EPFL

Thèse n° 10 282

Silicon CMOS and InGaAs(P)/InP SPADs for NIR/SWIR detection

Présentée le 12 avril 2024

Faculté des sciences et techniques de l'ingénieur
Laboratoire d'architecture quantique
Programme doctoral en microsystèmes et microélectronique

pour l'obtention du grade de Docteur ès Sciences

par

Utku KARACA

Acceptée sur proposition du jury

Prof. G. Boero, président du jury
Prof. E. Charbon, directeur de thèse
Dr A. Rochas, rapporteur
Prof. A. Gulinatti, rapporteur
Prof. J.-M. Sallese, rapporteur

École polytechnique fédérale de Lausanne

2024

To My Friends,
My Family &
My Ashley

Acknowledgements

First of all, I would like to express my sincere gratitude to my supervisor, Prof. Edoardo Charbon, for his extraordinary guidance and help in this work. I also want to thank him for giving me the chance to work in such an amazing laboratory. Although there was a one-year delay, I am very grateful that he still considered me for a PhD position and saw value in me.

I would like to thank Dr. Claudio Bruschini for supervising me with regular meetings and sharing his experience to improve my knowledge and the quality of this work.

I am thankful to Prof. Giovanni Boero for being the president of my thesis committee. It was also a great experience to teach with him during my PhD.

I would like to thank Prof. Angelo Gulinatti, Dr. Alexis Rochas, and Prof. Jean-Michel Sallese for their participation in my jury committee, evaluating my thesis, and providing me invaluable feedback to improve its quality.

I would like to express my sincere gratitude to my Master's thesis supervisor at Middle East Technical University, Prof. Serdar Kocaman, who is one of the most important people in my career. He taught me about semiconductors and photonics and gave me a chance to work with him on exciting projects that greatly helped me finish this thesis work as well.

I also owe a debt of gratitude to Ekin Kızılkan, with whom I have worked since 2015, performed projects together, and did a lot of brainstorming and discussions throughout our academic career, besides for his invaluable friendship in and outside the laboratory.

I would like to thank old and new AQUA lab members Halil Kerim Yildirim, Baris Can Efe, Francesco Gramuglia, Pouyan Keshavarzian, Ming-lo Wu, Paul Mos, Augusto Carimatto, Ivan Michel Antolovic, Andrea Ruffino, Vladimir Pesic, Jad Bensehir, Yasemin Uzun, Tommaso Milanese, Andrada Muntean, Andrei Ardelean, Preethi Padmanabhan, Yang Lin, Chufan Zhou, Arin Can Ulku, Batuhan Keskin, Won Yong Ha, Kodai Kaneyasu, Samuele Bisi, Jiuxuan Zhao, and Yating Zou for their friendship, technical discussions, and enjoyable moments outside the

lab. I am also very happy to meet our lab administrative assistants, Brigitte Khan and Begonia Tora, and their help regarding everything in the lab.

I would like to thank my friends Juan Francisco Gonzalez Marin, Cheol-Yeon Cheon, Ulas Dinc, Asli Sahin, Mustafa Kangul, Silan Sahan, Selen Keles, Deniz Yalcin, Andac Yigit, Eray Kacar, Fatih Kalyon, Yigit Ozer, Kubra Circir, Muammer Kozan, and Halit Dolas for their friendship and support.

Last but not least, I would like to express my love to my family and my fiancee Ashley for always being with me, for supporting me, and for their unconditional love that made me a better person and helped me finalize my PhD. They are my most valuable things, and without any of them, I could not finish this work.

Renens, 4 December 2023 U. K.

Abstract

Applications demanding imaging at low-light conditions at near-infrared (NIR) and short-wave infrared (SWIR) wavelengths, such as quantum information science, biophotonics, space imaging, and light detection and ranging (LiDAR), have accelerated the development of NIR/SWIR single-photon detectors. Up to date, there have been various detector types performing single-photon detection at infrared wavelengths. Among them, single-photon avalanche diodes (SPADs) have gained significant attention thanks to their low noise near room temperature (< 100 cps), high detection efficiencies (> 50%), and low timing jitter (< 100 ps). In addition, the integration of SPADs with standard CMOS technologies has paved the way for the design of low-cost, large pixel arrays with embedded photon-counting and timestamping circuitry.

For detection at NIR/SWIR, Si CMOS and InGaAs(P)/InP-based SPADs appear as the most promising material systems. Si CMOS has the advantage of yielding low noise and integrating SPADs with on-chip readout circuits; however, the low absorption coefficient towards NIR is a bottleneck to enhancing detection efficiencies. In this sense, the wide depletion region approach was investigated in this work by designing SPADs in a 110 nm CIS technology with a 10 μm active diameter. The implemented techniques of doping compensation and double multiplication region showed that the former can be utilized to extend depletion region widths, whereas the latter increases the total avalanche breakdown probability in a wide depletion SPAD, leading to high NIR efficiencies at a relatively lower breakdown voltage. In the fabricated SPADs with doping compensation, 7.3% PDP and 68 ps timing jitter at 850 nm and 5 V_{ex} were achieved under the noise of 962 cps. Thanks to a fully substrate-non-isolated structure with graded substrate doping and enhanced breakdown probabilities, 25.5% PDP at 850 nm and 5.5 V_{ex} were obtained in the device with the double multiplication region with a 295 cps noise, all at room temperature. However, the jitter at 850 nm was deteriorated to 236 ps due to many detected diffused carriers creating a diffusion peak in the timing histogram.

InGaAs(P)/InP-based SPADs were also designed, fabricated, and characterized, targeting 1.06 μm and 1.55 μm wavelengths. Planar device structures based on double zinc diffusions were implemented to define the active and guard ring regions of the SPADs. A comprehensive study was conducted via TCAD simulations to adjust the multiplication region thickness and the depth difference between two diffusions, which allowed for optimization of noise and active area uniformity. The numerical study also enabled the removal of floating guard rings,

shrinking pixel sizes for future array implementation. The noise of a 10 µm device with an InGaAsP absorber and a 1.5 µm multiplication region at 5 V_{ex} was 14.1 kcps, 5.5 kcps, and 2.75 kcps at 273K, 253K, and 225K, respectively, at 10 kHz frequency and 100 ns gate-on time. It was shown that the maximum operating frequency can be increased up to 200 kHz without suffering from the afterpulsing effect near room temperature. The PDP can be increased up to 36% at 9 V_{ex} and 1060 nm wavelength, and the jitter was reduced to 118.4 ps at 5 V_{ex} and 1060 nm. A 10 µm SPAD with an InGaAs absorber and 1.5 µm multiplication region achieved 20% PDP at 1550 nm and 61 kcps noise, both at 6 V_{ex}. The jitter of this device was 123 ps at 1550 nm and 5 V_{ex}.

Key words: Single-photon avalanche diodes (SPADs), near-infrared (NIR), short-wave infrared (SWIR), light detection and ranging (LiDAR), quantum key distribuiton (QKD), near-infrared optical tomography (NIROT), time-correlated single-photon counting (TCSPC), CMOS image sensor technology, doping compensation, double multiplication region, InGaAs(P)/InP, III-V, 3D integration

Résumé

Les applications exigeant l'imagerie dans des conditions de faible luminosité dans les longueurs d'onde du proche infrarouge (NIR) et de l'infrarouge à ondes courtes (SWIR), telles que la science de l'information quantique, la biophotonique, l'imagerie spatiale et la détection et la télémétrie de la lumière (LiDAR), ont accéléré le développement de Détecteurs à photon unique NIR/SWIR. Jusqu'à présent, il existe différents types de détecteurs effectuant une détection de photons uniques dans les longueurs d'onde infrarouges. Parmi elles, les diodes à avalanche à photon unique (SPAD) ont retenu l'attention grâce à leur faible bruit proche de la température ambiante (< 100 cps), leur efficacité de détection élevée (> 50%) et leur faible gigue temporelle (< 100 ps). De plus, l'intégration des SPAD avec les technologies CMOS standard a ouvert la voie à la conception de grands réseaux de pixels à faible coût avec des circuits intégrés de comptage de photons et d'horodatage.

Pour la détection en NIR/SWIR, les SPAD à base de Si CMOS et InGaAs(P)/InP apparaissent comme les systèmes de matériaux les plus prometteurs. Le Si CMOS présente l'avantage de produire un faible bruit et d'intégrer des SPAD avec des circuits de lecture sur puce ; cependant, le faible coefficient d'absorption vers le NIR constitue un goulot d'étranglement pour améliorer l'efficacité de la détection. En ce sens, l'approche de la région d'appauvrissement large a été étudiée dans ce travail en concevant des SPAD dans une technologie CIS de 110 nm avec un diamètre actif de 10 µm. Les techniques mises en œuvre de compensation de dopage et de région de double multiplication ont montré que la première peut être utilisée pour étendre la largeur de la région d'appauvrissement, tandis que la seconde augmente la probabilité totale de claquage par avalanche dans un SPAD à appauvrissement large, conduisant à des rendements NIR élevés à une tension de claquage relativement inférieure. Dans les SPAD fabriqués avec compensation de dopage, 7.3% PDP et 68 ps gigue à 850 nm et 5 V_{ex} ont été obtenus sous un bruit de 962 cps. Grâce à une structure entièrement non isolée avec un dopage de substrat gradué et des probabilités de rupture améliorées, 25.5% de PDP à 850 nm et 5.5 V_{ex} ont été obtenus dans le dispositif avec la région de double multiplication avec un bruit de 295 cps, le tout à température ambiante. Cependant, la gigue à 850 nm a été réduite à 236 ps en raison de nombreuses porteuses diffuses détectées, créant un pic de diffusion dans l'histogramme temporel.

Des SPAD basés sur InGaAs(P)/InP ont également été conçus, fabriqués et caractérisés, ciblant

des longueurs d'onde de 1.06 µm et 1.55 µm. Des structures de dispositifs planaires basées sur des doubles diffusions de zinc ont été mises en œuvre pour définir les régions actives et d'anneau de garde des SPAD. Une étude complète a été menée via des simulations TCAD pour ajuster l'épaisseur de la région de multiplication et la différence de profondeur entre deux diffusions, ce qui a permis d'optimiser l'uniformité du bruit et de la zone active. L'étude numérique a également permis de supprimer les anneaux de garde flottants, réduisant ainsi la taille des pixels pour la mise en œuvre future du réseau. Le bruit d'un appareil de 10 µm avec un absorbeur InGaAsP et une région de multiplication de 1.5 µm à 5 V_{ex} était de 14.1k cps, 5.5k cps et 2.75k cps à 273K, 253K et 225K, respectivement, à une fréquence de 10 kHz et à un temps d'activation de 100 ns. Il a été démontré que la fréquence de fonctionnement maximale peut être augmentée jusqu'à 200 kHz sans souffrir de l'effet de rémanence à proximité de la température ambiante. Le PDP peut être augmenté jusqu'à 36% à 9 V_{ex} et 1060 nm de longueur d'onde, et la gigue a été réduite à 118.4 ps à 5 V_{ex} et 1060 nm. Un SPAD de 10 µm avec un absorbeur InGaAs et une région de multiplication de 1.5 µm a atteint 20% de PDP à 1550 nm et un bruit de 61k cps, tous deux à 6 V_{ex}. La gigue de cet appareil était de 123 ps à 1550 nm et 5 V_{ex}.

Mots clefs : Diodes à avalanche à photon unique (SPAD), proche infrarouge (NIR), infrarouge à ondes courtes (SWIR), détection et télémétrie par la lumière (LiDAR), distribution de clés quantiques (QKD), tomographie optique dans le proche infrarouge (NIROT), comptage de photons uniques corrélés dans le temps (TCSPC), technologie des capteurs d'images CMOS, compensation du dopage, région de double multiplication, InGaAs(P)/InP, III-V, intégration 3D

Contents

List of Figures

List of Tables

Acronyms

2DEG two-dimensional electron gas

APDs avalanche photodiodes

Au gold

BSI back-side-illumination

BTB band-to-band

C_d diode junction capacitance

C_p parasitic capacitance

C_t total capacitance

CCD charge-coupled device

CIS CMOS Image Sensor

CMi Center of MicroNanoTechnology

CMOS complementary metal-oxide-semiconductor

cps counts per second

DCR dark count rate

DLTS deep-level transient spectroscopy

DMZn dimethylzinc

dToF direct ToF

E electric field magnitude

E_g bandgap energy

ebCMOS electron-bombarded CMOS

f frequency

FET field-effect transistor

FGRs floating guard rings

FOV field of view

FSI front-side-illumination

FWHM full-width-at-half-maximum

GaAsP gallium arsenide phosphide

Ge germanium

GHz gigahertz

GR guard ring

H_2O_2 hydrogen peroxide

H_3PO_4 phosphoric acid

HCl hydrochloric acid

HHb deoxygenated hemoglobin

HVNW high-voltage n-well

HVPW high-voltage p-well

I_q characteristic quench current

In indium

InGaAs indium gallium arsenide

InGaAsP indium gallium arsenide phosphide

InP indium phosphide

iToF indirect ToF

K Kelvin

LEDs light emitting diodes

LiDAR light detection and ranging

M measured count rate

MEMS micro-electromechanical systems

N real count rate

NbN niobium nitride

NbTiN niobium titanium nitride

NFAD negative feedback avalanche diode

Ni nickel

NIR Near-infrared

NIROT Near-infrared optical tomography

nm nanometer

O_2Hb oxygenated hemoglobin

P Phosphorus

PDP photon detection probability

PMTs photomultiplier tubes

QDs quantum dots

QE quantum efficiency

QKD quantum key distribution

R_d dynamic resistance

R_q ballast resistor

ROICs read-out integrated circuits

S switch

SACM separate absorption-charge-multiplication

SAG selective area growth

SEM scanning electron microscope

Si silicon

Si_3N_4 silicon nitride

SIMS secondary ion mass spectrometry

SiO_2 silicon oxide

SMU source measure unit

SNR signal-to-noise ratio

SNSPDs superconducting nanowire single-photon detectors

SPADs single-photon avalanche diodes

SQUID superconducting quantum interference device

SRH Shockley-Read-Hall

SWIR short-wave infrared

T temperature

TAT trap-assisted-tunneling

TCAD Technology Computer-Aided Design

TCCC time-correlated carrier counting

TCSPC time-correlated single-photon counting

TDCs time-to-digital converters

TESs transition-edge sensors

Ti titanium

ToF time-of-flight

UBM under bump metallization

UV ultraviolet

V_b avalanche breakdown voltage

V_{ex} excess bias voltage

V_{op} operating voltage

W tungsten

WSi tungsten silicide

Zn Zinc

Zn_3P_2 zinc phosphide

1 Introduction

Near-infrared (NIR) and short-wave infrared (SWIR) wavelengths offer advantageous features and contrast in imaging, where visible and thermal detection become ineffective. Thanks to their unique properties, NIR/SWIR photodetectors can be utilized in a wide range of applications, such as imaging through haze, dust, and fog [1], imaging of tissues and organs [2], hyperspectral imaging to identify moisture levels and contamination of agricultural products [3], [4], and in night vision through nightglow phenomena [5].

Single-photon detectors are a special category of photodetectors that enable the capture of individual photons and the counting of them accurately. These detectors are characterized by virtually zero readout noise and high timing accuracy, generally in the order of tens of picoseconds. If single-photon detectors only suffer from low dark noise, denoted as dark count rate, they enable Poisson-limited detection, which is why they have gained attention over the last decades in quantum imaging and quantum information science [6]–[9]. Since the last decade, many quantum information applications and quantum cryptography methods, for instance, quantum key distribution (QKD), have preferred utilizing NIR/SWIR single-photons to be able to provide secure communication [10]–[13]. However, quantum information science is not the only field that can benefit from NIR or SWIR single-photon detectors. Applications such as light detection and ranging (LiDAR) [14], [15], biomedical imaging [16], and space imaging [17] could also benefit from single-photon detectors in photon-starved scenarios or in complex time-of-flight (ToF) multi-bounce scenes. Hence, research in the field of NIR/SWIR single-photon detection is currently moving forward at a fast pace.

1.1 NIR/SWIR detection and its advantages

NIR/SWIR detectors and image sensors have become increasingly important in recent years due to the unique advantages they present compared to their visible and longer wavelength infrared counterparts. Essentially, NIR covers the wavelengths between 700 nanometer (nm) and 1000 nm, and the SWIR region corresponds to the 1000 nm to 3000 nm part of the wavelength spectrum. There are many light sources that emit photons at these wavelengths.

Although the photon flux is reduced with respect to visible wavelengths, the Sun is still a photon source at NIR/SWIR. Thus, the reflected sunlight from the objects can be detected to perform 2D imaging. This is also the reason why NIR/SWIR images resemble typical images taken by photographic instruments in use nowadays. Secondly, the primary photon source may be active illumination devices, such as lasers and light emitting diodes (LEDs). In particular, lasers utilized in fiber-optic telecommunication require SWIR sensors for detection as the optical fibers have the lowest loss at these wavelengths. Besides, the overtones of molecular vibrations can be absorbed in the NIR/SWIR band, which offers signature identification of molecules that contain chemical bonds between hydrogen and oxygen, carbon, and nitrogen. Hyperspectral imaging is hereby another significant field, especially in farming. Lastly, the radiance of the night sky, which is also known as nightglow, allows night vision in SWIR. The reflected Moon and the star lights can be employed as photon sources, for instance, to do surveillance at night. In addition to discussing the photon sources, the distinct advantages of building an optical sensor for NIR/SWIR detection are listed in Fig.1.1.

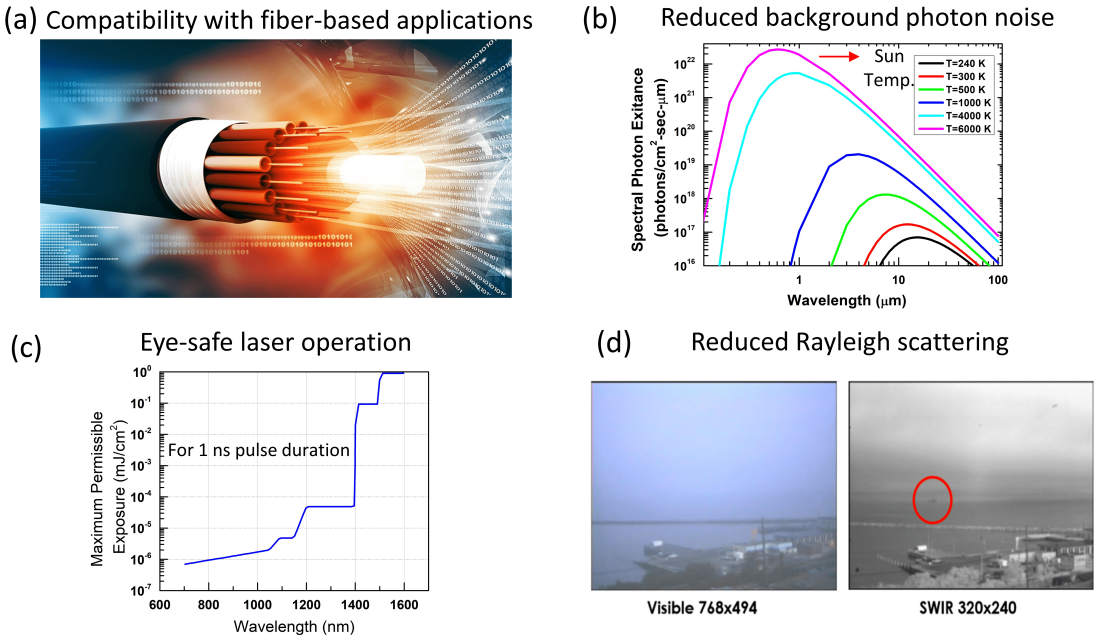


Figure 1.1: (a) Illustration of light propagation and data transmission in an optical fiber (Credits to Cadence). (b) Spectral photon exitance graphs belonging to various blackbody temperatures, including the Sun. (c) Maximum permissible exposure for eye-safe lasers for 1 ns pulse duration [18]. (d) Illustration of reduced Rayleigh scattering through images captured by a visible and a SWIR camera [19].

The first advantage of operating at NIR/SWIR wavelengths is to gain compatibility with fiber-based applications, as pictured in Fig. 1.1 (a). The standard fiber-optic telecommunication wavelengths are known to be 850 nm, 1310 nm, but mostly 1550 nm; all reside in the NIR or SWIR region. These wavelengths are chosen as they coincide with the minimum loss points for light propagation in an optical fiber. In general, a multi-mode 850 nm fiber has an

attenuation of 2.5 decibel/kilometer (dB/km), and single-mode 1310 nm and 1550 nm fibers have an optical loss of 0.3 and 0.2 dB/km, respectively [20]. Since dB is a logarithmic unit, even changing the communication wavelength from 1310 to 1550 can be interpreted as an almost 50% increase in transmission distance for the same data rate. Therefore, the most popular wavelength of operation in telecommunication is 1550 nm. Any optical communication system using one of these wavelengths requires a detector at the receiver side to convert this optical information into electrical one. Given that the optical loss is significant in a fiber-optic system and that single photons are utilized in an optical fiber-based quantum communication system to obtain a secure way of transmitting information through quantum states, designing single-photon detectors with high efficiency in NIR/SWIR is vital.

The second advantage of NIR/SWIR is that, in the case of using active illumination sources, the background photon noise exerted by the Sun is reduced. Free-space applications that need to be run under daylight conditions with a photon source have performance limitations due to the noise created by the enormous number of photons coming from the sunlight. The blackbody radiation of the Sun (a 6000 K object) in terms of spectral photon exitance is indicated in Fig.1.1 (b). As can be seen, the peak wavelength of the Sun resides in the visible region, whereas the number of photons from the sunlight decreases towards the infrared wavelengths. Hence, in order to reduce the background noise of the Sun, NIR/SWIR photon sources and detectors are excellent choices.

The third advantage is to acquire eye-safe operation for active illumination sources such as lasers and LEDs. Applications making use of these light sources, especially in free space, will have much less risk of damaging the human eye. As shown in Fig. 1.1 (c), the maximum permissible exposure of an 800 nm laser, which is still safe for the human eye, is 10^{-6} milijoule/centimeter^2 (mJ/cm^2)for 1 ns pulse duration [18]. However, laser power can be increased safely by 5× at 1100 nm, 50× at 1300 nm, and 10^6× at 1550 nm as seen from the graph. Using higher laser power levels at NIR/SWIR wavelengths also results in an increase in the maximum attainable depth since the laser photons can penetrate longer distances. This means that the field of view (FOV) of an imaging system can be improved.

The last advantage of moving to the NIR/SWIR bands is the reduced Rayleigh scattering. Rayleigh scattering occurs when light interacts with particles whose size is smaller than its wavelength. Therefore, it is a wavelength-dependent phenomenon that changes inversely with the fourth power of the wavelength ($1/\lambda^4$). In fact, this is the reason why the sky appears blue as the blue scatters more than the other colors. Thus, at longer wavelengths, Rayleigh scattering is less effective, which permits the light to travel more without perturbation and increases the FOV. A comparison of visible camera and SWIR camera images through the same hazy environment is provided in Fig.1.1 (d) [19]. As can be observed in these images, a visible camera cannot identify some of the objects under mist, whereas a clear vision of the shore in detail and even a ship at the farthest point of the FOV can be obtained with a SWIR camera thanks to less Rayleigh scattering. Hence, a NIR/SWIR imager provides better results through haze, which is an important property for applications doing 2D and 3D imaging.

1.2 Application areas of NIR/SWIR single-photon detectors

1.2.1 Light detection and ranging

LiDAR, which allows us to obtain 3D representations of environments and to construct high-resolution depth maps, has gained significant attention over the last decades. Autonomous vehicles, driving assistance, machine vision, virtual reality, robotics, and drone technologies are just some of the application areas of LiDAR, as illustrated in Fig. 1.2. The principle of depth sensing in a LiDAR system is based on recording the ToF of detected photons measured using direct and indirect methods [21]. In indirect ToF (iToF), the light source amplitude is modulated, and the phase shift between the emitted wave and reflected wave is measured to estimate the distance of an object. In order to increase the measurement range of iToF, one can decrease the frequency of the modulation, or use multiple frequencies to resolve ambiguity. This results in a trade-off between the distance, resolution, and complexity of the system. The main problem with iToF is the energy of the light source that is spread out in the period, thus creating the need for high light power to achieve reliable distance detection in the presence of high background illumination. Thus, highly collimated light can be used, like in the Bosch PLR15 telemeter, which achieves a point measurement at large distences. Or, alternatively, short distances can be used with a larger field of view [22]. Since a linear response is required due to the modulation, conventional photodiodes, or avalanche photodiodes (APDs) are preferred on the receiver side of an iToF system.

In direct ToF (dToF), typically a pulsed laser or LED is used as the illumination source. The laser is synchronized with the read-out integrated circuits (ROICs), and a kind of stopwatch starts to keep track of the time when the laser is fired. Then, the reflected light from the objects present in the surroundings is detected at the receiver side, and high time resolution ROIC translates the signal into time. Since the photons travel at the speed of light (c), the ToF information can be easily converted into distance (D) between the object and the receiver with the following formula:

$$D = c \times \frac{ToF}{2}. \tag{1.1}$$

As a result, according to the distance of different objects, a 3D depth map of the environment can be built based on timestamping. As shown in Fig.1.2, in a dToF LiDAR system with 20 ns timing resolution, targets that are 3 m apart from each other can be distinguished. If the timing resolution is decreased to 200 ps, the resolution in distance is increased to 3 cm, which would be a sufficient imaging resolution for most of the mentioned applications of LiDAR. Therefore, photomultiplier tubes (PMTs), single-photon avalanche diodes (SPADs) or superconducting nanowire single-photon detectors (SNSPDs) are considered for dToF imaging.

dToF LiDARs can be further divided into two categories: scanning and flash LiDARs. In a scanning LiDAR, the laser source scans the scene horizontally or vertically with the help

of mirrors or solid-state crystals, which are driven by actuators [23]–[25]. Since high laser power can be utilized for each shot and point in the FOV, high signal-to-noise ratio (SNR) can be achieved in scanning mode LiDARs. On the other hand, the dependence on the moving mechanical parts makes them bulky and can cause long-term reliability problems. Recently, micro-electromechanical systems (MEMS) based micromirrors have been integrated to address these issues and to make the system much smaller [26]. In a flash LiDAR, the whole target scene is illuminated in one shot instead of doing point-by-point scanning [27]–[29], which results in low-cost and reliable systems. In order to illuminate the FOV uniformly, an optical diffuser is used, which diverges the laser beam. Due to this light diffusion, laser power detected by the pixel array is weakened, which degrades the SNR and the resolution of the system over long distances. Yet, flash LiDARs are reliable systems for short-range applications.

Since LiDAR applications arise outdoors, the major problem becomes ambient Sun light, which introduces shot noise and false detections. However, at NIR/SWIR wavelengths, this problem becomes less of a concern thanks to the reduced background photon noise, as discussed. Similarly, to be able to use higher power for eye-safe lasers in the NIR/SWIR bands, the range and resolution of the LiDAR system can be improved as well.

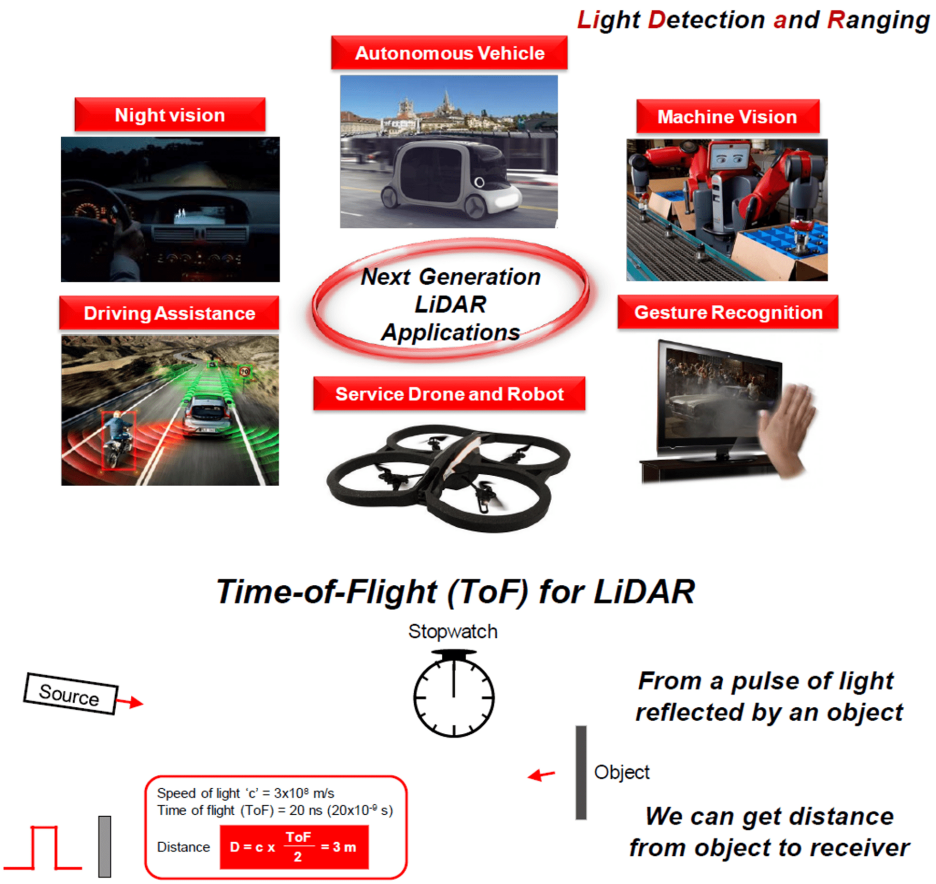


Figure 1.2: Applications and operation principles of LiDAR.

1.2.2 Quantum cryptography

Quantum cryptography is one of the most interesting and popular applications of quantum communication, which focuses on how to protect information in communication channels against eavesdropping. In one embodiment, known as QKD, one uses the quantum states of the photons to transfer information between two sides, possibly far from each other. Thanks to the quantum no-cloning phenomenon, these two individuals can determine if an outsider is trying to access the information in their communication link [30]. QKD systems can be implemented in free space, from the Earth to a satellite, or with optical fibers between two communication centers. Free space systems are beneficial to cover longer distances, such as in building an intercontinental network. However, their performance will be limited due to background noise coming from the sunlight. Conversely, fiber-based QKD offers reliable and stable operation regardless of the environmental conditions and the ability to use already-installed telecommunication infrastructure, but they cannot cover very long distances because of the fiber losses. Eventually, a hybrid system, which is composed of both free space and fiber-based communication, is inevitable in order to reach many users separated by long distances. With the latest technological advances, a 830 km fiber-based QKD link [31], and a hybrid 4600 km QKD system, which covers 2600 km in free space and 2000 km with fibers on the ground, have been established [32]. These demonstrations prove that secure communication and the internet for everyone is feasible in the near future. Another important parameter for a QKD framework is the key generation rate. To be able to satisfy the need for high data rates in some applications and to reach numerous users in the same network, high key generation rates are demanded. A typical QKD system can achieve a kilobits/second key rate, but recently, a record of 115.8 megabits/second key generation rate has been reported over a 10 km standard optical fiber, where the network also distributes keys over 328 km of low-loss fibers [33].

QKD systems demand robust and bright photon sources to generate a series of single photons since the purity and reliability of photons possess great importance. Hereby, single-photon detectors with high detection efficiencies have become another key element for these systems. Regarding the operation wavelength, 1550 nm has turned out to be more attractive as it gives compatibility to use existing telecommunication networks and to perform the process in daylight with much less background noise, which also leads to facilitating hybrid fiber-based and free space platforms. Most of the demonstrated QKD systems at 1550 nm utilize SNSPDs thanks to their very low noise and above 80% detection efficiency [31], [33]–[35]. A QKD system utilizing SNSPDs for high key generation rate at 1550 nm is presented in Fig.1.3 [33]. However, the necessity to cool these devices down to 4 Kelvin (K) appears to be the bottleneck for scalability, portability, and their integration into real user networks. Hence, indium gallium arsenide (InGaAs)/indium phosphide (InP) based SPADs, which can provide 40%-50% detection efficiency and can be operated with a thermoelectric cooler to achieve low dark counts, have been experimented with for their integration with QKD systems as well [32], [36]–[38]. As a result, developing single-photon detectors that can function near room temperature with high detection rates is critical for the achievability of secure communication in the form of QKD, whereas timestamping and timing resolution are not crucial parameters

for this application.

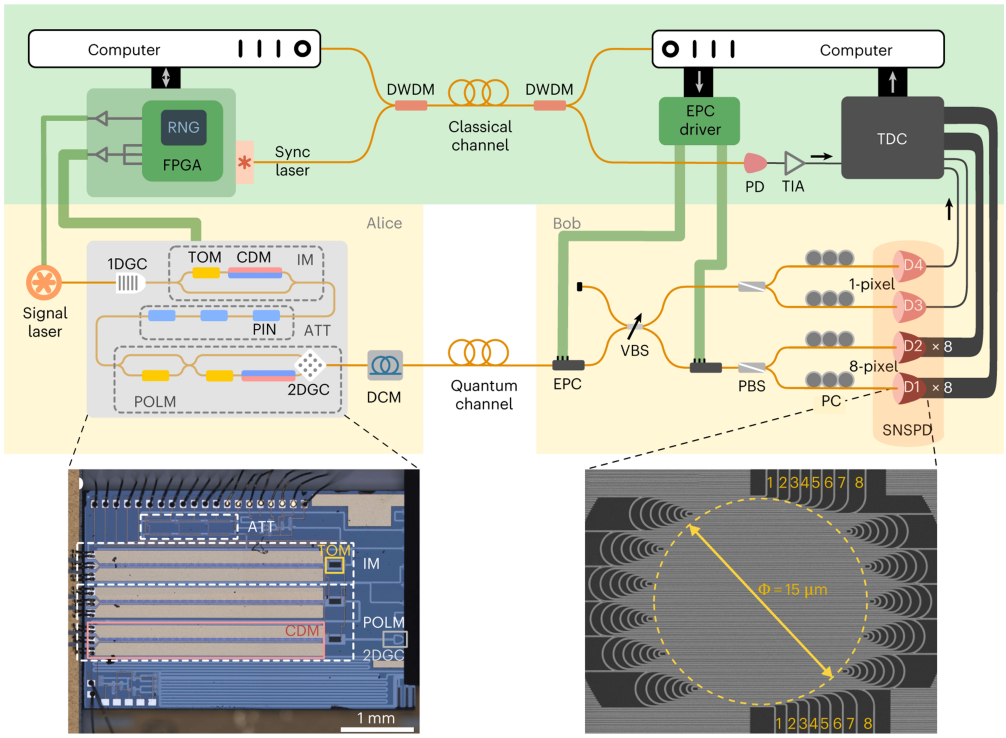


Figure 1.3: A QKD system, which is implemented for high key generation rate at 1550 nm, where 8 pixels SNSPD was used as the detector at the receiver side [33].

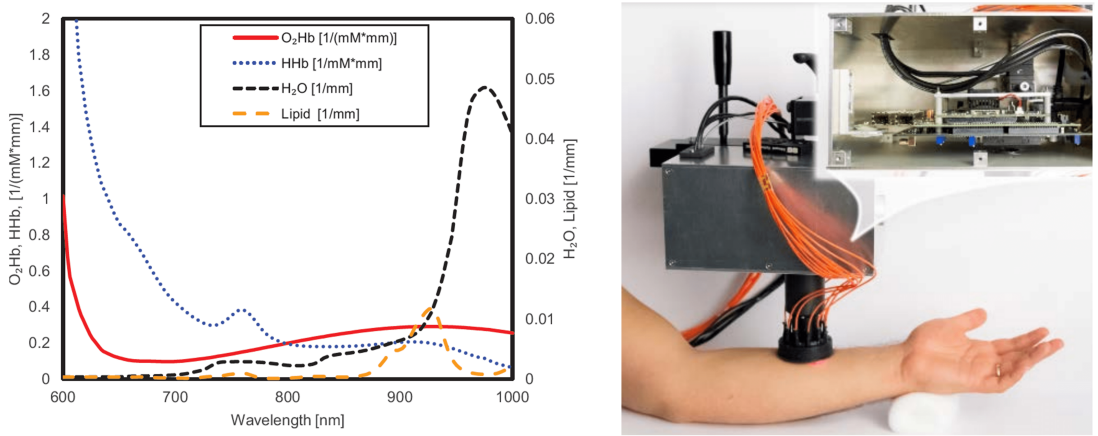
1.2.3 NIR optical tomography

Near-infrared optical tomography (NIROT), also called diffuse optical tomography, is an important subdomain of biomedical imaging that focuses on monitoring soft tissues like the brain or the breasts. In the human body, water and lipids are known to be the most common absorbers of light. In fact, these substances are reasonably transparent between 200 nm and 900 nm wavelengths. Conversely, for these wavelengths, hemoglobin, which is the blood cell that carries oxygen to the organs and tissues, becomes the main absorber in the forms of oxygenated hemoglobin (O_2Hb) and deoxygenated hemoglobin (HHb). However, the absorption of hemoglobin drops significantly between 650 nm and 900 nm, as depicted in Fig.1.4 (a) [39]. For this reason, 650 nm to 900 nm wavelengths are often defined as the biological optical window, where the light can penetrate deep inside the tissue (a few centimeters) and enable imaging.

In the case of less light absorption, scattering mechanisms play a role in determining light's journey. A high number of scattering events result in diffusive light in mediums like human tissues. Diffusive light can be modeled by the diffusion equation, which explains the angular dependency of the light intensity [40]. Hence, in practice, when a tissue is illuminated with a light source, the backscattered photons from the surface can be collected to construct

tomographic images and calculate the concentration of hemoglobin and the other absorbers. Eventually, measuring the oxygenation would be very beneficial in cancer treatments or preventing brain injuries in prematurely born infants.

In a typical NIROT setup, as can be seen in Fig.1.4 (b), optical probes are attached to the surface of a tissue for illumination [41]. The backscattered photons are then directed to a detector array via a lens. However, due to the limited surface area, the number of sources and detectors become limited, which is not desirable to increase and acquire more information from such a diffusive medium. Therefore, time-domain measurements have been integrated into NIROT systems to enhance the amount of information that can be obtained [16], [41], [42]. With the ToF information, the tissue scattering coefficient can be calculated, which also facilitates predicting precisely the concentrations of hemoglobin and other substances. Given that PMTs and SPADs can provide timing information based on the time-correlated single-photon counting (TCSPC) method and can operate at room temperature, they are mostly the preferred photodetector types, except that SPADs can be fabricated in large pixel arrays, which improves the spatial resolution of NIROT systems.



(a) (b)

Figure 1.4: (a) Absorption spectrum of the main absorbers in the human body, indicating an optical window at NIR for medical imaging. [39]. (b) A NIROT setup to reconstruct the features in the tissues [41].

1.3 Single-photon detector types

Single-photon detectors are basically the optical elements that convert the optical signal into the electrical one in the form of current. Since the magnitude of the electrical current generated by one photon is very low (1.602×10^{-19} A), single-photon detectors have either internal multiplication mechanisms (multiplication gain, photoconductive gain, etc.) or dedicated amplification circuits to create macroscopic current levels, which can be manipulated further by electronic circuitry. In the literature, there are several widely used single-photon detector

types. In the following subsections, the main detectors are categorized in terms of their operation mechanisms, and the pros and cons of each type are summarized.

1.3.1 Photomultiplier tubes

In a PMT, the incoming light excites the electrons in the absorber photocathode via the photoelectric effect, provided that the energy of the incident photons is higher than the work function of the photocathode material. The photogenerated electrons are then accelerated and directed onto the electron multipliers, also called dynodes, thanks to a focusing electrode. On the dynodes, the photogenerated electrons undergo a multiplication process through secondary electron emission. This secondary electron emission is repeated for several dynodes to achieve a total multiplication factor of approximately 10^6 [43]. Each dynode is biased at a very high voltage, which increases progressively towards the anode, to augment secondary electron emission and drive the electrons to the next dynode. At the final stage, a cloud of secondary electrons is collected from the anode. Fluctuation in time for electrons to form the electrical current causes time-broadened electrical signal output, so the timing jitter of PMTs should also be stated to determine the resolution of the TCSPC application. All these operations take place in vacuum to eliminate the random collision of electrons with the ambient atoms. Since the photocathode is made of a semiconductor material, the cutoff wavelength of operation is set by the material choice, which can be an alkali or III-V-based semiconductor and can cover a wide wavelength spectrum. An example schematic showing the basic operation principles of a PMT is depicted in Fig.1.5 [44]. Similar to PMTs, electron-bombarded sensors also use a photocathode and vacuum environment to detect single photons. The difference is that they use a sensor that can be made of silicon (Si) complementary metal-oxide-semiconductor (CMOS) or charge-coupled device (CCD) where a single photogenerated electron is accelerated onto it by a high voltage. This bombarded electron creates electron-hole pairs (one electron-hole pair per 3.7 eV) in Si. Since electron-bombarded CMOS (ebCMOS) imager lifetimes are restricted by the damage done to the sensor part, they are less utilized than PMTs.

In today's market, PMTs are commercially available [43]. Large-area PMTs (cm^2) made of gallium arsenide phosphide (GaAsP) and InGaAs for both visible and infrared single-photon detection have been demonstrated. With GaAsP photocathodes, photon detection efficiencies up to 40% around 500 nm wavelength and dark noise as low as 100 counts per second (cps) at room temperature have been achieved [43]. The timing jitter reported has been reduced to around 80 ps full-width-at-half-maximum (FWHM). Up to 16 channels of PMTs connected together have been successfully implemented. For the infrared wavelengths, 2% detection efficiency from 1000 to 1550 nm, 200×10^3 cps dark count at 200K and 80 ps jitter have also been obtained with InGaAs PMTs. The dark count rates can be further reduced by cooling down the devices, whereas photon detection efficiencies are limited by the efficiency of incident light knocking out the first electron from the photocathode. Regarding the alkali-based PMTs, a record of 43% efficiency at 350 nm, dark counts as low as 30 cps at room temperature, and

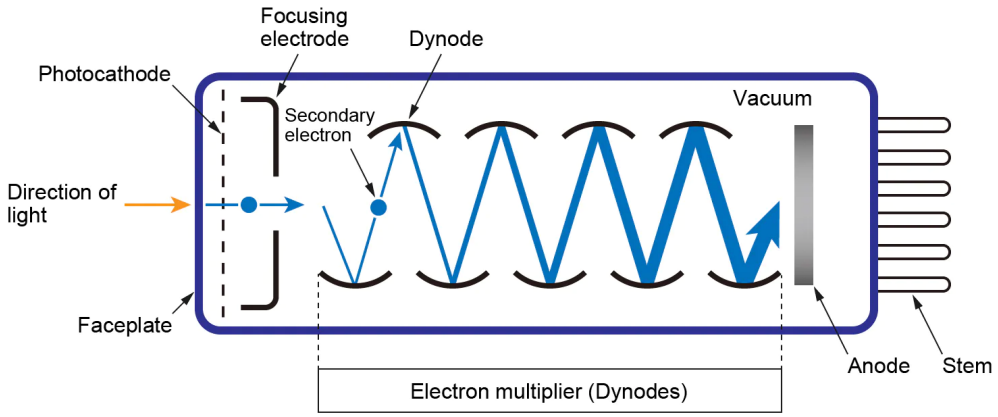


Figure 1.5: A schematic of a conventional photomultiplier tube [44].

45 ps jitter have been demonstrated by using an ultra bialkali-type material [43], [45]. A 64-channel PMT array has been fabricated and is available for customers. However, the major disadvantage of PMTs is the requirement of vacuum, which limits their lifetime and reliability. Also, as their operation is based on using very high voltages (at the kilovolt level) on dynodes, they become very fragile and expensive.

1.3.2 Single-photon avalanche diodes

SPADs are reverse-biased p-n junction photodiodes above their avalanche breakdown voltage. When the reverse bias voltage is increased to very high values, a very large electric field is formed in the depletion region of the diode. Under this circumstance, the carriers can acquire large kinetic energy, which results in breaking covalent bonds between lattice atoms by colliding with them. This process is called impact ionization [46]. Electron-hole pairs created by impact ionization can also create new pairs, and the reverse current of the diode grows like an avalanche, which introduces an internal gain, or multiplication factor, to the structure. The avalanche process is self-terminating if the diode is biased below but close to breakdown voltage. In this case, we refer to it as linear mode or proportional APDs, where there is a finite gain, typically 10 to a few hundred. It is self-sustaining if the diode is biased above breakdown because the electric field reaches a critical value for sustained avalanching. The operation above breakdown voltage is called Geiger mode, and diode current reaches macroscopic levels of a few mA with a very fast rising time. The I-V characteristics of a photodiode under various operation conditions is illustrated in Fig.1.6 (a). Although both linear mode and Geiger mode APDs can be used for single-photon detection, SPADs have become more widely used thanks to their very fast time response, low timing jitter, and easier integration with digital circuits.

SPADs can be designed with different material systems to target different parts of the wavelength spectrum. Si, III-V materials, and germanium (Ge) are currently leading the advance-

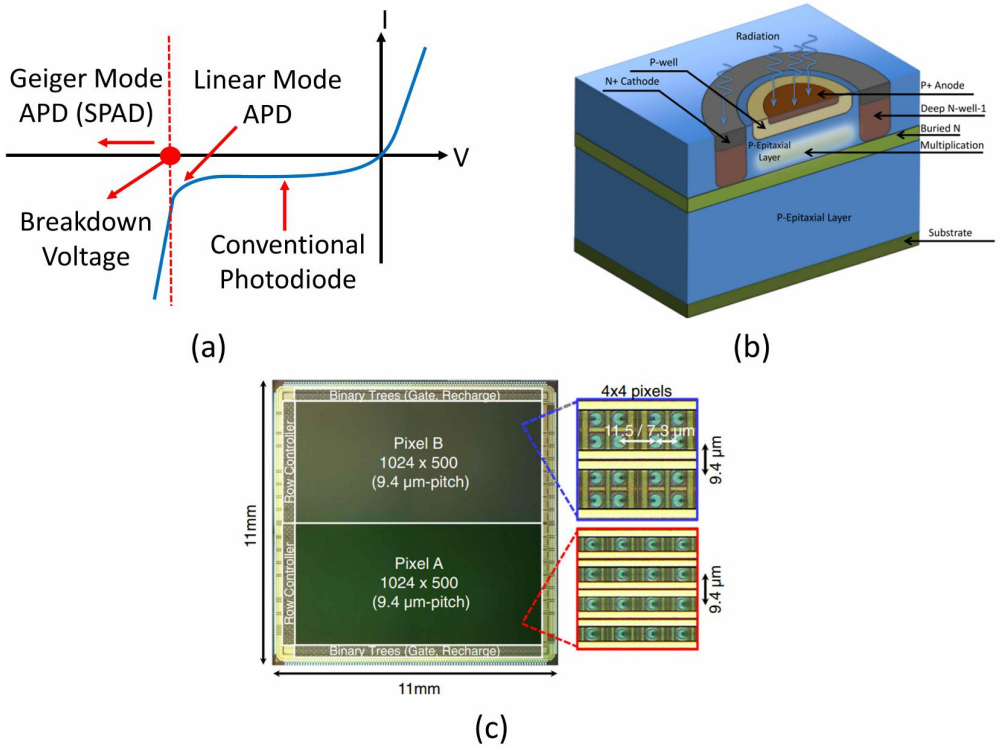


Figure 1.6: (a) Operation modes of a p-n junction photodiode. (b) Cross section of a conventional SPAD pixel in a CMOS technology [47]. (c) First megapixel SPAD array fabricated in 180 nm CMOS technology combined with time-gating circuit for 2D and 3D imaging applications [48].

ment of SPADs operating from ultraviolet (UV) to NIR/SWIR wavelengths. However, the cornerstone that paved the way for accelerating SPADs' development was the integration of SPADs with standard CMOS technologies [49], which is essentially utilized for designing analog and digital circuits. Mature Si CMOS processes have drastically increased the performance of SPADs in terms of noise, jitter, and detection efficiency and enabled to create low-cost large pixel arrays with a small pixel pitch. A cross section of a SPAD pixel in a CMOS technology is given in Fig.1.6 (b) [47]. Furthermore, CMOS technology allows to embed on-chip quenching, recharging, photon-counting, and timestamping circuits along with a SPAD pixel [50]. Even though they were not achieved simultaneously in a single device, in separate devices and current various CMOS nodes, less than 100 cps dark count rate (DCR) at room temperature [51], [52], more than 70% photon detection probability (PDP) [52]–[54] and less than 10 ps timing jitter [55] have been demonstrated for SPADs. Time-gating approach and time-to-digital converters (TDCs) have been successfully implemented for timestamping, which were also integrated with large format and megapixel SPAD arrays having less than 10 µm pixel pitch, as pictured in Fig.1.6 (c) [15], [26], [48], [56]–[60].

Although CMOS technologies offer very high performance, the cutoff wavelength of detection is limited to around 1100 nm due to the bandgap of silicon. Thus, for NIR and SWIR

detection, alternative material systems have been investigated. The most promising ones are InGaAs/InP-based and Ge-on-Si SPADs. InGaAs can provide high detection efficiencies thanks to its direct bandgap and, accordingly, high absorption coefficient. Its bandgap can also be tuned by adding phosphorus to the material to form indium gallium arsenide phosphide (InGaAsP). However, they can only be grown on lattice-matched InP substrates, thus requiring sophisticated techniques to integrate them with Si CMOS circuits. Therefore, the final module becomes more expensive than a Si SPAD product. On the other hand, Ge-on-Si SPADs offer on-chip integration of Ge SPADs with silicon circuits and photonics, whereas their detection performance will be worse than InGaAs because of its indirect bandgap, and their noise will also be degraded due to the defects introduced by the lattice mismatch between Ge and Si. In the last decade, high-performance InGaAs SPADs have been reported, where a few thousand cps DCR at 225K was measured, which can be easily achieved by thermoelectric coolers. More than 40% PDP at 1550 nm and less than 100 ps jitter have also been achieved [61]–[66]. Conversely, state-of-the-art Ge SPADs can reach 1.4×10^5 cps dark counts at 100K, 38% PDP at 1550 nm, and 300 ps jitter [67].

1.3.3 Superconducting nanowire single-photon detectors

This type of single-photon detector utilizes superconducting materials such as niobium nitride (NbN), niobium titanium nitride (NbTiN) or tungsten silicide (WSi), which are deposited and shaped to fabricate nanowires with electron-beam lithography on certain lattice-matched substrates such as silicon or sapphire. Below a specific temperature called the critical temperature (< typically 4K), superconducting behavior occurs, where material can be forced to switch to its resistive state if necessary current flows through the nanowire, which is called switching current. In the operation mechanism, upon the absorption of photons, a hot point is created on the device, which spreads along the nanowire and increases the total current density. When the current reaches the switching current value, a sharp voltage pulse is observed, which can be measured after amplification. Since the detection relies on a very narrow nm-sized wire, a meandering surface filling is preferred to widen the total photosensitive area, as shown in Fig. 1.7 (a). Also, in order to make the transition to the resistive state faster, SNSPDs are biased near their switching current value, which enhances jitter and detection efficiency. This scheme clearly states that SNSPDs can provide fast operation, low timing jitter, and high detection efficiencies. Dark counts are also very low due to cryogenic temperatures and are mostly limited by background photons.

SNSPDs may operate at infrared wavelengths, including 1550 nm for telecommunication and QKD, where they have high efficiencies > 90%, better than 100 ps jitter, and < 10 cps DCR. Recent breakthroughs were primarily thanks to the introduction of optical stacks, particularly Bragg reflectors and cavities, to enhance material absorption efficiency, as illustrated in Fig. 1.7 (b), and better fiber coupling [68], [69]. Currently, WSi SNSPDs with 93% detection efficiency, 1 cps dark count rate, 150 ps FWHM timing jitter at 3.6 μA biasing current and below 2K [69], NbN SNSPDs with 92.1% detection efficiency, 10 cps dark count rate and 79 ps FWHM timing

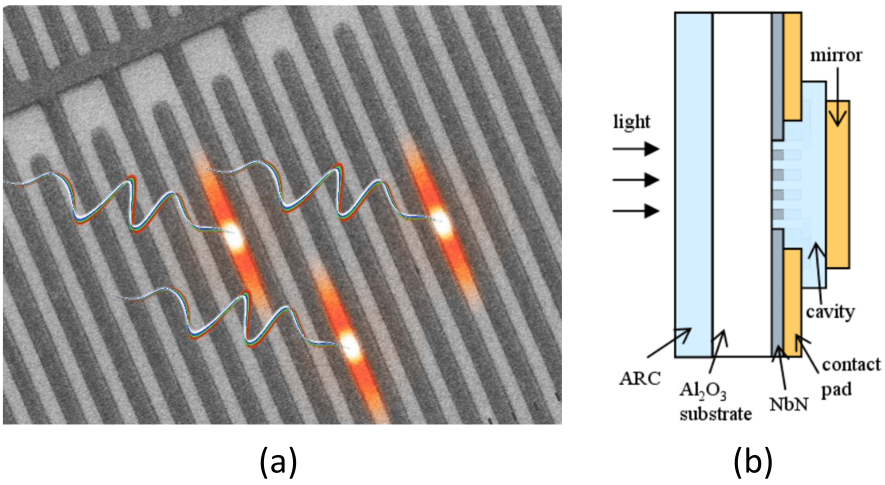


Figure 1.7: (a) Nanowire heat-up process via photon absorption (Credits to Duke University). (b) An SNSPD cross section integrated with an optical cavity and mirrors to enhance absorption efficiency [68].

jitter at 13 μA biasing current and below 2.1K [70], and NbTiN SNSPDs with 74% detection efficiency, 100 cps dark count rate and 68 ps FWHM timing jitter at 18 μA biasing current and below 2.3K [71] can be achieved. The major downside of SNSPDs is the requirement of cryogenic cooling, which is not practical for most of the applications. Also, creating large-format pixel arrays with SNSPDs is challenging due to biasing and cooling power limitations on the ROIC side. The current efforts have yielded an SNSPD array of 1024 pixels [72].

1.3.4 Superconducting transition-edge sensor single-photon detectors

Like SNSPDs, superconducting transition-edge sensors (TESs) make use of the perturbation of the resistance of superconducting materials with local heating. Unlike SNSPDs, TESs are voltage-biased devices. As demonstrated in Fig. 1.8, the absorbed photons start to heat the absorber material of the device and disturb the superconducting state of the TES [73]. As a result, the resistance of the TES also increases. Since the TES is biased with a constant voltage, this resistance change creates current, which can be measured through a typical superconducting quantum interference device (SQUID). SQUID in the scheme can be thought of as a thermometer. The variation in resistance depends on the absorbed photon energy, which is directly related to the number of absorbed photons in the case of a monochromatic light source. Thus, TESs are capable of resolving single photons in a quantum optical system [74], [75]. Eventually, the absorbed energy is dissipated out of the TES, usually with a thermal link, and the TES returns to its equilibrium state with a thermal time constant.

In TESs, in order to enhance the response speed and the energy resolution, materials with very low critical temperatures, such as titanium (Ti) or tungsten (W), are often chosen as an absorber [73]. These materials are also good candidates because they have low heat capacity,

which decreases noise by creating a large change in TES' resistance. However, when these thin film metals are utilized in a TES, the absorption efficiency and, accordingly, single-photon detection efficiency are deteriorated due to the reflections from the surface. Therefore, to boost the detection efficiency, TESs have been placed in multi-layer optical cavities, or the coupling efficiency between the optical fibers and the device can be improved. Thanks to these techniques, 98% single-photon detection efficiency has been reported for a Ti-based TES at 850 nm, operating around 300 mK [74]. For W-based TESs, 95% detection efficiency at 1556 nm has been demonstrated at 178 mK [75]. The main disadvantage of TESs is the requirement for a very advanced cooling system to reach 100 mK temperature levels. Moreover, the jitter of the reported devices is rather poor (100 ns), which is limited by the read-out circuit [74], [75]. Current efforts have yielded 4 ns jitter by optimizing the coupling inductance in the SQUID circuit to broaden signal bandwidth [76], but still, this is not very suitable for time-resolved applications. Lastly, the maximum achievable count rate is also low (around 100 kHz) for these detectors [7].

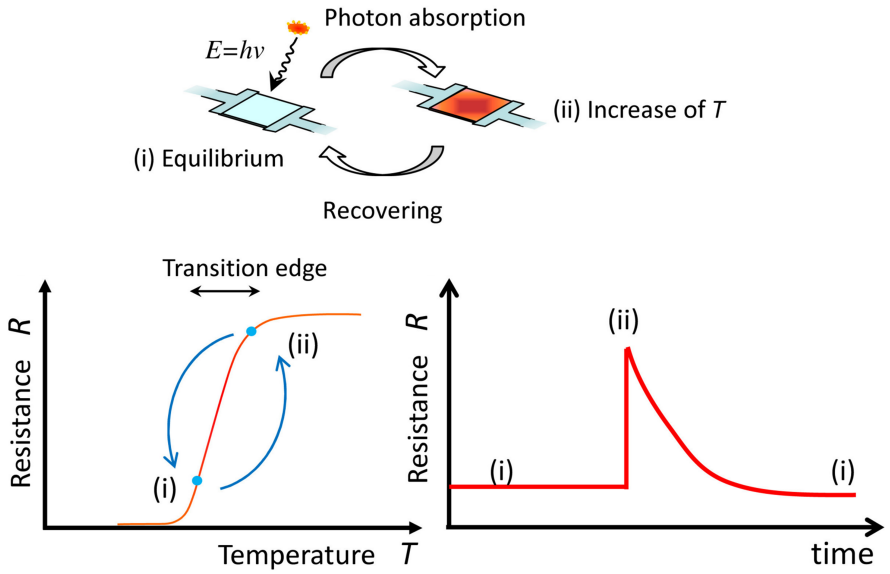


Figure 1.8: Operation mechanism and phases of a TES [73].

