Near Infra-red single-photon detection using Ge-on-Si heterostructures

Giuseppe Intermite

Submitted for the Degree of Doctor of Philosophy

Heriot-Watt University

School of Engineering and Physical Science

July 2015

The copyright in this thesis is owned by the author. Any quotation from the thesis or use of any of the information contained in it must acknowledge this thesis as the source of the quotation or information.

Abstract

This Thesis investigates the design of Ge-on-Si single-photon avalanche diode (SPAD) detectors combining the many advantages of low-noise Si single-photon avalanche multiplication with the infrared sensing capability of germanium. The devices were simulated by using electric field modelling software to predict key aspects of the device behaviour in terms of the current-voltage characteristic and electric field.

The devices were then characterised in terms of their single-photon performance. A 25 μm diameter device showed a single-photon detection efficiency of ~ 4 % at a wavelength of 1310 nm and a temperature of 100 K when biased at 10 % above the breakdown voltage. In the same condition, a dark count rate of ~ 6 Mcs^{-1} was measured. This resulted in the lowest noise equivalent power of ~ 1×10^{-14} $WHz^{-1/2}$ of Ge-on-Si SPADs reported in the scientific literature. At the longer wavelength of 1550 nm, the single-photon detection efficiency was reduced to ~ 0.1 % at 125 K and 6 % of relative excess bias. Although further investigation needs to be carried out, a potential major advantage of these devices compared to the InGaAs/InP SPADs could be that of reduced afterpulsing since a small increase (a factor of 2) in the normalised dark count rate was measured when the repetition rate was increased from 1 kHz to 1 MHz.

Finally, the fill-factor enhancement of 32 × 32 Si CMOS SPAD arrays resulting from the integration of high efficiency diffractive optical microlens arrays was investigated. A full characterisation of SPAD arrays integrating microlens arrays in terms of improvement factor and spatial uniformity of detection is presented for the first time in the scientific literature in a large spectral range (500-900 nm) and different f-numbers (from f/2 to f/22) by using a double telecentric imaging system. The highest improvement factor of ~16 was measured for a SPAD array integrating microlens arrays, combined with a very high spatial efficiency uniformity of between 2–6%.

Acknowledgements

One of the things that fascinating me about nature is, in general, its property of resilience. Because of this intrinsically positive property, the world around us evolves, changes, adapts and stands up. When I think to a single word that could fully describe my PhD, I then believe that “resilience” is the right one. It was the need of change that push me to start this wonderful but difficult, challenging but unpredictable experience. This thesis represents a scanning of my life in the recent years, and every chapter encodes a photograph of it which brought me to the beginning of a new growth. This was made possible thanks to the many people that I met during my way, and to which I will be forever grateful.

Firstly, my biggest thanks go to my supervisor Prof. Gerald S. Buller who believed in me from our first meeting and represented a reference point during the whole PhD. Thanks to him I had the opportunity to start this wonderful experience. Thanks for your support, guidance, persistence and for putting me on the right path when problems were behind the door.

A special thank goes to Dr. Ryan E. Warburton that guides me through the lab with a lot of patience. Thanks for your assistance, support and friendship. Thanks to Dr Peter Vines for his help, suggestion and patience. A big thanks goes to Dr. Aongus McCarthy, his passion for physics and his fairness let me understand which kind of scientist I would like to become in the future. A special thank goes to the rest of the Photon Counting group members present and past for many helpful discussion inside and outside the lab: Dr. Robert J. Collins, Aurora Maccarone, Nathan Gemmell, Ross J. Donaldson, Silvia Butera, Ximing Ren, Agata Pawlikowska, Kateryna Kuzmenko, Patrick J Clarke, Frauke Izdebsky and Dr. Nils Krichel.

Thanks also go out to the various collaborators with whom I have worked: Prof. Douglas J. Paul, Dr. Kevin Gallagher, and Ross J. Millar from the University of Glasgow, Dr Maksym Myronov and Phil Allred from the University of Warwick, Prof. Robert W. Kelsall from University of Leeds, Dr. Andrew Knights from McMaster University.

I would like to express my thanks to our collaborators at Politecnico di Milano, Prof. Franco Zappa and Dr. Federica Villa, for the Si CMOS SPAD arrays which were used

for the work within Chapter 6 and for guidance on the operation of SPAD sensors. I would also like to express my thanks to Dr. Andrew Waddie for the microlens arrays design and simulations, then to my colleagues Dr. Aongus McCarthy and Ximing Ren for the flip-chip bonding and post-bonding inspection of the microlens arrays

I want to thanks all my friends, but in particular Ehsan and Harry for all the unforgettable moments spent together.

Last but not least, a heartfelt thanks goes to my wife, Francesca. She has been the light, the strength, the passion, the kindness that pushes me towards my objectives. Wherever we will go, whatever will happen, I know for sure that we will be always together.

Finally, I want to thanks my family. Although, we did not meet often in the last few years, their support, encouragement, love has also been deep in my heart and memory.

ACADEMIC REGISTRY
Research Thesis Submission

Name:	Giuseppe Intermite		
School/PGI:	Engineering and Physical Sciences		
Version: *(i.e. First, Resubmission, Final)*	Final	Degree Sought (Award **and** Subject area)	PhD, Physics

Declaration

In accordance with the appropriate regulations I hereby submit my thesis and I declare that:

1) the thesis embodies the results of my own work and has been composed by myself
2) where appropriate, I have made acknowledgement of the work of others and have made reference to work carried out in collaboration with other persons
3) the thesis is the correct version of the thesis for submission and is the same version as any electronic versions submitted*.
4) my thesis for the award referred to, deposited in the Heriot-Watt University Library, should be made available for loan or photocopying and be available via the Institutional Repository, subject to such conditions as the Librarian may require
5) I understand that as a student of the University I am required to abide by the Regulations of the University and to conform to its discipline.

* *Please note that it is the responsibility of the candidate to ensure that the correct version of the thesis is submitted.*

Signature of Candidate:		Date:	

Submission

Submitted By *(name in capitals):*	Giuseppe Intermite
Signature of Individual Submitting:	
Date Submitted:	

For Completion in the Student Service Centre (SSC)

Received in the SSC by *(name in capitals):*			
Method of Submission *(Handed in to SSC; posted through internal/external mail):*			
E-thesis Submitted (***mandatory for final theses)***			
Signature:		Date:	

Table of Contents

Chapter 1 – Introduction 1
1.1 Reference 6
Chapter 2 – Physics of photodiodes 8
2.1 Introduction 8
2.2 Physics of p-n and p-i-n photodiodes 8
2.3 Avalanche photodiodes 16
2.3.1 Advantage of an APD over a p-i-n photodiode 21
2.4 Single-photon detectors 24
2.4.1 Photomultiplier tube (PMT) and Microchannel plates 24
2.4.2 The single-photon avalanche diode (SPAD) 26
2.4.2.1 Figures of merit of a SPAD 28
2.4.2.2 Quenching techniques for SPADs 31
2.5 Conclusion 33
2.6 Reference 34
Chapter 3 - Ge-on-Si Technology and Applications: Material selection, Epitaxial Growth and Review of Detectors 37
3.1 Introduction 37
3.2 Choice of material for near-infrared photodetectors on Si 38
3.3 Lattice mismatched near-infrared photodetectors on Si 42
3.3.1 $Si_{1-x}Ge_x$ strained layer 42
3.3.2 SiGeC structures 49
3.3.3 InGaAs structures 51
3.3.4 Pure Ge 55
3.4 Growth techniques for pure Ge-on-Si 58
3.4.1 Ge-on-Si heteroepitaxy 59
3.4.1.1 Two-temperature LT/HT Ge growth and cyclic thermal annealing 61
3.4.1.2 Selective epitaxial growth of Ge-on-Si 64
3.4.2 Effect of TDD on device performance 66
3.4.3 Effect of the Ge strain on the absorption at the NIR wavelengths 70
3.5 Ge-on-Si by wafer bonding 72
3.6 Ge-on-Si NIR photodetectors 73
3.6.1 Ge-on-Si p-i-n photodetectors 74
3.6.2 Waveguide integrated Ge-on-Si photodetectors 81
3.6.3 Ge-on-Si avalanche photodetectors 86
3.6.4 SPADs operating at infrared wavelengths 92
3.6.5 Superconducting Nanowire Single-Photon Detector (SNSPD) 102
3.6.6 Transition Edge Sensor 105
3.7 Conclusion 106
3.8 Reference 109

Chapter 4 - Ge-on-Si SACM structure: Design, modelling and SIMS measurements 121
4.1 Introduction 121
4.2 Design of a Separate Absorption, Charge and Multiplication Ge-on-Si structure 121
4.3 Ge-on-Si SPAD structure 125
4.4 Modelling of the Ge-on-Si hetero-structure 127
4.5 Effect of temperature on the breakdown voltage 133
4.6 Capacitance-Voltage Measurements on the 1st generation of Ge-on-Si SPAD structures 135
4.7 Impact of the Ge absorber layer doping on the SPAD structure 140
4.8 SIMS measurements on the 2nd generation of SPAD structure 141
4.9 Impact of the Si charge sheet and multiplication layer doping profile on the SPAD structure 146
4.10 3rd Generation of Ge-on-Si SPAD structure: SIMS measurements and modelling 151
4.11 Future development 161
4.12 Conclusion 164
4.13 Reference 165
Chapter 5 - Ge-on-Si Single-Photon Avalanche Diodes: Device fabrication, Characterisation and Performance 167
5.1 Introduction 167
5.2 Ge-on-Si SPAD growth 168
5.3 Ge-on-Si SPAD fabrication 169
5.4 The Current-Voltage Characteristic 172
5.4.1 I-V characteristic of sample 11-167 173
5.4.2 I-V characteristic of sample 11-141 177
5.4.3 Comparison between devices from samples 11-167 and 11-141 181
5.5 Single-Photon Characterisation 183
5.5.1 TCSPC technique 184
5.5.2 Experimental Setup and Methodology 188
5.5.3 Dark Count Rate 190
5.5.4 Single-Photon Detection efficiency 193
5.5.5 Noise Equivalent Power 197
5.5.6 Timing jitter 199
5.5.7 DCR versus frequency (afterpulsing) 203
5.5.8 Ge-on-Si SPADs in the scientific literature 205
5.6 Dark current analysis 206
5.7 Future work 212
5.8 Conclusion 214
5.9 Reference 215

Chapter 6 - Fill-factor recovery of Si CMOS SPAD array by using microlens integration: improvement factor and spatial uniformity characterisation 217
6.1 Introduction 217
6.2 Si SPAD: custom processing and CMOS fabrication 218
6.3 Single-photon Detection Efficiency and Fill-Factor recovery 220
6.4 HV-CMOS Si SPAD 223
6.4.1 Photon-counting mode of operation 226
6.5 Diffractive Microlens Array Design 228
6.5.1 Microlens performance 234
6.5.2 Microlens integration 235
6.6 Characterisation setup 237
6.7 Improvement Factor (IF) measurements 242
6.7.1 Methodology 242
6.7.2 Hot pixel removal 242
6.7.3 Results of the Improvement Factor 245
6.7.4 Improvement Factor comparison 249
6.8 Spatial Uniformity of the SPAD arrays 255
6.9 Future Outlook 260
6.10 Conclusions 262
6.11 Acknowledgements 263
Reference 263
Chapter 7 - Conclusions and Future work 266
7.1 Reference 270

List of Publications by the Candidate

Papers published in peer reviewed Journals

R. E. Warburton, **G. Intermite**, M. Myronov, P. Allred, D. R. Leadley, K. Gallacher, D. J. Paul, N. J. Pilgrim, L. J. M. Lever, Z. Ikonic, R. W. Kelsall, E. Haunte-Cerón, A. P. Knights, and G. S. Buller, *Ge-on-Si Single-Photon Avalanche Diode Detectors: Design, Modeling, Fabrication, and Characterization at Wavelengths of 1310 and 1550nm*, IEEE Transactions on Electron Devices, Vol. 60, No. 11, November 2013.

G. Intermite, R. E. Warburton, A. McCarthy, X. Ren, F. Villa, A. J. Waddie, M. R. Taghizadeh, F. Zappa, A. Tosi, and G.S. Buller, *Integration of diffractive microlens arrays for fill-factor improvement of Si CMOS single-photon avalanche diode detector arrays*, To be submitted to Optics Express in July-August 2015.

Peer reviewed Conference paper

G. Intermite, R. E. Warburton, A. McCarthy, X. Ren, F, Villa, A. J. Waddie, M.R. Taghizadeh, A. Tosi, F. Zappa and G. S. Buller, *Enhancing the fill-factor of CMOS SPAD arrays using microlens integration*, INVITED PAPER to SPIE Optics and Optoelectronics, Prague, Czech Republic, April 13-16, 2015.

G. Intermite, R.E. Warburton, M. Myronov, P. Allred, D.R. Leadley, K. Gallagher, D.J. Paul, N.J. Pilgrim, L.J.M. Lever, Z. Ikonic, R.W. Kelsall, and G.S. Buller, *Design and performance of a prototype mesa–geometry Ge–on–Si single–photon avalanche diode detector at 1310 nm and 1550 nm wavelengths*, The IEEE 10th International Conference on Group IV Photonics (GFP), Seoul, Korea, August 28-30, 2013.

G. Intermite, R.E. Warburton, M. Myronov, P. Allred, D.R. Leadley, K. Gallacher, D.J. Paul, N.J. Pilgrim, L.J.M. Lever, Z. Ikonic, R.W. Kelsall, and G.S. Buller, *Ge-on-Si single-photon avalanche diode near-IR detectors*, UK Semiconductor 2013, Sheffield, UK, July 3 - 4,2013.

Chapter 1 – Introduction

Exactly 50 years ago, Gordon Moore predicted that the number of transistors in an integrated circuit (IC) doubles roughly every eighteen months [1]. This prediction is better known as Moore's law. The main reason behind the success of the Moore's law is the technology scaling which has been favourably exploited from many silicon IC technologies, such as silicon complementary metal oxide semiconductor (Si-CMOS) technology. Therefore, technology scaling permits a high density of integration, tens of millions of CMOS transistor on chip, and today it is possible to speak of Deep Sub Micron (DSM) and Ultra-DSM technologies. However, as the feature size decreases below 100 nm, the gain in the performance obtained from shrinking the size of the devices is reduced. This is mainly due to two reasons: the RC delay and heat dissipation.

As predicted from the ITRS (International Technology Roadmap for Semiconductor), the metallic interconnection represents the main problem since the RC delay becomes very important, as shown in Figure 1 [2]. The RC time constant of a wire defines the maximum frequency (which is proportional to 1/RC) of the transmitted signal through a wire of a defined length. This depends on the resistance and capacitance per unit length. Both parameters, and hence the RC delay, increase with the shrinkage of the feature size. The RC delay does not only limit the overall frequency of the chip, but it also induces the de-synchronisation of the signals in the clock distribution, and increases the bit error rate [3].

The power dissipation in an information-processing system is a major limitation at many levels, including those related to CMOS chips. The ITRS stated that the amount of heat that can be removed from a chip in a cost-effective manner is about to saturate at approximately 200 W. Power dissipation is a problem that directly limits the performance of chips, and its increase is thus a significant factor in system economics and the environmental impact of information technology [4].

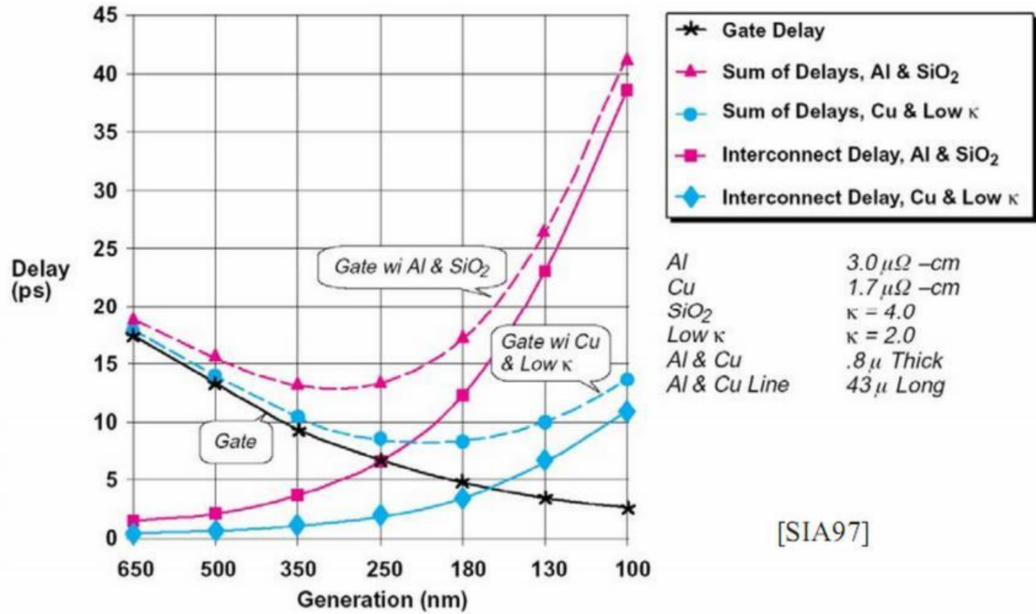


Figure 1. The gate delay and interconnect delay as a function of feature size for Al/SiO_2 and Cu/low k dielectric CMOS technology.

In this context, different solutions have been proposed to overcome these limitations. Silicon as emerged as an efficient material system for building photonics devices as well as electronic devices. The early applications of silicon photonics are in digital data communications. By substituting the metallic wires with optical waveguides, not only can the bandwidth of data transmission be increased, but the heat dissipation can be also greatly reduced, since, in principle, photons do not generate any heat when they propagate in a waveguide. The main advantage of silicon photonics that has emerged over the past 50 years is the possibility to develop processes that permit the re-use of CMOS fabrication infrastructure in order to build complex photonic circuits in which information is transferred seamlessly from the electronic to the optical domains and back again [5], [6]. The main constraints associated with CMOS refer to the materials that are not compatible with the CMOS process, and both process and circuits must be designed in such a way that processing them will not harm or contaminate the tools. Although different photonics devices have been fabricated and demonstrated separately, these devices can be interfaced to one another with standard optical fibres and connectors. However, high cost and loss emerge from the photonic packaging process which requires a severe alignment procedure with submicron accuracy. In order to achieve high performance and low-cost optical links, it is desirable to integrate optical components such as photodetectors, light sources and modulators on the same silicon chip, and silicon is a very attractive platform for photonic integrated circuits.

In addition, beyond data communications, there are a huge number of new applications that rely on silicon photonics being explored in both commercial and academic worlds such as quantum information and imaging technologies, condensed matter physics, biosensing, etc. Therefore, silicon photonics might offer a robust platform for the monolithic integration of optics and microelectronics on the same silicon chip by using the same fabrication facilities used for CMOS processes.

In particular, quantum technology exploits the quantum-mechanical phenomena to provide new techniques for sensing, measurements, information processing, data transmission and storage. This technology is driven by advances in technology and experimental capability obtained in the last decade or so. This is also confirmed by the considerable investment for the development of such technology. For example, it is worthwhile to mention the National Quantum Technology Programme from the UK Government with the aim of stimulating the development of commercial applications and markets for quantum technologies.

A direct consequence of this, is the demand of extremely sensitive devices capable of registering the elemental quantum of light, referred to as photons [7]. Such class of detectors, referred as single-photon detectors, are used in a wide range of time-correlated single-photon counting (TCSPC) applications such as quantum key distribution (QKD), time-of-flight, light detection and ranging (LiDAR), etc.. Different technologies have been used to detect single-photons such as photomultiplier tubes, microchannel plates, superconducting nanowires, and single-photon avalanche detectors (SPADs).

Among the various choices, SPADs tend to be preferred because of their intrinsic advantages that are typical of solid state devices, such as low power consumption, miniature size, low bias voltages, reduced magnetic field susceptibility, reliability, and low cost [8], [9]. Although Si represents the best material choice for all the aforementioned advantages, it is also the best material for avalanche multiplication due to its large ratio of electron to hole impact ionisation rates. Si SPADs now represent a commercially available technology, and can be classified into two distinct groups: those that are fabricated using customized processing techniques and those fabricated using CMOS compatible approaches.

Standard CMOS fabrication processes from the microelectronic industry offer significant advantages in terms of the routine on-chip integration with the electronics

required for photon-counting and photon-timing techniques. This efficient integration allows the possibility of fabrication of two-dimensional SPAD-based focal plane arrays. However, the detection of light in silicon is intrinsically limited to wavelengths below 1 μm. Therefore, for operation in the low loss fibre optical windows at around 1.3 μm and 1.55 μm other semiconductors must be used.

Near infrared single-photon detection is of interest in many applications. Although QKD can be accomplished at shorter wavelengths compatible with Si SPADs [10], it is preferable to work in the low attenuation windows of standard optical fibres [11]. Applications such laser ranging also benefit from the decreased attenuation through air as well as the low solar background radiation at these wavelengths [12]. Near-infrared is also used in the monitoring of microscopic biological system such as singlet oxygen generation ($^1O_2$) which fluoresces at a wavelength of 1270 nm beyond the detection range of Si SPADs [13]. Therefore, the development of low-noise SPADs in this spectral region is required. The best currently available practical detectors, operating at near-room temperature and a wavelengths in the range of 1 – 1.7 μm, are InGaAs/InP SPADs. These devices have demonstrated good single-photon performance, and they are commercially available from different sources such as Princeton Lightwave, MicroPhoton Device (MPD), etc. However, this class of detectors is not compatible with CMOS technology, they are expensive and one of their major drawbacks is represented by the deleterious effect of afterpulsing which significantly reduces the frequency of operation of the devices to be below 100 kHz.

Germanium has been shown to be a promising material for near-infrared detection on Si. The direct band gap of Ge is 0.8 eV, corresponding to 1550 nm. SiGe technology has already been applied for bipolar transistors, and therefore the material has proved to be fully compatible with Si electronic technology. A big challenge for the integration of Ge on Si is represented by the 4% lattice mismatch between the two materials. However, the development of SiGe buffer layers and two-temperature steps of pure Ge on Si has been used to overcome this problem. In terms of single-photon detectors, the first Ge-containing Si SPAD was demonstrated by Loudon *et al.* [14], who showed an improvement of detection efficiency at 1200 nm (above the Si wavelength cut-off). Pure Ge-on-Si SPADs have been reported on only a few occasions. In this context, the research presented in this thesis is aimed at filling this gap by investigating the design of a Ge-on-Si SPADs in order to combine the several advantages of low-noise Si single-photon avalanche multiplication with the infrared sensing capability of germanium.

In addition, Ge does not suffer from the same contamination processes as InGaAs, and its compatibility with the CMOS technology (mainly to make high-speed transistors) has been already demonstrated. Furthermore, Ge-on-Si CMOS imagers (768×600 pixels) for near-infrared detection has been fabricated and commercialised by NoblePeak Vision which is a company based in USA [17]. A monolithic CMOS imager with Ge detectors (incorporate in the CMOS process using selective epitaxial growth) fabricated within a conventional 180 nm CMOS foundry process was demonstrated [17]. Although each pixel consisted of a lateral p-i-n whole Ge detector, the main advantage arising from the integration of Ge with standard CMOS technologies was successfully proved.

Chapter 2 describes the physics of different photodetector structures and outlines the main parameters that must be considered during their design. The main focus is given to the avalanche photodiodes and single-photon detectors. Although, there are many common points between these two structures, they have been designed to satisfy different performance requirements. In particular, the key figures of merit of a SPAD such as dark count rate, single-photon detection efficiency, noise equivalent power, afterpulsing and jitter are then introduced.

Chapter 3 firstly provides an explanation of the choice of materials for near-infrared detection and then focuses on the integration between these materials and Si. In particular, it discusses the several problems which arise from the lattice mismatch, and then reviews the different solutions used to overcome this problem. Hence, the main attention is given to the Ge-on-Si hetero-epitaxial system. Finally, a literature review on the integration of near-infrared photodetectors on silicon is presented with the main emphasis given to the single-photon detector technologies which are used in the near-infrared wavelength range.

Chapter 4 describes the criteria that were used for the design of the Ge-on-Si SACM structure presented in this work. The devices were modelled using a commercially available software package (Silvaco ATLAS), and hence the results of simulations are presented. Since different Ge-on-Si SPAD generations were grown and analysed, experimental data are given for those obtained from the capacitance-voltage measurements and SIMS measurements used in the modelling software in order to evaluate their impact on the designed structure.

Chapter 5 firstly introduces the growth and fabrication process involved for the device development, and then describes the preliminary devices characterisation in terms of I-V characteristics in order to evaluate the main mechanisms which contributed to the leakage current. The Time Correlated Single-Photon Counting (TCSPC) setup used for the single-photon characterisation of these devices in terms of single-photon detection efficiency, dark count rate, noise equivalent power, time jitter and afterpulsing is then described.

Chapter 6 describes the main advantages introduced by the fabrication of Si SPAD by using the standard CMOS fabrication processes. From the early 2000s, several research groups have explored the design of a monolithically integrated single-photon imaging system in high-voltage (HV) and standard deep-sub-micron (DSM) CMOS technologies [15], [16]. Two-dimensional CMOS SPAD-based focal plane arrays generally suffer from low fill-factor of the detector photo-sensitive area compared to the overall detector pixel area. Diffractive microlens arrays have been used in this work with the aim of recovering this loss of sensitivity, and the microlens design and integration process is so described in the chapter. Additionally, the experimental setup used for the characterisation of the SPAD arrays integrating the microlens array is given, and the main finding in terms of improvement factor and spatial uniformity of detection are finally presented.

Chapter 7 summarises the conclusions pointed out for each chapter and describes future work in these areas.

1.1 Reference

[1] G. E. Moore, "Cramming more components onto integrated circuits," *IEEE Solid-State Circuits Soc. Newsl.*, vol. 38, no. 8, 1965.

[2] D. A. B. Miller and H. M. Ozaktas, "Limit to the Bit-Rate Capacity of Electrical Interconnects from the Aspect Ratio of the System Architecture," *J. Parallel Distrib. Comput.*, vol. 41, no. 1, pp. 42–52, 1997.

[3] R. Sridhar, "System-on-chip (SoC): clocking and synchronization issues," in *17th International Conference on VLSI Design, 2004. Proceedings*, 2004, pp. 520–527.

[4] D. A. B. Miller, "Device Requirements for Optical Interconnects to Silicon Chips," *Proc. IEEE*, vol. 97, no. 7, pp. 1166–1185, 2009.

