence issues. The initial approach involved leveraging the closed loop properties of the model, as discussed in **Chapter II**, **3**. By examining ***Eq. III-2***, governing the closed loop gains F_{xx}, it was observed that while the numerators may differ, the denominator plays a crucial role, as it is directly linked to the divergent nature of the avalanche process. It was noted that the gain initially exhibits a positive value and progressively increases until reaching infinity at $\beta_{\mathrm{n}} \cdot \beta_{\mathrm{p}} = 1$, after which it transitions to a negative value. This unique characteristic of the gain function served as the basis for establishing the quenching criteria of the model.

$$F_{\mathrm{xx}} = \frac{\mathrm{NUM}}{1 - \beta_{\mathrm{n}} \cdot \beta_{\mathrm{p}}}$$ ***Eq. III-2***

Upon careful examination of the transient response of the gain, as illustrated in ***Figure III-10***, a distinct pattern emerges. It is evident that when the quenching circuitry initiates the recharging process of the SPAD, the gain exhibits a gradual increase over time. This progressive increment in gain corresponds to the recovery of the SPAD from its quenched state. Subsequently, as the gain reaches a positive value again, the SPAD reenters an avalanche state, thereby restarting the avalanche process indefinitely.

To ensure effective quenching of the avalanche and prevent any residual charge from triggering subsequent avalanches, a quenching criterion is established. This criterion is based on the observation that once the gain becomes positive again, it implies that the SPAD has regained its ability to sustain an avalanche. Hence, the quenching criterion is defined as $\beta_{\mathrm{n}} \cdot \beta_{\mathrm{p}} = 0$ at the moment the gain turns positive. By setting $\beta_{\mathrm{n}} \cdot \beta_{\mathrm{p}}$ to zero, the avalanche is completely quenched, and the SPAD is set back to its non-avalanche state to enable retriggering of the device.

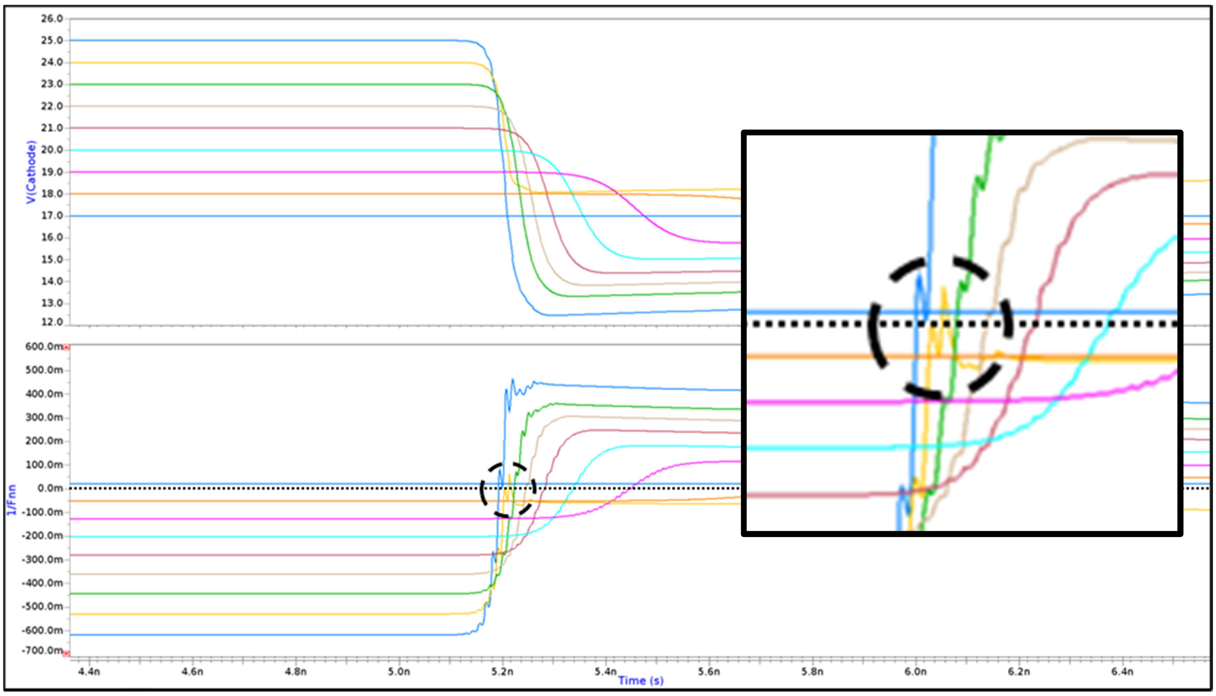

Figure III-10 Example of impact of gain oscillations (bottom) on cathode voltage (top)

Implementing this quenching solution has proven to be successful in achieving accurate quenching for the majority of scenarios. However, it is important to note that due to the presence of slight oscillations in the gain product $\beta_\mathrm{n} \cdot \beta_\mathrm{p}$, mainly linked to the highly unstable nature of the process, some situations may arise where these oscillations inadvertently trigger the avalanche quenching criteria erroneously. As a consequence, premature quenching of the SPAD can occur, impacting its performance and reliability. Therefore, further analysis and refinement of the quenching criteria are necessary to address these occasional inconsistencies and improve the overall robustness of the SPAD model.

As seen on the zoom on ***Figure III-10***, one can observe a notable oscillation pattern displayed by the yellow curve. This specific oscillation highlights a significant drawback of the previously mentioned gain quenching criterion, as the oscillation crossing the zero value inadvertently triggers the premature quenching of the SPAD. This unintended consequence compromises the accuracy and reliability of the quenching process.

Considering this limitation, an alternative solution was devised to address the issue, emphasizing the physical aspect of quenching that is closely linked to the electric charges retained within the SCR. Recognizing the crucial role of the remaining charges in the quenching mechanism, this second solution aims to refine the quenching process by incorporating the charge dynamics within the SPAD device.

By considering the dynamics of electric charges, the model takes into account the accumulation and dissipation of charges during the quenching phase. This approach provides a more comprehensive understanding of the quenching process and allows for a more accurate determination of when the SPAD has reached a fully quenched state.

The refined quenching solution incorporates parameters such as the residual charge present in the SCR during the avalanche and quenching process. The model can then precisely determine the point at which the accumulated charges have dissipated sufficiently to ensure the complete quenching of the SPAD, thereby preventing any inadvertent triggering of subsequent avalanches.

3 Model calibration and tuning

3.1 Charge quenching criterion

To address the issue of premature quenching of the SPAD, an alternative quenching criterion was developed, taking into account the charge dynamics resulting from the avalanche breakdown of the junction. This improved quenching criterion focuses on effectively managing the charge generated during the avalanche process to prevent retriggering of subsequent avalanches.

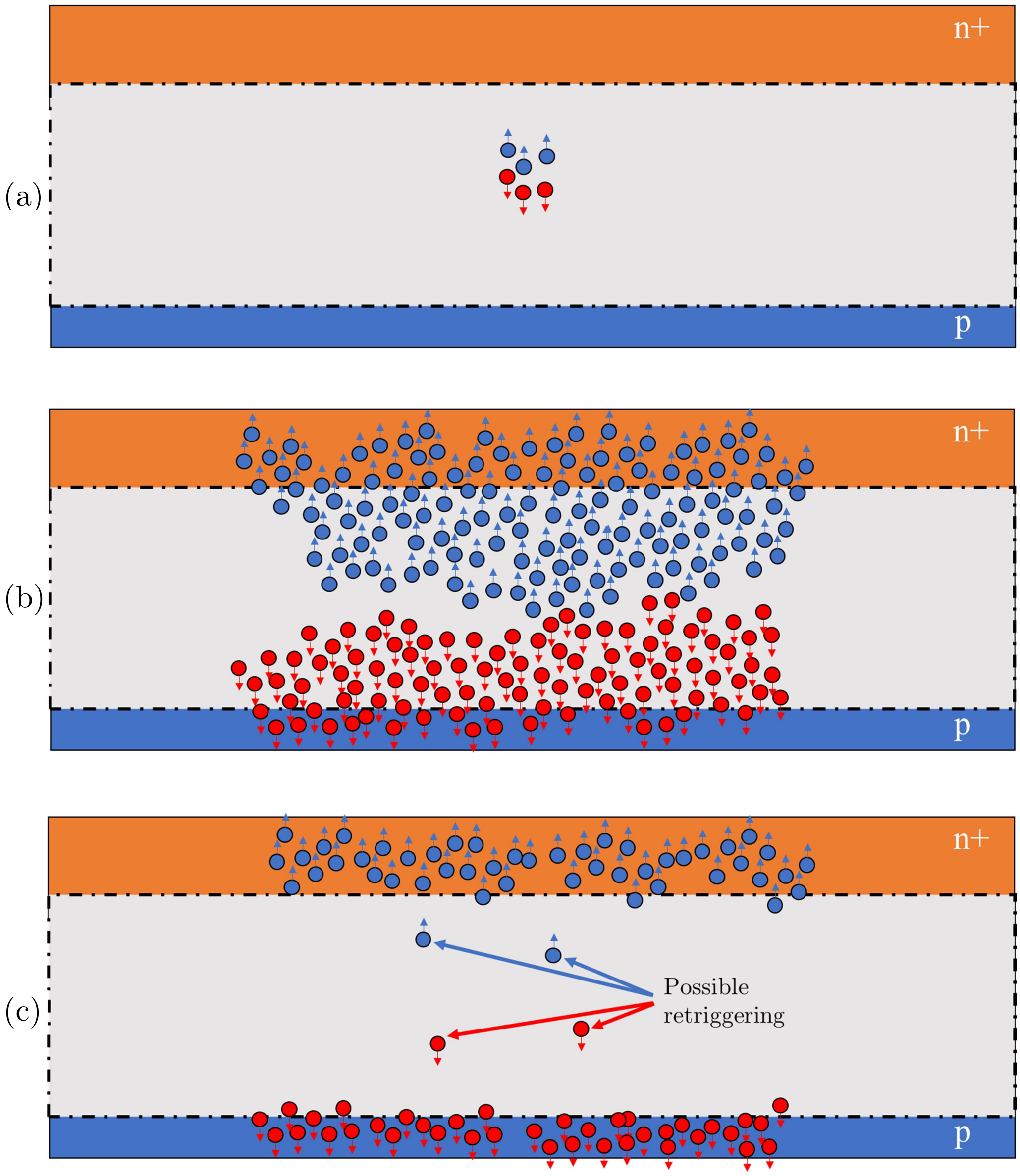

Figure III-11 : Avalanche breakdown from start (a), to peak amplitude (b), to end (c) showing the cause of possible retriggering.

Figure III-11 provides a simplified graphical representation of the events occurring in the multiplication region of the SPAD, depicting the stages from the initiation (a) to the peak amplitude (b) and finally the termination of an avalanche (c). Notably, in stage (c), a small fraction of charges remains uncollected outside the SCR. These residual charges possess the potential to reignite the avalanche process if they are not promptly evacuated. Hence, the aim of the quenching criterion is to limit the charge accumulation within the high electric field multiplication region to a predetermined threshold.

The charge-based quenching criterion ensures that the avalanche is terminated when the charge accumulated inside the multiplication region falls below a specified value. This threshold typically corresponds to a few hundred charges, equivalent to approximately 10^{15} C, to prevent prolonged convergence of the model. It is crucial to set the threshold at an appropriate level to avoid situations where the exponential decrease in carriers during the recharge phase fails to reach zero, leading to incorrect simulation results.

By incorporating the charge criteria into the model, the quenching process becomes more robust and realistic, reflecting the physical behavior of the SPAD. This refined approach enables accurate control of the avalanche termination and ensures that residual charges do not linger within the multiplication region, preventing undesired retriggering of subsequent avalanches. Moreover, it enhances the convergence of the model by effectively managing the carrier dynamics and accurately simulating the recharge phase.

The charge-based quenching criterion is a valuable addition to the SPAD model, as it addresses the limitations of the previous gain-based criterion and provides a more comprehensive understanding of the quenching process. This advancement contributes to the overall fidelity and reliability of the SPAD model, facilitating accurate predictions of its performance.

3.2 Non-quenching effect correction

The transition from a gain-based quenching criterion to a charge-based criterion in the SPAD model brings various advantages, particularly in terms of convergence and the overall functionality of the SPAD recharge cycle. However, this shift also revealed an interesting phenomenon where the SPAD undergoes triggering, generates an avalanche, but fails to quench and recharge properly. As a result, the device exhibits a constant current and voltage behavior akin to that of a Zener diode.

The occurrence of this behavior is contingent upon the value of the quenching resistor used in the model. If the resistor value is set too low, the asymptotic diode current I_{f} surpasses the latching current level I_{q}. Consequently, the avalanche sustains itself indefinitely, leading to a self-sustained current flow in the SPAD. This intriguing resemblance to a Zener diode highlights the interplay between the quenching resistor and the device's behavior.

It's worth noting that the behavior of the SPAD as a Zener diode-like voltage reference can be exploited in certain circuit applications. However, caution must be exercised to ensure that the power dissipated by the SPAD remains within safe limits, preventing potential permanent damage to the device.

The observation of this Zener-like behavior in the SPAD adds a layer of complexity and versatility to its potential applications. While the primary focus of the SPAD is single-photon detection, this unconventional characteristic opens up possibilities for its use as a voltage reference or in specialized circuit designs. It underscores the importance of considering and optimizing the quenching resistor value to achieve the desired operating characteristics and prevent unintended behavior.

Further investigation and characterization of the Zener-like behavior in the SPAD can provide valuable insights into its underlying mechanisms and potential applications. Additionally, exploring techniques to mitigate or control this behavior can enhance the versatility and reliability of the SPAD in various photonics and optoelectronics applications.

To ensure proper operation and avoid the unintended constant current behavior resembling that of a Zener diode, it is recommended to select high-value quenching resistors when working with SPAD diodes in applications. It should be noted that the constant current behavior is rarely desirable in SPAD applications, as there are other components more suitable for fulfilling that specific purpose.

When employing the gain-based quenching criterion, the Zener-like effect discussed earlier was absent. However, with the implementation of the improved charge-based quenching criterion, this effect has been fully incorporated into the SPAD model. Instead of relying on the gain criterion, the quenching decision is now based on the presence or absence of a specific amount of charge within the multiplication region of the SPAD. This charge-based quenching criterion has been demonstrated in simulations, as depicted in ***Figures III-12*** to ***III-14***.

The comparison between the two solutions clearly showcases the effective implementation of the Zener-like effect using the charge quenching criteria. Furthermore, the manifestation of this behavior is dependent on the chosen value of the quenching resistor. By varying the voltage in the simulations, covering a range from 17V to 25V, the impact of different voltage levels on the SPAD's behavior can be observed. It is worth noting that a reference voltage of 17V, deliberately chosen below the breakdown voltage of the SPAD, is included to monitor and ensure that the device does not trigger below its breakdown threshold.

These simulations provide valuable information on the correlation between the quenching resistor, voltage levels, and the resulting behavior of the SPAD. Careful selection of the quenching resistor value is crucial to achieve the desired functionality and prevent unwanted effects, such as the Zener-like behavior discussed.

In ***Figure III-12***, a comprehensive comparison is presented to demonstrate the impact of different quenching resistor values on the behavior of the SPAD when using the charge quenching criteria. It is evident that a quenching resistor value of Rq = 1k**Ω** is insufficient to effectively quench the SPAD, as indicated by the continuous operation of the device. In contrast, the gain criteria still manage to quench the

SPAD, even when it should not be quenched based on the charge criterion. This observation highlights the limitations of the gain-based quenching approach compared to the charge-based criterion.

Moving on to ***Figure III-13***, a critical turning point can be observed at a quenching resistor value of Rq = 100k**Ω**. At this threshold, the SPAD is only marginally quenched for VHV biasing voltages exceeding 18V. It is worth noting that the gain criterion (depicted at the bottom of the graph) consistently quenches the SPAD, irrespective of the quenching resistor value. Additionally, the graph reveals instances of premature quenching at specific VHV values, illustrating the issue previously discussed regarding the gain quenching criterion.

Figure III-14 shows the operation of the SPAD using the standard KNACK testchip's quenching resistor value, set at Rq = 758k**Ω**. At this resistor value, the SPAD is successfully quenched at all VHV biasing voltages tested. This finding confirms the effectiveness of the charge quenching criterion in achieving the desired quenching behavior across a range of voltage levels.

These detailed graphical representations offer a comprehensive understanding of the effects of different quenching resistor values at multiple biasing voltages on the resultant behavior of the SPAD. It is evident that the choice of quenching resistor plays a vital role in determining the quenching effectiveness and overall functionality of the SPAD. The comparison between the charge and gain criteria further emphasizes the advantages of the charge-based quenching approach, providing more accurate and reliable quenching behavior.

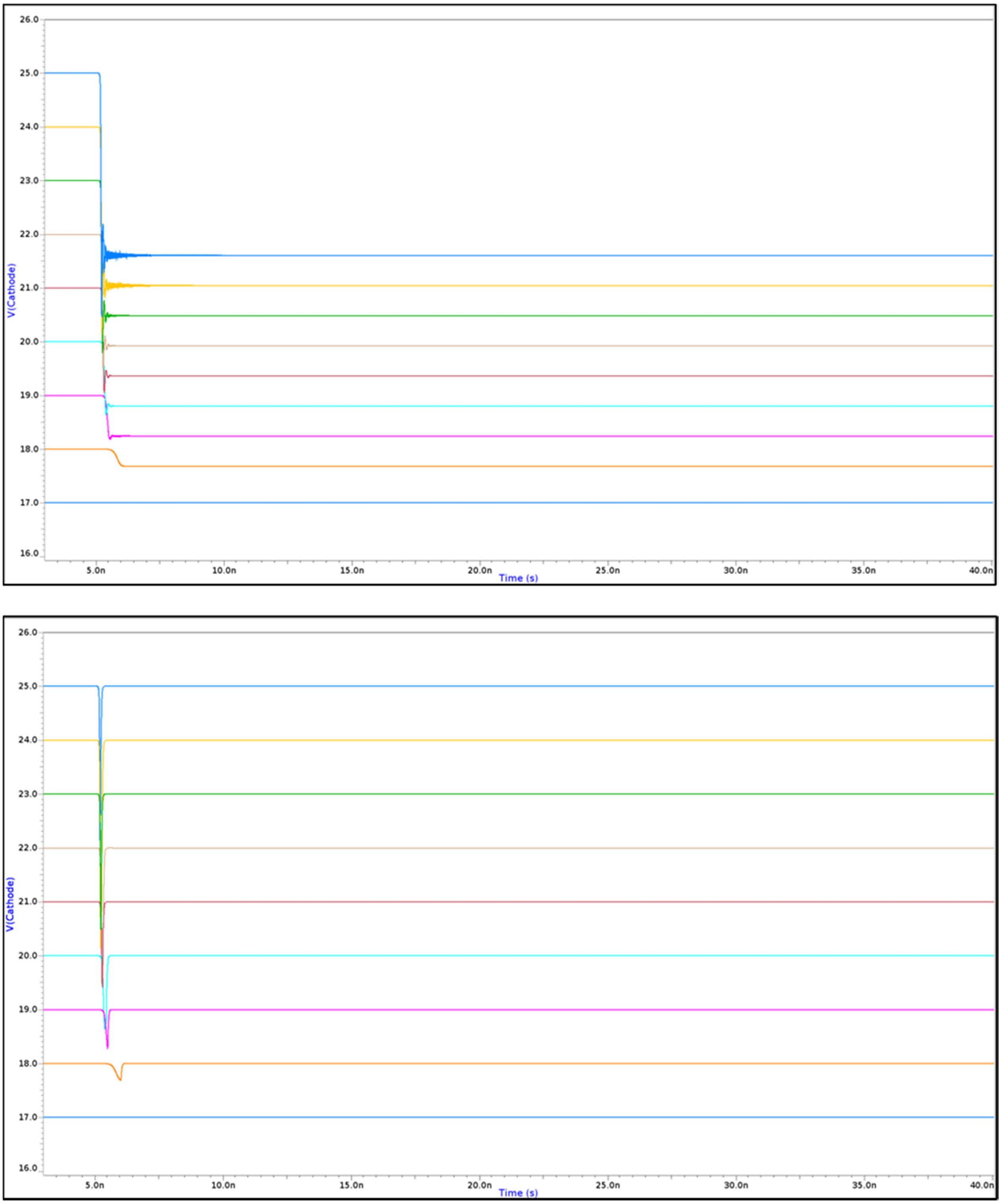

Figure III-12: SPAD quenching simulations at R_q = 1kΩ, using charge quenching criteria (top) and gain quenching criteria (bottom)