1.3.5 Quantum dot based single-photon detectors

In this detector type, there are several approaches to be able to utilize quantum dots (QDs) to sense single photons. In the first scheme, as shown in Fig.1.9 (a), QDs are self-assembled in a semiconductor absorber during the growth, and the absorber is also inserted into a field-effect transistor (FET) [77], [78]. Upon the absorption of the photons, the generated electrons are swept to two-dimensional electron gas (2DEG), which forms the channel between the source and the drain, whereas the holes drift to QDs and get trapped. Positively charged holes in QDs basically screen the electric field and increase the bias at the gate, which in turn also increases the channel current. Until the trapped hole gets recombined with the electron to

contribute to the channel current, the electron makes many round trips that introduce a photoconductive gain to the device. This means that a big rise in the current occurs over time, and the structure becomes highly sensitive to the photons. With this optically gated QDFET design, 68% internal quantum efficiency, which resulted in 2-3% overall detection efficiency at 805 nm and 0.003 cps noise at 4K, have been obtained. The main reason for the low photon detection efficiency is the relatively thin active absorber thickness. In addition, there were optical losses from the gate metal, which can be improved further with new design and fabrication techniques. The major downside of this type of detector is also the cooling requirement to extend recombination lifetimes and photoconductive gain.

In the second scheme, QDs are self-assembled in a resonant tunneling diode [79]. The cross section, SEM image of the fabricated device, and its energy band diagram are given in Fig.1.9 (b). The diode has a double-barrier tunnel layer, which limits the current between the emitter and the collector nodes. At certain forward bias voltages, a resonance tunneling current occurs when the conduction band energies behind the emitter and in the quantum well are aligned. The electrons trapped in QDs induce a potential and change this tunneling current. When photogenerated holes neutralize these trapped electrons, the tunneling current is further modified sharply, which permits the detection of single photons. The characterization of this device shows that 12% detection efficiency at 550 nm with only 0.002 cps dark noise can be achieved at 4 K. As in the first scheme, the cooling requirement is an obstacle for these detectors to integrate with circuits and construct large arrays. Besides, the device has a large 150 ns timing jitter, thus restricting its use in time-resolved applications.

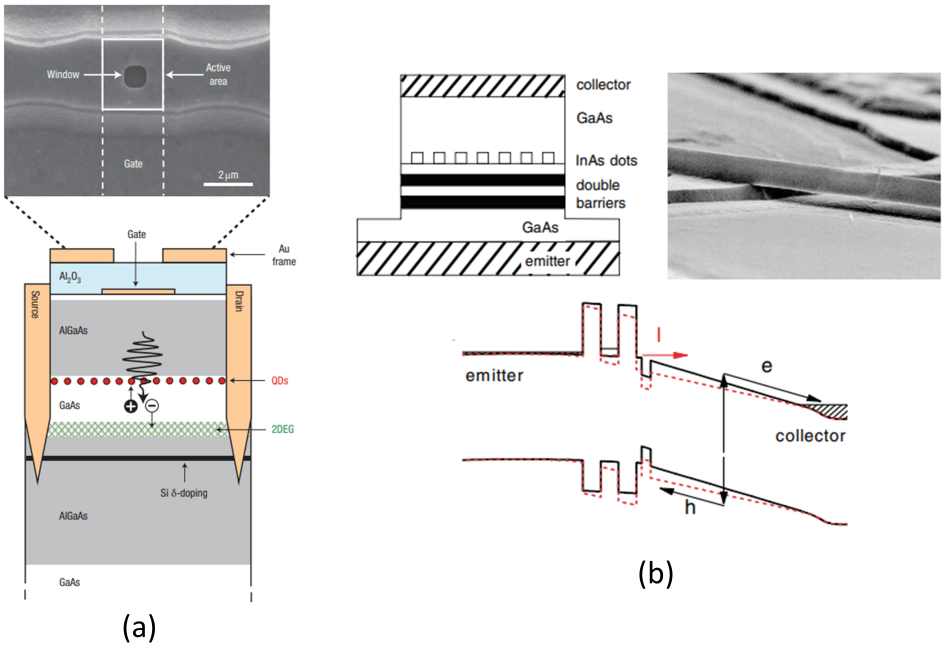


Figure 1.9: (a) An illustration of an optically gated QDFET [78]. (b) A QD resonant tunneling diode device [79].

1.3.6 Figures of merit comparison of single-photon detectors

To summarize the discussion, the figures of merit comparison of the presented detector types is provided in Table 1.1.

Table 1.1: Comparison of the state-of-the-art single-photon detectors

Detector Type	Detector Size	Operation Wavelength	Dark Counts	Detection Efficiency	Timing Jitter	Array Size
GaAsP PMT [43]	5 mm Diameter	280 nm to 720 nm	100 cps at 300K	40% at 500 nm	80 ps (FWHM)	16
Ultra Bialkali PMT [43]	23 × 23 mm^2	300 nm to 650 nm	30 cps at 300K	43% at 350 nm	45 ps (FWHM)	64
InGaAs/InP PMT [43]	3 × 8 mm^2	300 nm to 1700 nm	200×10^3 cps at 200K	2% at 1550 nm	80 ps (FWHM)	No Array
Si CMOS SPAD [56]	<16 µm	350 nm to 1100 nm	7.5 cps at 300K	50% at 520 nm	97.2 ps (FWHM)	512 × 512
Si CMOS SPAD [53]	8 µm Diameter	350 nm to 1100 nm	1.2×10^3 cps at 300K	72% at 560 nm	77 ps (FWHM)	No Array
Si CMOS SPAD [52]	<6.39 µm	350 nm to 1100 nm	1.8 cps at 300K	69.4% at 510 nm	100 ps (FWHM)	2072 × 1548
InGaAs/InP SPAD [61]	10 µm Diameter	920 to 1700 nm	20×10^3 cps at 225K	50% at 1550 nm	70 ps (FWHM)	No Array
InGaAsP/InP SPAD [62]	10 µm Diameter	920 to 1100 nm	500 cps at 253K	40% at 1064 nm	300 ps (FWHM)	256 × 64
InGaAs/InP SPAD [63]	25 µm Diameter	920 to 1700 nm	1×10^3 cps at 225K	30% at 1550 nm	46 ps (FWHM)	32 × 32
Ge-on-Si SPAD [67]	100 µm Diameter	Visible to 1550 nm	1.4×10^5 cps at 100K	38% at 1550 nm	300 ps (FWHM)	No Array
WSi SNSPD [69]	15 × 15 μm^2	1520 nm to 1610 nm	1 cps at 2K	93% at 1550 nm	150 ps (FWHM)	No Array
NbN SNSPD [70]	18 µm Diameter	1550 nm	10 cps at 2.1K	92.1% at 1550 nm	79 ps (FWHM)	No Array
NbTiN SNSPD [71]	15 × 15 μm^2	1550 nm	100 cps at 2.3K	74% at 1550 nm	68 ps (FWHM)	No Array
Ti-based TES [74]	10 × 10 μm^2	830 nm to 870 nm	Negligible at 300 mK	98% at 850 nm	100 ns (FWHM)	No Array
W-based TES [75]	25 × 25 μm^2	1310 nm to 1556 nm	Negligible at 178 mK	95% at 1556 nm	100 ns (FWHM)	No Array
QDFET [77], [78]	2 × 2.4 μm^2	805 nm	0.003 cps at 4K	2-3% at 805 nm	-	No Array
QD Tunneling diode [79]	1 × 1 μm^2	400 nm to 800 nm	0.002 cps at 4K	12% at 550 nm	150 ns (FWHM)	No Array

1.4 Motivation and the aims of the thesis

The motivation of this thesis is the analysis and characterization of NIR/SWIR single-photon detectors, which can operate at or near room temperature for integration in compact systems, particularly LiDAR and QKD applications. Both applications favor keeping the operation temperature as close as possible to room temperature to reduce cost and complexity. The trend in LiDAR systems is to operate in NIR/SWIR, so as to benefit from reduced background noise and higher laser power while satisfying eye-safety conditions. Scalability of SPADs, thus smaller pixel pitches, and timing resolution are other important factors that are required in LiDAR. For QKD, NIR/SWIR detection is required due to compatibility with low-loss optical fibers used in these systems. Conversely, timestamping and scalability are not crucial factors since QKD, in general, utilizes single devices without the need to access the ToF information. To improve the optical coupling between the fiber and the detector, larger device sizes are also preferable. Hence, timing jitter and the capability of creating SPAD arrays are the motivations of this thesis regarding only LiDAR applications, whereas achieving larger device sizes with low noise is the drive of this work considering only QKD.

In this sense, by interpreting Table 1.1, Si CMOS SPADs and InGaAs(P)/InP-based SPADs can meet these expectations for LiDAR and QKD at NIR/SWIR wavelengths. Even though SNSPDs and TESs can deliver better performance in terms of noise and detection efficiency if compared with Si and InGaAs/InP SPADs, cryogenic cooling and scalability concerns make them unsuitable in many applications. The main deficiency in CMOS SPADs is low PDP levels in the NIR due to reduced Si absorption. On the other hand, the main challenge in InGaAs(P)/InP SPADs is increased noise levels near room temperature owing to the lower bandgap compared to Si. Also, there is room for improvement in PDP for InGaAs(P)/InP SPADs since they have not been studied as extensively.

In this thesis, we focus on (a) developing new techniques to enhance NIR efficiency in Si CMOS SPADs, (b) optimizing noise and photon detection efficiency in InGaAs(P)/InP SPADs, and (c) shrinking pixel pitch to achieve large pixel arrays. The means to achieve these goals are extensive theoretical analysis, numerical simulations, and robust experimental activity on fabricated devices.

1.5 Contributions

The contributions of this thesis work can be listed as follows:

- A doping compensation technique was implemented for the first time to broaden the active region of Si SPADs in standard CMOS technologies, thus increasing NIR PDPs.
- The double multiplication region method was developed, which enables the designers to insert a second multiplication region into the depletion regions of the SPADs. This method has proven to enhance the total avalanche breakdown probability, leading to

higher visible/NIR PDPs at a relatively lower excess bias voltage in Si SPADs in standard CMOS technologies.

- A comprehensive study was conducted for InGaAs(P)/InP SPADs targeting 1.06 μm and 1.55 μm wavelengths to optimize the epitaxial structure for near room temperature operation with high PDPs and low timing jitter.

1.6 Thesis organization

Chapter 2 focuses on the operation principles, figures of merit, and metrology of SPAD devices. Chapter 3 gives an outlook on the methodology and current status of NIR-enhanced Si CMOS and NIR/SWIR InGaAs(P)/InP SPADs. Chapter 4 is dedicated to the proposed techniques and device structures to achieve a NIR sensitivity boost in Si CMOS SPADs, where a wide depletion region approach is investigated with a doping compensation technique and a novel double multiplication region method is described. In Chapter 5, the work performed for the development of NIR/SWIR InGaAs(P)/InP SPADs is explained. Devices targeting 1.06 μm and 1.55 μm cutoff wavelengths are demonstrated. Device designs, fabrication methods built up in the cleanroom, and the experimental results of the manufactured devices are presented. The integration process of InGaAs(P)/InP SPADs with Si platforms is expressed. Chapter 6 gives a summary, concluding remarks, and a future perspective of the work described in the thesis.

2 Figures of Merit and Metrology of SPADs

This chapter provides an overview of the physics behind the operation of SPAD devices, as well as information on their performance metrics and characterization methods.

2.1 SPAD's working principles

As briefly mentioned in the Introduction chapter, the generated carriers upon absorption of photons are accelerated by a very high electric field in the depletion region. With the increase in reverse bias, carriers can acquire enough energy to generate new electron-hole pairs through impact ionization when the device reaches its critical electric field [46]. The impact ionization is not a deterministic process but rather a statistical one. Therefore, to quantify the rate of ionization, the impact ionization coefficients are individually expressed for the electrons (α_e) and holes (α_h). Impact ionization coefficients basically correspond to the reciprocal of the mean free path between two ionization events. With the increase in electric fields, the ionization coefficients and rates are also enhanced, leading to a higher chance of triggering an avalanche [50], [80]. Moreover, the critical electric field and the ionization coefficients are linked to the type of material. For example, for Si, the critical breakdown field is about 3×10^5 V/cm where $\alpha_e > \alpha_h$ [81], [82], whereas for InP, it becomes 5×10^5 V/cm where $\alpha_h > \alpha_e$ [83]–[85]. The newly generated carriers by ionization can start new impact ionization processes as well, resulting in avalanche multiplication. SPADs are operated beyond this avalanche breakdown voltage (V_b) so that the avalanche multiplication process is self-sustaining. The biasing voltage beyond breakdown voltage is known as excess bias voltage (V_{ex}).

In the transient analysis of Geiger mode operation, there are five phases: seeding, build-up, spreading, quenching, and recharge [50], [86]. These phases are summarized in Fig. 2.1, which is a zoomed version of a dark I-V characteristic of a diode close to its avalanche breakdown voltage, as follows:

- Seeding: an electron-hole pair is generated, or either an electron or hole is injected into the depletion region upon photon absorption. At this moment, the occurrence

of the avalanche breakdown has a certain probability based on the impact ionization coefficients. The exact condition for the avalanche breakdown to occur will be discussed in the "Breakdown voltage" subsection.

- Build-up: assume that an avalanche is triggered during the seeding phase. Then, the local current density starts to increase rapidly with the positive feedback from the impact ionization process. In fact, the avalanche current flow is finite due to the space-charge effect in the depletion region [81]. The charge of the mobile ions in the depletion region creates an internal resistance for the diode, which causes a voltage drop across the device. Due to this negative feedback mechanism, the local potential can decrease to the breakdown voltage.

- Spreading: when the positive and negative feedback mechanisms are balanced, the avalanche spreads across the junction via the multiplication-assisted diffusion process [86], [87]. The speed of this spreading is about 10-20 $\mu m\ ns^{-1}$ [50], [86], [87]. At the end of this phase, the macroscopic current level is achieved, which can be detected externally.

- Quenching: after the registry of the photon hit, the device is quenched through a discharging element to stop the avalanche so that the SPAD does not get damaged because of overheating. Quenching is finalized when all the generated carriers leave the depletion region and the diode voltage goes back to V_b.

- Recharge: to restore the SPAD for the next photon detection, the same discharging path can be used to bring it back to the same initial biasing condition. This process is called recharging, and the SPAD is reset to its idle state.

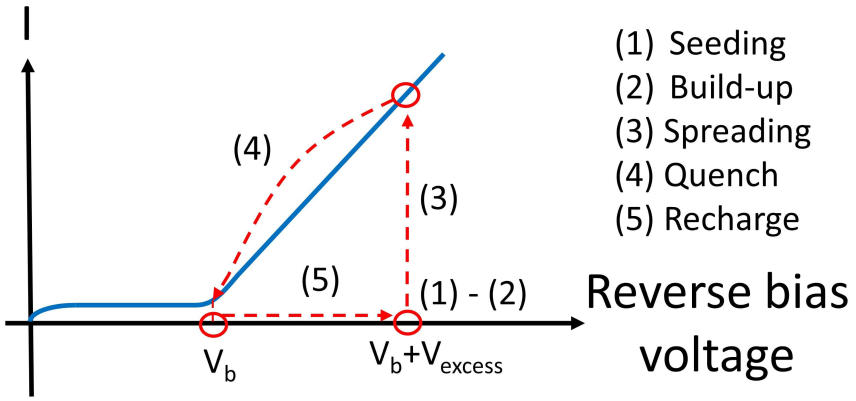


Figure 2.1: The phases of an avalanche process.

2.2 Quench and recharge techniques

As described, quenching and recharge are required to arm the SPAD for the next photon detection event. In principle, it is accepted that the SPAD is insensitive to the incoming

photons during quenching and recharge. Therefore, this total time duration is considered the "dead time" of the SPAD. Commonly, quenching and recharge can be passive or active [88]. The dead time and characterization of the devices are subjected to these techniques, which will be detailed in the "Figures of merit of the SPADs" section.

2.2.1 Passive technique

In passive mode, the SPAD is quenched and recharged through a ballast resistor (R_q), whose value can vary from 100 kΩ to several MΩ [86], [88]. R_q can be connected externally to the SPAD as a bulk resistor, or it can be implemented with a transistor operating in weak-inversion mode in a CMOS circuit. The schematic of a SPAD quenched and recharged passively is illustrated in Fig. 2.2 (a), under an applied operating voltage (V_{op}).

To elaborate on the quench and recharge mechanisms in the passive scheme, an equivalent circuit of the schematic is provided in Fig. 2.2 (b). The SPAD is represented at the upper part of the circuit on the right, which is composed of three parallel branches. The first branch includes the dynamic resistance (R_d), which is the resistance of the space-charge region of the diode combined with any ohmic contact resistance introduced during the fabrication. The avalanche current is modeled with a switch (S). On the onset of the avalanche, the switch becomes closed, which occurs when V_{op} exceeds V_b by V_{ex}, V_{op}= V_b + V_{ex}. The second branch has the diode junction capacitance (C_d), corresponding to the capacitance created across the depletion region of the diode due to the stored charges, which depends on the depletion width and the area of the device. The third branch indicates the parasitic capacitance (C_p) coming from the experimental setup. The lower part of the circuit contains the quenching resistor R_q.

In the absence of the avalanche current, the switch is open, and the applied V_{op} charges the capacitances C_d and C_p. As they are in parallel, the total capacitance (C_t) can be written as C_d + C_p. As soon as an avalanche is triggered while increasing the applied V_{op} above V_b, the switch is closed, and the avalanche current increases immediately with a negligible delay from the avalanche build-up dynamics. The avalanche current then starts to discharge C_t exponentially, thus dropping the voltage across the diode and, accordingly, the total current flowing. The time constant of this decay can be simply expressed as $C_t \times R_d$, assuming that $R_q >> R_d$ in the parallel configuration. Thus, at the steady-state, the diode voltage and current converge to [88]:

$$I_{steady\text{-}state} = \frac{V_{ex}}{R_d + R_q} \approx \frac{V_{ex}}{R_q}, \tag{2.1}$$

$$V_{steady\text{-}state} = V_b + R_d \times I_{steady\text{-}state}. \tag{2.2}$$

As shown in Eq. 2.2, $V_{steady\text{-}state}$ goes very near to V_b, provided that $R_q >> R_d$. This also means that less and less carriers are getting ionized while approaching $V_{steady\text{-}state}$, since the impact ionization rates depend on the voltage via the electric field magnitude. Hence, it can come

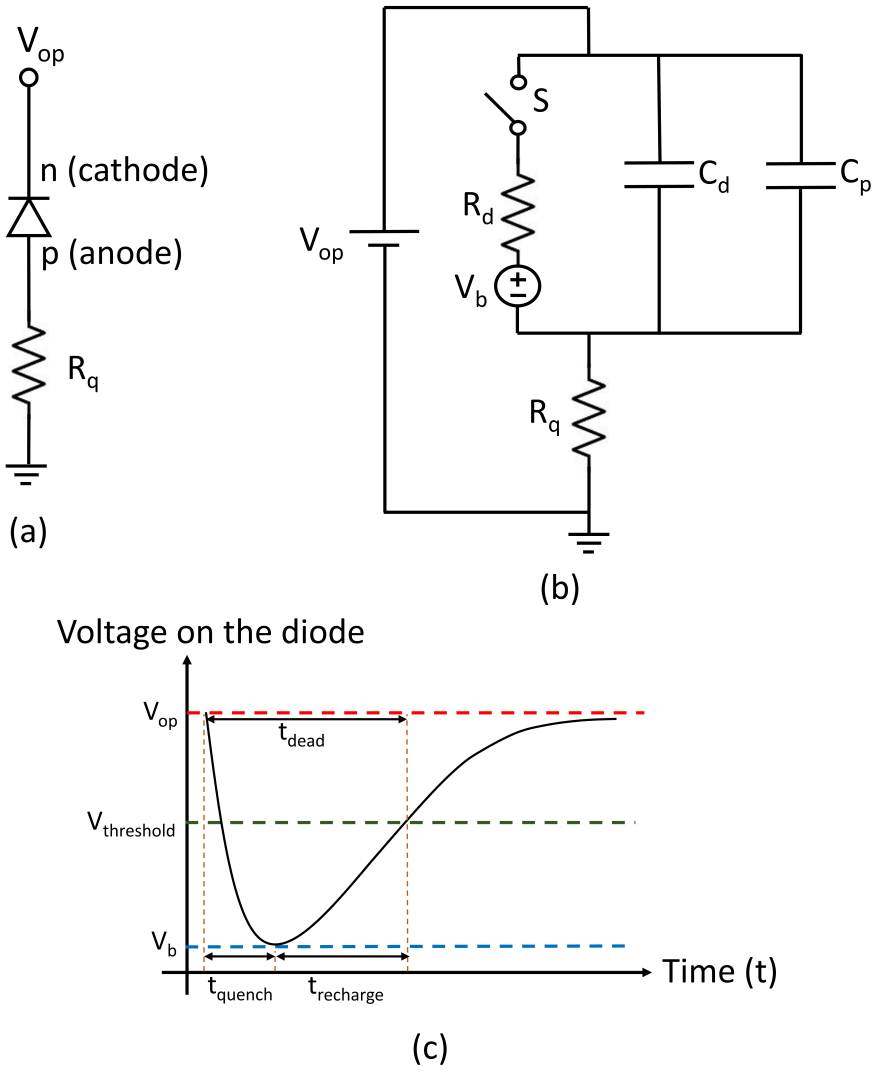


Figure 2.2: (a) A schematic of a SPAD quenched and recharged passively. (b) The equivalent circuit of a SPAD with passive quench and recharge. (c) The illustration of the quench and recharge phases in passive technique through the voltage developed on the diode over time.

to the point that none of the carriers go through the avalanche multiplication. Although it is not very well defined, it is commonly accepted that the multiplication factor drops to zero at the threshold current of 100 µA [88]. This current is also referred to as characteristic quench current (I_q) [63] or latching current as well [88]. It is declared that the avalanche process enters a self-quenching phase when I_q < 100 µA due to zero carrier multiplication. While measuring a SPAD with passive techniques, it is thereby vital that $I_{steady\text{-}state}$ is set to be below I_q to make sure that the avalanche is properly quenched. This is simply achieved by increasing R_q. For instance, the rule of thumb for selecting R_q is given as approximately 50 kΩ/V_{ex} for Si SPADs [88].

When the avalanche quenching process is completed, the switch goes back to its open state.

Then, the discharged capacitances are charged again, initially with the small $I_{steady\text{-}state}$. The recharging process is also exponential, with a time constant of $C_t \times R_q$, as the switch is open. This time constant is higher than the one in the quench phase, and it takes around 5 × time constant to fully charge the capacitances [88]. As a result, the recharging process takes much longer than the quenching process. To reduce the recharge time and the dead time of the SPAD, one needs to lower R_q. This is in contrast to the quench phase; thus, the value of R_q should be optimized such that the SPAD is quenched but also that the recharge time does not take a very long time, limiting the dead time. In addition, C_p is another restricting factor in the presented time constants, which necessitates that the parasitic effects also need to be minimized in the experimental setup. Finally, it should be noted that an avalanche can be triggered during the recharge since the SPAD stays still above V_b. While raising the voltage to V_{ex}, an avalanche multiplication can occur, but with a lower probability than in the fully charged case. If an avalanche is triggered and goes undetected during the recharge, this paralyzes the dead time of the SPAD. This effect will be further discussed later.

A summary of the quench and recharge phases with respect to time is given in Fig. 2.2 (c) by illustrating the voltage developed across the diode. As mentioned, the voltage is reduced very close to V_b at the end of the quench phase. The recharge time is defined according to a threshold voltage adjusted in the measurement equipment. The dead time of the SPAD is determined as the total time taken to fully quench and recharge the device for the next photon detection.

2.2.2 Active technique

The passive technique is easy to implement, but the disadvantage is having long and poorly controlled dead times. Therefore, more complex circuits can be designed to perform the quench and recharge operations in a shorter time and controlled manner, referred to as active quenching and/or recharge [88]. An example schematic of a SPAD quenched and recharged actively is depicted in Fig. 2.3 (a). In principle, when an avalanche is triggered, the large current flow is detected by a current-sensing circuit. The output of the current sensing circuit is connected to a controller circuit, which controls separate switches for quenching and recharging paths. The switch for quenching is closed after the avalanche generation, thus accelerating quenching by increasing the potential at the anode. Quenching ends when the voltage across the diode decreases to around V_b, and its switch opens. At this point, the controller can hold the SPAD off to precisely adjust the dead time. The hold-off period is followed by closing the switch for recharging, which grounds the anode to make the diode return to its initial biasing condition. Effectively, fast quenching and recharging are achieved by setting the anode potential properly with low resistance and capacitive effects between the diode and the anode to reduce the time constants. The timing diagram of the voltage across the SPAD is provided in Fig. 2.3 (b). As can be seen, the quench and recharge times are much shorter than the hold-off time, which allows for very good control over the total dead time of the SPAD. There are different ways of implementing fast active quenching, recharging, and

hold-off mechanisms, though this topic is beyond the scope of this work. Some examples can be found in [55], [88]–[91].

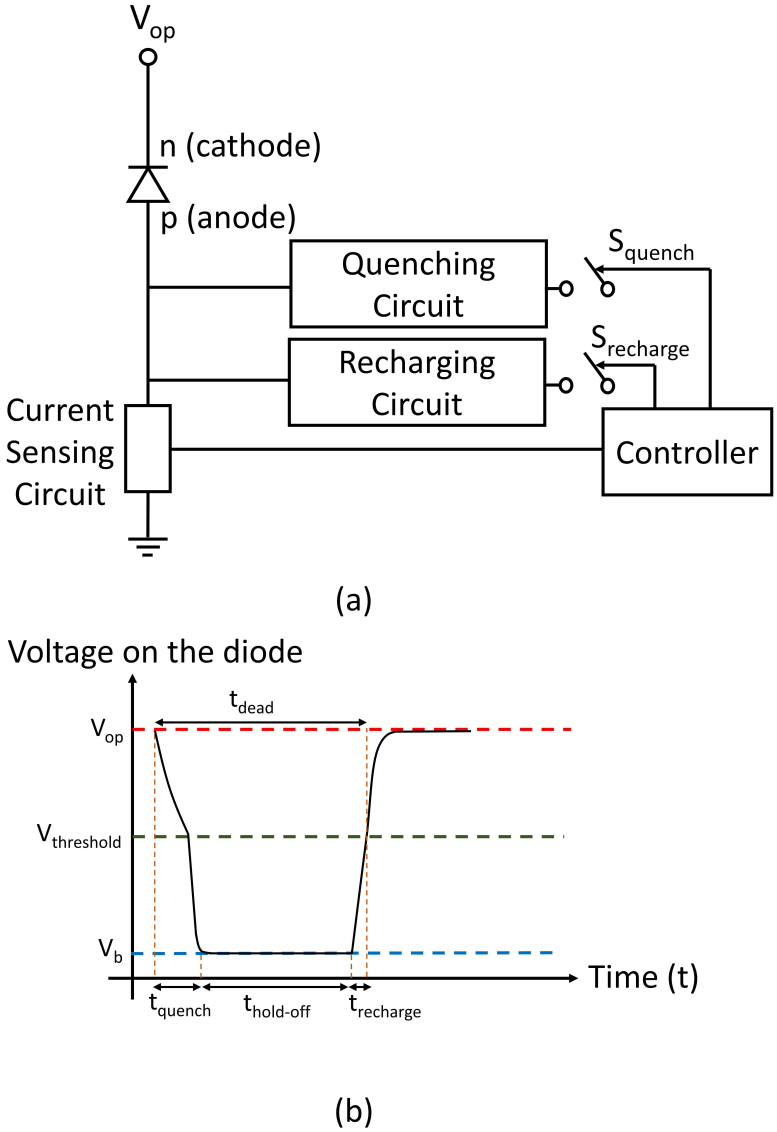


Figure 2.3: (a) A schematic of a SPAD quenched and recharged actively. (b) The illustration of the quench and recharge phases in active technique through the voltage developed on the diode in time.

2.3 Operation modes of the SPADs

The operation modes of a SPAD are going to be categorized in terms of biasing conditions. The first one is called free-running mode, where V_{op} is constantly applied at a given V_{ex}. In this case, the SPAD keeps firing in time, and detection is possible as long as the photons do not impinge during the dead time. Free-running mode is mostly utilized by Si SPADs that have

a very slight afterpulsing effect. Afterpulsing is going to be clarified later, but in short, it is a correlated noise mechanism where the trapped carriers during an avalanche can be released after a given relaxation time and create false counts.

The second operation mode is called gated mode or time-gating mode. Timing and voltage diagrams of biasing in gated mode are provided in Fig. 2.4 (a). As illustrated, V_{op} is applied such that the SPAD is periodically biased (V_{AC}) above V_b, and the photons are detected during a short and well-defined gate-on time. V_{DC} can be set to keep the SPAD below V_b for a certain gate-off time. Therefore, the frequency (f) of operation is expressed as 1/(gate-off time + gate-on time). The advantage of time-gating is that the gate-off time can be adjusted in such a way that trapped carriers are depopulated when the SPAD is off. This is a very effective technique to mitigate the afterpulsing effect. For instance, InGaAs(P)/InP-based SPADs benefit from time-gating since the trap density is generally high in InP material. Another reason is that InGaAs(P)/InP SPADs are usually cooled down to suppress the thermally generated carriers that create noise, and the trap lifetimes are extended at lower temperatures. Hence, time-gating is crucial to be able to bias the SPADs made of lower bandgap materials. To realize time gating experimentally, a circuit implementation is given in Fig. 2.4 (b). Here, a bias tee is utilized to combine the DC voltage and the AC gate pulses to constitute the final V_{op} to be applied to the cathode of the SPAD. However, on-chip time-gating circuits can also be designed.

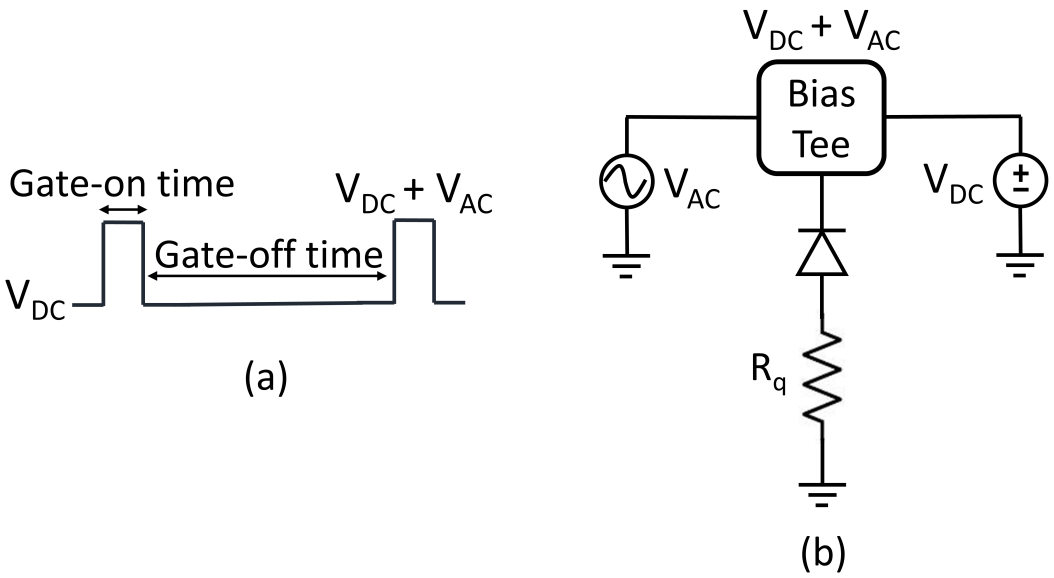


Figure 2.4: (a) Timing and voltage diagrams of gated or time-gating mode SPADs. (b) A circuit implementation of time-gating mode.

2.4 Figures of merit of the SPADs

2.4.1 Breakdown voltage

The first parameter to determine while characterizing a SPAD is the avalanche breakdown voltage. In theory, the breakdown voltage is defined at the point where the multiplication factor approaches infinity [81]. Infinite multiplication is achieved when the following condition is satisfied [50], [81]:

$$1 = \int_0^W \alpha \, dx. \tag{2.3}$$

This integral is known as the breakdown integral, where W represents the width of the depletion region and α is the ionization coefficient of the carriers, which is assumed to be the same for the electrons and holes [50]. In other words, the breakdown integral reaches unity at the breakdown voltage. In order to measure the breakdown voltage experimentally, the best way is to obtain the I-V characteristics of the diode. As indicated in Fig. 1.6 (a), the breakdown voltage can be identified as the voltage at which the current changes abruptly in an I-V curve. Therefore, a source measure unit (SMU) equipment that can apply and sweep the voltage over the SPAD and measure the output current of the diode simultaneously can be a simple and effective way to establish the breakdown voltage. Also, the avalanche breakdown voltage changes with temperature. The impact ionization coefficients worsen with the temperature rise due to the increased phonon scattering mechanisms [92], [93]. For this reason, the avalanche breakdown voltage decreases while cooling down the devices.

Another method to determine the breakdown voltage is called the light emission test. In the avalanche region, where carrier multiplication occurs, the carriers are accelerated due to very high electric fields. The acceleration of carriers can result in photon emission by hot carrier luminescence [63]. In principle, SPADs start to emit light at the avalanche breakdown voltage thanks to this hot carrier luminescence effect. Therefore, monitoring the SPADs while increasing the bias voltage can be a way of identifying their breakdown voltage. The light emission test is performed without quenching the SPAD, so that there can be a constant emission of photons. To be able to observe it, a microscope needs to focus on the SPAD area, and a camera captures emissions in the whole spectrum. Since this technique heavily relies on the microscope and camera capabilities, it becomes more of a rough method to determine the breakdown voltage.

In addition to the breakdown voltage, the light emission test is one of the few methods that can reveal spatial information about the SPAD's light-sensitive area. The images of the SPAD emission can indicate, for instance, the actual active area of the device, where multiplication occurs. The observed active area might differ from what is drawn in the layout, and it is very beneficial to quantify it to calculate the PDP correctly. Light emission tests can also show if the device suffers from premature edge breakdown. In premature edge breakdown, the

periphery of the SPAD goes into avalanche breakdown at lower voltages than the SPAD's main junction due to higher localized electric fields. High electric fields at the periphery can be an outcome of the junction curvature effect, where the implanted layers have steep edges [94]. Therefore, it is beneficial to design circular devices or to increase the radius of the curvature to achieve flatter junction edges [94], [95]. Another reason for premature edge breakdown is the high doping concentration around the periphery. This can be mitigated by designing a proper guard ring (GR) such that the peripheral junction becomes lightly doped and has a lower breakdown voltage [96]. A demonstration of light emission test results for a properly functioning SPAD and a SPAD with premature edge breakdown is given in Fig. 2.7 (a) and (b), respectively.

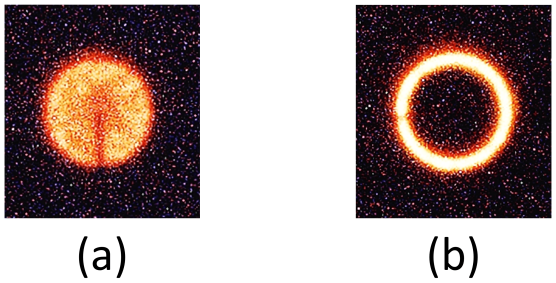


Figure 2.5: Light emission test result of a (a) properly functioning SPAD, (b) SPAD with premature edge breakdown.

2.4.2 Dark count rate

The primary noise of the SPAD is DCR, which is the mean value of the detected false counts when the SPAD is under dark conditions. The sources of the dark counts are thermally generated carriers and carriers generated through band-to-band (BTB) or trap-assisted-tunneling (TAT) mechanisms, which trigger an avalanche pulse [63], [97], [98]. Thermal generation can occur directly from the valence band to the conduction band (radiative). However, in most cases, it is dominated by the Shockley-Read-Hall (SRH) process due to the fact that the impurities or defects in the material introduce deep-level traps in the bandgap, which favor thermal generation through these states. The SRH generation is proportional to n_i/τ, where n_i is the intrinsic carrier concentration and τ is the carrier lifetime [81], [98]. n_i also strongly depends on the bandgap energy (E_g) and temperature (T) (α $e^{-E_g/T}$) [81]. As a result, with lower bandgap materials, such as InGaAs, the DCR is dominated by the SRH at room temperature. Another conclusion is that by lowering the temperature, the thermal generation and, accordingly, the DCR can be reduced as well. The carrier lifetime also plays an important role in the total SRH generation. The quality of the grown material determines the lifetimes, and deterioration of the lifetimes leads to a noisier device. On the other hand, tunneling generations occur since the valence and conduction bands are aligned at very high electric fields. This band alignment allows carriers to tunnel to the conduction band, which creates dark pulses. Tunneling generations are thereby more linked to the electric field magnitude

(E) in the depletion region [99]. Thus, thinner junctions and multiplication regions have the tendency to suffer from high DCR due to tunneling. Furthermore, the tunneling rate increases with the applied excess bias voltage, which is correlated to the electric field magnitude. However, they are almost insensitive to temperature changes. As a consequence, the DCR can be dominated by tunneling at higher excess bias and lower temperatures.

In order to measure the DCR of a free-running SPAD, it is sufficient to count the number of avalanche pulses in the dark. In the passive quench and recharge technique, counting can be performed by measuring the voltage on R_q via a high-speed oscilloscope or a pulse counter. For a chosen integration time, the DCR can then be calculated as:

$$DCR_{free\text{-}running\ mode} = \frac{Measured\ Avalanche\ Count\ in\ the\ Dark}{Integration\ Time}. \tag{2.4}$$

However, in a gated mode SPAD, the measured count rate in the dark (C_{dark}) has to be normalized by the gate-on time t_{on} and operation frequency f to account for the total time when the SPAD is ON. The normalized DCR in the gated mode is found as [100], [101]:

$$DCR_{gated\ mode} = \frac{C_{dark}}{f \times t_{on}}. \tag{2.5}$$

In this thesis, all Si SPADs were characterized in free-running mode, whereas InGaAs(P)/InP SPADs were measured in time-gating mode, and their count rates were normalized by the gate-on time.

In addition, while performing any counting measurement, two important factors need to be paid attention to: the integration time and dead time of the SPAD. First, it is vital to keep the integration as high as possible to measure the count rate accurately. Regarding the DCR, the occurrence of dark pulses follows a Poisson distribution [50]. In the case of low integration time, the measured counts fluctuate significantly, whose standard deviation is equal to the square root of the average number of dark pulses. Therefore, by increasing the integration time, the mean number of dark events will increase, and the effect of the fluctuations will be mitigated. Second, SPADs cannot trigger an avalanche during the dead time. This limits the maximum achievable count rate to 1/dead time in the measured count rate. There are two models to estimate the effect of dead time on the measured count rate: the paralyzable and non-paralyzable models [89], [102]. The paralyzable dead time model is associated with passive quenching and recharge techniques, where the dead time of the SPAD is not clearly defined by the circuits. In this configuration, the SPAD can still trigger an avalanche before reaching the threshold voltage, and these avalanche pulses can go undetected as they remain below the threshold voltage for the pulse reading. The demonstration of the extended dead time effect is provided in Fig. 2.6 (a). In the paralyzable model, the undetected pulses are assumed to prolong the dead time, and the relation between the measured count rate (M) and real count rate (N) is indicated as [89], [102]:

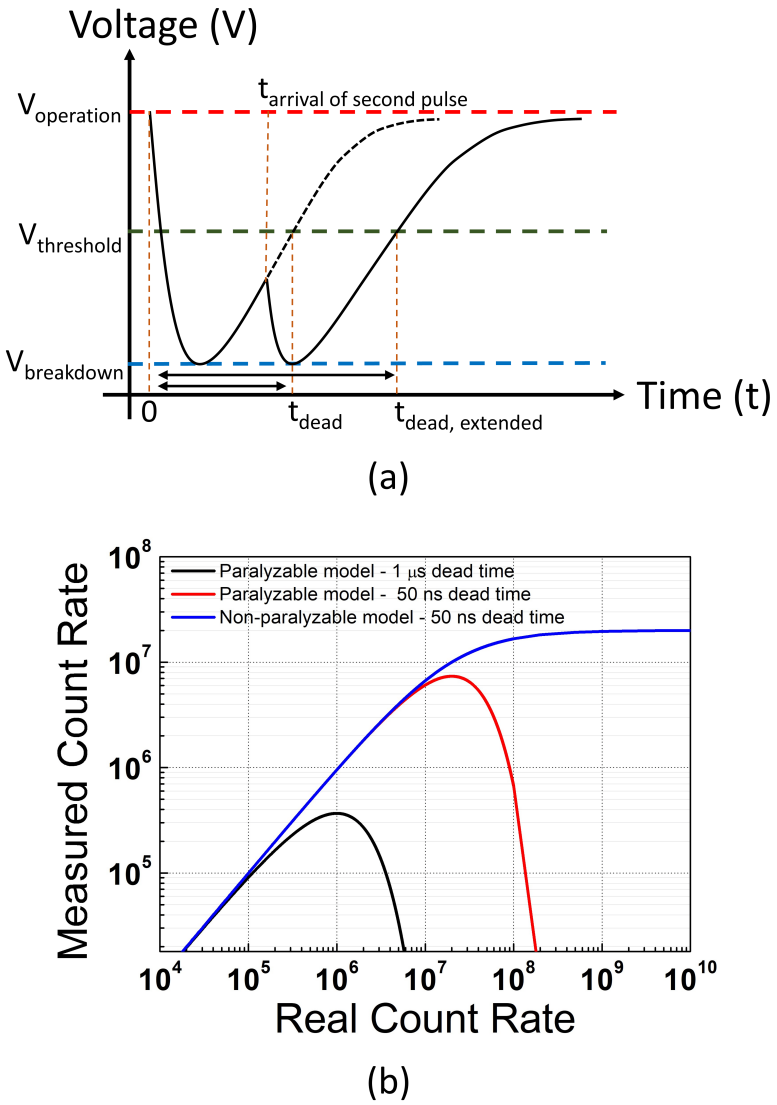


Figure 2.6: (a) Representation of the extended dead time effect modeled by the paralyzable dead time model. (b) The relation between the measured and real count rate according to the paralyzable and non-paralyzable dead time models under different count rates.

$$M = N \times e^{-N.t_{dead}}. \tag{2.6}$$

On the other hand, the non-paralyzable model is suitable for the active quenching and recharge technique, where the dead time of the SPAD is well controlled and the recharging process brings the SPAD rapidly back to the adjusted excess bias voltage as shown in Fig. 2.3 (b). Therefore, these devices do not suffer from the extended dead time effect and can reach the maximum detectable number of avalanche pulses limited by the dead time. The relation between the real and measured count rate in the non-paralyzable model is defined as

[89], [102]:

$$M = \frac{N}{1 + N \times t_{dead}}. \tag{2.7}$$

Fig. 2.6 (b) shows the comparison of the dead time models under various count rates. As depicted, the non-paralyzable model with a 50 ns dead time (active quenching and recharge) provides very accurate counting at low count rates and saturates at 1/dead time at high count rates, as expected. However, the measured count rate with the paralyzable model (passive quenching and recharge) deviates from the actual count rate at high values, even with a 50 ns dead time. The deteriorated dead times in the paralyzable model have severe effects on the measurement, as represented by a 1 µs dead time. Thus, the influence of the dead time while performing avalanche pulse counting has to be kept in mind to report accurate results.

2.4.3 Afterpulsing probability

In addition to the primary dark pulses, secondary pulses, or afterpulses, can contribute to the internal noise of a SPAD. In principle, when an avalanche occurs, large numbers of carriers are generated in the depletion region, which can get trapped in the defects of the material with different deep-energy levels. After their relaxation, if the trapped carriers are depopulated during the dead time of the SPAD, they flow through the device without any consequences. On the other hand, if carrier detrapping occurs when the SPAD is armed, it can trigger additional undesired pulses, which are called afterpulses. Afterpulsing is crucial in many time-sensitive experiments since correlated noise can adversely affect the DCR and PDP of SPADs. Also, the effects of afterpulsing are more severe at lower temperatures because the lifetimes of the trapped carriers elongate, and the probability of having a secondary pulse when the SPAD is armed increases.

There are several ways to mitigate this effect. First of all, by simply improving the material quality and decreasing the number of defects, afterpulsing can be reduced. Secondly, by extending the dead time or the gate-off time of the SPADs, the depopulation of all of the trapped carriers can be guaranteed when the SPAD is not armed [63], [97]. Although this is a very effective solution to overcome afterpulsing, it also degrades the maximum achievable counting rate of the device. Since some applications ask for higher count rates and operation frequencies, this method might not always be practical. Finally, by limiting the total amount of charge per avalanche pulse, the afterpulsing effect can be mitigated. As the number of trapped carriers is proportional to the total number of carriers passing through the depletion region, with less charge flow, the trapped number of carriers decreases. For instance, in gated mode SPADs, by shrinking the gate-on times to sub-nanoseconds, the total charge flowing and afterpulsing can be reduced [103], [104]. An improved version of this idea is biasing the devices with sinusoidal waves, which can provide gate widths in the range of ps. With sinusoidal gating, operating frequencies up to several gigahertz (GHz) with very low APP have

been demonstrated in InGaAs/InP SPADs [45], [105]. Another way to limit the number of carriers per avalanche pulse is to quench the SPADs faster [90], [91], [103]. In free-running mode, a comparator with a low detection threshold to sense the avalanche event, combined with a fast active quenching circuit, can diminish the total amount of flowing charge and thus APP. Therefore, with the passive quench technique, the APP could be overestimated. In gated mode SPAD, decreasing the fall times of the very short gates can rapidly quench the avalanche pulse. An advantage of reducing the charge flow is that reaching higher count rates is still feasible.

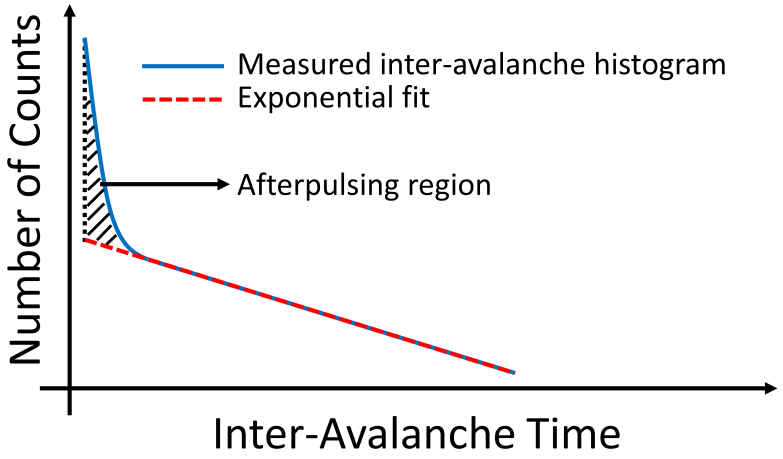


Figure 2.7: An example measured afterpulsing histogram with the exponential fit representing the case when there is no afterpulsing.

To estimate afterpulsing probability, the common method is to construct the inter-arrival time histogram, where time intervals are measured between subsequent avalanche pulses generated by the SPAD. The technique was first demonstrated as time-correlated carrier counting (TCCC) in [106]. In the absence of secondary pulses, the histogram follows a pure Poisson distribution. With afterpulsing, the inter-arrival time histogram deviates from pure exponential behavior due to the detrapping time of different defect centers. An example afterpulsing histogram is illustrated in Fig. 2.7. To quantify the APP of a SPAD, following the measurement of the histogram, a Poisson fit is applied for the time interval where all the trapped carriers are depopulated, and the histogram is based on only the primary dark counts. Since trap lifetimes can change from ns to a few µs in Si at room temperature, it would be safe to perform a fitting after 15 µs [107]. Regarding InGaAs/InP SPADs, it would be better to first sweep the hold-off time and measure DCR, as the trap lifetimes are highly dependent on the temperature. The resulting graph can show at which frequency afterpulsing starts. Accordingly, the fitting procedure can be performed at safe time intervals. After the fitting procedure, the APP can be calculated with the following formula:

$$APP = \frac{Secondary\ Pulse\ Count}{Total\ Avalanche\ Count}, \tag{2.8}$$

where the secondary pulse count corresponds to the difference between the measured curve and the fit curve, as shown as the black area in Fig. 2.7. While reporting the APP, the excess bias voltage of the device should also be provided, as the number of charges per avalanche pulse increases with voltage, which increases the APP.

2.4.4 Photon detection probability

PDP indicates how many of the incident photons can be detected by the detector. In principle, PDP is a product of internal quantum efficiency (QE) (η_{abs}), injection efficiency (η_{inj}) of the photogenerated carriers into the multiplication region, and the avalanche breakdown probability of these injected electron-hole pairs (P_j) as shown in the equation below [100]:

$$PDP = QE\ (\eta_{\text{abs}}) \times \eta_{\text{inj}} \times P_j. \tag{2.9}$$

QE depends on the absorber thickness, as the number of absorbed photons changes exponentially with the depth, following the Beer-Lambert law. Therefore, the absorber should be thick enough to absorb a sufficient amount of photons to obtain high PDP values. The injection efficiency is determined by drift and diffusion mechanisms. As in the case of InGaAs/InP SPADs, where the absorption and the multiplication volumes are separated, it is important to completely deplete the absorber to be able to transport the carriers efficiently into the multiplication zone by drifting them. Regarding Si SPADs, where the absorption and the multiplication occur in the same Si layer, wider depletion regions help to transfer the photogenerated carriers to the multiplication region. Thus, it is crucial to engineer the SPAD structures in terms of electric field profiles in order to optimize their PDP. The avalanche breakdown probability, on the other hand, is linked to the electric field magnitude in the multiplication region through the impact ionization coefficients, as explained previously. While increasing the excess bias voltage, the PDP thereby increases thanks to the enhanced impact ionization coefficients and the enhanced breakdown probabilities, until the point where these coefficients saturate. To sum up, the thickness of the absorber and the depletion region should be carefully designed together with the electric field profiles to maximize PDP.

In order to measure PDP as a function of the light absorption spectrum, a setup similar to the one in Fig. 2.8 (a) is utilized. In this configuration, a Xenon lamp ensures a broad-band spectrum for incident photons. The Xenon lamp light is then coupled to a monochromator, which provides high wavelength selectivity to select a single wavelength via diffraction gratings. The incident power and the number of photons are adjusted with a slit, which has a controllable aperture size with a micrometer for the attenuation. After that, an integrating sphere provides spatially uniform light at its output ports. A calibrated reference detector, which can be made of silicon or InGaAs depending on the wavelength of interest, is used to accurately estimate the impinging photon count rate on the SPAD. Reference detectors are usually conventional photodiodes, which measure the current under the given illumination.

As their responsivity (A/Watts) is a known parameter, measured current can be converted into photon power. The energy of one photon is also calculated by the Planck constant (h) × c/wavelength(λ). Dividing photon power by the energy of a single photon and normalizing it with the area of the reference detector results in the photon count rate per unit area, from which the photon count rate reaching the SPAD is computed by multiplying it with the active drawn area of the SPAD. Eventually, to obtain the PDP value of a free-running SPAD, the following equation is used:

$$PDP_{\text{free-running mode}} = \frac{Measured\ Count\ Rate - DCR}{Photon\ Count\ Rate}. \tag{2.10}$$

If the APP of the SPAD is known, the measured count rate can be further corrected by dividing it by (1+APP) and using this compensated value as the new measured count rate in Eq. 2.10. The most important thing to pay attention to while measuring PDP is to perform the experiment under low-light conditions in order not to saturate the count rate of the SPAD. As mentioned, the maximum count rate that can be achieved is limited by the dead time, so the photon count rate should be kept well below this value. If passive quenching/recharge is utilized, it should be made sure that the correct count rate is inserted in the PDP calculation. As explained in the DCR section, the measured value could deteriorate compared to the real count rate depending on the dead time in the paralyzable model. Moreover, the integration time should be set as high as possible. Similar to the DCR, the arrival times of the photons from an illumination source follow a Poisson distribution. In the case of high optical power, the variation in the number of photons impinging per unit time is very small. However, in the single-photon regime, the relative fluctuations in the number of detected photons will be important. These fluctuations are called shot noise. To alleviate the effect of shot noise on photon counting, the integration time should be kept high.

Regarding the SPADs that are operated in gated mode, the formulation changes. As the number of photons impinging in a particular time interval per gate pulse follows a Poisson distribution, the PDP calculation becomes:

$$PDP_{\text{gated mode}} = \frac{1}{\mu} ln \frac{1 - C_{\text{dark}}/f}{1 - C_{\text{total}}/f}, \tag{2.11}$$

where μ is the mean number of photons per gate pulse, and C_{total} is the total count rate measured under illumination [100], [101]. In addition, when a pulsed laser is used instead of a continuous light source, μ represents the mean number of photons per optical pulse. To accurately measure the PDP with a pulsed laser, the laser pulses should be synchronized with the gate pulses of the SPAD via the waveform generator, and then the optical pulses should be finely aligned with the gate pulses to make sure that every photon reaches the SPAD during this gate-on time. The optical setup utilized in this work for a PDP measurement with a pulsed fiber laser is demonstrated in Fig. 2.8 (b). The synchronized laser pulses are fed to a beam

splitter. 99% port of the beam splitter is directed to a calibrated power meter to calculate the number of photons per pulse at the 1% port. The power can be further attenuated with a calibrated attenuator to go into a photon-starving regime. Then, the laser beam is coupled to a confocal microscope with the fiber and is focused on the SPAD through a long-distance objective. Although the PDP measurement with a focused pulsed laser is more challenging than the continuous light source due to the optical setup construction, it yields more accurate results by removing the ambiguity about the size of the SPAD active area since the focused beam is smaller than the active area. In the case of using an integrating sphere with a Xenon lamp, the photons can get absorbed outside the SPAD area, and the photogenerated carriers can diffuse to the active region and increase the PDP by creating undesired pulses.

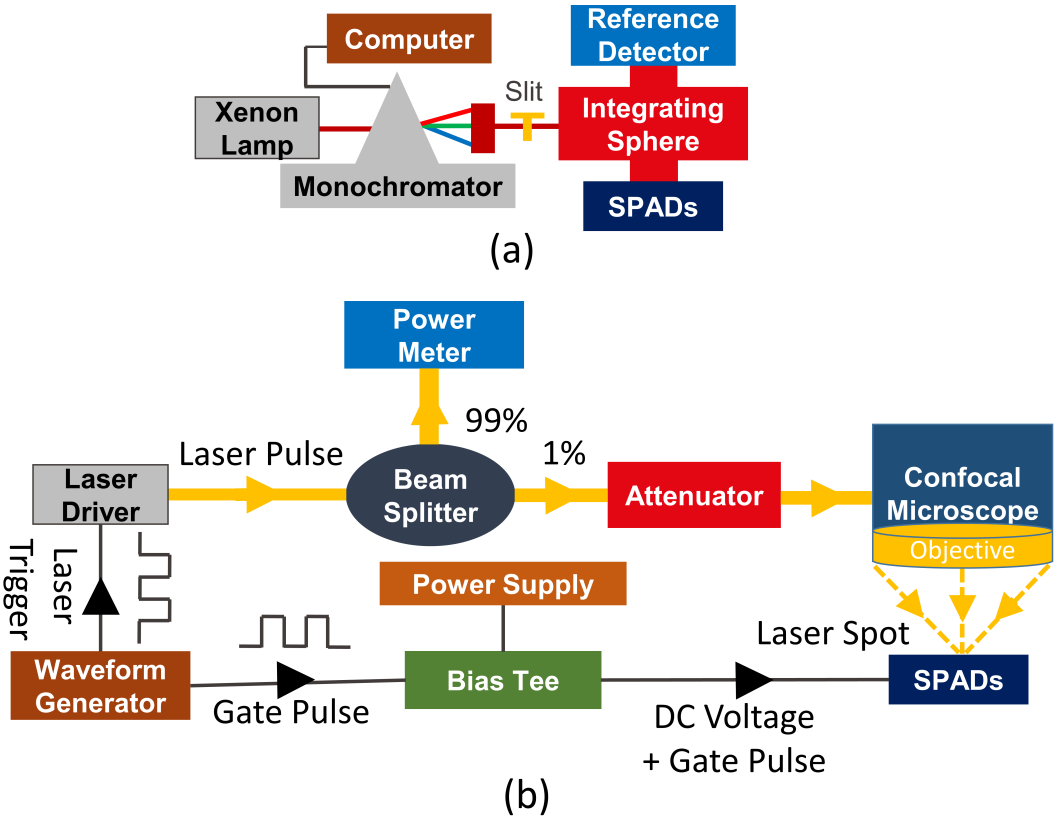


Figure 2.8: (a) The setup to characterize the PDP of the SPADs with a continuous light source. (b) The optical setup to measure the PDP of the SPADs with a focused pulsed laser beam.

2.4.5 Timing jitter

Timing resolution in a SPAD is characterized in terms of the temporal fluctuations between the time of arrival of photons and the photon detection time by the end-circuitry, also known as timing jitter. A typical jitter curve is composed of a Gaussian distribution around the peak of the detection and an exponential tail, as illustrated in Fig. 2.9 (a). Regarding the peak width of the jitter curve, there can be several factors contributing to it. It has been

shown that the timing jitter is inherently limited by the spatial randomness of the photon absorption process [87]. It means that there are going to be time differences in the transit times of the photogenerated carriers to the multiplication region due to the different locations of the absorption. However, timing jitter is usually dominated by the avalanche build-up and spreading dynamics. In the former, the randomness of the impact ionization process creates differences in the local growth of the first filament of current. The latter is caused by the randomness in the lateral spread of the avalanche from this localized seed point to the remaining high electric field regions via the multiplication-assisted diffusion process [87]. Photon-assisted spreading mechanisms also exist, where secondary photons emitted from the hot carriers can be reabsorbed and form multiple seed points in the junction. The contribution of the photon-assisted spreading has been shown to be negligible compared to the multiplication-assisted diffusion, whereas it might have an impact on timing jitter when the active volume of the device is larger [87]. Concerning the exponential tail, it stems from the diffusion of the photogenerated carriers from the neutral region to the depletion region [50]. The exponential tail widens the width of the timing histogram, and it can degrade the timing jitter of the SPAD depending on its time constant. Very slow and long tails might limit SPAD utilization in timing-critical applications that rely on timestamping from a fixed voltage threshold. Besides, since the absorption coefficient is wavelength-dependent, the shape of the diffusion tail changes with the incident wavelength as well. For instance, NIR photons are absorbed at relatively longer depths in Si SPADs; thus, a diffusion tail is expected in the timing histograms obtained at NIR wavelengths.