[5] Y. Arakawa, T. Nakamura, Y. Urino, and T. Fujita, "Silicon photonics for next generation system integration platform," *IEEE Commun. Mag.*, vol. 51, no. 3, pp. 72–77, 2013.

[6] M. Hochberg, N. C. Harris, R. Ding, Y. Zhang, A. Novack, Z. Xuan, and T. Baehr-Jones, "Silicon Photonics: The Next Fabless Semiconductor Industry," *IEEE Solid-State Circuits Mag.*, vol. 5, no. 1, pp. 48–58, 2013.

[7] R. H. Hadfield, “Single-photon detectors for optical quantum information applications,” *Nat. Photonics*, vol. 3, no. 12, pp. 696–705, 2009.
[8] G. S. Buller and R. J. Collins, “Single-photon generation and detection,” *Meas. Sci. Technol.*, vol. 21, p. 012002, 2010.
[9] S. D. Cova and M. Ghioni, “Single-Photon Counting Detectors,” *Breakthroughs*, p. 274, 2010.
[10] K. J. Gordon, V. Fernandez, P. D. Townsend, and G. S. Buller, “A short wavelength GigaHertz clocked fiber-optic quantum key distribution system,” *IEEE J. Quantum Electron.*, vol. 40, no. 7, pp. 900–908, 2004.
[11] G. S. Buller and R. J. Collins, “Single-Photon Detectors for Infrared Wavelengths in the Range 1–1.7 μm,” in *Advanced Photon Counting*, P. Kapusta, M. Wahl, and R. Erdmann, Eds. Springer International Publishing, pp. 43–69, 2014.
[12] R. E. Warburton, “Infrared Time-Correlated Single-Photon Counting,” Ph.D., Heriot-Watt University, 2008.
[13] N. R. Gemmell, A. McCarthy, B. Liu, M. G. Tanner, S. D. Dorenbos, V. Zwiller, M. S. Patterson, G. S. Buller, B. C. Wilson, and R. H. Hadfield, “Singlet oxygen luminescence detection with a fiber-coupled superconducting nanowire single-photon detector,” *Opt. Express*, vol. 21, no. 4, pp. 5005–5013, 2013.
[14] A. Y. Loudon, P. A. Hiskett, G. S. Buller, R. T. Carline, D. C. Herbert, W. Y. Leong, and J. G. Rarity, “Enhancement of the infrared detection efficiency of silicon photon-counting avalanche photodiodes by use of silicon germanium absorbing layers,” *Opt. Lett.*, vol. 27, no. 4, pp. 219–221, 2002.
[15] M. Gersbach, Y. Maruyama, R. Trimananda, M. W. Fishburn, D. Stoppa, J. A. Richardson, R. Walker, R. Henderson, and E. Charbon, “A Time-Resolved, Low-Noise Single-Photon Image Sensor Fabricated in Deep-Submicron CMOS Technology,” *IEEE J. Solid-State Circuits*, vol. 47, no. 6, pp. 1394–1407, Jun. 2012.
[16] F. Villa, R. Lussana, D. Bronzi, S. Tisa, A. Tosi, F. Zappa, A. Dalla Mora, D. Contini, D. Durini, S. Weyers, and W. Brockherde, “CMOS Imager With 1024 SPADs and TDCs for Single-Photon Timing and 3-D Time-of-Flight,” *IEEE J. Sel. Top. Quantum Electron.*, vol. 20, no. 6, pp. 364–373, Nov. 2014.
[17] C.A. King, C.S. Rafferty, US Patent 7,453,129, “Image sensor comprising isolated germanium photodetectors integrated with a silicon substrate and silicon circuitry”

Chapter 2 – Physics of photodiodes

2.1 Introduction

Photodiodes belong to the category of light sensors or detectors, whose task is to convert an optical signal into an electrical signal. As already discussed in Chapter 1, there are a huge number of applications that require that this conversion procedure is performed with certain requirements in terms of sensitivity, bandwidth, power, signal-to noise, etc. To meet these constraints in terms of performance, different photodiode structures have been proposed and demonstrated in the past century. Therefore, understanding the physics of these devices represents a key point for designing new structures that always satisfy more and more stringent requirements. Starting from the physics of the p-n junction, which represents the most basic semiconductor photodiode structure, this chapter will analyse the advantages or disadvantages of more complex structures, such as the p-i-n and the avalanche photodiodes. This analysis requires the introduction of their most important performance characteristics such as the leakage current, responsivity, bandwidth, signal-to-noise, etc. Finally, the single-photon detector will be introduced, which represents an avalanche photodiode operating in Geiger mode, that must satisfy new requirements in terms of dark count rate, single-photon detection efficiency, noise equivalent power, timing jitter and afterpulsing.

2.2 Physics of p-n and p-i-n photodiodes

Photodiodes convert a light signal to an electrical signal such as a voltage or current pulse. This conversion is achieved by the creation of free electron hole pairs (EHPs) by the absorption of photons, that is, the creation of electrons in the conduction band and holes in the valence band. A basic photodiode is represented by the p-n junction, which is created by putting two pieces of semiconductor in contact, a p type (with an acceptor concentration N_a) and an n type (with a donor concentration N_d).

In the equilibrium condition, the strong difference of doping concentration at the interface between the two materials leads to a diffusion of the charge carriers. Holes diffuse from the p-side to the n-side, while electrons diffuse in the opposite direction. In proximity to the interface, holes and electrons on the p and n-side, respectively, leave negative ion acceptors (N_a^-) and positive ion donors (N_d^+). Due to this phenomenon, a space charge region (or depletion region) is created at the junction, negative from the p-side, with an extension equal to W_p and, positive from the n-side (W_n). Therefore, an

electric field (E) is generated in the vicinity of the interface between the two regions, as shown in Figure 2.1.

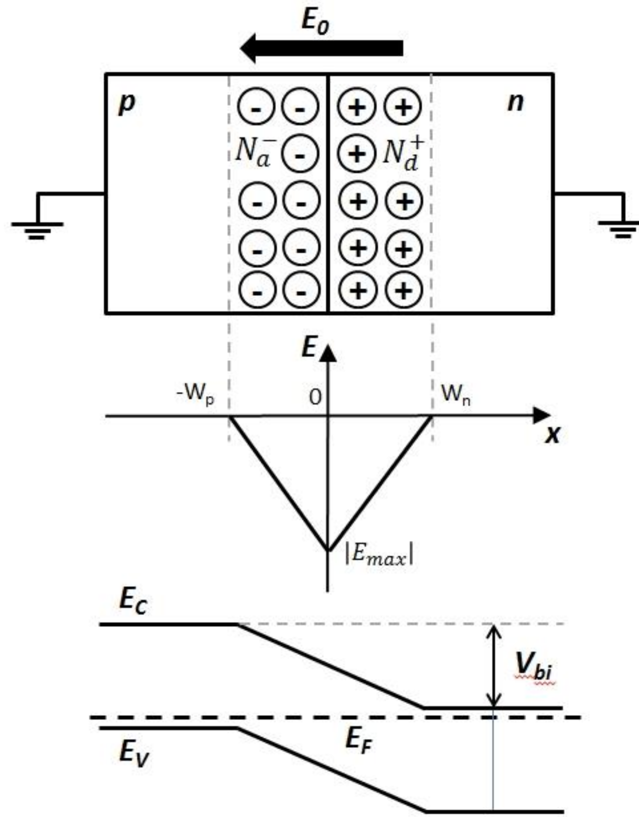


Figure 2.1. Schematic of a p-n junction at equilibrium.

Because of the effect of the electric field, two drift currents are created: holes drift from the right side to the left side (opposite to the hole diffusion component) and electrons drift in the opposite direction (opposite to the electron diffusion component). In equilibrium, the net flux of current through the junction is zero (there is a balance between the current drift component due to the electric field and, the current diffusion component due to different doping concentrations). The Fermi level is constant through the two semiconductor regions (Figure 2.1) and, this determines a built-in potential at the junction, which can be expressed by

$$V_{bi} = \frac{k_B T}{q} \ln\left(\frac{N_a N_d}{n_i^2}\right) \tag{2.1}$$

where k_B is the Boltzmann's constant, T the temperature and, n_i is the intrinsic carrier concentration. In one dimension, the profile of the electric field can be calculated by using Poisson's equation

$$\frac{dE}{dx} = \frac{\rho(x)}{\varepsilon} \tag{2.2}$$

where ρ is the density of space charge equal to qN_d and $-qN_a$ for the n and p side, respectively and, ε is the dielectric permittivity of the semiconductor. The electric field is then obtained integrating equation 2.2 over the space charge region with boundary conditions of $E = 0$ in the neutral regions (outside the space charge region in the p and n side). Considering the condition for which $N_a \gg N_d$ (asymmetric junction), the depletion region will be extended more on the n side than the p side leading to an electric field more confined in the n side.

If an external reverse bias V_r (negative on the p side and positive on the n side) is applied, the negative terminal will cause holes in the p-side to move away from the depletion region, which results in more exposed negative acceptor ions and thus wider depletion region. Similarly, the positive terminal will attract electrons away from the depletion region, which exposes more positive charged ions. The depletion width on the n-side therefore also widens. The electric field dropping across the junction is, therefore, the sum of the electric field associated to the depletion region (E_0) plus the electric field due to the external reverse bias (E), as shown in Figure 2.2. The voltage drops mainly across the resistive depletion region which becomes wider on both sides.

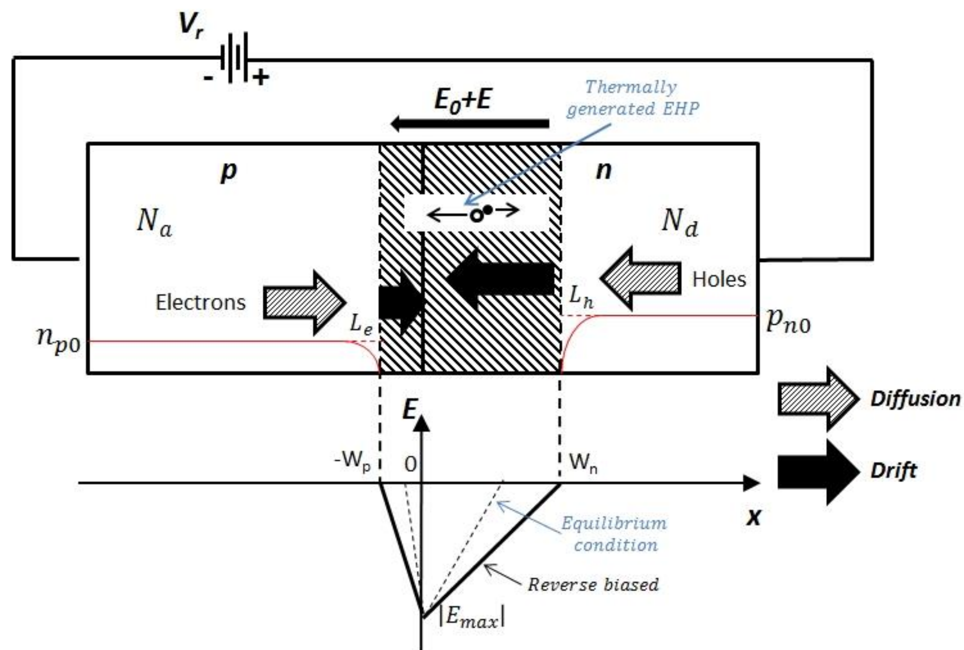


Figure 2.2. Schematic of a p-n junction with the associated electric field when a reverse bias V_r is applied.

There is a small reverse current due to two causes:

1. there is a small hole diffusion current from the bulk n-side, where the holes are at equilibrium with a concentration equal to p_{n0}, to the depletion region. The

reduction of the hole concentration near to the depletion region is related to their diffusion length L_h. Because of this small concentration gradient, a small hole diffusion current exists. Within the depletion region, these carriers are swept by the field across the depletion region over to the p-side. Similarly, for electrons on the p-side. This phenomenon leads to a reverse saturation current density, which can be expressed by

$$J_s = q\left(\frac{D_e}{L_e}n_{p0} + \frac{D_h}{L_h}p_{n0}\right) = q(\frac{D_e}{L_eN_d} + \frac{D_h}{L_hN_a})n_i^2 \quad (2.3)$$

Where D_e and D_h are the diffusion coefficients of electrons and holes, respectively. The value of this current depends only on the material properties via n_i, μ_e, μ_h, doping concentration.

2. The thermal generation of electron hole pairs (EHPs) in the depletion region contributes to this reverse current since the internal field in this layer will separate the electrons and holes and cause them to drift towards neutral regions. The value of this generation current (J_{gen}) within the depletion region can be calculated by using the Shockley-Read-Hall model and by defining the mean thermal generation time, τ_g, as

$$J_{gen} = \frac{qWn_i}{\tau_g} \quad (2.4)$$

It is dependent on the applied reverse bias through the depletion width W, which widens as the reverse bias is increased.

Therefore, the sum of these two contributions represents the total reverse current density, also called the leakage current. Considering a cross-sectional area equal to A, the total reverse current can be written as

$$I_{rev} = q\left(\frac{D_e}{L_eN_d} + \frac{D_h}{L_hN_a}\right)n_i^2A + \frac{qW}{\tau_g}n_iA \quad (2.5)$$

It is important to note that both contributions depend on the area of the device, the intrinsic carrier concentration (n_i), which is an intrinsic property of any semiconductor material and, on the temperature T since $n_i \sim exp(-E_g/2k_BT)$ [1]. This dependence on the temperature is very important and will be analysed in more detail in the following chapters. In real structures, however, there are other sources which increase this current

such as the surface leakage and tunnelling components that will be considered in the following chapters.

Leakage current is also called dark current because it represents the current that flows through a photodiode even without the injection of photons. Leakage current is, therefore, a source of noise in an optical receiver [2], which can degrade the signal to noise (S/N) ratio. Leakage current is also a source of power dissipation. As pointed out previously, in Chapter 1, power dissipation is becoming a stringent factor for CMOS design due to the high level of integration. Controlling power dissipation is a good reason for reducing leakage current. Moreover, the level of dark current is also an indication of good material quality and device fabrication.

Considering the structure shown in Figure 2.2 under a reverse bias, when a photon with energy greater than the bandgap energy is incident, it is absorbed and a photogenerated EHP is created, which is an electron in the conduction band and a hole in the valence band. The field in the depletion region then separates the EHP and, causes the carriers to drift in opposite directions until they reach the neutral regions. This generates a current, called the photocurrent I_{photo}. The generation of EHPs is directly related to the absorption of light and it can be defined as the optical generation rate, G, given by

$$G = -\frac{1}{A}\frac{dP_{inc}}{dx}\frac{1}{h\nu} = \frac{\alpha P_{inc}}{Ah\nu} \tag{2.6}$$

Where P_{inc} is the incident optical power, α is the absorption coefficient (see Chapter 3), and $h\nu$ is the photon energy. If the light is absorbed in the depletion region, assuming no recombination of the photo-generated EHPs in this layer, the photocurrent is simply given by the integral of the generation rate over the depletion region:

$$I_{photo} = -qA\int_{-W_p}^{W_n} G dx = q\frac{(1-R)P_{inc}}{h\nu}(1-e^{-\alpha W}) \tag{2.7}$$

also allowing for the reflectivity R at the semiconductor surface. It should be noted that if the light is incident on the neutral region to calculate the photocurrent, the continuity equations should also be solved for holes and electrons in the n-side and p-side, respectively. Light absorbed in the neutral region leads, however, to a very low level of photocurrent due to free carrier absorption and high recombination rates in these regions.

The basic p-n junction photodiode has, however, two main drawbacks (described below) that are substantially reduced by using the p-i-n photodiode configuration. A schematic of a typical p-i-n structure and the electric field profile in the device is shown in Figure 2.3.

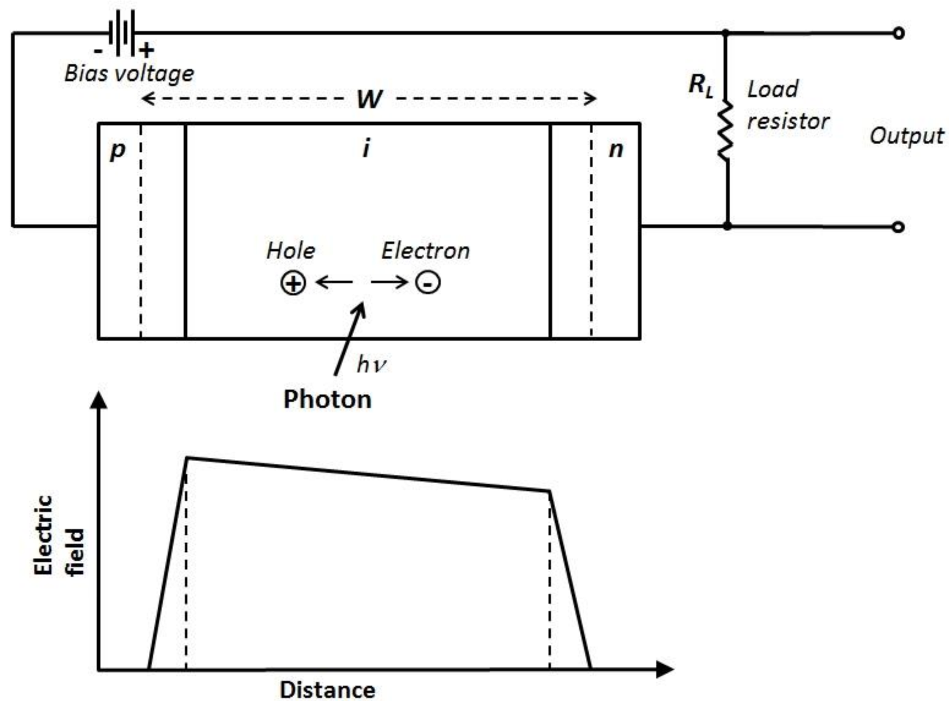


Figure 2.3. Schematic of a reverse biased p-i-n photodiode.

The particular geometry of a p-i-n device includes an intrinsic “i” region between the p and n region. When it is illuminated with energy greater than or equal to the bandgap of the semiconductor, electron-hole pairs (EHP) are generated. Due to the built-in potential, or by applying an external reverse bias, the intrinsic region is depleted (due to fairly low background doping level) and it has a high resistivity [3]. Because of this, the entire voltage drop takes place mainly in the “i” region, giving rise to a high electric field for the collection of the photo-generated EHP.

The main difference between these p-n and p-i-n structures is in terms of the depletion region, which is generally much narrower in the p-n structure than in the p-i-n structure. In the p-n structure there is more likely to be a large diffusion component that is due to the photons absorbed outside the depletion region. Minority carriers generated, electrons in the p-side and holes in the n-side, have to diffuse into the depletion region before they can drift to the n-side or p-side, respectively. Because diffusion is intrinsically a slow process, carriers take a long time to diffuse (about a nanosecond over 1 μm). In the p-i-n configuration, the depletion region can be designed to extend

across almost the whole length of the device, so that the drift component dominates over the diffusion component. In the depleted absorbing layer, the photo-generated carriers travel at their saturation velocity v_{sat}. The bandwidth of a p-i-n photodiode is determined by the carrier transit time, or by the RC time constant. Since transit time, junction capacitance, and quantum efficiency are interdependent, the design of a wide bandwidth p-i-n photodiode involves a performance tradeoff, and the depletion width W must be chosen depending on the requirements in terms on speed and sensitivity.

The performance of a photodiode is often characterised by the responsivity (R) that is defined as:

$$R = \frac{I_{photo}}{P_{optical}} = \frac{\eta q}{h\nu} \tag{2.8}$$

Where I_{photo} is the photogenerated current and $P_{optical}$ is the input optical power. The responsivity takes into account the energy of the incident photon. So, it takes into account the variation in energy implied by different wavelengths [4]. It is also related to the quantum efficiency η by the relation shown in Eq. 2.8. The quantum efficiency is the number of electron – hole carrier pairs generated compared to the number of incident photons of energy $h\nu$. Here, q is electron charge, h the Planck's constant and ν the frequency of the incident radiation.

In a p-n junction diode, the responsivity is limited by the diffusion lengths of the carriers. Minority carriers generated beyond a diffusion length from the depletion region recombine before they can reach the high field region.

In a p-i-n junction diode the responsivity is increased by adding an "i" region that will be always fully depleted so that any carriers generate in this region will be collected (assuming negligible recombination in the intrinsic region itself).

For a p-i-n photodiode with a totally depleted intrinsic layer, the external quantum efficiency, η_{ext}, can be written as

$$\eta_{ext} = (1 - R)(1 - e^{-\alpha d}) \tag{2.9}$$

where d is the width of the absorption region, α is the absorption coefficient, R is the reflectivity at the surface. The transit time limited -3 dB component of the bandwidth for a p-i-n having total depletion thickness W, is given by [5]

$$f_{tr} \cong \frac{3.5\bar{v}}{2\pi W}\ , \qquad \frac{1}{\bar{v}^4} = \frac{1}{2}\left(\frac{1}{v_e^4} + \frac{1}{v_h^4}\right) \tag{2.10}$$

where $\bar{v}$ is the average of saturation electron and hole velocity, and v_e and v_h are the electron and hole saturation velocities, respectively. The RC contribution to the bandwidth is given by

$$C = \frac{\varepsilon A}{W} \tag{2.11}$$

$$R = R_S + R_L \tag{2.12}$$

$$f_{RC} = \frac{1}{2\pi\tau_{RC}} = \frac{W}{2\pi R\varepsilon A} \propto \frac{W}{(R_S + R_L)A}\ , \qquad \tau_{RC} = RC \tag{2.13}$$

where ε is the dielectric constant of the intrinsic region and C is the intrinsic capacitance of the p-i-n structure. R_S is the series resistance including the semiconductor and the metal resistivity. R_L is the load resistance matching with the external circuit, which generally equates to 50 Ω. It is primarily determined by the photodiode junction capacitance thus can be controlled through adjusting the device area to a value that is consistent with a designed bandwidth. The total bandwidth of a p-i-n photodiode is expressed by the following equation

$$B = \frac{1}{\sqrt{\left(\frac{1}{f_{RC}}\right)^2 + \left(\frac{1}{f_{tr}}\right)^2}} \quad (Hz) \tag{2.14}$$

According to equation (2.9), both a low reflectivity R at the incident light surface and a large absorption region thickness are desirable to achieve a high quantum efficiency value and a low junction capacitance. However, a thicker absorber will decrease the photodiode bandwidth due to a longer carrier transit time. Therefore, as already mentioned, there exists a tradeoff between speed and quantum efficiency. This tradeoff can be addressed by decoupling light absorption from the carrier transport by using waveguide device structures or by increasing the responsivity by using a resonant cavity structure.

2.3 Avalanche photodiodes

Avalanche photodiodes (APDs) are another type of photodetector widely used in telecommunications systems. By utilising the impact ionisation process to provide an internal gain, APDs have the advantage of higher sensitivity over conventional p-i-n photodiodes, since the photocurrent is multiplied, when the overall noise in the receiver module is dominated by electronic noise.

In order for carrier multiplication to take place, the photogenerated carriers must traverse a region where a very high electric field is present. In this high-field region a photogenerated electron or hole can gain enough energy so that it ionises bound electrons in the valence band by colliding with them, resulting in an electron in the conduction band, leaving a hole in the valence band, and the original electron. This carrier multiplication mechanism is known as impact ionisation. The newly created carriers are also accelerated by the high electric field, potentially gaining enough energy to cause further impact ionisation events. This phenomenon is the avalanche effect and it is schematically illustrated in Figure 2.4.

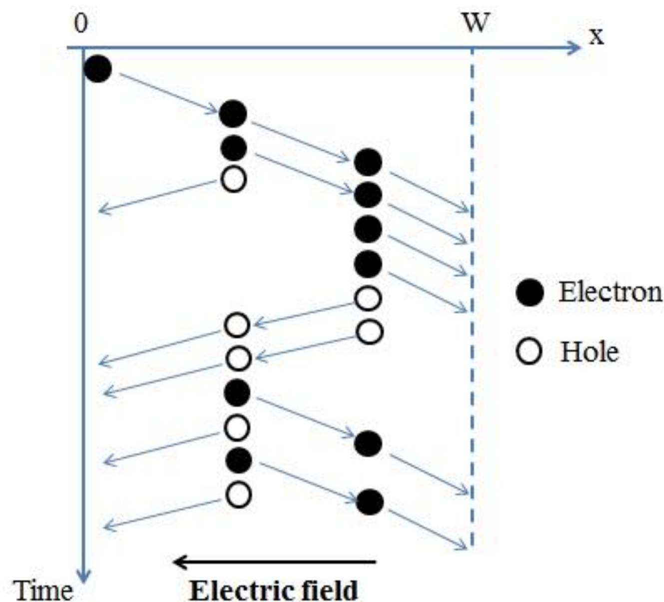


Figure 2.4. Schematic of multiplication process: a single carrier (electron) is injected at the edge of avalanche region under high electric field and a chain of impact ionisation events is triggered.

During the carrier acceleration process under the electric field, in addition to the impact ionisation process, the free carrier also encounters non-ionisation collisions with phonons, generally lowering the carrier's kinetic energy. Phonon scattering may involve

carriers gaining energy (phonon absorption), losing energy (phonon emission) and exchanging momentum (elastic).

The impact ionisation characteristics of electrons and holes in a semiconductor are generally characterised by the impact ionisation coefficients, α and β respectively. These electric field dependent coefficients describe the average number of electron-hole pairs created per unit distance travelled by an electron or hole in a uniform electric field and thus can be defined as the inverse of the mean distance between successive ionisation events. In general, α and β increase with the applied electric field because the carrier achieves the required energy for an impact ionisation event over a smaller distance.

In its simplest form, an APD has a similar device structure to a p-i-n photodiode. However, unlike the p-i-n, the APD is always biased at a voltage that is close to its breakdown. Through the carrier impact ionisation process, an APD provides more electron-hole pairs from the same amount of photogenerated carriers, compared to a p-i-n photodiode. Therefore, if we assume the same optical absorption volume dimension for both a p-i-n and an APD, the external quantum efficiency of the APD can be written as

$$\eta_{ext} = M \cdot (1 - R)(1 - e^{-\alpha d}) \tag{2.15}$$

where M is the avalanche gain obtained through carrier impact ionisation. The gain M is given by the expression

$$M = \frac{i_{photo} - i_{dark}}{i_{primary,photo} - i_{primary,dark}} \tag{2.16}$$

where i_{photo} and i_{dark} are the multiplied photocurrent and dark current, while $i_{primary,photo}$ and $i_{primary,dark}$ are the primary photocurrent and dark current measured prior the onset of carrier multiplication (M=1).