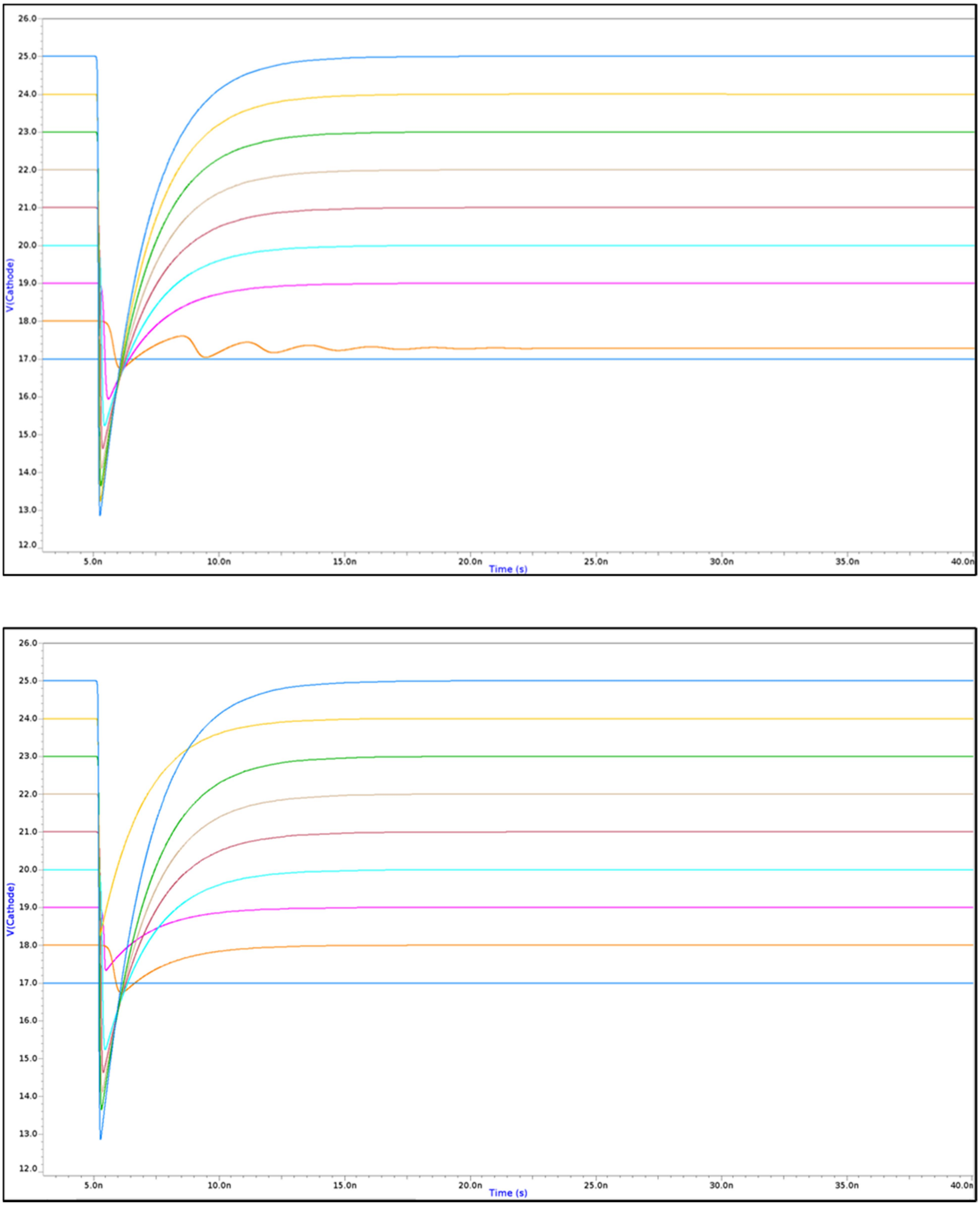

Figure III-13: SPAD quenching simulations at $R_q = 100k\Omega$, using charge quenching criteria (top) and gain quenching criteria (bottom)

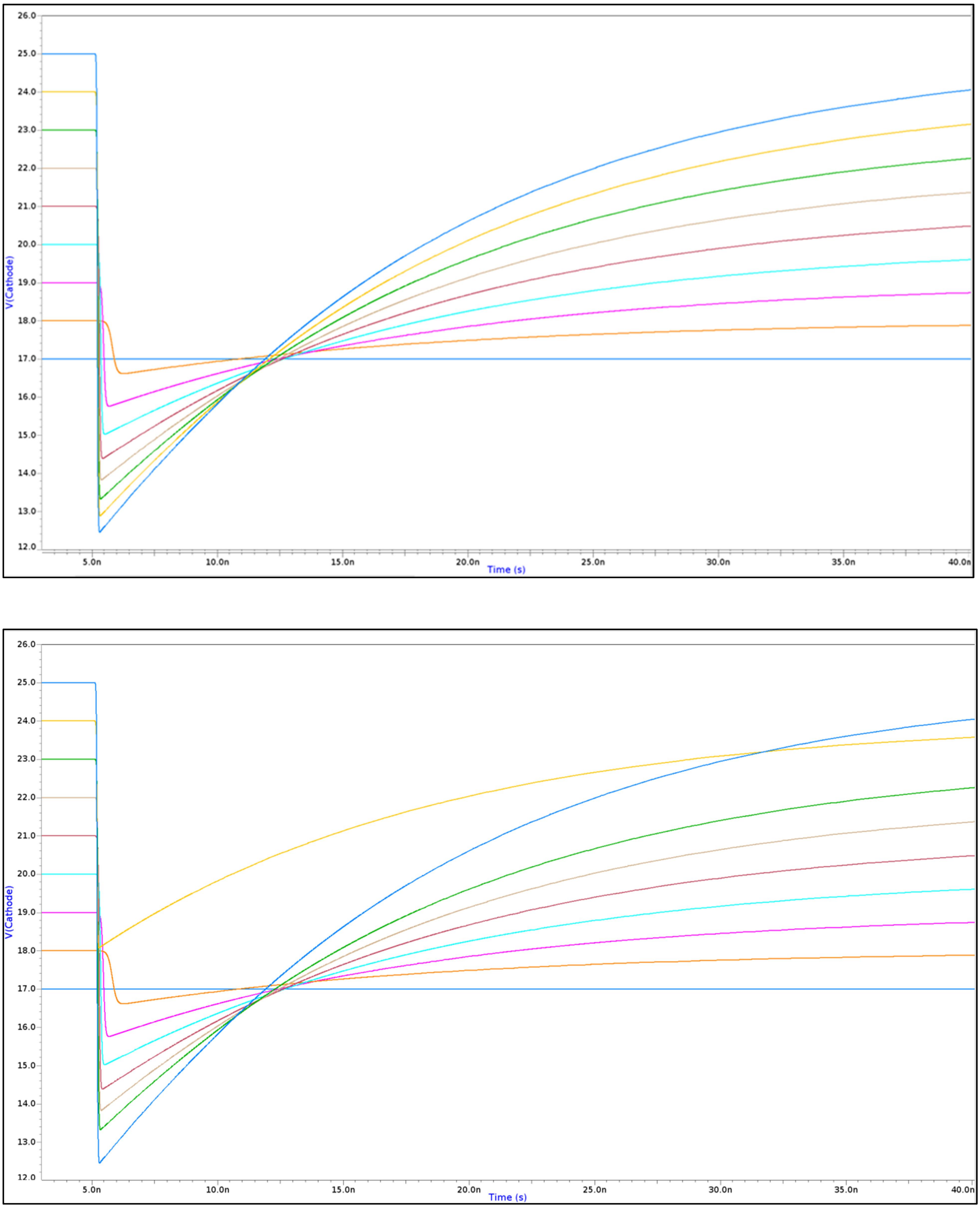

Figure III-14: SPAD quenching simulations at $R_q = 758k\Omega$, using charge quenching criteria (top) and gain quenching criteria (bottom)

3.3 Breakdown voltage modification

To address the discrepancy between the simulated and measured breakdown voltage of the SPAD, a thorough investigation was conducted, resulting in the implementation of corrective measures in the model. The theoretical breakdown voltage, determined through TCAD simulations [61], was initially estimated to be 17.2V. However, during the characterization process of the physical device, it became apparent that the actual breakdown voltage exceeded this value.

To rectify this disparity, a similar approach used for gain correction, as discussed in the previous section, was applied to adjust the breakdown voltage. Since the model presented in this paper derives its breakdown voltage ($V_{\mathbf{br}}$) directly from the closed loop gain values, a correction factor γ was introduced. Multiple values of γ were considered during the subsequent simulation analyses to determine the optimal correction factor, leading to a more accurate approximation of the measured behavior of the SPAD.

The simulation graphs presented in the following sections demonstrate the impact of different correction factors on the breakdown voltage of the SPAD. By carefully calibrating the correction factor, the model can be aligned with the measured data, enabling a more reliable representation of the SPAD's electrical characteristics. This iterative process of calibration and adjustment contributes to enhancing the accuracy and predictive capability of the model, ensuring its applicability in various practical scenarios, such as device optimization, system design, and performance evaluation.

In ***Figure III-15***, the impact of varying breakdown voltage ($V_{\mathbf{br}}$) values on the model's behavior is illustrated through a series of simulations. These simulations provided valuable insights, leading to two main observations that contribute to a comprehensive understanding of the SPAD's characteristics:

- The first observation highlights the relationship between breakdown voltage and the voltage drop on VPSNODE. As expected, the simulations revealed that a lower breakdown voltage corresponds to a higher voltage drop on

VPSNODE. This outcome is a natural consequence of the relationship between breakdown voltage and impact ionization gain values. At a fixed biasing voltage, a lower breakdown voltage leads to higher impact ionization gain, resulting in an amplified voltage drop across VPSNODE. This behavior aligns with the fundamental principles governing SPAD operation and serves as a confirmation of the model's accuracy in capturing the expected behavior.

- The second observation pertains to the variation in VPSNODE voltage drop as the breakdown voltage (V_{br}) of the SPAD is modified. Interestingly, the simulations exhibited a larger spread in VPSNODE voltage drop values as the breakdown voltage of the SPAD increased. This finding suggests that higher breakdown voltages introduce more pronounced variations in the voltage drop across VPSNODE. The reasons underlying this observation can be attributed to the intricate interplay between breakdown voltage, impact ionization gain, and biasing conditions. By analyzing the spread in VPSNODE voltage drop at different breakdown voltage settings, valuable insights can be gained regarding the sensitivity of the SPAD to variations in breakdown voltage and its potential implications on device performance.

By closely examining the simulation results presented in ***Figure III-15***, a deeper understanding of the SPAD's behavior under different breakdown voltage conditions is attained. These observations contribute to a comprehensive characterization of the SPAD, providing valuable information for device optimization to better fit the measurements.

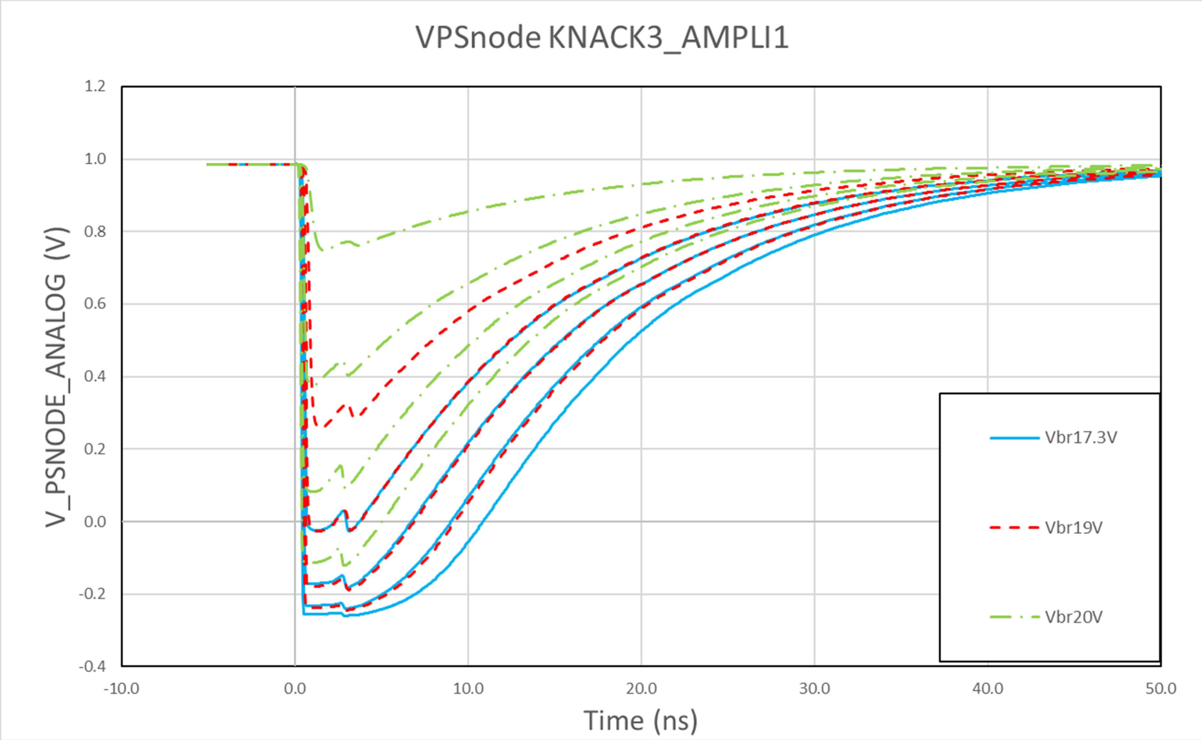

Figure III-15: VPSNode transient simulation responses for three different values of V_{br}, at four polarization voltages VHV = 22V, 23V, 24V, 25V.

After a meticulous process of fine-tuning the breakdown voltage to align the simulation curves with the experimental characterization results, the most accurate value for achieving a close fit was determined. ***Figure III-16*** showcases the adjusted simulation curves with a breakdown voltage $V_{br} = 20.5V$. The fit has been made to the first measurable value, VHV = 22V, and demonstrates an exceptional agreement with the measurement data. However, when examining the subsequent curves corresponding to different VHV values, it becomes evident that they exhibit a significant spread and deviate significantly from the measured results.

The observed discrepancy between the simulation and measurement data highlights the limitations of solely modifying the breakdown voltage in the device compact model. Therefore, an alternative modification, described in the subsequent section, was thoroughly considered to enhance the model's accuracy, and establish a closer correspondence with its physical counterpart: Electric Field Degradation (EFD).

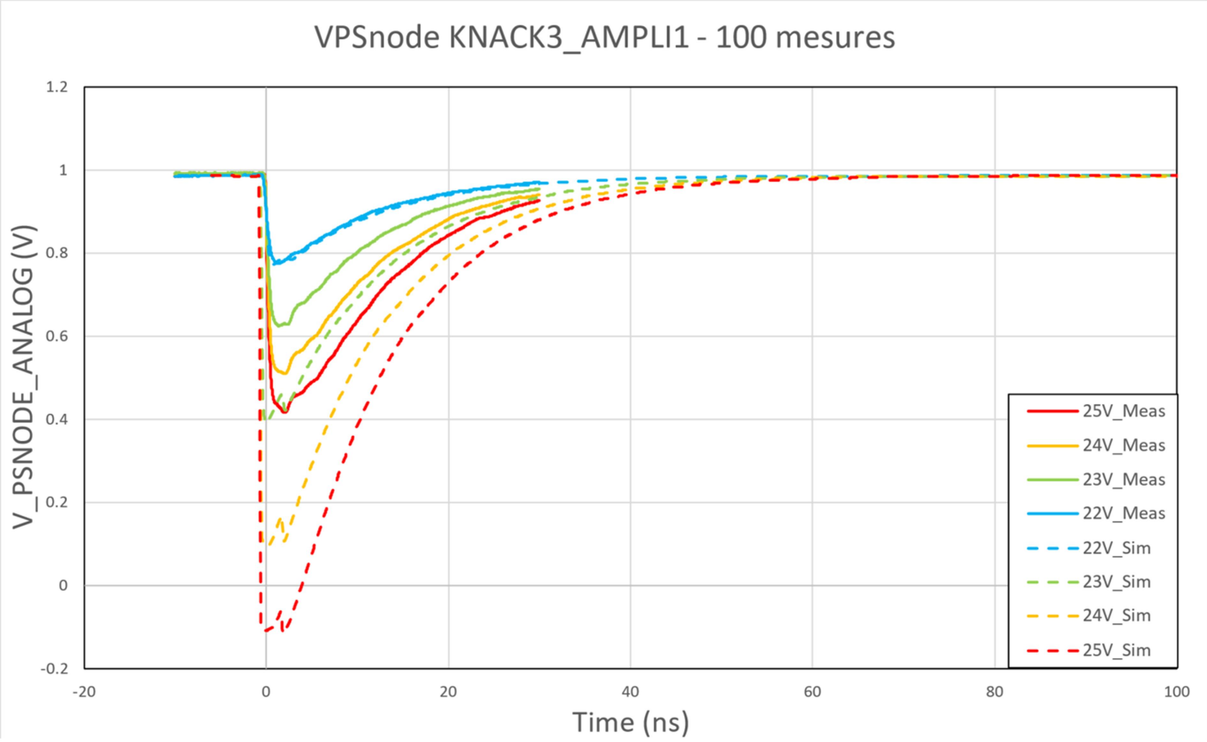

Figure III-16: `VPSNODE` transient comparison between measurements (solid) and simulations (dotted). Value of V_{br} set to 20.5V for the model.

By introducing the concept of EFD, a mechanism that accounts for the degradation of the electric field within the SPAD, an additional degree of sophistication was incorporated into the model. This modification aims to capture the non-ideal behavior of the device and address the remaining discrepancies between the simulation and measurement data. The inclusion of EFD allows for a more comprehensive representation of the physical processes occurring within the SPAD and provides a means to refine the model's predictive capabilities.

Through the integration of the EFD mechanism into the model, a more nuanced understanding of the SPAD's behavior should be achieved. This advancement not only aims to enhance the accuracy of the simulations but also facilitates a better alignment with the actual device's characteristics. The subsequent sections will delve into the details of the EFD implementation and its impact on the overall performance and agreement between the model and experimental data.

3.4 Electric Field Degradation

Following the initial calibration attempt, which involved modifying the breakdown voltage of the SPAD, it became apparent that this adjustment alone did not yield satisfactory results when compared to the experimental characterization data. The need for further improvement was evident, and a new perspective was developed to enhance the accuracy of the model: the consideration of a phenomenon occurring within the multiplication region known as Electric Field Degradation (EFD).