In order to measure timing jitter, TCSPC is generally utilized. A typical timing jitter measurement setup is shown in Fig. 2.9 (b). In this configuration, fiber-coupled monochromatic pulsed lasers are used as an illumination source. Then, uniform illumination is ensured through a collimator. The source power is attenuated to the single-photon regime thanks to absorptive neutral density filters. The resulting illumination is then directed to the SPAD through mirrors. It is important to tilt the stage of the SPAD so that there won't be any back-and-forth reflections between the SPAD and the mirror, which can cause multiple peaks in the histogram and degrade jitter. Afterwards, a high-speed digital oscilloscope can measure the time difference between the laser clock signal and the rising edge of avalanche pulses. The repeated timing measurements constitute the resulting timing histogram, whose FWHM is defined as the timing jitter of a SPAD. As depicted in Fig. 2.9 (a), to obtain the FWHM of the histogram, the time difference between two instants where the peak amplitude drops to its half is taken. In fact, there are other sources of jitter, such as laser jitter and jitter coming from electronics. By assuming the statistical independence of all these sources, the measured jitter can be formulated as:

$$FWHM_{measured}^{2} = FWHM_{SPAD}^{2} + FWHM_{electronics}^{2} + FWHM_{laser}^{2}. \tag{2.12}$$

Therefore, to obtain the SPAD jitter value, the laser and electronics jitter can be deconvolved from the measured value, if they are known. While reporting the jitter value, the excess bias

voltage of the device should also be stated, along with the laser wavelength. The excess bias voltage improves timing uncertainty by shortening the avalanche build-up time. Finally, it is crucial to synchronize the gate pulses and the laser pulses through a waveform generator when the SPAD is biased in gated mode, likewise in PDP measurements.

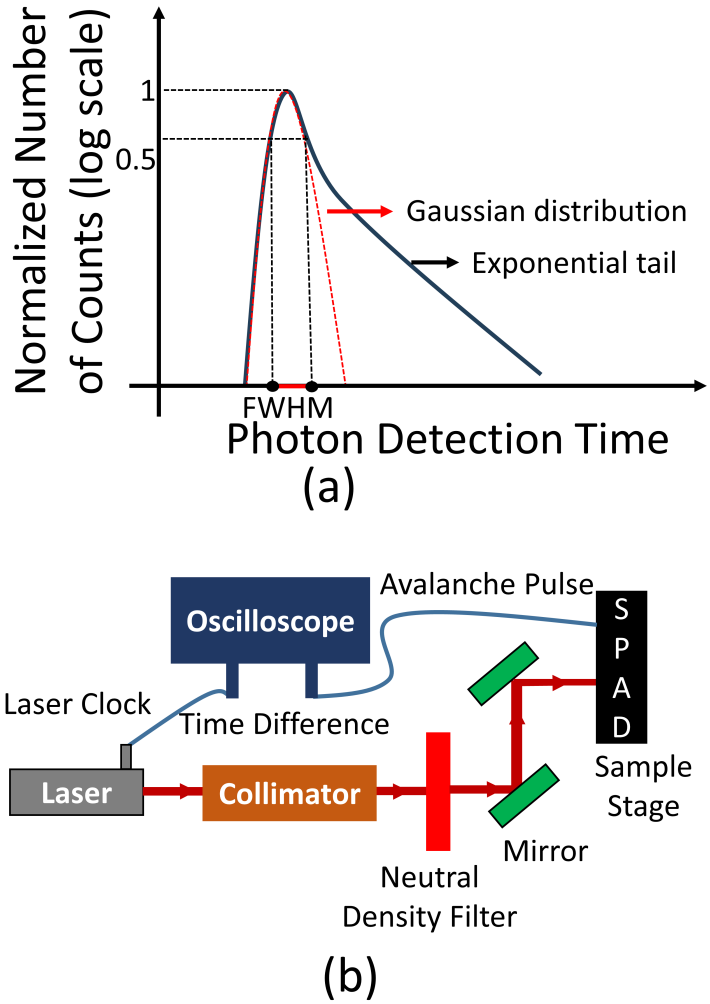


Figure 2.9: (a) A representative timing jitter histogram and its components. (b) A timing jitter measurement setup based on TCPSC technique.

3 NIR/SWIR SPAD Case Analysis

In order to obtain SPADs operating at NIR or SWIR wavelengths, there are different material systems, as indicated in Table 1.1 in Chapter 1. The operating temperature, achievability of pixel arrays, and integration with ROICs can be counted as the main criteria for being able to utilize these SPADs in a LiDAR, QKD, or NIROT system. In this sense, Si CMOS and InGaAs(P)/InP-based SPADs are promising candidates. As summarized in Chapter 1, Si SPADs can function at room temperature with low noise, constitute camera arrays, and allow the device to combine with circuits on the same CMOS chip. On the downside, the detection efficiencies in NIR drop due to reduced absorption. InGaAs(P)/InP SPADs give feasibility to building pixel arrays but require moderate cooling, which is actually attainable with a thermo-electric cooler, and 3D integration with ROICs since they are fabricated on InP substrates, not on a Si CMOS wafer. Therefore, these two material systems are going to be analyzed and investigated in this thesis to realize better-performing pixels and optimize pixel performance for NIR/SWIR detection. As shown in Table 1.1, Ge-on-Si SPADs are another alternative, but the requirement of cryogenic cooling (< 150 K) excludes them from use in the mentioned applications for the moment.

3.1 NIR Enhanced Si CMOS SPADs

Si, or more specifically CMOS nodes, are the most cost-efficient platforms for developing SPADs for NIR wavelengths, allowing the design and integration of devices and circuits on the same chip. However, the reduced absorption coefficient of Si in NIR prevents the devices to achieve high detection efficiencies at longer wavelengths. Besides, the bandgap of Si restricts the cutoff wavelength of detection to around 1.1 μm, thereby SWIR detection is effectively not possible with Si SPADs. The absorption coefficient (α) of Si with respect to wavelength is given in Fig. 3.1[108]. The figure indicates the drastic degradation in absorption of NIR wavelengths, such that the absorption coefficient drops 100× at 800 nm and 180× at 850 nm compared to the coefficient at 400 nm. The corresponding absorption depth ($1/\alpha$) is also depicted in the same figure in blue. According to Beer-Lambert law, which is formulated as:

$$I = I_0 e^{-\alpha x}, \tag{3.1}$$

where the incident light intensity (I_0) exponentially decreases inside the material with travel distance due to absorption [109]. At any wavelength and the corresponding absorption depth, the output intensity (I) reduces to 1/e of the input intensity; in other words, 64% of the incident light gets absorbed in the material at the absorption depth. Therefore, for Si, 64% of the light is absorbed when it travels 11.8 µm, 18.7 µm and 74.6 µm distances in the material at the wavelengths of 800 nm, 850 nm, and 960 nm, respectively. Following this analysis, up to date, there are four main ways of covering such long distances and enhancing the NIR efficiency of Si SPADs designed in any CMOS technology node, which are going to be discussed next.

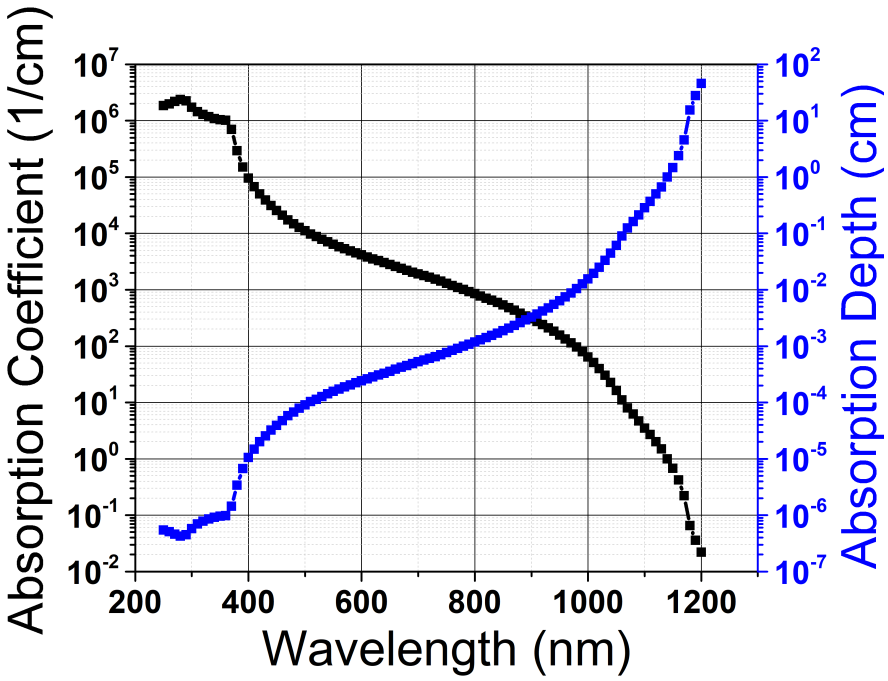


Figure 3.1: Absorption coefficients and depths of Si with respect to wavelengths.

3.1.1 Wide depletion region SPADs

As mentioned, the absorption depths of NIR light in Si are quite large due to the low absorption coefficients. Therefore, in most cases, the generated carriers deep inside the material need to diffuse to the depletion region to trigger an avalanche. However, the diffusion lengths of electrons and holes, for example, are in the order of 10-20 µm for the doping concentration of 10^{18} cm^{-3} [111], [112]. Hence, the depletion region, also called the active volume, where the high electric field is confined, can be extended to collect more photogenerated carriers thanks to the drift mechanism. By enlarging the width of the depletion region, more carriers can reach the multiplication region, thus increasing the NIR PDP of the devices. The depletion width (W) of a p-n junction diode is given as:

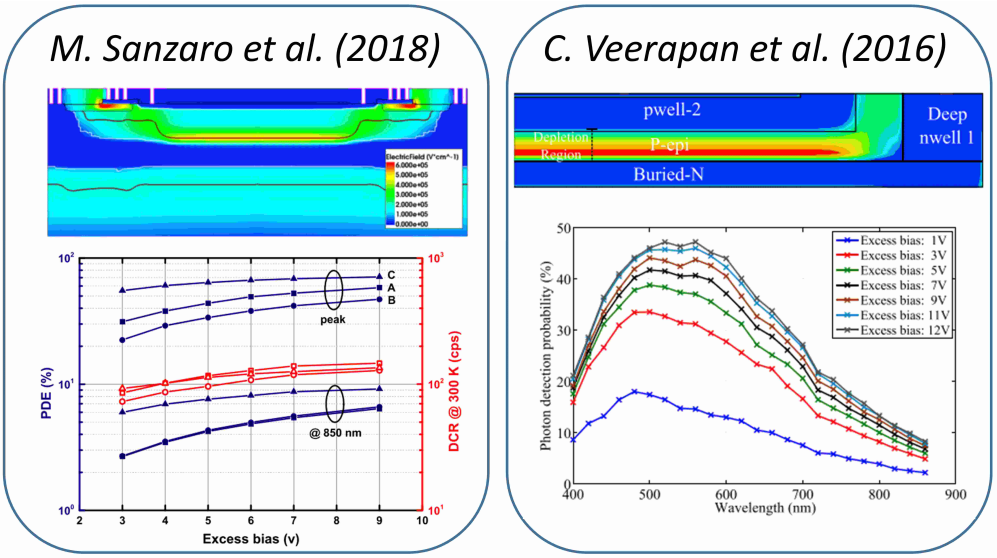


Figure 3.2: Wide depletion region SPADs demonstrated in the literature [51], [110].

$$W = \sqrt{\frac{2\epsilon}{q}\left(\frac{1}{N_A} + \frac{1}{N_D}\right)V}, \tag{3.2}$$

where ϵ, N_D and N_A correspond to the material permittivity, ionized donor, and ionized acceptor densities, respectively. It should be noted that this formula is valid only for constant doping concentrations. As displayed in Eq. 3.2, the methods to extend the depletion region width are to decrease doping concentrations and to increase the reverse bias voltage on the diode. In the former, doping densities are defined by the foundries, and they are mostly adjusted to develop electronics instead of optical devices. Therefore, achieving thicker depletion regions is not always straightforward with the available doping profiles for designers. Nevertheless, lower-doped layers should be preferred for a wider depletion region. In the latter, the maximum applied voltage is limited by the noise of the device; thus, it cannot be freely increased.

Two examples of state-of-the-art wide depletion region (≥ 1 µm) SPADs are illustrated in Fig. 3.2 in terms of electric field simulations and PDP spectrums [51], [110]. PDP curves were obtained with front-side-illumination (FSI). As can be observed, very high excess bias is required (~10 V) to increase the avalanche breakdown probability and detection efficiency in the wide depletion region approach [97], [113]. This is because the electric field stays lower at the breakdown voltage, and it rises slowly with increasing excess bias due to a thick depletion zone. As the magnitude of the electric field determines the electron-hole ionization coefficients and the probability of an avalanche, high excess bias is a requirement to enhance PDP in this approach. Also, the high excess bias need brings up the issue of voltage attenuation so that the SPAD's output voltage can be adapted to the digital circuit's low voltages. Furthermore, tunneling noise is inherently suppressed with this approach thanks to the reduced electric field magnitude. SPADs with wide depletion regions might be noisier due to augmented

SRH generation. The reason is that the probability of carriers being captured increases while traveling through a larger volume, which has a higher number of defect centers.

3.1.2 Back-side-illumination

In the back-side-illumination (BSI) approach, the SPADs are illuminated from the substrate side instead of the SPAD's front. In order to detect photogenerated carriers, the substrate should be thinned down so that the photons are absorbed closer to the device junction. There are several advantages of this approach for NIR sensitivity enhancement. First of all, with the BSI approach, SPADs and circuits can be developed independently on separate chips, and then these two different chips can be integrated with bump-based or wafer bonding. This allows designers to utilize different technologies on the bottom and top tiers and more optimized SPAD devices in terms of PDP in the NIR. A cross-section of this process, building a 3D stacked SPAD sensor with the SPAD on the top tier, is illustrated in Fig. 3.3 (a) [114]. Secondly, in FSI mode, the light can be reflected back from the metal layers and interconnects used to fabricate the devices, which causes a reduction in PDP. However, the light has an undisturbed path to reach the device junction in the BSI approach, preventing PDP lossses, as can also be seen in Fig. 3.3 (a). Although substrate thinning and 3D bonding procedures are costly, BSI SPADs have recently been investigated more and more to build SPAD arrays [15], [26], [114]. A 189×600 SPAD array fabricated in 90 nm and stacked with a circuitry designed in 40 nm CMOS technologies for automative LiDAR application by Sony is depicted in Fig. 3.3 (b).

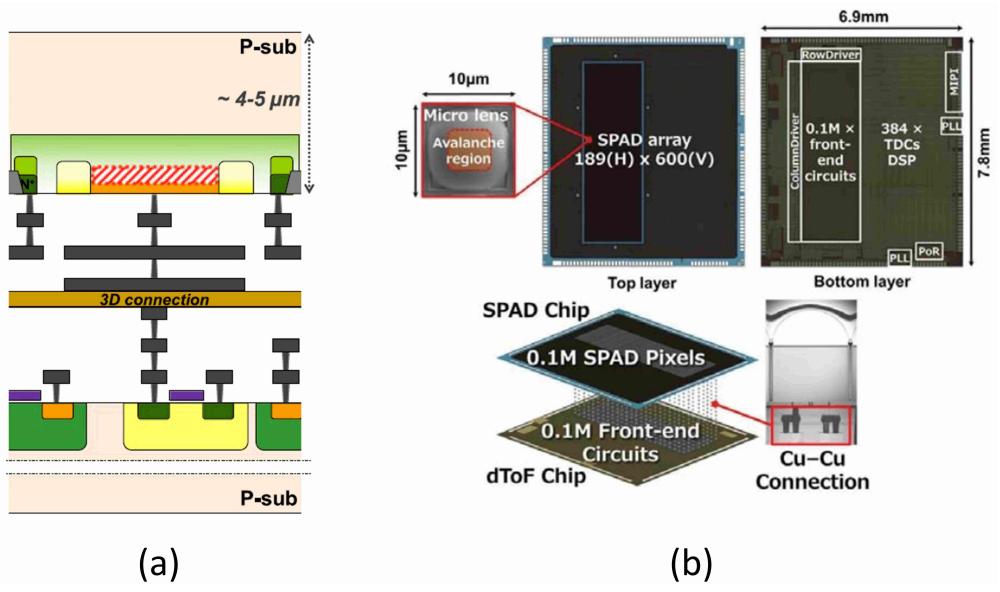


Figure 3.3: (a) Overview of the BSI approach in a cross-section [114]. (b) BSI-stacked large-format SPAD sensor with copper-to-copper bonding [26].

3.1.3 Electrical micro-lensing or charge-focusing technique

Another alternative to enhance NIR efficiency is to implement the technique called electrical micro-lensing or charge-focusing. In conventional SPADs, there exists only vertical electric field lines that sweep photogenerated carriers and collect them. Therefore, the photon-sensitive area is restricted by the horizontal width of the depletion or multiplication region. As the photon absorption also occurs outside of this active region, photogenerated carriers in those areas go undetected. In electrical micro-lensing, the electric field of the device is engineered in such a way that it has horizontal electric field lines that transit the carriers to the active region in the lateral direction as well. The schematic of the technique is provided in Fig. 3.4, which shows how the carriers generated at the peripheries are directed into a small multiplication region [115]. The electric field simulation of this device is also given in the same figure. The technique was first demonstrated as a micro-lensing effect, where the lateral transit of the carriers was achieved under a GR region [47]. Then, it was implemented as charge-focusing in the large-format and megapixel SPAD arrays by Canon [52], [115], [116].

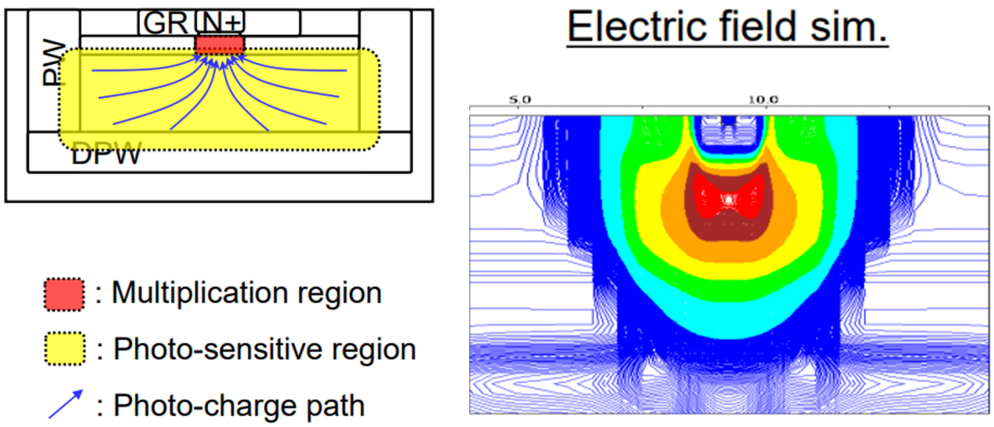


Figure 3.4: Electrical micro-lensing technique, also known as charge focusing, explained with a cross-section and electrical field simulations [115], [116].

3.1.4 Light-trapping structures

To boost the efficiency of specific wavelengths in NIR, implementing nano-structured light-trapping structures can be another method. In conventional devices, the light travels only vertically towards the deep inside of the substrate without any horizontal component. Since the absorption depth of NIR wavelengths is long, not all of the carriers can diffuse to the depletion region and undergo impact ionization. With the light-trapping structures, the vertical incident photons can be diffracted horizontally to travel, as if they are in a waveguide, and the light can be trapped between air-Si interfaces. This horizontal travel enhances the absorption significantly, given that the active diameter of the devices are larger than their depletion width. For instance, the light-trapping nanostructures can be formed by etching the SPAD surface as an array of inverse pyramids, as demonstrated in the cross-section and

scanning electron microscope (SEM) images in Fig. 3.5 (a) [117]. The period of the array can be adjusted to favor the diffraction of the target wavelength. Simulations of light absorption at 850 nm in a conventional SPAD and a SPAD with inverse pyramid-shaped light-trapping structures are illustrated in Fig. 3.5 (b). With this approach, the PDP can be enhanced from 13% to 32% at 850 nm.

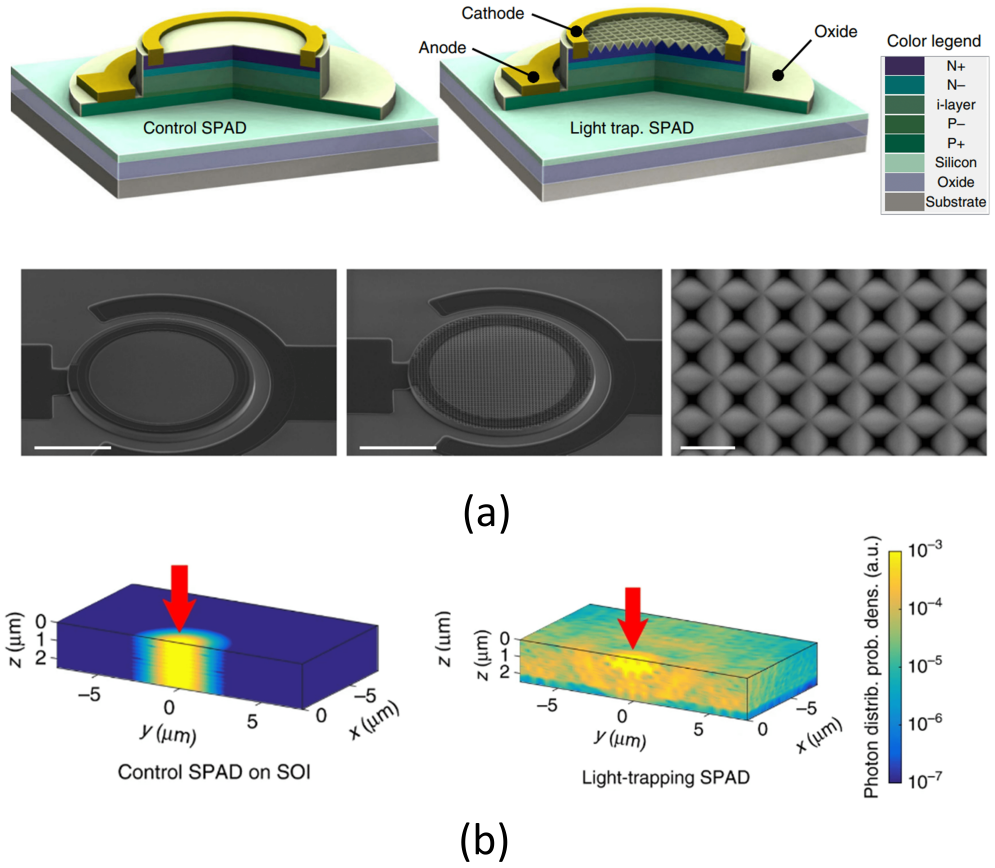


Figure 3.5: (a) The cross-section and SEM images of a conventional SPAD and a SPAD with inverse pyramid-shaped light-trapping structures [117]. (b) Photon distribution probability of a conventional SPAD and a SPAD with light-trapping nanostructures [117].

3.2 NIR/SWIR InGaAs(P)/InP SPADs

Since the absorption of Si is poor in NIR and the cutoff wavelength is limited up to 1.1 µm due to its bandgap (1.12 eV), alternative InGaAs(P)/InP-based SPADs have become the promising candidates for the spectrum going from 1.06 to 1.55 µm wavelengths, especially for LiDAR and QKD applications. In a SPAD utilizing this material system, the photons are absorbed in InGaAs(P) material. InGaAs is a compound semiconductor that is a combination of two other direct bandgap semiconductors, which are InAs and GaAs. Therefore, by changing the mole fraction of the group III materials, the bandgap of $In_xGa_{1-x}As$ can be tuned. The most special

mole fraction occurs for the $In_{0.53}Ga_{0.47}As$ case, where InGaAs becomes lattice matched to the commercially available InP substrates. This allows for the growth of InGaAs on InP without creating any strain, thus enabling the realization of high-performance heterojunction devices. For $In_{0.53}Ga_{0.47}As$ case, the cutoff wavelength of absorption is around 1.68 µm, corresponding to a 0.74 eV bandgap. In the rest of this thesis, InGaAs will be referring to this alloy. Due to the low bandgap of InGaAs, thermal generation increases, which deteriorates the DCR of the SPADs. Hence, moderate cooling up to 225 K can be required to lower the noise. In addition, the thermal noise can be mitigated by increasing the bandgap of InGaAs while sacrificing the cutoff wavelength. This can be achieved by incorporating Phosphorus (P) and tuning the mole fractions of group V materials ($InGaAs_yP_{1-y}$). The formula to estimate the bandgap (E_g) of $InGaAs_yP_{1-y}$ is [118]:

$$E_g = 1.344 - 0.738y + 0.138y^2. \tag{3.3}$$

Therefore, by selecting a proper y in Eq. 3.3, the detection wavelength can be restricted to 1.1 µm, in exchange for reducing the DCR. On the other hand, the disadvantages of InGaAs(P)/InP can be listed as follows: (a) InP substrates are more expensive than the Si substrates; (b) the integration of the SPADs with Si ROICs requires sophisticated techniques, which complicates the process and increases the total cost; (c) the material quality and trap densities are not as mature as in Si, which increases the effect of afterpulsing.

Regarding the afterpulsing effect, high APP in InGaAs(P) SPADs forces these devices to be biased in time-gating mode. The majority of the reported devices were characterized in time-gating mode [61], [63], [64], [91], [98], [100], [103], [104], [119]–[121], and this mode was also used in this thesis extensively thanks to an easier accessibility and the lack of very fast quenching circuits. However, time-gating mode, also known as synchronous detection, might not always be possible. Asynchronous detection, also known as free-running mode, where the time of arrival of photons is not known, might be required, especially in QKD and sometimes in LiDAR systems. In this sense, some techniques have been developed to achieve free-running operation. The first major method is the integration of an active quenching circuit, which can rapidly quench the avalanche pulses by lowering the voltage below breakdown fast, combined with placing this circuit in very close proximity to the SPAD, reducing parasitic capacitance and delays in the quenching circuit [122], [123]. The second one is to utilize the so-called negative feedback avalanche diode (NFAD) scheme. In this method, thin film resistors ($\approx 1\ M\Omega$) are monolithically integrated with the SPADs on the structure surface, thus reducing parasitic capacitance and providing very fast passive quenching [124]–[126]. Thanks to mainly these two techniques, the APP can be drastically reduced, thus enabling free-running operation in InGaAs(P)/InP-based SPADs.

To render InGaAs(P)/InP junctions operable as SPADs, some design constraints need to be taken into account:

- The avalanche multiplication cannot be performed in the low bandgap absorber layer; otherwise, the device is buried in the tunneling noise under very high electric fields. As a result, the absorber and multiplication layers need to be separated. As InP has a higher bandgap (1.34 eV), the natural choice is to confine the avalanche breakdown in an InP layer, which is already lattice-matched. Thanks to high bandgap InP, tunneling and DCR can be diminished significantly.

- Once the absorption and avalanche multiplication layers are separated, either electrons or holes should be selected for the multiplication since they travel in opposite directions under a given vertical electric field. In contrast to Si, the impact ionization of holes is greater than that of electrons in InP; thereby, the structure should favor the multiplication of the holes to achieve better detection efficiencies [83], [84], [98].

- To control and tailor the electric field in the epi-structure, an InP charge layer should be inserted between the absorber and the multiplication region. The doping concentration of the charge layer should be selected such that the electric field in the absorber should not exceed InGaAs' critical avalanche breakdown field of 2×10^5 V/cm under any bias condition [118]. However, the absorber layer should still be depleted at the breakdown voltage so that the photogenerated holes can be swept to the InP multiplication region by the drift mechanism. In other words, depleting the absorber enhances the carrier injection efficiency into the multiplication region and the PDP of the device, according to Eq. 2.9 in Chapter 2.

- In InP/InGaAs heterojunctions, a hole barrier occurs at the valence band. In order to avoid the hole pile-up and speed up the device operation, the bandgap between InP and InGaAs(P) can be tailored with an InGaAsP grading layer. The mole fractions of InGaAsP can be graded with the material flux during the growth such that InGaAsP evolves into InP (x = 1; y = 0) and InGaAs (x = 0.53; y = 1) at the corresponding interfaces. Thanks to the grading layer, the transit time of holes and the jitter of the SPAD will not be degraded.

- The doping of the InP avalanche region and InGaAs(P) absorption layer should be kept as minimal as possible to avoid a more peaked electric field and, accordingly, tunneling. However, these doping levels are restricted by the background doping during the growth process.

Hence, constructing this separate absorption-charge-multiplication (SACM) is vital to be able to operate InGaAs(P)-based SPADs. In addition to these design criteria, to construct the so-called p-i-n type photodiode of InGaAs(P)/InP SPADs, there are three main fabrication methods: double Zinc (Zn) diffusion [127], inverted mesa [98], and selective area growth (SAG) followed by a one-step Zn diffusion [128]. The difference between these techniques is in forming the p-side of the p-i-n photodiode, which will be discussed further in the following subsections.

3.2.1 Planar double Zn diffusion technique

In this technique, p-doping is not performed during epitaxial growth. Instead, a cap InP layer is grown intrinsically, but with an unavoidable background doping. Afterwards, the Zn diffusion process is performed on the InP cap layer from the top surface. Zn atoms displace In atoms and occupy these substitutional sites, which increases the acceptor density and realizes p-type doping for InP material [129]. Thus, the p-contact side of the p-i-n type photodiode is formed thanks to the Zn diffusion. The avalanche region in intrinsic InP then becomes sandwiched between the p-doped InP and n-doped charge layers. To accomplish Zn diffusion into InP, a silicon nitride (Si_3N_4) layer is deposited as a mask layer and etched to define the Zn diffusion areas. Zn sources are then deposited to the etched regions. The most commonly used Zn sources are dimethylzinc (DMZn) and zinc phosphide (Zn_3P_2) [130]–[133]. After the deposition of the source, Zn atoms diffuse into InP material with high temperature annealing around 500 ^oC [130].

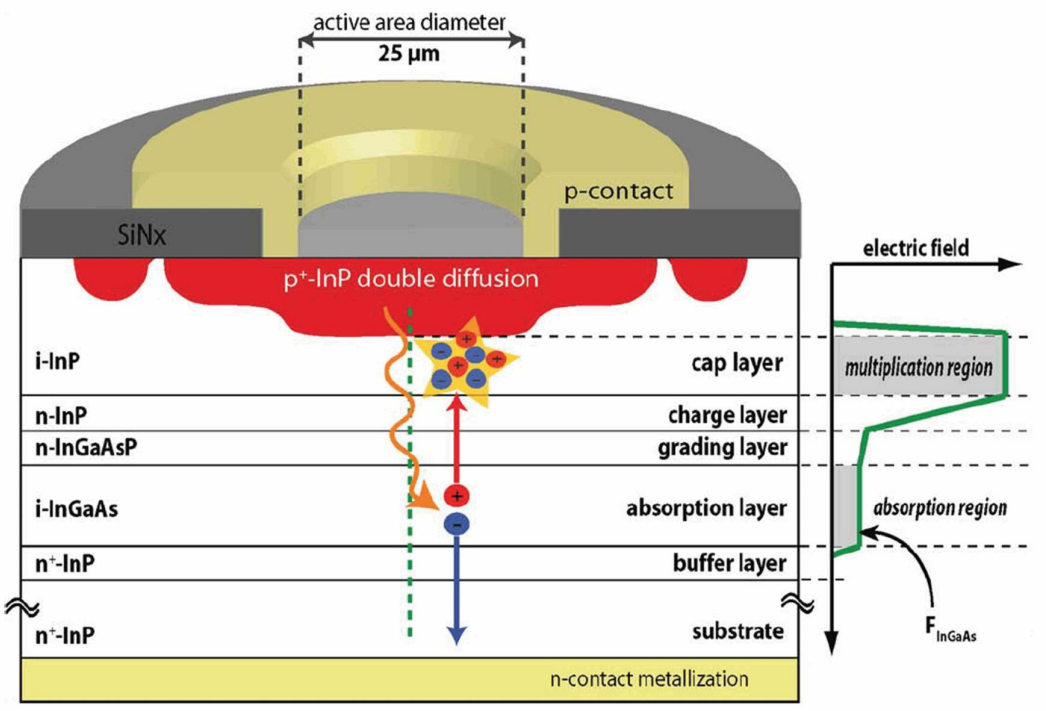


Figure 3.6: Cross-section of an InGaAs(P)/InP SPAD fabricated with the double Zn diffusion process [119].

A representation of the epitaxial structure of a SACM InGaAs(P)/InP SPAD is illustrated in Fig. 3.6 [119]. The Zn-diffused areas are shown in red in the i-InP layer. If Zn diffusion is performed only once, the device will suffer heavily from edge breakdown due to the junction curvature effect explained in Chapter 2. Therefore, to suppress the edge electric field, a GR can be implemented at the periphery of the active region with an additional Zn diffusion process [63], [127], [130]. Zn diffusion for the GR can reduce the edge effects through two mechanisms:

(1) By keeping the depth of the second diffusion shallower than the active region, the electric field at the center of the device is greatly enhanced, counteracting the field enhancements at the curvatures of the device edges [63], [127], [130]. As shown in Fig. 3.6, the active region is determined by the deeper Zn diffusion, whereas the shallow Zn diffusion corresponds to the GR. Since the multiplication region thicknesses are going to be the remaining thickness of the i-InP under the Zn-doped regions, a thinner multiplication region is constituted under the deep diffusion, leading to higher electric fields and a lower breakdown voltage at the center compared to the shallow-doped GR region. (2) If the shallow Zn diffusion for the GR is performed first, the curved edges of the GR diffusion can be graded and flattened thanks to the drive-in process of Zn atoms when there is no Zn source [127]. However, some groups also reported that the drive-in process is negligible at the same Zn diffusion temperature, and they achieved the same depth at Zn diffusion fronts after post-annealing while utilizing DMZn as the Zn source [130]. Therefore, the drive-in process might not always be prominent, depending on the fabrication process. Hence, optimizing Zn source, diffusion orders, depths, and times is critical to be able to operate InGaAs(P)/InP SPADs.

In addition, experimental results have demonstrated that the enhanced electric fields at the edges of the active or GR regions can still exist with the double Zn diffusion technique [63], [120]. Depending on the device design, the edge effects can get less effective with increasing V_{ex} on the active region. Nevertheless, an additional Zn diffusion is proposed to create floating guard rings (FGRs) to further suppress the edge effects, as indicated in Fig. 3.6 via the Zn diffused rings next to the shallowly-diffused GR regions [119], [127]. The working principle of FGRs is based on the fact that the depletion region around the main junction will be extended laterally as well with the increased V_{ex}. Since the FGRs are also highly doped, as in the shallow GR part, they form equipotential surfaces. When the depletion region reaches the FGRs, the equipotential lines become separated in the lateral direction from the curved junction to the semiconductor surface, which reduces the edge electric field magnitude [127]. Furthermore, it has been shown that the DCR and edge effect reduction can be optimized by properly biasing the FGRs [61]. Apart from controlling the electric potential at the junction curvatures, applying bias to FGRs decreases the DCR by mitigating the charge persistence effect. The photogenerated holes outside the InGaAs absorber region that is beneath the InP multiplication region can diffuse to the active InGaAs region and initiate more false counts, which is known as the charge persistent effect [134]. This effect can be a bottleneck in reducing noise at lower temperatures as the carrier lifetimes increase. These holes created outside the active region can be coupled to and collected from FGRs with the applied bias, thus preventing them from multiplying in the active region and having DCR degradation in the device [61].

Finally, it should be noted that any additional Zn diffusion to create GRs or FGRs limits the pixel pitch when an array of SPADs is intended to be fabricated. Hence, shrinking the dimensions of the active region and further structure and process optimizations are very important to increase the format of the SPAD cameras utilizing InGaAs(P)/InP materials.

3.2.2 Mesa technique

In the mesa technique, the p-contact is grown and doped (usually with boron) during the growth procedure. Therefore, the low-doped InP avalanche multiplication layer is also grown separately from the p-contact. In order to form the individual SPADs on the same wafer, a wet or dry etching process is then needed to achieve electrical and optical isolation between the neighboring devices. A cross-section of mesa-type InGaAs(P)/InP SPAD is given in Fig. 3.7 [98]. This device was grown as the vertically inverted shape of the double Zn diffusion device illustrated in Fig. 3.6. Due to the anisotropic etching of InP and InGaAs materials, inclined sidewalls are created at the device periphery. The slope of the sidewalls and the inverted-shaped structure help to suppress the edge breakdown in this technique. Thanks to the lack of Zn diffusion processes to form GRs, the minimum achievable pixel pitch is improved. However, the etching process tends to create more traps at the edges of the multiplication layer. If the edge effects are not completely eliminated with the slope of the sidewalls, the DCR and afterpulsing increase, leading to noisier devices than the Zn diffusion process. To reduce the number of traps at the device edges, passivation can be applied through various chemicals. For instance, polyimide was used in Fig. 3.7 to passivate the sidewalls. Hence, the mesa technique can provide easier fabrication and good pixel isolation in exchange for increased noise compared to planar Zn diffusion techniques.

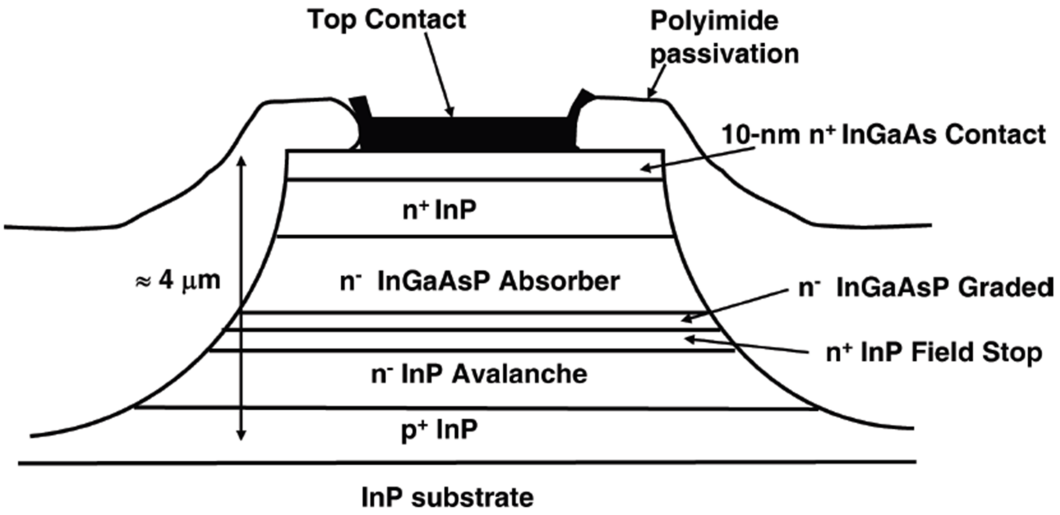


Figure 3.7: Cross-section of an InGaAs(P)/InP SPAD fabricated with the mesa process [98].

3.2.3 One-step Zn diffusion with SAG

To eliminate the additional Zn diffusion steps for GRs and potentially improve the pixel pitch of planar-type InGaAs(P)/InP SPADs, a SAG process has been recently proposed [128], [135]. In this technique, the growth procedure is the same as in the double Zn diffusion technique, where the p-type doping is not performed and a cap InP layer is grown with background doping. Afterwards, a SiN dielectric mask is used to grow another undoped InP layer by SAG on the

top of the active area. The thickness of the SAG layer gradually increases towards the edge of the SiN mask due to the enhancement of the growth rate closer to the mask edge, as shown in Fig. 3.8 [128], [135]. By utilizing the same dielectric mask, Zn diffusion is then performed after growing the SAG layer. The tapered surface of the SAG layer results in a diffusion profile whose depth towards the periphery of the active region is gradually decreased. This smooth transition at the edges of the active area ensures suppression of the edge electric fields and eliminates the edge breakdown. Surface laser scanning of SAG-based device revealed that the edge effects are indeed overcome and uniform response over the active area can be achieved [128], [135]. Hence, SAG epitaxy followed by only one-step Zn diffusion is another method to obtain InGaAs(P)/InP-based SPADs. On the other hand, side lobes next to the active region appeared in the laser scanning measurements, which are related to the crystal orientation of the SAG epitaxy. The occurrence of the side lobes still needs to be controlled and optimized to shrink the pixel sizes and pitch.

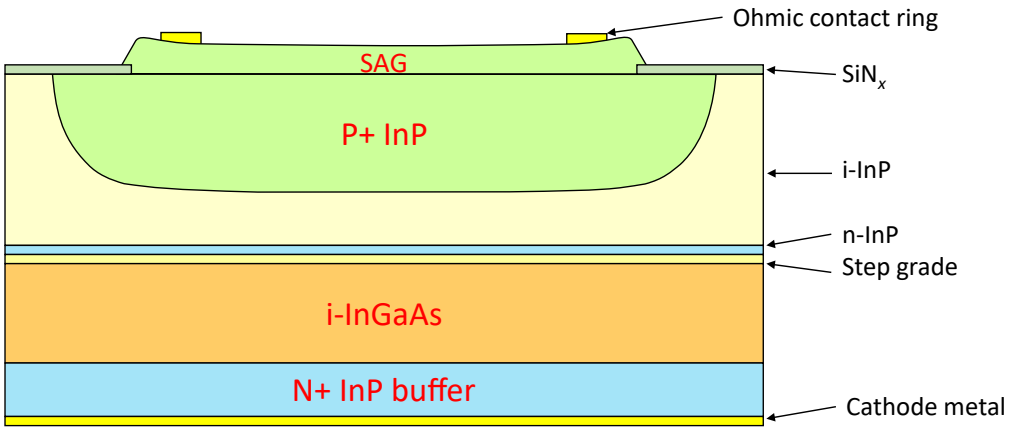


Figure 3.8: Cross-section of an InGaAs(P)/InP SPAD fabricated with the SAG followed by one-step Zn diffusion [128], [135].

4 New Techniques and Optimization for NIR Sensitivity Enhancement in Si CMOS SPADs

In this chapter, the work conducted to explore new methods for improving and optimizing the NIR sensitivity of Si SPADs in advanced technology nodes will be presented. The focus of the work is to make the wide depletion region approach more feasible and better-performing in CMOS technology nodes. For this purpose, FSI SPAD devices were designed in a 110 nm CMOS Image Sensor (CIS) technology via Synopsys Sentaurus Technology Computer-Aided Design (TCAD) numerical tool, and the fabricated devices were characterized after the tape-outs. The chapter is organized as follows: First of all, information will be provided about how the modeling and simulations were performed for Si SPAD devices in TCAD. Secondly, the two investigated techniques, called "doping compensation" and "double multiplication region", will be demonstrated [54]. It has been shown that the former can be utilized to enlarge the depletion region width of the SPADs. The latter, on the other hand, enhances the total avalanche breakdown probability by inserting a second multiplication region in a wide depletion region. Then, a guard ring optimization study will be explained for a substrate-isolated wide depletion region SPAD. Finally, the developed devices will be compared with the state-of-the-art FSI wide depletion region Si SPADs.

4.1 TCAD simulation methodology of Si SPADs

TCAD is a simulation environment to develop and optimize semiconductor devices and their fabrication processes. It is very crucial to simulate SPAD devices before the fabrication as well, since some structures might not yield an avalanche breakdown or might suffer from premature edge breakdown. In this sense, simulations are efficient ways of learning about the device operation, physics, or understanding the effects of each semiconductor or implantation layer on the performance parameters.

In TCAD, two main tools were utilized to numerically solve a device: "structure editor" and "device simulation". Each tool is controlled by a piece of code. The first tool is to define the device geometry along with the associated material types for each drawn layer. The doping concentration is also assigned in the structure editor to every semiconductor layer. In this

work, the exact doping concentrations were obtained from the foundry and integrated into the TCAD technology file. Another parameter while specifying the doping profiles is the curvature of the edge of the implanted layer. As explained in Chapter 2, steep edges can result in edge breakdown due to the junction curvature effect. As the curvature and the lateral diffusion information were not available from the foundry, a Gaussian function was used to describe the radius of the edge curvature and the lateral broadening of the implantation layers. Lastly, the meshing of the device is also defined here. Meshes are basically the smaller domains where the electrical and optical equations are solved. Therefore, mesh quality is important to converge the simulations to accurate solutions. In all of the simulations, triangular meshing was used in this work. Mesh density was also set to be denser in regions where the material or doping concentration changed, as well as in highly doped and high electric field regions and junction curvatures.

The second tool, device simulation, utilizes the finite element approach to solve coupled differential equations that describe the device's operation [136]. In order to obtain the electric field distribution, current, and electron and hole concentrations throughout the device, the drift-diffusion model is employed. This model consists of the current density (Eq. 4.1 and 4.2), electron-hole continuity (Eq. 4.3 and 4.4), and Poisson equations (Eq. 4.5), as shown below:

$$J_n = q\ n\ \mu_n\ E\ +\ q\ D_n\ \frac{dn}{dx} \tag{4.1}$$

$$J_p = q\ p\ \mu_p\ E\ -\ q\ D_p\ \frac{dp}{dx} \tag{4.2}$$

$$\frac{\partial n}{\partial t} = \frac{1}{q}\nabla.J_n\ +\ R_n \tag{4.3}$$

$$\frac{\partial p}{\partial t} = -\frac{1}{q}\nabla.J_p\ +\ R_p \tag{4.4}$$

$$\nabla\epsilon.\nabla\phi = -q(p - n + N_D - N_A). \tag{4.5}$$

In Eqs. 4.1 and 4.2, J_n and J_p indicate electron and hole current densities, μ_n and μ_p stand for electron and hole mobilities, D_n and D_p represent the electron and hole diffusion coefficients, and n and p belong to electron and hole densities. Electric potential is noted by ϕ. The net carrier generation-recombination rates are given by R_n and R_p in Eqs. 4.3 and 4.4. In TCAD, these equations are solved in a fully coupled way with Newton iteration. In this method, the device is analyzed under the initial bias voltage as a first step, and then the obtained results are used as guesses for the next step. The set of solutions eventually converges after many iterations, providing accurate results.

Although many parameters are unknowns to be solved, constants, or defined by the doping concentrations, the net generation-recombination rates require user input as well. In the simulations of Si-based SPADs, SRH and TAT generation mechanisms were considered while obtaining the I-V characteristics of the device. SRH generation is formulated as follows:

$$R_{SRH} = \frac{np - {n_i}^2}{\tau_h(n + n_i \times exp(\frac{E_{trap}}{kT})) \; + \; \tau_e(p + n_i \times exp(\frac{-E_{trap}}{kT}))}, \tag{4.6}$$

where τ_e and τ_h are the electron and hole lifetimes, respectively, E_{trap} is the trap energy level with respect to the mid-bandgap position, and k is the Boltzmann constant. In this work, τ_e was accepted as 1 ms and τ_h was adopted as 10 µs, whose values are based on the doping concentration of the utilized implantation layers and the lifetimes demonstrated in the previous works of [111], [137]. Also, the trap energy level is fixed exactly at the mid-bandgap. The TAT contribution is simulated via the Schenk model [138]. This model modifies the SRH generation with a field enhancement factor, which reduces the lifetimes in high electric field regions. The enhancement factor is adjusted by setting the lattice relaxation energy and the electro-optical frequency, for which the provided common values were selected [136], [138]. However, optimizing the lifetimes and TAT, and adding surface recombination to match the simulated I-V with the experimental data is beyond the scope of this work regarding Si-based SPADs.

To simulate the avalanche generation and breakdown voltage accurately, the impact ionization coefficients of Si should be defined appropriately. For this purpose, the Okuto-Crowell model was used in TCAD to insert ionization parameters [80], [136]. This model determines the variation of the ionization coefficients under high electric fields (> 10^5 V/cm). The generic formula to calculate the coefficients at room temperature is [80], [136]:

$$\alpha_{e,h} = a \times E \times exp(-(\frac{b^2}{E^2})). \tag{4.7}$$

In Eq. 4.7, the constants a and b are selected individually for the electron and hole to match the values with the ones reported in [80]. The dependence of α_e and α_h on the electric field is illustrated in Fig. 4.1. Then, in order to identify the avalanche breakdown voltage, the breakdown integral given in Eq. 2.3 in Chapter 2 was computed. However, since α_e and α_h are not equal anymore, the integral needs to be modified and separated for each carrier. The resulting integrals become:

$$I_e = \int_0^W \alpha_e(x) \times exp(-\int_x^W (\alpha_e(x') - \alpha_h(x'))dx')\,dx, \tag{4.8}$$

$$I_h = \int_0^W \alpha_h(x) \times exp(-\int_0^x (\alpha_h(x') - \alpha_e(x'))dx')\,dx, \tag{4.9}$$

where W corresponds to the width of the depletion region. The integration is performed along the electric field lines across the depletion region. Avalanche breakdown occurs if any of the integrals reaches one at any position of the device. Given that α_e is greater than α_h, as shown in Fig. 4.1, electrons have a higher probability of initiating the avalanche breakdown in Si

SPADs.

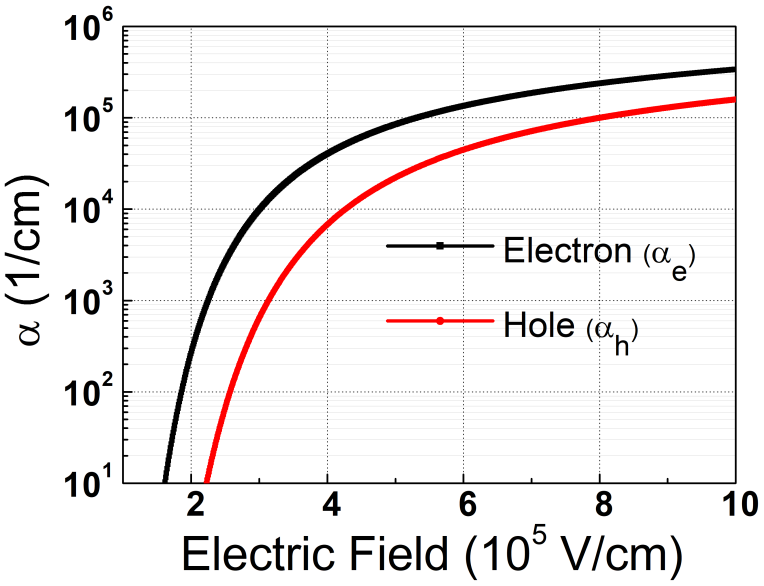


Figure 4.1: Electron and hole impact ionization coefficients utilized in TCAD for Si.

Lastly, to compute the avalanche breakdown probabilities of a device above the breakdown voltage, the McIntyre model was implemented in TCAD [139]. This model has the following coupled set of equations to be solved:

$$\frac{dP_{\mathrm{e}}}{dx} = (1 - P_{\mathrm{e}})\,\alpha_{\mathrm{e}}\,P_{\mathrm{j}}, \tag{4.10}$$

$$\frac{dP_{\mathrm{h}}}{dx} = -(1 - P_{\mathrm{h}})\,\alpha_{\mathrm{h}}\,P_{\mathrm{j}}, \tag{4.11}$$

$$P_{\mathrm{j}} = P_{\mathrm{e}} + P_{\mathrm{h}} - P_{\mathrm{e}}P_{\mathrm{h}} \tag{4.12}$$

under the boundary conditions $P_e(0)=0$ and $P_h(W)=0$, where P_e, P_h and P_j are the probabilities of avalanche breakdown for electron, hole, and electron-hole pair generated at position x. By integrating the two coupled differential equations simultaneously, the probability of triggering an avalanche at any position in the depletion region of the device can be found for both electrons and holes. Eq. 4.12 then combines these findings to provide the total avalanche breakdown probability of the device with respect to position.

4.2 Doping compensation technique

As mentioned, one of the main objectives of this work is to render the wide depletion region approach more viable to improve the NIR efficiency of CMOS-based SPADs. However, achieving a wide depletion region and low noise SPADs becomes challenging in more advanced CMOS technologies. Essentially, utilizing smaller technology nodes is vital to shrinking the pixel

sizes, enabling advanced functionality on a chip, higher resolution, lower power consumption, and multi-megapixel SPAD arrays. However, doping concentrations tend to increase in the reduced dimensions of smaller nodes [140], [141]. As a result of higher doping values, thinner depletion regions are formed, which restricts the detection of NIR wavelengths. Furthermore, higher electric fields created in thinner junctions can enhance tunneling contributions, deteriorating the noise of the devices. In this sense, developing new methods to enlarge the depletion region width of the devices is crucial to obtain high PDP at longer wavelengths. In order to achieve this aim, the doping compensation technique was utilized, which has proven to effectively alter the doping profiles of the existing and limited implantation layers to create a new junction with a wider depletion region [54].

In theory, when a semiconductor is doped with both donors and acceptors, the process is called doping compensation [81], [142]. In this special case, in addition to the regular ionization process of donors and acceptors, the electrons of the donor atoms also fill the vacancies created by acceptor atoms. This process is illustrated in an energy band diagram in Fig. 4.2. From the charge neutrality equations, the steady-state hole concentration under complete ionization is expressed as:

$$p = \frac{N_A - N_D + \sqrt{(N_A - N_D)^2 + 4n_i^2}}{2}. \tag{4.13}$$

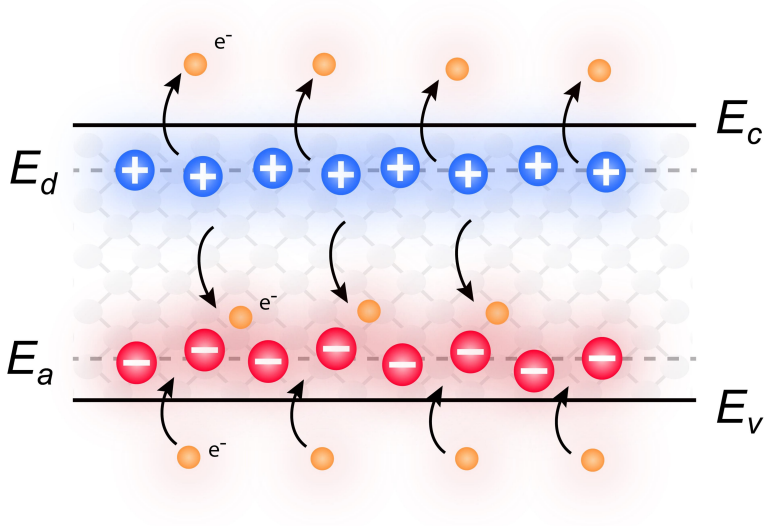


Figure 4.2: The energy band diagram of a compensated semiconductor.

From Eq. 4.13, it can be indicated that hole concentration in a compensated semiconductor can be controlled by impurity (dopant) concentrations. As described, the neutralization of some of the ionized acceptors by the ionized donors can lead to a different doping concentration. Therefore, although it may degrade the transport properties, doping compensation gives the freedom to adjust the doping profile of a semiconductor. This approach is applicable to any CMOS technology, and it gives designers an opportunity to obtain more desirable

doping profiles without any modifications to the fabrication process. Thus, it copes with the limitations of available foundry processes, and widening the possibility of implementing various doping concentrations. In this work, doping compensation has been implemented on the devices fabricated using an advanced 110 nm CIS technology. A similar approach has been applied in a linear mode APD to separate the multiplication and absorption regions of the device [143], which was then also characterized as a SPAD [144]. Here, the technique has been solely performed to design the multiplication region of a SPAD.

4.2.1 Device structures

To show the applicability of the technique, two types of devices were designed, with and without doping compensation. The cross-section of the structure without doping compensation can be seen in Fig. 4.3 (a). The device is based on the N^+/high-voltage p-well (HVPW) junction and has a circular shape with a 10 µm active diameter (HVPW width) with a virtual GR around the active area. Since both N^+ and HVPW layers have high doping concentrations, a thin depletion region and high tunneling noise are expected in this device. In order to modify the doping profile of the highly-doped HVPW layer and expand the depletion region, an n-type layer named high-voltage n-well (HVNW) was inserted to compensate for some of the ionized acceptors in the HVPW layer. The cross-section of the compensated SPAD is provided in Fig. 4.3 (b). All the layers and dimensions were kept the same with respect to the non-compensated device, except for the inclusion of HVNW to perform the compensation technique. The approach to realizing compensation is based on a reverse-type layer available in the technology without modifying the process. This HVNW layer existing in the technology node was chosen because it has the potential to effectively reduce the ionized acceptor concentration in HVPW thanks to its concentration and depth, hence achieving a wider depletion region. Therefore, doping compensation can help prevent the tunneling noise anticipated in the non-compensated device by enlarging the depletion width. Furthermore, the width of the HVNW layer was chosen to be wider than the HVPW so that the active area is defined by HVPW in both structures and narrower than N^+ so that edge breakdown does not occur.

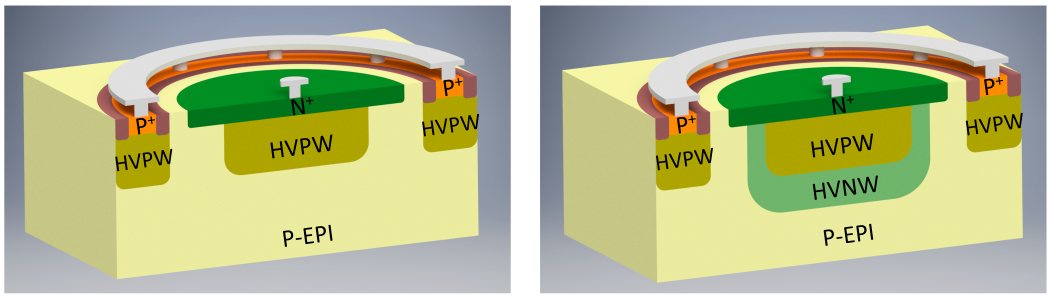


Figure 4.3: Cross-sections of (a) N^+/HVPW junction with a virtual GR and (b) N^+/HVPW junction with the HVNW doping compensation layer and a virtual GR.