The stochastic nature of the impact ionisation process leads to a notable spread of the total number of carriers generated by an injected electron-hole pair under a specific electric field, giving rise to excess noise which is characterised by the excess noise factor, F. The statistical nature of this impact ionisation process results in fluctuations in the number of secondary carriers (electron – hole pairs) generated from each parent carrier, which in turn leads to fluctuations in the APD gain [6]. Thus, while M can be

very large, the actual usable value of M for fibre optic receivers is normally limited to between 10 and 20 depending on the materials and device structures used. This gain fluctuation will cause variation in the APD output current and increase the APD's total shot noise power above that of the multiplied shot noise power. The excess noise factor, *F(M)*, is a measure of the standard deviation of the multiplication gain over its mean square value

$$F(M) = \frac{<M^2>}{<M>^2} \tag{2.17}$$

According to the local impact ionisation model developed by McIntyre [6] in which carrier impact ionisation coefficients were assumed to be only a function of electric field strength, *F(M)* can be written as

$$F(M) = kM + (1-k)\left(2 - \frac{1}{M}\right) \tag{2.18}$$

Where $k = \beta/\alpha$ $(\beta < \alpha)$ or $k = \alpha/\beta$ $(\alpha < \beta)$. Based on the local field theory, larger difference between carrier impact ionisation coefficients leads to lower excess noise in an APD, as shown in Figure 2.5.

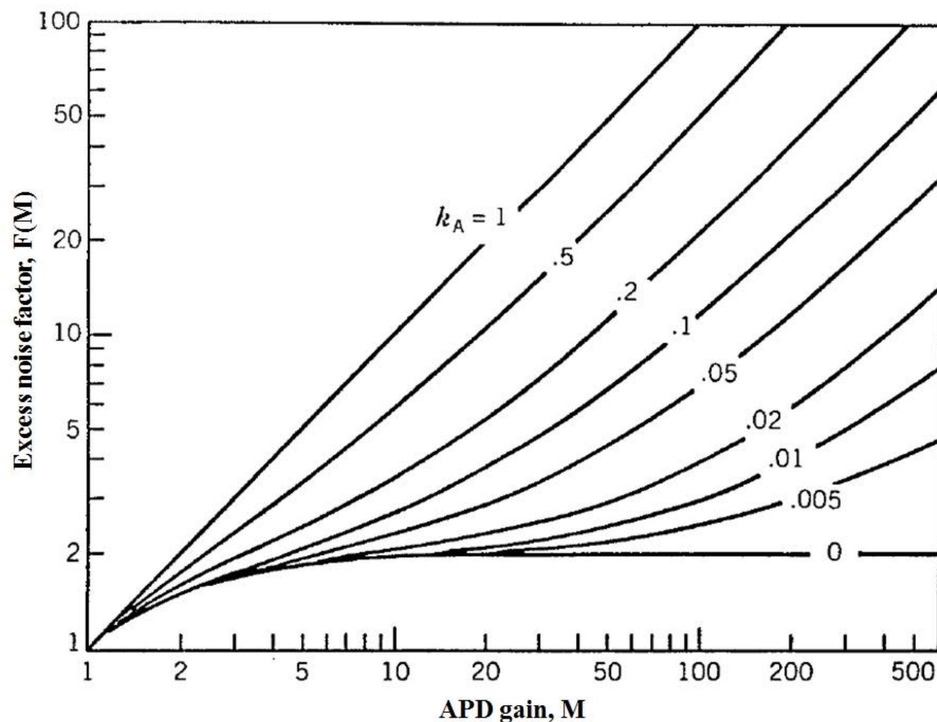


Figure 2.5. Excess noise factor predicted by McIntyre's local field theory as a function of gain (M) and various k ratios [3].

Another phenomenon associated with carrier multiplication is the effect of the gain process on bandwidth. Detailed research on this problem was performed by

Emmons [7], who showed that the frequency-dependent gain can be approximated by the expression

$$M(\omega) = \frac{M_0}{\sqrt{1 + (\omega M_0 k\tau)^2}} \tag{2.19}$$

where M_0 is the dc gain, and τ is approximately (within a factor of ~2) the carrier transit time across the multiplication region. Figure 2.6 shows the 3-dB bandwidth of the photodiode plotted as a function of the multiplication gain M_0 and the parameter k. Superimposed on these curves is the curve $M_0 = \alpha/\beta$.

Above this curve, where $M_0 < \alpha/\beta$, multiplication has little effect on bandwidth and current multiplication is obtained without any substantial reduction in bandwidth. It is important to note that this condition, which maximises the bandwidth in the avalanche region, also minimises the noise produced by the avalanche multiplication mechanism.

On the other hand, for $M_0 > \alpha/\beta$, the curves are straight lines, indicating a constant gain-bandwidth product (GBP), and a dependence of multiplication on bandwidth. Therefore, Emmons's finding imposed an upper limit on the APD bandwidth, in terms of the gain-bandwidth product.

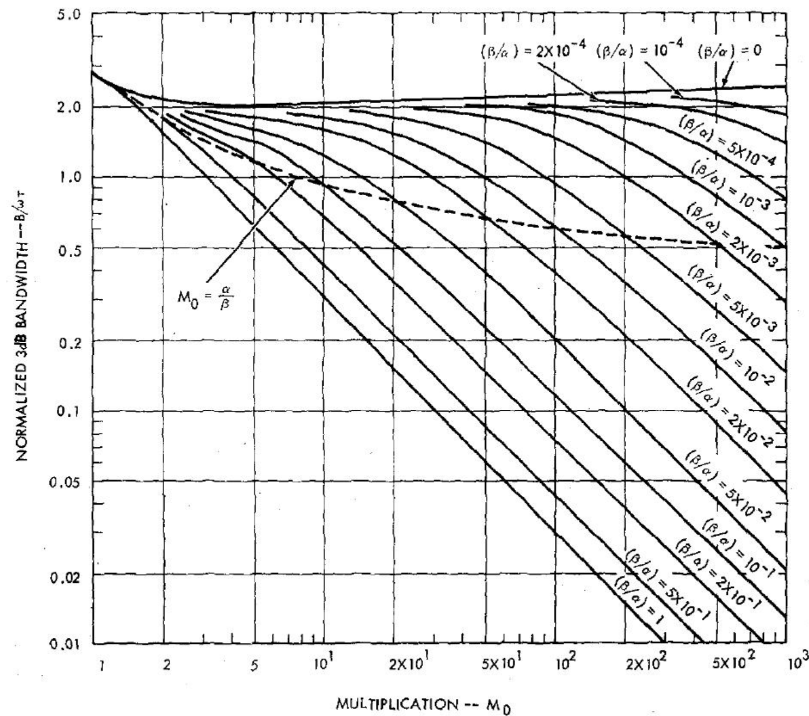


Figure 2.6. APD bandwidth as a function of multiplication gain and the parameter k, as predicted by local field theory [7].

According to the local field model, the performance of an APD is determined by its k value, which is an intrinsic property of each specific material. As the APD multiplication layer is thinned, a higher electric field is required to achieve a specific gain value than in a thicker device. The impact ionisation rate curves for electrons and holes merge at high electric field, causing k to approach unity. Therefore, it is expected that the excess noise would increase as the multiplication thickness decreases. The opposite, however, has been observed experimentally [8]. This discrepancy found a solution in the non-local model, which takes into account the history of the carrier prior to the impact ionisation event. The reduction of $F(M)$ in thin multiplication layer APDs has been attributed to the pronounced dead-space effect [9]. When a carrier, either a hole or an electron, initially enters the high field region it must travel for a certain distance, the so-called dead-space, before it gains enough energy from the electric field to undergo a subsequent impact ionisation event. In thick devices this distance is negligible compared to the total device thickness; therefore the dead-space plays a minor role in the noise characteristic of thick APDs. The effect of dead-space starts to influence the performance in thin multiplication regions for which the dimension of the dead-space becomes comparable. As a result, the impact ionisation events are more localised in a specific spatial range in the gain region and the impact ionisation events are more deterministic [10]. This reduces the variation in numbers of impact ionisation events every carrier will cause, which lowers the excess noise related to the ionisation process.

Therefore, to achieve low noise and high-speed APDs materials must be used that have lower ionisation coefficient ratio (k) value and thin multiplication regions (dead-space effect).

Another important point is that the gain mechanism of an avalanche photodiode is very temperature-sensitive because of the temperature dependence of the electron and hole ionisation rates. This temperature dependence is particularly critical at high bias voltage, where a small change in temperature can cause large variation in gain [2]. As the temperature in the semiconductor device is reduced, the lattice vibrational energy is also reduced, and so is the scattering of electrons and holes. This increases the energy of the carriers and hence their ionisation rates. As a consequence of the change in ionisation coefficients for a given temperature

decrease, the electric field will have to decrease to maintain breakdown. The value of the breakdown voltage drops with decreasing temperature, T, following the formula [11]

$$\frac{\Delta V_{BD}}{V_{BD,RT}} = \gamma T \tag{2.20}$$

where $V_{BD,RT}$ is the breakdown voltage at room temperature and γ is an index typical of the device under observation.

2.3.1 Advantage of an APD over a p-i-n photodiode

In fibre optic communication systems the photodiode is generally required to detect very weak optical signals. This requires the optimisation of the photodetector and its amplification circuit so that a given signal-to-noise ratio (S/N) is maintained. This ratio at the output of an optical receiver is defined by

$$\frac{S}{N} = \frac{signal\ power\ from\ photocurrent}{photodetector\ noise\ power + amplifier\ noise\ power} \tag{2.21}$$

It is clear that to obtain a high S/N ratio, the photodetector must have a high quantum efficiency to generate a large signal power and the photodetector and amplifier noise should be kept as low as possible.

For a p-i-n photodiode the mean square signal current $\langle i_s^2 \rangle$ is

$$\langle i_s^2 \rangle = \langle i_{ph}^2(t) \rangle \tag{2.22}$$

Where $i_{ph}(t)$ is the signal component related to the incident optical power by

$$i_{ph}(t) = \frac{q\eta_{ext}}{h\nu} P_{inc}(t) \tag{2.23}$$

For an APD

$$\langle i_s^2 \rangle = \langle i_{ph}^2(t) \rangle M^2 \tag{2.24}$$

the signal power term is boosted by a factor of M^2, representing the average of the avalanche gain, which greatly enhances the receiver sensitivity.

The noise current determines the minimum optical power level that can be detected, because the optical receiver must have the ability to discriminate between the minimum detectable input optical power and the noise generated by the integrated receiver itself within a working bandwidth. The noise is described in terms of noise power spectrum density $\langle i^2 \rangle$, which is the quadratic mean of the noisy current within a unity frequency bandwidth. Different noise sources affect the signal-to-noise ratio. For photodiodes operating at high frequencies, thermal noise and shot noise are two main types of noise sources.

Thermal or Johnson noise originates from the Brownian motion of carriers that exists in any type of resistive components. It is expressed by

$$\langle i_T^2 \rangle = \frac{4k_B T}{R_L} B \tag{2.25}$$

Where B is the operation bandwidth, T is the absolute temperature, k_B is Boltzmann's constant, and R_L is the load resistance.

The shot noise arises from the statistical nature of the production and collection of photoelectrons when an optical signal is incident on a photodetector, and it follows a Poisson distribution. It can be expressed by

$$\langle i_Q^2 \rangle = 2qI_{ph}BM^2F(M) \tag{2.26}$$

Where q is the electron charge, I_{ph} is the average photocurrent due to the signal power, $F(M)$ is the excess noise factor associated with the random nature of the avalanche process, and M is the avalanche gain. For p-i-n photodiodes M and $F(M)$ are unity.

As already mentioned, the dark current in a photodiode is a source of noise. This current is a combination of bulk and surface currents. The bulk component (i_{DB}) arises from electrons and holes which are thermally generated in the junction of a photodiode. In an avalanche photodiode these carriers also get accelerated by the high electric field, and are therefore multiplied. The mean square value of this current is given by

$$\langle i_{DB}^2 \rangle = 2qI_D M^2 F(M)B \tag{2.27}$$

where I_D is the primary (unmultiplied) detector bulk current. Regarding the noise associated to the surface leakage current (i_{DS}), that depends on surface defects, bias voltage, and surface area, it can be expressed by

$$\langle i_{DS}^2 \rangle = 2qI_L B \tag{2.28}$$

where I_L is the surface leakage current. This component is not affected by the avalanche gain, since the avalanche multiplication is a bulk effect.

Figure 2.7 shows a schematic of a detection system with the associated noise sources.

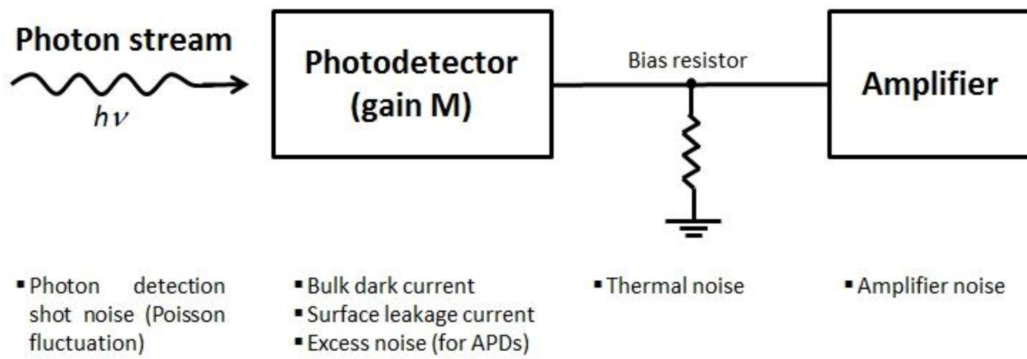


Figure 2.7. Schematic of a detection system with the related noise sources for each component.

Based on the above discussion, for the basic detection system shown in Figure 2.7, it is possible to write the S/N ratio as

$$\frac{S}{N} = \frac{\left(\frac{q\eta_{ext}}{h\nu} P_{inc}\right)^2}{\left[2q\left(I_{ph} + I_D\right)F(M) + 2q\frac{I_L}{M^2} + \frac{4k_B T}{R_L M^2}\right] B} \tag{2.29}$$

In the case of a p-i-n photodiode, where F(M) and M are unity, the dominating source of noise current is that of the detector load resistor (i_T). Therefore, it is desirable to have a high load resistance to reduce the thermal noise, but this increases the RC constant, which reduces the bandwidth of the detector and thus degrades the S/N ratio. For APDs, the thermal noise is reduced by a factor of M^2 which also enhances the S/N ratio. The total shot noise term is, however, multiplied by a factor F(M). A noise penalty therefore will occur as F(M) increases with M. Thus for a given set of operating conditions, there exists an optimum value of M for which the S/N ratio is a maximum. This is mainly true for low level light,

where the APDs show a high S/N ratio compared to p-i-n photodiodes. Furthermore, higher dark and surface leakage current levels set a limit of achievable sensitivity of the APD receiver even if F(M) is low. Consequently, these terms must be minimised to achieve high APD receiver sensitivity.

2.4 Single-photon detectors

A single-photon detector is an extremely sensitive device capable of registering the individual photons. Photons have so many properties that its detection can be employed in a wide range of applications such as quantum information, time-correlated single-photon counting (TCSPC) applications, etc. These applications demand stringent requirements in terms of signal-to-noise ratio, detection efficiency, spectral range, etc. It is because of all of these requirements that different technologies (photomultiplier tubes, semiconductor avalanche photodiodes, superconducting nanowires, etc.) have been employed to detect single-photons. In this section a review of these technologies is presented with a description of their main advantages and disadvantages, and is mainly focused on semiconductor single-photon avalanche photodiodes (SPADs), and on those technologies that enable the single-photon detection in in the infrared spectral range.

2.4.1 Photomultiplier tube (PMT) and Microchannel plates

A basic PMT consists of a vacuum tube with a photocathode for light absorption, from which electrons are liberated through the photoelectric effect (the energy of the incident photon must exceed the work function of the photocathode material) [12], [13]. Depending on the material composition of the photocathode, PMTs can be effective for detection of light at varying wavelengths. A schematic representation of a typical PMT structure is shown in Figure 2.8.

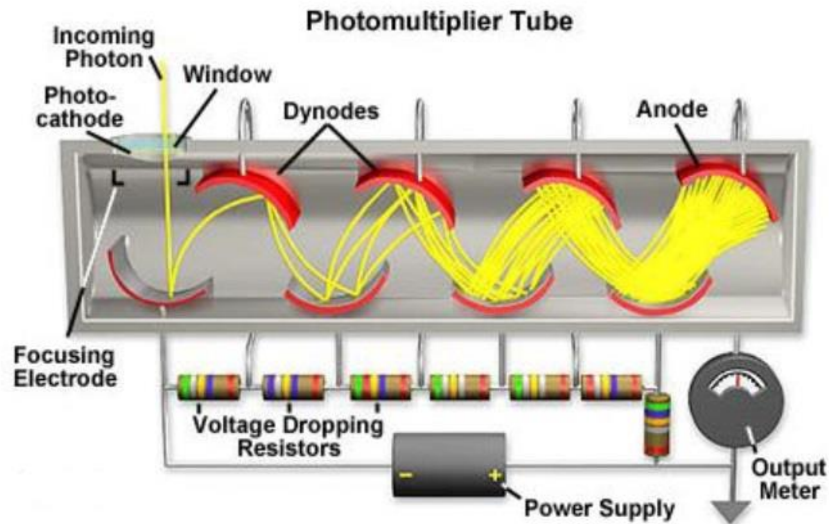


Figure 2.8. Schematic representation of a photomultiplier tube (PMT) [12].

Photomultipliers acquire light through a glass or quartz window that covers a photosensitive surface, called a photocathode, which then releases electrons that are multiplied by electrodes known as metal channel dynodes. At the end of the dynode chain is an anode or collection electrode. Over a very large range, the current flowing from the anode to ground is directly proportional to the photoelectron flux generated by the photocathode. Each of dynodes is biased at a greater positive voltage than the one before, producing a macroscopic current pulse of > 10^6 electrons.

There are, however, several disadvantages with such devices, including relatively large devices with poor mechanical stability and low SPDE (usually a few % in the infrared). PMTs also require high bias voltages in the order of 1 kV and their timing jitter (fluctuation of transit time) is typically in the region of 1 ns. Therefore, in many modern photon counting applications, PMTs are far from ideal where higher efficiencies and lower jitter are required. However, PMTs generally do not have a spectrally dependent instrumental response (unlike some semiconductor-based detectors), and much improvement has been seen in miniaturisation of the PMT packages [14].

An alternative configuration is the microchannel plate photomultiplier tube, see Figure 2.9, where glass capillaries are fused in parallel and coated with a secondary electron emitting material to achieve a single continuous dynode under a bias voltage [15]. Microchannel plate PMTs offer improved timing jitter over basic PMTs, down to ~20 ps at FWHM [15].

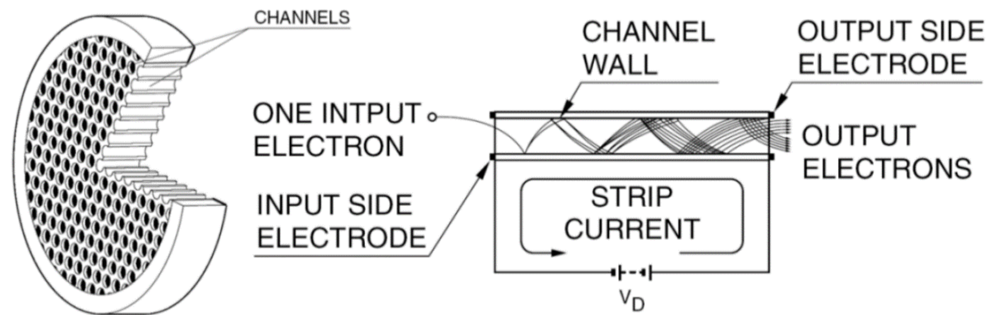


Figure 2.9. Schematic representation of a Microchannel plate [15].

2.4.2 The single-photon avalanche diode (SPAD)

Semiconductor avalanche photodiodes (APDs) have the typical advantages characterising the solid-state devices (small size, low bias voltage, low power consumption, and reliability). When the APD is operating in a linear mode (which means that the output current is proportional to the input optical power) it has an internal gain. However, this internal gain is not sufficient to detect single-photons. To overcome this problem the APD can be used above the avalanche breakdown, in the so-called Geiger mode of operation (since their operation principal is similar to that of the Geiger-Muller detectors, in which particle emission from radioactive materials gives rise to an avalanche of carriers from ionised gas atoms) [16]. When the APD is operating in Geiger mode it is called a Single Photon Avalanche Diode (SPAD) or triggered avalanche detector.

Investigation of the physics of avalanche breakdown in SPADs started back in the 1960s at the Shockley laboratory [17], [18], where it was observed that a p-n junction reverse biased above the breakdown level produced a macroscopic voltage pulse triggered by the absorption of single optical photons. Fundamental contributions to the understanding of the avalanche phenomenon and of its statistical properties were given by McIntyre and Webb in 1970s [19]–[21]. However, the first custom-designed SPAD was proposed and demonstrated by Cova *et al.* in 1981 [22], and it was a thin-junction Si SPAD. As pointed out in Figure 2.10, SPAD operation is radically different from that of ordinary APDs, which are biased near to the breakdown voltage V_{BD}, but below it. In APDs, avalanche multiplication is exploited to produce linear amplification of the primary photogenerated electrical signal.

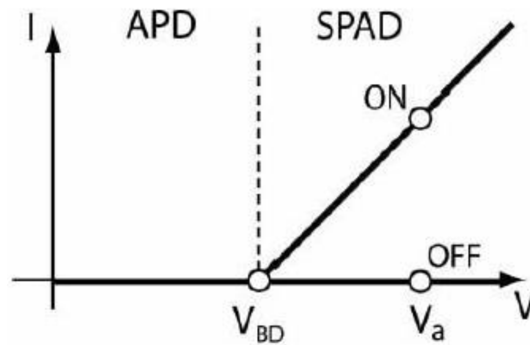


Figure 2.10. Schematic of current-voltage characteristic for a APD and SPAD operation in the reverse I-V characteristic of a p-n junction [22].

The behaviour of a SPAD is similar not to an amplifier, but to a bistable circuit as shown in Fig. 2.10. When the SPAD device is biased at voltage V_a above the V_{BD} (Geiger-mode) no current flows (OFF-STATE). In the junction depletion layer the electric field is very high, but no free carriers are present. When even a single carrier is injected into the high-field region, it is strongly accelerated and gains sufficient kinetic energy that, on collision with an atom in the lattice, it frees an electron from its bound state and promotes it into the conduction band, leaving a hole in the valence band. This phenomenon is known as impact ionisation, as described previously. These generated carriers may undergo further impact ionisation events, initiating an avalanche of carriers. A self-sustaining current (in the mA range) is then triggered (ON-STATE). The current is limited to a constant level by the space charge effect, which is proportional to the excess bias voltage $V_{EX} = V_a - V_{BD}$. The device remains in this ON-STATE until the avalanche is quenched by an external circuit, which drives the applied voltage down to below V_{BD}. The detector is insensitive to any subsequent photons arriving in the time interval from the avalanche onset to the voltage reset, which is the detector *dead time*. It is evident that describing the magnitude of the gain in the case of SPADs does not make sense, just as in the case of a bistable circuit.

It has already been shown in this chapter, that the performance of a photodiode is related to the intrinsic physical properties of the semiconductor material used (band gap, intrinsic carrier concentration, absorption coefficient, etc.). For a SPAD, these properties are also important in conjunction with other physical effects. It is possible to define key parameters for a SPAD as the single-photon detection efficiency (SPDE), dark count rate (DCR), noise equivalent power (NEP), afterpulsing and the timing jitter.

2.4.2.1 Figures of merit of a SPAD

In terms of SPDE, the detection of a single photon is not only related to its absorption in the detector active area and generation of a primary EHP. It is also mandatory that the primary carrier succeeds in triggering a self-sustaining avalanche. As studied mainly by McIntyre and Oldham [19], [20], [23], the avalanche triggering probability (defined as the probability that carriers, either holes or electrons, will initiate a self-sustaining avalanche) will first increase linearly with low excess bias voltages V_{EX} and then tends to saturate to 1 at high V_{EX}. This probability also depends on the electron and hole ionization coefficient α_e and α_h, respectively, and from their ratio $k = {}^{\alpha_h}/_{\alpha_e}$. These coefficients are not only a function of the electric field, but their values vary with the crystalline perfection of the sample, so that a reliable value of the ratio k can only be determined if α_e and α_h are measured in the same sample. In particular, as pointed out by McIntyre, in designing a single-photon detector, a material having a high ratio of the electron and hole ionisation coefficients (as for Si) should be used, provided that it is possible to ensure that most of the photogenerated carriers entering the multiplying region are those with the highest ionisation coefficient.

For a SPAD the SPDE will also depend on another two factors: the coupling efficiency at the air-semiconductor boundary, and the absorption of a photon in the absorption layer. The first factor is determined by the probability of the photon being coupled to the active area of the device and the probability that it passes the air-semiconductor boundary. Usually, an anti-reflection coating (ARC) layer is used to prevent reflection losses and enhance the coupling efficiency. The second term depends on the width of the absorption layer and the absorption coefficient of the semiconductor material of which it is composed and it follows equation 2.7, as shown previously.

A limiting factor for a SPAD is represented by the dark count rate (DCR). It is expressed in Hz or, more correctly, in counts/sec, and represents the number of times the detector is triggered during 1 second, in dark conditions, by noise events (originating from different sources). Similar to the dark current in an ordinary photodiode, the DCR represents the internal noise of a SPAD. Different mechanisms contribute to the DCR in a single-photon detector [24], [25]:

a) Carriers thermally generated in the active volume of the device;
b) Carriers generated by tunnelling processes;

c) Carriers emitted from trapping levels that were populated in previous avalanche pulses.

The first two processes produce avalanche events in the absence of photons and cause dark counts. In the third case process, it is necessary to previously have triggered an avalanche (by a photon detection or a dark count) in the detector, and this leads to afterpulsing, which will be discussed later.

The spontaneous thermal generation of dark counts is dominated by the Shockley-Read-Hall (SRH) effect, which is mainly due to local defects and thermal excitation of electrons into the conduction band from the valence band. This contribution can be significantly reduced by cooling the detector.