Electric Field Degradation is a physical phenomenon that arises due to the generation of an exceedingly large number of mobile carriers during the avalanche breakdown of the SPAD. When the device enters the avalanche breakdown state, electrons and holes experience rapid movement towards opposite sides within the high electric field of the multiplication region. This grouped movement of oppositely charged particles creates an electric dipole, an artificially formed configuration that produces an electric field opposite in direction to the electric field within the space charge region (SCR). Importantly, the strength of this opposing electric field is directly proportional to the number of carriers generated during the avalanche process, as a high electric field implies a stronger avalanche breakdown in the junction.

To provide a clearer understanding of this phenomenon, ***Figure III-17*** offers a visual representation of the entire volume of the SPAD device, illustrating the dynamic movements and distribution of carriers at two distinct moments in time. The figure highlights the significant carrier quantity and mobility within the multiplication region and its impact on the overall electric field distribution.

By incorporating the EFD into the model, a crucial element is introduced to capture the non-linear behavior observed within the multiplication region during avalanche breakdown. This inclusion enables a more accurate representation of the SPAD's performance by accounting for the intricate interplay between carrier movement, electric field distribution, and resultant changes in the electrical characteristics of the device.

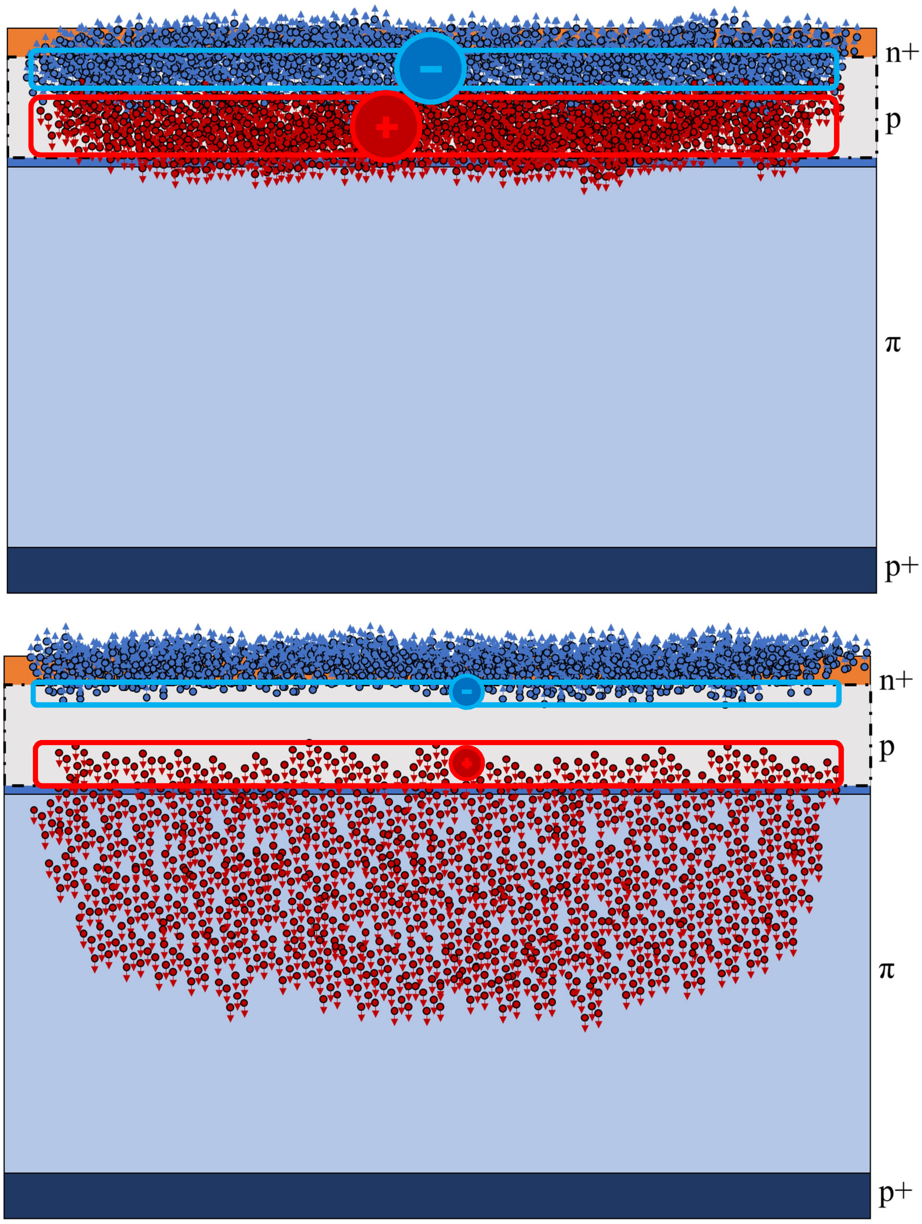

Figure III-17 Avalanche breakdown process with visible dipole creation via avalanche generated charge carriers at two different timestamps, at the peak of the avalanche and in the end of the avalanche.

With a refined model that contains the EFD, the aim is to better emulate the physical behavior of the SPAD during avalanche breakdown and improve the model's predictive capabilities. By considering the dynamic interaction between carrier movement, electric field distribution, and the underlying device characteristics, we can further refine the model and establish a closer alignment with the experimental observations obtained from characterization data.

The behavior of EFD during the avalanche breakdown of the SPAD is influenced by the amount of charge present within the multiplication region. At the peak of the avalanche, when the charge is maximized, a strong dipole is formed, leading to a significant electric field degradation. This degradation is a consequence of the induced dipole, which generates an electric field that opposes the electric field within the SCR. As the avalanche quenching process commences and the SPAD begins recharging, the contribution of EFD gradually weakens. Towards the end of the avalanche, the number of charges involved in EFD becomes negligible, no longer impacting the behavior of the SPAD.

To quantify the instantaneous charge within the SCR, a calculation method is employed, involving the integration of the instantaneous current flowing within the loop. This calculated charge is then subtracted from the collected current exiting the loop. By utilizing this approach, only the charges that actively participate in Electric Field Degradation are considered in the determination of the overall electric field within the SCR. The expression for the total instantaneous charge inside the SCR, taking into account both the positive and negative charges, is represented by the following equation ***Eq. III-3***:

$$\begin{aligned}\text{avltotcharge} &= \left(\int \text{V(ND[n\_loop])} - \int \text{V(NC[n\_loop])}\right) \\ &+ \left(\int \text{V(PD[n\_loop])} - \int \text{V(PC[n\_loop])}\right)\end{aligned} \quad \textbf{\textit{Eq. III-3}}$$

Here, the integration involves the voltage differences across the depletion regions of the n-type (ND and NC) and p-type (PD and PC) semiconductors, respectively. This charge value is subsequently used in the electric field expression, where it is subtracted from the contribution of the space charges within the SCR. As a result, the total electric field within the multiplication region experiences degradation as the avalanche progresses, leading to a weaker avalanche overall.

By accounting for the impact of EFD on the electric field distribution within the SPAD, the model becomes more comprehensive and accurate in capturing the device's behavior during avalanche breakdown. This refined understanding of the interplay between charge distribution, electric field degradation, and avalanche dynamics is crucial for improving the predictive capabilities of the model and achieving better alignment with experimental observations. In the subsequent sections, we will further explore the implications of EFD and delve into the mechanisms that govern its influence on the SPAD's performance.

In an effort to achieve the closest possible agreement with experimental measurements, a series of simulations were conducted, incorporating different levels of EFD. The simulations aimed to assess the impact of EFD on the behavior of `VPSNODE`, a critical parameter of the SPAD, under varying biasing conditions. ***Figure III-18***, presented below, illustrates the simulation results for VPSNODE at four different biasing voltages (ranging from 22V to 25V) with three distinct EFD percentage values.

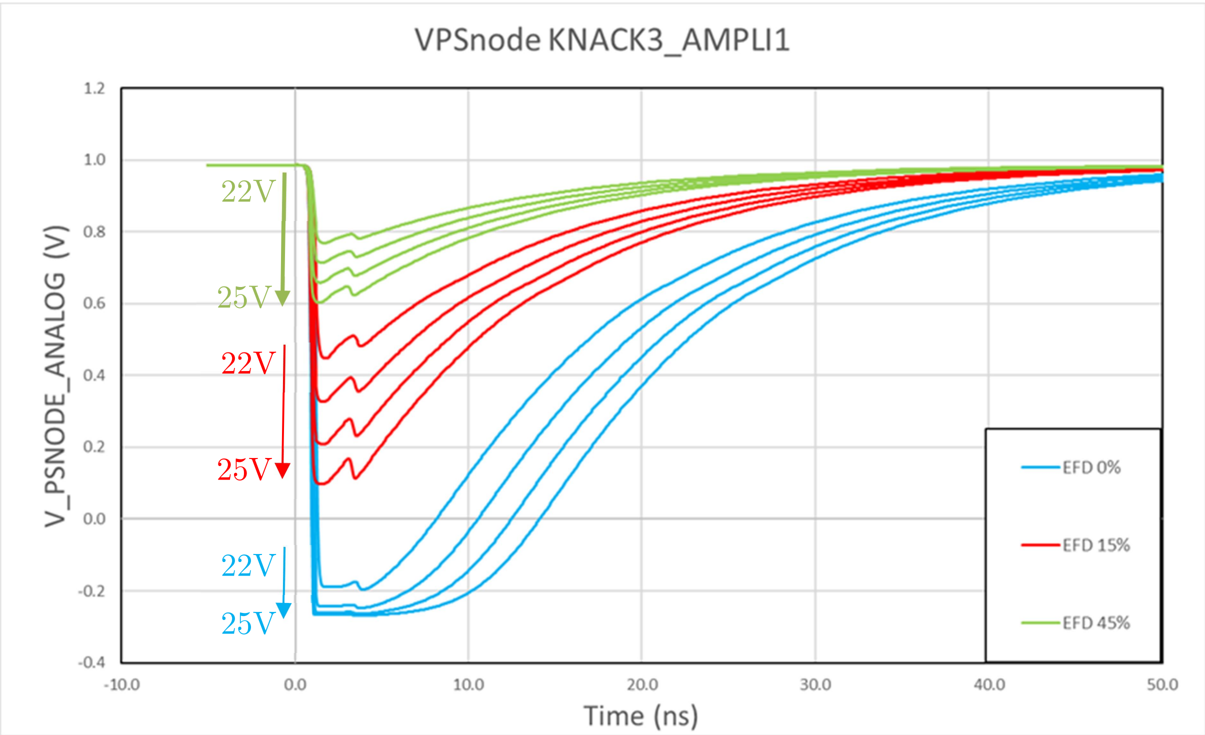

Figure III-18: **`VPSNODE`** ***simulation at three different EFD percentage values: 0% for a reference and both 15% and 45%.***

By systematically adjusting the EFD contribution, the simulations explored the effect of varying degrees of electric field degradation on the SPAD's performance. These different EFD percentages allowed for a comprehensive analysis of the relationship between biasing voltage, EFD, and the resulting behavior of VPSNODE. The objective was to identify the EFD percentage that would yield simulation results closest to the measured data, thus providing the most accurate representation of the SPAD's actual operation.

By comparing the simulated VPSNODE values at each biasing voltage across the different EFD percentages, valuable insights were gained into the influence of EFD on the SPAD's response. The simulations aimed to capture the nuances of how EFD affects the voltage at VPSNODE, considering both the amplitude and shape of the waveform.

The simulation results presented in ***Figure III-18*** serve as a visualization of the impact of different EFD percentage values on the SPAD's behavior under varying biasing conditions. The comparative analysis provides a valuable reference for selecting the optimal EFD percentage that best aligns the simulation results with the experimental measurements.

Upon analyzing the simulation results presented in ***Figure III-18***, several noteworthy observations can be made regarding the influence of the EFD contribution percentage on the behavior of VPSNODE under varying biasing voltages (VHV).

One prominent finding is the inverse relationship between the EFD contribution percentage and the overall voltage values observed at different VHV biasing voltages. As the EFD contribution percentage increases, the resulting voltage values at VPSNODE exhibit a decreasing trend across the range of biasing voltages considered. This behavior indicates that a higher EFD contribution leads to a reduction in the magnitude of the voltage drop at VPSNODE, implying a more efficient quenching of the SPAD and a better overall control of its behavior.

Furthermore, as the EFD contribution percentage increases, there is a notable reduction in the spread of voltage drops observed at different VHV biasing voltages. The spread refers to the variability or range of voltage values observed at VPSNODE. This reduction in spread suggests that a higher EFD contribution percentage contributes to a more consistent and stable behavior of the SPAD across different biasing conditions. By mitigating the variability in voltage drops, the EFD mechanism aids in achieving a more reliable and predictable performance of the SPAD.

It is worth mentioning that in the reference simulation where the EFD contribution percentage is set to 0%, the spread of voltage drops is not clearly visible. This is primarily due to the exceedingly high voltage drops observed, which saturate the RF amplifiers. The saturation of the amplifiers leads to a limited range of detectable voltage values, making it challenging to observe the spread. This highlights the importance of introducing an appropriate level of EFD contribution to maintain voltage values within a manageable range, ensuring reliable and accurate coincidence with measurements.

Upon analyzing the results of the adjusted model when compared to the measurements, as presented in ***Figure III-19***, it becomes apparent that achieving an accurate fit to the measured characteristics of the VPSNODE transient requires careful consideration of the EFD contribution percentage. The simulation results reveal that the most favorable fit, closely resembling the measured data, is obtained when the EFD contribution percentage is set to 45%. At this contribution level, the simulated curve aligns well with the measurements, particularly for the first biasing voltage of VHV = 22V.

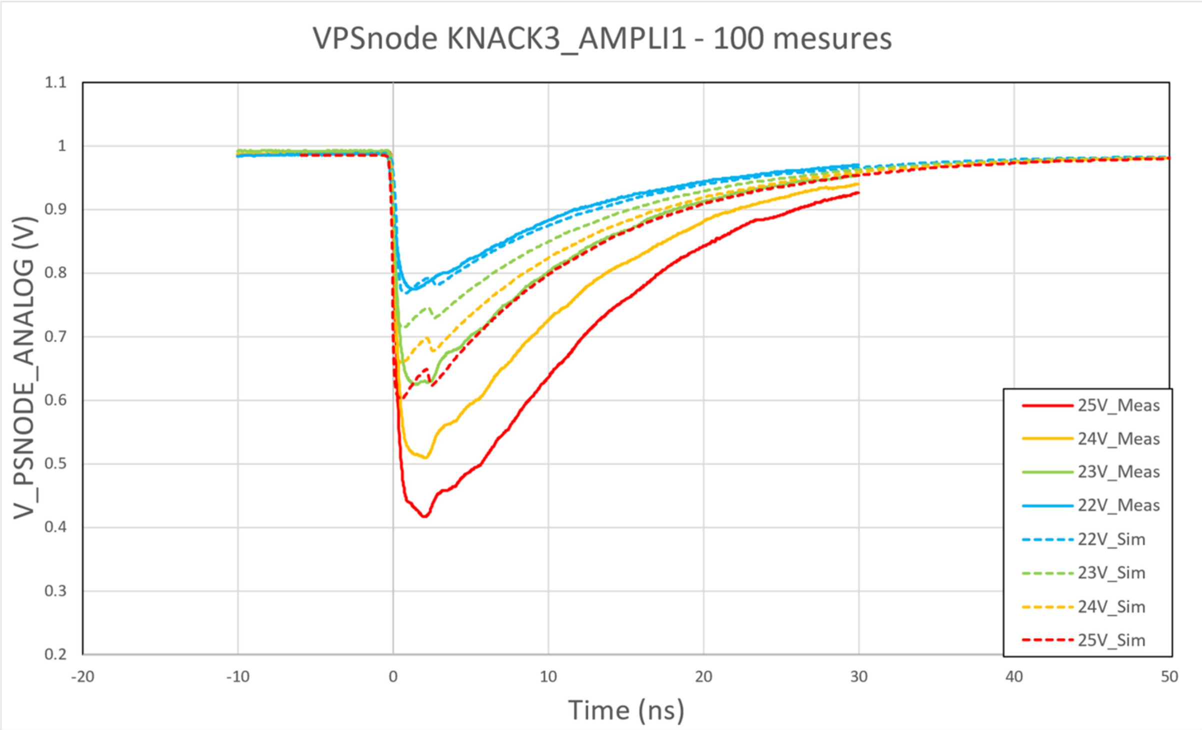

Figure III-19: Best compromise in EFD contribution percentage (45%) to fit the simulations with the first measurement transient ($VHV = 22V$)

However, it is important to note that while the simulation accurately captures the behavior of the SPAD for the initial biasing voltage, the subsequent voltage transients do not adequately describe the device's behavior. In these cases, the voltage swings are noticeably lower in magnitude and tightly clustered together, suggesting a deviation from the expected response. This indicates that the EFD phenomenon significantly influences the voltage swing values and their distribution, emphasizing its impact on the overall behavior of the SPAD.

Nevertheless, it is worth noting that similar to the breakdown voltage tuning discussed earlier, the integration of the EFD phenomenon alone is not sufficient to comprehensively describe the device's behavior and achieve a precise fit to the characterization results across the entire voltage range. While the EFD modification enhances the model's accuracy by accounting for the intricate mechanisms impacting the electric field within the multiplication region, it needs to be complemented by other adjustments.

Considering the limitations of relying solely on breakdown voltage tuning or EFD integration, a combined approach encompassing both aspects was pursued to enhance the model's accuracy and align it more closely with the characterization results. The subsequent section will delve into the details of this combined approach, providing insights into how the breakdown voltage tuning and EFD integration work in tandem to offer a more comprehensive understanding of the SPAD's behavior across the entire voltage range. By integrating both modifications, it is expected that the model will provide improved fidelity, capturing the intricacies of the SPAD's performance and achieving a better fit to the characterization data.

3.5 Combined tuning solution

As previously mentioned, breakdown voltage tuning and Eletric Field Degradation alone do not yield satisfactory results for the entire considered voltage range of the SPAD. When one of those parameters is fitted to measurements, the other becomes inaccurate. Therefore, conjointly using both calibrations is needed to achieve an accurate result realistically describing the SPAD's behavior.

The method used to obtain an accurate fit over the entire voltage range of the device consists of iteratively adjusting one parameter, followed by the other. First, the electric field degradation is set to 0% and the breakdown voltage is adjusted to fit the lower bound of the voltage range. Then, the EFD is progressively increased, and as noted in **Chapter III, subsection 3.4**, the EFD is influencing the spread of the SPAD's responses as the applied biasing voltage VHV increases. This property is used to increase the spread of the response curve progressively and iteratively as the lower bound is fitted to the experimental data.

Figure III-20 represents the final result of this fitting process with the combined effort of breakdown voltage and electric field degradation adjustments. The fit of those simulation results, when compared to the corresponding measurements, provides an accurate representation of the device's real response.