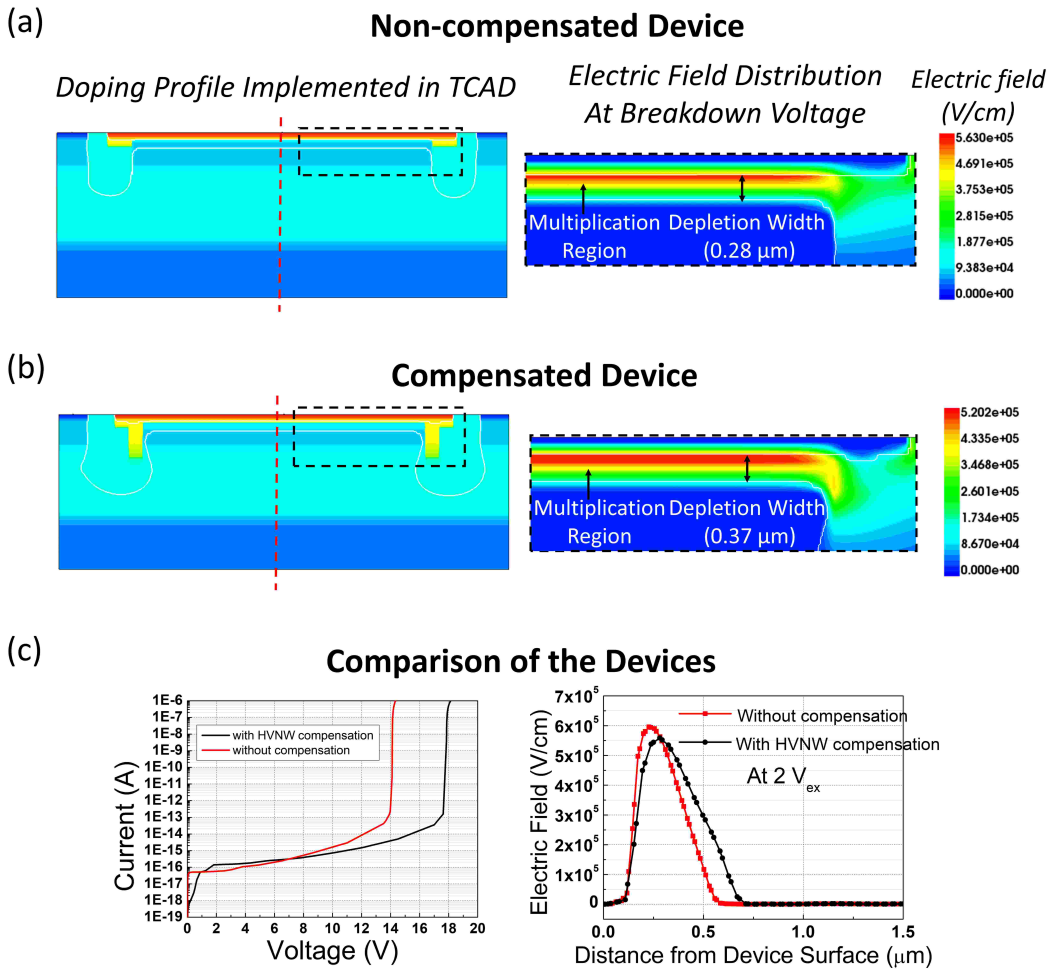


Figure 4.4: Doping profile and electric field distribution at breakdown voltage of the (a) non-compensated SPAD and (b) compensated SPAD with HVNW layer. (c) I-V curves and electric field magnitudes at 2 V_{ex} taken along the center of the devices.

4.2.2 TCAD simulation results

The proposed structures were designed and analyzed first in TCAD in terms of electric field distributions and I-V curves. The exact doping profiles of each of the employed layer taken from the foundry were integrated to TCAD. However, the doping values cannot be disclosed due to confidentiality reasons. Fig. 4.4 (a) shows the electric field distribution for the non-compensated device at the breakdown voltage. As demonstrated, the electric field is fully confined in the multiplication region without having edge breakdown. As the junction is constituted between two highly-doped layers, the depletion width of the device is 0.28 µm at breakdown, which is rather thin as expected. The thin depletion region leads to a very high electric field (5.63×10^5 V/cm) at the breakdown condition, which can lead to a degradation in the DCR of the device because of the increased trap-assisted tunneling rates. Then, the TCAD results belonging to compensated SPAD with HVNW are shown in Fig. 4.4 (b). In this structure,

the depletion region expands to 0.37 µm, while decreasing the peak electric field magnitude to 5.2×10^5 V/cm. The expansion of the depletion width is linked to the reduced acceptor level in the HVNW layer with compensation. Hence, with the compensated device, higher collection efficiency of the photogenerated carriers at NIR and lower trap-assisted tunneling noise can be achieved. Afterwards, the comparison of the proposed devices is presented in terms of I-V characteristics and electric field profile at 2 V_{ex} in Fig. 4.4 (c). As expected from the depletion region widths and peak electric field magnitudes, a relatively lower breakdown voltage of 14 V is observed in the non-compensated SPAD, whereas it increases to 17.6 V in the compensated device. Moreover, the electric field comparison graph indicates that the electric field profile becomes more uniform in the compensated device, and it is extended due to a wider depletion region. In conclusion, by extending the depletion region width and reducing the peak electric field magnitude, higher NIR efficiency and reduced trap-assisted tunneling noise are aimed at being achieved.

4.2.3 Characterization results

Following the designs, the structures were fabricated using a 110 nm CIS technology. A SPAD image after fabrication is given in Fig. 4.5 (a). The light emission test results of a compensated sample #3 are shown in Fig. 4.5 (b). These images were captured through a digital microscope (Hirox KH-8700) connected to a CCD camera. The images show that there is no edge breakdown, and the light-sensitive area has a 10 µm diameter, which is the same as in the drawn layout. Similarly, no edge breakdown issue was observed in the non-compensated device as well.

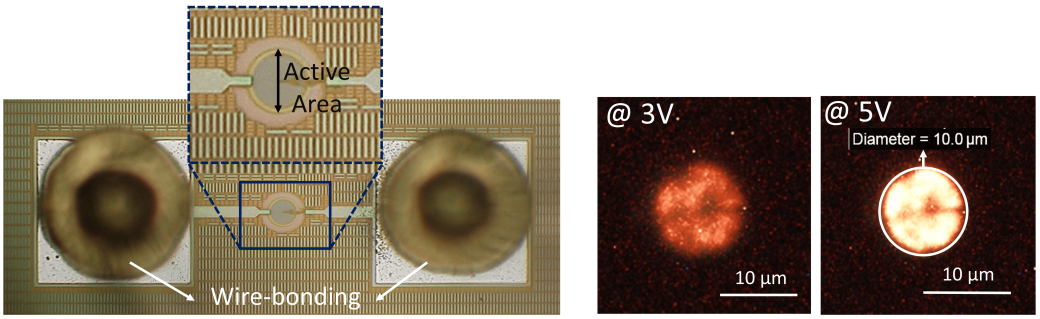


Figure 4.5: (a) Micrograph of a fabricated device. (b) Light emission test results of the compensated SPAD sample #3.

I-V characteristics of the fabricated N^+/HVPW junction-based SPADs with and without the compensation layer are given in Fig. 4.4 under both dark and illumination conditions. From dark I-V curves, the avalanche breakdown voltages are deduced as 14 V and 17.2 V, respectively. These values are well-matched to TCAD simulations. I-V curves also indicate that there is almost no difference in dark current while changing the compensation layers, except for the

variation in the breakdown voltage. Considering the slope and voltage dependence of these I-V curves, it appears that tunneling noise is not completely eliminated in the compensated device, even if the peak electric field magnitude is decreased. The current saturation above its breakdown voltage for each device is due to the series resistance of the device and the space-charge effect [81]. Finally, I-V curves under illumination demonstrate that both devices are sensitive to light.

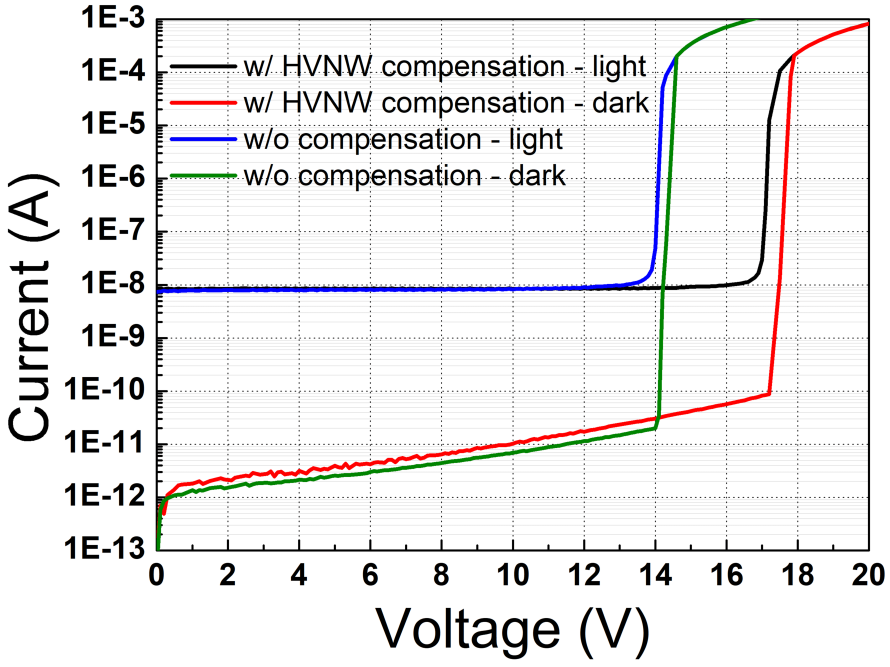


Figure 4.6: I-V characteristics of the devices with and without doping compensation under both light and dark conditions.

In the remaining measurements, SPADs were passively quenched and recharged with a 220 kΩ ballast resistor. The avalanche pulses were read through a digital oscilloscope (Teledyne LeCroy WavePro 760Zi-A) by measuring the voltage across this resistor from 50% of the peak level, as described in Chapter 2 under the "Passive technique" subsection. The devices were operated in free-running mode. DCR measurement results belonging to both types of devices and obtained from five different dies fabricated on the same wafer are provided in Fig. 4.7. These measurements were taken at room temperature. To be able to compare the devices, Sample #3 in both types was chosen since they correspond to the median of the measurements. The median DCR values at 2 V_{ex} are 1910 cps and 80 cps for the non-compensated and compensated devices, respectively. Therefore, a factor of 24× decrease is observed in the DCR of the compensated device at 2 V_{ex}. Above 2 V_{ex}, the noise of the non-compensated devices increases significantly (> 10k cps). The high DCR of non-compensated devices is attributed to enhanced trap-assisted tunneling noise due to the thin depletion region. The sudden increase in the DCR of the non-compensated devices between 1 V_{ex} and 2 V_{ex} also supports that trap-assisted tunneling contribution might be dominating the noise. This tunneling contribution

is mitigated in the compensated devices thanks to the wider depletion region.

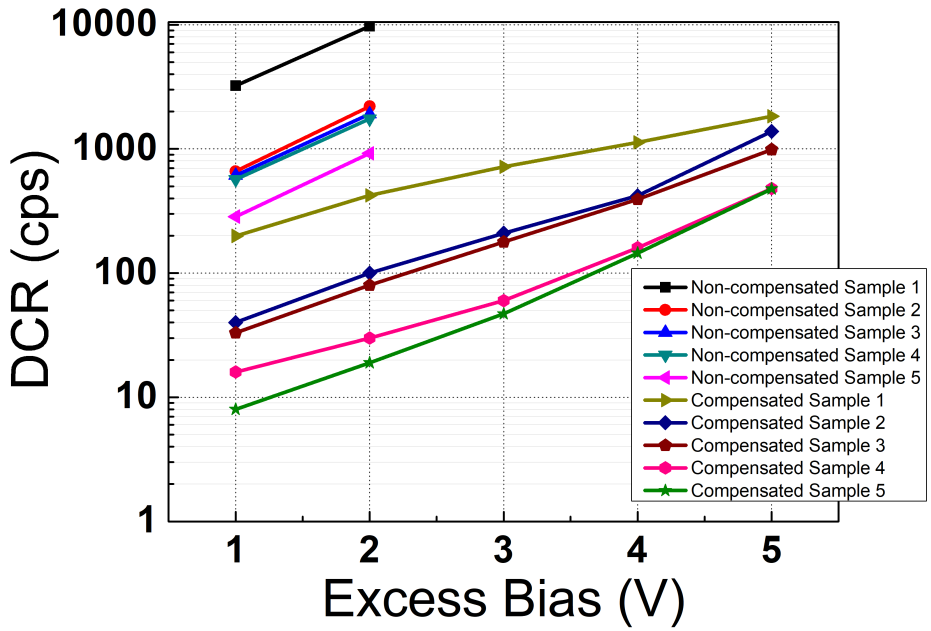


Figure 4.7: DCR versus excess bias voltage curves of non-compensated and compensated SPADs obtained from 5 different dies at room temperature.

PDP curves obtained for both types of devices are depicted in Fig. 4.8 (a). The measurements were taken under FSI configuration, and the area measured in the light emission tests was used to calculate the PDPs. A calibrated silicon photodiode (Hamamatsu S2281) was utilized to measure the precise incident light power. The peak PDPs achieved are 52.4% and 46.6% for the non-compensated and compensated devices, respectively, at 2 V_{ex}. Peak PDPs lie at 450 nm, which is rather a short wavelength and is a result of the fact that the multiplication region resides close to the device surface. At the same excess bias voltage, however, the compensated device achieves a higher PDP of 7.3% at 800 nm and 5% at 850 nm than that of 5.9% at 800 nm and 4.3% at 850 nm in the non-compensated device, thanks to the wider depletion region. Furthermore, the low noise in the compensated SPAD allows to measure the spectrum up to 5 V_{ex}. At 5 V_{ex}, the peak PDP reaches 73.9%, and NIR PDPs become 10.7% and 7.3% at 800 nm and 850 nm wavelengths. This indicates that the relative increase in the maximum achievable PDP is more than 40% for the peak and more than 75% for NIR wavelengths in the compensated device, owing to the reduced noise. The summary of the PDP comparison for the aforementioned wavelengths is provided in Fig. 4.8 (b). Still, low PDPs at longer wavelengths can be associated with the not-achieved very wide depletion region even after utilizing the compensation. Besides, not many of the diffused minority electrons seemed to be detected. This might be due to a slightly deeper HVNW compensation layer than HVPW, thermalizing the diffused electrons. After demonstrating that doping compensation can be utilized to mitigate noise and enhance NIR efficiencies, the remaining characterizations were performed for the compensated SPAD to prove that the technique does not deteriorate the rest of the

figures.

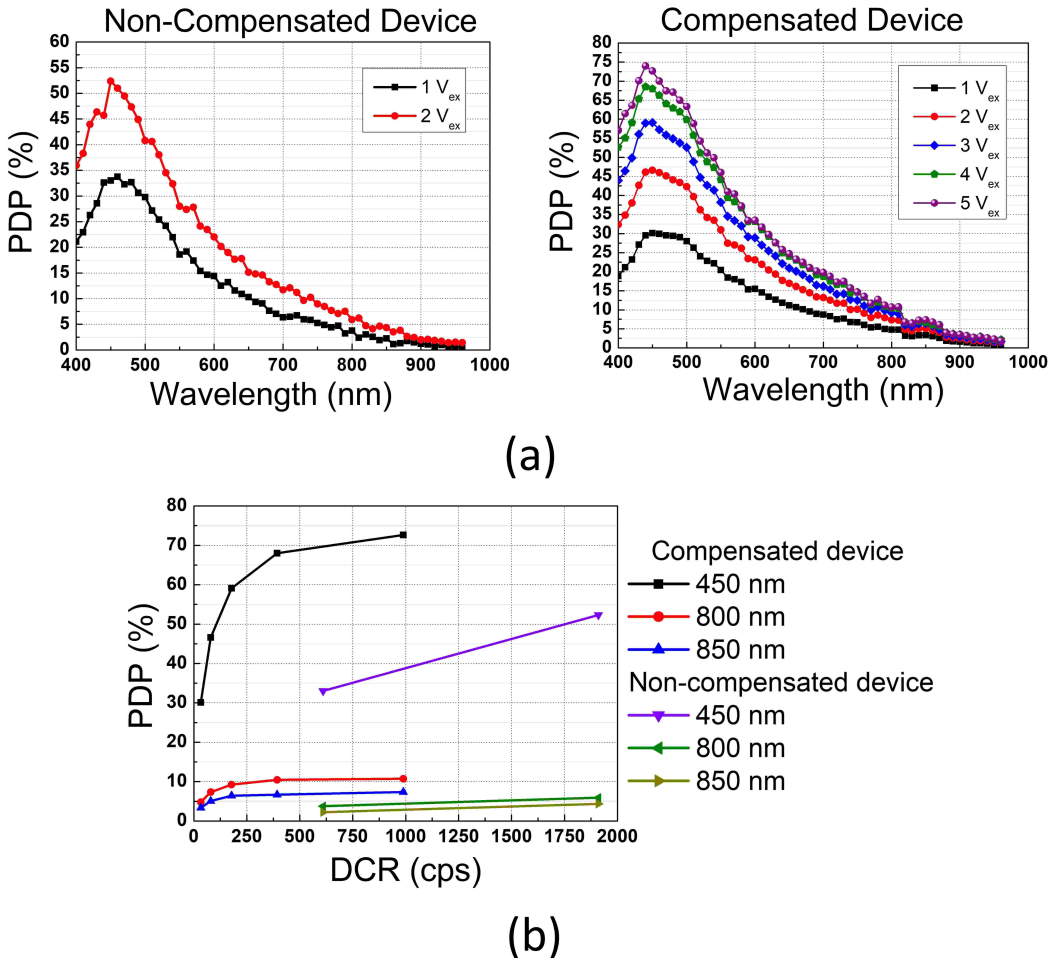


Figure 4.8: (a) Measured PDP spectrums of the compensated and non-compensated devices at various excess bias voltages. (b) PDP versus DCR comparison of the device at the peak and NIR wavelengths.

The variation of breakdown voltage with respect to temperature can be seen in Fig. 4.9 (a) for the compensated device. It is observed that the breakdown voltage changes 0.15 V per 10°C. DCR versus V_{ex} under various temperatures is plotted in Fig. 4.9 (b). At 30 °C, the device has 0.52 cps/µm^2, 2.15 cps/µm^2, and 12.26 cps/µm^2 normalized DCRs at 1 V_{ex}, 3 V_{ex} and 5 V_{ex}, respectively. It can be noticed that the slope of the curves becomes steeper below 30 °C, which indicates that trap-assisted tunneling contribution starts to dominate the total DCR. At 0 °C and 5 V_{ex}, the normalized DCR reduces to 3.41 cps/µm^2. Above 30 °C, the DCR seems to be dominated by the thermal generation up to 4 V_{ex}. Between 4 V_{ex} and 5 V_{ex}, the slope of the curves also becomes steeper, signifying that the tunneling contribution is taking over.

The measured afterpulsing histogram based on the inter-arrival time method is shown in Fig. 4.10. The measurement was performed at room temperature and 5 V_{ex}. The number of counts started to decrease below 3 µs due to passive quench and recharge through the external 220

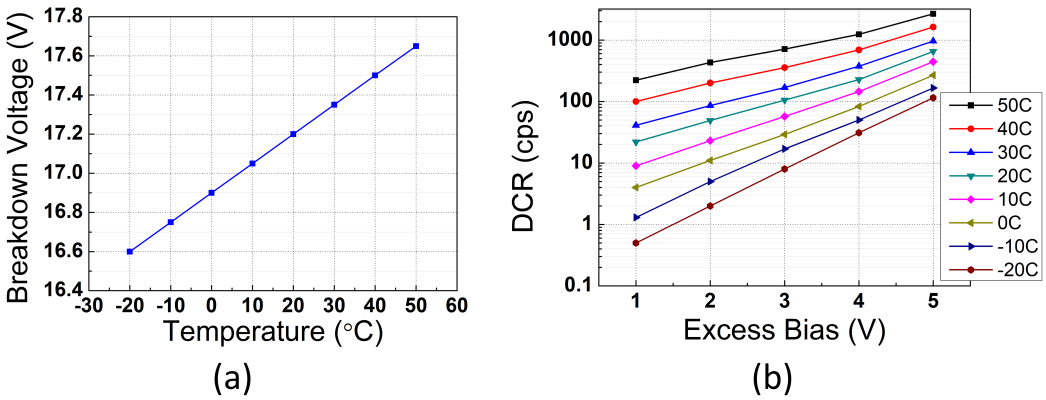


Figure 4.9: (a) Breakdown voltage variation with temperature. (b) DCR versus V_{ex} curves for different temperatures.

kΩ . Therefore, the dead time of the SPAD was determined to be 3 µs since it deviates from the Poissonian distribution. An exponential fit was then performed on the graph, as depicted, and the APP was calculated as 0.9%.

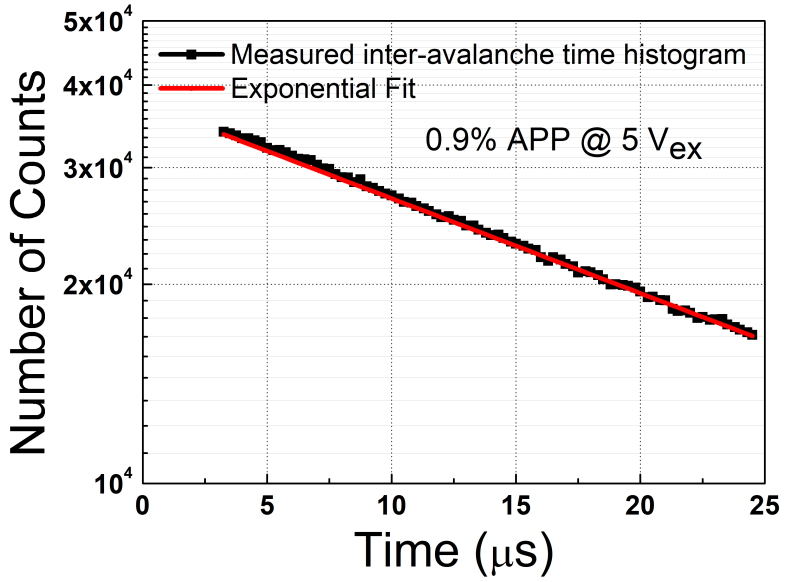


Figure 4.10: Inter-arrival time histogram of the compensated SPAD at 5 V_{ex}.

The timing jitter of the compensated SPAD was measured with an 850 nm (A.L.S. GmbH) pulsed laser source at 100 kHz repetition frequency. The histograms recorded at 850 nm wavelength and at 3 V_{ex} and 5 V_{ex} are illustrated in Fig. 4.11. FWHM of the histogram is obtained as 79 ps at 5 V_{ex}. After the deconvolution of the laser pulse, a SPAD jitter of 68 ps and 92 ps (FWHM) are obtained at 5 V_{ex} and 3 V_{ex}, respectively. The acquired FW1/100M value is also 371 ps at 5 V_{ex}. The timing histogram also demonstrated a very slow diffusion tail, as can be observed after 1.2 ns in Fig. 4.11. The magnitude of the tail indicates that many of the diffused electrons were indeed thermalized in the HVNW compensation layer and could not

reach the depletion region. The low jitter values make the device appealing for high timing resolution applications.

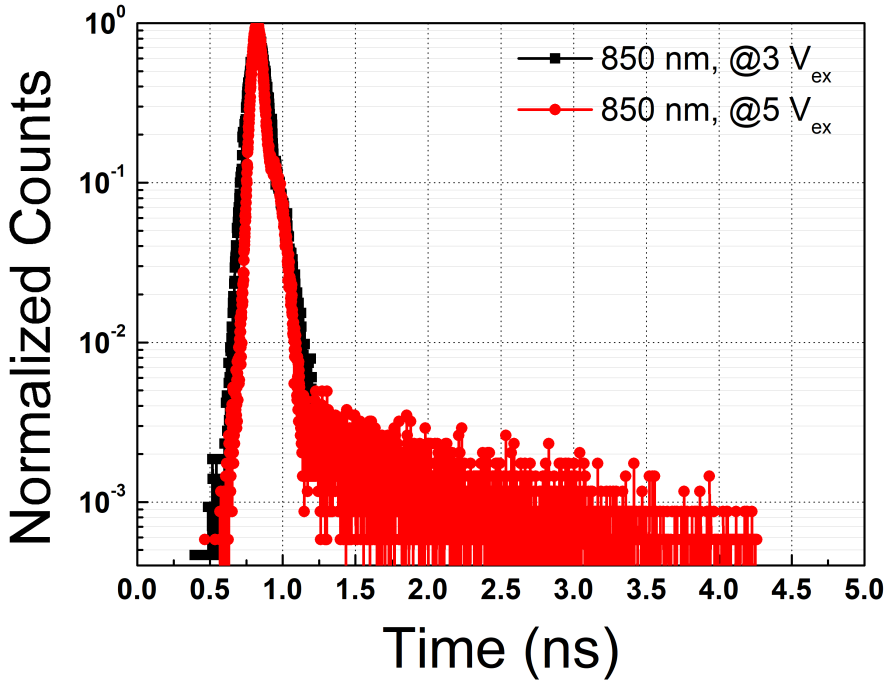


Figure 4.11: Timing jitter histograms of the compensated SPAD at 850 nm.

In conclusion, the achieved results demonstrate that the doping compensation technique is applicable to the design of SPADs in CMOS technologies [54]. The method gives designers more flexibility to alter the doping concentration of a layer they would like to utilize, enabling them to somewhat overcome the restricted doping profiles of the foundries. Here, it has been shown that doping compensation can be used to extend the depletion region width of the SPADs to enhance their efficiency at NIR. In addition to reaching higher efficiencies at NIR at the same excess bias voltage, the compensation also increased the maximum attainable PDP by lowering trap-assisted tunneling noise present in the non-compensated, thanks to a wider depletion region.

4.3 Double multiplication region technique

Another objective of this thesis was to improve the depletion region in SPADs. As discussed in Chapter 3, the wide depletion approach already requires a high excess bias voltage (~10 V) to increase the avalanche triggering probability and the PDP. The aim here is to find a new way to achieve similar or higher NIR efficiencies without applying more excess bias. For this purpose, the electric field simulations that were provided in previous works were carefully reviewed [51], [53], [110], [145]–[152]. As can be observed in Fig. 3.2 in Chapter 3, there is only one multiplication region for carriers in the depletion region, where the high electric

field is confined in these reported devices in the literature. Outside this multiplication region, the electric field decays linearly towards the edges of the depletion zone. Photon absorption at longer wavelengths and avalanche multiplication processes in a wide depletion region SPAD with a single multiplication region and p-substrate are illustrated in Fig. 4.12 (a). In this scheme, only photogenerated electrons can trigger an avalanche with a certain probability P_e, both in the case of photon absorption inside and outside the depletion zone. Photogenerated holes are swept to contacts located outside without undergoing avalanche multiplication. Following this analysis, inserting a second multiplication region was investigated and developed to increase the total avalanche breakdown probability P_j, thereby enhancing the PDP as indicated in Eq. 2.9. The scenario with a double multiplication region is described in Fig. 4.12 (b). As depicted in this scheme, double multiplication provides two advantages: (a) for the photons absorbed inside the depletion region, holes will have the chance to trigger an avalanche as well; (b) for the photons absorbed outside the depletion region, the electrons reaching the depletion will have the chance to undergo the avalanche process twice. Therefore, independently of the absorption location, the total avalanche breakdown probability increases for a photogenerated electron-hole pair, leading to similar or higher NIR PDPs at a lower excess bias voltage. The downside of this new method is that it can degrade the noise performance since the higher P_j also increases the number of dark pulses generated thermally or by tunneling. The technique was then implemented using the same 110 nm CIS technology for the first time and named *double multiplication region method* due to the existence of two distinct multiplication regions achieved in the same depletion region.

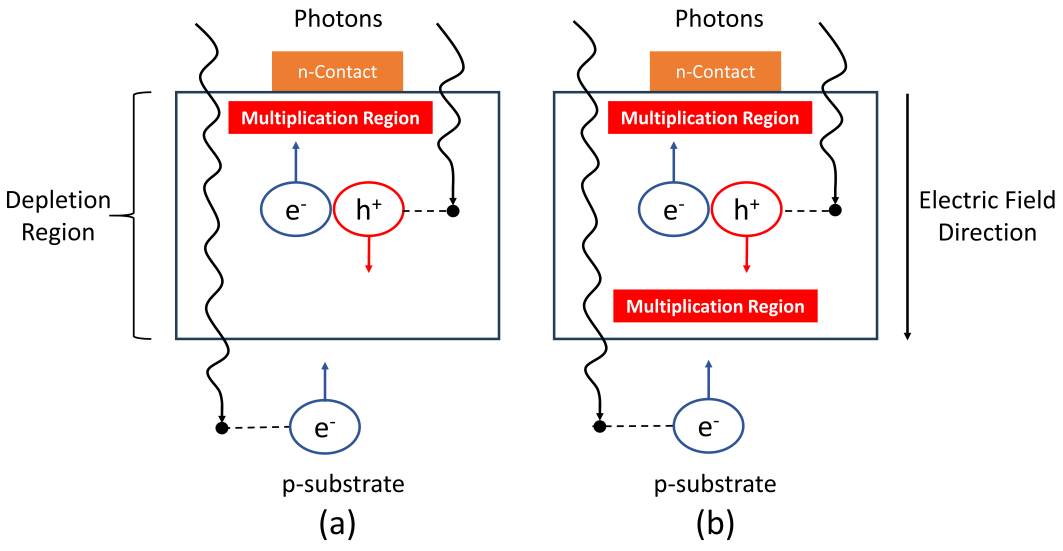


Figure 4.12: Photon absorption at longer wavelengths and avalanche multiplication processes in a wide depletion SPAD with (a) one multiplication region, (b) double multiplication region.

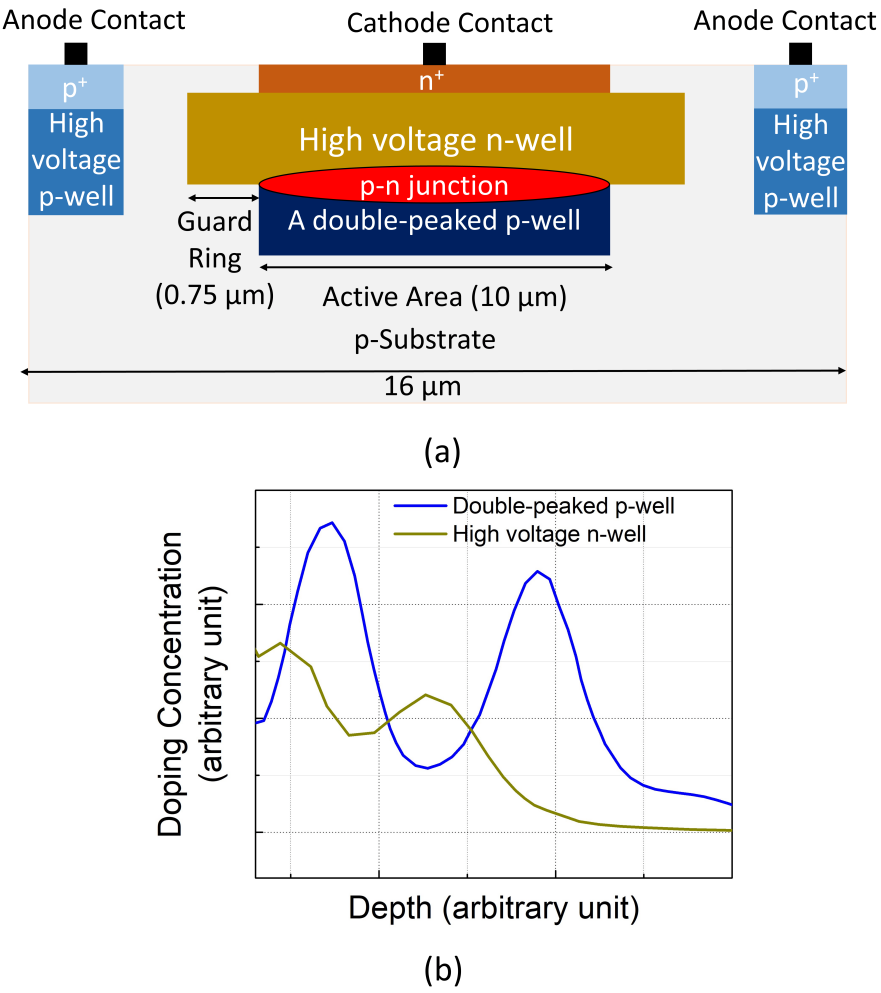


Figure 4.13: (a) Cross-section of the proposed device. (b) Doping profiles of the utilized layers in arbitrary units.

4.3.1 Device structure

Following this idea, a series of standard implants available in the foundry process were developed to achieve a double multiplication region. The cross-section of the proposed device can be seen in Fig. 4.13 (a). A substrate-non-isolated structure was preferred to collect the diffused carriers at longer wavelengths. Also, as indicated, the p-n junction is formed between the HVNW and a double-peaked p-well layer. The active area defined by the p-well has a 10 µm diameter, and there is a 0.75 µm virtual guard ring to prevent edge breakdown. From contact to contact, the whole SPAD covers 16 µm diameter. Moreover, the doping profiles of the utilized layers are given in Fig. 4.13 (b) in arbitrary units, due to confidentiality requirements from the foundry. In this configuration, the total doping concentration varies as n-p-n-p, which in fact should form two p-n junctions and depletion regions. The first intersection between the n and p layers should correspond to the first junction, and the third intersection should create

the second one. However, it is expected that the electric field is going to decay when n-doping exceeds p-doping in the middle of the device. If the second junction is located before the electric field reaches zero, then the two depletion regions, which have separate multiplication regions, should merge to constitute only one depletion region. This can be achieved provided that the n-doping and depth in the middle are suitable to not yield zero electric field before the second junction. Hence, it is possible that the second multiplication region in the same depletion zone can be achieved via this second peak in the doping profile of the p-well. Having a second peak in the p-well that exceeds the n-doping is thereby necessary to achieve a second multiplication region. Although this device was designed using the already existing implant layers in this technology node, a p-well with multiple peaks might not be common to find in the foundry processes at this point in time. As long as there is such an implant layer, the proposed method here can nevertheless be applicable to other CMOS technologies as well.

4.3.2 TCAD simulation results

In TCAD, in order to prove the existence of the second multiplication region and to show the avalanche breakdown probability enhancement with a double multiplication region method, a reference device was first created to replicate the previously designed wide depletion SPADs [51], [53], [110], [145]–[154] by simulating a SPAD using the technology's available layers. The cross-section of this wide depletion SPAD example is provided in Fig. 4.14 (d). The structure is based on HVNW/deep n-well/buried p-well layer implants, with the junction located between the deep n-well and the buried p-well. The design is similar to the SPAD demonstrated in [53], except that the buried p-well is used in the junction instead of the p-substrate. This structure with n^+ contact on top of the SPAD area was preferred to be able to sweep the carriers in the same direction as in the proposed double multiplication region SPAD. The electric field direction will thus not have an effect on the avalanche triggering probability simulations due to the impact ionization coefficient differences between the electrons and the holes. The breakdown voltage of this control SPAD was 31.5 V in TCAD. The electric field simulation results of the reference SPAD are seen in Fig. 4.14 (b), where the area highlighted in red corresponds to the multiplication region. As shown in this example device at both 1 V_{ex} and 5 V_{ex}, the depletion region contains only one multiplication region, and in the remaining space, the electric field decays towards the edges.

Afterwards, the proposed design was simulated, which has a 30.8 V breakdown voltage and a slightly shorter depletion region than the reference SPAD. The electric field results of the proposed SPAD are provided in Fig. 4.14 (a). As can be seen, there is only one depletion region, whose boundaries are set by the white lines and highlighted with a black arrow. As foreseen in the initial design, the implemented n-p-n-p-type doping profile actually formed two depletion regions, which merged into a single wide depletion region. This merging occurred because the n-doped region in between the two peaks of the p-well is quite narrow and stays depleted. When the polarity changes to n between x=0.3-0.5 µm, the electric field starts to decay, but it rises again before reaching zero, as can be observed in Fig. 4.14 (c), which is taken at the

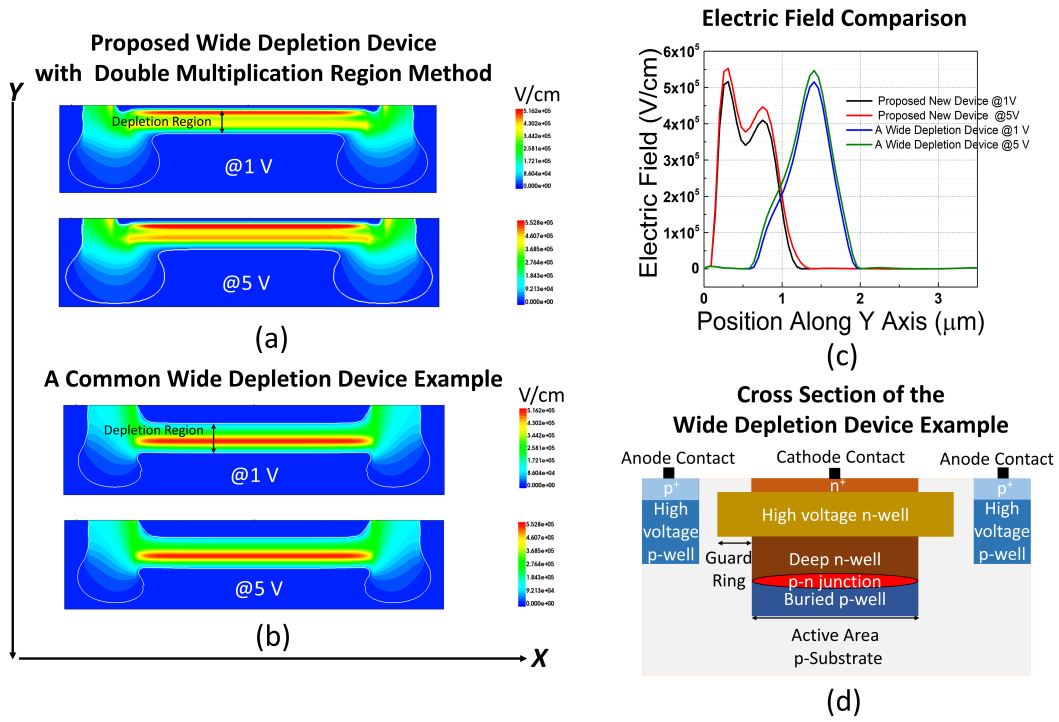


Figure 4.14: Electric field profiles of (a) the proposed device with the double multiplication region and (b) a SPAD with a wide depletion region. (c) Magnitude of the electric field at the center of each device along the y axis. (d) The cross-section of the reference SPAD with a wide depletion.

center of the devices along the y axis. This rise in the electric field, with the second peak in the p-well, formed the second multiplication region at the third intersection point. At 1 V_{ex}, in addition to the first multiplication region shown in red, the second multiplication region is highlighted in yellow, where the electric field is also above the critical field. At 5 V_{ex}, the second multiplication becomes painted in orange color. By interpreting Fig. 4.13 (b) and Fig. 4.14 (a) further, it is verified that the first multiplication region occurs at the first intersection point of the p and n-wells. Then, the electric field magnitude decreases, where the n-doping exceeds the p-doping value. Finally, the second multiplication region is located starting from the third intersection point between these two wells. The second multiplication region is also prominent in the proposed device, with the emerging second peak in Fig. 4.14 (c). In summary, by constituting a wide depletion region and inserting a second multiplication, it is aimed at achieving high PDP at longer wavelengths at relatively lower excess bias voltages, owing to boosting the avalanche triggering probabilities of the photogenerated electron-hole pairs.

As a side note, one can argue that the electric field magnitude, in fact, always stays above the critical electric field between the two peaks observed in the electric field profile, leading to an extended multiplication region instead of a double multiplication region. However, if there had been a longer electric field decay (let's say between x=0.3-0.75 μm instead of x=0.3-0.5 μm) provided that the electric field would stay above zero and go below the critical breakdown

field, this could have helped to extend the depletion region, and separate more clearly the two distinct multiplication regions. This would have rendered the name *double multiplication region* clearer. Since these are not custom-made doping layers in this foundry process, there is no flexibility to modify them. Still, having a shorter electric field decay for this type of structure does not mean that the multiplication regions are not separated and there is an extended multiplication region. There are two distinct multiplication regions created by multiple p-n junctions, but with a rather short separation between them.

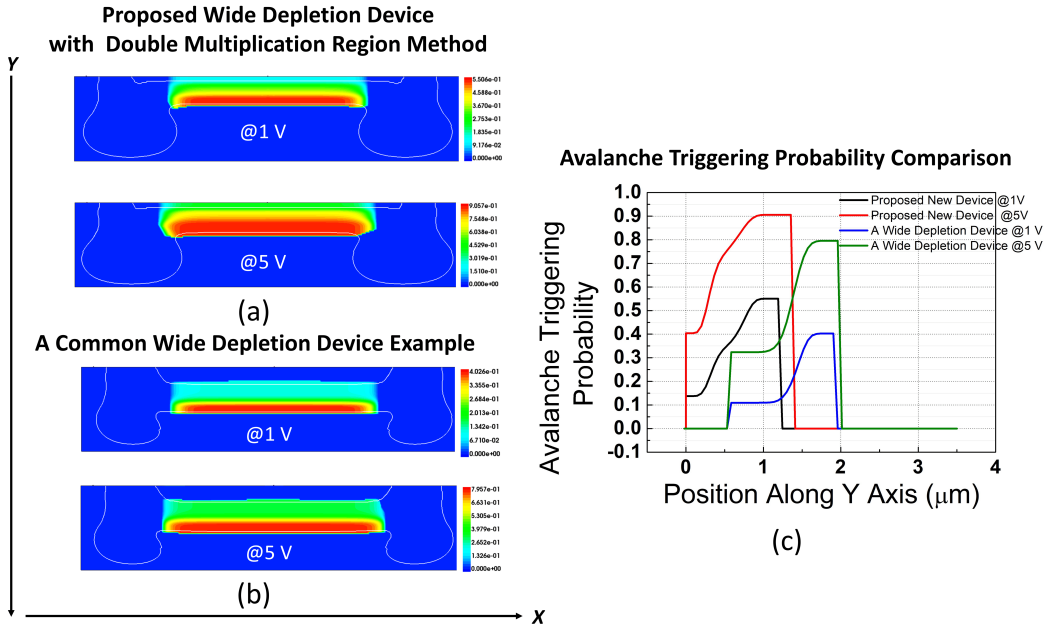


Figure 4.15: Electron-hole avalanche breakdown probabilities of (a) the proposed device with double multiplication region method, and (b) reference wide depletion region SPAD. (c) Values of the avalanche breakdown probabilities at the center of each device along the y axis.

The avalanche breakdown probability of each device was also calculated in TCAD to show the triggering probability enhancement with the double multiplication region technique. The results are provided in Fig. 4.15. The second multiplication region in the proposed device, as illustrated in Fig. 4.15 (a), significantly increased both the magnitude and volume of avalanche triggering probability from 1 V_{ex} to 5 V_{ex}, reaching 90% in half of the depletion region. In contrast, in the reference wide depletion region device, as shown in Fig. 4.15 (b), the high avalanche triggering probability is restricted due to only one multiplication region, and it rapidly decreases and remains low in a significant portion of the depletion region. The exact values of triggering probabilities at the centers of the devices along the y axis are given in Fig. 4.15 (c). The total area under the triggering probability curve of the reference SPAD achieved 0.74 μm at 5 V_{ex}. In the proposed device case, on the other hand, the total area was improved considerably, going from 0.45 μm at 1 V_{ex} to 1.03 μm at 5 V_{ex}. This emphasizes that the photo-generated electron-hole pairs produced in or reaching the depletion via diffusion will have a higher chance of being multiplied and detected by the circuitry.

To explain the findings more clearly, let us imagine the absorption of a photon in the middle of the depletion region in both devices. In the reference wide depletion SPAD, at x=1.2 µm, only the photogenerated holes will go through the impact ionization process in the multiplication region centered at x=1.4 µm. However, the electrons will not get multiplied due to low electric fields. Conversely, in the proposed double multiplication region SPAD at x=0.6 µm, both the holes and the electrons will be multiplied in the multiplication regions centered at x=0.75 µm and at x=0.3 µm, respectively. Due to the lack of a multiplication region for the electrons, the total avalanche triggering probability obtained for the wide depletion region SPAD example is only 33% in the middle of the device, the contribution of the only holes, whereas it reaches 77% in the proposed SPAD, as can be observed from Fig. 4.15(c), thanks to the existence of the second multiplication region. Furthermore, if an absorption outside the depletion region is considered, which is mostly the case for NIR wavelengths, the photogenerated electrons that are going to diffuse to the depletion region have 1.15× more probability to trigger an avalanche, as calculated from the edge of the depletion regions. Thus, operating double multiplication region devices at lower excess bias voltages is possible to achieve similar or higher NIR PDPs as in conventional wide depletion region SPADs.

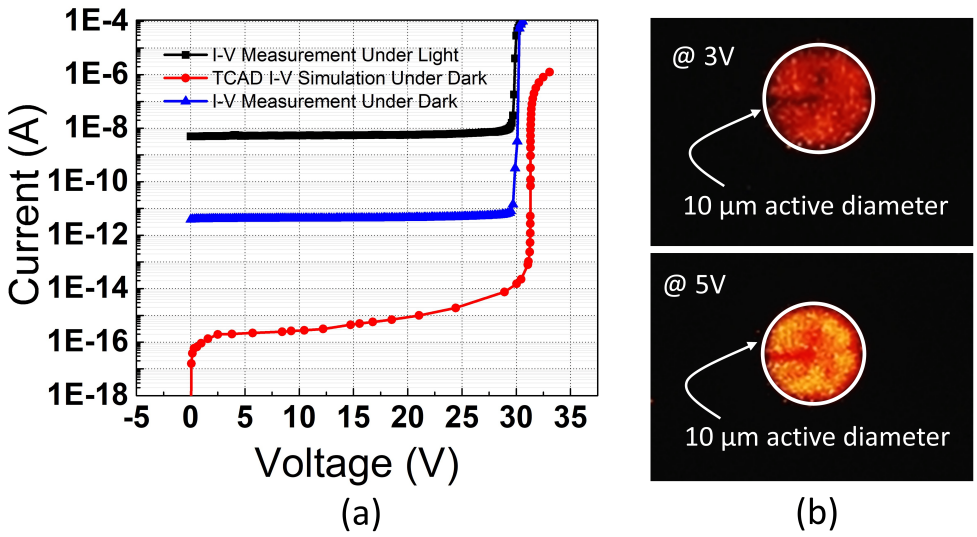


Figure 4.16: (a) Measured and simulated I-V characteristics of the proposed device. (b) Light emission tests at 3 V_{ex} and 5 V_{ex}.

4.3.3 Characterization results

In the fabricated devices with double multiplication region, the avalanche breakdown occurs at 29.8 V under both dark and ambient light conditions, as shown in Fig. 4.16 (a). The relatively high breakdown voltage is an inherent consequence of designing a wide depletion region device, as discussed before. The breakdown voltage of the device in TCAD simulations in the dark was identified as 30.8 V, as also depicted in the same figure. The small discrepancy between measurement and simulation may be due to some fabrication steps leading to

differences in the real and the numerically implemented doping depth and concentration, or due to deviations of the electron and hole impact ionization coefficients in the fabricated silicon from the theoretical values utilized in the Okuto-Crowell model. Furthermore, Fig. 4.16 (b) demonstrates the light emission tests of the device at 3 V_{ex} and 5 V_{ex}. The images prove that there is no edge breakdown in the proposed device, and the photoresponse is quite uniform over the active area. As a side note, the simulated reference wide depletion SPAD with a single multiplication region in Fig. 4.14 (d) suffered from edge breakdown after the fabrication. Therefore, characterization of that device is not possible at this point in time.

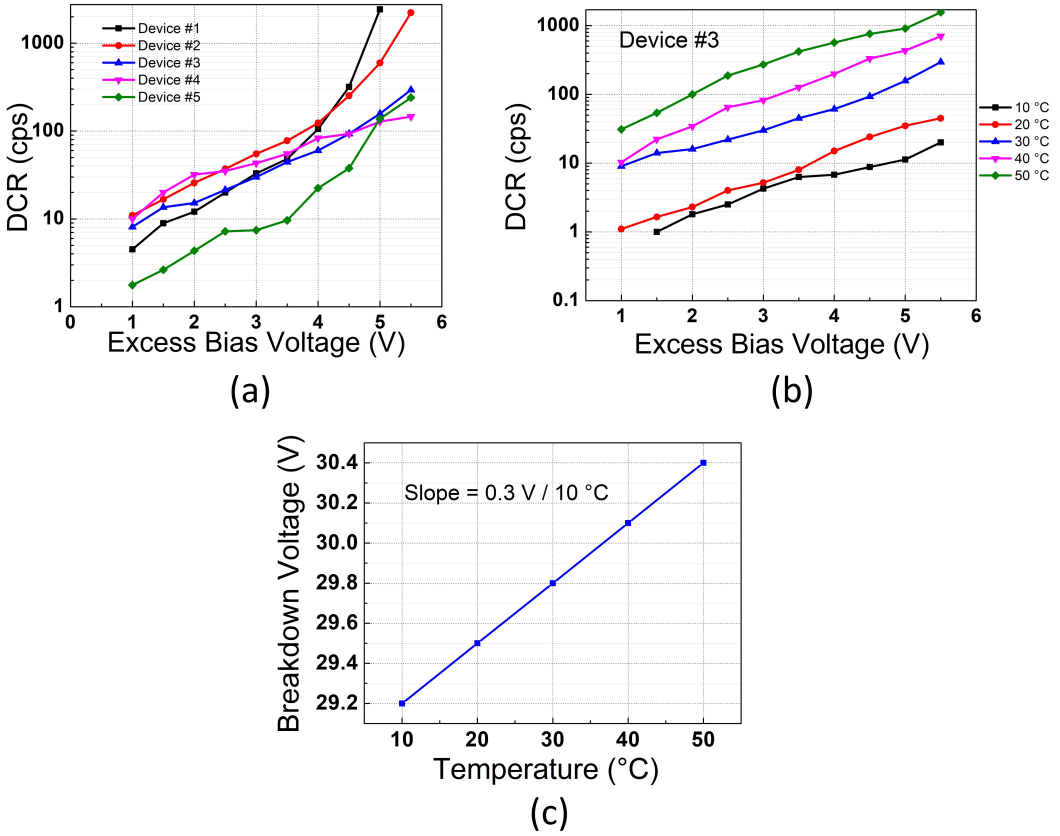


Figure 4.17: (a) DCR measurements of the device on 5 different dies for various excess bias voltages at room temperature. (b) DCR versus temperature analysis of the selected Device #3. (c) The change in breakdown voltage between 10 and 50 °C.

To operate the device in Geiger mode, it was quenched and recharged passively via a ballast resistor of 660 kΩ in free-running mode. Resistor values up to 330 kΩ were not able to quench this device; thus, a 660 kΩ resistor was utilized due to its proper functioning and availability. The generated pulses on the resistor were then counted with the same high-speed digital oscilloscope (Teledyne LeCroy WavePro 760Zi-A) from 50% of the peak magnitude level. With this configuration, the noise of the device was measured on five separate dies to verify its variation on the same wafer. DCR results at room temperature from these five dies are provided in Fig. 4.17 (a) for various applied excess bias voltages. It can be noted that

DCR is quite uniform up to 4 V_{ex}, whereas DCR might increase drastically after that. Beyond 5.5 V_{ex}, the noise increased enormously, and the devices became immeasurable. Device #3 was then selected to further characterize since it corresponds to the median of these DCR measurements.

Fig. 4.17 (b) shows the DCR of Device #3 as a function of temperature. Between 10 ^{o}C and 20 ^{o}C noise seems to be dominated by trap-assisted tunneling generation since similar DCR values were achieved at these temperatures. Starting from 20 ^{o}C, thermal generation begins to dominate, while at 30 ^{o}C, the device reaches 61 cps at 4 V_{ex} and 295 cps at 5.5 V_{ex}. The corresponding normalized DCR by the SPAD active area are 0.76 cps/μm^2 and 3.7 cps/μm^2 at the same excess biases. Due to the fact that the proposed device has two multiplication regions and enhanced avalanche triggering probability, the DCR might be higher than the reported wide depletion region SPADs at the same excess bias voltages. Above 30 °C, DCR increases significantly with thermal generation, attaining 570 cps and 1560 cps at 4 V_{ex} and 5.5 V_{ex}, respectively, at 50 °C. Breakdown voltage change with temperature is given in Fig. 4.17 (c). It was found that at each 10 ^{o}C, the breakdown voltage varies by about 300 mV, and the curve follows virtually ideal behavior according to the phonon scattering variation.

The afterpulsing histogram based on the inter-arrival time method is shown in Fig. 4.18. Because of passive quench and recharge, the dead time of the device was determined as around 7 µs. A fitting procedure based on the Poisson statistics was then performed, which is represented by the red curve on the same graph. For the fitting, the time interval between 15 and 25 µs was chosen, where the histogram follows an ideal exponential function. The APP was eventually calculated as 5.4% at the highest excess bias voltage of 5.5 V.

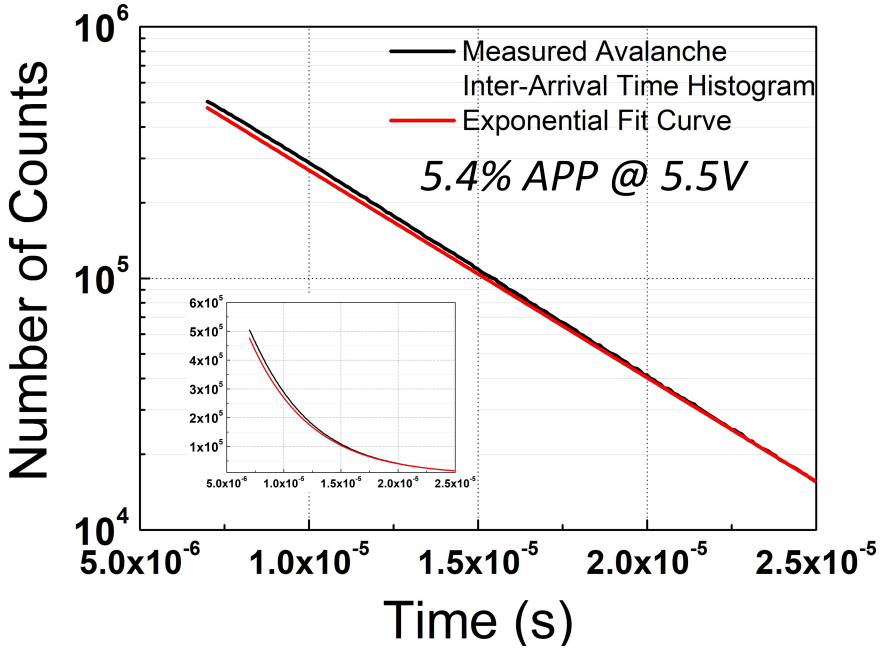


Figure 4.18: Inter-arrival time histogram of the device at 5.5 V_{ex}.

PDP measurements from 1 V_{ex} to 5.5 V_{ex} are shown in Fig. 4.19. The device was FSI, and 10-µm area measured in the light emission tests was used as the active diameter of the device to calculate PDP. As illustrated in the figure, the maximum PDP achieved at 5.5 V_{ex} is 78% at 500 nm. Since the edge of the depletion and the first multiplication regions are located very close to the surface, it favors the triggering of carriers at shorter wavelengths as well, thereby having the peak PDP at 500 nm. More importantly, high PDP was obtained at NIR wavelengths, such that they reach 25.5% at 5.5 V_{ex} and both at 800 nm and 850 nm wavelengths. This was achieved thanks to the wide depletion region formed, the non-isolated SPAD structure allowing and enhancing electron diffusion from the p-substrate with a graded doping profile, and the enhanced avalanche breakdown probability for these photogenerated diffusing electrons. Also, there are ripples all over the spectrum and some sudden PDP drops observed in the NIR. The reason for these might be the dielectric stack utilized on the SPAD in this tape-out, which is not yet fully optimized by the foundry and causes optical reflections and interferences. The same devices from different dies were double-checked, and the same ripples and peak locations were observed in other devices as well. The PDP achieved with the double multiplication region at relatively lower excess bias voltages are aligned with simulations and demonstrate the effectiveness of the technique for avalanche triggering probability enhancement. Therefore, the technique gives the possibility of addressing the high excess bias voltage needed to increase high NIR PDP.

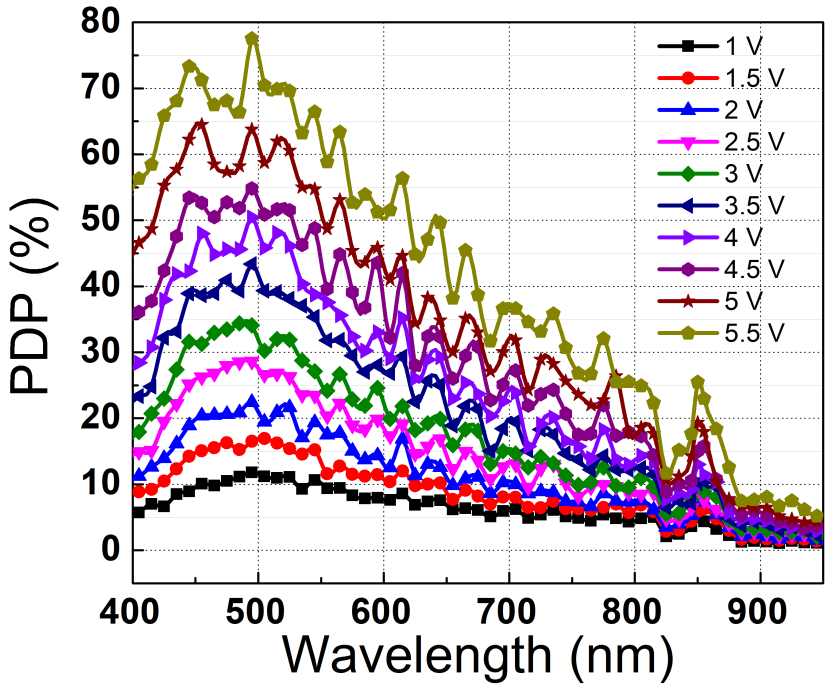


Figure 4.19: PDP measurements of the device from 1 V_{ex} to 5.5 V_{ex}.