The DCR of a SPAD also increases with the excess bias voltage V_{EX}. The rise is due not only to the avalanche triggering probability, which increases also the SPDE, but also to effects due to the electric field that enhance the rate of generation of carriers. In silicon, as well as in others semiconductors, a transitions of carriers from one band to another occur through the generation-recombination (GR) centres. At high electric fields, the emission from these centres to the band at higher energy can be enhanced by the Poole-Frenkel effect, giving rise to a field-assisted generation process [16]. At very high electric fields a tunnel-assisted direct band-to-band transition may also take place, generating free carriers in the junction, without the assistance of GR centres. This effect is not reduced by lowering the detector temperature (setting a limit to the DCR reduction by cooling), but the electric field should not be higher than the level necessary to obtain avalanche multiplication.

The noise in a SPAD is further increased by afterpulsing. Impurities and crystal defects cause not only GR centres at mid-gap, but also local levels at intermediate energy between mid-gap and band-edge, called deep levels. Some of these deep levels act as minority carrier traps. During a self-sustaining avalanche, the large charge flow through the device causes these trap centres to fill. After the avalanche is quenched, the trap centres start to release but these carriers can re-trigger the avalanche if the device is armed before all traps are depopulated, thereby generating an afterpulse correlated with previous avalanche pulse. This afterpulse contributes to the dark count rate of the detector.

The frequency at which the SPAD is gated must be sufficiently low to allow all the traps to empty (when the SPAD is quenched so that they cannot trigger a self-sustaining

avalanche). This effect is highly dependent on the excess bias. The higher the excess bias voltage, the more charge flows through the device, the greater the number of carriers that are trapped. Moreover, this effect does not improve at lower temperatures, because the emission lifetime of carriers from trapping levels exponentially increases with decreasing temperature. To reduce the effects of afterpulsing, the charge flowing through the device during an avalanche must be minimised, and this is commonly done using appropriate quenching circuits [26]. Hence, to reduce the afterpulsing probability in a SPAD device operating at cryogenic temperature a very long dead time is necessary (tens of microseconds or more). This leads to a strong reduction in the data collection rate, which represents the main limitation of SPADs operating at infrared wavelengths.

The most useful figure of merit of a SPAD is represented by the noise equivalent power (NEP) that takes into account both the SPDE and DCR. The NEP is defined as the signal power required to attain a unity signal to noise ratio within a 1-s integration time. To calculate the NEP the following equation is used:

$$NEP = \frac{h \cdot \nu}{SPDE} \sqrt{2DCR} \tag{2.30}$$

Where h is Planck's constant and ν is the optical frequency. The lower the NEP the more sensitive the device.

The timing resolution, also referred to as jitter, is typically described by the full-width half maximum (FWHM) of the SPAD timing response. It is measured in a time-correlated single photon counting (TCSPC) setup (see Chapter 5) when the detector is illuminated with a highly attenuated pulsed laser (~0.1 photons per pulse) with short pulse duration [27]. It represents the precision with which the arrival of a photon can be measured. The delay between the detection of an avalanche and a single photon in a SPAD detector can originate from various sources [25]:

a) the timing difference arising from photon absorption at different depths and the subsequent variance in diffusion time;
b) the stochastic nature of the multiplication process itself;
c) the lateral movement of the multiplication across the detector active area [28], which is strongly influenced by the location of the seed point of the avalanche. Consequently, jitter depends on the device diameter and smaller devices exhibit improved timing precision as the variation in the seed point is decreased.

In some cases an increase in the excess bias reduces the jitter. As the electric field increases, the carriers reach the high-field region quicker, hence reducing the jitter.

2.4.2.2 *Quenching techniques for SPADs*

As previously mentioned, when the SPAD operates in Geiger-mode, it is in OFF-STATE until a self-sustaining avalanche is triggered (ON-STATE) by a carrier. During the avalanche process, the detector cannot register incident photons, and the avalanche must be quenched and the SPAD reset to its initial, photo-sensitive state. The quenching operation is performed by an external circuit and it is possible to distinguish different quenching techniques: passive quenching, active quenching and gated-mode which has been used in this work to characterise the Ge-on-Si SPADs.

The passive quenching technique has been used in the early studies on avalanche breakdown in junctions [29]. A schematic of a passive quenching circuit is illustrated in Figure 2.11.

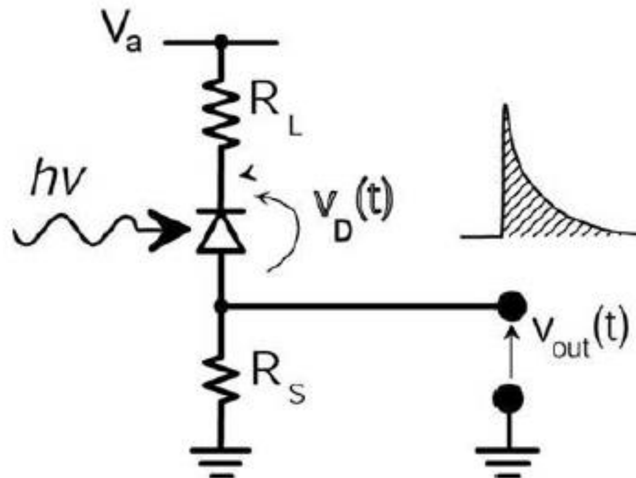


Figure 2.11. Schematic circuit diagram of a SPAD with a passive quenching circuit from [16] *where R_L (~500 kΩ) is the quenching resistor and V_a is a voltage above the breakdown.*

The SPAD is reverse biased above breakdown through a quenching resistor, R_L, of usually 500 kΩ or more [16] and the output is measured on the R_S (~50 Ω) resistance.

When the SPAD is in stand-by mode, there is no current flowing in the circuit and therefore the bias across the SPAD is equal to V_a. At the onset of the avalanche a current will start flowing in the resistor R_L, thus generating a reduction in the bias across the diode taking it below the breakdown voltage, quenching the avalanche. The quenching time is limited by the product of the SPAD's capacitance and the value of R_L [26], which can generate a very long recovery time after an avalanche, typically of the order

of a few microseconds. In these cases, this technique is only suitable when the count rate is lower than a few tens of kcounts^{-1}.

A solution which avoids the drawbacks of passive quenching circuits, slow voltage recovery and not well defined dead time, is to use the Active Quenching Circuit (AQC). A schematic of an AQC is shown in Figure 2.12.

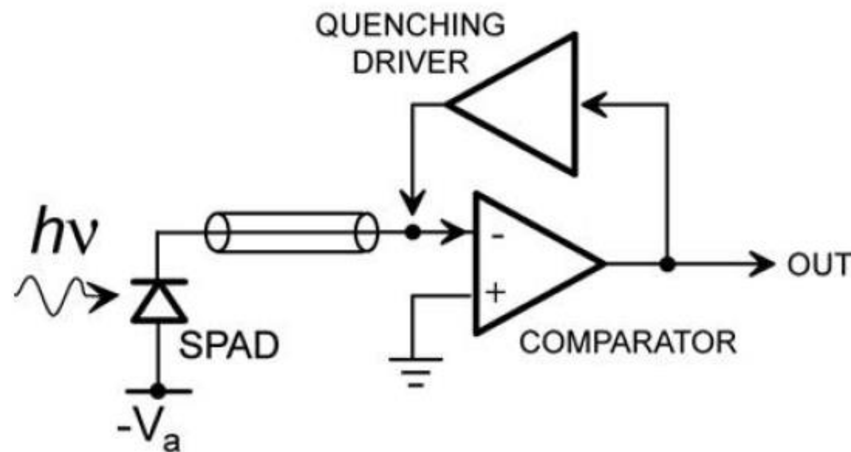


Figure 2.12. Schematic of an AQC from [26].

The fast comparator senses the rise of the avalanche pulse, then the comparator's output then switches the bias voltage source to less than V_{BD} to quench the avalanche. The voltage supply to the SPAD is kept below V_{BD} for a controlled period of time (hold-off) to allow the release of any trapped carriers to avoid the effects of afterpulsing. The output from the comparator is also used as the photon-counting signal for recording the arrival time of a photon since it is coincident with avalanche build-up. AQCs have several advantages, in particular, the recovery time is short (~3 ns or greater) allowing operation at high count rates. Moreover, this time could be set to longer values in order to reduce the effect of afterpulsing.

The last quenching technique is 'gated-mode'. This technique has been used throughout much of this work to characterise the designed Ge-on-Si SPADs. By using both passively and actively quenched circuits the SPAD can operate in free-running mode (it can detect photons at any time, except for the time necessary to reset the avalanche). In gated-mode operation, the SPAD is sensitive only during the gate-on window, with the SPAD disabled during the gate-off time interval. Figure 2.13 shows a schematic of this mode of operation.

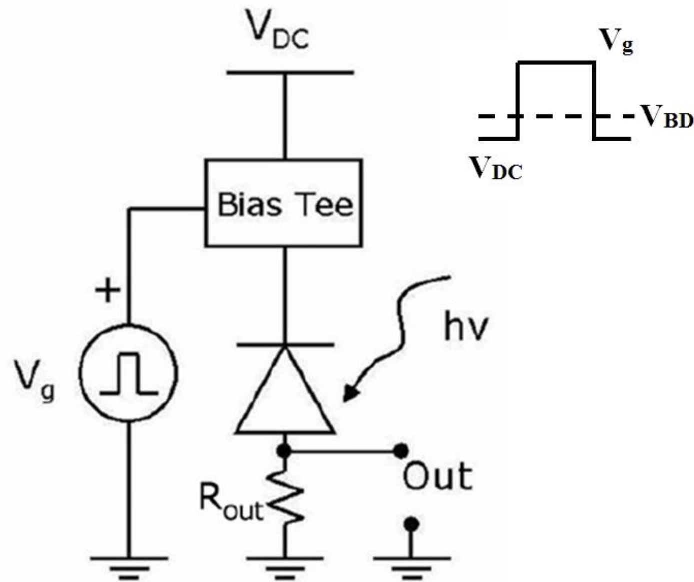


Figure 2.13. Gated mode circuit scheme [11]. *Vg is the gate voltage, V_{DC} is the constant bias, V_{BD} the breakdown voltage and R_{out} is the resistance of the output circuit, typically equivalent to 50 Ω.*

The device is DC biased slightly below the breakdown voltage. When a gate voltage (variable in duration, usually 10's of ns) is superimposed, it raises the SPAD voltage (V_{DC} + V_g) above the breakdown (Geiger-mode), in order that the SPAD can detect a single-photon. The end of the gate defines the time of quenching. Additionally, it is possible to introduce a hold-off time (or dead time) to ensure that the system returns to its initial condition and it can detect photons with a low afterpulsing probability. Gated mode quenching is most suited to characterisation since it allows the detector to be activated coincident with a highly attenuated laser pulse such that the SPAD is gated at the same repetition rate as the laser.

2.5 Conclusion

In this chapter, the physics of different photodiode structures have been introduced. The basic working principles of the p-n junction under equilibrium (fixed temperature, no external bias) and when reverse biased has been considered. In particular, the origin of the leakage or dark current has been explained for a reverse biased photodiode where the diffusion and thermal generation within the depletion region are the main mechanism at low voltages. The surface leakage and tunnelling component, however, have been intentionally omitted because their contribution to the total dark current depends on others factors (device geometry, material processing, bias condition, etc.) and will be discussed in the following chapter.

The p-i-n and avalanche photodiodes have been explained by using their most important figures of merit. In particular, it is clear from the signal-to-noise analysis that APDs are the natural choice for detecting low level light thanks to their internal gain mechanism. However, the gain must be chosen according to the requirements in terms of gain-bandwidth product due to the excess noise related to its gain mechanism. According to the local field theory this excess noise factor depends from the ionisation coefficients of electrons and holes, which are an intrinsic property of the material itself and are related to the strength of the electric field. However, the width of the multiplication region must be considered during the design of an APD. In particular, for thin multiplication region, the dead space effect, which is taken into account by the non-local field theory, plays an important role on the APD noise. This noise is reduced because in thin multiplication regions, the impact ionisation events are more deterministic. Therefore, to achieve low noise and high-speed APDs must be used materials that have lower ionisation coefficient ratio (k) value, and thin multiplication regions (dead-space effect).

Finally, SPADs were introduced and key operating conditions discussed. Whilst APDs are biased below the avalanche breakdown voltage, a SPAD is biased above the breakdown voltage in the so-called Geiger mode of operation. In this state, no current flows through the device, but a single carrier (dark or photo-generated) can generate, through the avalanche process, a self-sustaining current in the mA range. This state is referred as ON-STATE. The device remains in this state until the avalanche is quenched. A brief explanation of the various quenching circuits has been treated through the chapter. New key parameters (DCR, SPDE, NEP, afterpulsing, timing jitter) that are important for designing a SPAD have been defined.

2.6 Reference

[1] T. P. Pearsall and M. A. Pollack, “Chapter 2 Compound Semiconductor Photodiodes,” vol. 22, Part D, W. T. Tsang, Ed. Elsevier, 1985, pp. 173 – 245.

[2] G. Keiser, *Optical fiber-communication*, 2nd Editio. New York: Mc-Graw-Hill, 1991.

[3] G. P. Agrawal, *Fiber-Optic Communication Systems*, Third edit. John Wiley & Sons, Inc., 2002.

[4] H. J. R. Dutton, *Understanding Optical Communication*, First Edit. IBM Corporation, 1998.

[5] J. E. Bowers and J. Burrus, C.A., “Ultrawide-band long-wavelength p-i-n photodetectors,” *J. Light. Technol.*, vol. 5, no. 10, pp. 1339–1350, 1987.

[6] R. J. McIntyre, “Multiplication noise in uniform avalanche diodes,” *IEEE Trans. Electron Devices*, vol. 13, no. 1, pp. 164– 168, 1966.

[7] R. B. Emmons, "Avalanche Photodiode Frequency Response," *J. Appl. Phys.*, vol. 38, no. 9, pp. 3705–3714, 1967.

[8] J. C. Campbell, S. Demiguel, Feng Ma, A. Beck, Xiangyi Guo, Shuling Wang, Xiaoguang Zheng, Xiaowei Li, J. D. Beck, M. A. Kinch, A. Huntington, L. A. Coldren, J. Decobert, and N. Tscherptner, "Recent advances in avalanche photodiodes," *IEEE J. Sel. Top. Quantum Electron.*, vol. 10, no. 4, pp. 777– 787, 2004.

[9] P. Yuan, K. A. Anselm, C. Hu, H. Nie, C. Lenox, A. L. Holmes, B. G. Streetman, J. C. Campbell, and R. J. McIntyre, "A new look at impact ionization-Part II: Gain and noise in short avalanche photodiodes," *IEEE Trans. Electron Devices*, vol. 46, no. 8, pp. 1632–1639, 1999.

[10] J. C. Campbell, "Recent Advances in Telecommunications Avalanche Photodiodes," *J. Light. Technol.*, vol. 25, no. 1, pp. 109–121, Jan. 2007.

[11] S. Pellegrini, "InGaAs/InP Single-Photon Avalanche Diodes," Heriot-Watt University, 2005.

[12] Burle Industries Inc., *Photomultiplier Handbook*. pp. 8–80, 1980.

[13] R. Foord, R. Jones, C. J. Oliver, and E. R. Pike, "The Use of Photomultiplier Tubes for Photon Counting," *Appl. Opt.*, vol. 8, p. 1975, 1969.

[14] W. Becker, *Advanced Time-Correlated Single Photon Counting Techniques*. p. 389, 2005.

[15] H. Kume, K. Koyama, K. Nakatsugawa, S. Suzuki, and D. Fatlowitz, "Ultrafast microchannel plate photomultipliers," *Appl. Opt.*, vol. 27, pp. 1170–1178, 1988.

[16] S. Cova, M. Ghioni, A. Lotito, I. Rech, and F. Zappa, "Evolution and prospects for single-photon avalanche diodes and quenching circuits," *J. Mod. Opt.*, vol. 51, no. 9–10, pp. 1267–1288, 2004.

[17] A. Goetzberger, B. McDonald, R. H. Haitz, and R. M. Scarlett, "Avalanche Effects in Silicon p-n Junctions. II. Structurally Perfect Junctions," *J. Appl. Phys.*, vol. 34, no. 6, pp. 1591–1600, 1963.

[18] R. H. Haitz, "Mechanisms Contributing to the Noise Pulse Rate of Avalanche Diodes," *J. Appl. Phys.*, vol. 36, no. 10, pp. 3123–3131, 1965.

[19] R. J. McIntyre, "The distribution of gains in uniformly multiplying avalanche photodiodes: Theory," *IEEE Trans. Electron Devices*, vol. 19, no. 6, pp. 703–713, 1972.

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

[21] P. P. Webb, R. J. McIntyre, and J. Conradi, "Properties of Avalanche Photodiodes," vol. 35, pp. 234–278, 1974.

[22] S. Cova, A. Longoni, and A. Andreoni, "Towards picosecond resolution with single-photon avalanche diodes," *Rev. Sci. Instrum.*, vol. 52, no. 3, pp. 408–412, 1981.

[23] W. G. Oldham, R. R. Samuelson, and P. Antognetti, "Triggering phenomena in avalanche diodes," *IEEE Trans. Electron Devices*, vol. 19, no. 9, pp. 1056–1060, 1972.

[24] A. Lacaita, P. A. Francese, F. Zappa, and S. Cova, "Single-photon detection beyond 1 mm: performance of commercially available germanium photodiodes," *Appl. Opt.*, vol. 33, no. 30, pp. 6902–6918, 1994.

[25] P. A. Hiskett, G. S. Buller, A. Y. Loudon, J. M. Smith, I. Gontijo, A. C. Walker, P. D. Townsend, and M. J. Robertson, "Performance and Design of InGaAs /InP Photodiodes for Single-Photon Counting at 1.55 μm," *Appl. Opt.*, vol. 39, no. 36, pp. 6818–6829, 2000.

[26] S. Cova, M. Ghioni, A. Lacaita, C. Samori, and F. Zappa, “Avalanche photodiodes and quenching circuits for single-photon detection,” *Appl. Opt.*, vol. 35, no. 12, pp. 1956–1976, 1996.
[27] S. Pellegrini, R. E. Warburton, L. J. . Tan, Jo Shien Ng, A. B. Krysa, K. Groom, J. P. . David, S. Cova, M. J. Robertson, and G. S. Buller, “Design and performance of an InGaAs-InP single-photon avalanche diode detector,” *IEEE J. Quantum Electron.*, vol. 42, no. 4, pp. 397– 403, 2006.
[28] A. Spinelli and A. L. Lacaita, “Physics and numerical simulation of single photon avalanche diodes,” *IEEE Trans. Electron Devices*, vol. 44, no. 11, pp. 1931–1943, 1997.
[29] S. Cova, A. Longoni, and A. Andreoni, “Towards picosecond resolution with single-photon avalanche diodes,” *Rev. Sci. Instrum.*, vol. 52, no. 3, pp. 408–412, 1981.

Chapter 3 - Ge-on-Si Technology and Applications: Material selection, Epitaxial Growth and Review of Detectors

3.1 Introduction

The integration of optoelectronic devices on Si substrates has been an active research field from many years, as silicon offers a mature technology platform for the integration of low-loss optical waveguides with readout circuitry. Although Si acts as an excellent material for photodiode operation, efficient detection is generally confined to wavelengths below 1000 nm. As shown in Figure 3.1, for operation in the low loss fibre optical windows at around 1300 nm (second window) and 1550nm (third window, the "C" and "L" band), other semiconductors must be utilised.

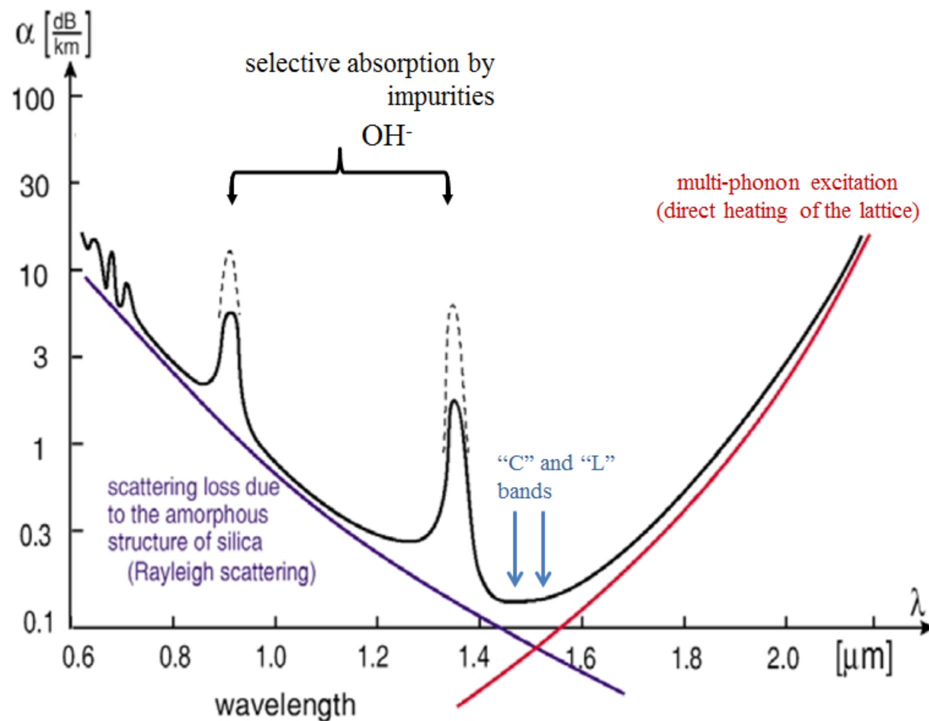


Figure 3.1. The spectral attenuation of telecommunications fibre [1].

Germanium (Ge) has excellent potential as a material for optical detection due to its possible integration with silicon (Si) technology and good optical absorption properties at the near-infrared wavelengths. In this chapter, a literature review of Ge-on-Si technology will be presented starting from the material selection to the epitaxial growth techniques. This will be followed by a justification of the selection of pure Ge as the material for near-infrared devices on Si. Finally, a review of different Ge-on-Si detectors technology will be presented with a main focus on avalanche photodiodes and their use as single-photon detectors.

3.2 Choice of material for near-infrared photodetectors on Si

As already discussed in Chapter 2, another parameter that influences the performance of a photodiode in terms of responsivity is the absorption coefficient (α). The optical radiation absorbed in a semiconductor material can be written as follows:

$$P(x) = P_0\left(1 - e^{-\alpha(\lambda)x}\right) \tag{3.1}$$

Here, $\alpha(\lambda)$ is the absorption coefficient at a wavelength λ, P_0 is the incident optical power and, $P(x)$ is the optical power absorbed at a distance x into the material. The expression for the photocurrent I_{photo}, resulting from optical absorption within the depletion region, is given by the expression 2.7. By using this expression, the responsivity can be rewritten as follows:

$$R = \frac{I_{photo}}{P_{optical}} = \frac{q}{h\nu}\left(1 - e^{-\alpha(\lambda)x}\right) \tag{3.2}$$

Here, q is the electronic charge, h is Planck's constant and ν is the frequency of incident radiation. In the above equation, the relationship between the responsivity and absorption coefficient clearly shows that the responsivity is a function of the wavelength and of the photodiode material (since different materials have different band-gap energies and then different absorption coefficients).

The dependence of the optical absorption coefficient on wavelength for different semiconductor materials is shown in Figure 3.2.

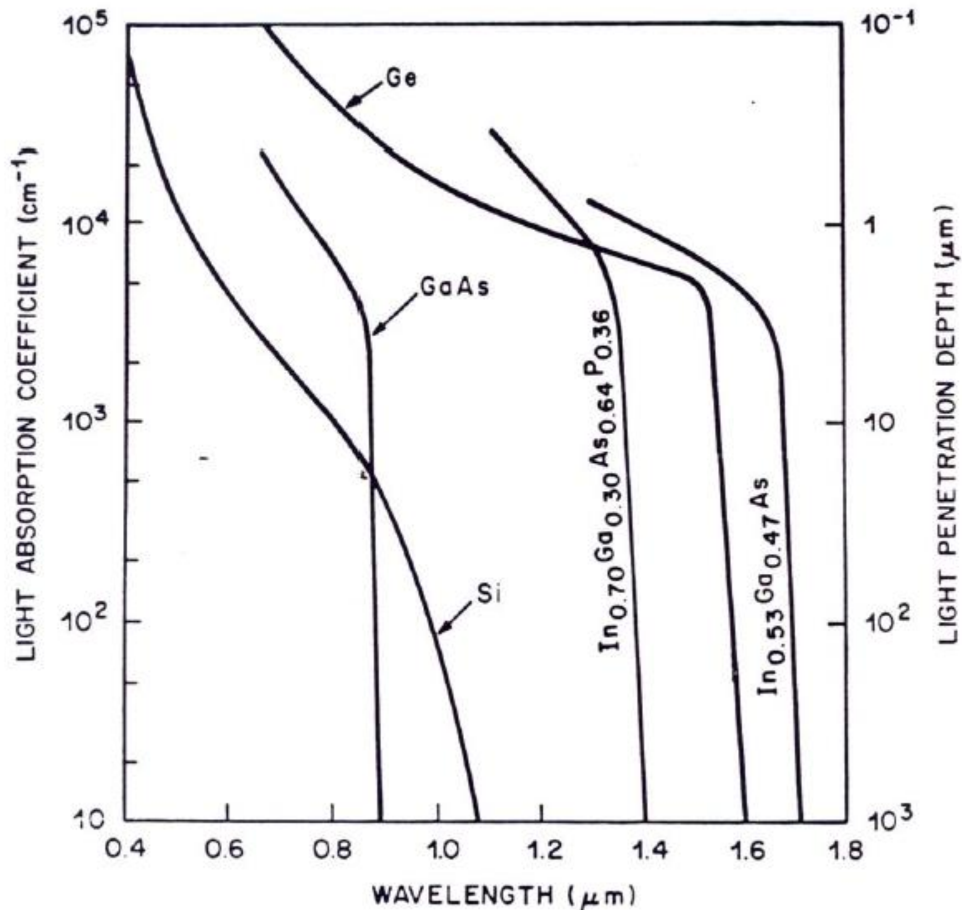


Figure 3.2. Wavelength dependence of the absorption coefficient for several semiconductor materials [1].