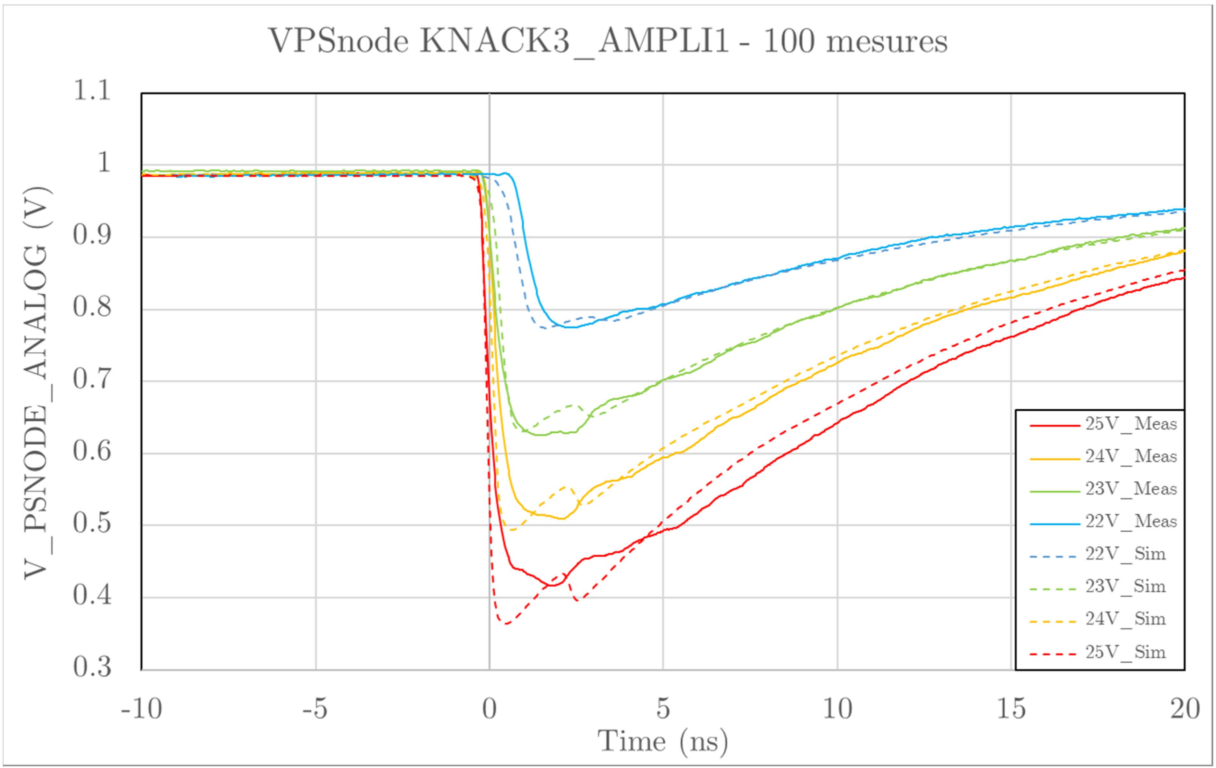

Figure III-20 : Combined tuning solution simulation results compared to measurements.

4 Parameter Extraction

Within the comprehensive calibration process of the model, one of the crucial stages in the design flow of compact modeling is parameter extraction. This extraction process is an essential component of the distributed model's calibration, ensuring accurate representation and predictive capability. To achieve this, the parameter extraction flow, as depicted in ***Figure III-21*** incorporates the utilization of three distinct sets of parameters, each serving a specific purpose and contributing to the overall accuracy of the model.

- The first set of parameters is extracted through careful analysis of AC C(V) and DC I(V) measurements conducted on a dedicated test structure. These measurements are essential for accurately describing the capacitance behavior of the device. The obtained capacitance values are then employed to calculate the electric field Fm within the multiplication region, a crucial factor in accurately simulating the device's performance.
- The second set of parameters is sourced from relevant literature, encompassing essential ionization coefficient parameters ($A_{\mathrm{n,p}}$, $B_{\mathrm{n,p}}$, $D_{\mathrm{n,p}}$) and carrier saturation velocities ($v_{\mathrm{n,p}_{\mathrm{sat}}}$). These parameters are well-documented and widely accepted within the scientific community. Incorporating these literature-based parameters ensures the model captures the fundamental physical characteristics of the device, aligning it with established knowledge and empirical data.
- The last set of parameters is derived from transient measurements of the analog voltage swings observed at the anode and cathode terminals. These measurements provide critical insights into the dynamic behavior of the device and enable adjustments of the impact ionization parameters $\beta_{\mathrm{n,p}}$ with γ_{W} and the electric field with the field degradation parameter avltotcharge. By refining these parameters based on the transient measurements, the model can more

accurately capture the intricate interplay between voltage dynamics and ionization processes, resulting in improved predictive capabilities.

The collective utilization of these three sets of parameters in the parameter extraction flow ensures a comprehensive and accurate calibration of the compact model. By combining the information obtained from dedicated test structures, established literature, and transient measurements, the model can effectively simulate the complex behavior of the device under various operating conditions. This meticulous parameter extraction process enhances the model's reliability and predictive power, making it a valuable tool in the analysis and design of advanced electronic devices.

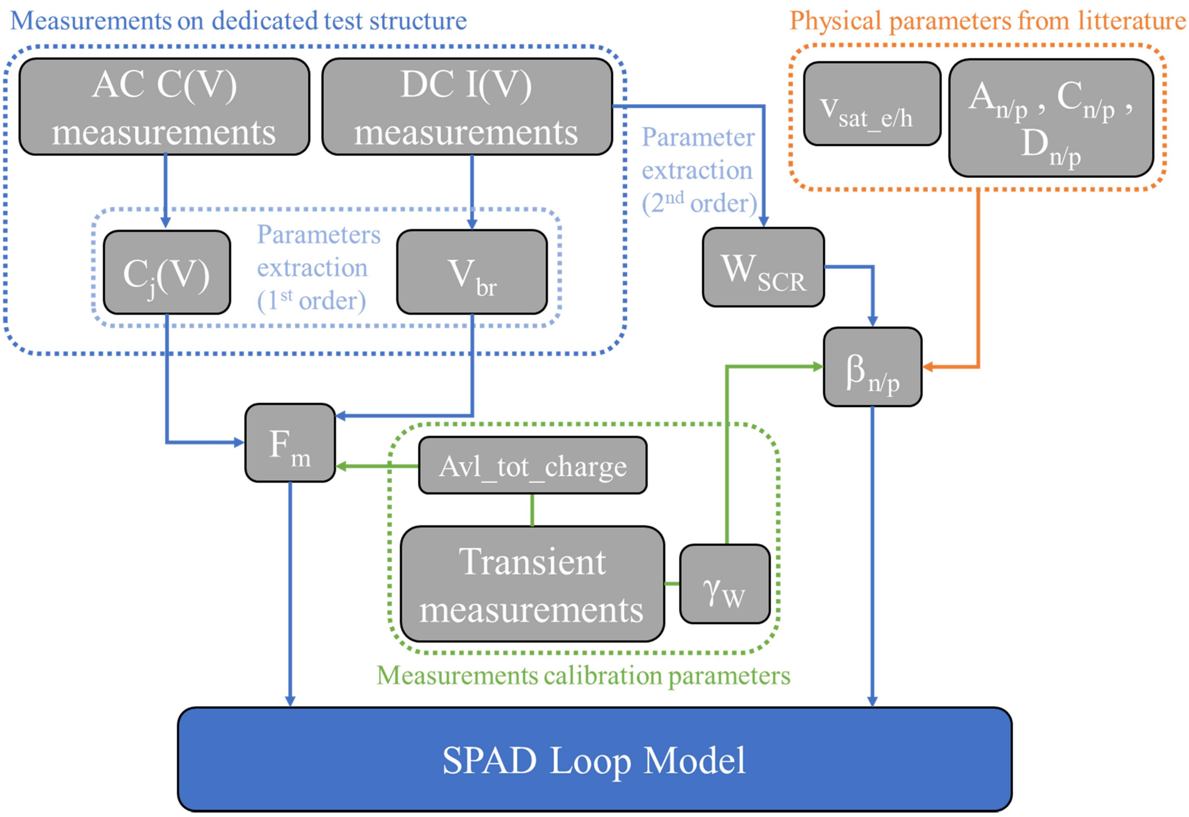

Figure III-21 Model extraction flow and different contributions of parameter sets coupled with their extraction method

4.1 Physical parameters from literature

As the model utilizes several physical equations to describe the behavior of the SPAD device, the first source of parameters was naturally scientific literature. As mentioned in the bibliography, one of the most important parameters for a functional SPAD model was the impact ionization coefficients. These coefficients are used to describe the avalanche breakdown mechanics inside the SCR of the diode and are a function of the internal electric field. In the developed model, coefficients found by Massey [25], which integrate a function of temperature, were used to describe the carrier multiplication.

Other important physical constants extracted from literature were the electron and hole saturation velocities. As the model uses a loop-based approach, taking into account the transit time of both electrons and holes, these parameters were essential for an accurate description of the carrier dynamics. As the electric field inside the multiplication region of the device is very high, the carriers experience an enormous amount of acceleration. This eventually leads to the carriers traveling at their respective saturation velocities $v_{\mathrm{n,p}_{\mathrm{sat}}}$.

4.2 DC I(V) and AC C(V) extracted parameters

An essential aspect of the model's parameter set was the extraction of various parameters from both DC I(V) measurements for static behavior parameters and from AC C(V) measurements for dynamic behavior ones. The two primary parameters of the model that were thus extracted and calibrated were the breakdown voltage of the device and its capacitance. This extraction was based on both TCAD simulations and measurements on the device test structure. The extraction relied mainly on figures of merit extracted from the SPAD's digital response, which did not require the complex and high-speed high-precision circuitry of the "KNACK" test structure that was not available at the time.

The extraction of the breakdown voltage and capacitance parameters was a critical step in accurately modeling the SPAD device's behavior. The breakdown voltage is the voltage at which the device experiences avalanche breakdown, and it is a crucial parameter for understanding the device's performance and overall behavior. The capacitance, on the other hand, is a measure of the device's ability to store charge, and it is essential for understanding the device's dynamic behavior.

To extract these parameters, both TCAD simulations and measurements on the device test structure were used. The TCAD simulations [61] provided a theoretical understanding of the device's behavior, mainly on its internal electric field, while the measurements on the device test structure provided data that could be used to calibrate the model. The figures of merit extracted from the SPAD's digital response were used to validate the accuracy of the model and ensure that it could accurately predict the device's behavior under different conditions.

4.2.1 DC I(V)

Extensive DC measurements were conducted on the studied device to extract its I(V) characteristic. These measurements provided valuable insights into the device's behavior, and several key parameters could be extracted from the I(V) curves, as shown in ***Figure III-22***. The parameters extracted from the I(V) curves included those linked to the reverse bias current, recombination current, and other critical aspects of the device's behavior.

One of the most crucial parameters extracted from the I(V) curves was the breakdown voltage of the device. This parameter is essential for describing the device's behavior under avalanche breakdown conditions. The change of slope, marked at the lower inflection point of the I(V) characteristic, indicates the beginning of avalanche breakdown inside the diode and is, therefore, a clear indicator of the device's breakdown voltage.

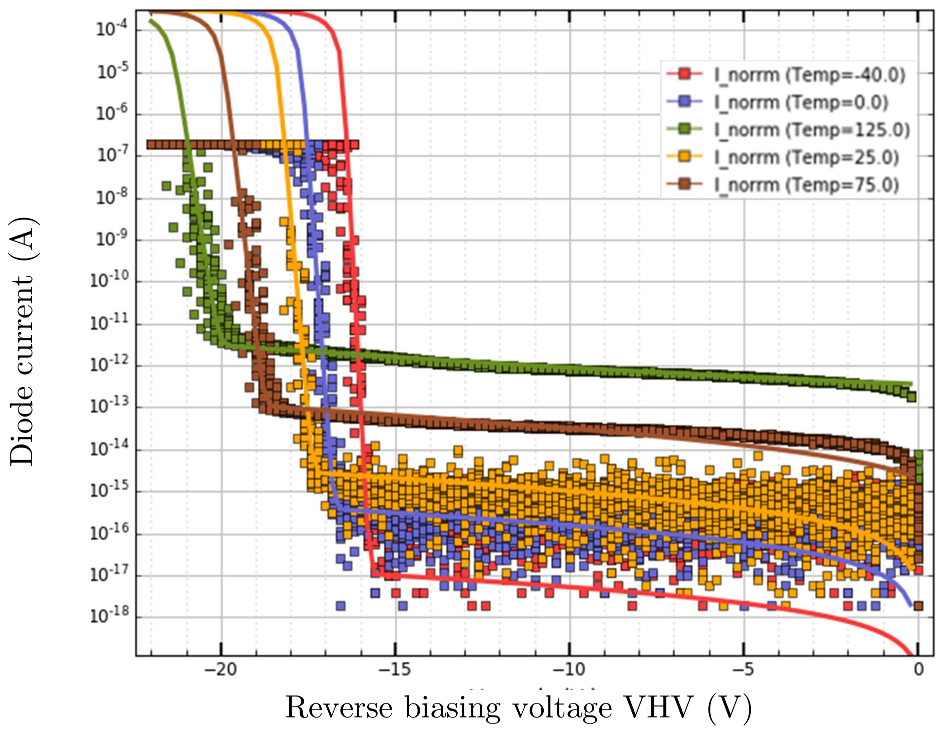

Figure III-22 I(V) SPAD characteristics, lower inflection point marks breakdown voltage region

The breakdown voltage can also be further confirmed by using TCAD data on the electric field inside the space charge region as a function of the applied reverse biasing voltage, as depicted in ***Figure III-23***. The first step is to trace the asymptotes of both regimes, the low inverse current regime and the avalanche breakdown "Geiger mode" regime. From there, the intersection point of those asymptotes is projected on the x-axis of the graph, providing an indication of the avalanche breakdown of the SPAD.

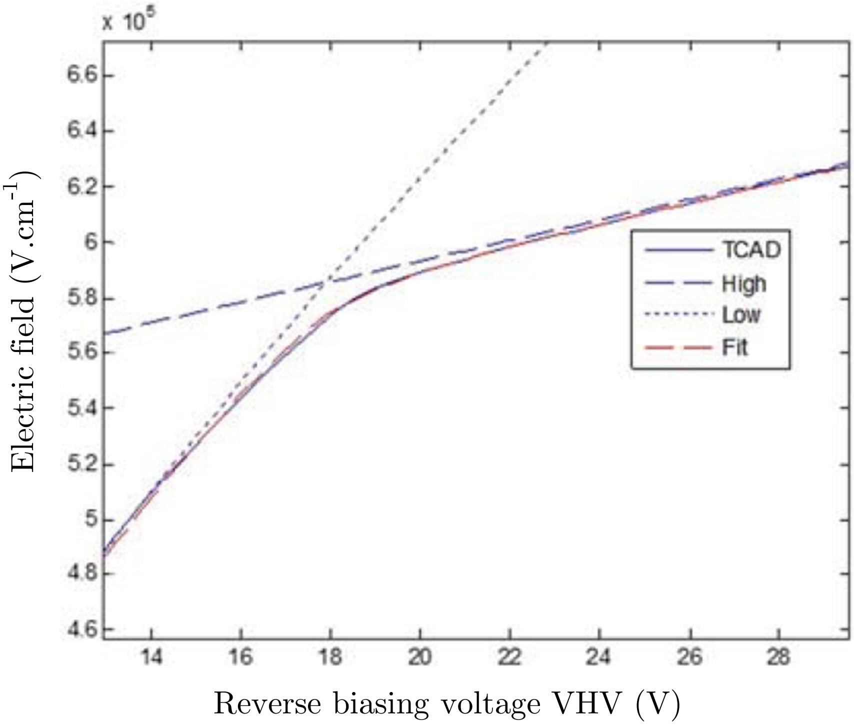

Figure III-23 Basic model of the electric field (red dashed line) as implemented in the SPAD model, calibrated on TCAD (solid blue). Two electric field regimes are considered (high and low). [67]

4.2.2 AC C(V)

The next step, and one of the most crucial ones in terms of modeling SPAD's avalanche breakdown behavior, is the electric field inside its space charge region. As the different n and p doping concentrations in the diode's junction create this depletion region, an electric field is formed across it. In the case of SPADs, this electric field is both very intense ($> 10^5$ V.cm^{-1}) and crucial in the impact ionization mechanism, leading to the avalanche breakdown.

Modeling this breakdown phenomenon demands precise information on the electric field inside the multiplication region of the device, hence the need for an accurate method to extract this field's values. To achieve this, a method based on the charges generated inside the junction, thus based on the capacitance of the device, was developed.

AC measurements were performed on the device, and C(V) curves were extracted. These curves served as a reference point to adjust the model's capacitance with various parameters. The junction capacitance of the SPAD is described as its charge as a function of the applied voltage at its terminals. Integrating this charge can, in turn, accurately describe the capacitance of the device. The description of the junction capacitance in this model is derived from the description of the [diode level2] model, readily available in the Eldo simulator documentation, with ***Eq. III-4 to III-9***:

IF $Vd \leq Fcj$:

$$Q_J = \frac{C_{J0}.pj}{1-mj}.\left(1-\left(1-\frac{Vd}{pj}\right)^{1-mj}\right)$$ ***Eq. III-4***

ELSE

$$Q_J = C_J\left(f1+\frac{1}{f2}.\left(f3.(Vd-Fcj)+\left(0.5\frac{mj}{pj}\right).(Vd^2-Fcj^2)\right)\right)$$ ***Eq. III-5***

WITH

$$Fcj = fc.Vd$$ ***Eq. III-6***

$$f1 = \frac{pj}{1-mj}\left(1-(1-fc)^{1-mj}\right)$$ ***Eq. III-7***

$$f2 = (1-fc)^{1+mj}$$ ***Eq. III-8***

$$f3 = 1 - fc.(1 + mj) \quad \textbf{\textit{Eq. III-9}}$$

The set of parameters pj, mj, fc, and the constant $CJ0$ are crucial for the calculation of the charge and, therefore, the overall capacitance of the device. $CJ0$ is the value of the capacitance at 0V biasing voltage ($Vd = 0$), mj is the gradation coefficient, pj represents the junction potential at equilibrium, and fc is the depletion coefficient. These parameters are used and extracted from the C(V) curves to properly calibrate the SPAD model's capacitance.

Figure III-24 shows the calibrated capacitance of the device and the corners represented by the over- and underestimation of the measurements made. This leads to overall maximal and minimal capacitance values to establish upper and lower boundaries.

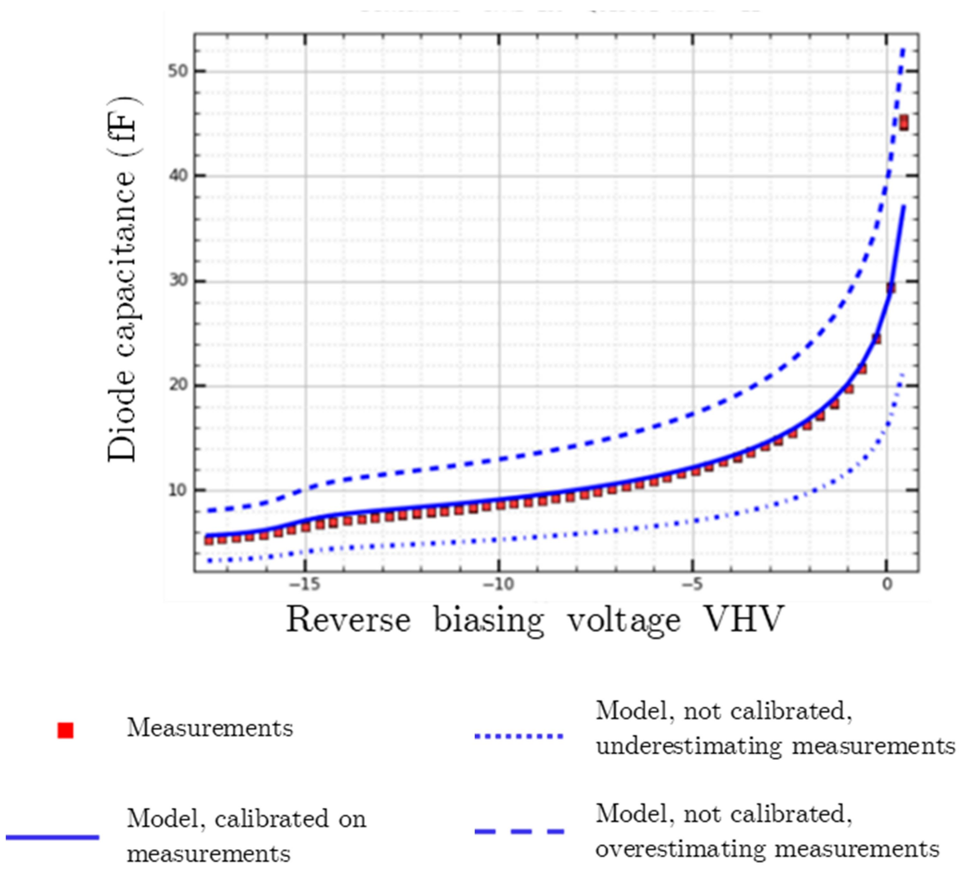

Figure III-24: SPAD Capacitance values, measurements compared to simulated model

4.3 Transient adjustment

The last parameter extractions and model adjustments are the calibration of the transient characteristics of the simulated model. This model tuning has been introduced in **Chapter III**, **Subchapter 3.3-3.5**. By combining the adjustment of both the effective breakdown voltage and the EFD contribution percentage of the model, the simulated response was adjusted to fit the characterization data. The temperature at which the simulations are conducted, 80°C, are determined by temperature measurements on the test chip, thanks to a thermal camera.