The timing histogram of the device obtained with the same 850 nm pulsed laser is shown in Fig. 4.20. It was made sure that the device was operating in a single-photon regime (≈ 0.1 photon per pulse), while operating at 100 kHz repetition frequency. Due to the many detected diffused

electrons with the substrate non-isolated SPAD structure at NIR, a diffusion peak was observed in the timing histogram [155]. Essentially, the first peak with a smaller magnitude and width is the contribution of the depletion region, whereas the second peak and very long diffusion tails are coming from the carrier diffusion. Therefore, a deteriorated timing jitter of 240 ps FWHM was measured at 5.5 V_{ex} in this device with double multiplication region. After deconvolving the laser jitter, the jitter of the SPAD was determined as 237 ps. The acquired FW1/100M value was also 1.33 ns at 5.5 V_{ex}. Furthermore, exponential tails with different time constants were observed, which can be associated with the change in the carrier lifetimes in the substrate. The doping of the substrate is not constant; it has very low doping towards the surface for several µm, whereas it is followed by a transition region, and then doping becomes very high at deeper parts. In TCAD, only the very low-doped part of the substrate was included. Since the doping modifies the mobility, which also alters the diffusivity of the carriers, tails with different time constants might be observed. Besides, there can be an additional electric field region between the lower-doped p-epi and highly-doped p substrate, in which the electrons are accelerated before starting diffusion, which can lead to a sharper diffusion peak. The doping profile of the substrate cannot be disclosed due to foundry confidentiality.

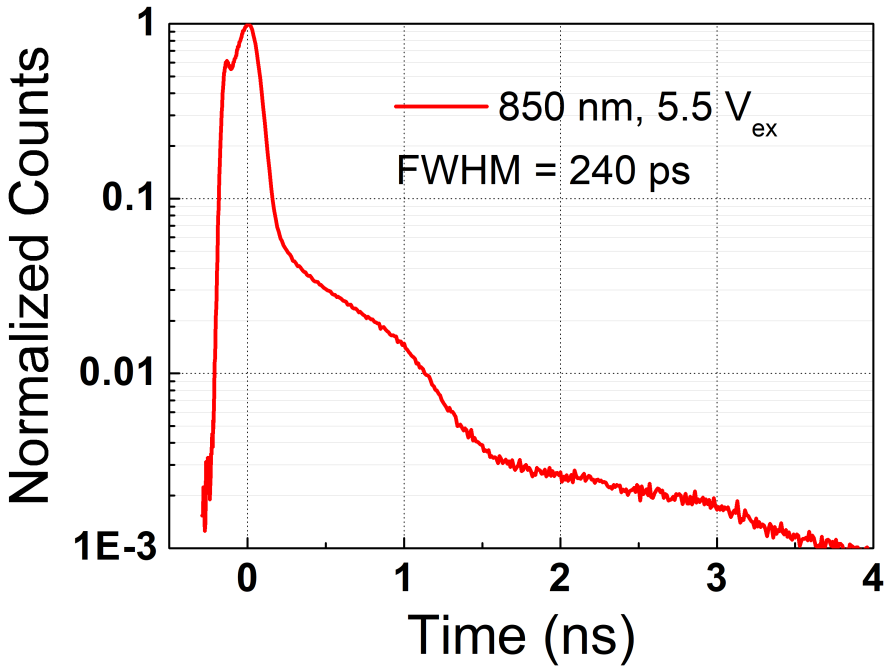


Figure 4.20: Jitter measurement of the device at 850 nm and at 5.5 V_{ex}.

To conclude, a new technique, named *double multiplication region*, was proposed to achieve high NIR PDP in the wide depletion region SPADs at relatively lower excess bias voltages. As opposed to the conventional wide depletion devices that have only one multiplication region, the technique aims at inserting a second multiplication region. As demonstrated in the avalanche breakdown probability simulations in TCAD in Fig. 4.15, second multiplication region increases the triggering probability significantly. The reason for this increase is also

explained in Fig. 4.12. To achieve the second multiplication region, a double-peaked p-well was used in junction with the HVNW layer, where two peaks result in two p-n junctions and, consequently, two multiplication regions. Between these two peaks, n-doping exceeds p-doping to cause electric field decay and separate the two distinct multiplication regions. The characterization results of this device indicate that high PDPs at NIR can indeed be obtained at 5.5 V_{ex}, reaching 25.5% at 850 nm while keeping the noise at an acceptable level of 295 cps. The peak PDP also increases significantly, attaining 78% at 500 nm. High PDPs were achieved thanks to the substrate-non-isolated and wide depletion region SPAD structure, with a higher breakdown probability for the carriers and a graded substrate doping profile. Reducing the need for high excess bias voltages was addressed owing to the enhanced avalanche triggering probabilities at a given excess bias. However, the jitter of the SPAD deteriorated to 236 ps due to the detected diffused carriers from the substrate.

4.4 A guard ring study for an isolated wide depletion region SPAD

The new doping compensation and double multiplication region concepts presented so far were demonstrated on structures that are not fully isolated from the substrate, where the active area and substrate have a direct electrical connection [156]. On the other hand, this might complicate the integration of the SPADs with the front-end circuits on the same substrate. Without electrical isolation, SPADs and circuits might affect each other's operation. The isolation can be implemented by adding buried or deep n-well layers between the active area and the substrate [156]. The addition of these layers is applicable to both the compensated device and the device with the double multiplication region, since the utilized layers are not deeper than these layers. However, PDP performance, especially at NIR for the double multiplication region SPAD, would not be the same and could be degraded. Therefore, substrate-isolated wide depletion region SPADs were also developed, and various guard ring types were studied to optimize the noise and efficiency of the devices fabricated in 110 nm CIS technology.

4.4.1 Device structures

The cross-sections of the designed substrate-isolated SPADs are shown in Fig. 4.21. The p-n junction is located between the p-epi/p^+ and HVNW layers. Thanks to the low-doped p-epi closer to the device surface and slightly doped HVNW layers, the extension of the depletion region is expected towards the deep inside the substrate, resulting in a wide depletion region SPAD. The active area of the device drawn in the layout was 10 µm. The isolation of the SPAD from the substrate was achieved through the buried n-well layer. To optimize the electric field at the edges and the noise of the SPAD, three types of GR structures were investigated: p-substrate itself, p-well, and HVPW. The dimension of the GR was 1 µm for the p-well and HVPW cases, whereas there was no additional implantation for the p-epi GR case.

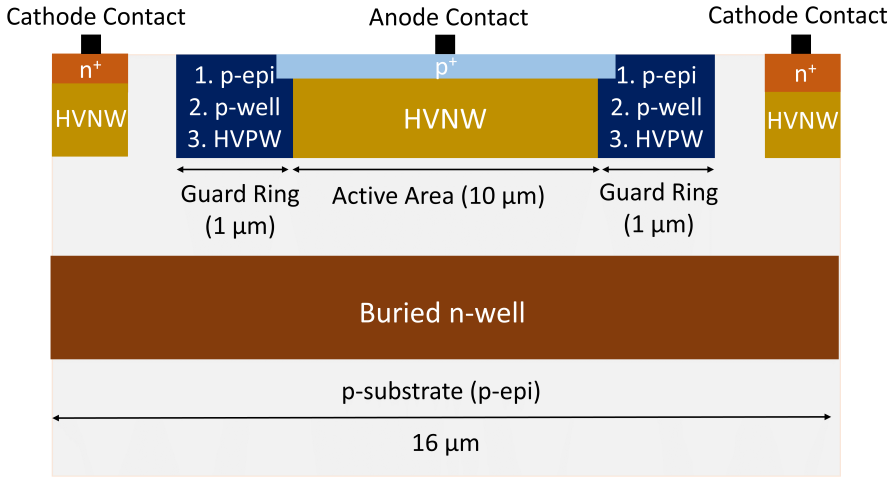


Figure 4.21: Cross-sections of the substrate-isolated SPADs.

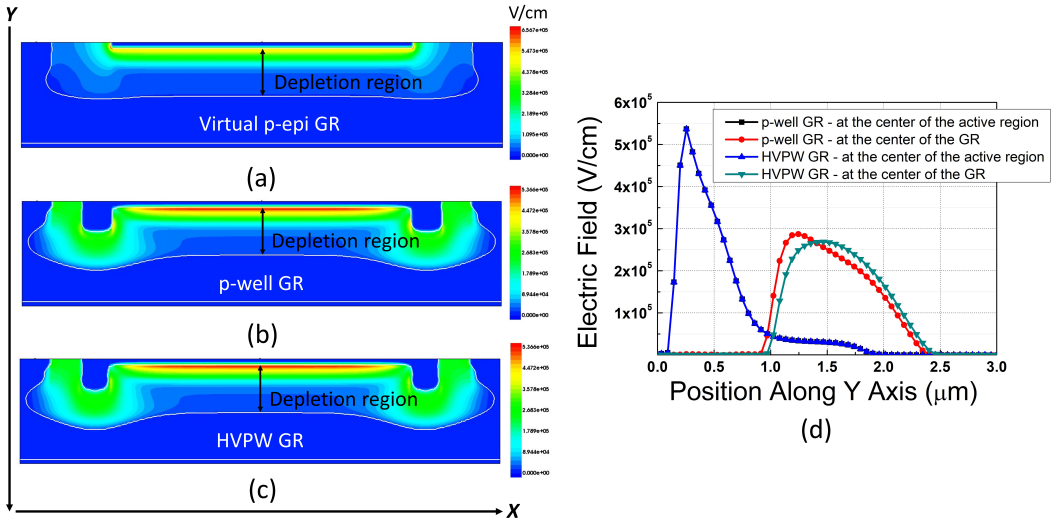


Figure 4.22: Electric field profiles of the device with (a) p-epi GR, (b) p-well GR, and (c) HVPW GR at the breakdown voltage. (d) Electric field magnitude of the devices taken at the center of the active region and GR.

4.4.2 TCAD simulation results

The simulated electric field profiles of the same device with the aforementioned GR types are provided in Fig. 4.22 (a)-(c). The first device with the virtual p-epi GR suffered from edge breakdown, as can be observed in Fig. 4.22 (a). This result indicates that simpler virtual GR structures are not always feasible in designing SPADs, and it is sometimes required that different GR structures need to be investigated. In this sense, low-doped p-well and HVPW layers were integrated into the periphery of the active region to lower the electric field magnitude at the edges, preventing the device from entering into the breakdown before the main junction. Simulation results belonging to p-well and HVPW type GRs are given in

Fig. 4.22 (b) and (c), respectively. Both structures have a depletion region width of around 1.7 µm at the breakdown voltage, and utilized GR layers seem to be effective in mitigating edge breakdown. The breakdown voltage of both devices is 19.3 V. The multiplication region shown in red is formed at the junction of p^+ and HVNW layers. The depletion region is expanded by utilizing the junction of low-doped p-substrate region and HVNW. The electric field magnitudes taken at the center of the active and GR regions along the y axis are illustrated in Fig. 4.22 (d). As expected, the electric field at the peripheries is reduced in both devices, attaining lower values than the main junction. Besides, the electric field magnitudes at the center of the junction have the same profile in p-well and HVPW GR types, leading to the same breakdown voltage of the device. In the GR regions, it is also observed that the p-well type has a slightly higher electric field due to its higher doping concentration than HVPW.

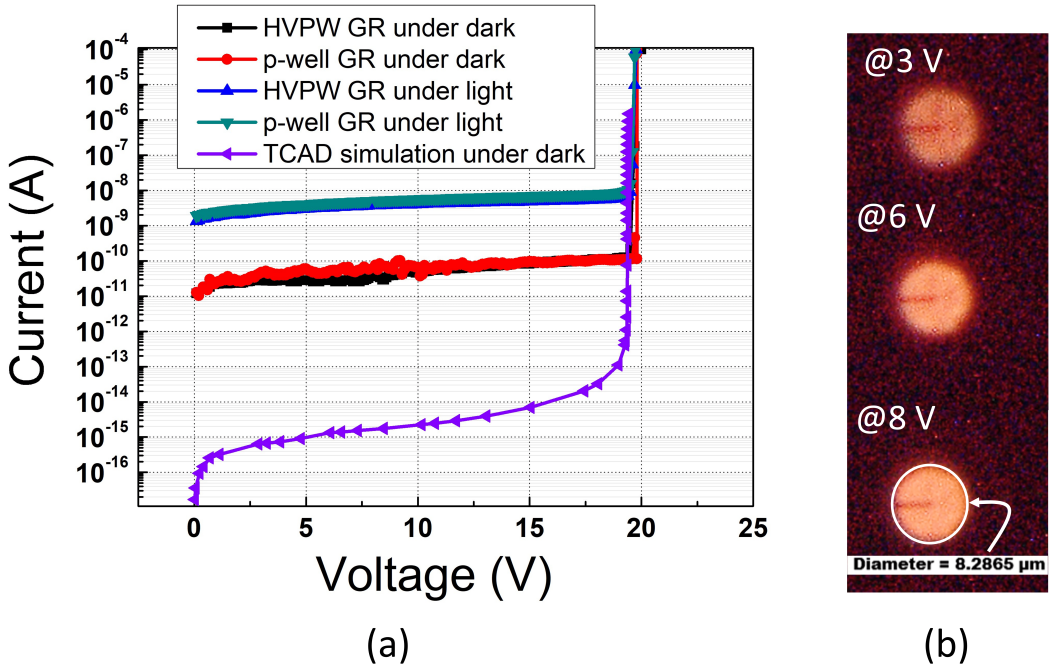


Figure 4.23: (a) I-V measurements and simulation of the SPAD with p-well and HVPW GRs under dark and light conditions. (b) Light emission tests of the device with p-well GR at various excess bias voltages.

4.4.3 Characterization results

I-V characterization of the fabricated devices with p-well and HVPW GRs is given in Fig. 4.23 (a). The breakdown voltage of both devices is 19.5 V in the measurements under light since the main junction remains the same. A relatively smaller breakdown voltage for such a wide depletion region device was achieved since the multiplication region is formed via the p^+/HVNW junction, whereas the depletion region is extended thanks to the p-epi/HVNW junction. As mentioned and illustrated, the simulated breakdown voltage was 19.3 V in TCAD, which is very close to the fabricated devices. The light emission test results of the device with

the p-well GR are provided in Fig. 4.23 (b). No edge breakdown was observed in the device, indicating the effectiveness of the GR. However, it was noticed that the active area of the device was around 8.3 µm, in contrast to the drawn 10 µm active diameter defined by the HVNW layer in the layout. Similar results were obtained for the same device on different dies and for the device HVPW GR as well. Although the reason for this deviation is not very clear, it might be related to the lateral diffusion of the dopant during the annealing procedure. In the PDP calculations in the future, 8.3 µm diameter observed in these light emission tests was thereby used to report accurate results. Furthermore, the device with the p-epi GR was also fabricated, and as TCAD simulations indicated, the device suffered from edge breakdown, whose light emission test was actually given in Fig. 2.5 (b) in Chapter 2.

To characterize the SPADs, they were quenched and recharged passively through a ballast resistor of 660 kΩ . The pulses were counted with the Teledyne LeCroy WavePro 760Zi-A digital oscilloscope from 50% of the peak magnitude level. The devices were operated in free-running mode. DCR results at room temperature from five dies of the SPAD with the p-well and HVPW GRs are depicted in Fig. 4.24 from 1 V_{ex} to 8 V_{ex}. Sample #3s correspond to the median of each device. The median DCRs are 114 cps and 5100 cps at 6 V_{ex} for the p-well and HVPW GR, respectively. SPAD with the HVPW GR is much noisier, and the standard deviation of the DCR is around 4000 cps, which is 64× larger than the standard deviation (62.5 cps) of the SPAD with the p-well GR at 3 V_{ex}. Therefore, in the remaining measurements, Sample #3 from the device with the p-well GR was utilized.

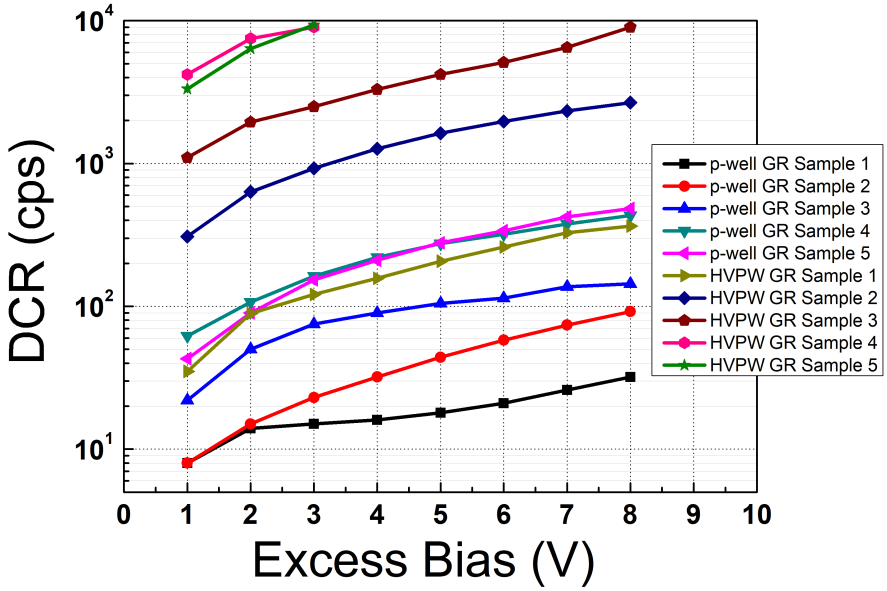


Figure 4.24: DCR versus excess bias voltage curves of the p^+/HVNW SPAD with the p-well and HVPW GRs obtained from 5 different dies at room temperature.

PDP measurements of Sample#3 from the SPAD with the p-well GR are given in Fig. 4.25. In the calculation, 8.3 µm device diameter was used, as obtained in the light emission tests. Peak PDP of the device is at 460 nm, and reaches 57% at 8 V_{ex}. The lower PDP acquired compared to

the SPAD designed with doping compensation and double multiplication region techniques is related to the fact that the p^+ contact on the active region changes the electric field direction and sweeps the holes instead of electrons to the multiplication region. Hence, the avalanche breakdown probability becomes lower due to smaller hole ionization coefficient. The PDPs of the device at 800 nm and 850 nm are, respectively, 11% and 7% at 8 V_{ex}. As explained, at longer wavelengths, the photogenerated electrons cannot reach the depletion region and undergo the impact ionization process due to substrate isolation, which limits the PDP at these wavelengths.

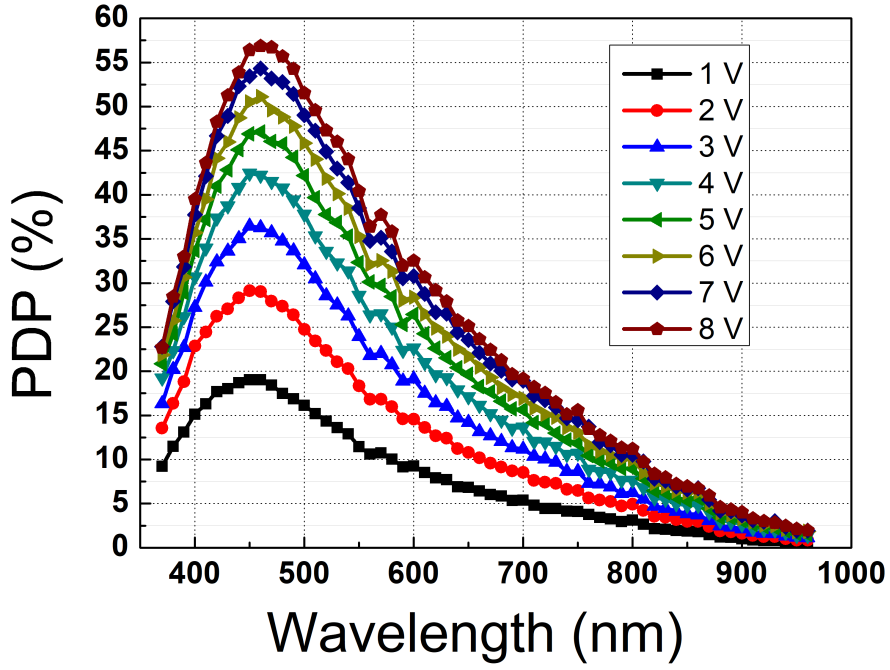


Figure 4.25: PDP spectrum of the SPAD with the p-well GR from 1 V_{ex} to 8 V_{ex}.

The timing histograms of the device with the p-well GR are demonstrated in Fig. 4.26. The measurements were taken with the same 850 nm pulsed laser, as in the previous devices. The FWHM of the timing curves are 99 ps and 71 ps at 4 V_{ex} and 7 V_{ex}, respectively. The SPAD jitter is obtained as 90 ps and 58.5 ps after the deconvolution of the laser jitter, at the same excess bias voltages. The small exponential tail with a time constant of 63 ps, as calculated from the curve at 4 V_{ex}, shows that many of the diffused carriers indeed could not reach the depletion region and were thermalized at the buried n-well layer.

To summarize, a substrate-isolated wide depletion region SPAD was also designed and fabricated. Isolation was achieved with the buried n-well implantation beneath the main junction. The multiplication region was formed between p^+ and HVNW layers. The depletion region was extended thanks to the junction with p-epi itself. Three types of GR structures were investigated to be able to operate the device and optimize its performance. Virtual p-epi GR showed edge breakdown in both simulations and light emission tests. HVPW GR yielded

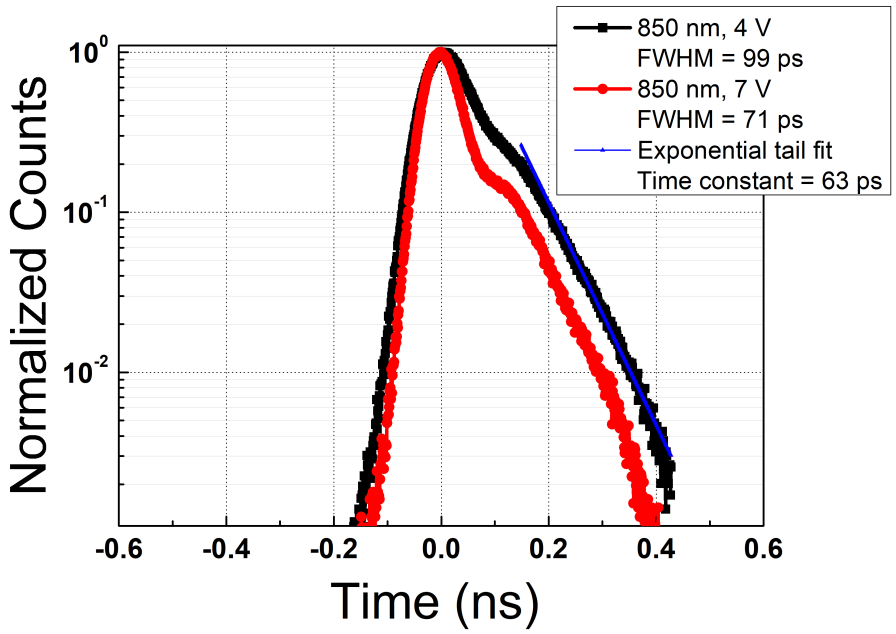


Figure 4.26: Jitter measurements of the device with the p-well GR at 850 nm.

noisy devices. Therefore, this study shows the importance of designing a proper GR to be able to operate a wide depletion region SPAD under optimum conditions. The devices with the p-well GR achieved 144 cps DCR, 57% peak PDP at 460 nm and 7% PDP at 850 nm, and 58.5 ps timing jitter at 850 nm, all at 8 V_{ex}. The lower NIR PDPs achieved, compared to the previously mentioned devices, are due to: (a) photogenerated holes with a lower ionization coefficient compared with the electrons going through impact ionization [I]; (b) the effect of the buried n-well layer, which blocks the diffused carriers from the substrate.

4.5 Comparison with the state-of-the-art

We compare the SPADs described in this section, which include doping compensation and double multiplication, and GR-optimized substrate-isolated SPADs described here with the state-of-the-art Si SPADs designed in CMOS technologies. Fig. 4.27 shows the results of the comparison, whereas FSI wide depletion region SPADs were selected [51], [53], [110], [145]–[154] as a reference. Since the aims of this thesis are to make the wide depletion region approach viable in advanced CMOS technologies and to improve performance in SPADs for the NIR wavelength detection, we focus on PDP and DCR as a function of excess bias and wavelength. In Fig. 4.27 (a), PDP at 850 nm are compared with respect to the normalized DCR with the device area. The PDP of the compensated SPAD and substrate-isolated SPAD at NIR is not greatly enhanced. For the former, it was linked to a rather thin junction and slightly deeper HVNW compensation layer, blocking the collection of diffused carriers. For the

[I] This is due to the direction of the electric field in the depletion.

latter, it was purely due to the substrate isolation not allowing carrier diffusion, despite a wide depletion region. The noise of these two devices appears to be higher than the average as well, because of tunneling and because the GR layer is still not best-suited for the goal. Yet, doping compensation can be utilized to design wider depletion regions, which can both reduce the DCR and enhance NIR PDP further. On the other hand, the SPAD designed with the double multiplication region technique achieved high NIR PDP at moderate noise levels. Higher PDP were obtained thanks to the perfectly non-isolated substrate with a graded doping profile, enabling efficient collection of the diffused carriers and augmented avalanche breakdown probabilities for the photogenerated carriers in the depletion region and electrons reaching from the substrate. The maximum achieved 850 nm PDP level in the double multiplication SPAD is closer to other demonstrated substrate non-isolated SPADs, as in Refs. [53], [145], [148]. Similar to the designed SPADs in this work, in Refs. [53], [145], [148], a doping gradient exists in the substrate, enhancing carrier migration from the substrate. Furthermore, very high PDP at NIR (≥ 30%) can be achieved with very wide depletion regions (~10 µm), as shown in Refs. [150], [152], while keeping the normalized noise at low levels. Beside, in Fig. 4.27 (b), 850 nm PDP versus V_{ex} comparison is made. As mentioned, the usual way to achieve high NIR PDP in wide depletion region SPADs is to increase V_{ex}. It is because the electric field magnitude and breakdown probabilities stay lower compared to the thin junctions. This was also the case for the doping-compensated SPAD and substrate-isolated SPAD. However, the demonstrated SPAD with double multiplication region achieved similar or higher NIR PDP at relatively lower V_{ex}, owing to the second multiplication region boosting the total avalanche breakdown probability.

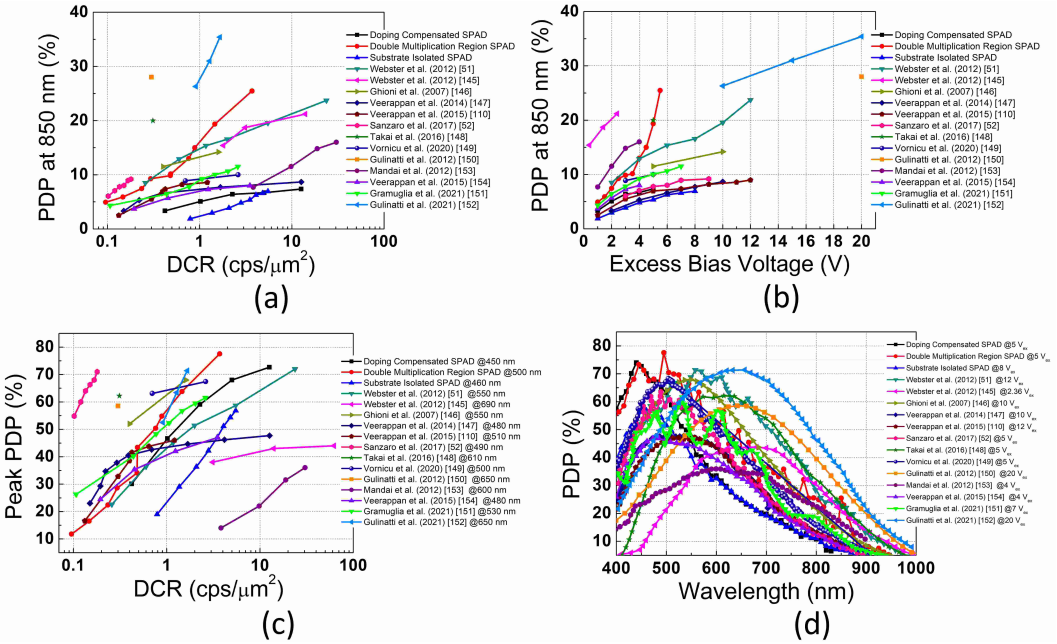


Figure 4.27: (a) PDP at 850 nm versus normalized DCR, (b) PDP at 850 nm versus V_{ex}, (c) peak PDP versus normalized DCR, and (d) PDP spectrum comparisons of the developed SPADs with FSI state-of-the-art wide depletion region SPADs.

In addition, a peak PDP versus normalized DCR comparison is given in Fig. 4.27 (c). Very high peak PDP (> 70%) was obtained with the doping compensation and double multiplication region methods. Peak PDP occurred at shorter wavelengths since the junctions are close to the device surface. Doping-compensated SPAD favors electron multiplication, whereas the double multiplication region can trigger both electrons and holes at peak wavelengths, which are the main reasons for reaching high PDP. Measured peak PDP is at the state-of-the-art level[51], [53], [146], [149], [152]. The full PDP spectrums of each device at their maximum operable V_{ex} are also provided in Fig. 4.27 (d) as a reference.

Regarding jitter, a comparison with the state-of-the-art is given, where not all of the referred devices reported jitter at the NIR wavelengths. Demonstrated jitter values at NIR are 35 ps at 820 nm [146], 60 ps at 850 nm [51], 92 ps at 850 nm [149], 93 ps at 820 nm [150], 52 ps at 780 nm [151], and 95 ps at 820 nm [152]. Doping-compensated (68 ps) and substrate-isolated SPADs (58.5 ps) provide a jitter at 850 nm that is in the range of state-of-the-art. However, the SPAD designed with the double multiplication region technique showed a degraded timing jitter of 236 ps at 850 nm due to the substrate-non-isolated structure with graded substrate doping, resulting in a diffusion peak from the many diffused electrons in the timing histogram.

5 InGaAs(P)/InP-based SPAD Development Targeting 1.06 µm and 1.55 µm Wavelengths

In this chapter, we report a study of InGaAs(P)/InP-based SPADs targeting 1.06 µm and 1.55 µm wavelengths. The motivation was to shrink down pixel sizes to around 10 µm by optimizing the planar double Zn diffusion technique while simultaneously achieving low noise and jitter and high PDP. For this purpose, the devices were first designed and analyzed in TCAD. Then, the wafers were grown, and Zn diffusion processes were performed in a foundry since the growth and Zn diffusion capabilities were not available in the school's facilities. Afterwards, the fabrication procedure for the etching and metallization of the devices was developed and finalized in the EPFL cleanroom (Center of MicroNanoTechnology (CMi)). The chapter is organized as follows: First, the modeling and simulation methodology for InGaAs(P)/InP SPADs in TCAD will be expressed. Then, the devices designed separately for 1.06 µm and 1.55 µm wavelengths will be demonstrated. The devices will also be compared with the state-of-the-art. Finally, the work for developing the integration of InGaAs(P)/InP SPAD arrays with the Si ROICs will be provided. Due to the lack of ROICs, Si fan-outs with metal pads were designed and fabricated to perform the integration process.

5.1 TCAD simulation methodology of InGaAs(P)/InP-based SPADs

Similar to the simulation of Si SPADs, coupled electron-hole continuity, Poisson, and current density equations were solved to obtain the electric field distribution, current, and electron and hole concentrations in each device. Triangular meshing was utilized, and mesh density was set to be denser where there is a material or doping concentration change, as well as in high electric field regions and junction curvatures.

Regarding the lifetimes of the carriers, the provided background doping of the n-type InGaAs absorber layer by the foundry after the epitaxial growth was around 7.5×10^{14}. Therefore, 5 µs lifetime was chosen for both electrons and holes in InGaAs, according to the study performed in Ref. [157]. The trap energy level for SRH was set at the mid-bandgap as well. The tunneling and impact ionization in InGaAs were not modeled, provided that the electric field was kept below the critical breakdown field (2×10^5 V/cm [118]) in the fabricated devices thanks to

the numerical simulations. For the $InGaAs_{0.35}P_{0.65}$ absorber with 1.1 µm cutoff wavelength, the carrier lifetimes of 20 µs extracted from the InGaAsP/InP SPAD in Ref. [98] were utilized. For the InP material, the most important region is the multiplication region; thereby, its doping concentration determines the carrier lifetimes. The background of the n-type InP multiplication region was around 6.8×10^{14} in the fabricated devices, which was also provided by the foundry. Hence, the carrier lifetimes were selected as 100 ns according to the data published in Ref. [158]. Mid-bandgap energy was used for the SRH generation as well. To include the TAT contribution in the InP avalanche region via the Schenk model mentioned in Chapter 4, 0.5 and 0.1 eV were inserted, respectively, for the Huang–Rhys factor (S) and the effective phonon energy ($\hbar\omega$), whose multiplication product defines the lattice relaxation energy [136], [138]. Besides, to compute the electro-optical frequency, the electron and hole tunneling masses were entered as 0.09 [136], [138]. The favored trap location for the TAT was set to 0.75 E_g (1 eV), which is associated with the phosphorus vacancy in InP [98], [119], [159]. All these values for the TAT modeling in InP were adopted in accordance with the previous simulation works in Refs. [98], [119], [160]. However, optimizing the lifetimes and TAT in InGaAs(P)/InP-based SPADs to match the simulated I-V with the experimental data is beyond the scope of this thesis and will be dealt with in a future work.

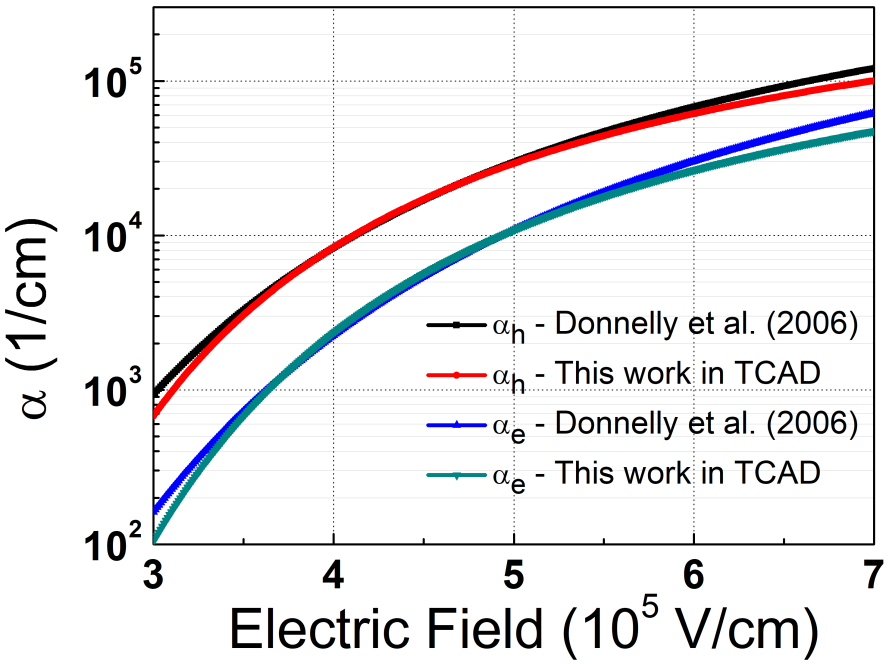


Figure 5.1: InP electron and hole impact ionization coefficients that are extracted from Ref. [98] and utilized in TCAD in this work.

In order to determine the avalanche breakdown voltage, the same Okuto-Crowell model was used as in Si SPADs [80]. However, the impact ionization coefficients for the carriers had to be modified. For this purpose, the model proposed in Ref. [98] was adopted and evaluated for the electric field range between $3\text{-}7\times10^5$, as shown in Fig. 5.1. The critical avalanche breakdown

field in InP is 5×10^5 V/cm [118]. After that, the parameters of the Okuto-Crowell formula (Eq. 4.7) implemented in TCAD were fitted to match the impact ionization coefficients of both carriers with Ref. [98]. The fitted ionization coefficients of the holes and electrons are depicted in Fig. 5.1 as well. As expected, holes have higher coefficients than the electrons as reported in the other works [83], [84]. Moreover, the same McIntyre model with the set of Eqs. 4.10, 4.11, and 4.12 was used to determine the avalanche breakdown probabilities beyond the breakdown voltage [139].

5.2 Planar InGaAsP/InP SPADs for 1.06 µm wavelength detection

The first type of InGaAsP/InP SPAD design had a 1.1 µm cutoff wavelength (1.12 eV bandgap) and aimed at detection of 1.06 µm wavelength to benefit particularly from high power 1064 nm pulsed Nd:YAG lasers for LiDAR applications. By increasing the bandgap compared to the InGaAs absorber with a 1.7 µm cutoff, thermal generation and, accordingly, the DCR were intended to be reduced, thus enabling operation at room temperature or between 0-20 °C in time-gating mode. Furthermore, they can be an alternative to Si SPADs for 960 nm detection, which recently drew a lot of attention again for LiDAR applications. In this work, various 4-inch wafers with different multiplication regions were designed. The main focus was on the device with a 10 µm active diameter and 2.5 µm GRs to shrink the pixel sizes for future array fabrication, considering LiDAR applications. However, 25 µm devices were also fabricated and measured since larger diameter devices are of interest in QKD to achieve better optical coupling efficiency. 0.25 µm and 0.5 µm depth differences between shallow and deep Zn diffusions were implemented to suppress edge effects for 10 µm devices while keeping their noise low, which allowed us to completely remove the FGRs and reduce pixel pitches. A first prototype of 1×128 pixel array was also fabricated and flip-chip bonded with in-house fabricated Si fanouts to optimize the process for integration with Si ROICs in the near future.

5.2.1 Device structures

The cross-section of the devices with the corresponding doping concentration of each epi-layer is given in Fig. 5.2. The devices were created with the planar double Zn diffusion process. The intrinsic layers' background doping layers were set by the foundry, whereas the charge layer doping was determined based on the simulations in TCAD. In the SPADs with 1.1 µm cutoff wavelength, no grading layer is shown in the epi-structure since no InGaAsP grading with several steps was intentionally grown, such as the one in InGaAs-InP interfaces. The hole barrier is very small between InP and InGaAsP ($\approx$ 0.1 eV), and it is eliminated with a natural grading that occurs while shifting from InGaAsP to InP growth, as explained by the foundry. The InGaAs etch stop layer was also inserted into the structure in the case of removing the InP substrate to extend the wavelength of operation below 920 nm, which is the cutoff wavelength of InP. SiN was used as a physical mask for Zn diffusion, and then it was left on the top surface for passivation and to provide long-term stability. The achieved peak doping concentration in

a Zn-diffused area after one diffusion process was indicated as 3×10^{18} cm^{-3} by the foundry.

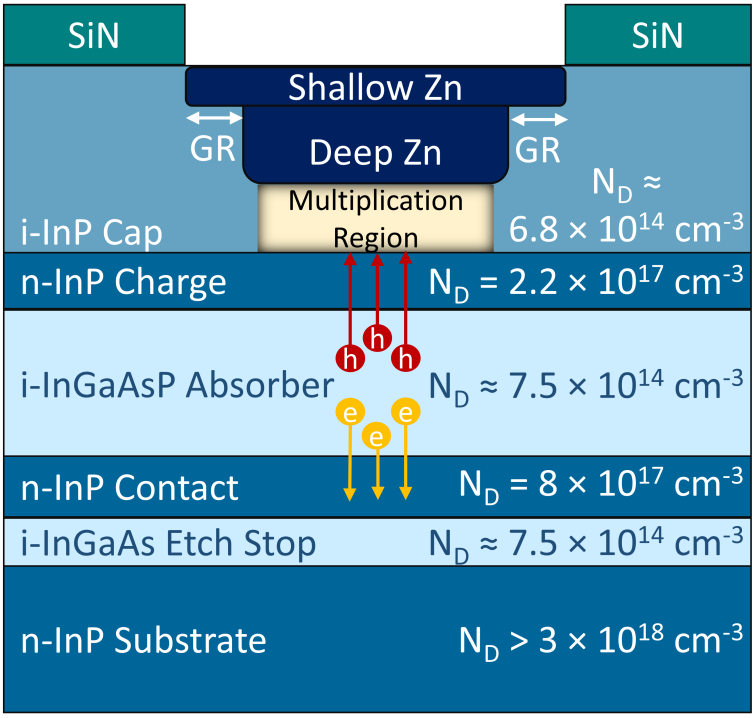


Figure 5.2: The cross-section of 1.1 µm cutoff InGaAsP/InP SPADs.

The thicknesses of the layers are as follows:

- The cap InP layer was 3 µm. By varying the diffusion time of the deep Zn, the multiplication thicknesses were planned to be 1.5 µm, 1 µm and 0.75 µm over three different wafers. For the shallow Zn, depth differences of 0.5 µm was chosen between the shallow and deep Zn diffusions for each multiplication region thickness to not suffer from drive-in of Zn atoms at GRs, which is going to be explained further later. The width of the active region (multiplication region) was also altered from 10 µm to 25 µm.
- The charge layer was 0.1 µm. The product of the charge layer thickness and the doping determines the electric field drop between the multiplication region and the absorber. It was ensured by the simulations that the InGaAsP absorber was completely depleted before the avalanche breakdown to sweep the photogenerated holes effectively.
- The thickness of the InGaAsP absorber was selected based on its absorption coefficient and Beer-Lambert law (3.1). The absorption coefficient is modeled as:

$$\alpha_{\text{InGaAsP}} = 3.6\times10^{4} cm^{-1}/eV^{1/2} \times (h\nu_{(\text{photon})} - E_{g(\text{InGaAsP})}\,(eV))^{1/2}, \tag{5.1}$$

as indicated in Ref. [98]. According to Eq. 5.1, the absorption coefficient of InGaAsP with respect to the wavelength is illustrated in Fig. 5.3. The absorption coefficient at

1.06 µm is around 7500 cm^{-1}. Therefore, the absorber thickness was chosen as 1 µm, which guarantees absorption of 53% of the photons. Enlarging the absorber will result in an increase in PDP, but thermal generation will also be enhanced due to a larger active volume, thus increasing DCR. Since close to room temperature operation is crucial for the applications, 1 µm absorber was decided in the first runs. Considering this tradeoff and the first characterization results, a thicker absorber could be implemented in the future.

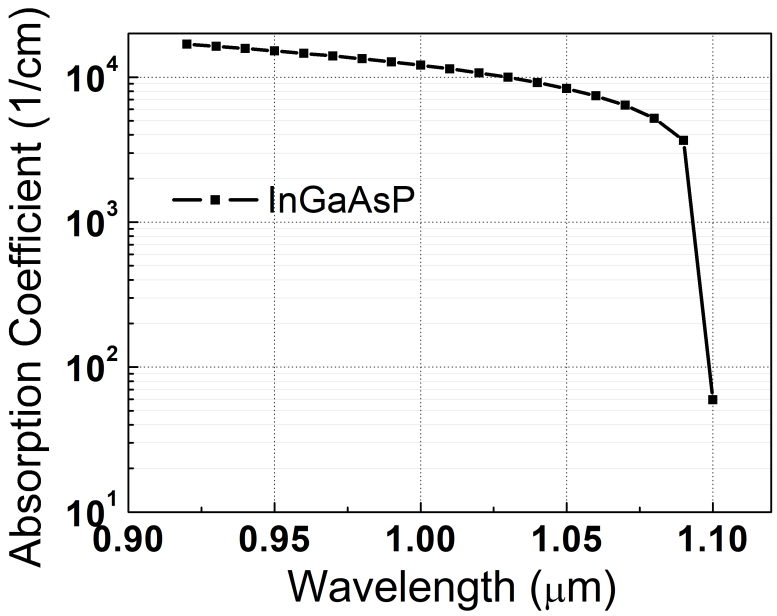


Figure 5.3: The absorption coefficient of InGaAsP with respect to wavelength.

- n-contact InP and InGaAs etch stop layers were 0.1 µm. The subsrate was 625 ± 25 µm.

5.2.2 TCAD simulation results

The defined cross-section in Fig. 5.2 was implemented in TCAD with the given doping concentrations and thicknesses. The simulation results were obtained for the SPADs with 10 µm active region diameter and 2.5 µm GR width, as they were the smallest sizes manufactured in the tape-out and future interest for a SPAD array. Although the simulations were mostly focused on the multiplication region thickness of 1.5 µm for the optimization of the layers, the electric field distribution and I-V curves of 1 µm and 0.75 µm multiplication region devices were also simulated and will be presented. Regarding the doping concentration of the Zn-diffused regions, the Zn concentrations obtained by secondary ion mass spectrometry (SIMS) measurements in Refs. [130], [161] were extracted. Both works provide Zn concentration with respect to the depth after a single Zn diffusion performed in InP. Knowing that the peak Zn concentration should be around 3×10^{18} in fabricated devices, according to the information from the foundry, extracted Zn concentrations were normalized, and the resulting Zn profile

given in Fig. 5.4 was used for each diffusion. Moreover, the edges of the Zn-diffused regions were modeled. As the edge electric fields highly affect the breakdown voltage, rectangular box-like Zn diffusion could not be utilized. Therefore, the edges of the active and GR were curved thanks to the utilized error function of factor 1.5 while defining the Zn doping profile edges in TCAD. This more realistic junction curvature matches more with the SEM images of the Zn diffusion curvature profiles demonstrated in Refs. [121], [161], allowing for more accurate results in the simulations.

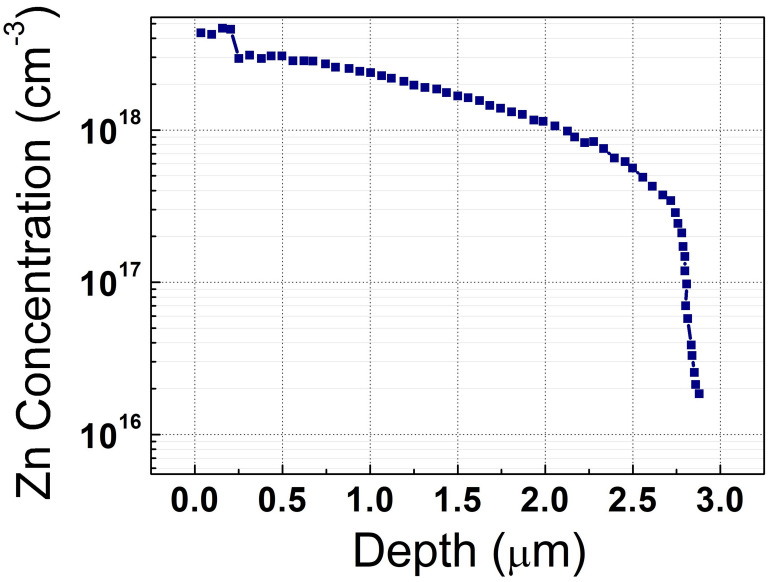


Figure 5.4: Zn diffusion profile utilized in TCAD.

The electric field simulation of the device with a 1.5 µm multiplication region and 0.5 µm depth difference between the shallow and deep diffusions is illustrated in Fig. 5.5 (a) at the breakdown voltage. The charge layer was doped 2.2×10^{17} cm^{-3}. Similar to Si SPADs, the white lines on the figure represent the depletion region boundaries. The figure shows that the high electric field is confined in the InP multiplication region. However, the edge effects are not fully mitigated, and the electric field is still slightly enhanced at the edges of the active region. Similar findings for the edge electric fields with double Zn diffusion were reported in Refs. [61], [63], [104], [120]. Furthermore, as indicated, the absorber is fully depleted at the breakdown voltage, and the depletion region is extended to the interface between the absorber and n-InP contact. To elaborate on the charge doping decision, a sweep of charge doping was performed for the device with a 1.5 µm multiplication region. Fig. 5.5 (b) demonstrates the electric field distribution at the center of the device for three different charge dopings at the corresponding breakdown voltage. As can be observed, the charge doping does not modify the field in the InP multiplication region, but the field in the absorber alters depending on the charge density that screens the electric field. The doping of 2×10^{17} cm^{-3} puts the absorber electric field closer to the breakdown field of InGaAsP ($\approx 2\times10^{5}$ V/cm), and the effect of tunneling in the absorber might increase the DCR. On the other hand, the doping of 2.5×10^{17} cm^{-3} reduces the electric field in the absorber effectively, which can eliminate the tunneling noise from the absorber

while degrading the jitter of the device. Therefore, the charge doping of 2.2×10^{17} cm^{-3} was preferred in all the devices presented in this work. Finally, the electric field distribution of the device for the multiplication region thicknesses of 1.5 µm, 1 µm, and 0.75 µm is given in Fig. 5.5 (c) at the corresponding breakdown voltage. It shows that expanding the multiplication region decreases the electric field in both the multiplication region and the absorber. In other words, lower noise is expected with a thicker multiplication region, whereas the PDP and jitter of the device will deteriorate in return. The breakdown voltage of the device with 1.5 µm multiplication region will also be the highest given that the electric in InP is the lowest.

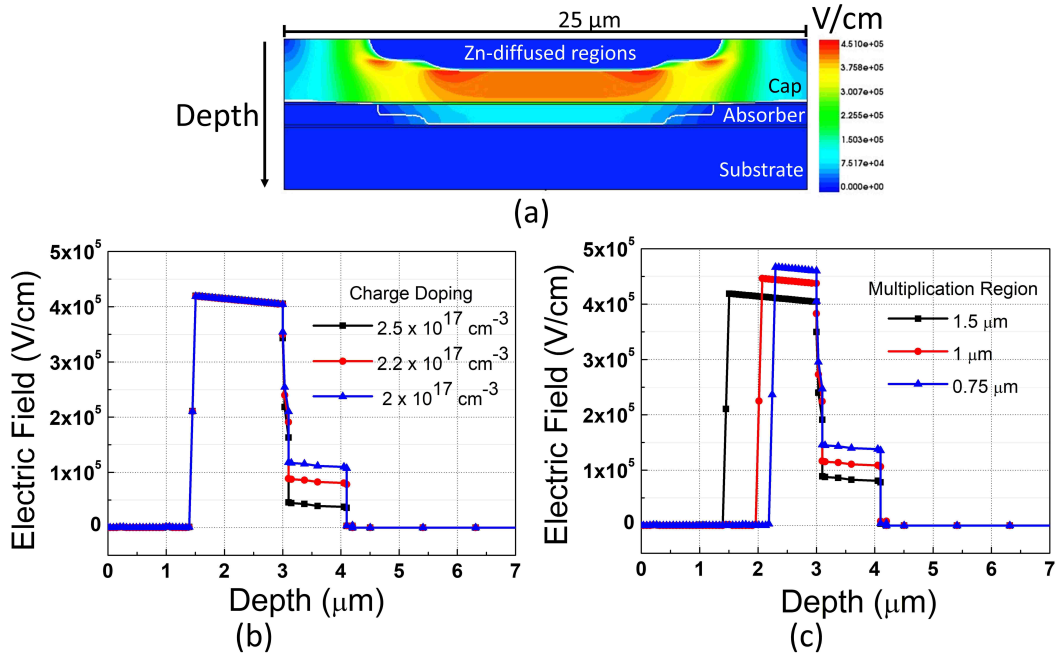


Figure 5.5: (a) The electric field distribution of the device with a 1.5 µm multiplication region at the breakdown voltage. (b) Charge doping sweep for the device with a 1.5 µm multiplication region at the breakdown voltage. (c) Multiplication region thickness sweep for InGaAsP/InP SPADs.

The dark I-V curves of the SPADs with different multiplication regions are given in Fig. 5.6. The breakdown voltages are 73.2 V for a multiplication region thickness of 1.5 µm, 57 V for 1 µm, and 50.9 V for 0.75 µm at 300K. As expected, thicker multiplication regions yield higher breakdown voltages. The punch-through voltages, where the depletion region extends to the absorber layer, can also be seen in the simulations. They appear as sudden increases in the I-V curves, which happen before the breakdown of the devices. The corresponding punch-through voltages for the same multiplication regions are 52.9 V, 34.1 V, and 26.3 V, respectively. 2.2×10^{17} cm^{-3} charge doping ensures that punch-through occurs before the avalanche breakdown, and the absorber becomes completely depleted while operating the device as a SPAD. Hence, there will not be a degradation of PDP or jitter.

The hole avalanche breakdown probabilities of the device with a 1.5 µm multiplication region

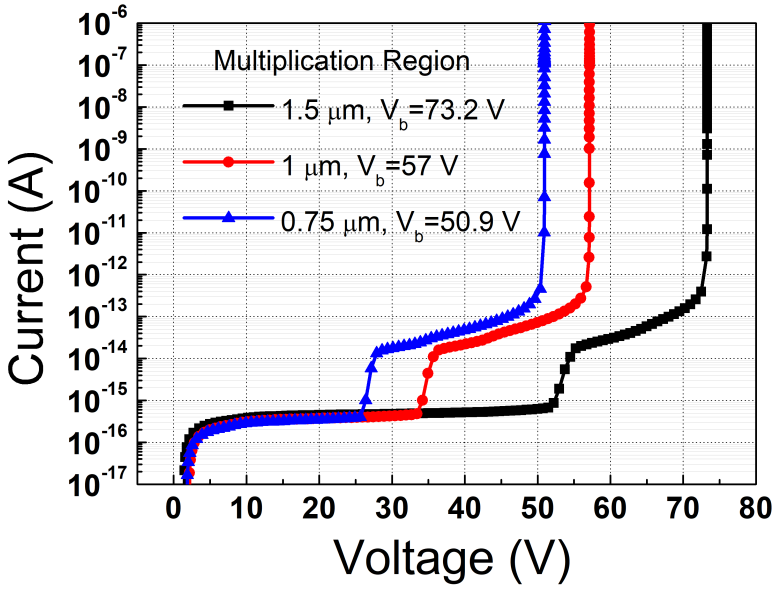


Figure 5.6: The dark I-V curves of the devices with the 1.5 µm, 1 µm, and 0.75 µm multiplication region thicknesses at 300K in TCAD.

and 0.5 µm depth difference between Zn diffusions are depicted in Fig. 5.7. In Fig. 5.7 (a), the distributions of the hole breakdown probabilities at 1 V_{ex}, 5 V_{ex}, and 9 V_{ex} are illustrated on the device cross-section. Because of the presence of edge effects, breakdown can only occur at the edges of the multiplication region at 1 V_{ex}. While increasing the bias voltage towards 5 V_{ex} and 9 V_{ex}, the electric field at the center of the device is enhanced, thus the breakdown integral reaches 1, enabling the hole breakdown at the center of the device as well. Besides, since this is an electrical simulation, it indicates hole breakdown both in the InGaAsP absorber and InP multiplication region. However, when the device is illuminated with a 1060 nm laser, the light is only absorbed in the absorber due to the bandgap restriction of InP. Consequently, the hole breakdown probability will only exist for the holes generated in the absorber. In Fig. 5.7 (b), the breakdown probabilities at the center and edge of the active region are plotted with respect to the depth. At the center of the device, hole breakdown is near zero for 1 V_{ex} and 2 V_{ex}. Starting from 3 V_{ex}, the breakdown integral can effectively become equal to 1, allowing to ionize holes at the center. Breakdown probability reaches around 0.04%, 48%, and 76% at 2 V_{ex}, 5 V_{ex}, and 9 V_{ex}, respectively. At the edges of the active region, the breakdown probability stays higher than the center for all excess bias voltages because of the junction curvature effect. For the same excess bias voltages, the probability reaches 34%, 64%, and 83% at the active region edges. The results indicate that the device response is non-uniform at 2 V_{ex}, yet the probability difference between the center and edges becomes less than 25% above 5 V_{ex}, providing more uniform photo-response over the active region.

The hole avalanche breakdown probabilities were also computed for the same 1.5 µm multiplication region device, but with a 1.25 µm shallow diffusion depth, yielding a 0.25 µm depth differences between the Zn diffusions. The results of this device are illustrated in Fig. 5.8.

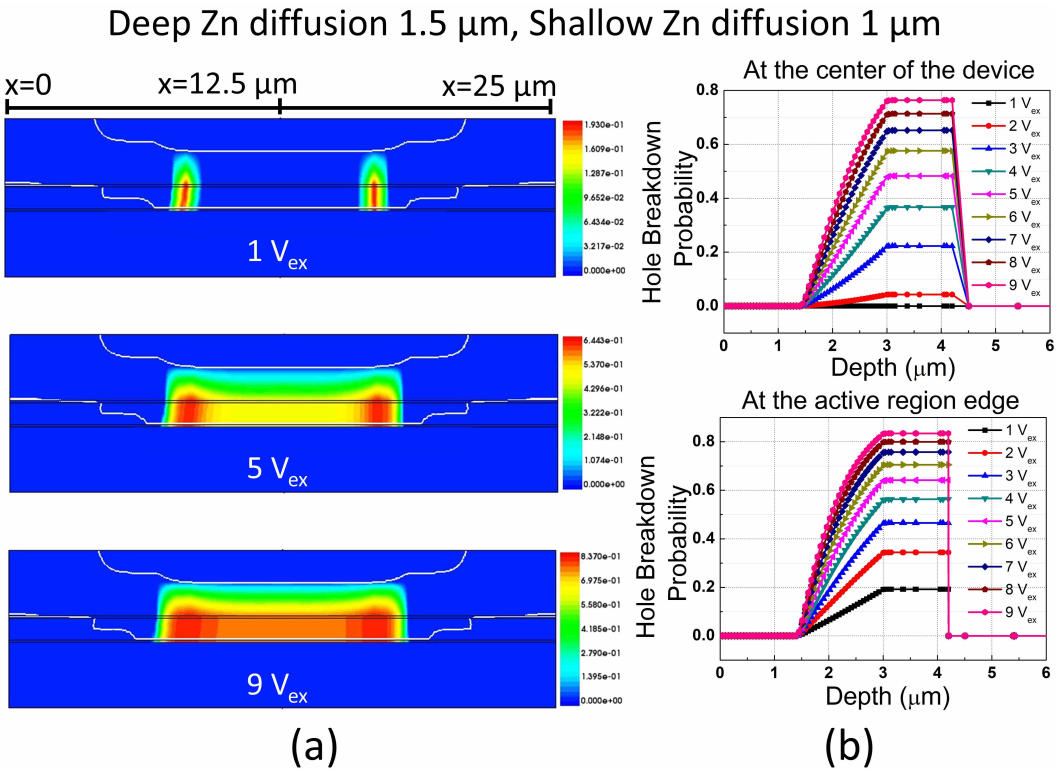


Figure 5.7: (a) Hole avalanche breakdown probability distribution of the device with a 1.5 µm multiplication region and 0.5 µm depth difference between Zn diffusions at 1 V_{ex}, 5 V_{ex}, and 9 V_{ex}. (b) Hole avalanche breakdown probabilities taken at the center and edge of the active region for various excess bias voltages.