As shown in Figure 3.2, semiconductor materials have a cut-off wavelength that depends on their energy gap. For wavelengths longer than the cut-off wavelength, the photon energy is insufficient to excite an electron from the valence band to the conduction band. Conversely, at the lower-wavelength end, the photo-current can be reduced since the high absorption means that photons are absorbed very close to the photodetector surface where the recombination time of the EHPs can be very short.

Figure 3.2 also makes clear that depending on the wavelength at which the photodetector has to operate it is necessary to choose the correct semiconductor material.

For operation in the second window (at wavelengths around 1310 nm) and third window, "C" band (1530 - 1565 nm) and "L" band (1565 - 1625 nm), the choice of the absorbing material is between $In_{0.53}Ga_{0.47}As$ and Ge because of their high absorption coefficients (~10^4 cm^{-1}) at these wavelengths. Although $In_{0.53}Ga_{0.47}As$ and Ge are

obvious candidates as absorbers in high efficiency near-infrared photodetectors, their integration with Si technology needs to be carefully evaluated.

The lattice constant '*a*' of Ge (*a*=0.565 nm) and $In_{0.53}Ga_{0.47}As$ (*a*=0.587 nm) [2] are both very different from that of Si (a=0.543 nm). This is also shown in Figure 3.3.

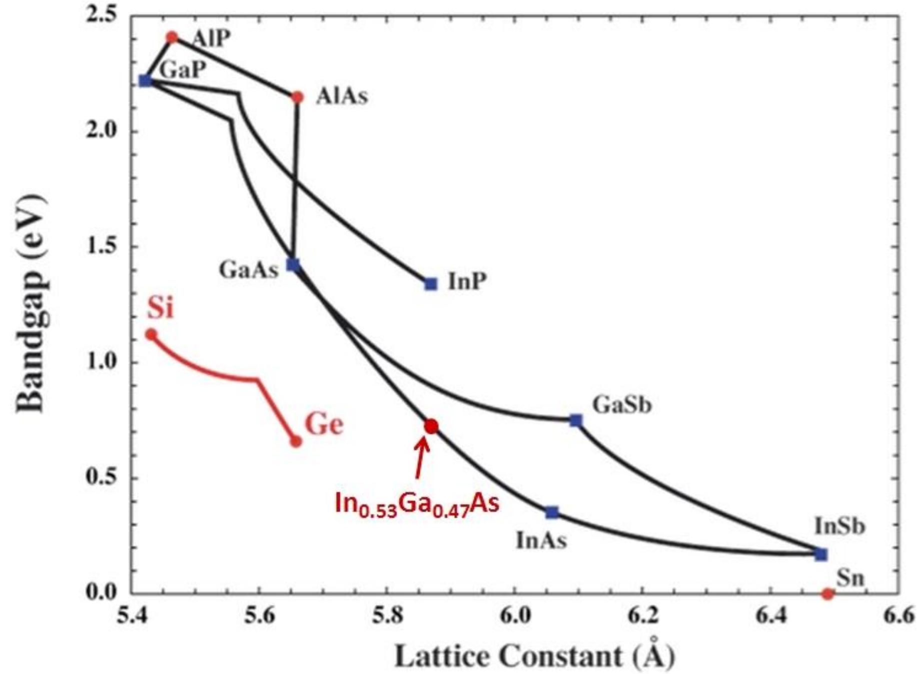


Figure 3.3. Bandgap versus lattice constant for different semiconductor materials.

Although $In_{0.53}Ga_{0.47}As$ and Ge photodiodes are typically made on lattice matched substrates such as InP and Ge, respectively, as explained in Chapter 1 there are a number of advantages if near-infrared photodiodes can be integrated on Si.

The lattice mismatches between Si and $In_{0.53}Ga_{0.47}As$, and Ge, are of the order of ~8% and ~4.2%, respectively, and create two main problems [3], [4], [5]:

1. High surface roughness due to island growth;
2. The introduction of high densities of misfit-dislocations (MD) and threading dislocations (TD) in the epilayers.

Regarding the first point, it is well-known that Ge on Si epitaxy shows a Stranski-Krastanov growth mode, or layer-plus-island growth as shown in Figure 3.4. Initially, Ge grows in a layer-by-layer style, but beyond a few monolayers, Ge continues to grow through the nucleation of germanium "islands" because of the large lattice mismatch. Such growth results in high surface roughness of the Ge epitaxial film, which is not desirable. High surface roughness degrades heterojunctions and causes difficulties in process integration.

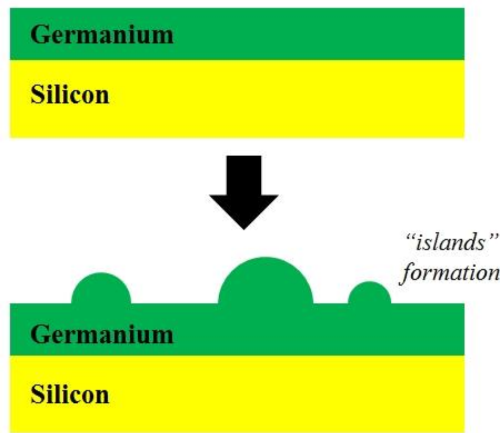


Figure 3.4. Schematic of the Stranski-Krastanov Ge-on-Si heteroepitaxy growth mode; Ge film shows 2-D growth for the first few monolayers, then starts forming islands making a 3-D surface.

High densities of threading dislocation can degrade device performance and compromise device reliability [6]. Misfit-dislocations are caused by the relaxation of the epilayer due to the lattice mismatch with the Si substrate and there are several sources of misfit dislocation formation. Dislocations from missing, or dangling, bonds in the lattice between the two layers (substrate and epilayer), or substrate inhomogeneities such as impurities, dust particles, residual oxide, or mechanical damage are examples. In 1970, Matthews first explained how a TD can extend into a MD [7]. If the elastic energy released by TD glide is larger than the energy to create MD at the interface, TD segment will glide laterally, creating MD at the interface. This is schematically shown in Figure 3.5.

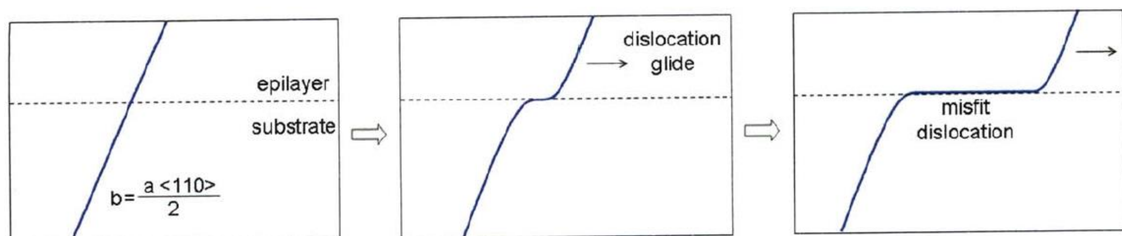


Figure 3.5. Schematic diagram showing generation of a misfit dislocation from a threading dislocation [7].

Once long lengths of MD are formed, they interact with each other, adding more MDs as a result. They are typically confined between the epilayers and the Si substrate. For every misfit dislocation there will always be two threading dislocations at the ends of the misfit. Dislocations have to either form a loop or terminate at a free surface. The epilayer surface is always the nearest free surface, and these threading dislocations typically thread to this surface. Since devices are usually built near the surface, these

threading dislocations cannot be avoided and may affect the performance of devices reducing carrier lifetime, carrier mobility [8] and compromising device reliability.

3.3 Lattice mismatched near-infrared photodetectors on Si

Silicon is the most mature microelectronic technology for integrated circuits due to a number of properties of the material and, more importantly, thanks to the advantage of having two compatible good insulators, silicon dioxide (SiO_2) and silicon nitride (Si_3N_4). To detect light in the near-infrared regime, near-infrared absorbing materials, for example Ge and $In_{0.53}Ga_{0.47}As$, need to be used, as shown in Figure 3.2. From the previous section, it is clear that the lattice mismatch between Si and $In_{0.53}Ga_{0.47}As$, or Ge, causes different problems that compromise the device performance and reliability. To overcome these problems, and integrate high-performance $In_{0.53}Ga_{0.47}As$ or Ge near-infrared photodetectors on Si, different solutions have been proposed during the last 30 years in order to solve the problems related to threading-dislocations and surface roughness. In this section a literature review of the work carried out in this field is presented, followed by a justification for the selection of pure Ge as the material for near-infrared photodetectors on Si.

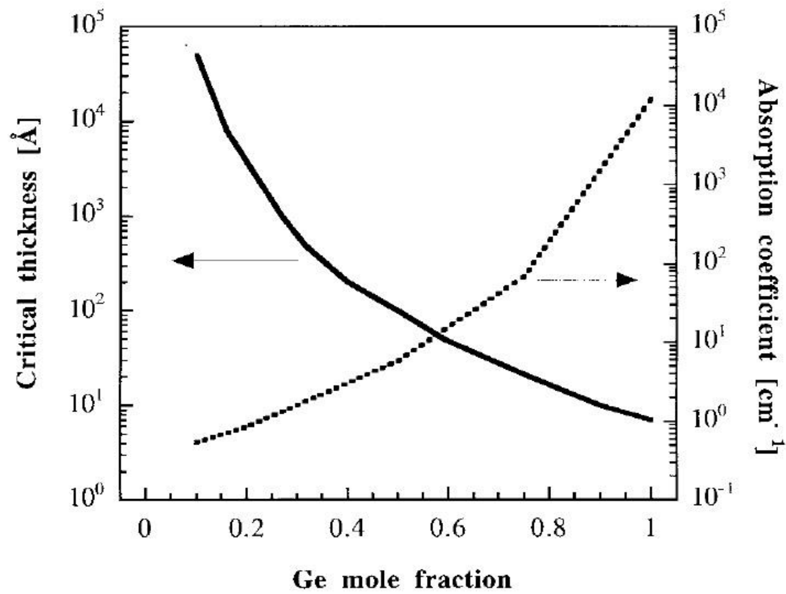
3.3.1 $Si_{1-x}Ge_x$ strained layer

SiGe was initially proposed for applications in microelectronics to overcome the problems of low mobility and low velocity saturation in Si homojunctions that allowed other III-V semiconductors (e.g. GaAs, InP) to dominate a number of areas in analogue electronics [9].

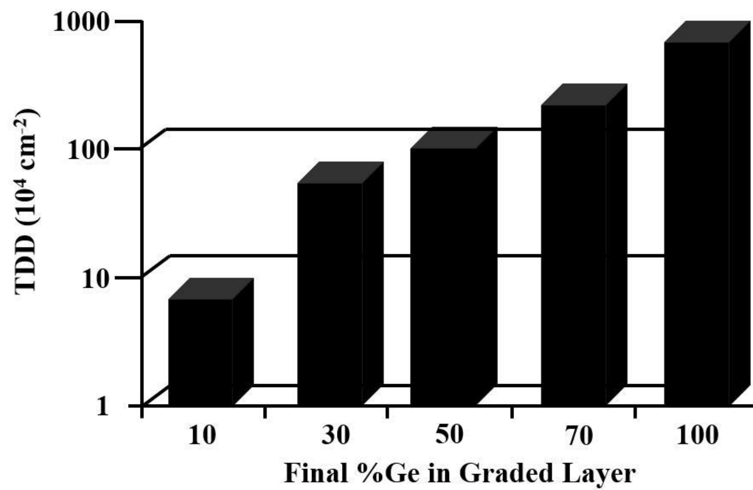
The first mention of SiGe devices was in 1957 [10], where the idea of a SiGe based heterojunction bipolar transistor (HBT) was discussed. Although the theory was described, the first epitaxial growth of SiGe heterostructures was not demonstrated until 1975 by Erich Kasper and colleagues at the AEG Research Centre in Germany using molecular beam epitaxy (MBE) [11]. With the advent of the SiGe HBT and first sale in 1998, the market of SiGe devices for radio-frequency (RF) application started to increase at a rate of 30% per annum [12]. The real strength of SiGe technology lies in its ability to integrate analogue, RF and digital electronics on a single chip using existing CMOS fabrication facilities. This is not possible with any other technologies (e.g. GaAs).

Although SiGe technology has found its success in the microelectronic area, its use in photonics applications can be dated to 1984. The first Ge photodetector for near-infrared application grown directly on Si, by molecular beam epitaxy (MBE), was reported by Luryi *et al.* [13]. To minimize the effect of dislocations, the active detector structure was moved away from the Ge-Si interface by inserting a highly doped germanium buffer layer of the same conductivity type as that in the underlying silicon substrate. However, a high density of threading dislocation (~10^9 cm^{-2}) was measured by transmission electron microscopy (TEM) at ~1 μm into the germanium layer. These p-i-n detectors demonstrated an external quantum efficiency of 40 % at a wavelength of 1450 nm in photovoltaic mode (no external bias applied). The dark current density was 50 mA/cm^2. The authors also stated that the high leakage current in reverse bias in their device was related to the quality of germanium.

The high density of threading dislocation measured by Luryi *et al.*, was probably due to the thickness of the SiGe epilayer. As the thickness of the epitaxial layer is increased, there exists a maximum thickness, called the critical thickness h_c, above which defects appear, in this case misfit dislocations, which act to relieve the strain in the epitaxial film. Different theories have been developed to predict the critical thickness of the strained epitaxial layer based on energy considerations [14], [15]. This thickness will first of all depend on the lattice misfit, f_m, equal to $((a_s-a_f)/a_f)$, with a_s and a_f the stress-free substrate and epilayer lattice constant respectively, where the substrate thickness is assumed to be much larger than the film thickness [5]. In particular, positive values correspond with tensile stress in the film ($a_f < a_s$), for example, $Si_{1-x}Ge_x$ on a Ge substrate, while negative values lead to compressive stress ($a_f > a_s$) in the case of $Si_{1-x}Ge_x$ on a Si(001) substrate. In addition, the maximum thickness h_c of the SiGe alloy layer depends on the Ge fraction x, decreasing with increasing Ge content, as shown in Figure 3.6a [16]. It is clear that the thickness required for 100 % absorption at even the shorter telecommunications wavelength of 1300 nm with pure Ge ($\alpha = 10^4$ cm^{-1}) would be well beyond the corresponding critical value. At the same time, the density of threading dislocations (TDD) increases with the Ge content in the top layer, as is illustrated in Figure 3.6b [17].



(a)



(b)

Figure 3.6. (a) Critical thickness of a SiGe strained layer grown on silicon as a function of the Ge mole fraction in the alloy (left axes). Corresponding absorption coefficient α at 1320 nm (right axis) [16]. *(b) Plot of the average threading dislocation density observed in the top uniform cap layer grown on relaxed SiGe layers graded to different final Ge concentration* [5].

After demonstrating the integration of a Ge photodetector directly on Si in 1984, Luryi *et al.* proposed the use of strained-layer $Si_{1-x}Ge_x$/Si superlattice (SLS) waveguide structure for an infrared photodetector in 1986 [18]. Their proposal was based on the theoretical bandgap calculation by People [19], whereby the bandgap narrowing in a strained Ge_xSi_{1-x} alloy grown on a Si (100) substrate is substantially reduced in comparison with the unstrained alloy. This reduction in the bandgap was also demonstrated by Lang *et al.* [20]. However, the Ge_xSi_{1-x} alloy remains an indirect-gap

semiconductor at all values of x and the absorption coefficient is low, $\alpha < 10^2$ cm^{-1}, at the near-infrared wavelengths. Because of this, they proposed the use of a waveguide geometry to increase the interaction length and overcome the low absorption coefficient at wavelengths of interest for fibre-optic communications.

Although, Luryi *et al.* demonstrated the first infrared photodetector made by using a SiGe graded buffer layer, pioneering MBE work by Baribeau *et al.* [21]–[23] demonstrated that the defect density could be significantly reduced by using a graded SiGe buffer layer. They verified that the density of misfit dislocations increases with epilayer thickness and Ge concentration. The TDD was measured using plan-view TEM and it was estimated to be ~5×10^8 cm^{-2}. SLS structures were also grown, but they resulted in lower quality films without any defect improvement in the Ge epilayer. They also investigated the direct deposition of pure Ge on Si (001) for different growth temperatures, 285 °C, 610 °C and 700 °C, respectively. The best results were obtained for the intermediate temperature, and a TDD of ~2×10^7 cm^{-2} was measured by using scanning electron microscopy (SEM) and etch-pit density (EPD). They also studied post-growth annealing treatments verifying a significant improvement in the crystalline quality of all Ge epilayers. Annealing was performed for 30 minutes at 700 °C and a reduction of the TDD by almost one order of magnitude was observed. They explained this reduction on the basis of the thermal mismatch between the two materials, which is a function of temperature. Moreover, a two temperature growth process was proposed, where the Ge is initially grown at low temperature, to permit a smooth surface morphology and then the temperature was raised to favour a reduction in dislocation density in the epilayer.

Due to these advantages, strained layer $Si_{1-x}Ge_x$ MQW SLS detectors [24]–[27], [28], waveguide p-i-n [29], [29]–[32], waveguide avalanche photodetectors [33]–[35], and single-photon avalanche detectors [36] were demonstrated.

A new method for reducing the threading dislocation in Ge on Si structures using a SiGe buffer layer was proposed by Currie *et al.* in 1998 [37]. In this method, a chemical-mechanical polishing (CMP) step was introduced during the growth of the SiGe graded layer at a fraction of 50 %. Then a regrowth step was performed with an increasing fraction of Ge. The addition of this step stopped the increase in threading dislocations and allows dislocations to relieve the strain introduced in the subsequent growth. A reduction in TDD of one order of magnitude from 10^7 to 10^6 cm^{-2} was

demonstrated. This method was first used by Samavedam *et al.* to make p-n mesa Ge on Si diodes using graded SiGe layers [38]. Devices were fabricated using an ultrahigh vacuum chemical vapour deposition (UHVCVD) reactor. A responsivity of 0.133 A/W was measured at 0 V, with an external quantum efficiency of 12.6 % at a wavelength of 1300 nm. A low dark current density of 0.15 mA/cm^2 at an applied reverse bias of -1 V was calculated. Moreover, the author stated that the leakage current was mainly due to a bulk effect within the device rather that from surface or edge effects, mainly because of the small band gap of Ge. Later in 2001, Giovane *et al.* also applied this method to fabricate n-i-p $Si_{0.75}Ge_{0.25}$ photodiodes in a UHVCVD reactor [27]. Mesa devices with an area ranging from 0.1 mm^2 to 10 mm^2 were characterized. Moreover, samples with various threading dislocation densities (2×10^5, 2×10^6, and 9×10^6 cm^{-2}) were prepared by growing step graded SiGe buffer layers with three different grading rates (6 %/μm, 10 %/μm, and 17 %/μm). They demonstrated that the defect states related to threading dislocation act as generation and recombination centres in these devices, and the bulk leakage current correlates directly with the TDD. In fact, they measured an increase in bulk leakage current densities of 0.02, 0.32, and 0.78 mA/cm^2 as the TDD increases. Similar results were also previously demonstrated by Ross *et al.* for higher TDDs [39].

Although the reduction in threading dislocations by using a SiGe graded layer was verified by many authors and different devices were demonstrated, to increase the absorption at the infrared wavelength the concentration of Ge needs to be increased. However, the maximum thickness of the Ge_xSi_{1-x} layers decreases with an increasing Ge fraction x. As pointed out by Temkin *et al.* for a Ge concentration of x=0.6 the absorption coefficient is equal to 21 cm^{-1} at a wavelength of 1300 nm, and decreases to 2.5 cm^{-1} at a wavelength of 1550 nm [34]. Detectors made from strained layer $Ge_{1-x}Si_x$/Si superlattices require very long absorption lengths to provide reasonable detector responsivities.

In terms of single-photon avalanche detectors (SPADs), pioneering work on this device was done by Loudon *et al.* in 2002. A separate absorption and multiplication (SAM) region device structure was employed in the design of this $Si/Si_{1-x}Ge_x$ MQW SPAD. 25 period superlattice layers with a Ge concentration equal to 0.3 were grown for the absorption region. The multiplication region consists of 700 nm thick intrinsic Si grown on top of a highly doped Si substrate. Because of the low Ge concentration, the maximum absorption wavelength was limited to ~1200 nm, and single-photon counting

was performed at wavelengths of 826 and 1210 nm. The performance of this device was compared with an all-Si control sample, the structure of which was identical except that for the absorbing region which consisted only of Si without the MQW. Low detection efficiency equal to 0.013 % was measured at a wavelength of 1210 nm (at 3.6 % of excess bias and 300 K). At the same wavelength the all-Si control sample showed a detection efficiency 30 times lower. A noise equivalent power (NEP) of 5×10^{-12} $WHz^{-1/2}$ at 1210 nm was obtained for the MQW structure, while it was ~1×10^{-11} $WHz^{-1/2}$ for the all-Si control sample. Despite the low Ge concentration, the authors demonstrated, for the first time, an enhancement of infrared photon counting performance in SPADs incorporating Si multiplication layers.

Table I summarises the performance of various infrared detectors based on strained layer SiGe MQW SLS structures.

Structure	Efficiency η_{ext}	Responsivity [A/W]	Dark Current Density [mA/cm^2]	Wavelength [nm]	Reference
p-i-n pure Ge	40 %	NA	50	1450	[13]
SiGe MQW detectors	59 %	0.003 @ 0 V 0.62 @ - 7 V	NA	1300	[24]
SiGe p-i-n MQW	1 %	0.01	3	1300	[25]
SiGe n-i-p MQW	NA	0.5	NA	1100 – 1600	[26]
SiGe p-n diodes	12.6 %	0.133	0.15	1300	[38]
SiGe n-i-p	NA	NA	0.02	NA	[27]
SiGe p-n diodes	NA	0.37 @ 0 V 0.57 @ -2 V	13	1300	[28]
SiGe MQW waveguide	10 %	1.1	NA	1300	[33]
SiGe p-i-n MQW waveguide	12 %	0.43	0.1	1100	[30]
SiGe p-i-n MQW waveguide	7 %	NA	2.7	1320	[29]
SiGe n-i-p MQW waveguide	12 %	NA	100	1282	[31]
SiGe n-i-p MQW waveguide	18.2 %	0.08	NA	1550	[32]

Table I. Summary of SiGe/Si photodetectors in the near-infrared.

Although strained-layer $Si_{1-x}Ge_x$ SLS have been employed by many authors to overcome the problem related to the lattice mismatch between Si and Ge, their main disadvantage is the low absorption coefficients. To solve this problem, the Ge concentration must be increased, but there are two main limitations:

1. the critical thickness decreases as the Ge content is increased (see Figure 3.6a) and it is too small for efficient optical detection;
2. TDD increases as the Ge content is increased (see Figure 3.6c) due to the increased lattice mismatch.

By using waveguide structures, it is possible to improve the overall absorptance at wavelengths around 1300 nm, but the photoresponse at wavelengths near 1500nm still remains very low.

3.3.2 SiGeC structures

To overcome the aforementioned limitation of the strained-layer SiGe SLS structures different authors have tried to use the SiGeC ternary system, and its optical and electronic properties have been studied [40]. The basic idea is to introduce a small lattice constant element (such as C) into the SiGe alloy to compensate for the large lattice constant of Ge as compared with Si. The growth of this ternary system was demonstrated using MBE [41], [42] or CVD [43], [44] growth. However, there are different problems related to the growth of this ternary system:

a) the high mismatch between the C and Si lattice;
b) low solubility of C in Si;
c) and silicon carbide (SiC) precipitation, which was found to negatively affect the electrical properties of the structure.

The third problem can be avoided using low growth temperature and a lower C concentration which, however, increases the lattice mismatch between Ge and Si. Therefore, a tradeoff between these parameters should be found. Amour *et al.* demonstrated that adding C to the SiGe alloy increased the critical thickness, and also reduced the strain between the SiGe and Si layers. However, they observed a slight increase in the band gap as C was added, but they stated that this bandgap increase was much less than it would be if the strain was reduced simply by reducing the Ge concentration in SiGe without adding C [45]. This bandgap widening was experimentally verified by Soref *et al.* [46] who grew different waveguide structures using the ternary system $Si_{1-x-y}Ge_xC_y$ as a waveguide layer on top of a Si substrate. These structures were identical except for the C concentration. Their findings are summarised in Table II.

C concentration	Bandgap E_g $Si_{1-x-y}Ge_xC_y$
0 %	1.13 eV
2 %	0.96 eV
4 %	0.87 eV
9 %	1.02 eV
11 %	1.15 eV

Table II. Bandgap, E_g, of the ternary system $Si_{1-x-y}Ge_xC_y$ as a function of the C concentration [46].

As shown from the bandgap value reported in Table II, as the C concentration is increased from 0 % to 4 % the bandgap reduces, but as the C concentration is increased above 9 % the bandgap starts to increase. This implies that some matched SiGeC compositions are transparent at both 1300 and 1550 nm, while other matched compositions are transparent at 1550 nm only.

Only rarely have SiGeC photodiodes been reported in the scientific literature. Normal incidence n-i-p photodiodes have been demonstrated by Huang *et al.* [47]. 80 nm thick $Si_{0.25}Ge_{0.6}C_{0.15}$ was used as absorption layer. The dark current density measured at a reverse bias of 0.5 V was 7 mA/cm^2, while the external quantum efficiency was limited to 1 % at a 1300 nm wavelength. The authors stated that the low photoresponse is mainly due to the indirect bandgap of the SiGeC alloy and its small thickness. They proposed a waveguide geometry that was later demonstrated [48]. The photodetector was integrated on a waveguide400 μm long. The dark current density at a reverse bias of 0.3 V was 4 mA/cm^2, while the external quantum efficiency was 8 % and 0.2 % at 1300 nm and 1550 nm wavelengths, respectively. They effectively showed an improvement of the external quantum efficient compared to the normal incidence structure, but the absorption at 1550 nm was very low (the absorption coefficient was 10 cm^{-1} at 1550 nm).

Table III summarised these results.

Structure	Efficiency ηext	Responsivity	Dark Current Density [mA/cm^2]	Wavelength [nm]	Reference
SiGeC normal incidence nip	1 %	NA	7	1300	[47]
SiGeC nip waveguide	8 % @ 1300 nm 0.2 % @ 1550nm	NA	4	1300 – 1500	[48]

Table III. Summary of SiGeC/Si photodetectors in the near-infrared.

The ternary system SiGeC was introduced to alleviate the lattice mismatch between Ge and Si, and it has been proved to show a reduction in threading dislocation density compared to the SiGe system [49]. However, this reduction in TDD did not help to reduce the dark current density and SiGeC based photodiodes provide very little improvement than the SiGe approach. Several disadvantages are present due to the low C concentration: its indirect bandgap, its contamination problem, and its low absorption coefficient have limited its use to make near-infrared photodetectors.