The first phase in this tuning and parameter extraction is to fit the model's response to the first exploitable transient measurement, in the case of this study, the first exploitable transient was at a biasing voltage `VHV` of 22V. This adjustment is made by tuning the value of the γ_W coefficient in the same way as presented in **Chapter II**, **3.3**, with the exception being that the transient response is considered as a reference, not the DC I(V) response. Added to that, in this first step, the EFD contribution is set to 0% initially. Once the simulated curve coincides with the measured transient, the second step of the adjustment is executed and the value of the EFD is increased.

Based on the result of the first breakdown voltage calibration, the spread between the different transients at varying `VHV` was observed to be too low, as shown in ***Figure III-16*** of **Chapter II**, **3.3**. To increase this spread, the EFD contribution must be gradually adjusted by varying the fraction of `avltotcharge` influencing the electric field. This second adjustment phase therefore consists in the incremental increase of the EFD contribution, proceeding by dichotomy. The first tested value is a contribution of 50%. Then, the breakdown voltage is adjusted as per the first phase to fit the simulated response on the first measurement curve (`VHV` = 22V). If the spread is still too low, the EFD is increased to 75%, if the spread is too high, it is reduced to 25%. This EFD adjustment followed by the breakdown voltage tuning are repeated alternatively until the simulated transient responses coincide with the characterization data over the entire voltage range the device has been tested at.

Conclusion

Building upon the framework of the Single-Loop base model and guided by the analysis of the SPAD architecture of the studied device, the distributed model shows great potential of further development.

The structure and architecture of the SPAD played a crucial role in the definition of the outline of the model. In the case of this study, the device is a reach-through type of SPAD, thus being divided into two distinct regions. First a low electric field absorption region which is especially wide to enable a good photon absorption and second, the multiplication region, where impact ionization occurs.

After establishing the base distribution of the model, based on those two regions, several key improvements to the model have been implemented to enhance the stability and accuracy of the model. The first implementation was the change of quenching criteria from a gain-based criterion to a charge-based criteria. The reason being the gain oscillations of the model causing premature and unwanted quenching. Then, linked to this change to a charge-based criterion, a non-quenching effect was also corrected. Indeed, by fixing a charge threshold below which the device quenches as normal, the model can accurately simulate the response when too low of a quenching resistor is used in the quenching circuitry. Indeed, when this case emerges, residual charges inside the device are not be properly evacuated and thus the SPAD does not quench properly, which could now be replicated by the model. Finally, as charges are generated in a great number when avalanche breakdown occurs, the electric field is influenced by the mobile carrier induced dipole. This electric field degradation played an important role in the following tuning of the model.

The next crucial step in evolving the distributed model from its basic form consisted in the tuning of the previously mentioned electric field degradation and breakdown voltage. Thanks to a clearly executed iterative adjustment method, the model is able to accurately simulate the transient behavior of the device that was measured through the expertly design test circuit.

Nevertheless, the model remains incomplete in terms of stochastic effects such as jitter, amplitude fluctuations, DCR and Afterpulsing. Due to unexpected problems in both the development of the model and the testchip arrival, the focus was shifted to improve the model's mean response and solidify its base to avoid future convergence or outlier problems. Several perspectives, either not entirely implemented and calibrated or in the beginning of the development process have already been explored and will be developed in the following **Chapter IV**.

Chapter IV Model Perspectives and conclusions

1 Perspectives

1.1 Delayed hole collection at the Anode

1.1.1 Observations

During the comprehensive characterization of the Knack test circuit, which incorporates the SPAD device, a notable observation emerged from the analysis of the transient Anode voltage measurements, depicted in Fig. IV-1. After approximately 3ns, a secondary peak in the voltage waveform becomes apparent, signaling the presence of an intriguing phenomenon changing the overall shape of the voltage transient.

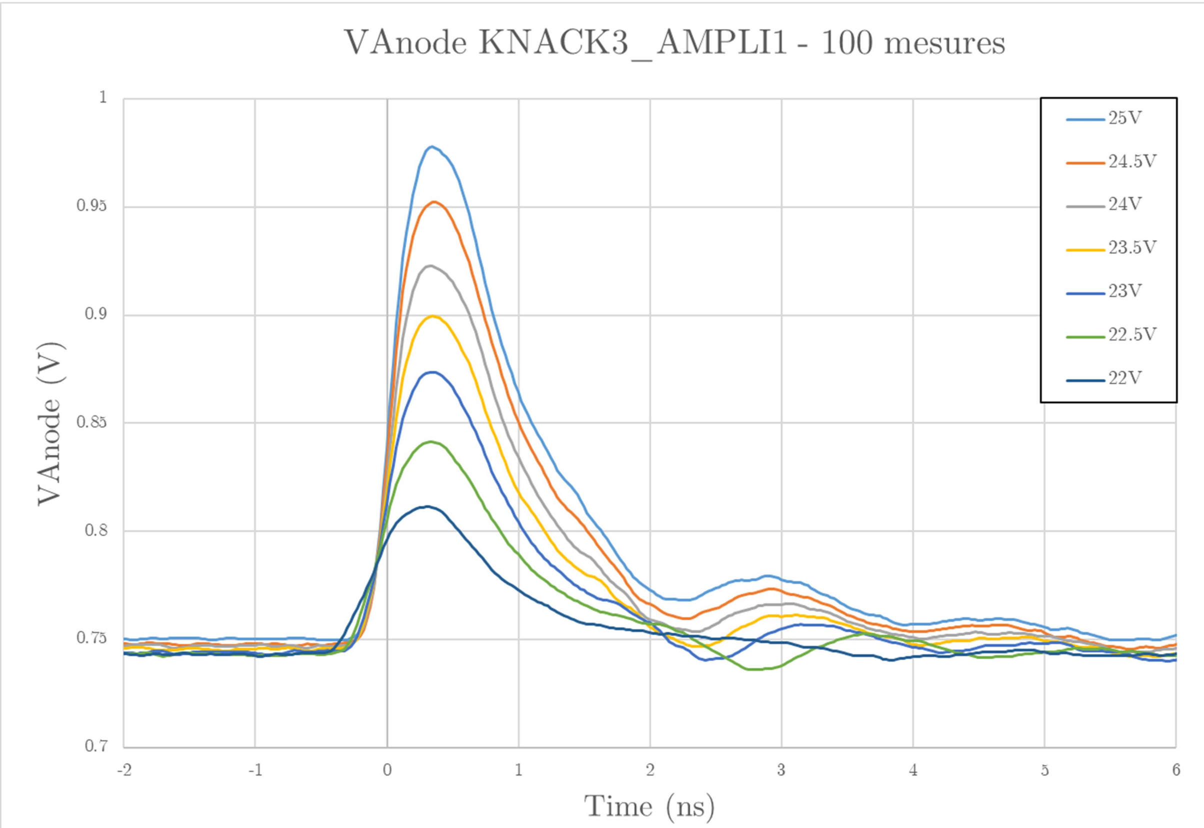

Figure IV-1: Anode transient voltage of the reference KNACK circuit over the entire active voltage range (22V-25V)

To shed light on this observation, a thorough examination was conducted, encompassing a comparative analysis between the measurement results and TCAD simulations of the SPAD device. This investigation aimed to establish a plausible explanation for the occurrence of the secondary peak in the transient Anode voltage.

Based on the findings, it was deduced that the secondary peak is primarily attributed to the delayed collection of holes generated during the avalanche breakdown of the SPAD. The unique structural characteristics of the SPAD, particularly its reach-through nature, play a pivotal role in this phenomenon. When an avalanche breakdown transpires, electrons generated within the device are swiftly collected by the cathode. This expedient collection process is facilitated by the relatively short distance that these carriers must traverse, namely the narrow n+ region, as they exit the high electric field multiplication region.

Conversely, the holes generated during the avalanche breakdown encounter a distinct journey. They must traverse a wider intrinsic π region, where the electric field is comparatively low. As a consequence, the holes undergo partial carrier diffusion and exhibit a relatively sluggish drifting motion towards the anode. This slower migration of holes results in a delayed peak in the Anode voltage waveform, which is distinguishable from the primary peak primarily caused by the swift and quasi-total collection of electrons.

It is important to note that the delayed peak stemming from the holes' delayed collection is attenuated compared to the primary peak attributed to electrons. This attenuation is a direct consequence of the unique characteristics of carrier diffusion and drifting behavior exhibited by the holes as they traverse the intrinsic π region.

1.1.2 Interpretations and conclusions

To begin with, the simulations were conducted by the STMicroelectronics TCAD team [63], which specializes in developing and utilizing software tools for advanced semiconductor device simulation. Those simulations, in addition to Monte Carlo simulations conducted by the same team [64], were aimed at studying the behavior of carriers after they exit the multiplication region of the device. This region is where the avalanche breakdown process occurs, leading to the generation of a large number of electron-hole pairs.

To analyze the behavior of carriers, the TCAD simulations utilized a Fokker Planck-based approach. This is a mathematical model that describes the diffusion of particles in a medium. In this case, it was used to model the movement of electrons and holes in the device.

Figure IV-2 provides a visual representation of the simulation results. The figure shows two different electron impact points inside the device. During the avalanche breakdown process, the holes generated during the avalanche breakdown take those same paths when drifting away from the multiplication region (highlighted in red). When the maximum number of carriers have been generated, the strongest electric field degradation (EFD) contribution is achieved. At this point, electrons are still being accumulated and captured at the cathode, while holes are slowly starting to drift and diffuse inside the single-photon avalanche diode (SPAD) and towards the

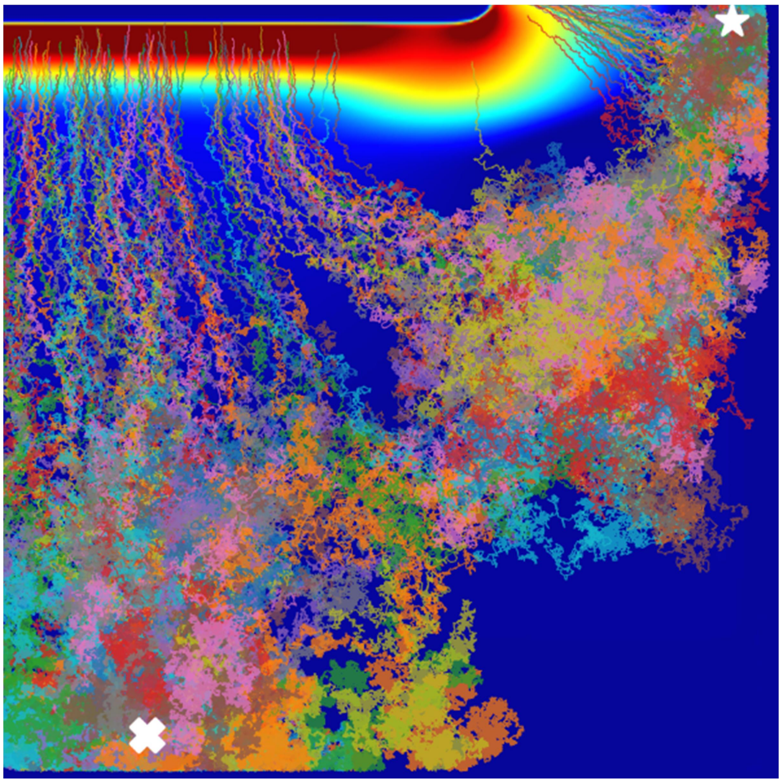

Figure IV-2: Electron trajectories inside the device departing from two photon impact points.

anode.

In a second stage, nearly all electrons have been collected by the cathode, and the holes are diffused in the entire lower intrinsic region of the SPAD. They then start partially drifting towards the anode, where they are slowly collected. However, only a fraction of the drifted holes is collected at the anode, resulting in a weak secondary current.

The simulations have provided valuable information on the behavior of carriers in the multiplication and absorption regions of the device. Specifically, they have helped to understand the travel paths and collection of carriers, which are critical factors in determining the performance of the device. The simulations have also highlighted the importance of optimizing the collection efficiency of holes at the anode to improve the overall performance of the device.

Overall, the collaboration between our team and the TCAD team, along with the use of advanced simulation tools, has enabled a more exhaustive understanding of the behavior of carriers in the different regions of the SPAD. The findings obtained will be instrumental in improving the performance of the device and advancing the field of semiconductor device technology in terms of device modelling.

1.2 Statistical jitter modelling

As a baseline multiple loop model has been established and validated by experimental confrontation, the statistical phenomenon of jitter is a prominent perspective for future implementation.

Jitter is manifested by a statistical distribution of the SPAD's response time. This distribution originates from the device's internal structure and spatial distribution. As the SPAD's space charge region is divided into two distinct parts, absorption, and multiplication region, this leads to a particular behavior of the device. Incoming photons can be absorbed over the entire SCR and two cases emerge:

- The photon is absorbed directly inside the multiplication region, thus generating the resulting electron hole pair in a critically high electric field environment. As a result, a rapid response of the SPAD through avalanche breakdown is set off. The multiplication region being narrow, the jitter of the response is very low as the electron hole pair will rapidly multiply through impact ionization no matter where the photon is absorbed.

- The photon is absorbed inside the much larger, low electric field, absorption region. The field inside this region will, once the electron hole pair is photoelectrically generated, slowly drift the electron towards the multiplication region. Consequently, depending on the area within the absorption region where the photon is absorbed, the resulting response of the device is more delayed, the farther the absorption is away from the multiplication region.

Figure IV-3 shows a representation of the current responses of the device when triggered by photon absorption in different areas of its SCR. As it is visible, the closer the electron and hole generation happens to the multiplication region, the faster the response time of the SPAD's current.

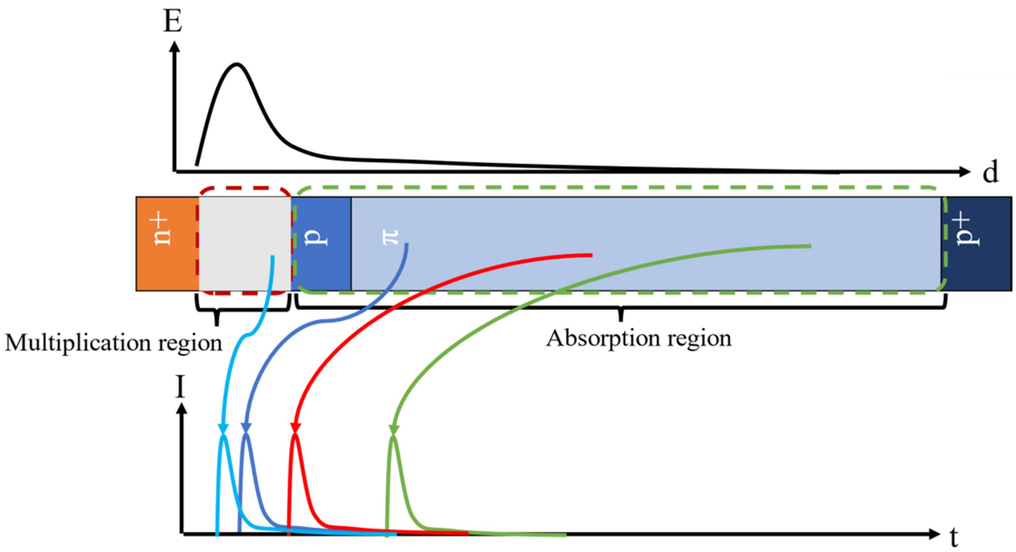

Figure IV-3 Visual representation of the SPAD's SCR, its different regions and corresponding electric field. Four different photon impact points and the resulting current response in time are depicted below.

The integration of this phenomenon is intrinsic to the Multiple Loop Model structure of the previous part. As an extension of the Single Loop Model, the ability to chain multiple loops together would permit jitter modeling in a physically accurate way.

As the multiplication region is very narrow in the reach-through type SPAD that was studied, only a single loop would be necessary to describe its contribution to the device's behavior.

The absorption region on the other hand, being considerably larger, would be segmented into several different sections, each representing a potential photon impact area. If high jitter resolution and precise statistical modelling is needed, the number of loops can be increased, whereas for less timing sensitive applications, 3 to 4 loops in the absorption region would give the designer sufficient information on the boundaries of the jitter.

1.3 Statistical amplitude modelling

Similarly to the statistical distribution of jitter, the amplitude of the SPAD output pulses is highly statistical and can thus vary from output to output, as shown in the measurements seen on ***Figure IV-4***.

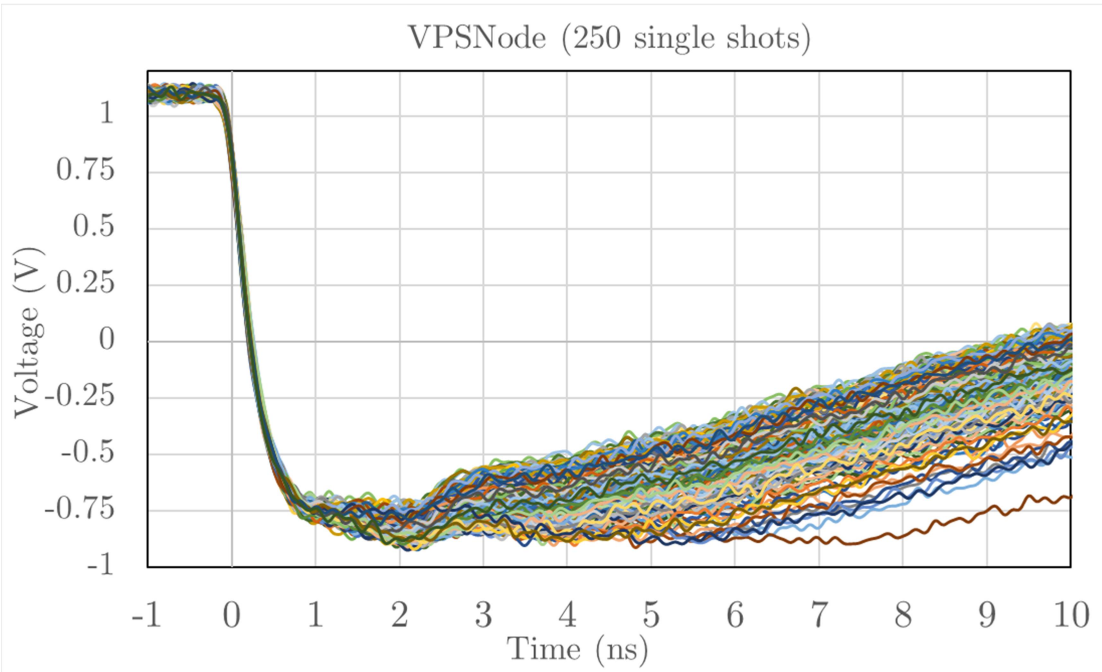

Figure IV-4 VPSNode graphs, superposition of 250 single shots at biasing of 25V

The variation of the peak voltage of the VPSNode response, when the same biasing conditions are applied, is of multiple nature.