The comparison of 0.25 µm and 0.5 µm depth difference from the Figs. 5.7 (a) and 5.8 (a) shows that more uniform device response can be achieved at the center of the device by a deeper GR depth. For instance, the device with a 0.5 µm depth difference has only 0.04% breakdown probability at 2 V_{ex}, whereas the 0.25 µm depth difference enables it to achieve 22% probability according to Fig. 5.8 (b); thus resulting in a flatter response over the active region. The probability difference between the center and edges becomes less than 15% above 5 V_{ex} in the 0.25 µm difference device, which was 25% in the 0.5 µm one. Besides, the hole breakdown probability is higher in the 0.25 µm difference device at all excess bias voltages, giving also a chance of increasing the PDP by reducing the depth difference. However, the downside of the 0.25 µm difference device is that the edges of the GR regions can enter into avalanche breakdown while increasing V_{ex}, as shown with the orange rectangle at 7 V_{ex} in Fig. 5.8 (a). Having the breakdown at the GRs can greatly increase the DCR of the SPADs, which is not desirable for particularly high temperature operations. To sum up, by increasing the shallow diffusion depth and reducing the depth difference between the GRs and active region, the edge effects can be mitigated further, providing more uniform photo-response over the SPAD area and enhanced PDP. On the other hand, there is a risk of breakdown occurring at GRs as well, which can deteriorate the noise of the SPAD.

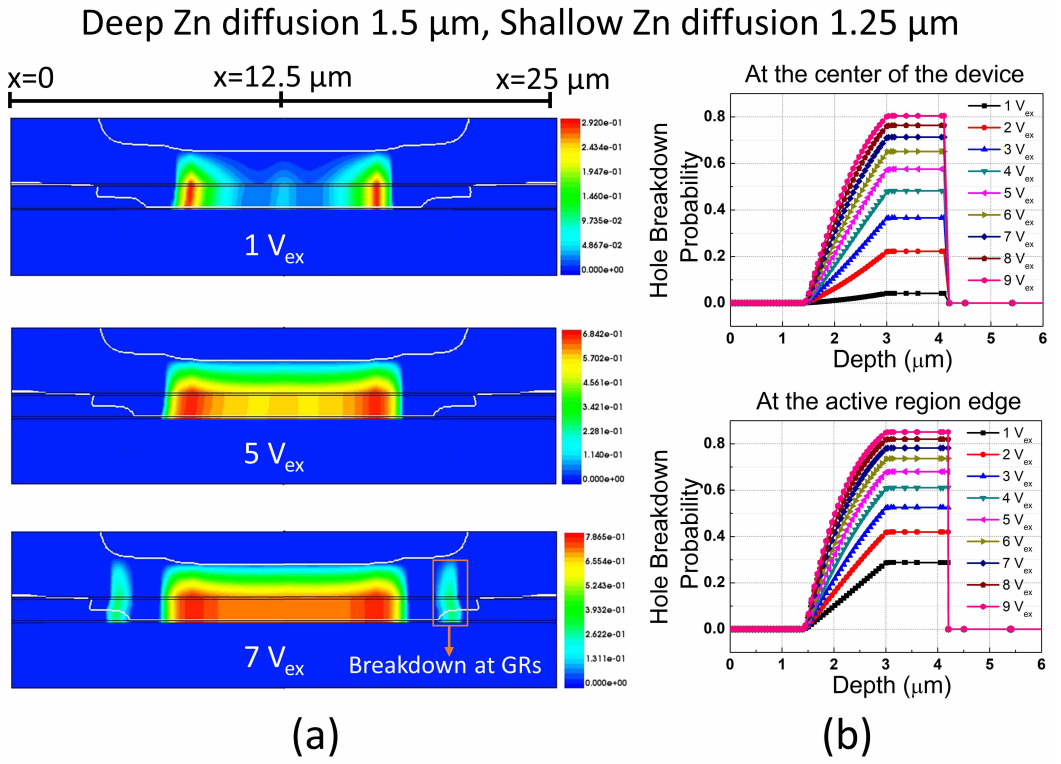


Figure 5.8: (a) Hole avalanche breakdown probability distribution of the device with a 1.5 µm multiplication region and 0.25 µm depth difference between Zn diffusions at 1 V_{ex}, 5 V_{ex}, and 7 V_{ex}. (b) Hole avalanche breakdown probabilities taken at the center and edge of the active region for various excess bias voltages.

5.2.3 Fabrication procedure

As mentioned, the growth and Zn diffusion of the devices were performed by the foundry on 4-inch wafers. To optimize the Zn diffusion process, diffusion orders, times, and depths were discussed with the foundry in two different tape-outs. The physical chromium masks were in-house manufactured in the EPFL cleanroom and were used to shape SiN before the deep and shallow Zn diffusions in the foundry process. The foundry first deposited SiN on the epi and used one of the physical masks to etch SiN. Afterwards, the Zn source was deposited on the etched region, followed by the annealing process to diffuse Zn atoms inside the epi. Exactly the same procedure was applied for the second Zn diffusion, thanks to the second physical mask. An optical image of arrays of 10 µm active diameter SPADs with a 2.5 µm GR that were defined by deep and shallow diffusions is provided in Fig. 5.9 (b). After receiving the wafers, etchings and metallizations were finalized in the cleanroom. The cross-section of the devices after the fabrication of the SPADs is given in Fig. 5.9 (a). Even though the measurements were taken in BSI mode, the structure was initially planned to be compatible with both FSI and BSI modes. Developing and testing the samples under BSI mode is important since the SPADs will be illuminated from the back after their integration with ROICs in later stages. Therefore,

the n-contact cathode metal was carried to the top surface thanks to the ladder-shaped wet etching applied in three photolithography steps [162]. To etch InP and InGaAsP layers, mixtures of phosphoric acid (H_3PO_4):hydrochloric acid (HCl) and H_3PO_4:hydrogen peroxide (H_2O_2) selective etchants were utilized, respectively. For the metallization, 250 nm/1000 nm Ti/gold (Au) were deposited with electron-beam evaporation technique. To make the SPADs FSI-compatible and keep the FSI option for the characterization if it is going to be required, the anode metal was partially touching the deep Zn-diffused area. A microscope image belonging to some fabricated text pixels is shown in Fig. 5.9 (c).

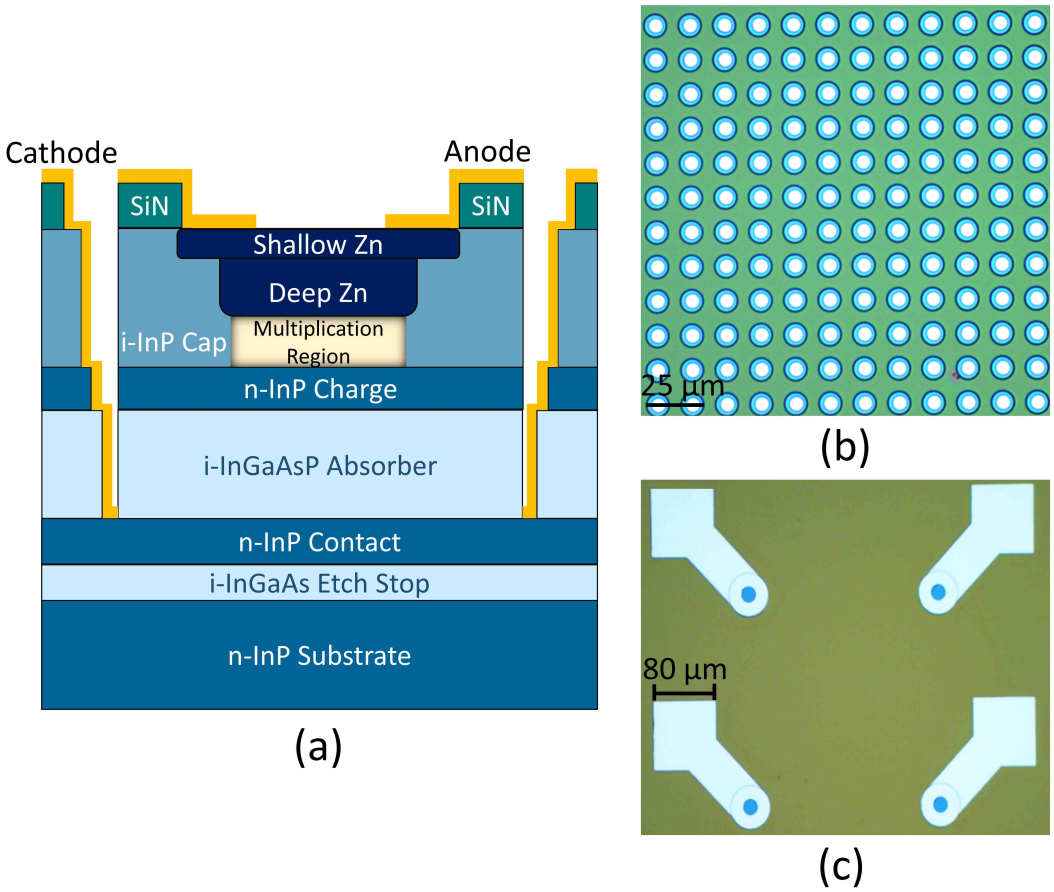


Figure 5.9: (a) The cross-section of the InGaAsP/InP SPADs after finalizing the fabrication. (b) A microscope image of arrays of 10 µm active diameter SPADs with a 2.5 µm GR after performing the deep and shallow Zn diffusion. (c) A microscope image of metallized text pixels.

5.2.4 Characterization results

First tape-out results

In the first tape-out, it was decided that the shallow Zn diffusion would be performed first, followed by deep Zn diffusion. Indeed, deep Zn diffusion depth could be controlled accurately

by performing it at last. After receiving the wafers and dicing them, SEM analysis was first conducted to verify Zn diffusion profiles. To do that, an array of SPADs with a 10 µm active diameter and 2.5 µm GR diameter was broken from multiple points with a diamond cutter, and it was expected that the broken crystal line would go through the center of at least a SPAD. As can be seen in Fig. 5.10, the broken crystal line intersected with a SPAD so that its cross-section could be analyzed via SEM. According to SEM analysis, the thicknesses of the layers were as designed. The shallow diffusion depth was planned at 1.5 µm, whereas the deep diffusion was targeted at 2 µm. The deep diffusion depth seemed to match with the planned value. However, the shallow Zn diffusion to form the GR was extended and merged with the deep Zn diffusion at the same depth, thus not constituting GRs for the active region of the SPAD. This phenomenon is called drive-in, where the Zn atoms at the GRs diffuse more during the second temperature process, although there is no Zn source at the GRs during the deep diffusion [127]. As mentioned in Chapter 3, some groups reported the existence of the drive-in process, and some did not. Here, the drive-in process is significant in this foundry process. Therefore, the structure in the first tape-out was not a success as it was going to suffer greatly from the edge effects in the case of a lack of GRs.

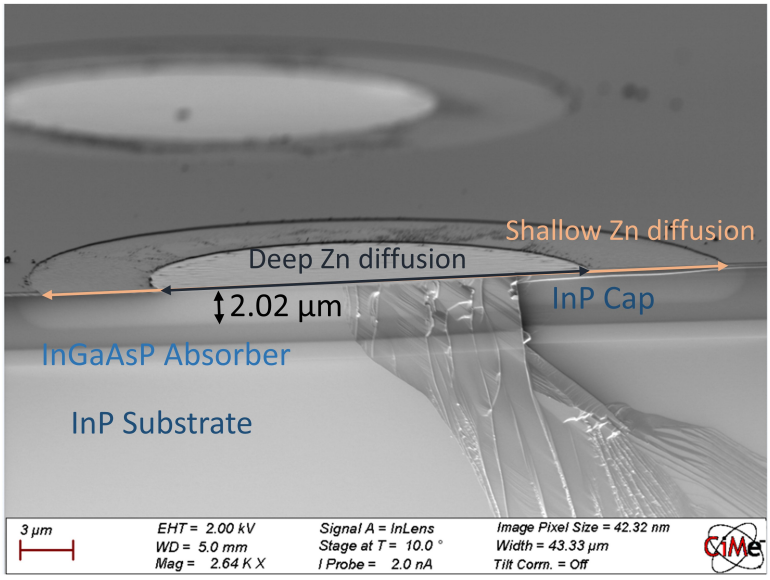


Figure 5.10: An SEM image of the cross-section of a SPAD with a 10 µm active, 2.5 µm GR diameter, and 1.5 µm shallow, 2 µm deep Zn diffusions.

Nevertheless, some characterization results from the first tape-out were obtained, which are given in Fig. 5.11. Dark and light-on I-V curves of the device with a 1 µm multiplication region, 10 µm active, and 2.5 µm GR diameter can be observed in Fig. 5.11 (a). The breakdown and punch-through voltages were identified as 59.5 V and 23.6 V, respectively. The multiplication region of this device was around 1 µm, as indicated in the SEM image. The breakdown and punch-through voltages are deviated from the simulated SPAD with 1 µm multiplication region in TCAD. It is observed in the fabricated SPAD that the breakdown voltage is higher, whereas the punch-through voltage is lower than in simulations. The deviation in the breakdown voltage

might be related to the modeling of the impact ionization coefficients. Also, the decrease in the charge doping density or the background doping of the InP cap layer in the fabricated device can cause this deviation [119]. However, the reason for the big difference in the punch-through voltages is not exactly known yet but might also be related to the background doping of the InGaAs absorber. Therefore, more precise control in the foundry process might be needed to adjust the charge doping density or the background doping of the layers to achieve the desired performance.

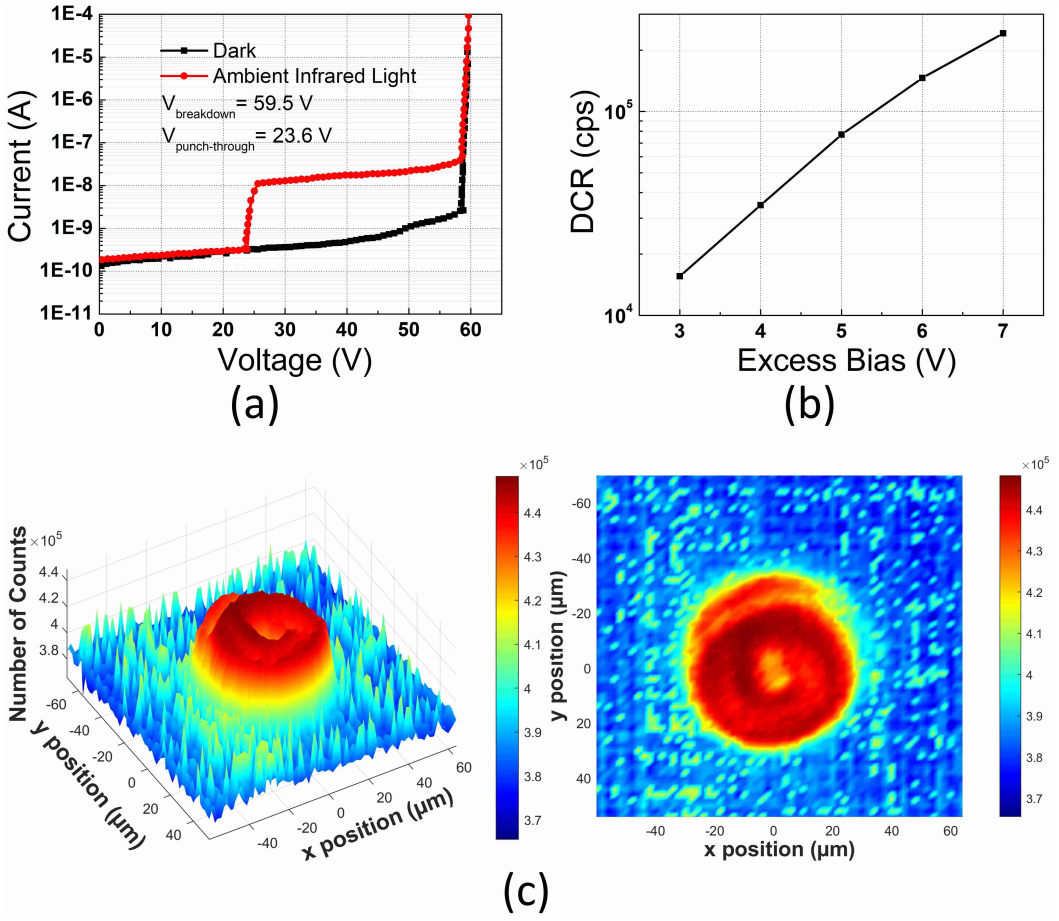


Figure 5.11: (a) I-V curves of a SPAD with a 10 µm active and 2.5 µm GR diameter. (b) DCR measurement of a 10 µm device at room temperature with 10 kHz gating frequency and 100 ns gate-on time. (c) Active area scanning of a SPAD with a 25 µm active and 40 µm GR diameter with a 1060 nm pulsed laser at 5 V_{ex} and room temperature.

DCR measurements obtained at room temperature and in gated mode with 10 kHz operation frequency and 100 ns gate-on time are provided in Fig. 5.11 (b) for the 10 µm device. All the obtained DCR values were normalized by the gate-on time, as described in Eq. 2.5. The DCR reaches around 70 kcps at 5 V_{ex}. In Fig. 5.11 (c), an active area scanning of a device with a 25 µm active diameter and 7.5 µm GR was performed at room temperature and 5 V_{ex} by utilizing the setup configuration illustrated in Fig. 2.8 in Chapter 2. A 1060 nm pulsed laser

was synchronized with a 10 kHz gating frequency and 100 ns gate-on time using a waveform generator. The avalanche pulses were read from a high-speed digital oscilloscope. The scanning results indicate that the center of the device was indeed about 10% less responsive to the light compared to the edges of the active region. The electric field and hole breakdown probabilities were enhanced at the edges of the shallow diffusion region, thus causing a non-uniformity in the device response even at 5 V_{ex}. This result is all associated with the drive-in phenomenon in the shallow diffusion regions pushing the diffusion depth to the deep diffusion front, hence preventing effective GRs around the active region. Because of these characterization results, a second tape-out was made to fix the Zn diffusion problem and achieve a uniform response over the SPAD area.

Second tape-out results

The discussions with the foundry to solve the Zn diffusion problem led to interchanging Zn diffusions to be able to achieve the desired profiles for double Zn diffusion. Hence, in the second tape-out, the deep Zn diffusion was performed first, followed by the shallow diffusion. In this way, there would not be a drive-in phenomenon occurring during the shallow diffusion, since it is isolated. However, the drive-in would happen for the deep Zn diffusion in the second annealing process. Therefore, the diffusion time for the deep Zn diffusion was also reduced to compensate for the drive-in process, such that the total diffusion time (deep+shallow) would yield the planned diffusion depth for the deep diffusion. Thus, changing diffusion orders gives the possibility of mitigating the drive-in in GRs and providing the desired double Zn diffusion profile, but controlling deep Zn diffusion depth and multiplication region thickness becomes more challenging.

As in the first tape-out, arrays of 10 µm diameter devices with 2.5 µm GR were broken with a diamond cutter to access the cross-section. SEM images of three different devices are given in Fig. 5.12. As indicated in the SEM images, the pictures were taken with in-lens detectors and a 3 kV accelerating potential. Thanks to changing Zn diffusion order, correct double diffusion profiles were obtained in the second tape-out, as observed in the images. The devices designed to achieve a 1.5 µm multiplication region (Fig. 5.12 (a)) seemed to have the correct deep diffusion depth of 1.5 µm, but with a rather shorter shallow Zn diffusion depth of 0.86 µm (1 µm in the design). The devices designed to have a 1 µm multiplication region (Fig. 5.12 (b)) resulted in a 1.7 µm deep diffusion, corresponding to a 1.3 µm multiplication region. The shallow diffusion depth of these devices was measured as 1.16 µm. The last structure designed with a 0.75 µm multiplication region (Fig. (5.12 (c)) achieved the correct Zn diffusion depth of 2.25 µm, whereas the shallow diffusion depth was 1.91 µm, which is 0.15 µm deeper than the designed depth. In conclusion, the desired double Zn diffusion profiles, where depth is deeper in the active region and shallower in GRs, were achieved in the second tape-out, but the diffusion depths were sometimes deviated from the planned values.

The I-V curves of the devices are provided in Fig. 5.13. The dark I-V curves were obtained for 10 µm and 25 µm active diameter devices with 2.5 µm GR width. The photocurrents were

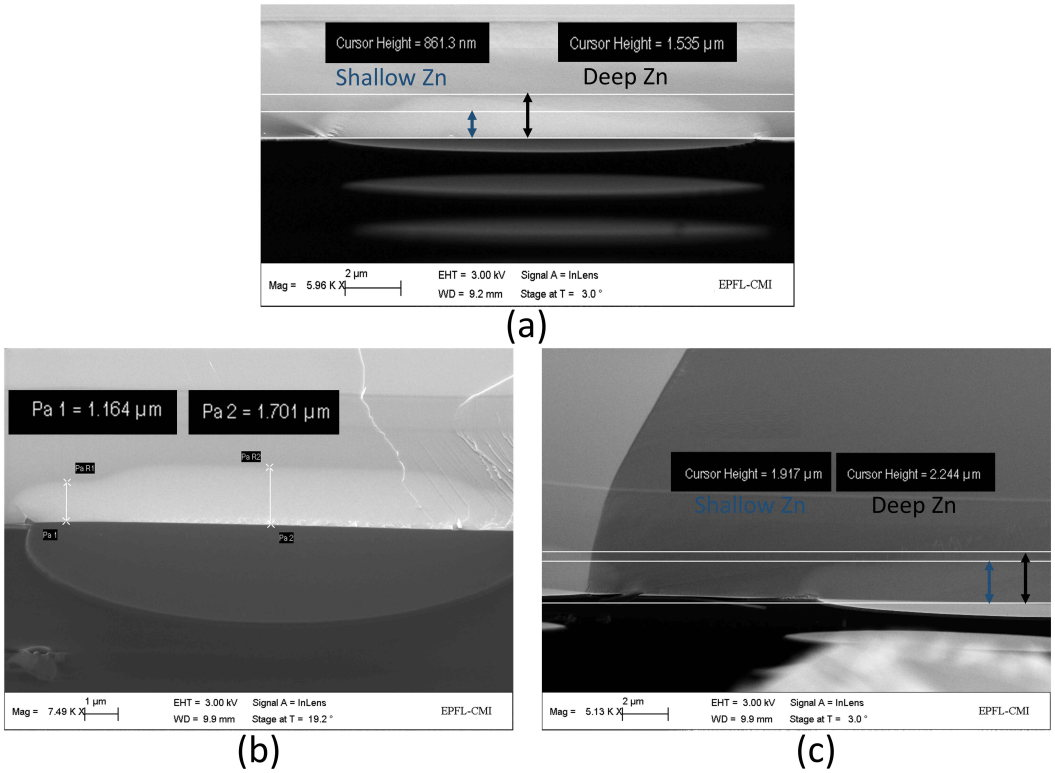


Figure 5.12: Double Zn diffusion profiles of devices with (a) 1.5 µm deep and 1 µm shallow diffusions, (b) 2 µm deep and 1.5 µm shallow diffusions, and (c) 2.25 µm deep and 1.75 µm shallow diffusions.

acquired from 10 µm devices under infrared light. The breakdown voltages in 10 µm devices are smaller than 25 µm devices in each multiplication region. This can be associated with the deep diffusion depth difference of the Zn atoms while shrinking the mask openings; smaller mask sizes result in increased Zn diffusion depth [104]. The breakdown voltage of the SPADs with a 1.5 µm multiplication region is 73 V for 10 µm active diameter. This is only 0.2 V less than the simulated value in TCAD. The punch-through voltage of this device is 27.2 V, which is almost 26 V less than the simulated value. The reason for the big difference in the measured and simulated punch-through voltages is yet unknown at this point in time; however, it might be related to the difference in the InGaAs background doping value utilized in the simulations and in the fabricated devices. The breakdown and punch-through voltages in 10 µm SPADs with a 1.3 µm multiplication region are 61.4 V and 17.2 V. Lastly, these voltages shrink to 54.2 V and 10.3 V for the same size SPADs with a 0.75 µm multiplication region. Observing the punch-through voltage guarantees that the absorber is depleted in each device, preventing degradation in the jitter and PDP of the SPADs. On the other hand, low punch-through voltages mean that the electric field in the absorber is enhanced, which might trigger tunneling noise in InGaAsP as well and increase the DCR. As a final note, the variation of the breakdown voltage with respect to temperature was changing between 0.75 V/10K and 0.9 V/10K among the devices for each wafer.

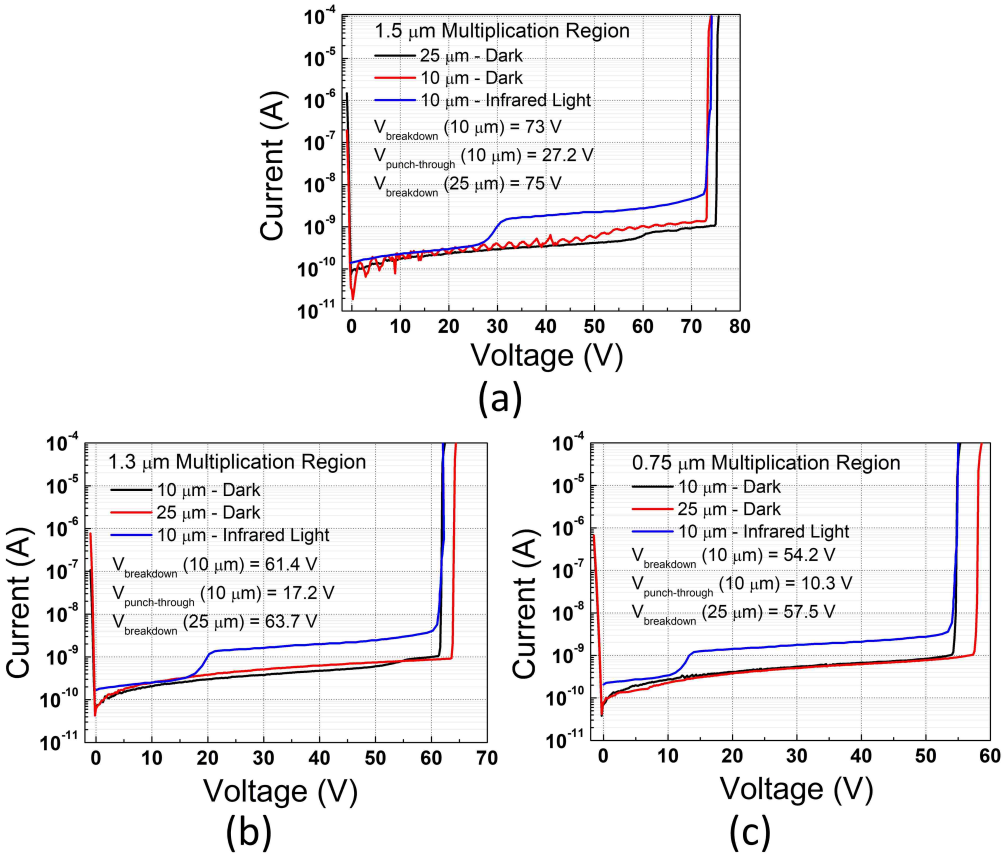


Figure 5.13: I-V curves of the devices with (a) 1.5 µm multiplication region, (b) 1.3 µm multiplication region, and (c) 0.75 µm multiplication region.

Active area scanning of a 10 µm diameter SPAD having a 1.3 µm multiplication region is illustrated in Fig.5.14. The measurements were taken with a 1060 nm pulsed laser at room temperature. The SPAD was biased with the circuit presented in Fig. 2.4 (b), while 10 kHz gating frequency and 100 ns gate-on time were applied to the AC port of the bias tee, and the DC port was set to 1 V below the breakdown voltage. The SPAD was quenched and recharged passively with a 50 kΩ external resistor. The device was illuminated from the backside, and the focused beam size was around 1 µm. The side and top views of the scanning results acquired with 1 µm step set on the motorized stage and from 2 V_{ex} to 8 V_{ex} can be seen in Fig. 5.14 (a) and (b), respectively. Along with the hole breakdown probability simulations in Fig. 5.7, edge effects were observed in the fabricated SPADs, particularly at low excess bias voltages. At 2 V_{ex}, the SPAD's response was low at the center of the active region, and edge effects were dominating. With the increase in the voltage, the electric field and hole breakdown probabilities at the center were enhanced, providing a more uniform response. Unlike the first tape-out, there was no non-responsive spot in the middle of the active region. This was achieved by better control of shallow Zn diffusion depth, leading to effective GRs around the active region's periphery. Moreover, the SPAD's response became almost flat over the active area at 6 V_{ex}. Also, similar results in the active area scanning of a 10 µm SPAD with a 1.5 µm

multiplication region were obtained.

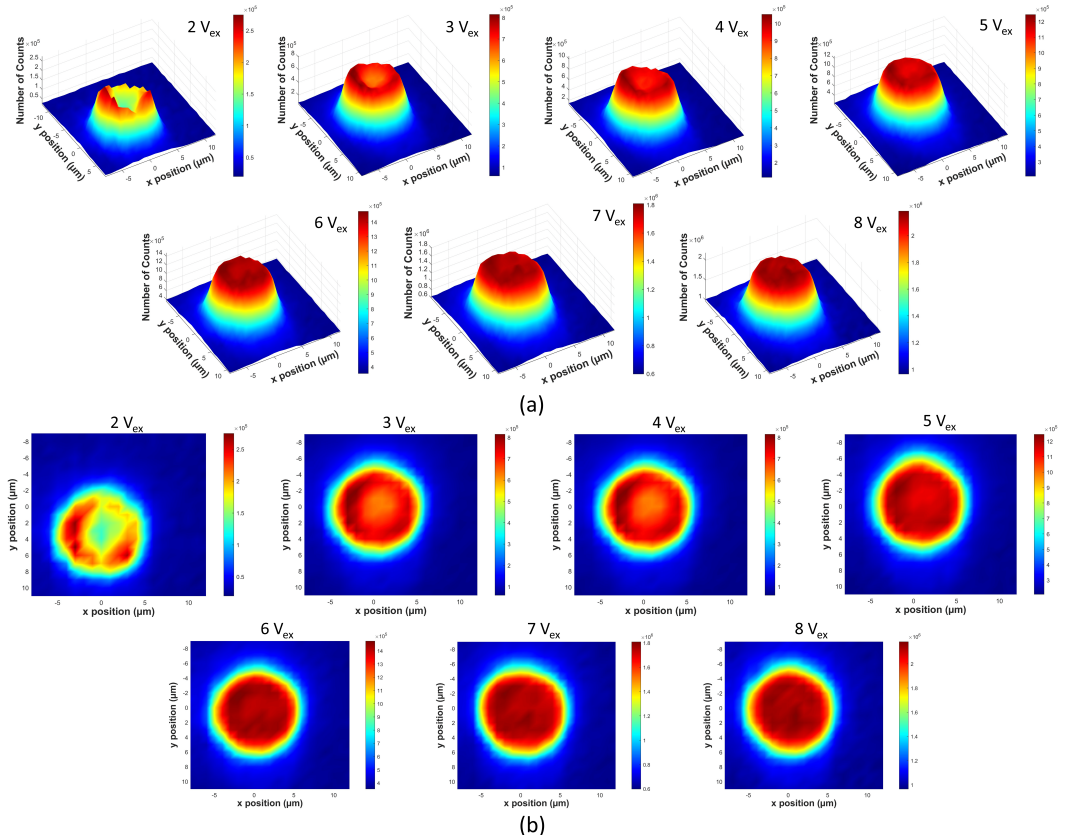


Figure 5.14: Active area scanning of a 10 µm diameter SPAD with a 1.3 µm multiplication region shown from the (a) side view, (b) top view from 2 V_{ex} to 8 V_{ex}.

After demonstrating an achievable uniform photo-response, the gating frequency sweep was performed for 10 µm diameter devices at various temperatures to identify the afterpulsing-free operating frequencies. While decreasing the gate-off time of the SPADs, the neighboring gates tend to fire more avalanche pulses if the afterpulsing effect is significant. This can be determined by the increase in the DCR of the SPADs. DCR versus the gate-off time plots of each device are given in Fig. 5.15, where the SPADs were biased at 5 V_{ex} with 100 ns gate-on time. For each multiplication region thickness, the devices showed low afterpulsing-probability, for up to 400-500 kHz frequency (2-2.5 µs gate-off time) at room temperature. Increasing the gating frequency above these frequencies resulted in a substantial increase in afterpulsing and DCR. To indicate the effect of gate-on times on the afterpulsing, 20 ns gate-on time was also checked for the 1.5 µm multiplication region device at room temperature, as depicted in Fig. 5.15 (a). Utilizing shorter gate-on times lowered the afterpulsing probability thanks to the lower amount of charges flowing during an avalanche, which can be observed with the decrease in the DCR at higher frequencies. Furthermore, the effect of temperature can be seen in each graph as well. As expected, the maximum operable frequency of the SPADs decreased at lower temperatures since the carrier lifetimes prolonged, enhancing the firing

of an avalanche at the subsequent gates. The devices tended to suffer from high afterpulsing around 100-200 kHz gating frequency at 225K. Thanks to the findings in these measurements, 10 kHz gating frequency was utilized in the rest of the characterization of each device to suppress afterpulsing alltogether.

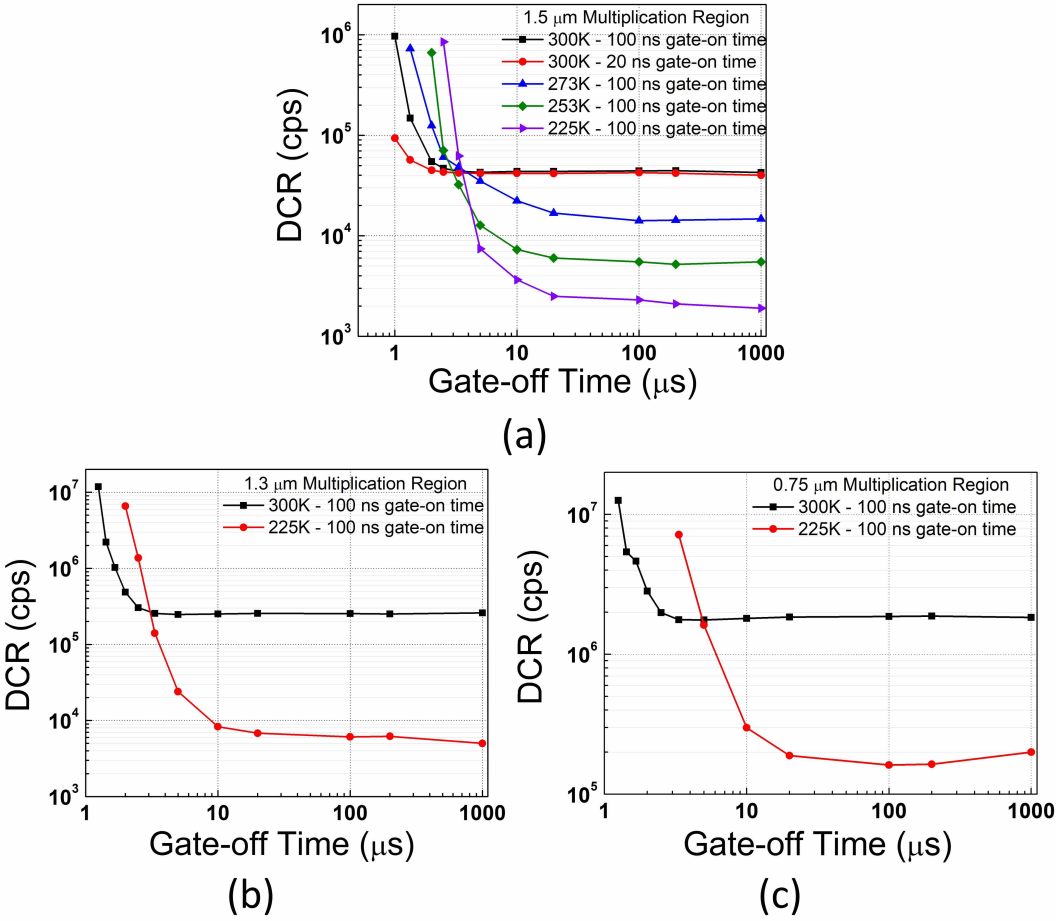


Figure 5.15: Gate-off time sweeps of 10 µm diameter devices with 2.5 µm GRs and (a) 1.5 µm multiplication region, (b) 1.3 µm multiplication region, and (c) 0.75 µm multiplication region at 5 V_{ex} and various temperatures.

The DCR of 10 µm and 25 µm diameter devices with 2.5 µm GRs for each multiplication region and at various temperatures are illustrated in Fig. 5.16. Including the DCR, in all of the remaining measurements, devices were operated at 10 kHz gating frequency and with 100 ns gate-on time while the DC bias was 1 V below the breakdown, and were quenched and recharged with a 50 kΩ ballast resistor. Similarly, all the reported DCR values were normalized by the gate-on time to account for the total SPAD ON time according to Eq. 2.5. Three devices were measured from 10 µm active diameter at 300K, and then Sample #2s (S2) were further characterized in the rest of the work for each multiplication type since they correspond to the median. At 300K and 5 V_{ex}, the median DCRs were 53 kcps for a multiplication thickness of 1.5 µm, 302 kcps for 1.3 µm, and 2130 kcps for 0.75 µm in 10 µm devices. Regarding 25

µm diameter devices, the DCR increases by around 2.8× compared to 10 µm ones, which means that the DCR does not exactly scale with the area of the device. Lower DCR per unit area was observed in the larger diameter devices. The DCR of the 10 µm SPAD with 1.5 µm multiplication region at 5 V_{ex} reduces to 14.1 kcps at 273K, 5.5 kcps at 253K, and 2.75 kcps at 225K, owing to decreased thermal generation. The same devices with 1.3 µm and 0.75 µm multiplication regions attained 23 kcps and 526 kcps DCR at 253K and 5V_{ex}. As expected, SPADs with a 0.75 µm multiplication region are the most noisy devices, since the electric field magnitude in the active volume is the highest, leading to higher hole breakdown probability and enhanced tunneling noise in InP. Due to tunneling, the reduction in the DCR is less dependent on temperature for the 0.75 µm multiplication region SPADs. SPADs with 1.5 µm and 1.3 µm multiplication thicknesses provide low-noise SPADs at room temperature and also at temperatures that can be achieved by utilizing a thermoelectric cooler.

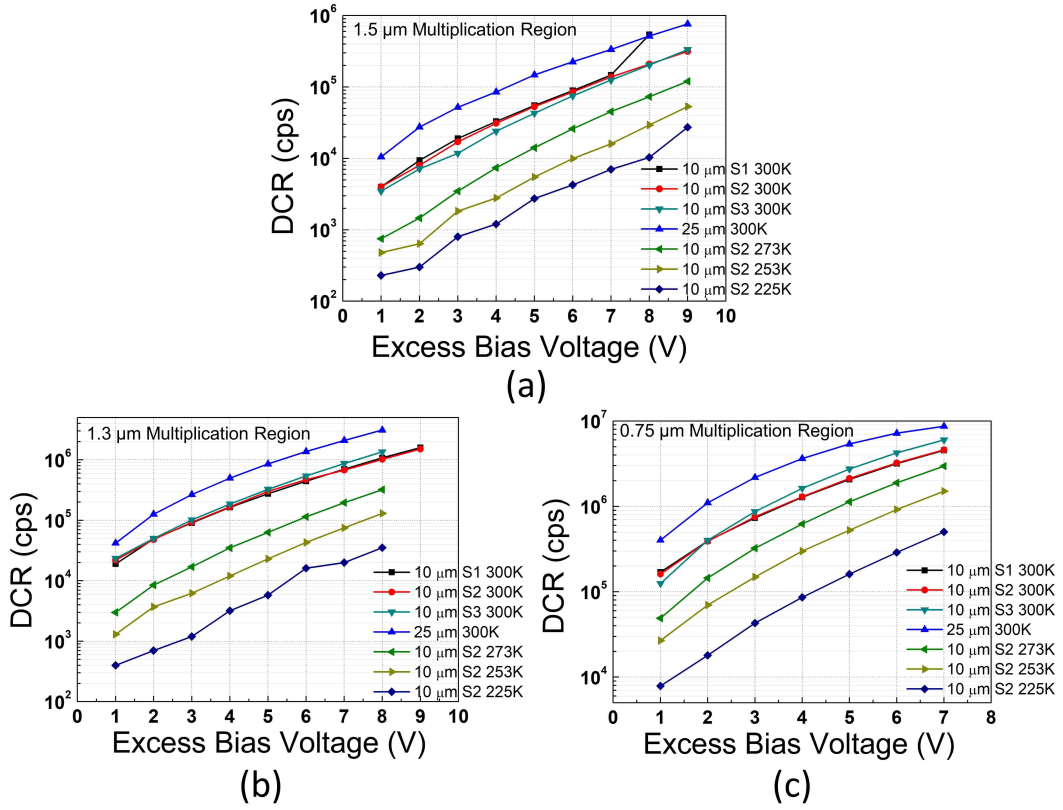


Figure 5.16: The DCRs of 10 µm and 25 µm devices with 2.5 µm GRs and (a) 1.5 µm multiplication region, (b) 1.3 µm multiplication region, and (c) 0.75 µm multiplication region at various temperatures.

The PDP spectrum for each device type is plotted in Fig. 5.17. The measurements were performed using the setup in Fig. 2.8 (a) with a monochromator and Xenon continuous light source. A calibrated Si photodiode (Thorlabs S130C) was utilized to calculate the incident number of photons at each wavelength. The calculations made use of the formula presented in Eq. 2.11 in Chapter 2 for gated mode SPADs. 10 µm devices were illuminated from the backside

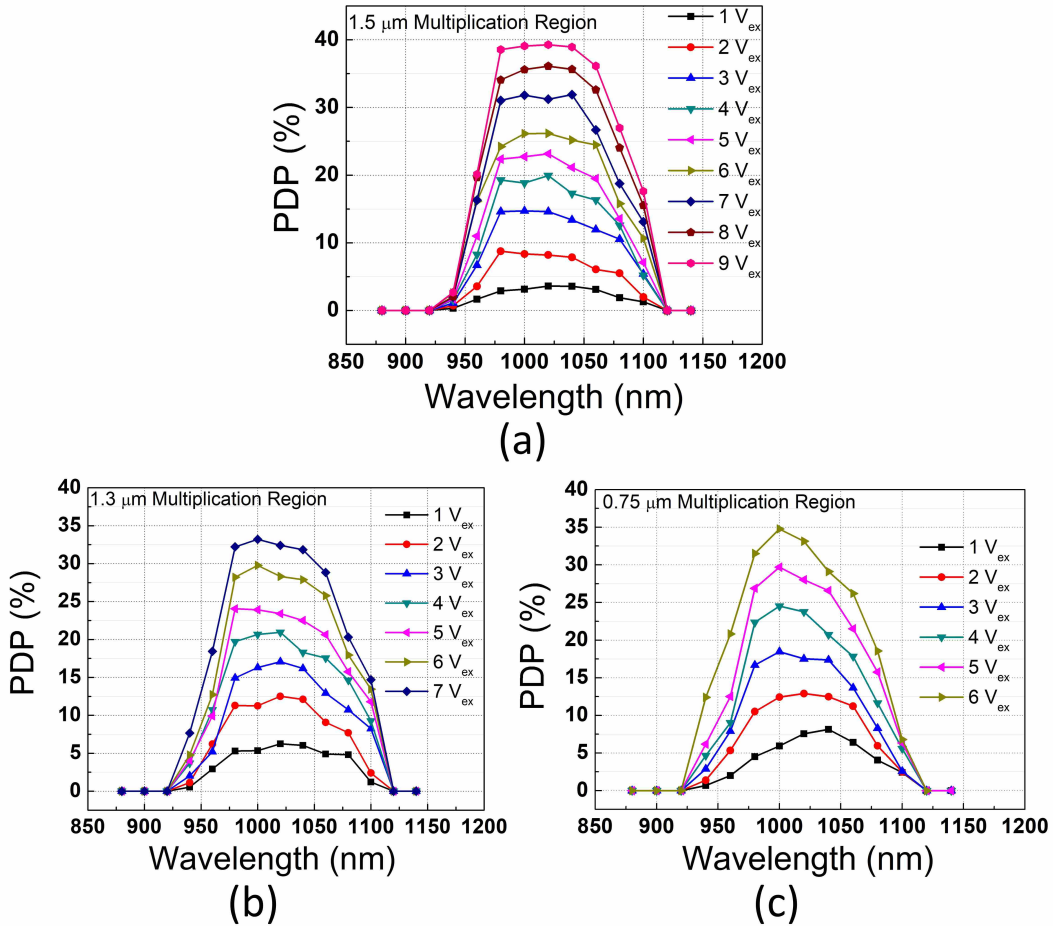


Figure 5.17: The PDP spectrums of 10 µm devices with 2.5 µm GRs and (a) 1.5 µm multiplication region, (b) 1.3 µm multiplication region, and (c) 0.75 µm multiplication region at room temperature.

at room temperature. Since the thick InP substrate absorbs the photons up to 920 nm, there was no SPAD response below 920 nm wavelength in BSI mode. Due to the noise limitations at room temperature, the maximum applied V_{ex} was 7 V and 6 V for 1.3 µm and 0.75 µm multiplication regions to be able to differentiate the photon counts. In the PDP calculations, the active diameter was accepted as 10 µm as designed in the layout and obtained in the laser scanning graphs. The PDPs at 1.06 µm wavelength and 5 V_{ex} are 19.5% for a multiplication thickness of 1.5 µm, 20.4% for 1.3 µm, and 21.5% for 0.75 µm. The peak PDP occurred around 1 µm wavelength, reaching 39% at 9 V_{ex}, 33.2% at 7 V_{ex}, and 34.9% at 6 V_{ex} in the SPADs with 1.5 µm, 1.3 µm, and 0.75 µm multiplication regions. As expected from theory and from TCAD simulations, the shorter avalanche region provided higher PDPs at the same excess bias voltages, thanks to the higher breakdown probabilities. In return, the DCR increases while shrinking the avalanche region thickness, as demonstrated in Fig. 5.16.

The timing histograms of the SPADs are given in Fig. 5.18. The histograms were acquired

with the same 1060 nm pulsed laser at 5 V_{ex} and room temperature, while the laser beam was focused at the center of the active region of the devices from the backside of the substrate. The FWHMs of the histograms were obtained as 125 ps for a multiplication region 1.5 µm thick, 117 ps for 1.3 µm, and 93 ps for 0.75 µm. After the deconvolution of the laser pulse, the SPAD jitter was found to be 118.4 ps, 110 ps, and 84 ps FWHM, respectively. Timing jitter dependence on avalanche region thickness is due to the enhanced electric field that shrinks the avalanche build-up times, as explained in Chapter 2. Moreover, owing to the fully depleted InGaAsP absorber, there was no significant exponential tail observed in the recorded timing plots. The results prove that the developed SPADs are compatible with timing-critical applications.

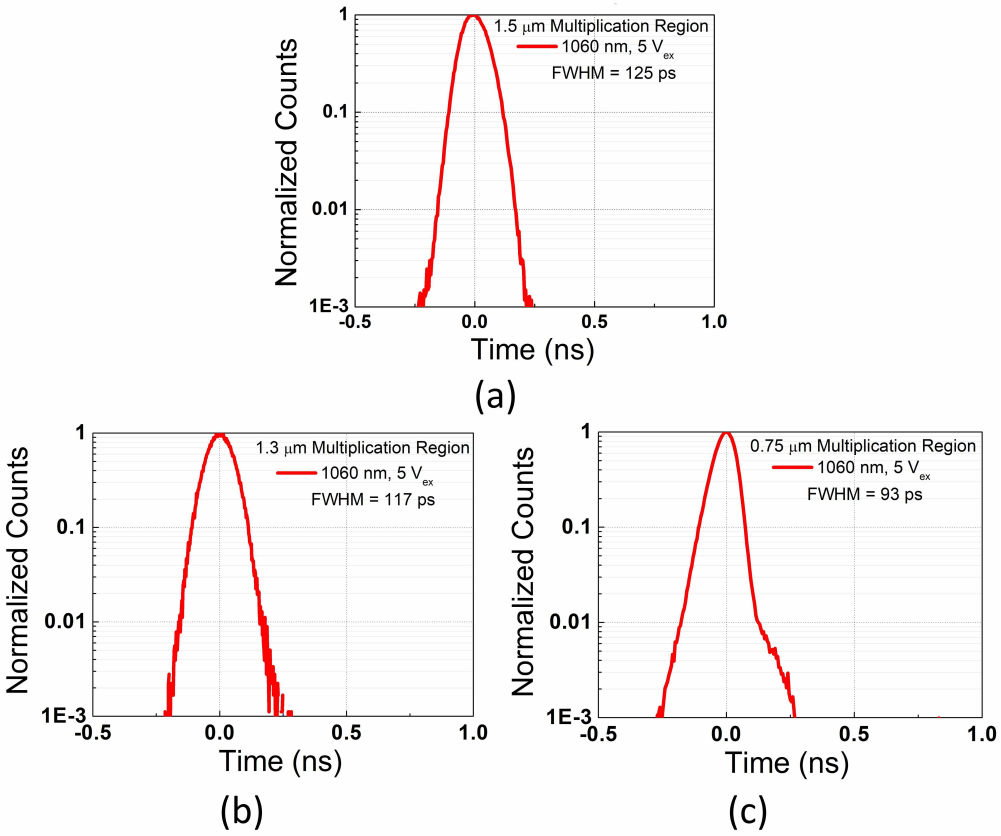


Figure 5.18: The timing histograms of 10 µm devices with 2.5 µm GRs and (a) 1.5 µm multiplication region, (b) 1.3 µm multiplication region, and (c) 0.75 µm multiplication region at 5 V_{ex} and room temperature.

The avalanche inter-arrival time histograms of the developed SPADs were also measured and are illustrated in Fig. 5.19. All the SPADs showed no afterpulsing at 10 kHz gating frequency, as demonstrated in Fig. 5.15. Therefore, the histograms were obtained at higher frequencies to see the effect of afterpulsing. A 10 µm device with a 1.5 µm multiplication region was measured at 500 kHz and 200 kHz gating frequencies at 300K, 5 V_{ex} and 100 ns gate-on time, whose histograms can be seen in Fig. 5.19 (a) and (b). The exponential fit was made on each histogram, representing the Poissonian distribution of avalanche inter-arrival times. The calculated APPs

for 500 kHz and 200 kHz gating were 11.1% and 5.8%, respectively. These findings indicate that the SPADs can be operated with low-enough afterpulsing at such high frequencies for 100 ns gate-on times. The temperature sweep for 200 kHz gating frequency was also performed for the same device and conditions. As shown in Fig. 5.19 (c) and (d), the APP increases to 12.8% and 33.5% at 273K and 253K due to the enhanced carrier lifetimes, as discussed. Hence, the APP is still at acceptable levels for 273K operation at 200 kHz frequency. Furthermore, the inter-avalanche time histogram of a 10 µm device with a 1.3 µm multiplication region was measured to check the difference between wafers. At 300K, 5 V_{ex} and 200 kHz gating, the APP was found to be 8.7%, which is close to the SPAD with a 1.5 µm multiplication region and indicates similar carrier lifetimes in the other wafers.

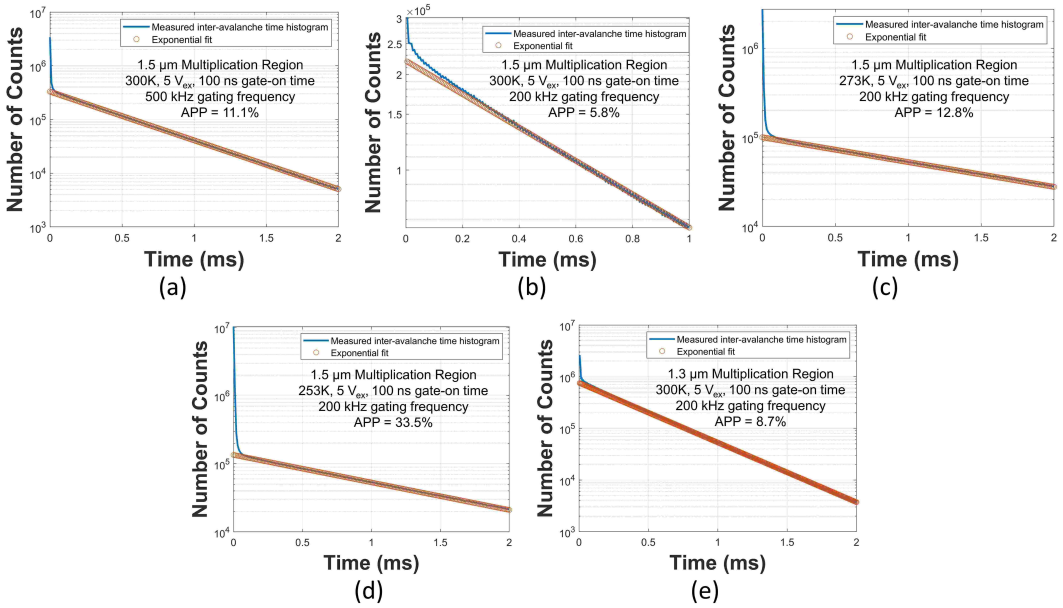


Figure 5.19: The avalanche inter-arrival time histograms of a 10 µm device with 2.5 µm GRs and a 1.5 µm multiplication region at 5 V_{ex}, 100 ns gate-on time, (a) 300K and 500 kHz, (b) 300K and 200 kHz, (c) 273K and 200 kHz, and (d) 253K and 200 kHz. (e) The same measurement for a 10 µm device with 2.5 µm GRs and a 1.3 µm multiplication region at 5 V_{ex}, 100 ns gate-on time, 300K, and 200 kHz.

Finally, a SPAD with a 0.25 µm depth difference between the deep and shallow Zn diffusion was characterized. This SPAD had 1.5 µm deep and 1.25 µm shallow Zn diffusions, and it was the fabricated version of the SPAD whose breakdown probability simulations were presented in Fig. 5.8. The SEM analysis of the fabricated device demonstrated that the deep and shallow diffusion depths were around 1.4 µm and 1.1 µm, respectively, as shown in Fig. 5.20 (a). The breakdown and punch-through voltages were identified as 76 V and 30.5 V at 300K from the I-V characteristics given in Fig. 5.20 (b). The breakdown voltage was 3 V higher than the simulated 1.5 µm multiplication region SPAD due to the deviation in the deep Zn depth in reality. The DCR measurements obtained for 10 µm active region and 2.5 µm GR width SPADs at 10 kHz gating frequency and 100 ns gate-on time at 300K are provided in Fig. 5.20 (c). Compared

to the SPAD with the 1.5 µm multiplication region and 0.5 µm depth difference in diffusions, the DCR was at similar levels up to 3 V_{ex}, whereas the DCR rose suddenly starting from 4 V_{ex}. Based on the hole breakdown probability analysis given in Fig. 5.8, this sudden increase was associated with the GR regions entering into breakdown while increasing V_{ex}. Due to the excessive noise observed in this device, the device with a 0.5 µm difference in Zn diffusions was considered the more optimized SPAD type for 10 µm SPADs with 2.5 µm GRs, and was characterized fully, as illustrated in the above parts. Therefore, the results indicate that to shrink pixel sizes and to achieve large-format SPAD arrays, preferring a higher depth difference between the shallow and deep Zn diffusion is preferable.

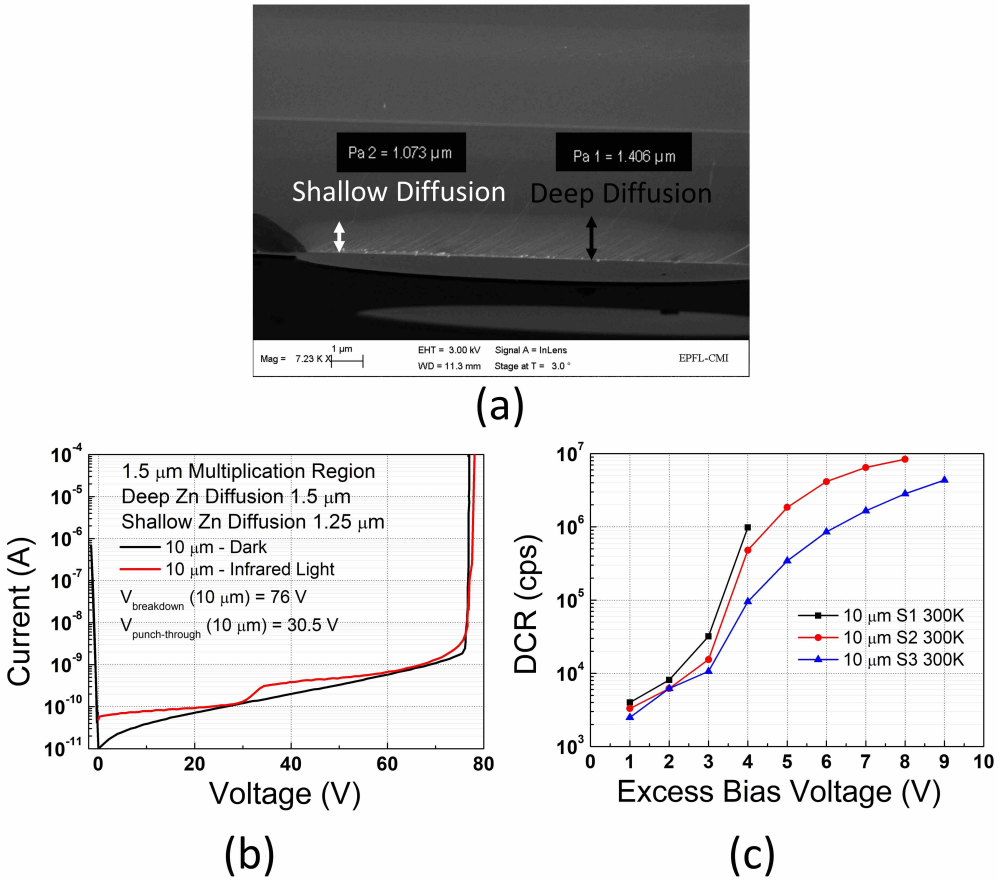


Figure 5.20: (a) The SEM analysis, (b) I-V characteristics at 300K, and (c) DCR measurements taken at 300K, 10 kHz gating frequency, and with 100 ns gate-on time of the SPAD with the 1.5 µm deep and 1.25 µm shallow Zn diffusions.