3.3.3 InGaAs structures

Another approach to combine near-infrared photodetectors with Si is the integration of InGaAs on Si. The main challenge is represented by the 8 % lattice mismatch between these two materials. However, the absorption coefficient of $In_{0.53}Ga_{0.47}As$ at the infrared wavelengths (1300 and 1550 nm) is ideal to make near-infrared detectors, as shown in Figure 3.2. Despite this large lattice mismatch, different techniques, like epitaxial growth and wafer bonding, have been successfully demonstrated during the last 30 years to make high efficiency InGaAs detectors integrated on Si.

In terms of epitaxial growth, different devices were demonstrated [50], [51]. For example, Gao *et al.* demonstrated a normal-incidence p-i-n InGaAs photodetector integrated on Si [51]. Devices showed a dark current density of 64 mA/cm^2 at a reverse bias of 1 V, with a responsivity of 0.57 A/W at 1550 nm wavelength.

Although the direct epitaxial growth is a more straightforward solution for the monolithic integration of III-V devices with silicon microelectronics for medium/large scale integration, significant improvements in the integration of InGaAs on Si have been

obtained using the wafer bonding or fusion techniques. This technique was first demonstrated in 1990 by Liau and Mull and consisted of fusing together centimetre-size InP and GaAs wafers [53]. A reactor, which can press the wafers together at elevated temperatures (~800 °C), was designed and diodes were also fabricated and characterized. Following this, InGaAs/Si p-i-n [54]–[59], InGaAs/SOI waveguide p-i-n [60], [61], InGaAs/Si APD and SPAD [54], [62], [63] have been fabricated showing good performance in terms of low leakage current density and high responsivity at the infrared wavelengths.

Most of these InGaAs/Si structures are firstly grown by epitaxial growth on top of a InP substrate to which InGaAs is lattice matched. This initial epitaxial step is performed using different reactors, like metal-organic CVD (MOCVD) or MBE, then the bonding is obtained by placing the Si wafer and InGaAs/InP structure in direct contact under pressure at a temperature, between 600 and 700 °C in a H_2 atmosphere. After the fusion, the InP substrate is selectively removed from the InGaAs epitaxial layer and devices are made using standard photolithography techniques.

In [54] a wafer bonded InGaAs/Si p-i-n mesa-geometry detector was developed. The wafer bonding process was employed to join the InP wafer (on which the InGaAs p-i-n epilayers were grown) and the Si wafer. Because of the 7.7 % lattice mismatch between the InP and the Si wafers, misfit dislocations were confined at the bonding interface to relax the strain between the two wafers. A low dark current density of 0.0057 mA/cm^2 at a reverse bias of 5 V was measured, with a responsivity of 1 A/W (corresponding to an external quantum efficiency of ~80 %) at a wavelength of 1550nm. Performance of these devices was slightly improved by Levine *et al.* who demonstrated, in two different papers [55], [56], InGaAs p-i-n mesa devices with dark current densities of 0.0025 and 0.00014 mA/cm^2 respectively, at a reverse bias of 4 V,. However, no data on responsivity at the infrared wavelengths were reported. It is worthwhile mentioning the InGaAs/Si p-i-n planar-geometry devices reported by Pauchard *et al.* that demonstrated a dark current density of 0.012 mA/cm^2 with a responsivity of 0.85 and 0.8 A/W at 1310 and 1550 nm wavelengths, respectively [59]. These devices also demonstrated a fabrication yield in excess of 90 %.

An important step forward on the integration of III-V semiconductors with SOI wafers was obtained at the Interuniversity Microelectronics Centre (IMEC) in Belgium. They developed a new bonding technique, to fuse together small pieces of unprocessed

InGaAs/InP optoelectronics dies with the processed SOI waveguide circuitry. The proposed process is illustrated in Figure 3.7.

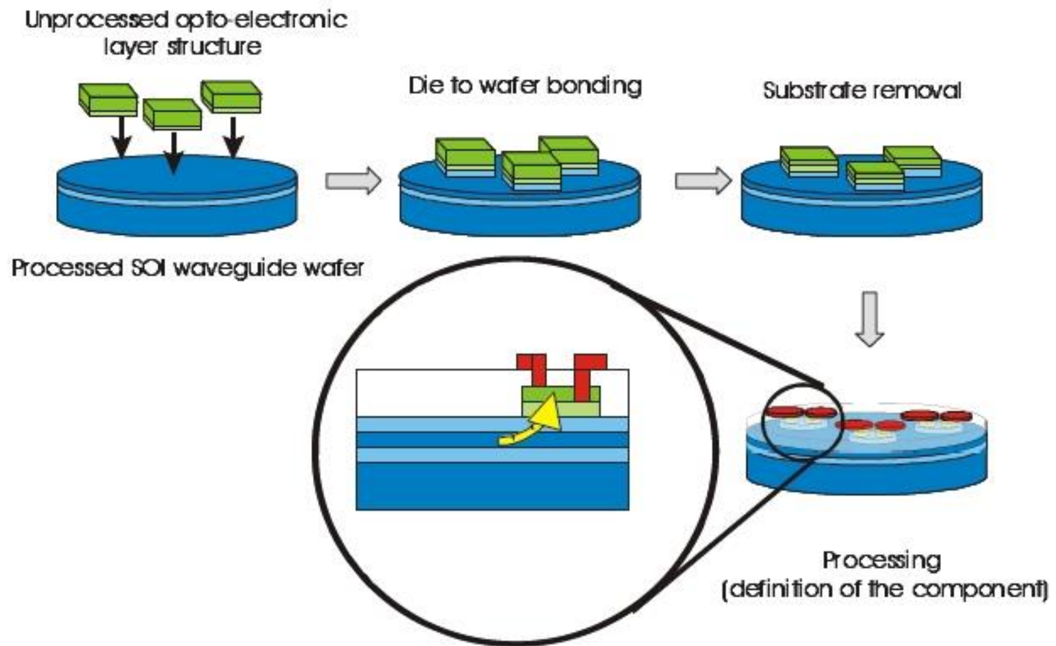


Figure 3.7. Proposed process flow for the integration of III-V semiconductors and SOI waveguide circuitry [60].

Integration of InGaAsP photodetectors on SOI was demonstrated [60]. The measured dark current density was 0.3 mA/cm^2 at a reverse bias of 1 V, with a responsivity of 0.02 A/W at 1550 nm wavelength. This performance was later improved in [61], and a very low dark current of 10 pA (corresponding to 0.00833 mA/cm^2) at a reverse bias of 0.5 V, with a responsivity of 1.1 A/W (corresponding to an external quantum efficiency of 88 %) were measured.

By using the wafer bonding technique different InGaAs/Si avalanche photodiodes (APD) were also demonstrated [62]–[64]. All of these structures were based on the Separate Absorption and Multiplication (SAM) configuration, where an InGaAs layer was used as the absorption layer (because of its absorbance at the infrared wavelengths) and the Si layer was used for the multiplication layer (due to the excellent avalanche properties of Si). Some of these APDs have been also characterized as SPADs, but their performance will be evaluated in the later sections of this chapter after introducing the concept of single-photon detectors.

Table IV summarises the photodetector performance obtained by integrating InGaAs on Si.

Structure	Efficiency η_{ext}	Responsivity [A/W]	Dark Current Density [mA/cm^2]	Wavelength [nm]	Reference
InGaAs/Si normal-incidence mesa p-i-n (epitaxially growth)	NA	0.57	64	1550	[51]
InGaAs/Si mesa p-i-n (wafer bonded)	80 %	1	0.0057	1550	[54]
InGaAs/Si mesa p-i-n (wafer bonded)	NA	NA	0.0025	NA	[55]
InGaAs/Si mesa p-i-n (wafer bonded)	NA	NA	0.00014	NA	[56]
InGaAs/Si mesa p-i-n (wafer bonded)	NA	0.3	4.8	1320	[57]
InGaAs/Si planar p-i-n (wafer bonded)	NA	0.54	0.5	1310	[58]
InGaAs/Si planar p-i-n (wafer bonded)	NA	0.85@ 1310 nm 0.8 @ 1550 nm	0.012	1310 - 1550	[59]
InGaAs/SOI waveguide p-i-n (die to wafer bonded)	NA	0.02	0.3	1550	[60]
InGaAs/SOI waveguide p-i-n (die to wafer bonded)	NA	1.1	NA	1550	[61]
InGaAs/Si SAM APD (wafer bonded)	NA	0.64	0.04	1310	[62]

Table IV. Summary of InGaAs/Si photodetectors in the near-infrared.

The integration of InGaAs on Si has demonstrated photodetectors with better performance, in terms of leakage current density and responsivity, than the SiGe/Si and SiGeC/Si counterparts for detection at near-infrared wavelengths. Hovewer, despite the intrinsic advantages of epitaxial growth, in terms of monolithic integration of III-V devices with silicon microelectronics for medium/large scale integration, the large lattice mismatch and the cross contamination problems (since all the elements in InGaAs are electrically-active dopants in Si) still constrain the performance of these devices. Moreover, the low leakage current density and good responsivity obtained by using the wafer bonding techniques has not been demonstrated with devices grown epitaxially. The wafer bonding approach, however, often suffers from relatively poor mechanical strength and thermal stability due to the weak bonding mechanism. It still

requires the handling of InP substrate that are fragile and expensive compared to Si wafers. This technique does not provide the possibility of selective introduction of InGaAs material on Si in small areas, therefore its applicability in a CMOS foundry is difficult for medium/large scale integration.

3.3.4 Pure Ge

Although SiGe and SiGeC layers have been used to make near-infrared photodetectors, their limited performance at these wavelengths (and especially at 1550 nm) and the high leakage current density have constrained its use in silicon photonics platforms. On the other hand, InGaAs structures have shown good infrared performance and low leakage current, but their integration and contamination problems with Si technology are still an open question. As shown in Figure 3.2, pure Ge, however, has an absorption coefficient that is almost comparable to that of InGaAs at infrared wavelengths.

The Ge band structure is shown in Figure 3.8. Its optical and electronic properties have been studied extensively by different authors [65]–[67].

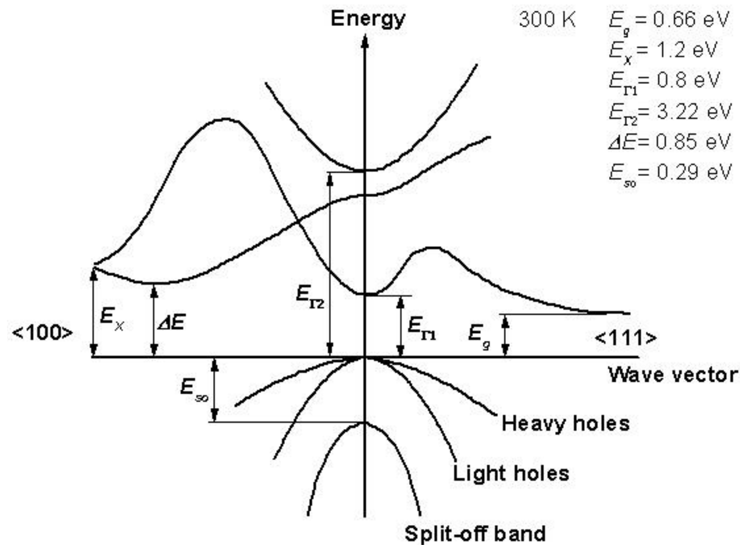


Figure 3.8. Band structure of Ge calculated by Chelikowsky and Cohen [65].

Ge is considered a pseudo-indirect band gap material. Its indirect bandgap E_g is equal to 0.66 eV, but it also has a direct bandgap that is slightly larger and equal to 0.80 eV located at the Γ-valley. Due to this, electrons in the valence band can be easily promoted into the conduction band by absorbing optical power, whose energy is equal or greater than the bandgap energy. This direct transition is a two-particle process, photon and electron-hole, and it is the mechanism that gives direct band-gap

semiconductors materials (e.g. InGaAs) a high absorption coefficient. On the other hand, Si is an indirect bandgap material, and this process is a three-particle process that requires a photon, an electron and a phonon (lattice vibration) for momentum conservation and it is less efficient compared to the direct transition. Due to its direct bandgap, Ge can absorb light efficiently at the infrared wavelengths, as shown in Figure 3.2.

Some of the most important optical, electronic and thermal properties at room temperature (300 K) of Ge are shown in Table V and a comparison with Si and InGaAs is also reported.

	Ge	**Si**	**$In_{0.53}Ga_{0.47}As$**
QUANTITY	**VALUE**	**VALUE**	**VALUE**
Dielectric constant	16.2	11.7	13.9
Lattice constant	5.658 Å	5.431 Å	5.8687 Å
Energy gap (indirect)	0.661 eV	1.12 eV	
Energy gap (direct)	0.8 eV		0.74 eV
Intrinsic carrier concentration	2.4×10^{13} cm^{-3}	1.0×10^{10} cm^{-3}	6.3×10^{11} cm^{-3}
Electron Mobility	3900 cm^{2}V^{-1}s^{-1}	1400 cm^{2}V^{-1}s^{-1}	12000 cm^{2}V^{-1}s^{-1}
Hole Mobility	1900 cm^{2}V^{-1}s^{-1}	450 cm^{2}V^{-1}s^{-1}	300 cm^{2}V^{-1}s^{-1}
Breakdown Field	~ 1×10^5 Vcm^{-1}	~ 3×10^5 Vcm^{-1}	~ 2×10^5 Vcm^{-1}
Refractive index	4	3.42	3.43
Linear Thermal expansion	5.9×10^{-6} K	2.6×10^{-6} K	5.66×10^{-6} K
Melting point	937 °C	1415 °C	

Table V. Summary of the optical, electronic and thermal properties at 300 K of Ge, Si and InGaAs.

Table V summarises the intrinsic properties of these semiconductors. It is important to note the difference, which has been already mentioned in this chapter, of lattice constant between the Ge and Si. A point that it is important to underline is the higher intrinsic carrier concentration (n_i) of Ge than InGaAs, which is an intrinsic property of the material itself.

Considering a p-n junction diode, the leakage current can be divided in two components, diffusion and generation [68], as already discussed in Chapter 2. Both leakage components are directly proportional to the intrinsic carrier concentration, as shown by the following equations:

$$J_{diffusion} = qD_n \frac{n_i^2}{N_A L_n}\left[exp\left(\frac{qV}{kT}\right) - 1\right]$$

(3.3)

$$J_{generation} = q\frac{n_i}{\tau_g} W\left[exp\left(\frac{qV}{2kT}\right) - 1\right]$$

where q is the electron charge, D_n is the diffusion constant of electrons, N_A is the acceptor density, L_n is the diffusion length of electrons, τ_g is the generation lifetime and W is the depletion width. The diffusion component is due to generation in the neutral region and diffusion to the depletion region, while the second component is due to the generation in the depletion region.

At room temperature the intrinsic carrier concentration of Ge is higher than InGaAs (the ratio is ~40). Considering two identical photodiode geometries fabricated from Ge and InGaAs, for low defect density material the leakage current will be mainly dominated by the diffusion component resulting in a leakage current of a Ge photodiode ~1600 times higher than that of an InGaAs photodiode. In the case of a high defect concentration, the generation component will dominate, and assuming an equal depletion width W, the leakage current of a Ge photodiode will be ~40 times higher than an InGaAs photodiode.

However, pure Ge remains the best choice for near-infrared photodetectors on Si for a number of reasons:

a) Ge photodetectors can be integrated with Si, SiO_2 or Si_3N_4 waveguides.
b) Ge does not suffer from the same contamination processes as InGaAs, and its compatibility with the CMOS technology (mainly to make high-speed transistors) has been already demonstrated.
c) Selective area epitaxial growth of pure Ge on Si through patterned SiO_2 has been already demonstrated giving the capability for large scale integration with microelectronics technology, while the selective epitaxial growth of InGaAs is

made difficult due to the large lattice mismatch, whilst wafer bonding techniques are not suitable.

Moreover, it has been shown previously that many authors reported a reduction in leakage current as the density of threading dislocations is decreased. Although this has been demonstrated, the surface passivation and the geometry of a device (mesa or planar) should be taken into account. In small size devices (mesa or planar), the surface leakage component dominates rather than the bulk leakage component [69]. This aspect will be discussed in the next sections and Chapter 5. Therefore, materials with the lowest threading-dislocation densities do not always make the best photodetectors, but surface passivation needs also to be carefully evaluated, and this is especially true for Ge where passivation approaches are still being researched.

The key challenge for the integration of pure Ge on Si is the epitaxial growth of Ge-epilayers on Si with low surface roughness and low threading dislocation densities. These aspects will be evaluated in the next section and a literature review of different Ge-on-Si infrared detectors will be also presented.

3.4 Growth techniques for pure Ge-on-Si

As with InGaAs, pure Ge can be incorporated on Si with growth by epitaxial growth or bonding techniques. Different epitaxial techniques have been used to grow smooth surfaces with low threading dislocation densities of pure Ge on Si. Among these, the most commonly used ones are: Molecular Beam Epitaxy (MBE) and Chemical Vapour Deposition (CVD).

MBE is a versatile ultra-high vacuum technique (UHV - $\sim 10^{-10}$ mbar) for the epitaxial growth of semiconductor, metal and insulator thin films. In MBE, the film crystallises via reactions between thermal-energy molecular beams of the constituent elements and a substrate surface which is maintained at an elevated temperature under UHV conditions. The composition of the grown epilayer and its doping level depend on the relative arrival rates of the constituent elements and dopants, which in turn depend on the evaporation rates of the appropriate sources [70]. MBE, however, suffers from contamination issues (material is deposited on the walls of the growth chamber) compared to CVD which has a low background contamination and is more uniform, mainly due to the higher development budgets and its use to develop production tools.

CVD involves flowing a precursor gas or gases into a chamber containing one or more heated objects to be coated. Chemical reactions occur on and near the hot surfaces, resulting in the deposition of a thin film on the surface. This is accompanied by the production of chemical by-products that are extracted out of the chamber along with unreacted precursor gases [71]. There are many variants of CVD due to the large variety of materials deposited and the wide range of application. In particular, the designed Ge-on-Si structures of this thesis work have been grown by Low-Pressure CVD (LP-CVD). Depending on the pressure (P), CVD can be classified as

- Reduced-Pressure CVD (RP-CVD) for $133\ mbar > P > 1.3\ mbar$
- LP-CVD for $13\ mbar > P > 13 \times 10^{-3}\ mbar$;
- UHV-CVD for $P \sim 1.3 \times 10^{-7}\ mbar$.

There are also a variety of enhanced CVD processes, which involve the use of plasma, ions, photons, lasers to increase deposition rates and/or lower deposition temperatures. CVD has a number of advantages for depositing thin films. The CVD films are generally quite uniform (the film thickness on the sidewalls of features is comparable to the thickness on the top). Another advantage of CVD is that, in addition to the wide variety of materials that can be deposited, they can be deposited with very high purity. Other advantages include relatively high deposition rates, the fact that CVD often does not require high vacuum, and the reproducibility that it is required for high yield production. There are, however, a number of disadvantages. One of the primary disadvantages is that the precursors can be very toxic, explosive, or corrosive and sometimes quite costly. The other major disadvantage is the fact that the films are usually deposited at elevated temperatures leading to restrictions on the kind of substrates that can be coated. More importantly, it leads to stresses in films deposited on materials with different thermal expansion coefficients, which can cause mechanical instabilities in the deposited films.

3.4.1 Ge-on-Si heteroepitaxy

The early stages of Ge on Si epitaxy, MBE was used by many authors to fabricate various infrared detectors [13], [22], [24]–[26], [30]–[34]. The performance of these devices has been shown in the previous section. MBE has been mainly used to grow Ge on Si heterostructures using SiGe buffer layers, in order to reduce the threading dislocation density at the SiGe/Si interface. This reduction in TDD was demonstrated, but the price paid in term of low absorption coefficient (because of the indirect bandgap of the SiGe alloy) was too high to make an efficient infrared photodetector. Different

authors, however, have demonstrated the epitaxial growth of pure Ge on Si without using SiGe buffer layers [23], [72], [73], [74], [75]. Baribeau *et al.* investigated the direct deposition of pure Ge on Si (001) by MBE for different growth temperatures, 285 °C, 610 °C and 700 °C, respectively. The best results were obtained for the intermediate temperature, and a TDD of ~2×10^{7} cm^{-2} was measured. They also studied post-growth annealing treatments confirming a significant improvement in the crystalline quality of all Ge epilayers. Annealing was performed for 30 min at 700 °C and a reduction of the TDD by almost one order of magnitude was observed. This TDD reduction was also verified by Fukuda *et al.* using an annealing temperature of 680 °C. Malta *et al.* demonstrated the epitaxial growth of a 2.5 μm-thick Ge layer on Si using MBE [72], [76]. Ge growth was initiated at a temperature of 500 °C and then raised to 900 °C. They obtained very low TDDs in the range of 1-3 $\times 10^{5}$ cm^{-2} that were confined at the Ge/Si interface without propagating to the free surface of the Ge epitaxial film. Although SiGe buffer layers were not used, due to the high growth temperature (near to the Ge melting point, see Table V), they observed a melting effect of Ge at the interface and subsequent alloying with the Si substrate.

On the other hand, Liu *et al.* demonstrated an epitaxial growth of a Ge thin film (~200 nm) on Si by MBE at low temperature (370 °C) [73]. This low growth temperature was used to reduce the tensile strain, which is introduced by thermal mismatch between Ge and Si. They measured, however, a high TDD of ~4×10^{10} cm^{-2}.

The aforementioned CVD techniques have also been used to grow Ge on Si using SiGe buffer layers, and different infrared detectors were fabricated [27], [28], [38]. Extensive research has been shown to reduce the TDD by Currie *et al.* [37], as shown previously (see paragraph 3.3.1). However, the graded SiGe buffer layer method usually requires a thick buffer for pure Ge epitaxy on Si, while in silicon photonics technology it is preferable to fabricate the Ge detectors close to Si, mainly in a waveguide design to facilitate the light coupling.

To prevent and reduce the TDDs during the growth of pure Ge on Si, other techniques have been developed during the last 20 years: the two temperature method, cyclic thermal annealing and selective epitaxial growth (SEG). These techniques will be described in the following sections.

3.4.1.1 *Two-temperature LT/HT Ge growth and cyclic thermal annealing*

The two-step low temperature/high temperature (LT/HT) method was introduced by Fan *et al.* in 1986, for the reduction of threading-dislocations in the heteroepitaxial growth of GaAs on Si [77]. The lattice mismatch in the GaAs/Si system is about 4 %, and it is very similar to that between Ge and Si. This TDD reduction was obtained by cyclic interruption of the vapour phase growth process, where the samples were cooled to room temperature and heated back to the growth temperature. Following this, different authors reported the epitaxial growth of GaAs on Si by a two-step process [78]–[80].

This TDD reduction was also demonstrated by Lee *et al.* [78]. A buffer layer 0.3 μm-thick GaAs was grown on Si at low substrate temperatures between 475 and 550 °C. The substrate temperature was raised to 580 °C to grow a 3 μm-thick GaAs epilayer. Then annealing at 850 °C was carried out for 15 minutes. Samples with and without annealing were analysed by TEM microscopy. Samples that were not annealed showed many defects propagating from the GaAs/Si heterointerface into the thick epilayer region. The measured TDD near the epilayer surface was ~10^9 cm^{-2}. On the other hand, annealed samples showed a reduction of two orders of magnitude in TDD (~10^7 cm^{-2}).

An improvement of one order of magnitude in TDD was obtained by Yamaguchi *et al.* [79]. Firstly, a 10 nm thin GaAs buffer layer was grown at 400 °C, then a second layer of about 2 μm was grown at 700 °C. The growth was interrupted and the substrate temperature was lowered to near room temperature. The sample was heated to 700 – 900 °C and annealed for 1-15 min. The substrate temperature was lowered to 700 °C and growth was resumed to obtain a total GaAs film thickness of 3.5 – 4 μm. The annealing treatment reduced the TDD, which was measured to be about 2-3 × 10^6 cm^{-2}.

Based on the above discussion, the research carried out for the GaAs/Si system demonstrated that the two-step method combined with the thermal annealing was able to produce GaAs on Si with low TDDs as low as 10^6 cm^{-2}.

Although Baribeau *et al.* proposed to use two temperature steps to grow pure Ge on Si in 1986, the first use was in 1998 [81], [82]. Colace *et al.* proposed and fabricated a metal-semiconductor-metal (MSM) near-infrared photodetector, growing pure Ge on Si by using the two-step temperature method in a UHV-CVD reactor. In the two-step Ge growth procedure, firstly, after thorough cleaning, the substrate was maintained at low

temperature (~300 – 400 °C), and a thin layer of Ge buffer layer (~50 – 100 nm) was grown to prevent strain release through undesirable island growth. Secondly, the substrate temperature was elevated to ~500 – 700 °C and a thick Ge layer with reduced threading dislocation density was grown on top of the low-temperature thin Ge buffer layer.

Although Colace *et al.* demonstrated an efficient pure Ge on Si near-infrared detector, they reported a high density of recombination centres (~10^{14} cm^{-3}) that suggested a high TDD. An improvement to the two-step method was carried out by Luan *et al.* [4], [83]. This improvement consisted of adding a cyclic thermal annealing step after the growth to reduce the high TDDs. A 30 nm-thick Ge buffer layer was grown on a Si substrate at 350 °C. The temperature was raised to 600 °C to deposit a 1 μm-thick Ge layer. After the two-step growth, wafers were cyclically annealed between a high annealing temperature (T_H) and a low annealing temperature (T_L). Their major findings are summarised in Table VI.

Sample ID	A	B	C	D	E
T_H(°C)/time (min)	NA	900 °C/10 min	900 °C/100 min	900 °C/10 min	900 °C/10 min
T_L(°C)/time (min)	NA	100 °C/10 min	100 °C/10 min	100 °C/min	780 °C/10 min
Number of annealing cycles	NA	1	1	10	10
Threading-dislocation density (cm^{-2})	$9.5\pm0.4\times10^8$	$7.9\pm0.6\times10^7$	$7.8\pm0.5\times10^7$	$5.2\pm0.6\times10^7$	$2.3\pm0.2\times10^7$

Table VI. Summary of annealing parameter and TDDs obtained in [4].