Figure IV-4 shows superimposed graphs of 250 single shots at 25V biasing voltage VHV. Two phenomena are clearly shown:

- The peak voltage drop has a fluctuation of 0.25V which represents approximately 12-15% of statistical variability.
- The recharge tail of the device's response presents a high fluctuation at both the plateau's duration with a variation from 2 up to 6ns, and the voltage amplitude which has a variation of 0.5V, corresponding to a consequent 25-30% variation.

Thanks to insights provided by TCAD [63], those two observations can be explained by corresponding observations made on the avalanche breakdown process.

First, the fluctuations in the avalanche amplitude can be caused by the localized difference in impact ionization coefficients, depending on the area of the high electric field multiplication regions, in which the avalanche breakdown occurs. Those small coefficient variations can result in a subsequent variation in the avalanche breakdown intensity, thus in- or decreasing the device's output response voltage drop.

Then, if the second observation of variable plateau durations and recharge voltages is considered, the cause is linked to the horizontal distribution of the avalanche phenomena. The avalanche breakdown typically extends itself over a narrow region localized around the initially generated electron-hole pair (or drifted electron).

Occasionally, a secondary avalanche can occur, not far besides the primary avalanche as shown in ***Figure IV-5***, therefore prolonging the avalanche breakdown time above its primary duration. This can lead to a plateau in the voltage drop on the SPAD output as previously seen in ***Figure IV-4***. The secondary avalanche is thus sustaining the impact ionization at a steady rate, delaying the quenching of the device.

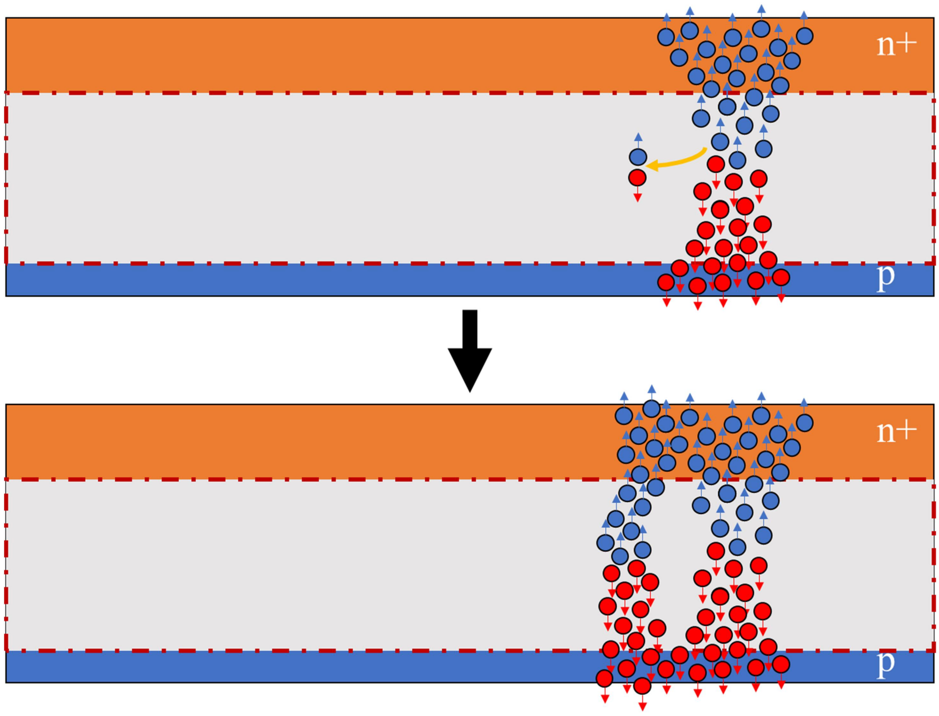

Figure IV-5 Visual representation of secondary avalanche breakdown occurence

2 General Conclusions

The main goal of this thesis was to develop a new approach in modelling Single Photon Avalanche Diode in order to provide circuit designers with an accurate and close to physics model. This approach has been decomposed into two main design steps:

A Single-Loop approach was first developed, translating the core physical phenomenon of impact ionization into a functional diagram of the mechanism. It was made possible by analyzing impact ionization's exponential increase of the number of implicated carriers, taking inspiration from control theory and its description of closed loop systems. This assimilation was especially close when considering divergent closed loop systems that are able to sustain the increase of the output signal. Once this convergence of ideas was translated into a graphical looped representation, the first bone of the model's skeleton was established. Following that, the implementation of the code came with several problems, mainly concerning the impact ionization coefficients that had to be weighted and adjusted to the distance that is considered. In the case of the Single-Loop approach, this resulted in the weighing of the coefficient by the narrow multiplication region width.

The framework of this new modelling approach has then provided proof of concept by the comparison of its simulated response to measurements on the physical device. This proof was further confirmed by the publication and presentation of a conference paper at ESSDERC/ESSCIRC 2022 [62]. The confrontation of the simulation results from this first step of the modelling approach to the real device's response was carried out on several key figures of merit like Charge Per Pulse or Pulse Width. The accuracy of the results was a milestone that lead to the extension of the single loop based modelling approach to a distributed modelling approach, using multiple elementary loops.

Subsequently, as mentioned in the previous paragraph, a distributed model was developed upon the results and structure of the Single-Loop model. The key idea behind this distributed modelling approach was the spatialization of the model, based on the device's structural architecture. By studying the structure of the modelled SPAD, using the schematic sections of the device and also taking into account physical simulations, we were able to deduce that it is a reach-through SPAD architecture. This structure has two distinct sections; multiplication and absorption region, which behave very differently. On the one hand is the absorption region which is a wide region that is mainly destined to maximize photon absorption and thus, improve the device's detection capabilities. The multiplication region on the other hand, is a narrow section of the device where the very high electric field ($> 10^5$ $V.cm^{-1}$) enables impact ionization, the mechanism responsible for the SPAD's special behavior.

The resulting distributed model, starting at a single loop for the multiplication region to describe the area where the impact ionization phenomenon is occurring, and a single loop for the absorption region to first model the mean response of the device. This takes into account both the transit time of the electron to the high electric field multiplication region and also the transit time of the hole flow from this same region towards the anode. Several key improvements were also made to the model's structure in order to correct undesired previous behavior and improve its overall stability and robustness. A charge based quenching criteria helped resolve internal oscillation issues occurring in some cases, enhancing reliability of the model. Another addition was the integration of the non-quenching behavior that can occur when the circuit's quenching resistor has too low of a value. Finally, conjointly with breakdown voltage tuning, the addition of the electric field degradation phenomenon lead to accurate simulation results when compared to the transient measurements of the SPAD. A publication in the IEEE journal "Transactions on Electron Devices" [65] further corroborates the model's relevance and potential.

When combined, those two stages of constructing the presented model and their corresponding experimental validation enable us to confirm the potential of this modelling approach. Some very interesting perspectives can be raised to improve upon the existing model. The first of those is linked to the delayed hole collection at the anode. Measured current transients of the device showed that there is a secondary delayed current peak in the mean behavior, which can be assimilated to the hole current traversing the low electric field absorption region to be collected by the anode terminal. Then, there is statistical jitter modelling and also statistical amplitude modelling that can be physically and spatially translated and represented by the model, due to its spatial modelling architecture.

Provided the initial stakes of the thesis, the scope and focus have evolved over time. As the PhD went on, the project evolved from a broader SPAD model to a deeper focus on the spatially distributed modelling method. Due to the divergent and internally interconnected nature of this method, the need for a solid and reliable model basis was crucial. When represented graphically, the model seems straightforward and simple in appearance, but all internal nodes and equations are tightly linked to one another and thus, the debugging of the core was challenging. Due to the structure of the developed model, Afterpulsing and DCR can be easily implemented as separate configurable building blocks. The output of those modules would be connected to a module that determines the location of the photon absorption as a function of wavelength and the device's material properties. All those modules can be designed individually and independently from each other before connecting them to the main model. Once assembled, as in ***Figure IV-6***, the complete model would be able to entire simulate the behavior of the SPAD, from DCR, to Afterpulsing and all other statistical phenomena (timing jitter, amplitude).

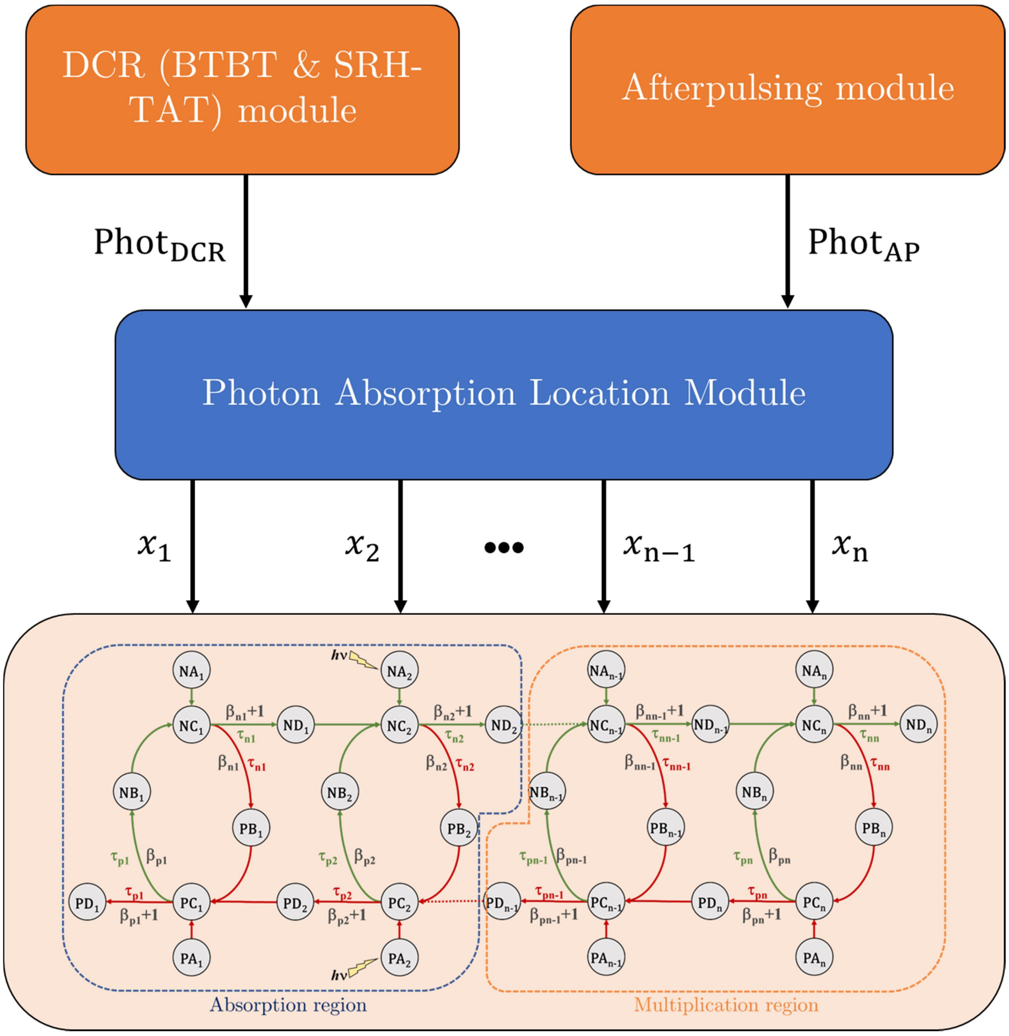

Figure IV-6 Final model description with perspective modules

The innovative spatially distributed approach developed in the scope of this PhD is aimed to diversify the portfolio of available modelling approaches and offer a new way of describing SPAD devices. As previously described, the results of this study and development are promising in terms of results when compared to characterization data. Combined with the potential of all the perspective implementations and modules that can be added to the existing model structure, this approach can have a great potential in high precision and high-performance detection systems. In such systems, fidelity to the real device can be crucial, and as the model can be modulated by adding more loops resulting in a more precise description of jitter, photon arrival location dependency etc., further providing potential utility to the model.

References

[1] A. D. Mora, A. Tosi, S. Tisa, and F. Zappa, "Single-Photon Avalanche Diode Model for Circuit Simulations," *IEEE Photonics Technol. Lett.*, vol. 19, no. 23, pp. 1922–1924, 2007, doi: 10.1109/LPT.2007.908768.

[2] I. Malass, W. Uhring, J.-P. Le Normand, N. Dumas, and F. Dadouche, "Evaluation of size influence on performance figures of a Single Photon Avalanche Diode fabricated in a 180 nm standard CMOS technology," *Analog Integr. Circuits Signal Process.*, pp. 1–8, mai 2016, doi: 10.1007/s10470-016-0763-8.

[3] L. Pancheri, D. Stoppa, and G. D. Betta, "Characterization and Modeling of Breakdown Probability in Sub-Micrometer CMOS SPADs," *IEEE J. Sel. Top. Quantum Electron.*, vol. 20, no. 6, pp. 328–335, 2014, doi: 10.1109/JSTQE.2014.2327791.

[4] F. Sun, Y. Xu, Z. Wu, and J. Zhang, "A Simple Analytic Modeling Method for SPAD Timing Jitter Prediction," *IEEE J. Electron Devices Soc.*, vol. 7, pp. 261–267, 2019, doi: 10.1109/JEDS.2019.2895151.

[5] R. Kazma, O. Rossetto, and G. Sicard, "High level model of SPAD based pixel," in *2015 11th Conference on Ph.D. Research in Microelectronics and Electronics (PRIME)*, 2015, pp. 125–128. doi: 10.1109/PRIME.2015.7251350.

[6] F. Villa, Y. Zou, A. D. Mora, A. Tosi, and F. Zappa, "SPICE Electrical Models and Simulations of Silicon Photomultipliers," *IEEE Trans. Nucl. Sci.*, vol. 62, no. 5, pp. 1950–1960, 2015, doi: 10.1109/TNS.2015.2477716.

[7] A. Dieguez, O. Alonso, E. Vilella, and A. Vilà, "A Verilog-A model for the design of devices for fluorescence life-time measurement with CMOS SPADs," in *2015 International Conference on Synthesis, Modeling, Analysis and Simulation Methods and Applications to Circuit Design (SMACD)*, 2015, pp. 1–4. doi: 10.1109/SMACD.2015.7301680.

[8] L. Zheng, J. Tian, Z. Weng, H. Hu, J. Wu, and W. Sun, "An Improved Convergent Model for Single-Photon Avalanche Diodes," *IEEE Photonics*

Technol. Lett., vol. 29, no. 10, pp. 798–801, 2017, doi: 10.1109/LPT.2017.2685680.

[9] G. Giustolisi, R. Mita, and G. Palumbo, ''Behavioral modeling of statistical phenomena of single-photon avalanche diodes,'' *Int. J. Circuit Theory Appl.*, vol. 40, 2012, doi: 10.1002/cta.748.

[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, 2016, doi: 10.1109/TED.2016.2537879.

[11] T. Zhao, F. Sun, Y. Xu, B. Li, and D. Li, ''An improved behavioral simulation model for afterpulsing phenomenon,'' in *2018 China Semiconductor Technology International Conference (CSTIC)*, 2018, pp. 1–3. doi: 10.1109/CSTIC.2018.8369313.

[12] J. Conradi, ''The distribution of gains in uniformly multiplying avalanche photodiodes: Experimental,'' *IEEE Trans. Electron Devices*, vol. 19, no. 6, pp. 713–718, 1972, doi: 10.1109/T-ED.1972.17486.

[13] J. Léonard *et al.*, ''High-throughput time-correlated single photon counting,'' *Lab Chip*, vol. 14, no. 22, pp. 4338–4343, 2014, doi: 10.1039/C4LC00780H.

[14] B. Chauvenet, C. Bobin, and J. Bouchard, ''A general dead-time correction method based on live-time stamping. Application to the measurement of short-lived radionuclides,'' *Appl. Radiat. Isot.*, vol. 130, pp. 238–244, 2017, doi: https://doi.org/10.1016/j.apradiso.2017.10.013.

[15] F. L. R. Vinagre and C. Conde, ''Method for effective dead time measurement in counting systems,'' *Nucl. Instrum. Methods Phys. Res. Sect. Accel. Spectrometers Detect. Assoc. Equip.*, vol. 462, pp. 555–560, Apr. 2001, doi: 10.1016/S0168-9002(01)00179-6.

[16] M. Dolatpoor Lakeh *et al.*, ''An Ultrafast Active Quenching Active Reset Circuit with 50% SPAD Afterpulsing Reduction in a 28 nm FD-SOI CMOS Technology Using Body Biasing Technique,'' *Sensors*, vol. 21, no. 12, 2021, doi: 10.3390/s21124014.

[17] M. Moreno-García, R. del Río, Ó. Guerra, and Á. Rodríguez-Vázquez, "5×5 SPAD matrices for the study of the trade-offs between fill factor, dark count rate and crosstalk in the design of CMOS image sensors," in *2014 10th Conference on Ph.D. Research in Microelectronics and Electronics (PRIME)*, 2014, pp. 1–4. doi: 10.1109/PRIME.2014.6872752.

[18] L. V. Keldysh, "Concerning the Theory of Impact Ionization in Semiconductors," *Sov. J. Exp. Theor. Phys.*, vol. 21, p. 1135, décembre 1965.

[19] C. L. Anderson and C. R. Crowell, "Threshold Energies for Electron-Hole Pair Production by Impact Ionization in Semiconductors," *Phys Rev B*, vol. 5, no. 6, pp. 2267–2272, Mar. 1972, doi: 10.1103/PhysRevB.5.2267.

[20] M. Levinshtein, J. Kostamovaara, and S. Vainshtein, *Breakdown phenomena in semiconductors*, vol. 36. World Scientific Publishing Company, 2005.

[21] N. W. Ashcroft and N. D. Mermin, *Solid State Physics*. Harcourt Inc., 1976.

[22] S. Reggiani, M. Valdinoci, L. Colalongo, M. Rudan, and G. Baccarani, "An Analytical, Temperature-dependent Model for Majority- and Minority-carrier Mobility in Silicon Devices," *VLSI Des.*, vol. 10, pp. 467–483, 2000.

[23] G. Giustolisi, R. Mita, and G. Palumbo, "Verilog-A modeling of SPAD statistical phenomena," in *2011 IEEE International Symposium of Circuits and Systems (ISCAS)*, 2011, pp. 773–776. doi: 10.1109/ISCAS.2011.5937680.

[24] Y. Okuto and C. R. Crowell, "Ionization coefficients in semiconductors: A nonlocalized property," *\prb*, vol. 10, no. 10, pp. 4284–4296, Nov. 1974, doi: 10.1103/PhysRevB.10.4284.

[25] D. J. Massey, J. P. R. David, and G. J. Rees, "Temperature Dependence of Impact Ionization in Submicrometer Silicon Devices," *IEEE Trans. Electron Devices*, vol. 53, no. 9, pp. 2328–2334, 2006, doi: 10.1109/TED.2006.881010.

[26] W. N. Grant, "Electron and hole ionization rates in epitaxial silicon at high electric fields," *Solid-State Electron.*, vol. 16, no. 10, pp. 1189–1203, 1973, doi: https://doi.org/10.1016/0038-1101(73)90147-0.

[27] R. J. McIntyre, "On the avalanche initiation probability of avalanche diodes above the breakdown voltage," *IEEE Trans. Electron Devices*, vol. 20, no. 7, pp. 637–641, 1973, doi: 10.1109/T-ED.1973.17715.