5.2.5 Comparison with the state-of-the-art

The three devices we developed with 1.5 µm, 1.3 µm, and 0.75 µm multiplication region thickness and 0.5 µm depth difference between the shallow and deep Zn diffusion are compared with the state-of-the-art InGaAsP/InP SPADs aiming for the detection of 1.06 µm wavelength

in Fig. 5.21 [98], [163]–[166]. The figure is plotted as PDP versus DCR. The reason for not using the normalized DCR is that the DCR does not scale with active region area, resulting in a lower normalized DCR in larger devices, as demonstrated in the previous sections. Instead, the active diameter of each device is given in the figure. The figure indicates the DCR is at acceptable levels for the 1.5 µm device at 300K and 253K, and for the 1.3 µm device at 253K. However, the DCR of the device with 0.75 µm multiplication is high, in exchange for very low timing jitter, as expected for a thin avalanche region. Better noise performance was achieved in Refs. [165], [166]. The former utilized a similar multiplication region of ≈ 1.6 µm, but with a bigger depth difference of 1.8 µm in Zn diffusions, lowering DCR further. The latter utilized a 1 µm multiplication region, but optimized the thickness and doping levels of the epi-structure to reduce the noise. Therefore, there is still room to improve the DCR in the developed SPADs. Furthermore, the maximum achieved PDP from the 1.5 µm multiplication device is around 36% at 9 V_{ex}. Higher PDP can be achieved with a thicker absorber or metal reflector allowing photons to double-pass the absorber, such as the SPADs demonstrated in Refs. [98], [165], [166] with 1.5 µm, 1.25 µm, and 1.5 µm absorbers, respectively. Particularly, Ref. [166] obtained a remarkable PDP of 70% with a 1.5 µm absorber, which indicates that either the metal reflector is very efficient or the avalanche triggering and the carrier injection efficiencies reached around 100% with their epi-structure. In addition, the major advantage of the developed SPADs is their very small device sizes of 10 µm, which is important to build larger pixel arrays with reduced pixel pitches. Another advantage is the lower timing jitter, proving their better applicability to timing-critical applications.

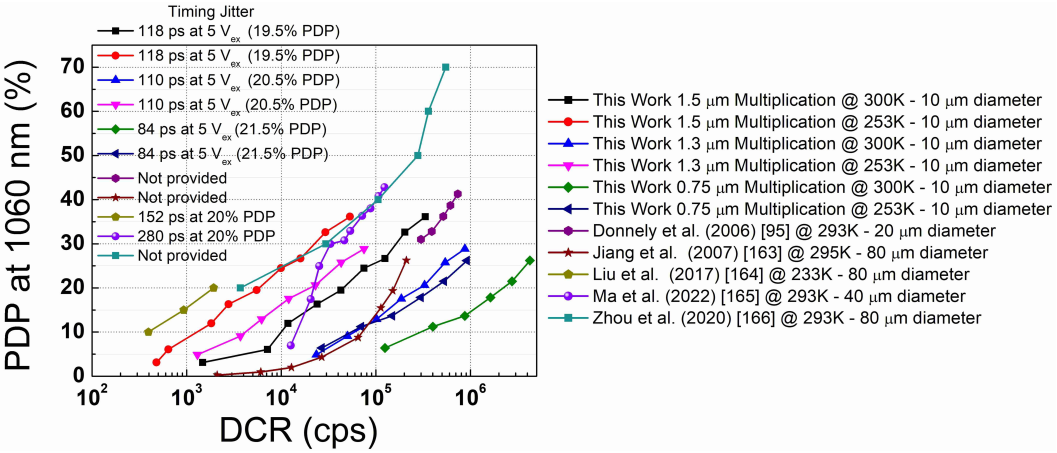


Figure 5.21: The DCR, PDP, and timing jitter comparison of the proposed InGaAsP/InP SPADs with the state-of-the-art for 1.06 µwavelength detection.

5.3 Planar InGaAs/InP SPADs for 1.55 µm wavelength detection

The second type of SPADs had an InGaAs absorber with a 1.7 µm cutoff wavelength (0.73 eV bandgap) and aimed at detection of 1.55 µm wavelength. LiDAR applications can benefit from eye-safe and very high-power lasers manufactured at this wavelength. As also mentioned,

1.55 µm wavelength provides the lowest losses in fiber-optics and is commonly preferred for fiber-based telecommunication and QKD systems. So far, SPADs with a 10 µm diameter were characterized to achieve lower noise; however, 25 µm diameter devices will be fabricated and investigated in the future since QKD can demand larger devices for better optical coupling. Compared to InGaAsP absorber SPADs with the 1.1 µm cutoff wavelength, the lower bandgap of the InGaAs absorber enhances thermal generation, thus the DCR increases in InGaAs/InP SPADs, requiring cooling to suppress thermal generation. Cooling is the price to pay to extend the wavelength of detection to 1.55 µm. In this thesis, InGaAs SPADs with a 1.5 µm multiplication region and 0.25 µm depth difference between shallow and deep diffusion were analyzed. 0.25 µm depth difference between the active and GR regions was preferred since it can provide a more uniform device response and higher PDP according to the TCAD simulations. Yet, this depth difference was problematic in InGaAsP devices in terms of noise. Therefore, various GR widths were also implemented to optimize the structure.

5.3.1 Device structures

The cross-section of the InGaAs/InP SPADs with the implemented doping concentration of each epi-layer is given in Fig. 5.22. As in 1.1 µm cutoff devices, the devices were designed based on the planar double Zn diffusion process without FGRs, and the intrinsic layers' background dopings were adjusted by the foundry. The charge layer doping was the same as in the InGaAsP SPADs, which is 2.2×10^{17} cm^{-3}. Similarly, the InGaAs etch stop layer was used in the case of removing the InP substrate to extend the wavelength of operation below 920 nm. SiN was passivating the top surface for long-term stability. The Zn diffusion procedure was exactly the same as described in the InGaAsP SPADs. The only additional layer in the epitaxy is the InGaAsP grading layer between the InP and InGaAs layers. A grading layer is necessary in this design since the hole barrier reaches 0.4 eV between the InP-InGaAs heterojunction, and it has to be eliminated to overcome hole pile-up and jitter degradation. The thickness of the epi-layers is the same in InGaAsP SPADs, and the grading layer is 0.1 µm. The depth difference between shallow and deep diffusion was 0.25 µm to achieve a more uniform active area response and enhance the PDP at 1.55 µm wavelength. The multiplication region thickness was fixed, and it is 1.5 µm. The absorption coefficient of InGaAs at 1.55 µm is around 5800 cm^{-1} [167]. Therefore, according to Beer-Lambert's law, the absorption is 45% of photons for 1 µm absorber thickness. As the thermal generation and DCR will be augmented in InGaAs SPADs, the absorber thickness was not extended at this point before measuring the DCR of the devices. The active diameter of the fabricated devices was 10 µm. The 0.25 µm depth difference augmented the DCR in the InGaAsP SPADs, as explained in the previous section, due to the GR regions entering into breakdown. Therefore, the GR size was varied in InGaAs SPADs from 2.5 µm to 7.5 µm to find the optimum value for the 0.25 µm Zn diffusion difference in double Zn diffusion. As also mentioned before, a larger mask opening results in a reduction in the Zn diffusion depth, and the variation of the GR sizes was aimed at benefiting this property.

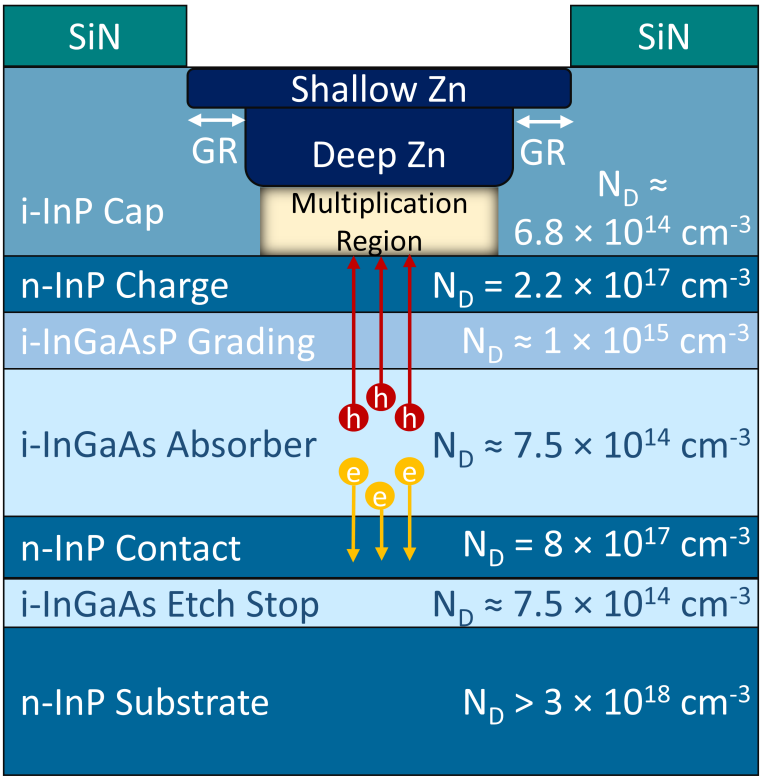


Figure 5.22: The cross-section of the 1.7 µm cutoff InGaAs/InP SPADs.

5.3.2 TCAD simulation results

The electric field distribution of the described device with a 10 µm active diameter and 2.5 µm GRs in TCAD is given in Fig. 5.23 (a), which was obtained at the breakdown voltage. It shows that the high electric field is confined in the InP multiplication region beneath the Zn-diffused regions. The absorber is also fully depleted at the breakdown, as can be observed from the white depletion region lines extending to the n-contact InP layer. Compared to 1.5 µm deep and 1 µm shallow diffusion in Fig. 5.5 (a), 1.25 µm shallow diffusion enhances the electric field at the edges of the GR regions. This is also in line with the hole breakdown probability simulations of the same device with the InGaAsP absorber in Fig. 5.8. Hence, the GRs might contribute to the DCR after a certain excess bias voltage, like the presented InGaAsP SPAD in Fig. 5.20. To prevent GRs from entering into breakdown, various GR sizes were fabricated, where increasing the mask opening of the GRs can slightly decrease the shallow Zn diffusion depth, enabling finely tuning the electric fields at the GRs' edges to optimize the device.

The energy band diagram simulation of the device is also provided in Fig. 5.23 (b) at the breakdown voltage. This simulation was made to demonstrate the importance of the grading layer in the design of InGaAs SPADs. As the holes are subjected to avalanche multiplication, the hole barrier that exists in the InGaAs-InP heterojunction has to be removed. The 0.4 eV hole barrier can be seen in the zoomed version when there is no grading layer. By inserting a

0.1 µm InGaAsP grading layer, where the composition linearly changed from InP to InGaAs, the barrier at the valence band can be eliminated, as shown in the figure. Thus, in the fabricated devices, an InGaAsP grading layer was utilized, in which the foundry graded the InGaAsP composition in several steps.

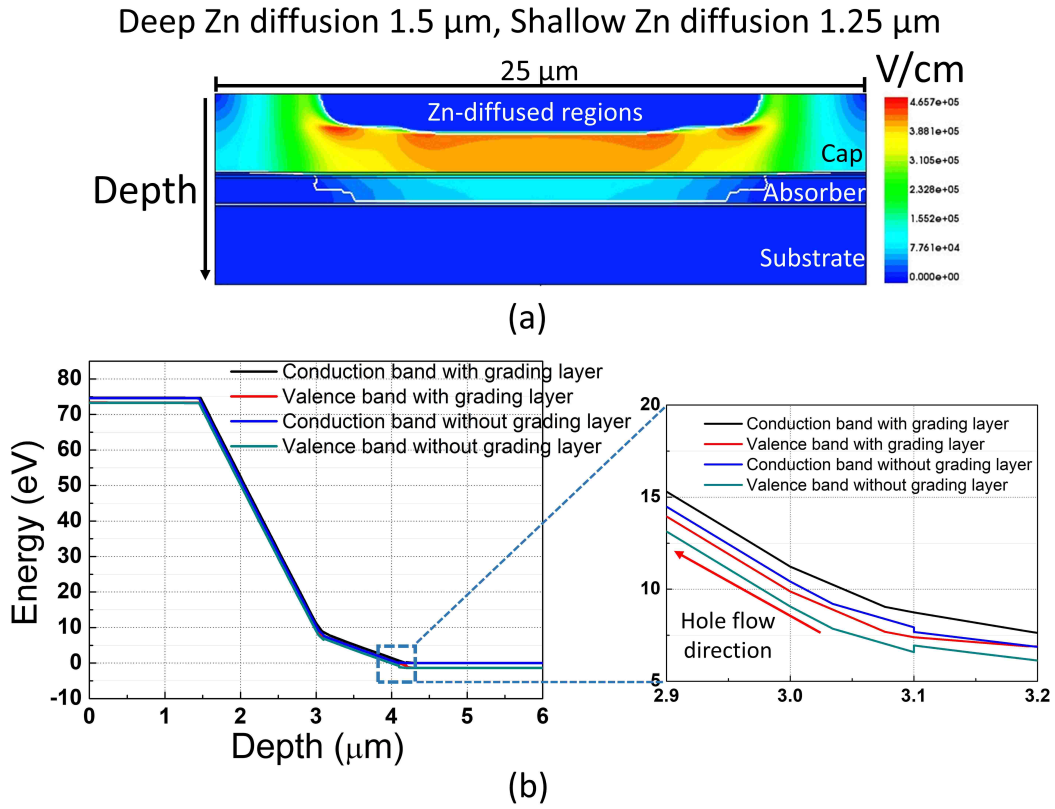


Figure 5.23: (a) The electric field distribution and (b) the energy band diagram with and without the grading layer of a 10 µm InGaAs SPAD with 1.5 µm deep, 1.25 µm shallow diffusion, and 2.5 µm GR width at the breakdown voltage.

The simulated dark I-V curve of the described InGaAs SPAD is depicted in Fig. 5.24. The breakdown voltage of the device was 73.2 V at 300K. The punch-through voltage was 51 V, ensuring the depletion of the absorber during Geiger mode operation.

The hole avalanche breakdown probabilities of the SPAD are given in Fig. 5.25. Similarly to the InGaAsP SPAD with a 0.25 µm depth difference between diffusions, the device has a more uniform response compared to the 0.5 µm difference, achieving a higher breakdown probability at the center of the device at 1 V_{ex}. While increasing the voltage, it was also observed that the GR regions break down in InGaAs SPADs with a 0.25 µm depth difference, like in InGaAsP SPADs. GR regions started to yield avalanche multiplication at around 5 V_{ex} in InGaAs. Furthermore, the whole SPAD area, including GRs, contributed to the breakdown probability at 9 V_{ex}. Although the device response is flatter at the center and edges of the active region, as shown in Fig. 5.25 (b), the breakdown probability outside the active region could

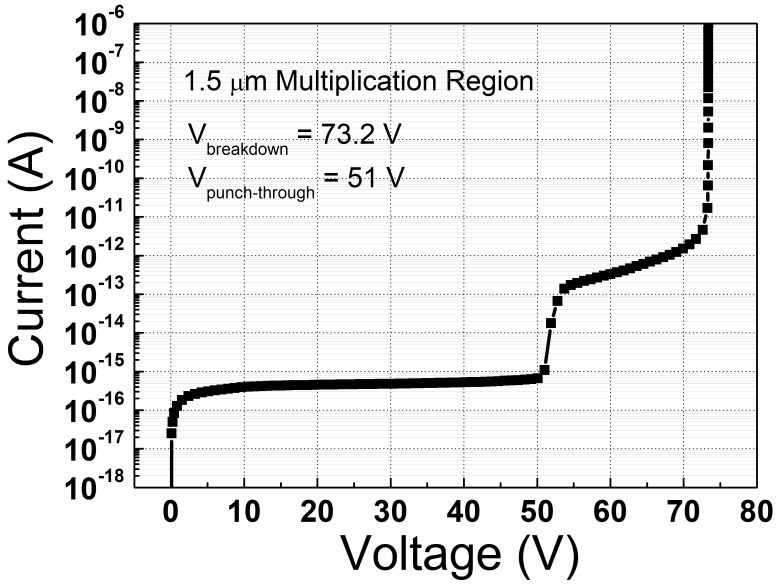


Figure 5.24: The simulated dark I-V curve of the described 10 µm InGaAs SPAD with 2.5 µm GRs at 300K.

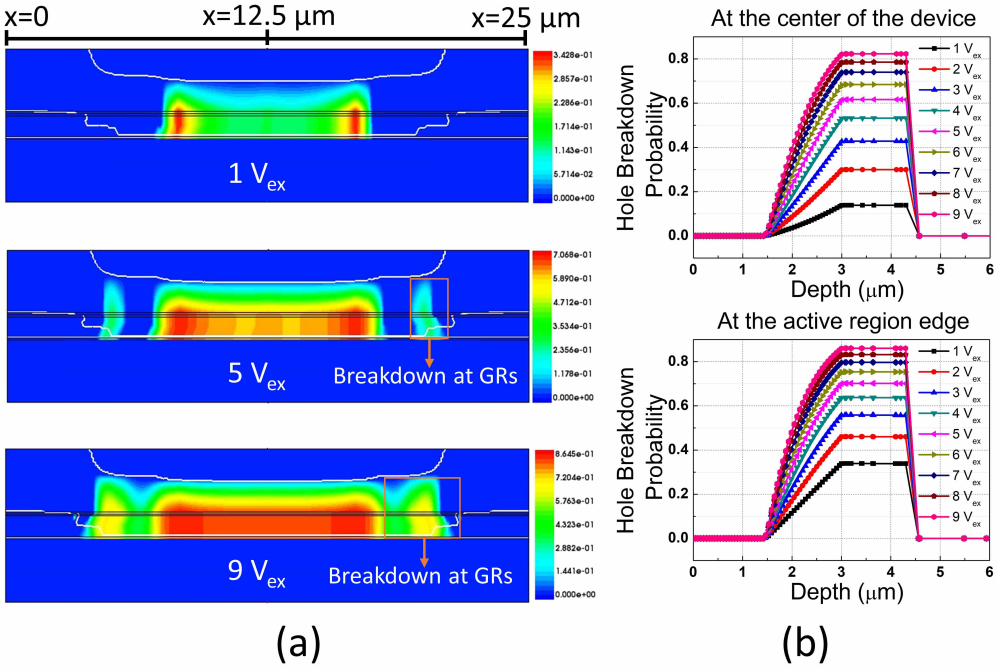


Figure 5.25: (a) Hole avalanche breakdown probability distribution of the device with a 1.5 µm multiplication region and 0.25 µm depth difference between Zn diffusions at 1 V_{ex}, 5 V_{ex}, and 9 V_{ex}. (b) Hole avalanche breakdown probabilities taken at the center and edge of the active region for various excess bias voltages.

greatly increase DCR, limiting the maximum allowable excess bias voltage, and thus limiting the PDP of the device. Therefore, several GR widths were considered in the fabricated device

to slightly tune the depth of GRs and to be able to increase the PDP.

5.3.3 Fabrication procedure

The fabrication procedure was the same as in 1.1 µm cutoff InGaAsP/InP SPADs, resulting in a structure compatible with both FSI and BSI modes by carrying the cathode to the top surface with a ladder-shaped etching in three steps, as shown in Fig. 5.9. In order to etch InGaAsP and InGaAs, a dionized water-diluted H_3PO_4:H_2PO_2 mixture was used. For the metallization, 250 nm/1000 nm Ti/Au were deposited.

5.3.4 Characterization results

The SEM analysis was performed first on the broken 10 µm diameter SPADs with 2.5 µm GR width. The SEM image in Fig. 5.26 indicates the achieved deep and shallow diffusion depths were 1.48 µm and 1.27 µm, which are close to the planned diffusion depths of 1.5 µm and 1.25 µm. The SEM images of the SPADs with wider GR widths could not be obtained since the arrays from which the samples can be broken were initially designed only with 2.5 µm GR width. However, as reported in the literature in Refs. [61], [104], lower shallow diffusion depths are expected with larger GR widths.

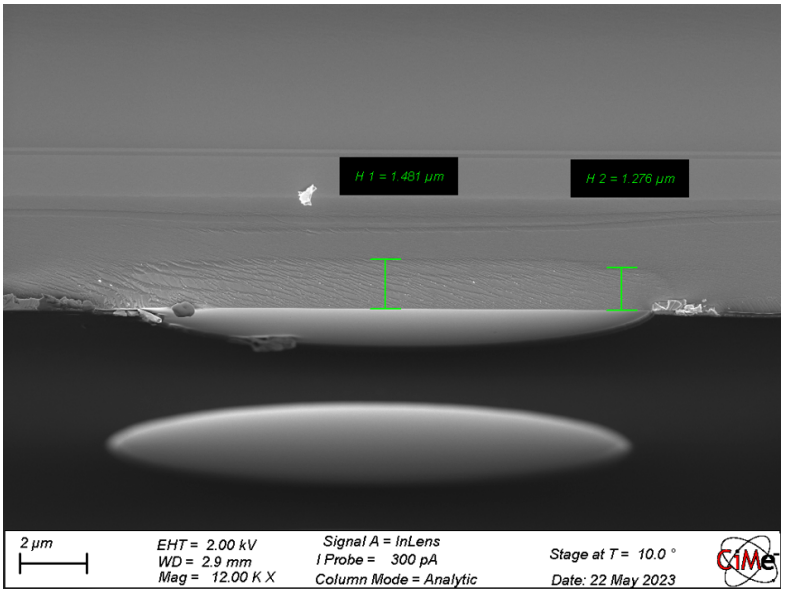


Figure 5.26: The SEM analysis of the described InGaAs with a 10 µm diameter and 2.5 µm GRs.

The I-V characteristics of the device at 300K is given in Fig. 5.27. The breakdown voltage is 73.5 V, which is only 0.3 V higher than the simulated value. The punch-through voltage is 34.5 V in the ambient infrared light, which ensures that the absorber is depleted while operating the device above the breakdown voltage. Similar to InGaAsP SPADs, a lower punch-through voltage was observed in this InGaAs SPAD compared to the simulation, whose reason is not

exactly clear at this point of time, but might be related to the difference in the absorber background doping utilized in the simulation and in the fabricated devices. Also, the change in the breakdown voltage with respect to temperature was measured as 0.9 V/10K in InGaAs SPADs.

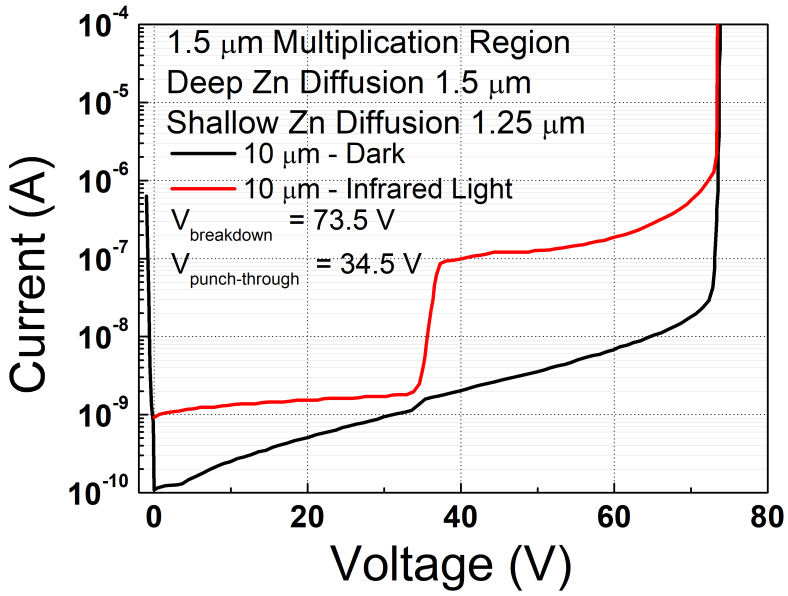


Figure 5.27: The I-V characteristics of the InGaAs SPAD with a 10 µm diameter and 2.5 µm GRs at 300K.

The DCR of 10 µm diameter SPADs with various GR widths are provided in Fig. 5.28 (a) at room temperature. Since breakdown at GRs is expected while increasing V_{ex} according to TCAD simulations, different GR sizes were analyzed, knowing that lower shallow diffusion depths are expected with larger GR widths. The SPADs were operated with a 50 kΩ passive resistor, 10 kHz gating frequency, and 100 ns gate-on time, while the DC voltage of the bias tee was kept 1 V below the breakdown voltage. The results showed that the DCR is higher in 2.5 µm and 5 µm GR widths compared to the 7.5 µm one. Higher DCR can be attributed to the higher electric field at the GR edges, contributing to the avalanche multiplication in 0.25 µm Zn diffusion depth difference devices. However, it was observed that the high electric field at GRs can be suppressed with 7.5 µm GRs, resulting in lower noise SPADs. As mentioned, this can be possible with shallower Zn diffusion at GRs while increasing mask sizes [61], [104]. Therefore, the SPADs with 7.5 µm GRs were focused on in the remaining measurements as they provide more optimized noise performance. Fig. 5.28 (b) demonstrates the variation of DCR with respect to temperature for a 10 µm SPAD with 7.5 µm GRs. Measurements were performed under the same conditions of 10 kHz gating and 100 ns gate-on time. As expected, the DCR is higher in the InGaAs SPADs than the InGaAsP SPADs because of the lower bandgap, increasing the thermal generation. At 5 V_{ex}, the DCR attained 8.6×10^6 cps, 2.79×10^6 cps, 6.52×10^5 cps, and 42.8k cps at 300K, 273K, 253K, and 225K, respectively. In Fig. 5.28 (c), the gating frequency sweep is shown from 273K to 225K at 5 V_{ex} for a SPAD with 7.5 µm GRs. As can be seen, the 10

kHz gating frequency provides afterpulsing-free operation for all temperatures. At 273K and 253K, the devices can be operated with low afterpulsing between 75 kHz and 100 kHz. Low afterpulsing at 225K is achievable with 20 ns gate-on time at 50 kHz. However, under 100 ns gate-on, the maximum operating frequency suffering from afterpulsing at 225K is around 10 kHz.

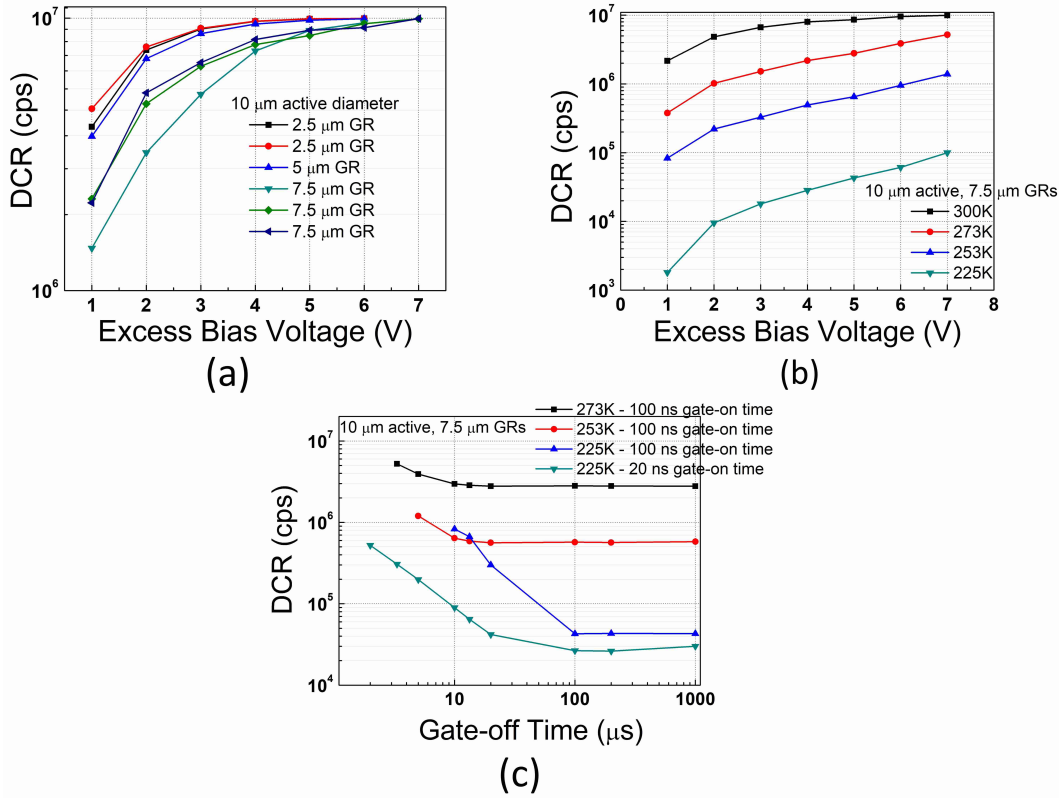


Figure 5.28: (a) DCR versus V_{ex} analysis of 10 µm SPADs for various GR widths at 10 kHz gating with 100 ns gate-on time. (b) The DCR of a 10 µm SPAD with 7.5 µm GRs under various temperatures. (c) Gating frequency sweep of a SPAD with 7.5 µm GRs from 273K to 225K.

Active area scanning of a 10 µm diameter SPAD with 7.5 µm GRs is illustrated in Fig.5.29 (a) and (b), at 3 V_{ex} and 5 V_{ex}. The measurements were taken with a 1550 nm pulsed laser, at 10 kHz gating frequency and 100 ns gate-on time at room temperature. The SPAD was quenched and recharged passively with a 50 kΩ ballast resistor. The illumination was from the backside, and the focused beam size was around 1 µm. The step size of the motorized stage was 1 µm during the scanning. The measured size of the SPAD was also around 10 µm, matching the design in the layout. The transition zone that occurred around the active area with a gradually decreasing count rate is probably given by the laser beam spot size. Thanks to the smaller diffusion depth difference between the deep and shallow Zn diffusions, there were no observed edge effects even at a low bias of 3 V_{ex}. In other words, a more uniform active area was achieved with a 0.25 µm depth difference compared to InGaAsP SPADs with a 0.5 µm depth difference, as demonstrated in Figs. 5.14 and 5.29.

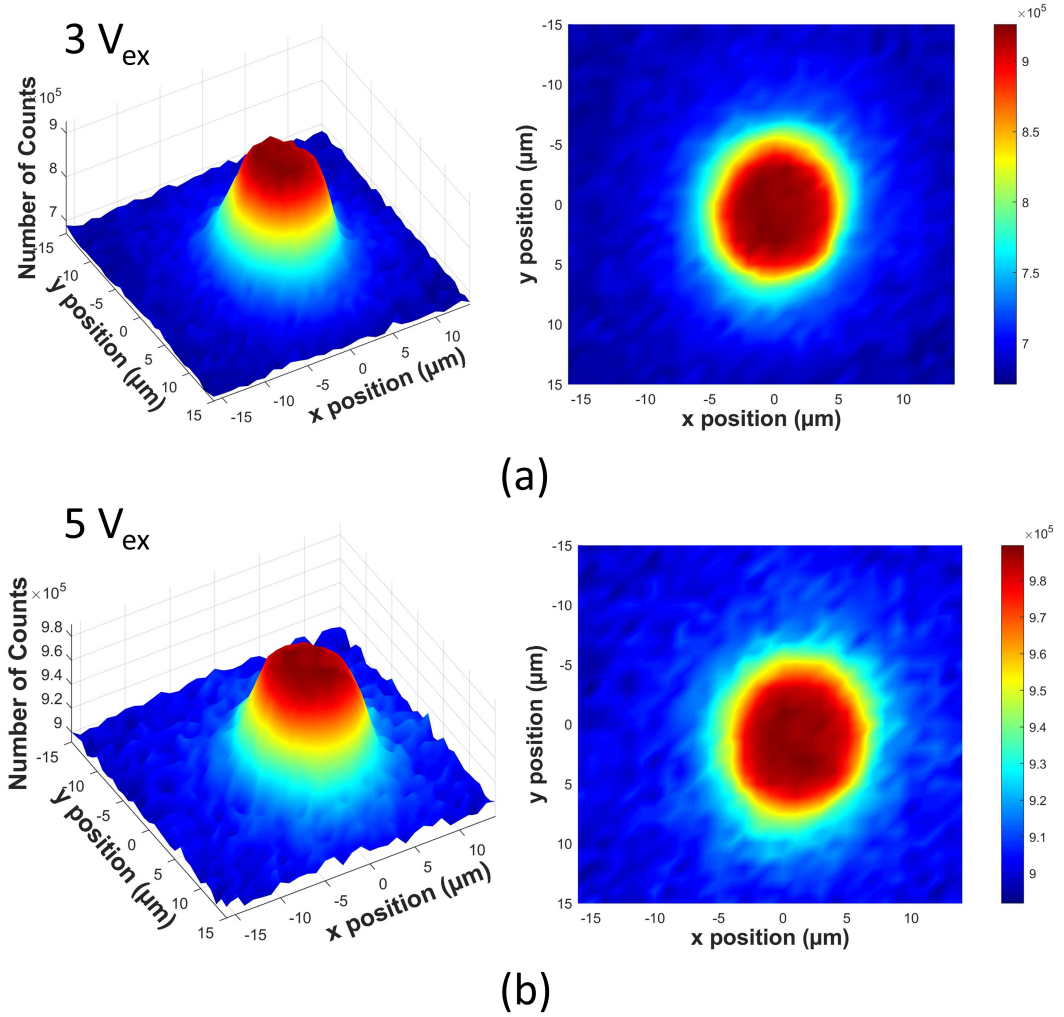


Figure 5.29: The active area scanning of a 10 µm InGaAs/InP SPAD with 7.5 µm GRs with a 1550 nm pulsed laser at room temperature and (a) 3 V_{ex} and (b) 5 V_{ex}.

The PDP measurements of the SPAD with a 10 µm diameter and 7.5 µm GRs at room temperature are illustrated in Fig. 5.30. They were taken with the same 1550 nm laser while it was focused at the center of the active region from the back-side at 10 kHz and 100 ns gate-on time. The setup presented and described in Fig. 2.8 was utilized to measure PDP. The 99% port of the beam splitter was connected to a calibrated Thorlabs power meter (PM20A) to measure the input laser power. A calibrated Thorlabs InGaAs photodiode (FGA21) was used to calculate the mean number of photons per optical pulse after the power reduction at the 1% beam splitter port, attenuator, and confocal microscope's objective. The measurements were eventually performed with 0.15 photons per optical pulse. Eq. 2.11 was then used to calculate the PDPs in time-gating mode. The PDP at 1550 nm reaches around 15% and 20% at 5 V_{ex} and 6 V_{ex}, respectively.

The timing histogram of the device is given in Fig. 5.31. The histogram was also obtained

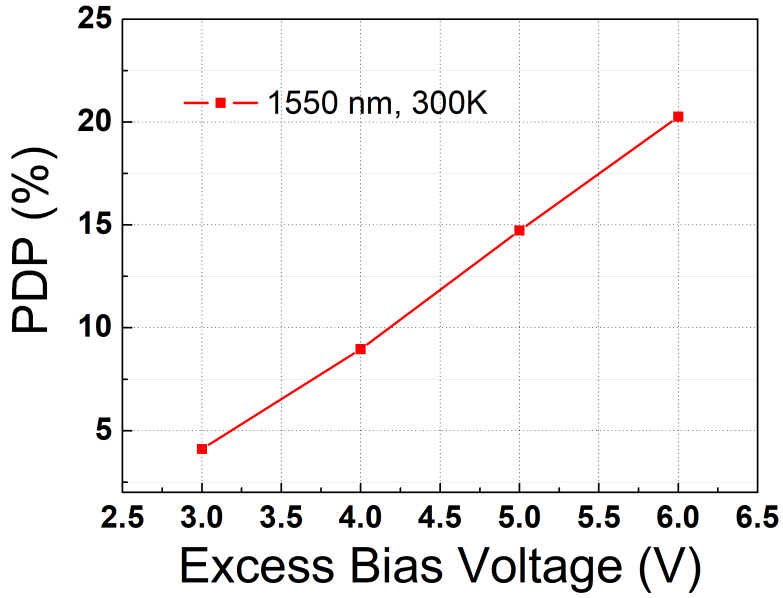


Figure 5.30: The PDP measurements of a 10 µm InGaAs/InP SPAD with 7.5 µm GRs at 1550 nm and room temperature for various excess bias voltages.

with the 1550 nm pulsed laser while the beam was focused in the middle of the active area at room temperature and 5 V_{ex}. The laser was synchronized with the waveform generator at 10 kHz. The FWHM of the histograms was found to be 130 ps. The deconvolution of the laser pulse yields a timing jitter of 123 ps FWHM. Besides, an exponential tail was observed in the histogram, even though the absorber layer was expected to be depleted. The tail might stem from the grading layer, whose material composition was changed in several steps but not in a perfectly smooth manner during the growth. This might result in a deviation in the transition time of the carriers from the absorber to the multiplication region. The time constant of the tail was calculated at around 82 ps from the exponential fit.

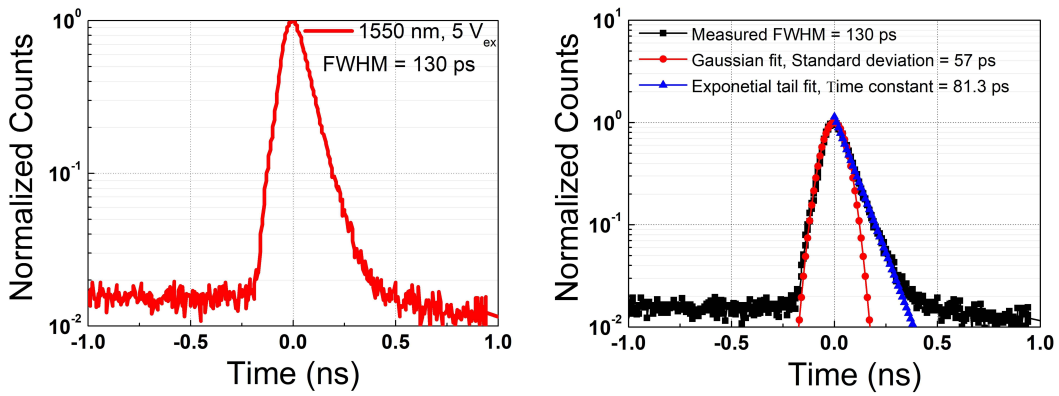


Figure 5.31: The timing histogram of a 10 µm InGaAs/InP SPAD with 7.5 µm GRs at 1550 nm, room temperature and 5_{ex}.

Lastly, the avalanche inter-arrival time histogram of the device is given in Fig. 5.32. The measurement was performed at 273K, 100 kHz gating frequency, and 100 ns gate-on time. The APP was found to be 8.4% after performing the exponential fit. The result indicates both that the device can be operated at 100 kHz without suffering much from afterpulsing and that there was a near-zero contribution of afterpulsing effect to the PDP measurement, which was performed at 300K and 10 kHz gating.

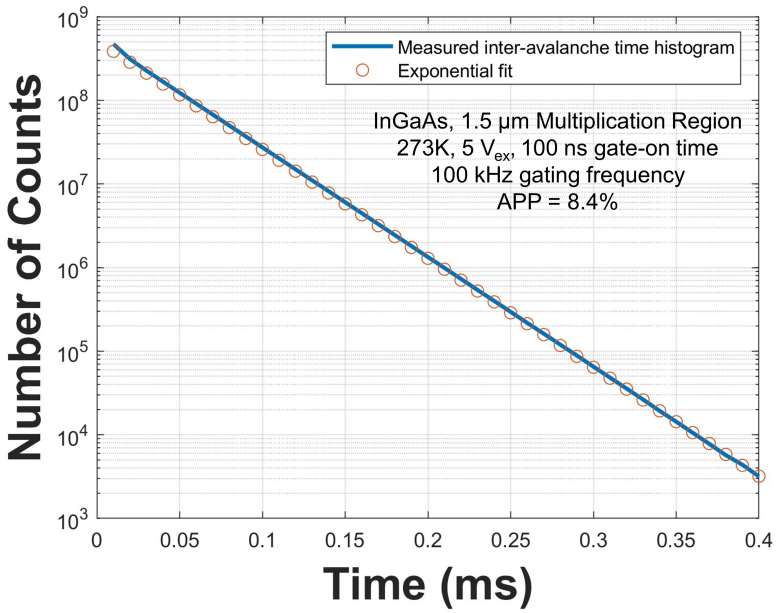


Figure 5.32: The avalanche inter-arrival time histogram of a 10 µm device with 7.5 µm GRs at 5 V_{ex}, 273K, and 100 kHz gating frequency with 100 ns gate-on time.

5.3.5 Comparison with the state-of-the-art

The comparison of the designed and characterized InGaAs SPAD with the state-of-the-art InGaAs SPADs targeting 1.55 µm detection is given in Fig. 5.33 [66], [104], [168]–[172]. The comparison is made at around 250K for all the devices. As can be seen, the developed device is noisier than the state-of-the-art. This is partially due to keeping the Zn diffusion depth difference smaller, and better results can be obtained by a shallower GR depth, as demonstrated in InGaAsP SPADs. Yet, epi-structure optimization, particularly charge doping and the width of the multiplication region, is required to further bring down the noise levels to the state-of-the-art. The achieved 20% peak PDP is somewhat closer to the best PDP obtained, which is in the range of 25%-30%. It should be noted that the measured PDP in this thesis were obtained without any back-reflector metal unlike the reported devices [66], [104], [168]–[172]. In Ref. [98], it was calculated that a 90% back-reflector can increase the efficiency by up to 15%. Also, the device in Ref. [66], which reached a very high PDP (> 40%), suffered from high afterpulsing. At 45% PDP, the afterpulsing was around 12.5% in that device. Moreover, in the demonstrated device, higher PDP can be achieved by cooling down the device, which

would allow to increase V_{ex} without saturating the count rate, which was the case at room temperature after 6 V_{ex}. However, the gating frequency should be reduced at lower temperatures due to the more severe afterpulsing effect. Finally, the active diameter of the SPAD is one of the smallest devices fabricated, while the timing jitter is also better than most of the state-of-the-art InGaAs SPADs.

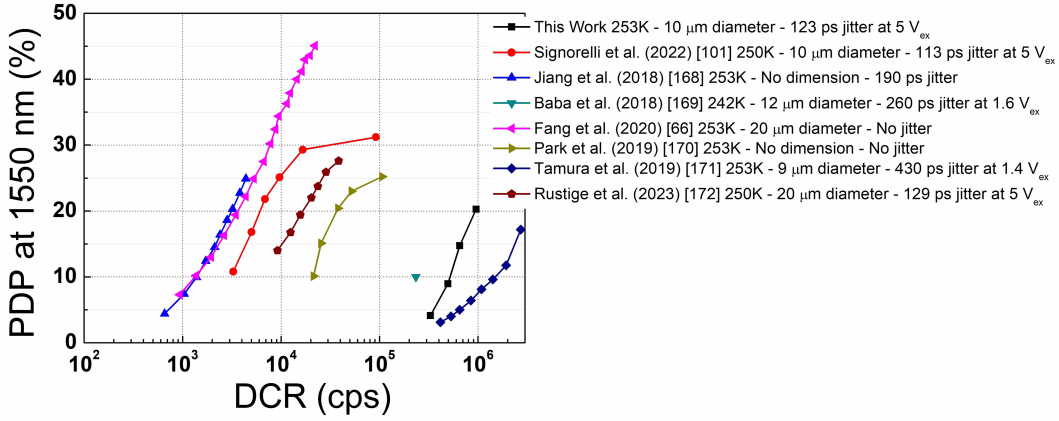


Figure 5.33: The DCR, PDP, and timing jitter comparison of the proposed InGaAsP/InP SPADs with the state-of-the-art for 1.55 µwavelength detection.

5.4 Development for the 3D Integration of InGaAs(P)/SPADs with Si ROICs

The last goal regarding InGaAs(P)/InP-based SPADs was to develop a methodology for integrating them with Si-based ROICs. Since the demonstrated SPADs were grown on InP substrates, they have to be somehow bonded with the Si chips in a way that large SPAD pixel arrays connected to fast ROICs can be realized. The motivation behind this study was to be able to use these devices in LiDAR applications. In this sense, the 3D flip-chip bonding technique based on indium (In) bumps was investigated. The schematic of the flip-chip bonding is given in Fig. 5.34. The idea is to create In bumps on both the InGaAs(P) and Si chips, and to flip upside down either the Si or InGaAs(P) chip so that the separate parts can be aligned with a high-precision flip-chip bonding machine [173]. A flip-chip machine has essentially two mechanical parts, named the chuck and arm. Each part is capable of holding a chip, and the arm part can pick up a chip by vacuuming it and can lift it up. Then, a microscope goes in between the chuck and arm, enabling to monitor both chips simultaneously and perform the alignment procedure. The bumps can then be soldered together thanks to the applied force and temperature of the bonding machine. In addition, it is possible that the bumps are placed only to the InGaAs(P) side, and bump-to-metal path bonding is achieved. However, a wide enough gap between two chips is also desired to be able to inject underfill epoxy between the bumps to increase the long-term stability of the chips. Therefore, the bumps were intended to be deposited on both chips in this work. Another important point is to make sure that In bumps do not destroy the functionality of each chip. It is known that In atoms tend to diffuse

inside the epitaxy and change the doping concentration of the layers, which can alter the operation of the chips [162]. Hence, the under bump metallization (UBM) is vital to prevent In atoms diffusing. In this thesis, a Ti/nickel (Ni)/Au metal stack was deposited on both chips as shown in Fig., where Ni forms a diffusion barrier for In atoms [174].

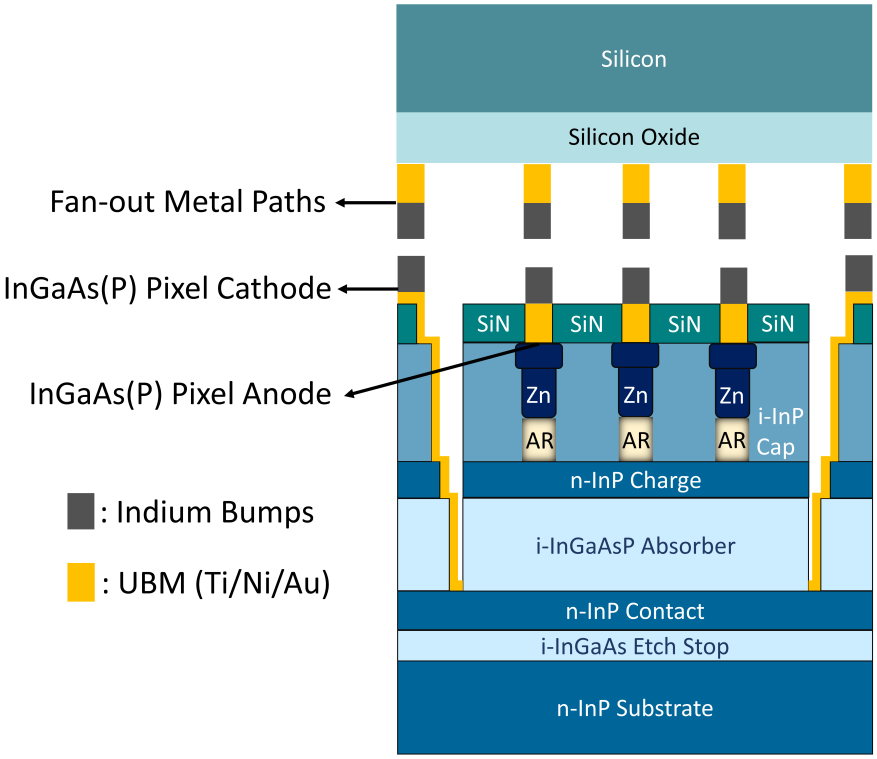


Figure 5.34: A schematic of the flip-chip bonding technique applied for the integration of the InGaAs(P) SPADs with Si chips.

To achieve this aim, first of all, a process was developed to achieve at least 4 µm thick In bumps so that the neighboring pixels do not short each other and there is enough gap between the chips for underfill injection. However, this thickness is far thicker than the regular thin metal film depositions (in nm range), so a special fabrication procedure is required to create thick bumps. For this purpose, a photoresist type that can reach around 8 µm thickness after spin coating was selected and deposited. Followed by photolithography to form 10 µm diameter bumps, 25 nm/35 nm/100 nm Ti/Ni/Au metal stack was deposited for the UBM. To deposit In, a thermal evaporator was utilized, where the final In thickness depends on the amount of In metal loaded in the boat. SEM images of the first In deposition is given in Fig. 5.35. The chip holder was rotated to achieve uniform In thickness. The first run did not yield enough In material in the middle of the bump, as can be seen. Therefore, in the second run, the amount of In in the boat was increased, and the wafer holder was not rotated to determine the maximum In thickness that could be achieved. As shown in Fig. 5.35, very thick In bumps (>10 µm) can be obtained with this recipe. However, the uniformity of the bumps was not sufficient, and some falling pieces were observed, which can electrically short neighboring

bumps. Thus, in the final run, the same amount of In material was used while the wafer holder was rotating at a medium speed. The deposition was stopped when 4 µm thickness was read from a calibrated crystal. As indicated in the SEM pictures, 4-5 µm thick, fully filled, and uniformly distributed In bumps were created in the final run, which are suitable for flip-chip bonding.

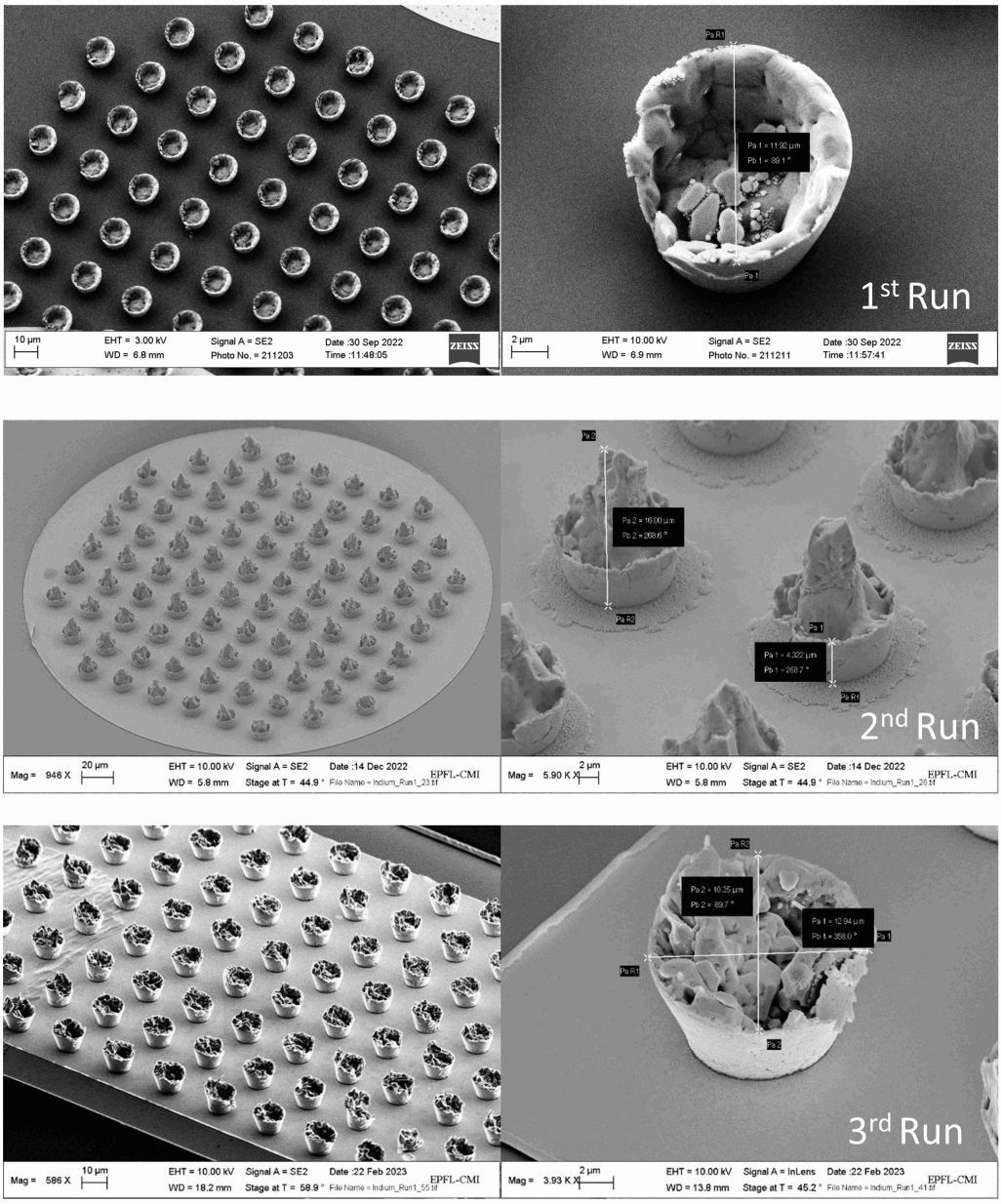


Figure 5.35: The process of In bump fabrication development, illustrated with SEM images.

After developing In bumps, they were deposited on a fabricated 1×128 InGaAsP SPAD array with 25 µm active diameter and 60 µm pixel pitch. SPADs had a 1.3 µm multiplication region with a 0.5 µm depth difference between Zn diffusions. A microscope image of a fabricated

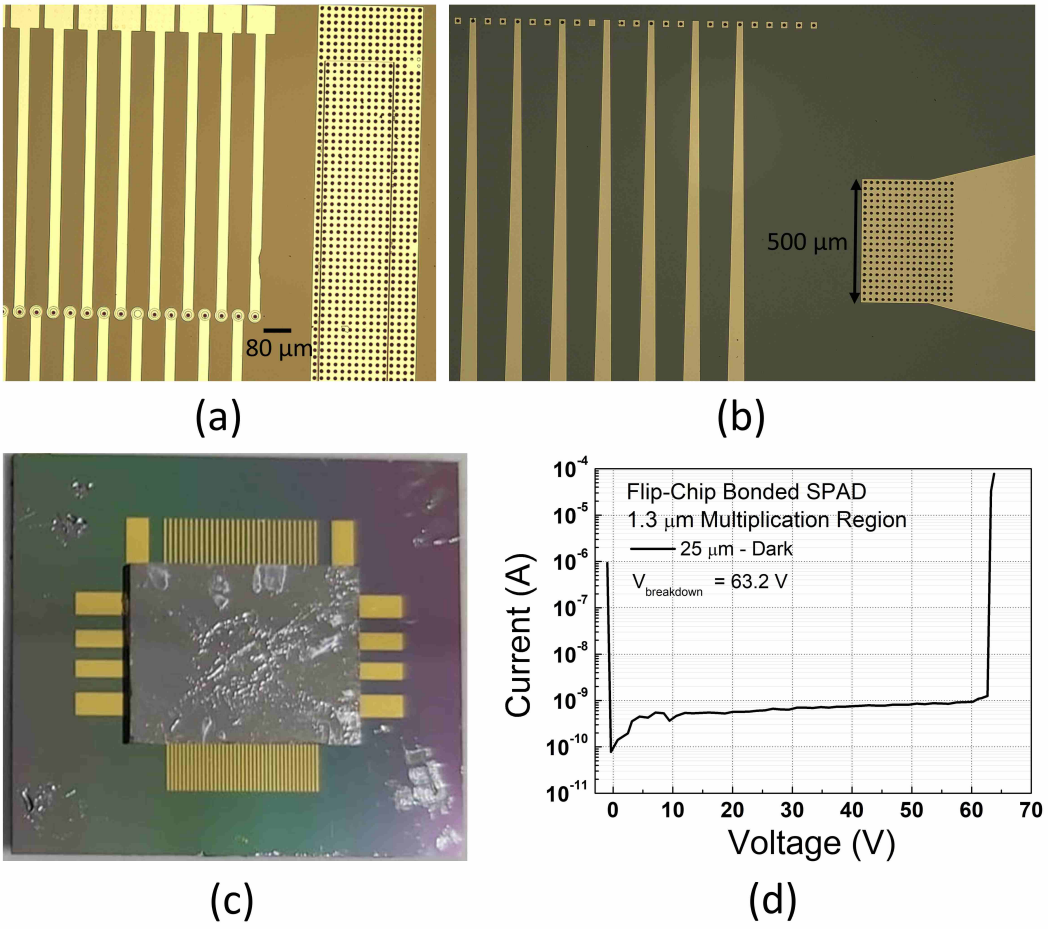


Figure 5.36: (a) A fabricated 1×128 InGaAsP SPAD array with 25 µm active diameter and 60 µm pixel pitch. (b) A fabricated Si fan-out. (c) An image of an InGaAsP SPAD array (on top) flip-chip-bonded with a Si fan-out (at the bottom). (d) An I-V characteristic from a flip-chip-bonded 25 µm diameter SPAD.

SPAD array is shown in Fig 5.36 (a). For the Si chip, a Si fan-out was first fabricated in the cleanroom, as there was no available ROIC at this point in time for this array format. The fan-outs consist of metal pads with UBM and In bumps, where the metal pads are aligned with the anode and cathode metals at the InGaAsP chip. Also, the fan-outs were fabricated on Si substrate with 4 µm thick grown silicon oxide (SiO_2) on top of it by wet oxidation. SiO_2 would provide electrical isolation between the metal pads. The anode and cathode metals on the fan-out were also extended so that they reach outside the area defined by the InGaAsP chip. Thanks to this extension of metal pads, wirebonding can be performed on the pads to measure the chips. A microscope image of a fabricated Si fan-out is illustrated in Fig. 5.36 (b). Upon finalizing the fabrication, the chips were taken to a company producing flip-chip bonding machines and providing service. After several trials on-site, the bonding recipe was optimized. In the 4-minute recipe, 0.12 grams of force were applied per bump throughout the entire recipe. The InGaAsP chip was placed on the chuck, whose temperature was raised

to 160 ∘C after the first minute. Si fan-out was picked up by the arm, whose temperature was increased to 140 °C after the second minute. The melting temperature of the indium is around 157 °C. Therefore, the idea was to slightly melt the bumps on one side and solder the bumps with the applied force. As a result, an image of the flip-chip bonded chips is given in Fig. 5.36 (c). To demonstrate that the bonded chip is working, an I-V curve of a SPAD pixel was obtained, as shown in Fig. 5.36 (d). The observed avalanche breakdown proves that the SPAD array is functional after the bonding process. The developed 3D bonding methodology is promising to integrate InGaAs(P) SPAD arrays with Si ROICs in the near f uture.