As shown in Table VI, increasing the number of annealing cycles proved more effective in reducing TDDs. Sample E, which was cyclically annealed between T_H = 900 °C and T_L = 780 °C had the lowest TDD of 2.3×10^7 cm^{-2}.

Shah *et al.* studied the reduction in TDDs as a function of the Ge buffer layer and Ge epilayer thicknesses [84]. Ge on Si heteroepitaxy was performed using a RP-CVD reactor. The LT Ge buffer layer was grown at 400 °C and its thickness varied in the range of 30 to ~150 nm, while the HT Ge layer was grown at 670 °C and its thickness was limited to a maximum of ~ 1.2 μm. Post growth *in situ* anneals of 830 °C for 10 min were performed. Figure 3.9 summarises their results.

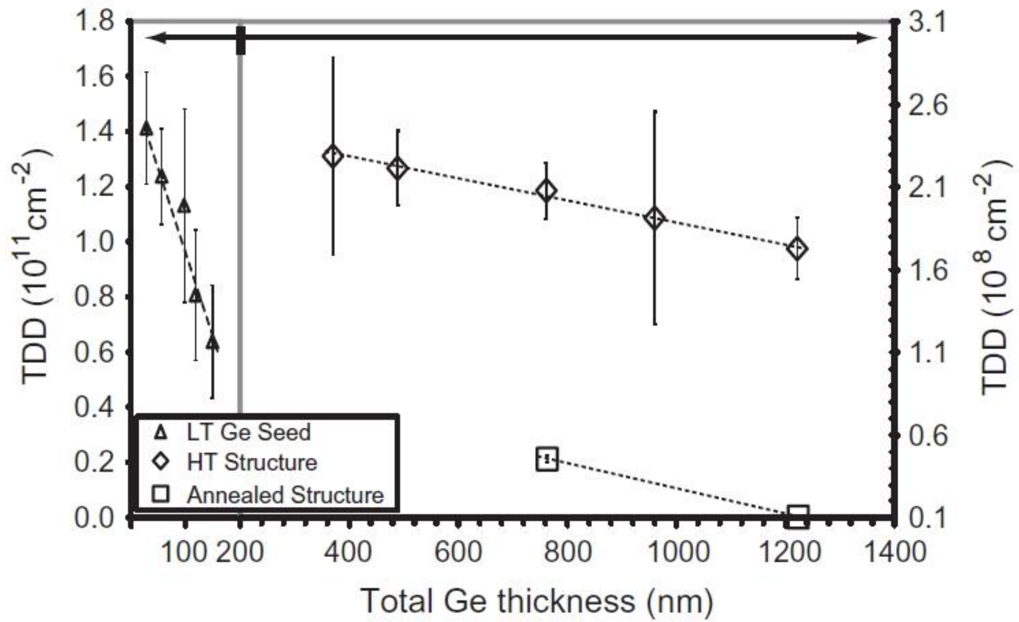


Figure 3.9. The variation of TDD as a function of the LT Ge buffer layer thickness, HT Ge thickness and annealing parameters [84].

A high TDD was generated in the LT Ge buffer layer, and this TDD was slightly reduced when its thickness was increased. The purpose of the LT buffer layer is to keep the system planar and generate dislocations. When the HT Ge layer was grown, the threading dislocation extended also in to this layer, as shown in previous sections. However, as the HT Ge thickness was increased there was a reduction in TDD. This reduction was more consistent when annealing was performed.

As demonstrated by Yamaguchi *et al.* for the GaAs/Si system, the two-step method combined with cyclic annealing also showed a reduction of TDDs on the Ge/Si heteroepitaxy. After the works of Colace and Luan many research groups applied the two-temperature method combined with the cyclic thermal annealing to fabricate various Ge-on-Si near-infrared detectors (the performance of these will be evaluated later in this chapter) and different reactors have been used (MBE, RP-CVD, UHV-CVD, LEPECVD). Table VII summarises, in terms of growth parameter and TDDs, some of the work published on the heteroepitaxy of pure Ge on Si using the two temperature method.

	Thickness of Ge seed layer (μm) / Temperature of growth (°C)	Thickness of Ge layer (μm) / Temperature of growth (°C)	Temperature of annealing (°C)	Annealing cycles	TDD (cm^{-2})	Reference
RPCVD	0.1 / 400	1 / 670	830	- NA	~ 2×10^7	[84]
	0.025 / 400	1.6 / 750	750 - 875	- NA	< 2×10^8	[85]
	0.04 / 400	0.3 / 730	NA	1	- NA	[86]
	0.1 / 400	1.2 / 670	~900	1	~ 10^7	[87]
	0.1 / 400	1.2-1.7 / 650	- NA	- NA	- NA	[88]
UHVCVD	0.03 / 350	1/600	780 - 900	10	2.3×10^7	[4]
	0.05 / 350	0.4 / 600	780 - 900	1	1×10^8	[89]
	0.06 / 360	1.1 / 730	650 - 850	- NA	- NA	[90]
LEPECVD	0.03 / 350	1 / 600	850 - 900	3	2.5×10^7	[91]

Table VII. Summary of Ge on Si heteroepitaxy by two-step LT/HT Ge growth and cyclic thermal annealing.

The two-step method has also been used to grow Ge epilayer on Si using a SiGe buffer layer deposited at low temperature. The role of the SiGe buffer layer was to confine and lower the dislocation density by reducing the lattice mismatch between the Ge and Si. No significant improvements, however, have been reported in terms of TDDs and near-infrared photodetector performance [92], [93].

3.4.1.2 Selective epitaxial growth of Ge-on-Si

Another approach to reduce the TDDs is selective epitaxial growth (SEG). This technology allows the growth of Ge directly on Si wafers with a patterned SiO_2 (or Si_3N_4) mask. Combined with the two-temperature LT/HT method and cyclic thermal annealing, reductions in the TDD in the film have been demonstrated. When the SEG was applied to a small area, threading arms of misfit dislocations glide and terminate at the edge of the growth area reducing the overall TDD. This reduction happens because during the Ge-on-Si heteroepitaxy the misfit dislocations lie at the Si-Ge interface, and the threading arms climb, propagating at a 45 degree angle from the Si substrate, and terminate at the oxide sidewalls [94]. This process, called epitaxial necking, is shown in Figure 3.10.

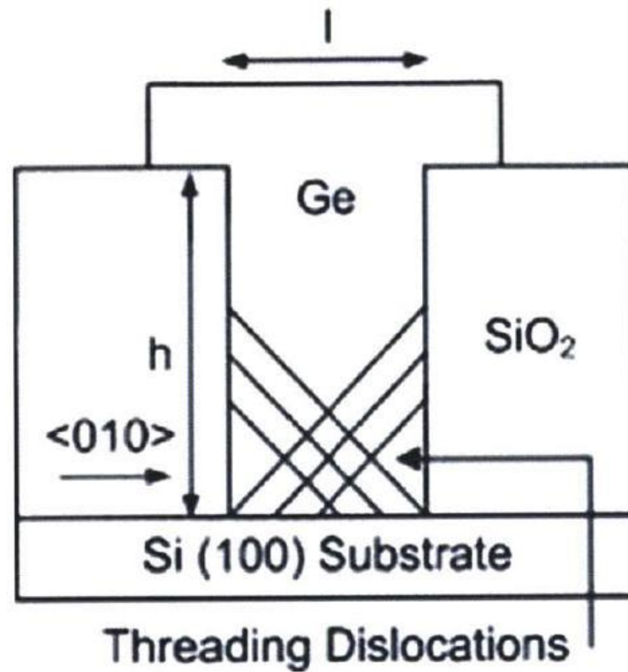


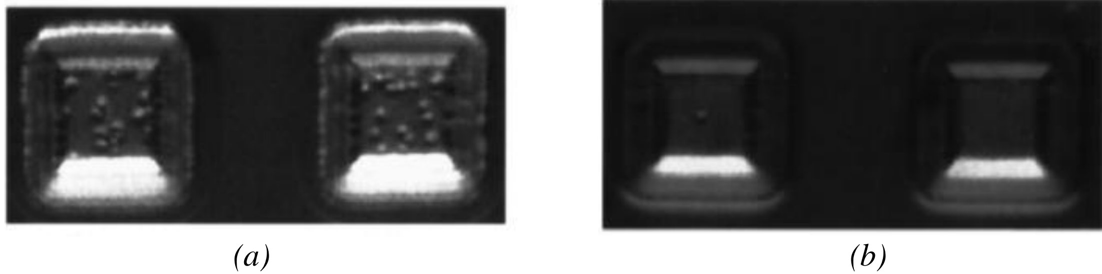
Figure 3.10. Cross-section diagram demonstrating the principles of epitaxial necking showing zero threading dislocations at the Ge film surface [94].

There is also an added benefit called aspect ratio trapping (ART), which occurs when the Ge film width is comparable to the layer thickness. This advantage has been demonstrated by different authors [95], [96]. Park *et al.* demonstrated defect-free germanium growth in different aspect ratio (AR = trench height/width) SiO_2 trenches on silicon [95]. The maximum Ge thickness was kept to 450 nm. They demonstrated that for trenches having an AR > 1 all the dislocations originating at the Si/Ge interface were trapped at the oxide sidewall and were confined near the interface without propagating in the Ge epilayer. On the other hand, trenches with an AR < 1 showed some dislocation terminations at the Ge surface. In both cases, however, no defect generation along the SiO_2 sidewall was observed. Similar results were also obtained by Wang *et al.* [96]. However, they also observed that growing selective Ge in SiO_2 trenches by using the two temperature method and cyclic thermal annealing was more effective for reducing the TDDs. Threading dislocations were confined at the Si/Ge interface, leaving a defect-free Ge surface, and a TDD of $\sim 1 \times 10^7$ cm^{-2} was measured at the interface.

However, ART techniques are dependent on critical nano-scale patterning and selective growth over window sizes in the 100 nm range.

The first SEG on patterned SiO_2/Si wafers of a 10 μm-side square mesa Ge-on-Si was demonstrated by Luan *et al.* [4]. The thickness of the oxide was 1 μm. The two-step

process, followed by 10 minute annealing at 900 °C, Figure 3.11a, and cyclic annealing between 900 °C and 100 °C, Figure 3.11b, was used.



(a) (b)

Figure 3.11. 10 μm square mesa Ge selectively grown on patterned SiO_2/Si wafers followed by (a) 10 minute annealing at 900 °C, and (b) 10 annealing cycles between T_H = 900 °C and T_L = 100 °C.

The measured TDD was 4.3 × 10^7 cm^{-2}, Figure 3.11a, and 2.3 × 10^6 cm^{-2}, Figure 3.11b. They also observed many threading dislocation-free Ge mesas in their samples.

Similar results were later obtained by Sammak *et al.* on 10 μm square mesas and 1 μm-thick SEG Ge on Si [97].

Various near-infrared photodetector have been demonstrated using this technology, and their performance will be evaluated later in this chapter. Moreover, it should be noted that the advantage of SEG is not only the reduction of TDDs, but more important is the possibility to selectively introduce Ge into CMOS technology for large scale integration with other Si photonics component (e.g. waveguides, modulators) that make this technology very practical and powerful.

3.4.2 Effect of TDD on device performance

As discussed in several occasions throughout this chapter, the lattice mismatch in the Ge/Si system (or GaAs/Si or InGaAs/Si) causes a high density of defects at the heterointerface. The density of these defects is quantified through the TDD. The TDD reduction has been studied by many authors and different solutions have been proposed. The importance of reducing the TDD is because defects at the heterointerface and in the epilayer degrade the performance of the device and compromise their reliability.

Dislocations are known to increase the dark current of a device. As shown in section 3.3, many authors have demonstrated that a reduction of TDD helped to reduce the dark current of a device. A more detailed investigation on the correlation between threading dislocation density and leakage current of SiGe p-i-n diodes was carried out by Giovane

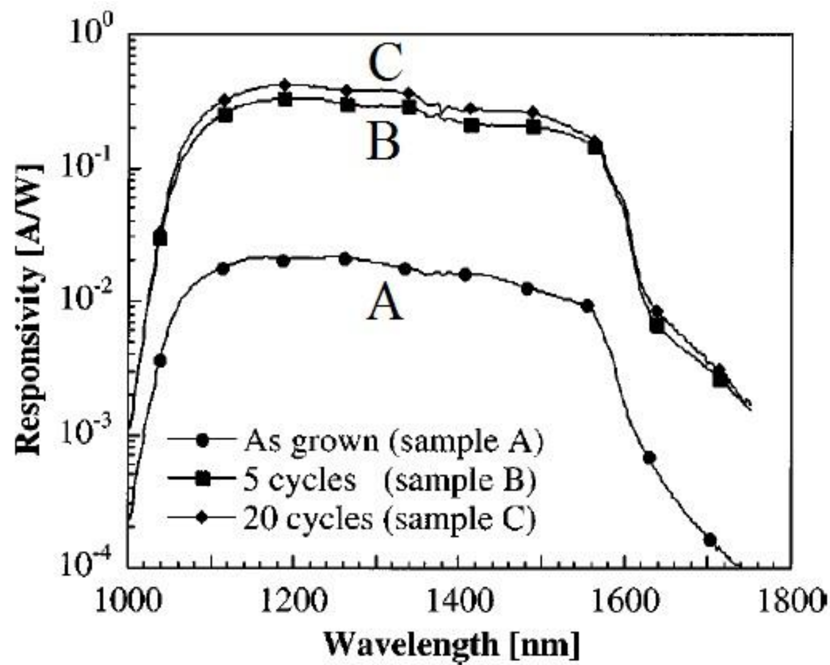
et al. [27]. They demonstrated that defect states related to threading dislocations act as generation and recombination centre in these devices, and the bulk leakage current correlates directly with the TDD. In fact, they measured an increase in bulk leakage current densities of 0.02, 0.32, and 0.78 mA/cm^2 as the TDD increased (2×10^5, 2×10^6, and 9×10^6 cm^{-2}, respectively).

The effect of TDD on the performance of a photodetector was also demonstrated by Colace *et al.* [98]. Here, the devices consisted of pure Ge on Si grown by a two-step UHV/CVD process that were treated in different ways, as reported in Table VIII. Mesa heterojunction diodes with areas ranging from 4×10^{-4} to 0.1 cm^2 were fabricated.

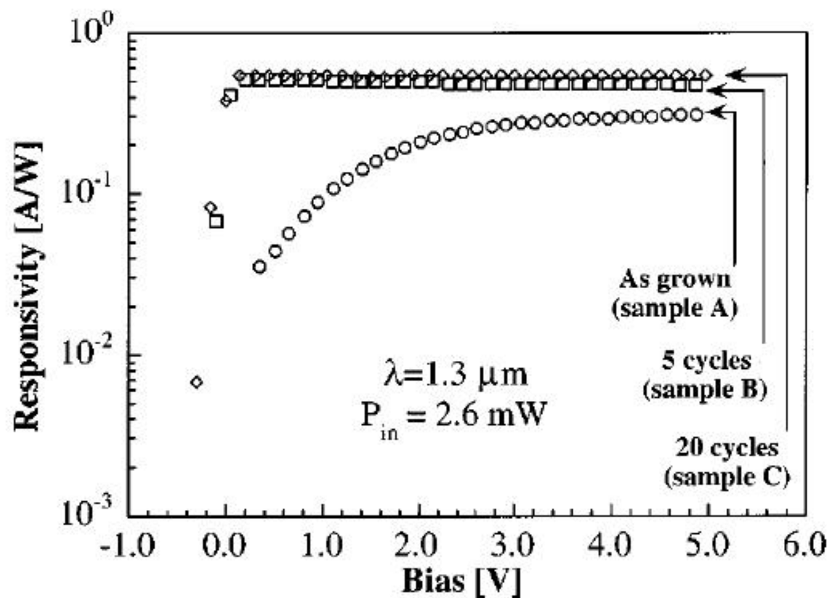
Sample ID	A	B	C
Cyclic annealing temperatures (°C)	NA	900/780	900/780
Number of annealing cycles	0	5	20
Threading-dislocation density (cm^{-2})	$9.5 \pm 0.4 \times 10^8$	$2.7 \pm 0.1 \times 10^7$	$1.6 \pm 0.1 \times 10^7$
Mobility-lifetime product (cm^2/V)	7×10^{-8}	2×10^{-7}	3×10^{-6}

Table VIII. Summary of process parameters as reported in reference [98].

The responsivity was measured in short circuit mode at different reverse biases, as shown in Figure 3.12a-b.



(a)



(b)

Figure 3.12. (a) Short-circuit spectral responsivity of mesa heterojunction photodetectors made from Ge grown on Si treated with different post-growth annealing treatments. (b) Responsivity at 1300 nm at different reverse biases as reported in [98].

A large improvement in responsivity was measured, in both operation modes, in the annealed samples (samples B and C) when compared to that measured from the as-grown samples. The measured mobility for sample C was 3500 cm^2/Vs, and the calculated carrier lifetime was 0.8 ns. This is three or four orders of magnitude shorter than that of bulk Ge. They attributed the short carrier lifetime to recombination at

threading-dislocation in the Ge epilayer and misfit dislocations at the Ge/Si interface. They confirmed the beneficial effect of post-growth annealing treatments on reducing TDD and, hence, improving the electrical and optical properties of their devices.

Further research on the effect of TDDs on the dark current were performed by DiLello *et al.* in 2012 [69]. Planar n-i-p photodiode were fabricated in a RP-CVD reactor. 1 μm-thick undoped Ge was grown on top of a p+ Si substrate. The Ge was grown by the two-step method. Samples underwent annealing between 450 and 800°C to reduce the threading dislocations. The annealing step was performed for 4 cycles, 2 cycles or none at all to study the effect of threading dislocation on device dark current density. The Ge was then implanted with phosphorous to create a shallow n+ region for the top contact and the devices were passivated with SiO_2. Different devices, with areas ranging between 5×5 to 300×300 μm^2, were fabricated. The measured TDD, for devices that were not annealed, was ~3×10^8 cm^{-2}, while after 4 annealing cycles it was ~5×10^7 cm^{-2}. The effect of annealing on the dark current of a 300×300 μm^2 device is shown in Figure 3.13.

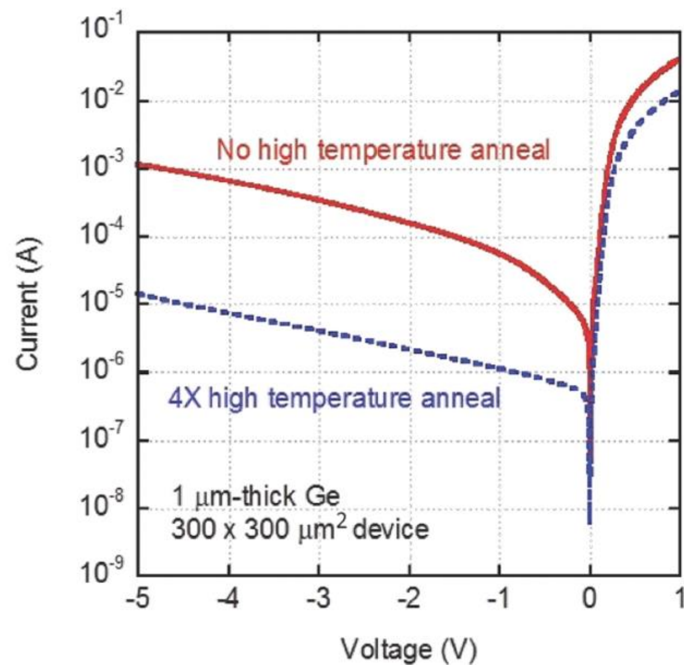


Figure 3.13. Measured dark current for a 300×300 μm^2 device. The solid line (red) represents the dark current that did not receive any annealing treatment after the growth. The dot line (blue) is the dark current for the same size device after 4 cycles annealing between 450 and 800 °C.

At -1 V, a reduction of dark current by a factor of 45 was demonstrated, most likely due to the reduction in threading defects. This reduction was not consistent over all sizes.

Small devices showed a reduction of only a factor of 15, because the effects of the surface dominate the effects of the defects in the bulk material of the device.

The effect of threading dislocations on the performance of a photodetector has been evaluated in terms of its effect on the device photoresponse and dark current. As pointed out previously in this chapter, leakage current is a source of noise and power dissipation in an optical receiver. Hence, its reduction is important. Although threading dislocations play an important role in this reduction, others factors (e.g. semiconductor material, device geometry and surface passivation) also have an impact on the leakage current. Intrinsically, Ge is a narrow bandgap material and effect on the device dark current has more impact than on InGaAs, as shown previously. Device geometry needs to be also taken in account, and the surface passivation has also a major contribution. As shown by DiLello [69], on smaller size devices, this contribution dominates the device dark current, so that the passivation of Ge become very important at these small dimensions. Although different materials (e.g. SiO_2, Si_3N_4, GeO_2, α-Si) can be used to passivate the Ge surface, this research field is still open, and the best material for its passivation has not been found.

Moreover, the aforementioned analyses have been done at low reverse biases. For devices like APDs and single-photon detectors, which operate at much higher voltages, the impact of threading dislocations needs to be evaluated. At higher voltages and, hence, high electric fields (near or above the material breakdown field), others mechanism, like band-to-band tunnelling or trap assisted tunnelling contribute to the device dark current. To date, the correlation between these mechanisms and the threading-dislocation, at higher voltages, has not been evaluated. For a single-photon detector, there is also a correlation between the device dark current and the device dark count rate (DCR), so it is better to keep the dark current at the lowest possible level to have a low DCR.

3.4.3 Effect of the Ge strain on the absorption at the NIR wavelengths

The 4 % $\left(f = \frac{a_{Si}-a_{Ge}}{a_{Ge}}\right)$ lattice mismatch between Ge and Si causes a high density of misfit dislocations and threading-dislocation in the epilayer. Another process that generates strain in the Ge film is due to the temperature at which the growth is performed. For a thickness > 1 µm, the Ge film is nearly completely relaxed at the growth temperature (600 – 700 °C). Upon cooling to room temperature, tensile strain

can be accumulated in the Ge film due to the larger thermal expansion coefficient of the Ge film compared with Si substrate. The thermal expansion coefficients of Ge and Si at temperature T (°C), are given by [99], [100] respectively:

$$\alpha_{Ge}(T) = 6.050 \times 10^{-6} + 3.60 \times 10^{-9}T - 0.35 \times 10^{-12}T^{2} \quad (°\mathrm{C}^{-1}) \qquad (9)$$

$$\alpha_{Si}(T) = \{3.725 \times [1 - \exp(-5.88 \times 10^{-3}(T + 149.15))] + 5.548 \times 10^{-4}T\} \times 10^{-6} \quad (°\mathrm{C}^{-1})$$

Many authors demonstrated that this thermal strain causes a tensile strain between 0.20 – 0.32 % in the Ge film [101], [102]. The first positive effect of this ~0.2 % tensile strain in the Ge film, is a reduction of the Ge bandgap of ~0.03 eV (from 0.8 eV for bulk Ge to ~0.77 eV for Ge epitaxially grown on Si). This was first demonstrated by Ishikawa *et al.*, who verified this bandgap reduction by measuring the absorption coefficient of 1 μm-thick Ge epitaxially grown on Si by two-step method in a UHV/CVD reactor, as shown in Figure 3.14.

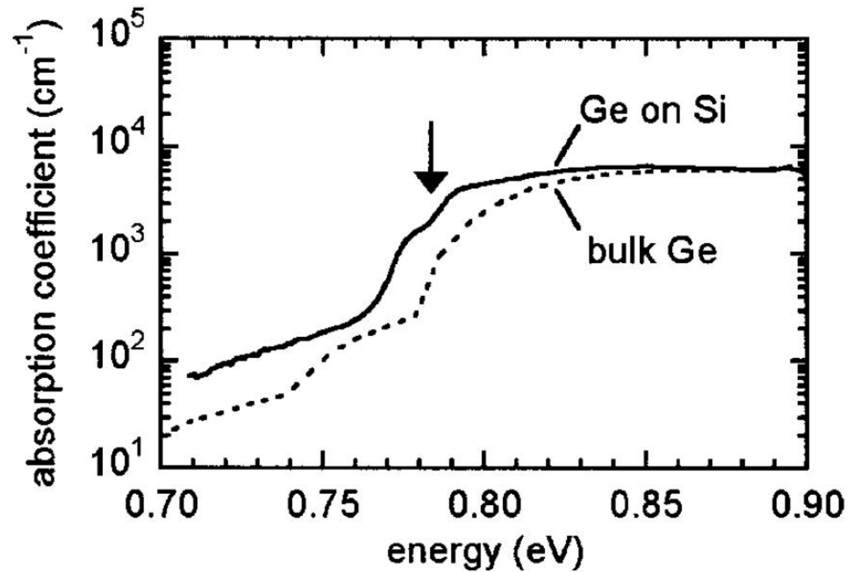


Figure 3.14. Comparison of the absorption coefficient for Ge-on-Si (solid line) and bulk Ge (dashed line) [101].

The advantage of bandgap reduction is to enhance the absorption coefficient of bulk Ge from 840 cm^{-1} to 3300 cm^{-1} at 1550 nm wavelength [103]. It is also clear that the detection range is extended toward smaller energies (longer wavelengths).

Additionally, when Ge is cooled, its absorption coefficient drops very quickly at the infrared wavelengths (mainly at 1550 nm) [104], as shown in Figure 3.15.