[28] A. Spinelli, A. Pacelli, and A. L. Lacaita, "Dead space approximation for impact ionization in silicon," *Appl. Phys. Lett.*, vol. 69, no. 24, Art. no. 24, décembre 1996, doi: 10.1063/1.117196.

[29] A. Panglosse, "Modélisation pour la simulation et la prédiction des performances des photodiodes à avalanche en mode Geiger pour Lidars spatiaux," Université de Toulouse, Institut Supérieur de l'Aéronautique et de l'Espace, 2019.

[30] J. Marsland, "Comparison of different models of non-local impact ionization for low noise avalanche photodiodes," *Phys. Status Solidi C*, vol. 8, pp. 2577–2581, 2011, doi: 10.1002/pssc.201084034.

[31] O.-H. Kwon, M. Hayat, J. Campbell, B. Saleh, and M. Teich, "Effect of stochastic dead space on noise in avalanche photodiodes," *Electron Devices IEEE Trans. On*, vol. 51, pp. 693–700, 2004, doi: 10.1109/TED.2004.825798.

[32] C. Tan, J. P. R. David, G. Rees, R. Tozer, and D. Herbert, "Treatment of soft threshold in impact ionization," *J. Appl. Phys.*, vol. 90, pp. 2538–2543, 2001, doi: 10.1063/1.1388865.

[33] J. Marsland, "Nonlocal impact ionisation coefficients derived from Monte Carlo calculations," *Electron. Lett.*, vol. 38, pp. 55–57, 2002, doi: 10.1049/el:20020025.

[34] S. Plimmer, J. P. R. David, B. Jacob, and G. Rees, "Impact ionization probabilities as functions of two-dimensional space and time," *J. Appl. Phys. - J APPL PHYS*, vol. 89, pp. 2742–2751, 2001, doi: 10.1063/1.1342190.

[35] Q. Hernández, D. Gutiérrez, and A. Jarabo, "A Computational Model of a Single-Photon Avalanche Diode Sensor for Transient Imaging," 2017.

[36] W. G. Oldham, R. R. Samuelson, and P. Antognetti, "Triggering phenomena in avalanche diodes," *IEEE Trans. Electron Devices*, vol. 19, no. 9, pp. 1056–1060, 1972, doi: 10.1109/T-ED.1972.17544.

[37] M. M. Hayat *et al.*, ''Breakdown probabilities for thin heterostructure avalanche photodiodes,'' *IEEE J. Quantum Electron.*, vol. 39, no. 1, pp. 179–185, 2003, doi: 10.1109/JQE.2002.806217.

[38] M. Estribeau, ''Analyse et modélisation de la fonction de transfert de modulation des capteurs d'images à pixels actifs CMOS,'' Ecole nationale supérieure de l'aéronautique et de l'espace, 2004.

[39] A. C. Therrien, B. Bérubé, S. A. Charlebois, R. Lecomte, R. Fontaine, and J. Pratte, ''Modeling of Single Photon Avalanche Diode Array Detectors for PET Applications,'' *IEEE Trans. Nucl. Sci.*, vol. 61, no. 1, pp. 14–22, 2014, doi: 10.1109/TNS.2013.2293426.

[40] A. Sugimura, ''Auger recombination effect on threshold current of InGaAsP quantum well lasers,'' *IEEE J. Quantum Electron.*, vol. 19, no. 6, pp. 932–941, 1983, doi: 10.1109/JQE.1983.1071956.

[41] S. K. Saha, A. M. Farhan, S. I. Reba, S. I. Ferdaus, and M. I. B. Chowdhury, ''An analytical model of dark saturation current of silicon solar cell considering both SRH and Auger recombination,'' in *2011 IEEE Regional Symposium on Micro and Nano Electronics*, 2011, pp. 9–13. doi: 10.1109/RSM.2011.6088280.

[42] D. Mathiot, ''Cours de L3-ESA, Matériaux et composants,'' 2017.

[43] M. Mazzillo *et al.*, ''Quantum Detection Efficiency in Geiger Mode Avalanche Photodiodes,'' *IEEE Trans. Nucl. Sci.*, vol. 55, no. 6, pp. 3620–3625, 2008, doi: 10.1109/TNS.2008.2007483.

[44] J. López-Martínez, I. Vornicu, R. Carmona-Galan, and A. Rodriguez-Vazquez, ''An Experimentally-Validated Verilog-A SPAD Model Extracted from TCAD Simulation,'' 2018, pp. 137–140. doi: 10.1109/ICECS.2018.8617962.

[45] M. Perenzoni, ''Single-Photon Avalanche Diode-Based Detection and Imaging: Bringing the Photodiode Out of Its Comfort Zone,'' *IEEE Solid-State Circuits Mag.*, vol. 10, no. 3, pp. 26–34, 2018, doi: 10.1109/MSSC.2018.2844602.

[46] T. C. de Albuquerque *et al.*, ''Integration of SPAD in 28nm FDSOI CMOS technology,'' in *2018 48th European Solid-State Device Research Conference (ESSDERC)*, 2018, pp. 82–85. doi: 10.1109/ESSDERC.2018.8486852.

[47] A. Panglosse, P. Martin-Gonthier, O. Marcelot, C. Virmontois, O. Saint-Pé, and P. Magnan, "Dark Count Rate Modeling in Single-Photon Avalanche Diodes," *IEEE Trans. Circuits Syst. Regul. Pap.*, vol. 67, no. 5, pp. 1507–1515, 2020, doi: 10.1109/TCSI.2020.2971108.

[48] H. S. Benett and C. L. Wilson, "Statistical comparisons of data on band-gapnarrowing in heavily doped silicon : Electrical and optical measurements," *J. Appl. Phys.*, vol. 55.10, pp. 3582–3587, 1984.

[49] G. A. M. Hurkx, H. C. de Graaff, W. J. Kloosterman, and M. P. G. Knuvers, "A new analytical diode model including tunneling and avalanche breakdown," *IEEE Trans. Electron Devices*, vol. 39, no. 9, pp. 2090–2098, 1992, doi: 10.1109/16.155882.

[50] A. Mitonneau, A. Mircea, G. M. Martin, and D. Pons, "Electron and hole capture cross-sections at deep centers in gallium arsenide," *Rev. Phys. Appl.*, vol. 14, no. 10, pp. 853–861, 1979, doi: 10.1051/rphysap:019790014010085300.

[51] G. A. M. Hurkx, D. B. M. Klaassen, 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, 1992, doi: 10.1109/16.121690.

[52] M. Ghioni, A. Gulinatti, I. Rech, F. Zappa, and S. Cova, "Progress in Silicon Single-Photon Avalanche Diodes," *IEEE J. Sel. Top. Quantum Electron.*, vol. 13, no. 4, pp. 852–862, Jul. 2007, doi: 10.1109/JSTQE.2007.902088.

[53] H. Qiuyang, X. Yue, and Z. Fei-fei, "An accurate simulation model for single-photon avalanche diodes including important statistical effects," *J. Semicond.*, vol. 34, p. 104007, 2013.

[54] D. Shin, B. Park, Y. Chae, and I. Yun, "The Effect of a Deep Virtual Guard Ring on the Device Characteristics of Silicon Single Photon Avalanche Diodes," *IEEE Trans. Electron Devices*, vol. 66, no. 7, pp. 2986–2991, 2019, doi: 10.1109/TED.2019.2913714.

[55] A. Gulinatti, F. Ceccarelli, M. Ghioni, and I. Rech, *Custom silicon technology for SPAD-arrays with red-enhanced sensitivity and low timing jitter*. 2020.

[56] S. Cova, M. Ghioni, A. Lacaita, C. Samori, and F. Zappa, "Avalanche photodiodes and quenching circuits for single-photon detection," *Appl Opt*, vol. 35, no. 12, pp. 1956–1976, avril 1996, doi: 10.1364/AO.35.001956.

[57] S. Isaak, M. C. Pitter, S. Bull, and I. Harrison, "Design and characterisation of 16×1 parallel outputs SPAD array in 0.18 um CMOS technology," in *2010 IEEE Asia Pacific Conference on Circuits and Systems*, 2010, pp. 979–982. doi: 10.1109/APCCAS.2010.5774894.

[58] M. Stipčević, D. Wang, and R. Ursin, "Characterization of a Commercially Available Large Area, High Detection Efficiency Single-Photon Avalanche Diode," *J. Light. Technol.*, vol. 31, no. 23, pp. 3591–3596, 2013, doi: 10.1109/JLT.2013.2286422.

[59] F. Acerbi *et al.*, "Characterization of Single-Photon Time Resolution: From Single SPAD to Silicon Photomultiplier," *IEEE Trans. Nucl. Sci.*, vol. 61, no. 5, pp. 2678–2686, 2014, doi: 10.1109/TNS.2014.2347131.

[60] D. Rideau *et al.*, "Direct Measurements and Modeling of Avalanche Dynamics and Quenching in SPADs," in *Essderc 2023, Lisbon, Portugal*, Lisbonne, Portugal, 2023. doi: 10.1109/ESSDERC59256.2023.10268474.

[61] Y. oussaiti *et al.*, "Verilog-A model for avalanche dynamics and quenching in Single-Photon Avalanche Diodes," in *2020 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, 2020, pp. 145–148. doi: 10.23919/SISPAD49475.2020.9241648.

[62] S. Rink *et al.*, "A self-sustaining Single Photon Avalanche Diode Model," in *ESSDERC 2022 - IEEE 52nd European Solid-State Device Research Conference (ESSDERC)*, 2022, pp. 281–284. doi: 10.1109/ESSDERC55479.2022.9947120.

[63] R. Helleboid *et al.*, "A Fokker–Planck-based Monte Carlo method for electronic transport and avalanche simulation in single-photon avalanche diodes," *J. Phys. Appl. Phys.*, vol. 55, no. 50, p. 505102, Oct. 2022, doi: 10.1088/1361-6463/ac9b6a.

[64] D. Rideau *et al.*, "Multiscale SPAD modeling: from Monte Carlo to SPICE simulations," in *SPIE Defense + Commercial Sensing*, SPIE, May 2023, pp. 12512–12. doi: 10.1117/12.2665099.

[65] S. Rink *et al.*, "A Spatially Distributed Single Photon Avalanche Diode Verilog-A Compact Model," *IEEE Trans. Electron Devices*, 2023, doi: 10.1109/TED.2023.3311423.

List of publications

Papers in international peer revied journals

[1] **S. Rink** *et al.*, "A Spatially Distributed Single Photon Avalanche Diode Verilog-A Compact Model," *IEEE Trans. Electron Devices*, 2023, doi: 10.1109/TED.2023.3311423.

[2] M. Dolatpoor Lakeh *et al.*, "An Ultrafast Active Quenching Active Reset Circuit with 50% SPAD Afterpulsing Reduction in a 28 nm FD-SOI CMOS Technology Using Body Biasing Technique," *Sensors*, vol. 21, no. 12, 2021, doi: 10.3390/s21124014. **Co-author**

Papers in international conferences with proceedings

[3] **S. Rink** *et al.*, "A self-sustaining Single Photon Avalanche Diode Model," in *ESSDERC 2022 - IEEE 52nd European Solid-State Device Research Conference (ESSDERC)*, 2022, pp. 281–284. doi: 10.1109/ESSDERC55479.2022.9947120

[4] D. Rideau *et al.*, "Direct Measurements and Modeling of Avalanche Dynamics and Quenching in SPADs," in *Essderc 2023, Lisbon, Portugal*, Lisbonne, Portugal, 2023. doi: 10.1109/ESSDERC59256.2023.10268474. **Co-Author**

[5] D. Rideau *et al.*, "Multiscale SPAD modeling: from Monte Carlo to SPICE simulations," in *SPIE Defense + Commercial Sensing*, SPIE, May 2023, pp. 12512–12. doi: 10.1117/12.2665099 **Co-Author**

Sven RINK

Modelization of Single-Photon Avalanche Diode (SPAD) for the Design of Integrated Imager Pixels in advanced CMOS technology

Résumé

RESUME GLOBAL

Les applications d'imageurs à base de matrices de pixel utilisant des diodes à avalanche à photon unique de type SPAD intégrées en technologie CMOS se sont considérablement développés durant la dernière décennie avec des applications en imagerie et détection dans les domaines du biomédical, de l'environnement, de l'industrie et de l'automobile. La diode SPAD est une photodiode utilisée dans un domaine de fonctionnement atypique pour la plupart des dispositifs semiconducteurs et l'écriture d'un modèle permettant de la simuler dans un environnement de conceptionmicroélectronique est nécessaire pour répondre aux besoins des concepteurs d'imageurs CMOS à base de matrices de SPAD. Un modèle distribué a été conçu afin de répondre à ces exigeances, basé sur une structure en boucles qui traduit le phénomène d'ionisation par impact, phénomène physique au cœur du fonctionnement du dispositif. La validation expérimentale de ce modèle est faite par le biais d'une structure de test solide, développée au cours d'un projet parallèle entre STMicroelectronics et le laboratoire ICube. C'est grâce à cette structure de test et les mesures transitoires résultantes dont nous avons bénéficié, que le modèle a été calibré afin d'obtenir un comportement fidèle au dispositif réel.

CONTEXTE ET OBJECTIFS DE LA THESE

L'objectif principal de ce projet de modélisation est de répondre aux besoins des designers travaillant sur les circuits d'imageurs à base de diodes SPAD. La réponse impulsionnelle d'un pixel donné est déterminée selon une multitude de facteurs, incluant la réponse statique (DC), dynamique (AC), le comportement statistique temporel du dispositif ainsi que le temps de réponse des circuits de contrôle et de lecture. Tous ces effets doivent être modélisés de façon précise et implémentés dans un environnement de design de circuit pour atteindre les objectifs suivants :

- L'optimisation de l'implémentation des SPADs individuelles : ceci implique l'optimisation des figures de mérite des SPADs, incluant la probabilité de detection de photons (PDP), la résolution temporelle à photon unique (SPTR), le jitter, le temps mort et l'afterpulsing. De plus, l'optimisation de l'électronique du pixel telle que le facteur de remplissage (FF), l'efficacité de détection de photons (PDE), la capacité d'entrée, la bande passante et d'autres paramètre reste critique.

- L'évaluation de la performance d'une matrice de pixels. Une fois que les SPADs et pixels individuels ont été optimisés, il est nécessaire d'évaluer la performance de la matrice de pixels dans son intégralité. Ceci implique la détermination de la sensibilité de la matrice de pixel et l'identification de limitations pouvant empêcher ou entraver

le bon fonctionnement du dispositif tel que les phénomènes de couplage optique ou électrique (crosstalk).

En modélisant de façon précise le dispositif et en implémentant tous ces facteurs dans l'environnement de design de circuit, il est possible de développer des circuits d'imageurs hautement sensibles et précis basés sur des diodes SPAD. Ceci est critique dans une multitude d'applications, incluant notamment l'imagerie médicale, les capteurs à distance et la recherche scientifique.

L'approche de modélisation utilisée dans ce projet est un processus complexe et multifacettes incluant un panel large de techniques et de méthodes. La première étape dans ce processus va être de conduire une analyse poussée des phénomènes physiques qui font partie et dictent le fonctionnement des SPAD. Ceci nécessite une revue de la littérature et l'utilisation d'outils de simulation avancés pour modéliser le comportement de la SPAD sous différentes conditions d'utilisation.

Une fois que les effets physiques ont été analyses et modélisés, la prochaine étape sera le développement d'un modèle analytique pouvant être utilisé et intégré dans un environnement de design de circuit. Ceci impliquera l'utilisation de résultats de simulation afin de dériver un modèle décrivant de façon précise la SPAD sous différentes conditions. Ce modèle sera ensuite validé en utilisant des données expérimentales qui seront obtenues en utilisant un panel élargi de méthodes de caractérisation optoélectroniques ultra rapides, basées sur l'utilisation de lasers femtoseconde.

La validation expérimentale sera basée sur des techniques de photoluminescence résolues dans le domaine temporel, le comptage de photons corrélés en temps ainsi que d'autres méthodes avancées. Ces dernières ont été développées et implémentées au laboratoire ICube dans le cadre du projet Nano2022, projet qui est subventionné par le gouvernement français et visant à consolider l'économie nationale de l'industrie des composants électroniques.

Pour s'assurer que l'approche de modélisation est valide dans un environnement de design de circuits d'imagerie, deux niveaux de modélisation seront considérés. Le premier niveau se concentrera sur la modélisation d'une diode SPAD élémentaire qui sera la priorité de cette thèse. Le second niveau s'appuiera sur le premier et consistera en la modélisation d'une matrice de pixels, ce qui est une tâche plus complexe nécessitant l'intégration de multiples SPAD dans un unique système d'imageur.

Le langage comportemental Verilog-A sera utilisé pour la modélisation, étant donné qu'il est compatible avec les simulateurs communément utilisés dans le design de circuits analogiques (eldo, spectre…). Ceci permettra au modèle analytique d'être intégré dans une grande panoplie de design de circuits d'imageurs, incluant ceux utilisés dans l'imagerie médicale, la recherche scientifique et les applications industrielles.

En globalité, l'approche utilisée dans ce projet est un pas critique dans le développement de circuits d'imagerie hautement précis et fiables, basés sur les diodes SPAD. En consolidant l'utilisation de ces modèles de simulation dans l'industrie, ce projet aidera à promouvoir l'innovation et la croissance économique en France.

LA DIODE SPAD

La diode SPAD est une photodiode utilisée dans un domaine de fonctionnement atypique pour la plupart des dispositifs semiconducteurs : une polarisation inverse est appliquée au-delà de la tension de claquage de la jonction, et génère un état métastable en présence de porteurs libres générés soit par l'absorption d'un photon incident, soit par d'autres mécanismes (Génération-recombinaison, effet tunnel, ...).

Alors que l'absorption de chaque photon permet en quelques picosecondes (ps) la création d'une paire électron-trou, la polarisation à la limite du claquage impose un champ électrique favorable à la multiplication des paires électrons-trous via le phénomène d'ionisation par impact ; le courant est ainsi amplifié et devient observable à l'échelle macroscopique pour chaque photon absorbé, après quelques dizaines à centaines de ps. En fonction du champ électrique appliqué, le facteur d'amplification varie sur plusieurs décades, ce qui impose de limiter le courant par une électronique de contrôle, non seulement pour éviter la dégradation du SPAD, mais également pour éteindre l'avalanche dès la détection de la première impulsion lumineuse par l'électronique de lecture, et réinitialiser la polarisation du SPAD avant la prochaine impulsion lumineuse, le tout en quelques nanosecondes. Alors que le régime d'ionisation par impact permet d'augmenter la sensibilité de la diode SPAD en tant que détecteur de photon, l'électronique de contrôle permet de contrôler le courant transitoire d'avalanche (amplification, extinction, réinitialisation). La vitesse de déclenchement du SPAD associée au temps de réponse de l'électronique de contrôle déterminent la résolution temporelle de chaque pixel d'imageur.

Enfin, les processus physiques d'absorption de photon, d'ionisation par impact étant de nature stochastique, la réponse temporelle du SPAD à une absorption de photon non seulement a une composante déterministe qui dépend de paramètres structurels (tels que le choix du matériau et des dimensions de la zone d'absorption), mais est également exposée à une fluctuation (Jitter) qui doit être prise en compte dans la définition de la résolution temporelle minimum de chaque pixel.

STRUCTURE DU DISPOSITIF

La structure de la SPAD développée par STMicroelectronics, étudiée et modélisée dans le cadre de cette thèse est de type « reach-through » (avec une zone de très fort champ en surface). La structure est divisée en deux parties bien distinctes :

- Une zone d'absorption, dans laquelle un faible champ électrique entraîne les porteurs (carrier drift). Cette zone est généralement très large, permettant une absorption améliorée générant des porteurs qui vont potentiellement être entraînés dans la zone de multiplication pour y générer une avalanche.