6 Conclusions and Future Perspective

6.1 Conclusions and the summary of the works

In this thesis, Si CMOS and InGaAs(P)/InP-based SPAD devices that can be used in NIR/SWIR applications, such as LiDAR, QKD, or NIROT, were investigated. These two material systems were selected because they allow for creating large-format SPAD arrays operating at room temperature or under moderate cooling with a thermoelectric cooler. In both systems, the modeling and simulations were performed on TCAD to numerically develop the device structures. Then, the taped-out or fabricated devices were characterized in detail, in terms of I-V, DCR, PDP, timing jitter, active area scanning, and afterpulsing, to demonstrate that the device can effectively function in Geiger mode.

Regarding Si CMOS SPADs, the wide depletion region approach that is used to enhance NIR efficiencies was focused on. However, achieving wide depletion devices is challenging in more advanced CMOS technology due to the increased doping concentrations of the layers. Besides, they require very high V_{ex} to achieve high PDPs, which can complicate the integration with circuits. Therefore, the goal was to make the wide depletion region approach more feasible and better-performing by designing SPADs using a 110 nm CIS technology.

- First, the doping compensation technique was implemented, which enables designers to modify the restricted doping profiles provided by foundries. In this work, a non-compensated N^+/HVPW junction and the same junction, where the HVPW was compensated by the HVNW layer were analyzed as SPADs. Essentially, the reverse-polarized HVNW layer reduces the p-doping of the HVPW, extending the depletion layer width of the p-n junction. Thus, it was shown that doping compensation can be a method to create wider depletion region devices in advanced technologies, especially with TCAD simulations and PDP measurements. The comparison of the compensated and non-compensated devices indicated that higher NIR PDPs can be obtained with compensation, as well as that the tunneling noise in thinner junctions can be mitigated thanks to the wider depletion region. Reduced noise also allows the device to

be biased at higher V_{ex}, hence increasing the maximum achievable PDP. At 2 V_{ex}, the compensated device achieved a PDP of 7.3% at 800 nm and 5% at 850 nm, whereas the non-compensated device attained PDPs of 5.9% at 800 nm and 4.3% at 850 nm. At 5 V_{ex}, the peak PDP reaches 73.9% at 440 nm, and NIR PDPs become 10.7% and 7.3% at 800 nm and 850 nm wavelengths in the compensated device, with a noise of 962 cps at room temperature and a timing jitter of 68 ps at 850 nm.

- Secondly, a new double multiplication region concept was demonstrated. The idea behind this technique was to insert a second multiplication region in a wide depletion region to enhance the total avalanche breakdown probability. In conventional wide depletion devices, there is only one multiplication region where the magnitude of the confined electric field is relatively lower than at thin junctions, and a very high excess bias voltage is needed to achieve a high electric field and consequently, high breakdown probabilities. In the double multiplication region scenario, a second avalanche multiplication region allows (a) to trigger both electrons and holes photogenerated in the depletion region, and (b) to give the diffused carriers a second chance to undergo avalanche breakdown. Therefore, with this technique, similar NIR PDP levels at a relatively lower V_{ex} can be achieved. The downside of the technique could be the increased DCR since the total avalanche breakdown probability is also enhanced for thermally generated carriers or carriers generated via tunneling. In this work, it was shown that inserting a second multiplication region in a depletion region is indeed possible. Thanks to a double-peaked p-well layer, an n-p-n-p type structure was formed in junction with the HVNW layer. This structure should effectively create two p-n junctions, leading to two multiplication and two depletion regions. However, it was demonstrated that if the thickness of the second n polarity is thin enough, two depletion regions are merged, and two high-electric field-confined regions are separated in the same depletion zone. The DCR of the fabricated device was 60 cps and 295 cps at 4 V_{ex} and 5.5 V_{ex}, respectively, at room temperature. The NIR PDPs were around 25.5% at both 800 nm and 850 nm wavelengths, while the peak PDP reached 78% at 500 nm. High PDPs achieved were thanks to a wider depletion region, a substrate non-isolated structure allowing to collect diffused electrons from the neutral regions, and enhanced avalanche breakdown probabilities via the double multiplication region technique. However, the timing jitter was deteriorated, reaching 236 ps at 850 nm, due to the diffused carriers, which created a diffusion peak in the histogram.

- Finally, a guard ring optimization study was conducted on a substrate-isolated wide depletion region SPAD. Isolation was achieved with the buried n-well implantation beneath the main junction of p^+/HVNW. The depletion region was further extended thanks to the junction of HVNW/p-epi. Three types of GR structures were investigated to be able to operate the device and optimize its performance. Virtual p-epi GR showed edge breakdown in both simulations and light emission tests. HVPW GR yielded noisy devices, which might be related to the increased trap density with the implementation of this layer. The devices with the p-well GR achieved 144 cps DCR, 57% peak and 7%

PDPs at 460 and 850 nm wavelengths, and 58.5 ps timing jitter at 850 nm, all at 8 V_{ex}. The lower PDP achieved compared to the previously mentioned devices is because of the photogenerated holes going through the impact ionization in the depletion region, which have a lower ionization coefficient than the electrons, and also due to the substrate non-isolated structure. This study indicated the importance of designing a proper GR to be able to operate a wide depletion region SPAD under optimum conditions.

Regarding InGaAs(P)/InP SPADs, 1.06 μm and 1.55 μm wavelength detection was targeted.

- In order to reduce the noise for 1.06 μm detection, the bandgap of InGaAs was increased by the incorporation of P atoms, thereby obtaining an InGaAsP absorber with a 1.1 μm cutoff wavelength. The devices were grown without any p-doped layers. The p-doping and p-n junctions were realized by the double Zn diffusion technique, forming planar structures. 10 μm active diameter SPADs with 2.5 μm GRs were more focused on, since they were the smallest manufacturable size. Careful modeling and simulations enabled the removal of FGRs, which have more potential to shrink pixel pitch. In that sense, in InGaAsP SPADs, the 0.25 μm and 0.5 μm depth differences between the deep and shallow diffusions were investigated. TCAD simulations always showed slightly enhanced electric fields at the edges of the active region due to junction curvature effect, independent of the depth difference. However, it was found out that the avalanche breakdown probability difference between the center and active region edge becomes more significant while increasing the Zn diffusion depth difference. In other words, a more uniform photoresponse over the active area can be achieved with a 0.25 μm depth difference. On the other hand, it was discovered that the edges of the GRs might go into avalanche breakdown for the 0.25 μm depth difference device, and avalanche contribution at the GRs can deteriorate the noise performance of the SPADs. It was experimentally shown that indeed the DCR greatly increases in the 0.25 μm depth difference device after several V_{ex}. In the SPADs with a 0.5 μm depth difference, the noise was lowered, and uniform active area response was obtained at 5 V_{ex} to 6 V_{ex}. Since noise is very important for near room temperature operation, three different multiplication region thicknesses were analyzed for the SPADs with the 0.5 μm depth difference. At 300K and 5 V_{ex}, a DCR of 53 kcps for a multiplication thickness of 1.5 μm, 302 kcps for 1.3 μm, and 2130 kcps for 0.75 μm were obtained, while operating in time-gating mode with 10 kHz frequency and 100 ns gate-on time. The DCR was reduced to 14.1 kcps at 273K, 5.5 kcps at 253K, and 2.75 kcps at 225K, for the 1.5 μm multiplication region device at 5 V_{ex}. The maximum operating frequency can be increased up to 400-500 kHz for each multiplication region thickness while ensuring low afterpulsing probability at room temperature. The calculated APPs for 500 kHz and 200 kHz gating frequencies were 11.1% and 5.8% at room temperature. The PDPs at 1.06 μm wavelength and 5 V_{ex} are 19.5% for a multiplication thickness of 1.5 μm, 20.4% for 1.3 μm, and 21.5% for 0.75 μm. A high PDP of 36% at 9 V_{ex} was achieved in the SPAD with a 1.5 μm multiplication at the same wavelength. The timing jitters were acquired as 118.4 ps, 110

ps, and 84 ps FWHM for the corresponding multiplication region thicknesses.

- For the 1.55 μm detection, the absorber material was InGaAs with a 1.7 μm cutoff wavelength. Therefore, the increase in the DCR was expected due to enhanced thermal generation in this SPAD. 1.5 μm multiplication region with 0.25 μm depth difference in Zn diffusions were implemented in the InGaAs SPAD. A smaller depth difference was preferred to achieve a more uniform photoresponse and to enhance the PDP. In order to mitigate the avalanche breakdown at GRs, 7.5 μm was identified as the optimum GR width for 10 μm active diameter devices. In this device, the DCR attained 8.6×10^6 cps at 300K, 2.79×10^6 cps at 273K, 6.52×10^5 cps at 253K, and 42.8 kcps at 225K, with 10 kHz gating and 100 ns gate-on time. Active area scanning showed that the device response was very uniform at both 3 V_{ex} and 5 V_{ex}. The PDP of the device at 1550 nm reached around 15% and 20% at 5 V_{ex} and 6 V_{ex}. The timing jitter was 123 ps at 5 V_{ex} and 1550 nm wavelength and had an exponential tail with an 82 ps time constant. Although the other metrics were acceptable, it was concluded that the DCR of this device is higher than the state-of-the-art, and further optimization in the epi-structure.

- Finally, a 3D integration procedure for the InGaAs(P)/InP-based SPADs with Si ROICs has been developed, and several 1×128 pixel array were successfully flip-chip bonded. This study is going to form a basis to bond 96×96 InGaAs(P) SPAD array with an in-house developed Si ROIC.

6.2 Future perspective

In Si CMOS SPADs, the demonstrated doping compensation and double multiplication region techniques were implemented only in a 110 nm CIS technology. Both techniques can be investigated in other CMOS nodes to prove their efficiency in designing CMOS-based SPADs with enhanced NIR PDPs. For instance, doping compensation could yield in a wider depletion region SPADs so that the increase in NIR PDP can be seen more clearly while comparing it with a non-compensated SPAD. Likewise, with the implementation of the double multiplication region method in other CMOS technologies, reducing the need for high excess bias voltage or increasing the maximum achievable NIR PDPs in the wide depletion region approach can be further verified. As performed in the simulations, if a control SPAD with one multiplication region can be fabricated in the same technology, the comparison of the characterization results with a control SPAD that has a similar depletion region width would also be a nice follow-up to show the proposed technique's effectiveness. Another future work of Si SPADs would be integrating them with on-chip quench and recharge circuits, creating arrays of SPADs, and optimizing the pixel pitch and structure for a low optical crosstalk probability to be utilized in LiDAR or NIROT applications. Larger diameter devices can also be designed and fabricated, considering QKD applications.

In InGaAs(P)/InP SPADs, the effort has been mostly made to optimize the Zn diffusion process and profiles. However, DCR measurements showed that the noise can be further reduced.

First, the trap concentration can be measured, for example, by utilizing deep-level transient spectroscopy (DLTS). This would definitely reveal the material quality and whether thermal (SRH) generation is dominating the DCR. In the case of high trap concentration, additional annealing processes can be considered to further mitigate it, or other foundries that can grow better-quality InGaAs(P) and InP can be investigated. If the trap concentration is at acceptable levels, epi-layer optimization is required, as it means that TAT is dominating the DCR, which is related to the electric field magnitude in the multiplication and absorber regions. Therefore, efforts can be made at the foundry side to decrease the background doping concentration of InP, which results in lower peak electric fields in the multiplication region. Moreover, particularly the charge layer doping sweep can be investigated since it adjusts the electric field in the absorber and should be precisely controlled. Even a small deviation in the charge doping can affect the absorber electric field significantly, as demonstrated in TCAD simulations. Optimizing the charge layer doping could mitigate tunneling that might be coming from the InGaAs(P) absorber. After optimizing the epi-structure and material quality, arrays of InGaAs(P)/InP-based SPADs can be fabricated, which are going to be integrated with Si ROICs to be utilized in LiDAR applications. Optical crosstalk characterization and techniques to reduce it can be performed and studied. As also mentioned, larger diameter devices targeting the 1550 nm wavelength can be investigated for better optical coupling in QKD applications.

Bibliography

[1] R. G. Driggers, V. Hodgkin, and R. Vollmerhausen, "What good is swir? passive day comparison of vis, nir, and swir", in *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXIV*, SPIE, vol. 8706, 2013, pp. 187–201.

[2] M. P. Hansen and D. S. Malchow, "Overview of swir detectors, cameras, and applications", in *Thermosense Xxx*, SPIE, vol. 6939, 2008, pp. 94–104.

[3] A. López-Maestresalas, J. C. Keresztes, M. Goodarzi, S. Arazuri, C. Jarén, and W. Saeys, "Non-destructive detection of blackspot in potatoes by vis-nir and swir hyperspectral imaging", *Food control*, vol. 70, pp. 229–241, 2016.

[4] A. Siedliska, P. Baranowski, M. Zubik, W. Mazurek, and B. Sosnowska, "Detection of fungal infections in strawberry fruit by vnir/swir hyperspectral imaging", *Postharvest Biology and Technology*, vol. 139, pp. 115–126, 2018.

[5] L. J. Kozlowski, W. E. Tennant, M. Zandian, J. M. Arias, and J. G. Pasko, "Swir staring fpa performance at room temperature", in *Infrared Detectors and Focal Plane Arrays IV*, SPIE, vol. 2746, 1996, pp. 93–100.

[6] R. H. Hadfield, "Single-photon detectors for optical quantum information applications", *Nature photonics*, vol. 3, no. 12, pp. 696–705, 2009.

[7] M. D. Eisaman, J. Fan, A. Migdall, and S. V. Polyakov, "Invited review article: single-photon sources and detectors", *Review of scientific instruments*, vol. 82, no. 7, p. 071 101, 2011.

[8] C. H. Bennett, "Quantum cryptography using any two nonorthogonal states", *Physical review letters*, vol. 68, no. 21, p. 3121, 1992.

[9] H. Inamori, N. Lütkenhaus, and D. Mayers, "Unconditional security of practical quantum key distribution", *The European Physical Journal D*, vol. 41, pp. 599–627, 2007.

[10] E. Wu, V. Jacques, H. Zeng, P. Grangier, F. Treussart, and J.-F. Roch, "Narrow-band single-photon emission in the near infrared for quantum key distribution", *Optics express*, vol. 14, no. 3, pp. 1296–1303, 2006.

[11] T. Heindel, C. A. Kessler, M. Rau, *et al.*, "Quantum key distribution using quantum dot single-photon emitting diodes in the red and near infrared spectral range", *New Journal of Physics*, vol. 14, no. 8, p. 083 001, 2012.

[12] N. Peters, P. Toliver, T. Chapuran, *et al.*, "Dense wavelength multiplexing of 1550 nm qkd with strong classical channels in reconfigurable networking environments", *New Journal of physics*, vol. 11, no. 4, p. 045 012, 2009.

[13] R. H. Hadfield, J. L. Habif, J. Schlafer, R. E. Schwall, and S. W. Nam, "Quantum key distribution at 1550nm with twin superconducting single-photon detectors", *Applied physics letters*, vol. 89, no. 24, 2006.

[14] J. Tachella, Y. Altmann, N. Mellado, *et al.*, "Real-time 3d reconstruction from single-photon lidar data using plug-and-play point cloud denoisers", *Nature communications*, vol. 10, no. 1, p. 4984, 2019.

[15] S. W. Hutchings, N. Johnston, I. Gyongy, *et al.*, "A reconfigurable 3-d-stacked spad imager with in-pixel histogramming for flash lidar or high-speed time-of-flight imaging", *IEEE Journal of Solid-State Circuits*, vol. 54, no. 11, pp. 2947–2956, 2019.

[16] J. M. Pavia, M. Scandini, S. Lindner, M. Wolf, and E. Charbon, "A 1 ×400 backside-illuminated spad sensor with 49.7 ps resolution, 30 pj/sample tdcs fabricated in 3d cmos technology for near-infrared optical tomography", *IEEE Journal of Solid-State Circuits*, vol. 50, no. 10, pp. 2406–2418, 2015.

[17] L. You, J. Quan, Y. Wang, *et al.*, "Superconducting nanowire single photon detection system for space applications", *Optics Express*, vol. 26, no. 3, pp. 2965–2971, 2018.

[18] M. H. Smith, R. L. Fork, and S. T. Cole, "Safe delivery of optical power from space", *Optics Express*, vol. 8, no. 10, pp. 537–546, 2001.

[19] D. Malchow, J. Battaglia, R. Brubaker, and M. Ettenberg, "High speed short wave infrared (swir) imaging and range gating cameras", in *Thermosense XXIX*, SPIE, vol. 6541, 2007, pp. 36–47.

[20] K. Miah and D. K. Potter, "A review of hybrid fiber-optic distributed simultaneous vibration and temperature sensing technology and its geophysical applications", *Sensors*, vol. 17, no. 11, p. 2511, 2017.

[21] K. Yoshioka, "A tutorial and review of automobile direct tof lidar socs: evolution of next-generation lidars", *IEICE Transactions on Electronics*, vol. 105, no. 10, pp. 534–543, 2022.

[22] C. S. Bamji, P. O'Connor, T. Elkhatib, *et al.*, "A 0.13 μm cmos system-on-chip for a 512 × 424 time-of-flight image sensor with multi-frequency photo-demodulation up to 130 mhz and 2 gs/s adc", *IEEE Journal of Solid-State Circuits*, vol. 50, no. 1, pp. 303–319, 2014.

[23] C. Niclass, M. Soga, H. Matsubara, S. Kato, and M. Kagami, "A 100-m range 10-frame/s 340 × 96-pixel time-of-flight depth sensor in 0.18-μm cmos", *IEEE Journal of Solid-State Circuits*, vol. 48, no. 2, pp. 559–572, 2012.

[24] A. R. Ximenes, P. Padmanabhan, M.-J. Lee, Y. Yamashita, D.-N. Yaung, and E. Charbon, "A 256 × 256 45/65nm 3d-stacked spad-based direct tof image sensor for lidar applications with optical polar modulation for up to 18.6 db interference suppression", in *2018 IEEE International Solid-State Circuits Conference-(ISSCC)*, IEEE, 2018, pp. 96–98.

[25] K. Yoshioka, H. Kubota, T. Fukushima, *et al.*, "A 20-ch tdc/adc hybrid architecture lidar soc for 240 × 96 pixel 200-m range imaging with smart accumulation technique and residue quantizing sar adc", *IEEE Journal of Solid-State Circuits*, vol. 53, no. 11, pp. 3026–3038, 2018.

[26] O. Kumagai, J. Ohmachi, M. Matsumura, *et al.*, "A 189× 600 back-illuminated stacked spad direct time-of-flight depth sensor for automotive lidar systems", in *2021 IEEE International Solid-State Circuits Conference (ISSCC)*, IEEE, vol. 64, 2021, pp. 110–112.

[27] C. Zhang, S. Lindner, I. M. Antolović, J. M. Pavia, M. Wolf, and E. Charbon, "A 30-frames/s,252 ×144 spad flash lidar with 1728 dual-clock 48.8-ps tdcs, and pixel-wise integrated histogramming", *IEEE Journal of Solid-State Circuits*, vol. 54, no. 4, pp. 1137–1151, 2018.

[28] P. Padmanabhan, C. Zhang, M. Cazzaniga, *et al.*, "7.4 a 256 × 128 3d-stacked (45nm) spad flash lidar with 7-level coincidence detection and progressive gating for 100m range and 10klux background light", in *2021 IEEE International Solid-State Circuits Conference (ISSCC)*, IEEE, vol. 64, 2021, pp. 111–113.

[29] S. Lindner, C. Zhang, I. M. Antolovic, M. Wolf, and E. Charbon, "A 252 × 144 spad pixel flash lidar with 1728 dual-clock 48.8 ps tdcs, integrated histogramming and 14.9-to-1 compression in 180nm cmos technology", in *2018 IEEE Symposium on VLSI Circuits*, IEEE, 2018, pp. 69–70.

[30] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, "Quantum cryptography", *Reviews of modern physics*, vol. 74, no. 1, p. 145, 2002.

[31] S. Wang, Z.-Q. Yin, D.-Y. He, *et al.*, "Twin-field quantum key distribution over 830-km fibre", *Nature photonics*, vol. 16, no. 2, pp. 154–161, 2022.

[32] Y.-A. Chen, Q. Zhang, T.-Y. Chen, *et al.*, "An integrated space-to-ground quantum communication network over 4,600 kilometres", *Nature*, vol. 589, no. 7841, pp. 214–219, 2021.

[33] W. Li, L. Zhang, H. Tan, *et al.*, "High-rate quantum key distribution exceeding 110 mb s–1", *Nature Photonics*, vol. 17, no. 5, pp. 416–421, 2023.

[34] A. Boaron, G. Boso, D. Rusca, *et al.*, "Secure quantum key distribution over 421 km of optical fiber", *Physical review letters*, vol. 121, no. 19, p. 190 502, 2018.

[35] M. Avesani, L. Calderaro, M. Schiavon, *et al.*, "Full daylight quantum-key-distribution at 1550 nm enabled by integrated silicon photonics", *npj Quantum Information*, vol. 7, no. 1, p. 93, 2021.

[36] P. A. Hiskett, G. Bonfrate, G. S. Buller, and P. D. Townsend, "Eighty kilometre transmission experiment using an ingaas/inp spad-based quantum cryptography receiver operating at 1.55 μm", *Journal of Modern Optics*, vol. 48, no. 13, pp. 1957–1966, 2001.

[37] N. Namekata, H. Takesue, T. Honjo, Y. Tokura, and S. Inoue, "High-rate quantum key distribution over 100 km using ultra-low-noise, 2-ghz sinusoidally gated ingaas/inp avalanche photodiodes", *Optics express*, vol. 19, no. 11, pp. 10 632–10 639, 2011.

[38] D. Ribezzo, M. Zahidy, G. Lemmi, *et al.*, "Quantum key distribution over 100 km underwater optical fiber assisted by a fast-gated single-photon detector", *arXiv preprint arXiv:2303.01449*, 2023.

[39] S. Wray, M. Cope, D. T. Delpy, J. S. Wyatt, and E. O. R. Reynolds, "Characterization of the near infrared absorption spectra of cytochrome aa3 and haemoglobin for the non-invasive monitoring of cerebral oxygenation", *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, vol. 933, no. 1, pp. 184–192, 1988.

[40] S. R. Arridge, "Optical tomography in medical imaging", *Inverse problems*, vol. 15, no. 2, R41, 1999.

[41] A. Kalyanov, J. Jiang, S. Lindner, *et al.*, "Time domain near-infrared optical tomography with time-of-flight spad camera: the new generation", in *Optical Tomography and Spectroscopy*, Optica Publishing Group, 2018, OF4D–5.

[42] M. S. Patterson, B. Chance, and B. C. Wilson, "Time resolved reflectance and transmittance for the noninvasive measurement of tissue optical properties", *Applied optics*, vol. 28, no. 12, pp. 2331–2336, 1989.

[43] H. Photonics, *Photomultiplier tubes basics and applications*. [Online]. Available: https://www.hamamatsu.com/content/dam/hamamatsu-photonics/sites/documents/99_SALES_LIBRARY/etd/PMT_handbook_v4E.pdf.

[44] M. P. Inc., *Photomultiplier tube (pmt)*. [Online]. Available: https://www.matsusada.com/application/ps/photomultiplier_tubes/.

[45] K. Nakamura, Y. Hamana, Y. Ishigami, and T. Matsui, "Latest bialkali photocathode with ultra high sensitivity", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 623, no. 1, pp. 276–278, 2010.

[46] B. G. Streetman and S. Banerjee, *Solid state electronic devices*. Pearson Prentice Hall, 2006.

[47] C. Veerappan, Y. Maruyama, and E. Charbon, "Silicon integrated electrical micro-lens for cmos spads based on avalanche propagation phenomenon", in *Int. Image Sensor Workshop*, 2013.

[48] K. Morimoto, A. Ardelean, M.-L. Wu, *et al.*, "Megapixel time-gated spad image sensor for 2d and 3d imaging applications", *Optica*, vol. 7, no. 4, pp. 346–354, 2020.

[49] A. Rochas, "Single photon avalanche diodes in cmos technology", PhD thesis, Ecole Polytechnique´ Fédérale de Lausanne, Lausanne, Switzerland, 2003.

[50] E. Charbon, "Single-photon imaging in complementary metal oxide semiconductor processes", *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 372, no. 2012, p. 20 130 100, 2014.

[51] M. Sanzaro, P. Gattari, F. Villa, A. Tosi, G. Croce, and F. Zappa, "Single-photon avalanche diodes in a 0.16 μm bcd technology with sharp timing response and red-enhanced sensitivity", *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 24, no. 2, pp. 1–9, 2017.

[52] K. Morimoto, J. Iwata, M. Shinohara, *et al.*, "3.2 megapixel 3d-stacked charge focusing spad for low-light imaging and depth sensing", in *2021 IEEE International Electron Devices Meeting (IEDM)*, IEEE, 2021, pp. 20–2.

[53] E. A. Webster, L. A. Grant, and R. K. Henderson, "A high-performance single-photon avalanche diode in 130-nm cmos imaging technology", *IEEE Electron Device Letters*, vol. 33, no. 11, pp. 1589–1591, 2012.

[54] M.-J. Lee, U. Karaca, E. Kizilkan, C. Bruschini, and E. Charbon, "A 73% peak pdp single-photon avalanche diode implemented in 110 nm cis technology with doping compensation", *IEEE Journal of Selected Topics in Quantum Electronics*, 2023.

[55] F. Gramuglia, M.-L. Wu, C. Bruschini, M.-J. Lee, and E. Charbon, "A low-noise cmos spad pixel with 12.1 ps sptr and 3 ns dead time", *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 28, no. 2: Optical Detectors, pp. 1–9, 2021.

[56] A. C. Ulku, C. Bruschini, I. M. Antolović, *et al.*, "A 512×512 spad image sensor with integrated gating for widefield flim", *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 25, no. 1, pp. 1–12, 2018.

[57] I. Gyongy, N. Calder, A. Davies, *et al.*, "A 256× 256, 100-kfps, 61% fill-factor spad image sensor for time-resolved microscopy applications", *IEEE Transactions on Electron Devices*, vol. 65, no. 2, pp. 547–554, 2017.

[58] M. Perenzoni, N. Massari, D. Perenzoni, L. Gasparini, and D. Stoppa, "A 160× 120 pixel analog-counting single-photon imager with time-gating and self-referenced column-parallel a/d conversion for fluorescence lifetime imaging", *IEEE Journal of Solid-State Circuits*, vol. 51, no. 1, pp. 155–167, 2015.

[59] R. K. Henderson, N. Johnston, H. Chen, *et al.*, "A 192× 128 time correlated single photon counting imager in 40nm cmos technology", in *ESSCIRC 2018-IEEE 44th European Solid State Circuits Conference (ESSCIRC)*, IEEE, 2018, pp. 54–57.

[60] Y. Hirose, S. Koyama, T. Okino, *et al.*, "5.6 a 400× 400-pixel 6μm-pitch vertical avalanche photodiodes cmos image sensor based on 150ps-fast capacitive relaxation quenching in geiger mode for synthesis of arbitrary gain images", in *2019 IEEE International Solid-State Circuits Conference-(ISSCC)*, IEEE, 2019, pp. 104–106.

[61] F. Signorelli, F. Telesca, E. Conca, *et al.*, "Ingaas/inp spad detecting single photons at 1550 nm with up to 50% efficiency and low noise", in *2021 IEEE International Electron Devices Meeting (IEDM)*, IEEE, 2021, pp. 20–3.

[62] B. F. Aull, E. K. Duerr, J. P. Frechette, K. A. McIntosh, D. R. Schuette, and R. D. Younger, "Large-format geiger-mode avalanche photodiode arrays and readout circuits", *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 24, no. 2, pp. 1–10, 2017.

[63] M. A. Itzler, X. Jiang, M. Entwistle, *et al.*, "Advances in ingaasp-based avalanche diode single photon detectors", *Journal of Modern Optics*, vol. 58, no. 3-4, pp. 174–200, 2011.

[64] M. A. Itzler, M. Entwistle, U. Krishnamachari, *et al.*, "Swir geiger-mode apd detectors and cameras for 3d imaging", in *Advanced Photon Counting Techniques VIII*, SPIE, vol. 9114, 2014, pp. 37–48.

[65] S.-H. Baek, S.-C. Yang, C.-Y. Park, C.-W. Park, S.-B. Cho, and S.-W. Ryu, "Room temperature quantum key distribution characteristics of low-noise ingaas/inp single-photon avalanche diode", *Journal of the Korean Physical Society*, vol. 78, no. 7, pp. 634–641, 2021.

[66] Y.-Q. Fang, W. Chen, T.-H. Ao, *et al.*, "Ingaas/inp single-photon detectors with 60% detection efficiency at 1550 nm", *Review of Scientific Instruments*, vol. 91, no. 8, 2020.

[67] P. Vines, K. Kuzmenko, J. Kirdoda, *et al.*, "High performance planar germanium-on-silicon single-photon avalanche diode detectors", *Nature communications*, vol. 10, no. 1, p. 1086, 2019.

[68] K. M. Rosfjord, J. K. Yang, E. A. Dauler, *et al.*, "Nanowire single-photon detector with an integrated optical cavity and anti-reflection coating", *Optics express*, vol. 14, no. 2, pp. 527–534, 2006.

[69] F. Marsili, V. B. Verma, J. A. Stern, *et al.*, "Detecting single infrared photons with 93% system efficiency", *Nature Photonics*, vol. 7, no. 3, pp. 210–214, 2013.

[70] W. Zhang, L. You, H. Li, *et al.*, "Nbn superconducting nanowire single photon detector with efficiency over 90% at 1550 nm wavelength operational at compact cryocooler temperature", *Science China Physics, Mechanics & Astronomy*, vol. 60, pp. 1–10, 2017.

[71] S. Miki, T. Yamashita, H. Terai, and Z. Wang, "High performance fiber-coupled nbtin superconducting nanowire single photon detectors with gifford-mcmahon cryocooler", *Optics express*, vol. 21, no. 8, pp. 10 208–10 214, 2013.

[72] L.-D. Kong, H. Wang, Q.-Y. Zhao, *et al.*, "Readout-efficient superconducting nanowire single-photon imager with orthogonal time–amplitude multiplexing by hotspot quantization", *Nature Photonics*, vol. 17, no. 1, pp. 65–72, 2023.

[73] D. Fukuda, "Single-photon measurement techniques with a superconducting transition edge sensor", *IEICE Transactions on Electronics*, vol. 102, no. 3, pp. 230–234, 2019.

[74] D. Fukuda, G. Fujii, T. Numata, *et al.*, "Titanium-based transition-edge photon number resolving detector with 98% detection efficiency with index-matched small-gap fiber coupling", *Optics express*, vol. 19, no. 2, pp. 870–875, 2011.

[75] A. E. Lita, A. J. Miller, and S. W. Nam, "Counting near-infrared single-photons with 95% efficiency", *Optics express*, vol. 16, no. 5, pp. 3032–3040, 2008.

[76] A. Lamas-Linares, B. Calkins, N. A. Tomlin, *et al.*, "Nanosecond-scale timing jitter for single photon detection in transition edge sensors", *Applied Physics Letters*, vol. 102, no. 23, 2013.

[77] E. Gansen, M. A. Rowe, M. Greene, *et al.*, "Photon-number-discriminating detection using a quantum-dot, optically gated, field-effect transistor", *Nature photonics*, vol. 1, no. 10, pp. 585–588, 2007.

[78] M. A. Rowe, E. Gansen, M. Greene, *et al.*, "Single-photon detection using a quantum dot optically gated field-effect transistor with high internal quantum efficiency", *Applied physics letters*, vol. 89, no. 25, 2006.

[79] J. Blakesley, P. See, A. Shields, *et al.*, "Efficient single photon detection by quantum dot resonant tunneling diodes", *Physical review letters*, vol. 94, no. 6, p. 067 401, 2005.

[80] Y. Okuto and C. Crowell, "Ionization coefficients in semiconductors: a nonlocalized property", *Physical review B*, vol. 10, no. 10, p. 4284, 1974.

[81] S. M. Sze, Y. Li, and K. K. Ng, *Physics of semiconductor devices*. John wiley & sons, 2021.

[82] W. Maes, K. De Meyer, and R. Van Overstraeten, "Impact ionization in silicon: a review and update", *Solid-State Electronics*, vol. 33, no. 6, pp. 705–718, 1990.

[83] L. Cook, G. Bulman, and G. Stillman, "Electron and hole impact ionization coefficients in inp determined by photomultiplication measurements", *Applied Physics Letters*, vol. 40, no. 7, pp. 589–591, 1982.

[84] C. Armiento and S. Groves, "Impact ionization in (100),(110), and (111) oriented inp avalanche photodiodes", *Applied Physics Letters*, vol. 43, no. 2, pp. 198–200, 1983.

[85] A. Kyuregyan and S. Yurkov, "Room-temperature avalanche breakdown voltages of p-n# 3 junctions made of si, ge, sic, gaas, gap, and inp", *Soviet Physics–Semiconductors(English Translation)*, vol. 23, no. 10, pp. 1126–31, 1989.

[86] M. W. Fishburn, *Fundamentals of CMOS single-photon avalanche diodes*. fishburn, 2012.

[87] A. Spinelli and A. L. Lacaita, "Physics and numerical simulation of single photon avalanche diodes", *IEEE Transactions on Electron Devices*, vol. 44, no. 11, pp. 1931–1943, 1997.

[88] S. Cova, M. Ghioni, A. Lacaita, C. Samori, and F. Zappa, "Avalanche photodiodes and quenching circuits for single-photon detection", *Applied optics*, vol. 35, no. 12, pp. 1956–1976, 1996.

[89] A. Eisele, R. Henderson, B. Schmidtke, *et al.*, "185 mhz count rate 139 db dynamic range single-photon avalanche diode with active quenching circuit in 130 nm cmos technology", in *Proc. Int. Image Sensor Workshop*, 2011, pp. 278–280.

[90] D. Bronzi, S. Tisa, F. Villa, S. Bellisai, A. Tosi, and F. Zappa, "Fast sensing and quenching of cmos spads for minimal afterpulsing effects", *IEEE Photonics Technology Letters*, vol. 25, no. 8, pp. 776–779, 2013.

[91] F. Acerbi, A. Della Frera, A. Tosi, and F. Zappa, "Fast active quenching circuit for reducing avalanche charge and afterpulsing in ingaas/inp single-photon avalanche diode", *IEEE Journal of Quantum Electronics*, vol. 49, no. 7, pp. 563–569, 2013.

[92] L. Keldysh, "Concerning the theory of impact ionization in semiconductors", *Sov. Phys. JETP*, vol. 21, no. 6, p. 1135, 1965.

[93] C. Groves, R. Ghin, J. David, and G. Rees, "Temperature dependence of impact ionization in gaas", *IEEE Transactions on electron devices*, vol. 50, no. 10, pp. 2027–2031, 2003.

[94] S. Sze and G. Gibbons, "Effect of junction curvature on breakdown voltage in semiconductors", *Solid-state electronics*, vol. 9, no. 9, pp. 831–845, 1966.

[95] M. S. Adler, V. A. Temple, A. P. Ferro, and R. Rustay, "Theory and breakdown voltage for planar devices with a single field limiting ring", *IEEE Transactions on electron Devices*, vol. 24, no. 2, pp. 107–113, 1977.

[96] J. A. Richardson, E. A. Webster, L. A. Grant, and R. K. Henderson, "Scaleable single-photon avalanche diode structures in nanometer cmos technology", *IEEE Transactions on Electron Devices*, vol. 58, no. 7, pp. 2028–2035, 2011.

[97] F. Zappa, S. Tisa, A. Tosi, and S. Cova, "Principles and features of single-photon avalanche diode arrays", *Sensors and Actuators A: Physical*, vol. 140, no. 1, pp. 103–112, 2007.

[98] J. P. Donnelly, E. K. Duerr, K. A. McIntosh, *et al.*, "Design considerations for 1.06-*mu* m ingaasp–inp geiger-mode avalanche photodiodes", *IEEE Journal of Quantum Electronics*, vol. 42, no. 8, pp. 797–809, 2006.

[99] G. A. M. Hurkx, H. De Graaff, W. Kloosterman, and M. Knuvers, "A new analytical diode model including tunneling and avalanche breakdown", *IEEE Transactions on electron devices*, vol. 39, no. 9, pp. 2090–2098, 1992.

[100] J. Zhang, M. A. Itzler, H. Zbinden, and J.-W. Pan, "Advances in ingaas/inp single-photon detector systems for quantum communication", *Light: Science & Applications*, vol. 4, no. 5, e286–e286, 2015.

[101] C. Hu, X. Zheng, J. C. Campbell, B. M. Onat, X. Jiang, and M. A. Itzler, "Characterization of an ingaas/inp-based single-photon avalanche diode with gated-passive quenching with active reset circuit", *Journal of Modern Optics*, vol. 58, no. 3-4, pp. 201–209, 2011.

[102] L. Neri, S. Tudisco, F. Musumeci, *et al.*, "Note: dead time causes and correction method for single photon avalanche diode devices", *Review of Scientific Instruments*, vol. 81, no. 8, 2010.

[103] A. Tosi, F. Acerbi, M. Anti, and F. Zappa, "Ingaas/inp single-photon avalanche diode with reduced afterpulsing and sharp timing response with 30 ps tail", *IEEE Journal of quantum electronics*, vol. 48, no. 9, pp. 1227–1232, 2012.

[104] F. Signorelli, F. Telesca, E. Conca, *et al.*, "Low-noise ingaas/inp single-photon avalanche diodes for fiber-based and free-space applications", *IEEE journal of selected topics in quantum electronics*, vol. 28, no. 2: Optical Detectors, pp. 1–10, 2021.

[105] A. Restelli, J. C. Bienfang, and A. L. Migdall, "Single-photon detection efficiency up to 50% at 1310 nm with an ingaas/inp avalanche diode gated at 1.25 ghz", *Applied Physics Letters*, vol. 102, no. 14, 2013.

[106] S. Cova, A. Lacaita, and G. Ripamonti, "Trapping phenomena in avalanche photodiodes on nanosecond scale", *IEEE Electron device letters*, vol. 12, no. 12, pp. 685–687, 1991.

[107] A. Giudice, M. Ghioni, S. Cova, and F. Zappa, "A process and deep level evaluation tool: afterpulsing in avalanche junctions", in *ESSDERC'03. 33rd Conference on European Solid-State Device Research, 2003.*, IEEE, 2003, pp. 347–350.

[108] M. A. Green and M. J. Keevers, "Optical properties of intrinsic silicon at 300 k", *Progress in Photovoltaics: Research and applications*, vol. 3, no. 3, pp. 189–192, 1995.

[109] J. I. Pankove, *Optical processes in semiconductors*. Courier Corporation, 1975.

[110] C. Veerappan and E. Charbon, "A low dark count pin diode based spad in cmos technology", *IEEE transactions on electron devices*, vol. 63, no. 1, pp. 65–71, 2015.

[111] J. A. del Alamo and R. M. Swanson, "Modelling of minority-carrier transport in heavily doped silicon emitters", *Solid-State Electronics*, vol. 30, no. 11, pp. 1127–1136, 1987.

[112] M. Tyagi and R. Van Overstraeten, "Minority carrier recombination in heavily-doped silicon", *Solid-State Electronics*, vol. 26, no. 6, pp. 577–597, 1983.

[113] W. G. Oldham, R. R. Samuelson, and P. Antognetti, "Triggering phenomena in avalanche diodes", *IEEE Transactions on electron devices*, vol. 19, no. 9, pp. 1056–1060, 1972.

[114] M.-J. Lee, A. R. Ximenes, P. Padmanabhan, *et al.*, "High-performance back-illuminated three-dimensional stacked single-photon avalanche diode implemented in 45-nm cmos technology", *IEEE Journal of selected topics in quantum electronics*, vol. 24, no. 6, pp. 1–9, 2018.

[115] K. Morimoto *et al.*, "Charge-focusing spad image sensors for low light imaging applications", in *Int. SPAD Sensor workshop*, 2020.

[116] Y. Ota, K. Morimoto, T. Sasago, *et al.*, "A 0.37 w 143db-dynamic-range 1mpixel backside-illuminated charge-focusing spad image sensor with pixel-wise exposure control and adaptive clocked recharging", in *2022 IEEE International Solid-State Circuits Conference (ISSCC)*, IEEE, vol. 65, 2022, pp. 94–96.

[117] K. Zang, X. Jiang, Y. Huo, *et al.*, "Silicon single-photon avalanche diodes with nano-structured light trapping", *Nature Communications*, vol. 8, no. 1, p. 628, 2017.

[118] Y. A. Goldberg and N. Schmidt, "Handbook series on semiconductor parameters", *vol*, vol. 2, pp. 1–36, 1999.

[119] F. Acerbi, M. Anti, A. Tosi, and F. Zappa, "Design criteria for ingaas/inp single-photon avalanche diode", *IEEE Photonics Journal*, vol. 5, no. 2, pp. 6 800 209–6 800 209, 2013.

[120] A. Tosi, F. Acerbi, A. Dalla Mora, M. Itzler, and X. Jiang, "Active area uniformity of ingaas/inp single-photon avalanche diodes", *IEEE Photonics Journal*, vol. 3, no. 1, pp. 31–41, 2010.

[121] A. Tosi, N. Calandri, M. Sanzaro, and F. Acerbi, "Low-noise, low-jitter, high detection efficiency ingaas/inp single-photon avalanche diode", *IEEE Journal of selected topics in quantum electronics*, vol. 20, no. 6, pp. 192–197, 2014.

[122] R. Thew, D. Stucki, J.-D. Gautier, H. Zbinden, and A. Rochas, "Free-running ingaas/ inp avalanche photodiode with active quenching for single photon counting at telecom wavelengths", *Applied Physics Letters*, vol. 91, no. 20, 2007.

[123] A. Rochas, C. Guillaume-Gentil, J.-D. Gautier, *et al.*, "Asic for high-speed-gating and free running operation of spads", in *Photon Counting Applications, Quantum Optics, and Quantum Cryptography*, SPIE, vol. 6583, 2007, pp. 143–152.

[124] C. Yu, M. Shangguan, H. Xia, J. Zhang, X. Dou, and J.-W. Pan, "Fully integrated free-running ingaas/inp single-photon detector for accurate lidar applications", *Optics express*, vol. 25, no. 13, pp. 14 611–14 620, 2017.

[125] B. Korzh, N. Walenta, T. Lunghi, N. Gisin, and H. Zbinden, "Free-running ingaas single photon detector with 1 dark count per second at 10% efficiency", *Applied Physics Letters*, vol. 104, no. 8, 2014.

[126] T. Lunghi, C. Barreiro, O. Guinnard, *et al.*, "Free-running single-photon detection based on a negative feedback ingaas apd", *Journal of Modern Optics*, vol. 59, no. 17, pp. 1481–1488, 2012.

[127] Y. Liu, S. R. Forrest, J. Hladky, M. Lange, G. H. Olsen, and D. Ackley, "A planar inp/ingaas avalanche photodiode with floating guard ring and double diffused junction", *Journal of Lightwave Technology*, vol. 10, no. 2, pp. 182–193, 1992.

[128] E. Kizilkan, U. Karaca, V. Pešić, *et al.*, "Guard-ring-free ingaas/inp single-photon avalanche diode based on a novel one-step zn-diffusion technique", *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 28, no. 5: Lidars and Photonic Radars, pp. 1–9, 2022.

[129] G. Van Gurp, T. Van Dongen, G. Fontijn, J. Jacobs, and D. Tjaden, "Interstitial and substitutional zn in inp and ingaasp", *Journal of applied physics*, vol. 65, no. 2, pp. 553–560, 1989.

[130] F. Acerbi, A. Tosi, and F. Zappa, "Growths and diffusions for ingaas/inp single-photon avalanche diodes", *Sensors and Actuators A: Physical*, vol. 201, pp. 207–213, 2013.

[131] O. Pitts, M. Hisko, W. Benyon, S. Raymond, and A. SpringThorpe, "Optimization of mocvd-diffused p-inp for planar avalanche photodiodes", *Journal of crystal growth*, vol. 393, pp. 85–88, 2014.

[132] H. Ando, N. Susa, and H. Kanbe, "Low-temperature zn-and cd-diffusion profiles in inp and formation of guard ring in inp avalanche photodiodes", *IEEE Transactions on Electron Devices*, vol. 29, no. 9, pp. 1408–1413, 1982.

[133] I. Yun and K.-S. Hyun, "Zinc diffusion process investigation of inp-based test structures for high-speed avalanche photodiode fabrication", *Microelectronics journal*, vol. 31, no. 8, pp. 635–639, 2000.

[134] N. Calandri, M. Sanzaro, A. Tosi, and F. Zappa, "Charge persistence in ingaas/inp single-photon avalanche diodes", *IEEE Journal of Quantum Electronics*, vol. 52, no. 3, pp. 1–7, 2016.

[135] O. Pitts, M. Hisko, W. Benyon, G. Bonneville, C. Storey, and A. SpringThorpe, "Planar avalanche photodiodes with edge breakdown suppression using a novel selective area growth based process", *Journal of Crystal Growth*, vol. 470, pp. 149–153, 2017.

[136] S. D. U. Guide and D. U. Guide, "Version n-2017.09", *Mountain View, CA, USA*, 2017.

[137] J. Fossum and D. Lee, "A physical model for the dependence of carrier lifetime on doping density in nondegenerate silicon", *Solid-State Electronics*, vol. 25, no. 8, pp. 741–747, 1982.

[138] A. Schenk, "A model for the field and temperature dependence of shockley-read-hall lifetimes in silicon", *Solid-State Electronics*, vol. 35, no. 11, pp. 1585–1596, 1992.

[139] R. J. McIntyre, "On the avalanche initiation probability of avalanche diodes above the breakdown voltage", *IEEE Transactions on Electron Devices*, vol. 20, no. 7, pp. 637–641, 1973.

[140] R. H. Dennard, F. H. Gaensslen, L. Kuhn, and H. Yu, "Design of micron mos switching devices", *IEEE Solid-State Circuits Society Newsletter*, vol. 12, no. 1, pp. 35–35, 2007.

[141] M. Bohr, "A 30 year retrospective on dennard's mosfet scaling paper", *IEEE Solid-State Circuits Society Newsletter*, vol. 12, no. 1, pp. 11–13, 2007.

[142] B. G. Streetman, S. Banerjee, *et al.*, *Solid state electronic devices*. Prentice hall New Jersey, 2000, vol. 4.

[143] B. Steindl, R. Enne, S. Schidl, and H. Zimmermann, "Linear mode avalanche photodiode with high responsivity integrated in high-voltage cmos", *IEEE Electron Device Letters*, vol. 35, no. 9, pp. 897–899, 2014.

[144] B. Steindl, R. Enne, and H. Zimmermann, "Thick detection zone single-photon avalanche diode fabricated in 0.35 μ m complementary metal-oxide semiconductors", *Optical Engineering*, vol. 54, no. 5, pp. 050 503–050 503, 2015.

[145] E. A. Webster, J. A. Richardson, L. A. Grant, D. Renshaw, and R. K. Henderson, "A single-photon avalanche diode in 90-nm cmos imaging technology with 44% photon detection efficiency at 690 nm", *IEEE Electron Device Letters*, vol. 33, no. 5, pp. 694–696, 2012.

[146] M. Ghioni, A. Gulinatti, I. Rech, F. Zappa, and S. Cova, "Progress in silicon single-photon avalanche diodes", *IEEE Journal of selected topics in quantum electronics*, vol. 13, no. 4, pp. 852–862, 2007.

[147] C. Veerappan and E. Charbon, "A substrate isolated cmos spad enabling wide spectral response and low electrical crosstalk", *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 20, no. 6, pp. 299–305, 2014.

[148] I. Takai, H. Matsubara, M. Soga, M. Ohta, M. Ogawa, and T. Yamashita, "Single-photon avalanche diode with enhanced nir-sensitivity for automotive lidar systems", *Sensors*, vol. 16, no. 4, p. 459, 2016.

[149] I. Vornicu, J. M. Lopez-Martinez, F. N. Bandi, R. C. Galan, and A. Rodriguez-Vazquez, "Design of high-efficiency spads for lidar applications in 110nm cis technology", *IEEE Sensors Journal*, vol. 21, no. 4, pp. 4776–4785, 2020.

[150] A. Gulinatti, I. Rech, F. Panzeri, *et al.*, "New silicon spad technology for enhanced red-sensitivity, high-resolution timing and system integration", *Journal of Modern Optics*, vol. 59, no. 17, pp. 1489–1499, 2012.

[151] F. Gramuglia, P. Keshavarzian, E. Kizilkan, *et al.*, "Engineering breakdown probability profile for pdp and dcr optimization in a spad fabricated in a standard 55 nm bcd process", *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 28, no. 2: Optical Detectors, pp. 1–10, 2021.

[152] A. Gulinatti, F. Ceccarelli, M. Ghioni, and I. Rech, "Custom silicon technology for spad-arrays with red-enhanced sensitivity and low timing jitter", *Optics Express*, vol. 29, no. 3, pp. 4559–4581, 2021.

[153] S. Mandai, M. W. Fishburn, Y. Maruyama, and E. Charbon, "A wide spectral range single-photon avalanche diode fabricated in an advanced 180 nm cmos technology", *Optics express*, vol. 20, no. 6, pp. 5849–5857, 2012.

[154] C. Veerappan and E. Charbon, "Cmos spad based on photo-carrier diffusion achieving pdp> 40% from 440 to 580 nm at 4 v excess bias", *IEEE Photonics Technology Letters*, vol. 27, no. 23, pp. 2445–2448, 2015.

[155] A. Gulinatti, I. Rech, M. Assanelli, M. Ghioni, and S. Cova, "A physically based model for evaluating the photon detection efficiency and the temporal response of spad detectors", *Journal of Modern Optics*, vol. 58, no. 3-4, pp. 210–224, 2011.

[156] F. Ceccarelli, G. Acconcia, A. Gulinatti, M. Ghioni, I. Rech, and R. Osellame, "Recent advances and future perspectives of single-photon avalanche diodes for quantum photonics applications", *Advanced Quantum Technologies*, vol. 4, no. 2, p. 2 000 102, 2021.

[157] R. Ahrenkiel, R. Ellingson, S. Johnston, and M. Wanlass, "Recombination lifetime of in 0.53 ga 0.47 as as a function of doping density", *Applied Physics Letters*, vol. 72, no. 26, pp. 3470–3472, 1998.

[158] A. Liu and Y. Rosenwaks, "Excess carriers lifetime in inp single crystals: radiative versus nonradiative recombination", *Journal of applied physics*, vol. 86, no. 1, pp. 430–437, 1999.

[159] S. Verghese, J. P. Donnelly, E. K. Duerr, *et al.*, "Arrays of inp-based avalanche photodiodes for photon counting", *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 13, no. 4, pp. 870–886, 2007.

[160] E. Kizilkan, U. Karaca, V. Pešić, *et al.*, "Extended temperature modeling of ingaas/inp spads", in *ESSDERC 2023-IEEE 53rd European Solid-State Device Research Conference (ESSDERC)*, IEEE, 2023, pp. 140–143.

[161] O. Salehzadeh, G. Bonneville, O. Pitts, and A. Springthorpe, "Edge breakdown suppression of avalanche photodiodes using zn diffusion and selective area growth", *IEEE Photonics Technology Letters*, vol. 31, no. 10, pp. 767–770, 2019.

[162] U. Karaca, "High performance near/short wavelength infrared megapixel ingaas focal plane array fabrication development and new design proposals", M.S. thesis, Middle East Technical University, 2018.

[163] X. Jiang, M. A. Itzler, R. Ben-Michael, and K. Slomkowski, "Ingaasp–inp avalanche photodiodes for single photon detection", *IEEE Journal of selected topics in quantum electronics*, vol. 13, no. 4, pp. 895–905, 2007.

[164] J. Liu, T. Zhang, Y. Li, *et al.*, "Design and characterization of free-running ingaasp single-photon detector with active-quenching technique", *Journal of Applied Physics*, vol. 122, no. 1, 2017.

[165] Y. Ma, Y. Gu, X. Li, *et al.*, "Ingaasp/inp geiger-mode avalanche photodiode towards sub-khz dark count rate at 20% photon detection efficiency", *Journal of Lightwave Technology*, vol. 40, no. 22, pp. 7364–7374, 2022.

[166] M. Zhou, W. Wang, H. Qu, *et al.*, "Ingaasp/inp single photon avalanche diodes with ultra-high photon detection efficiency", *Optical and Quantum Electronics*, vol. 52, pp. 1–9, 2020.

[167] P. C. Eng, S. Song, and B. Ping, "State-of-the-art photodetectors for optoelectronic integration at telecommunication wavelength", *Nanophotonics*, vol. 4, no. 3, pp. 277–302, 2015.

[168] X. Jiang, S. Wilton, I. Kudryashov, *et al.*, "Ingaasp/inp geiger-mode apd-based lidar", in *Optical Sensing, Imaging, and Photon Counting: From X-Rays to THz*, SPIE, vol. 10729, 2018, pp. 33–44.

[169] T. Baba, Y. Suzuki, K. Makino, *et al.*, "Development of an ingaas spad 2d array for flash lidar", in *Quantum Sensing and Nano Electronics and Photonics XV*, SPIE, vol. 10540, 2018, pp. 81–94.

[170] C. Park, S.-B. Cho, C.-Y. Park, S. Baek, and S.-K. Han, "Dual anode single-photon avalanche diode for high-speed and low-noise geiger-mode operation", *Optics express*, vol. 27, no. 13, pp. 18 201–18 209, 2019.

[171] Y. Tamura, Y. Suzuki, T. Fujita, *et al.*, "Development of ingaas mppc for nir photon counting applications", in *Optical Components and Materials XV*, SPIE, vol. 10528, 2018, pp. 172–183.

[172] P. Rustige, J. Krause, L. Eckoldt, P. Runge, and M. Schell, "Guard-ring free ingaas/inp single photon avalanche diodes for c-band quantum communication", in *2023 Optical Fiber Communications Conference and Exhibition (OFC)*, IEEE, 2023, pp. 1–3.

[173] A. Rogalski, "Infrared detectors: an overview", *Infrared physics & technology*, vol. 43, no. 3-5, pp. 187–210, 2002.

[174] W. W. So, S. Choe, R. Chuang, and C. C. Lee, "An effective diffusion barrier metallization process on copper", *Thin Solid Films*, vol. 376, no. 1-2, pp. 164–169, 2000.

CIRRICULUM VITAE

utku.karaca@epfl.ch
Rue du Simplon 26, 1020 Renens VD, Switzerland

EDUCATION

École Polytechnique of Fédérale de Lausanne — Lausanne, Switzerland
Doctor of Philosophy in Microsystems and Microelectronics — Present
- Focus: Si CMOS and InGaAs/InP Single-Photon Avalanche Diodes Development for NIR/SWIR Applications

Middle East Technical University — Ankara, Turkey
Master's Degree in Electrical and Electronics Engineering — July 2018
- Thesis: High Performance NIR/SWIR Megapixel InGaAs Focal Plane Array Fabrication Development and New Design Proposals

Middle East Technical University — Ankara, Turkey
Bachelor's Degree in Electrical and Electronics Engineering — June 2016

WORK EXPERIENCE

École Polytechnique of Fédérale de Lausanne — Neuchâtel, Switzerland
Advanced Quantum Architecture Laboratory (AQUA) — August 2019 - Present
Doctoral Assistant (Supervisor: Prof. Edoardo Charbon)

École Polytechnique of Fédérale de Lausanne — Lausanne, Switzerland
Laboratory of Nanoscale Electronics (LANES) — August 2018 - August 2019
Doctoral Assistant (Supervisor: Prof. Andras Kis)

Middle East Technical University — Ankara, Turkey
Quantum Devices and Nanophotonics Research Laboratory — January 2016 – July 2018
Research Engineer (Supervisor: Assoc. Prof. Serdar Kocaman**)**

PEER-REVIEWED JOURNAL ARTICLES

- Lee, M-J.*, **Karaca, U.***, et al., "A 73% peak PDP single-photon avalanche diode implemented in 110 nm CIS technology with doping compensation". **Journal of Selected Topics in Quantum Electronics (JSTQE)**, (2023)
- Kizilkan, E.*, **Karaca, U.***, et al., "Guard-ring-free InGaAs/InP single-photon avalanche diode based on a novel one-step Zn-diffusion technique". **JSTQE**, (2022)
- Uzgur, F., **Karaca, U.**, Kızılkan, E., & Kocaman, S., "All InGaAs unipolar barrier infrared detectors". **Transactions on Electron Devices**, (2018)
- Govdeli, A., Sarihan, M. C., **Karaca, U.**, & Kocaman, S., "Integrated optical modulator based on transition between photonic bands". **Scientific Reports**, (2018)
  *First authors with equal contributions

CONFERENCE PROCEEDINGS

- **Karaca, U.**, Kizilkan, E., Bruschini, C., & Charbon, E., "A NIR Enhanced SPAD Fabricated in 110 nm CIS Technology with 78% PDP at 500 nm**". International Image Sensor Workshop (IISW)**, (2023)
- **Karaca, U.**, Govdeli, A., & Kocaman, S., "Hybrid integration of broadband silicon modulators and InGaAs photodetectors". **IEEE Photonics Conference (IPC)**, (2018)
- Govdeli, A., Sarihan, M. C., **Karaca, U.**, & Kocaman, S., "Band-to-Band transition based on-chip optical modulator. **IEEE Photonics Conference (IPC)**, (2018)
- Uzgur, F., **Karaca, U.**, Kızılkan, E., & Kocaman, S., "Al/Sb free InGaAs unipolar barrier infrared detectors". **SPIE Infrared Technology and Applications XLIII**, (2017)

UNDER REVISION ARTICLES & PROCEEDINGS

- **Karaca, U.***, Kizilkan, E.*, Bruschini, C., & Charbon, E., "A new double multiplication region method to design high sensitivity and wide spectrum SPADs in standard CMOS technologies". **Scientific Reports**, (2023)
- **Karaca, U.**, Kizilkan, E., Bruschini, C., & Charbon, E., "10 μm-diameter InGaAsP/InP SPADs for 1060 nm detection with low noise, uniform photoresponse, and low timing jitter". **International SPAD Sensor Workshop (ISSW)**, (2023)

CERTIFICATES

Microelectronics Support Centre, UKRI-STFC Rutherford Appleton Laboratory, Attendant 2021
Topic: Introduction to Technology CAD (TCAD)

PRINCE2® Foundation, Attendant 2021
Topic: Project Management