- Une zone de multiplication, dans laquelle régit un fort champ électrique (E), entraînant ainsi l'ionisation par impact des porteurs et donc la génération du phénomène d'avalanche. Cette zone est généralement étroite par rapport à la zone d'absorption.

Afin de décrire au mieux le comportement de cette structure, nous avons choisi, grâce au modèle distribué en profondeur, de commencer à implémenter une solution de modèle à deux boucles. Une boucle qui modélise l'absorption ainsi que la propagation des porteurs par le champ électrique vers la zone de multiplication d'une part (pour les électrons, nous remarquons ici que cette structure N+/P/π/P+ génère une avalanche principalement par la multiplication de l'électron), et la collection retardée du courant de trous d'autre part.

APPROCHE DE MODELISATION

Le modèle développé dans le cadre de ce projet se veut au plus proche de la réalité physique en décrivant le phénomène d'ionisation par impact le plus justement possible, tout en permettant à l'utilisateur du modèle de choisir le degré de précision qu'il souhaite. Ce modèle développé est basé sur une structure bouclée.

Le déclenchement d'avalanche (Trigger), qu'il soit d'origine photonique, thermique ou par effet tunnel, arrive au niveau des nœuds NA et PA. Dépendant de la nature du trigger, on injecte une charge élémentaire coté N (NA) si un électron est généré et coté P (PA) si un trou est généré. Les nœuds NB et PB sont des nœuds de multiplication de porteurs et contribuent au gain de boucle du système Les nœuds NC et PC somment les contributions de nœuds NA, NB et PA, PB respectivement après un délai qui est le temps de transit des porteurs dans la zone de charge d'espace (ZCE).

En prenant en compte la nature cyclique et récurrente du phénomène d'ionisation par impact, le modèle a été structuré sous forme de boucle, permettant ainsi une représentation compréhensible du comportement de la diode à avalanche à photon unique (SPAD). Cette architecture de modèle sous forme de boucle comprend deux chemins de porteurs distincts au sein du dispositif : les électrons (en vert) et les trous (en rouge).

Il est important de noter que cette structure interne du modèle est uniquement connecté au modèle global par les deux nœuds de sortie qui servent d'extrémités de la zone de charge d'espace. Cette boucle élémentaire englobe divers nœuds essentiels qui ont chacun leur propre fonction et contribution au fonctionnement de la représentation de la SPAD :

-NA et PA : ces nœuds correspondent aux points d'injection initiaux des paires électron/trou et sont donc utilisés pour représenter l'effet photoélectrique de l'absorption de photon, résultant en la génération d'une paire électron/trou. Ces nœuds permettent également l'injection dans le modèle de charges induites par d'autres phénomènes physiques tels que la libération de porteurs piégés (Afterpulsing) ou encore les porteurs générés par effet thermique (Shockley-Reed-Hall). Cette versatilité permet une modélisation de la SPAD sous diverses conditions de fonctionnement.

-NC et PC : ces nœuds servent de nœuds de sommation qui sont responsables de regrouper et sommer les courants des nœuds NA/NB et PA/PB respectivement. A chaque itération dans la boucle, ces nœuds accumulent et combinent les courants entrants permettant ainsi l'analyse et la modélisation de la dynamique des flux de porteurs au sein du dispositif.

-NB et PB : ces noeuds sont les noeuds de multiplication et comme l'indique leur nom, ils servent à la multiplication des porteurs. Leur rôle est donc la multiplication des courants de l'itération précédente en utilisant les coefficients d'ionisation. Au travers de ce processus de multiplication, le modèle prend en compte l'effet d'ionisation par impact simulant la multiplication exponentielle des porteurs lors de leur traversée de la SPAD.

-ND et PD : ces nœuds servent de nœuds de sortie. ND et PD jouent un rôle crucial dans la fonctionnalité globale du modèle. Ces nœuds facilitent l'extraction des courants de l'itération courante et l'injectent dans les terminaux de sortie du modèle, permettant la représentation du comportement global de la SPAD.

En incorporant ces divers nœuds dans la structure en boucles, le modèle compact encapsule les aspects essentiels de la dynamique des porteurs, de leur multiplication ainsi que du flux de courant au sein du dispositif. La nature interconnectée de ces nœuds permet de simuler de façon fine l'interaction des charges, facilitant la compréhension plus profonde de la performance de la SPAD en fonction des différentes conditions d'opération. Pour conclure, le modèle compact est basé sur une architecture à base de boucles qui est fortement corrélée au phénomène sous-jacent de l'ionisation par impact.

MODELE DISTRIBUE

C'est sur la base de cette boucle élémentaire que la suite du modèle a été construite, en assemblant plusieurs boucles afin de générer un modèle distribué dans l'espace. Cela permet d'avoir des sous-sections au sein du modèle qui sont modulables selon la zone de la SPAD dans laquelle on se situe : zone de multiplication ou zone d'absorption :

- La zone d'absorption est une zone de faible champ électrique et de largeur conséquente. Son rôle est de capter un maximum de photons afin de générer par effet photoélectrique des paires électron/trou. Une fois généré, ces porteurs vont être entraînés par le champ électrique vers l'anode pour les trous, et vers la cathode et donc la zone de multiplication pour les électrons. La largeur de cette zone permet un rendement quantique, une proportion de photons captés générant une avalanche, accrue, améliorant le fonctionnement du dispositif de façon conséquente.

-La zone de multiplication est une zone de fort champ électrique avec des champs pouvant atteindre les 5-8x10^5 V.cm^-1. Ce très fort champ électrique au sein de cette zone très étroite (quelques centaines de nanomètres) est la principale cause de l'effet d'avalanche des diodes SPAD. L'entrainement conséquent des porteurs dans la zone de multiplication leur fait atteindre leur vitesse de saturation et donc une grande quantité d'énergie. Cette énergie va, lors d'impacts des porteurs avec les atomes de silicium environnants, permettre d'éjecter des électrons de valence de ces atomes, créant ainsi une nouvelle paire électron/trou, qui va à son tour pouvoir ioniser.

RESULTATS

Après une première validation de la preuve de concept du modèle en boucles, notamment par une publication en 2022 d'un article de conférence IEEE (ESSDERC), le modèle distribué a été calibré et paramétré sur les mesures effectuées sur la structure de test « KNACK » développée avec STMicroelectronics. Grâce à ce circuit, notamment par les amplificateurs RF intégrés, les mesures possibles sur la SPAD ont été largement étendues.

A l'aide des mesures de précision faites avec ce circuit, nous avons pu établir les caractéristiques transitoires de la SPAD. Ceci nous a permis d'ajuster notre modèle afin qu'il reproduise de façon précise les réponses transitoires du dispositif étudié. En faisant la comparaison entre les courbes de mesures et les courbes simulées du modèle ajusté sur la plage de fonctionnement de la diode SPAD, nous avons constaté que les résultats sont très proches des courbes mesurées, ce qui permet de dire que le modèle est fidèle au fonctionnement réel de la diode.

CONCLUSIONS

L'objectif principal de cette these était de developer une nouvelle approche dans la modélisation de diode à avalanche à photon unique afin de fournir au designers de circuits un modèle précis et proche de la physique. Cette approche a été décomposée en deux principales étapes de modélisation :

-Une approche à simple boucle a été développée dans un premier temps, traduisant le phénomène physique au cœur des SPAD, l'ionisation par impact, en un diagramme fonctionnel du mécanisme. Ceci a été rendu possible en analysant l'augmentation exponentielle du nombre de porteurs impliqués dans l'ionisation par impact tout en s'inspirant des concepts d'automatique et sa description de systèmes à boucle fermée. La proximité avec un système à boucle fermée est d'autant plus proche si l'on considère les systèmes à boucle fermée divergents, étant capables d'entretenir l'amplification exponentielle d'un signal.

Une fois que la convergence de ces différentes idées a été traduite en une représentation graphique, le premier os du squelette du modèle a été établi. A la suite de cela, l'implémentation du code a subi quelques problèmes, notamment concernant les coefficients d'ionisation par impact qui ont eu besoin de subir une calibration afin d'ajuster leur poids dans la multiplication des porteurs. Dans le cadre de l'approche à simple boucle, ceci s'est traduit par la pondération des coefficients avec l'étroite largeur de la zone de multiplication. Le cadre de cette nouvelle approche a servi de preuve de concept en comparant la réponse simulée du modèle avec les mesures sur le dispositif physique. Cette preuve de concept a notamment été validé au travers de la publication et présentation d'un article de revue IEEE à l'ESSDERC/ESSCIRC 2022. La confrontation des résultats de simulation de ce premier pas dans l'approche de modélisation avec les réponses réelles de la SPAD s'est faite sur quelques figures de mérite clés tels que le Charge Per Pulse (CPP) ou la largeur de pulse (Pulse Width). La précision des résultats obtenus a servi de premier jalon qui a permis l'extension de l'approche à simple boucle à une approche de modélisation distribuée, utilisant plusieurs boucles élémentaires interconnectées.

-Suite à cela, comme mentionné dans le precedent paragraphe, un modèle distribué a été développé en se basant sur les résultats et la structure du modèle à simple boucle. L'idée principale de cette approche distribuée étant de créer une spatialisation du dispositif en se basant sur l'architecture physique du dispositif étudié. En étudiant la structure de la SPAD, en utilisant des coupes schématiques du dispositif ainsi que de simulations physiques, nous avons pu conclure que la SPAD étudiée avait une architecture dite « reach-through ». Cette structure dispose de deux sections distinctes : une région de multiplication et une région d'absorption, qui ont tous deux des comportements et fonctions bien distincts. D'une part, la zone d'absorption et une région large et principalement destinée à maximiser l'absorption de photons et de ce fait, améliorer la capacité de détection du dispositif. D'autre part, la zone de multiplication est une zone bien plus étroite (quelques centaines de nm), régie par un champ électrique très intense (> 10^5 V.cm-1) permettant l'ionisation par impact causée par les porteurs, le mécanisme responsable du comportement si particulier des diodes SPAD.

Le modèle ainsi obtenu, en prenant une boucle pour la zone de multiplication pour décrire la zone dans laquelle se produit l'ionisation par impact, et une boucle pour la zone d'absorption pour une première modélisation de la réponse moyenne de la SPAD. Cette boucle de zone d'absorption permet de prendre en compte le temps de transit des électrons vers la zone de multiplication ainsi que le temps de transit des trous de la zone de multiplication vers l'anode. Plusieurs améliorations clés ont également été faites sur la structure du modèle et son fonctionnement afin d'améliorer sa stabilité et robustesse. Un critère de charge pour le rechargement de la SPAD a résolu des problèmes d'oscillation internes et a ainsi amélioré la performance du modèle. Une autre intégration a été l'intégration du phénomène de non-recharge du dispositif qui peut se produire lorsque la résistance de recharge est trop faible. Finalement, en croisant l'ajustement de la tension de claquage de la diode avec l'ajout et l'ajustement de la dégradation du champ électrique causé par le grand nombre de porteurs, nous avons pu obtenir des résultats précis et cohérents lors de la comparaison des simulations avec les résultats de caractérisation transitoire. L'importance et la validité de ces résultats a été corroborée par la publication d'un article de revue IEEE au journal TED (Transactions on Electron Devices), accentuant la pertinance et le potentiel du modèle.

En combinant les deux étapes de construction du modèle présenté dans cette thèse, ainsi que les validations expérimentales correspondantes, nous avons pu confirmer l'importance et le potentiel de cette approche de modélisation en boucles. Quelques perspectives intéressantes peuvent être relevées afin d'améliorer et développer le modèle existant. En premier lieu, il y'a la collecte des trous retardée à l'anode. Des courbes de mesure transitoire révèlent en effet que le dispositif génère un pic de courant secondaire visible dans son comportement moyen, ce qui peut être assimilé au courant de trou traversant la zone d'absorption de faible champ pour être ensuite collecté par l'anode. Ensuite, il y'a également la modélisation statistique du jitter ainsi que de l'amplitude du pic de courant d'avalanche qui peuvent être traduit de façon physique et spatiale au travers du modèle en boucles, dû à son architecture spatialisée inhérente.

En se basant sur les objectifs initiaux de la these, le scope et le focus ont évolués au cours du temps. Au fil de la thèse, le projet a évolué d'un modèle de SPAD au sens large à un focus particulier sur la méthode modélisation distribuée dans l'espace. A cause de la nature divergente et interconnectée en interne de cette méthode, la nécessité d'une base de modèle solide et fiable était cruciale. La représentation graphique du modèle semble simple en apparence, mais tous les nœuds internes ainsi que toutes les équations sont étroitement liés, rendant ainsi le debug du code difficile. Au vu de la structure du modèle distribué, l'Afterpulsing et le compte d'obscurité (DCR) peuvent être facilement intégrés sous la forme de blocs séparés configurables. La sortie de ces modules serait connectée a un module déterminant la localisation de l'absorption du photon (ou la génération du porteur) en fonction de la longueur d'onde et/ou des propriétés du matériau. Tous ces modules peuvent être développés séparément et de façon indépendante avant la connexion avec le modèle principal. Une fois assemblés, le modèle complet serait donc capable de simuler le comportement complet de la SPAD, prenant en compte le DCR, l'Afterpulsing et tous les autres phénomènes stochastiques entrant en compte (jitter, amplitude…)

L'approche innovante de modélisation spatialement distribuée développée au cours de cette thèse vise à diversifier le portfolio d'approches de modélisation de SPAD disponibles et offre une nouvelle manière de décrire les dispositifs de type SPAD. Comme décrit précédemment, les résultats de cette étude et du développement sont prometteurs tant en termes de résultats de simulation que de par le potentiel qu'il représente. Les perspectives de ce modèle sont multiples et permettraient un grand potentiel dans les systèmes de détection dans des applications de pointe nécessitant une haute précision et une haute performance. Dans ces systèmes, la fidelité du modèle au dispositif réel est cruciale, et de par sa modularité, le modèle peut s'adapter au détriment du temps de simulation en augmentant le nombre de boucles, résultant en une description précise de jitter, de localisation d'absorption de photon etc. .

Résumé en anglais

Applications based on Pixel matrix imagers using CMOS technology and SPAD-type avalanche diodes have significantly expanded in the last decade with applications in biomedical imaging, environmental monitoring, industrial settings, and automotive systems. The SPAD diode operates in a unique mode compared to most semiconductor devices, requiring a model for simulation in a microelectronic design environment to meet the needs of CMOS imager designers using SPAD arrays. A distributed model, based on a loop structure reflecting the impact ionization phenomenon at the core of the device's operation, has been developed to address these requirements.

The main goal of this thesis was to develop a new approach in modelling Single Photon Avalanche Diode in order to provide circuit designers with an accurate and close to physics model. This approach has been decomposed into two main design steps:

A Single-Loop approach was first developed, translating the core physical phenomenon of impact ionization into a functional diagram of the mechanism. It was made possible by analyzing impact ionization's exponential increase of the number of implicated carriers, taking inspiration from control theory and its description of closed loop systems. This assimilation was especially close when considering divergent closed loop systems that are able to sustain the increase of the output signal. Once this convergence of ideas was translated into a graphical looped representation, the first bone of the model's skeleton was established. Following that, the implementation of the code came with several problems, mainly concerning the impact ionization coefficients that had to be weighted and adjusted to the distance that is considered. In the case of the Single-Loop approach, this resulted in the weighing of the coefficient by the narrow multiplication region width.

The framework of this new modelling approach has then provided proof of concept by the comparison of its simulated response to measurements on the physical device. This proof was further confirmed by the publication and presentation of a conference paper at ESSDERC/ESSCIRC 2022. The confrontation of the simulation results from this first step of the modelling approach to the real device's response was carried out on several key figures of merit like Charge Per Pulse or Pulse Width. The accuracy of the results was a milestone that lead to the extension of the single loop based modelling approach to a distributed modelling approach, using multiple elementary loops.

Subsequently, as mentioned in the previous paragraph, a distributed model was developed upon the results and structure of the Single-Loop model. The key idea behind this distributed modelling approach was the spatialization of the model, based on the device's structural architecture. By studying the structure of the modelled SPAD, using the schematic sections of the device and also taking into account physical simulations, we were able to deduce that it is a reach-through SPAD architecture. This structure has two distinct sections; multiplication and absorption region, which behave very differently. On the one hand is the absorption region which is a wide region that is mainly destined to maximize photon absorption and thus, improve the device's detection capabilities. The multiplication region on the

other hand, is a narrow section of the device where the very high electric field (> 105 V.cm-1) enables impact ionization, the mechanism responsible for the SPAD's special behavior.

The resulting distributed model, starting at a single loop for the multiplication region to describe the area where the impact ionization phenomenon is occurring, and a single loop for the absorption region to first model the mean response of the device. This takes into account both the transit time of the electron to the high electric field multiplication region and also the transit time of the hole flow from this same region towards the anode. Several key improvements were also made to the model's structure in order to correct undesired previous behavior and improve its overall stability and robustness. A charge based quenching criteria helped resolve internal oscillation issues occurring in some cases, enhancing reliability of the model. Another addition was the integration of the non-quenching behavior that can occur when the circuit's quenching resistor has too low of a value. Finally, conjointly with breakdown voltage tuning, the addition of the electric field degradation phenomenon lead to accurate simulation results when compared to the transient measurements of the SPAD. A publication in the IEEE journal "Transactions on Electron Devices" further corroborates the model's relevance and potential.

When combined, those two stages of constructing the presented model and their corresponding experimental validation enable us to confirm the potential of this modelling approach. Some very interesting perspectives can be raised to improve upon the existing model. The first of those is linked to the delayed hole collection at the anode. Measured current transients of the device showed that there is a secondary delayed current peak in the mean behavior, which can be assimilated to the hole current traversing the low electric field absorption region to be collected by the anode terminal. Then, there is statistical jitter modelling and also statistical amplitude modelling that can be physically and spatially translated and represented by the model, due to its spatial modelling architecture.

Provided the initial stakes of the thesis, the scope and focus have evolved over time. As the PhD went on, the project evolved from a broader SPAD model to a deeper focus on the spatially distributed modelling method. Due to the divergent and internally interconnected nature of this method, the need for a solid and reliable model basis was crucial. When represented graphically, the model seems straightforward and simple in appearance, but all internal nodes and equations are tightly linked to one another and thus, the debugging of the core was challenging. Due to the structure of the developed model, Afterpulsing and DCR can be easily implemented as separate configurable building blocks. The output of those modules would be connected to a module that determines the location of the photon absorption as a function of wavelength and the device's material properties. All those modules can be designed individually and independently from each other before connecting them to the main model. Once assembled, the complete model would be able to entire simulate the behavior of the SPAD, from DCR, to Afterpulsing and all other statistical phenomena (timing jitter, amplitude).

The innovative spatially distributed approach developed in the scope of this PhD is aimed to diversify the portfolio of available modelling approaches and offer a new way of describing SPAD devices. As previously described, the results of this study and development are promising in terms of results when compared to characterization data. Combined with the potential of all the perspective implementations and modules that can be added to the existing model structure, this approach can have a great potential in high precision and high-performance detection systems. In such systems, fidelity to the real device can be crucial, and as the model can be modulated by adding more loops resulting in a more precise description of jitter, photon arrival location dependency etc., further providing potential utility to the model.

Experimental validation of this model was carried out through a robust test structure developed in collaboration between STMicroelectronics and the ICube laboratory. Calibration of the model was achieved using this test structure and the obtained transient measurements to ensure accurate representation of the actual device's behavior.

Keywords : SPAD, Single-Photon Avalanche Diode, CMOS, Verilog-A, Compact Modelling