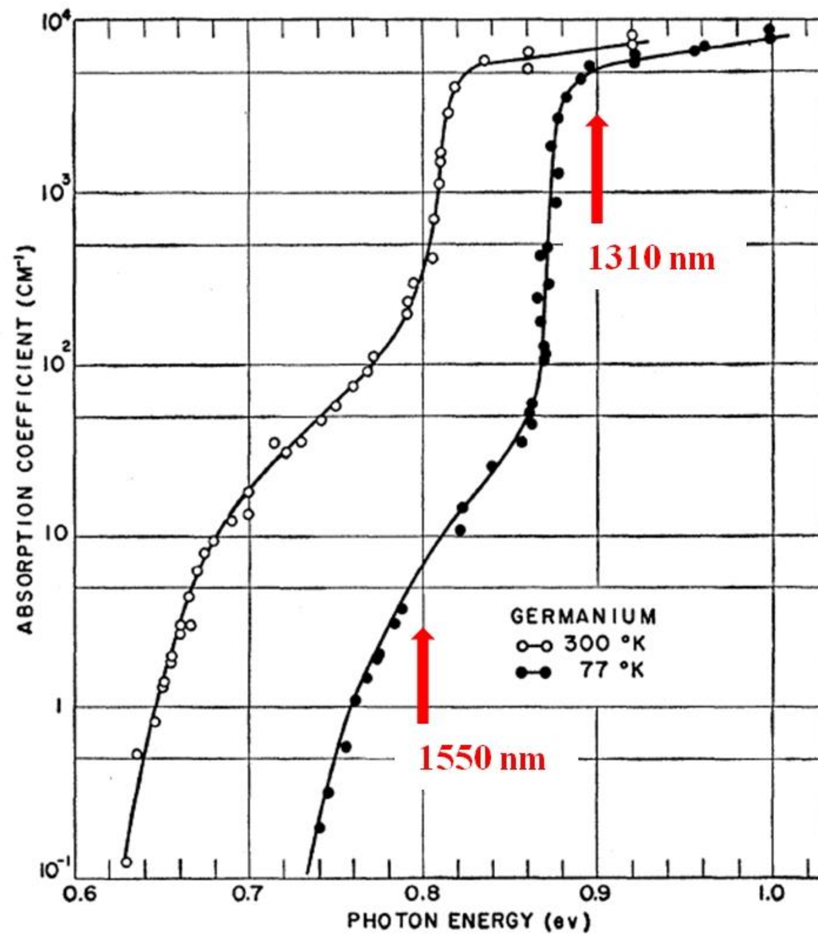


Figure 3.15. Absorption spectra of single crystal Ge at 77 K and 300 K [104].

Therefore, for this reason also the effect of strain on Ge becomes important. If the strain can be further enhanced it will also enhance the absorption coefficient at infrared wavelengths, further shrinking the Ge bandgap. This enhancement could have a significant impact on devices like Ge-based lasers and photodetectors that could be integrated on a Si photonics platform.

3.5 Ge-on-Si by wafer bonding

As with the InGaAs/Si system, wafer bonding techniques can be also applied for the integration of germanium on silicon. The limitations of these techniques are this intrinsic complexity on integration in a CMOS foundry, poor mechanical strength, thermal stability, the handling of Ge substrates that are more expensive than Si substrates and, they do not give the possibility to selectively introduce Ge on Si in small areas. Although these techniques have not been used in this thesis, a brief literature review is presented.

Bonding techniques have been used in the Ge/Si system to overcome the limitation on the devices performance due to 4 % lattice mismatch between these two semiconductor materials. There are not many papers in the literature about devices fabricated using these techniques. While in the InGaAs/Si system there is a clear advantage of using wafer bonding, because the TDDs are much higher in this system when epitaxially grown (due to the 8 % lattice mismatch between InGaAs and Si), its complexity and the technological progress on the Ge-on-Si heteroepitaxy have limited its use on the integration of Ge on Si.

In 2013, Gity *et al.* reported the fabrication of a p-n Ge/Si heterojunction photodiode by wafer bonding [106]. An n+ Si substrate and a p- Ge substrate were bonded together. The bonding was followed by two 24-hour anneal step at 200 °C and 300 °C to enhance the bond strength. Then the Ge wafer was thinned leaving a 5.4 μm thick Ge layer. Mesa p-n diodes ranging from 20 to 500 μm were fabricated. A challenge in this kind of structure is the thickness of the bonded interface which should be minimized because it affects the carrier transport across the interface. In this case, a 2 nm-thick amorphous interfacial region was observed, plus additional regions at the interface on the Ge side. A high dark current of 25 mA/cm^2 was measured for a 500 μm diameter mesa with responsivity at 1550 nm wavelength of 1.6 A/W. However, the performance of these devices was mainly limited by the interfacial traps.

As it will be shown in the next section the performance of Ge-on-Si wafer bonding photodetectors is worse than Ge-on-Si detectors made by heteroepitaxy.

3.6 Ge-on-Si NIR photodetectors

After the introduction of the two-temperature method and annealing in Ge-on-Si heteroepitaxy demonstrated by Colace *et al.* in 2000 [98], different Ge-on-Si infrared photodetectors have been demonstrated in the last 15 years. Research has been focused on the fabrication of efficient detectors at infrared wavelengths with low levels of leakage current. P-i-n, waveguide Ge detectors, and APDs have been successfully demonstrated. A literature review of these detectors is presented with the main focus on their performance.

3.6.1 Ge-on-Si p-i-n photodetectors

One of the most used configurations in the literature for photodetectors is the p-i-n or n-i-p structure. This configuration is mainly used in optical receivers to convert the optical signal in an electrical system. Their main requirements are: low leakage current, which influences the S/N ratio and the power consumption; high efficiency at infrared wavelengths; and high speed to satisfy the requirements in terms of high bandwidth. Although the market at infrared wavelengths is mainly dominated by III-V photodetectors, thanks to the advantages obtained in the heteroepitaxial systems a lot of effort has been put in to demonstrate low leakage current and high responsivity Ge-on-Si detectors that can be integrated within a CMOS process and with other silicon photonics components such as modulators and waveguides.

The first Ge-on-Si p-i-n photodetector was demonstrated by Colace *et al.* [98] in 2000. Heteroepitaxy was performed in a UHV CVD reactor using the two-step method and mesa p-i-n diodes with areas ranging between 4×10^{-4} and 0.1 cm^2 were fabricated. No passivation layer was used. The devices exhibited a dark current density (DCD) of 30 mA/cm^2 with a responsivity of 0.55 A/W at 1310 nm (corresponding to an external quantum efficiency of 52 %) measured in short circuit mode. One year later, they also demonstrated p-i-n and n-i-p detectors fabricated using n Si or n+ Si and p Si or p+ Si substrates, respectively [107]. The same heteroepitaxy steps were also used for these devices. For the p-i-n structure the Ge was implanted with boron, while for the n-i-p structure the Ge was implanted with phosphorous to make the top contact in both configurations. The implant was activated at 600 °C. Mesas ranging between 4×10^{-4} and 0.1 cm^2 were fabricated. The lowest DCD of 20 mA/cm^2 was measured on the n-i-p structure grown on the heavy doped p+ substrate. These photodiodes exhibit a lower responsivity of 0.3 and 0.2 A/W at 1300 and 1550 nm wavelengths, respectively.

In 2002, Fama' *et al.* demonstrated n-i-p photodetectors [108]. 4 μm-thick Ge was grown on top of a highly doped p-type Si substrate on a UHV/CVD reactor using the two-step method. 10 annealing cycles were performed at a temperature ranging between 780 and 900 °C to reduce the TDDs. The Ge layer was then implanted with phosphorous to create the n region and to make a low resistivity top contact. Mesa diameters ranging between 135 and 585 μm were fabricated by photolithography. The measured DCD was 15 mA/cm^2 and 8.3 mA/cm^2 for a 585 and 135 μm mesa diameter, respectively. They also observed a linear scaling of dark current with area, which

confirmed that the main mechanism contributing to the leakage current was carrier recombination-generation (R-G) in the bulk Ge. The measured responsivity was 0.89 and 0.75 A/W at 1310 and 1550 nm wavelengths, respectively.

Bandaru *et al.* fabricated p-Ge/n-Si photodetectors using MBE [109] in 2004. A 200 nm-thick p-Ge layer (10^{17} cm^{-3}) was grown on top of a n-Si substrate (10^{15} cm^{-3}) using a low temperature (< 450 °C) process. 10 nm p+ Ge was also grown on top of Ge for the contact. Square mesas from 400 μm to 5 mm were defined. Because of the high doping in the Ge, the depletion region extended mainly in the Si layer giving a low DCD of 0.3 mA/cm^2 at a reverse bias of 1V. Devices were not characterised in terms of their responsivity at infrared wavelengths and, due to the low temperature growth the Ge quality observed was poor.

In 2005, Dosunmu *et al.* designed and fabricated a resonant cavity Ge-on-SOI photodetector, whose structure is shown in Figure 3.16 [110]. For the SOI wafer, a 340 nm thick Si device layer, which is the top layer of a SOI substrate, and 200 nm-thick SiO_2 were chosen giving a reflectivity of 55 % at 1550 nm wavelength. Prior to the 1450 nm-thick Ge growth using a low temperature buffer layer, the Si device layer was implanted with boron to form the p-contact. Circular mesas ranging between 10 and 78 μm in diameter were fabricated and a SiO_2 anti reflection coating was also used and devices were back-illuminated. The measured responsivity was 0.73 A/W at 1538 nm wavelength and at a reverse bias of 0.5 V. The DCD was not stated in the paper.

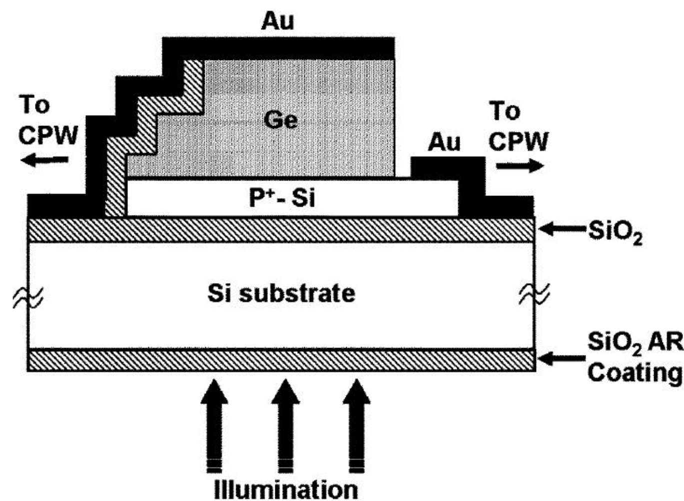


Figure 3.16. Cross-sectional view of the Ge-on-SOI resonant cavity photodetectors proposed by Dosnumu [110].

The same year, a vertical-incidence Ge-on-Si photodiode grown by MBE, which structure is shown in Figure 3.17, was proposed by Jutzi *et al.* [111].

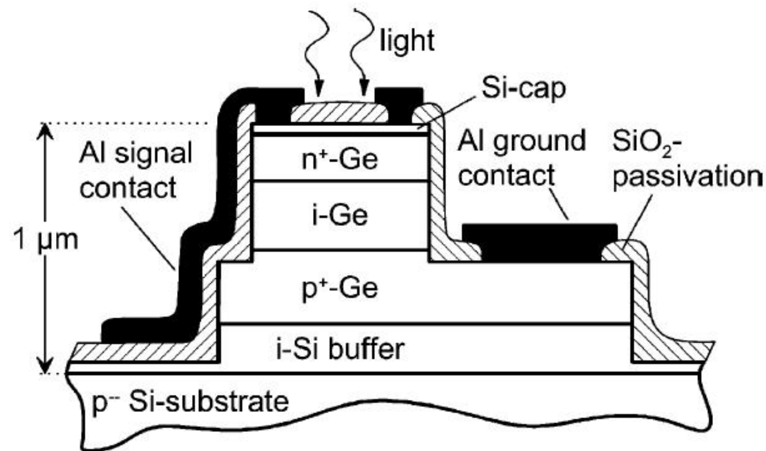


Figure 3.17. Cross-sectional view of the vertical-incidence Ge-on-Si photodiode proposed by Jutzi [111].

300 nm-thick p+ Ge was grown on top of a Si buffer layer, and then 300 nm i-Ge was grown at low temperature (300 °C) followed from 200 nm-thick n+ Ge. SiO_2 was deposited by plasma enhanced CVD (PECVD) for passivation. Mesa diameters ranging between 10 and 80 μm were defined by photolithography. A very high DCD of 100 mA/cm^2 at a reverse bias of 1 V was measured with an external quantum efficiency of 16 % and 2.8 % at 1298 nm and 1550 nm wavelengths, respectively.

Liu *et al.* demonstrated a 0.25 % tensile strained Ge-on-Si p-i-n photodetector [103]. As described previously, the 0.25 % strain was obtained by thermal mismatch and by adding a $TiSi_2$ backside silicidation. The i-Ge layer was 2.35 μm-thick and the measured DCD was 10 mA/cm^2. The main effect of the strain is on the Ge absorption coefficient which was demonstrated by measuring responsivities of 0.422 and 0.52 A/W at 1550 nm for devices without (0.2 % strain) and with (0.25 % strain) backside silicidation, respectively. Because the strain also shifts the Ge wavelength cut off to longer wavelengths, responsivities of 0.048 and 0.1 A/W were measured for structures with 0.2 % and 0.25 % strain, respectively.

Due to the high annealing temperatures typically used in Ge heteroepitaxy, which is incompatible with the standard CMOS process, Colace *et al.* fabricated Ge-on-Si p-i-n photodetectors using a low temperature buffer layer (without annealing) to keep the maximum temperature below 600 °C in 2006 [112]. Devices were grown in a UHV CVD reactor and square mesas ranging from 20 to 80 μm were defined by photolithography. Devices showed a much higher DCD (200 mA/cm^2) than previously

reported annealed devices. The measured responsivities at 1300 and 1550 nm were 0.4 and 0.2 A/W, respectively.

In 2006, Morse *et al.* proposed Ge-on-Si p-i-n detectors passivated using amorphous Si (a:Si) [87]. 1.2 μm-thick Ge was grown by two step process followed by a 900 °C annealing. Circular mesas of diameters ranging from 15 to 250 μm were fabricated. The top surface of Ge was implanted with boron to define the p+ Ge region (dopants were activated at 650 °C). A DCD of 6.4 mA/cm^2 at a reverse bias of 1 V was measured and a linear scaling of dark current with area of the device was also observed (bulk effect in the Ge). The calculated responsivity was 0.45 A/W at 1310 nm wavelength.

Colace *et al.* demonstrated Ge-on-Si p-i-n detectors with low dark current in 2007 [6]. Compared to the other structures presented by the same author, a highly doped p+ Ge buffer layer was grown, at low temperature, on top of a p+ Si substrate. This layer was introduced to compensate the acceptor-like defects that arise from the lattice mismatch between Ge and Si which, in turn, introduce deep electronic states within the bandgap. The 1 μm-thick i-Ge was grown at higher temperature followed by a low temperature deposition (400 °C) of a 200 nm-thick n+ Ge. Mesas ranging from 20 to 80 μm were fabricated. A low DCD of 1 mA/cm^2 at a reverse bias of 1 V was measured with responsivities of 0.3 and 0.2 A/W at 1310 and 1550 nm wavelengths.

A further improvement on the DCD of these devices was also demonstrated by Isella *et al.* in 2007 [113]. Devices were grown by low-energy plasma-enhanced CVD (LEPECVD). LEPECVD has the advantage that a thick Ge layer can be deposited in a few minutes at 600 °C which is fully compatible with CMOS processing without using a buffer layer. Different photodiodes were grown on top of a n-Si substrate and the i-Ge thickness was varied from 1 to 3 μm. Devices were annealed between the growth temperature and 780 °C, which is incompatible with the CMOS process, to reduce the TDDs. After annealing, 200 nm-thick p+ Ge was deposited. The same structure with 1 μm-thick i-Ge was also fabricated without any annealing step to evaluate the effect of TDD on the device performances. Circular mesas of diameter ranging from 0.5 to 10 mm were fabricated, and a SiO_2 layer was deposited to passivate the devices. The DCD measured, in the annealed samples, was 0.4 mA/cm^2 at a reverse bias of 1.5 V, while it was two orders of magnitude higher in the un-annealed devices. The increased in DCD for the devices with a thicker i-Ge layer was due to the increasing number of R-G centres. The lowest DCD of 0.08 mA/cm^2 (at -1.5 V) was measured in the 1 μm-thick i-

Ge. An increase in responsivity with increasing i-Ge thickness was also measured. The responsivities at 0 V for the 1 μm-thick i-Ge were 0.17 and 0.14 A/W at 1310 nm and 1550 nm respectively. At 0 V the responsivity of the 3 μm-thick i-Ge was lower due to the strength of the built-in electric field being too low to effectively collect the photogenerated carriers in the thick absorption layer. On the other hand, as the reverse voltage was increased the 3 μm-thick sample showed the higher responsivities of 0.39 and 0.47 A/W (at -3 V) at 1310 nm and 1550 nm wavelengths, respectively.

One year later, the same group proposed a comparison based on the same structure, as above, but with two different Si substrates, n+ Si and n Si, respectively [114]. The same growth steps were performed, varying only the diameters of the mesa from 25 μm to 3 mm. Devices fabricated on the n Si substrate showed lower DCD than the devices fabricated on the n+ Si substrate. The DCD were 0.5 and 0.3 mA/cm^2 (at -1 V) for the 25 and 500 μm devices, respectively, showing that the perimeter component of the dark current cannot be neglected with respect to the bulk component. On the other hand, samples on the n+ Si substrate showed the highest responsivity, 0.32 A/W at 0 V and 1550 nm wavelength, because the higher doping created a larger electric field and hence good carrier collection.

The lowest reported DCD for any Ge-on-Si photodetector was demonstrated by the same authors in 2009 [115]. Two different structures were proposed, p-i-n and n-i-p, with and without annealing. The annealed p-i-n structure demonstrated a DCD of 0.041 mA/cm^2 (at -1 V), while the unannealed showed a value of 4.6 mA/cm^2 (at -1 V). The measured DCD on the annealed n-i-p samples was 0.001 mA/cm^2 (at -1 V), demonstrating the lowest DCD, while it increased to 2 mA/cm^2 (at -1 V) on the unannealed samples, probably because of the higher TDD.

In 2010, Sorianello *et al.* demonstrated Ge-on-Si p-n diodes where a layer of Ge 200 nm-thick was grown on top of a Si or SOI substrate using thermal evaporation [116]. They used this technique because of the low temperature of deposition (from 225 to 400 °C) required. They verified that to obtain acceptable monocrystalline Ge a temperature around 400 °C should be used. For temperatures below 225 °C the Ge was amorphous, while above 450 °C the Ge was polycrystalline. Normal incidence detectors were fabricated showing a relatively high DCD of 2 mA/cm^2 at -1 V. Because of the low Ge thickness the measured responsivity was only 0.002 A/W at 1550 nm. Later in 2012, they tried to improve the process by adding a thermal diffusion step of phosphorous

spin-on-dopant to compensate the acceptor states introduced by dislocations [117]. This step was added to avoid annealing and keep the thermal budget of the process to a low level (less than 600 °C). A layer 1.3 μm-thick of Ge was deposited at 300 °C by thermal evaporation and then P doping was added by spin-on-dopant at 580 °C. The measured DCD was as high as 200 mA/cm^2. However, the responsivity was slightly improved to 0.11 A/W at 1550 nm, due to the wider depletion region obtained by compensating the acceptor states with P spin-on-dopant.

In 2012, DiLello *et al.* investigated the mechanisms of dark current in planar Ge-on-Si photodiodes [69]. Devices with a Ge thickness varying between 1 and 2 μm were fabricated on top of a p+ Si substrate by using the two step growth in a RPCVD reactor. Devices were annealed between 450 and 850 °C. Ge was then implanted with P, which was activated at 550 °C. The Ge surface was passivated using 100 nm-thick SiO_2, and Ti/Al was used to form both contacts. They also introduced a further step after metallisation called post-metallisation annealing (PMA), which was performed at different temperatures ranging from 300 to 425 °C. Devices ranging from 5 to 300 μm were defined.

They observed that in small devices (5 – 20 μm) the dark current scaled with the perimeter of the device, while in big devices (100 – 300 μm) the dark current scaled with the area of the device. Moreover, they demonstrated that the PMA step, at 425 °C, reduced the dark current of the small devices (10×10 μm^2) by a factor of ~1000, while for larger devices (100×100 μm^2) this reduction was a factor of ~140. The small device showed a DCD of 10 mA/cm^2 with PMA, while it increased to 100 mA/cm^2 without PMA.

This PMA step was introduced because dangling bonds at germanium/insulator interfaces are always negatively charged [118]. Without PMS, there was a large depletion region at the Ge surface that created a region with a high electric field and hence increased the leakage current. After PMA, holes were drawn to the surface of Ge, reducing the depletion layer and hence reducing the electric field. To demonstrate this, they grew an identical structure where the top Ge layer was slightly doped p-type (~5×10^{17} cm^{-3}) and then implanted with P. The dark current was reduced by adding the p-type layer by nearly the same amount as the PMA. In terms of responsivity, the PMA did not have any effect, and a value of 0.4 A/W at 1550 nm was obtained.

Table IX summarises the performance of Ge-on-Si p-i-n and n-i-p discussed in this section.

Structure	Epitaxial growth technique	Responsivity [A/W]	Dark Current Density [mA/cm^2]	Wavelength [nm]	Reference
Ge-on-Si mesa p-i-n detector	UHV/CVD two-temperature process	0.55	30	1300	[98]
Ge-on-Si mesa n-i-p detector	UHV/CVD two-temperature process	0.3@1310 nm 0.2 @1550 nm	20	1300 – 1500	[107]
Ge-on-Si mesa n-i-p detector	UHV/CVD two-temperature process	0.89 @1310 nm 0.75 @1550 nm	8.3	NA	[108]
Ge-on-Si mesa p-n detector	MBE	NA	0.3	1300 – 1500	[109]
Ge-on-SOI mesa p-i-n detector	NA	0.73	NA	1538	[110]
Ge-on-Si mesa p-i-n detector	MBE	η_{ext} = 16 % @1298 nm η_{ext} = 2.8 % @1552 nm	100	1298 – 1552	[111]
Ge-on-Si mesa p-i-n detector	UHV/CVD two-temperature process	0.4@1310 nm 0.2 @1550 nm	200	1310 – 1550	[112]
Ge-on-Si mesa p-i-n detector	UHV/CVD two-temperature process	0.45 @1310 nm	6.4	1310	[87]
Ge-on-Si mesa p-i-n detector	RPCVD two-temperature process	0.3 @1310 nm 0.2 @1550 nm	1	1310 – 1550	[6]
Ge-on-Si mesa p-i-n detector	LEPECVD no buffer	0.17 @1310 nm 0.14 @1550 nm	0.4	1310 – 1550	[113]
Ge-on-Si mesa p-i-n detector	LEPECVD no buffer	0.32 @1550 nm	0.5	1550	[114]
Ge-on-Si mesa p-i-n detector	RPCVD two-temperature process	0.95 @1310 nm 0.74 @1550 nm	18.5	1310 – 1550	[119]

Structure	Epitaxial growth technique	Responsivity [A/W]	Dark Current Density [mA/cm^2]	Wavelength [nm]	Reference
Ge-on-Si mesa p-i-n detector	LEPECVD no buffer	0.4 @1550 nm	0.041	1550	[91]
Ge-on-Si mesa p-i-n detector	UHV/CVD two-temperature process	0.23 @1550 nm	10	1550	[120]
Ge-on-SOI mesa p-n detector	Thermal evaporation	0.002 @1550 nm	2	1550	[116]
Ge-on-Si planar n-i-p detector	RPCVD two-temperature process	0.4 @1550 nm	10	1550	[69]
Ge-on-Si mesa p-n detector	Thermal evaporation	0.11 @1550 nm	200	1550	[117]
Ge-on-Si mesa p-i-n tensile strained detector	UHV/CVD two-temperature process	0.6 @1310 nm 0.52 @1550 nm 0.1 @1620 nm	10	1310 – 1620	[103]

Table IX. Summary of Ge-on-Si photodetectors in the near-infrared grown by hetero-epitaxy.

Thanks to the technological progress in Ge-on-Si heteroepitaxy, many different Ge-on-Si p-i-n or n-i-p photodetectors have been successfully demonstrated in the last 20 years, as shown in Table IX. The performance of these devices is comparable in terms of DCD and responsivity to the InGaAs/Si detectors fabricated by wafer bonding. Although the InGaAs/Si detectors demonstrated a slightly lower leakage current than the Ge-on-Si detectors (probably due to the intrinsic property of the materials themselves), the possibility of introducing Ge in a CMOS process due to its compatibility and to fabricate detectors by heteroepitaxy using the standard reactor used in CMOS represent a big advantage for this class of near-infrared detector.

3.6.2 Waveguide integrated Ge-on-Si photodetectors

The demonstration of the integration of Ge-on-Si photodetectors with waveguides is very important for many reasons. First of all, this configuration gives the possibility to overcome the bandwidth-efficiency trade-off in normal incidence detectors. The light signal is delivered to the device by an in-plane optical waveguide rather than top-down, permitting the bandwidth and efficiency to be determined almost independently. The efficiency is independent of the thickness of the absorbing layer, but is instead

dependent on waveguide length. Secondly, this configuration represents a further step on the monolithic integration of Si photonics components with electronics device for large scale integration.

In 2007, Ahn *et al.* demonstrated a Ge p-i-n photodetector that was monolithically integrated with silicon oxynitride (SiON) and silicon nitride (SiN_x) waveguides [90]. Devices (structure in Figure 3.18) were grown in a UHV CVD reactor using the two-temperature process. Devices were defined by an etching process and then SiO_2 was deposited followed by a CMP step to planarise the top surface. SiON or SiN waveguides were fabricated on top of the devices.

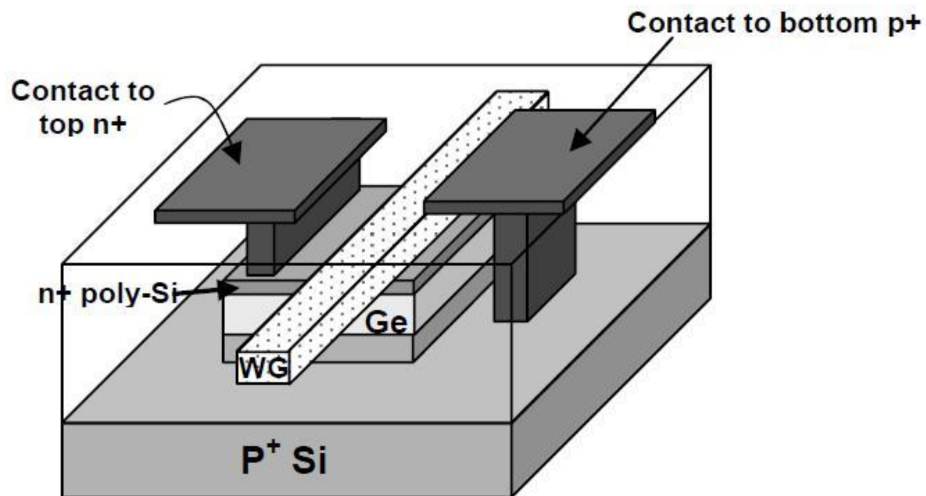


Figure 3.18. Schematic structure of the waveguide integrated Ge p-i-n photodetector [90].

The measured DCD at -0.5 V was 410 mA/cm^2. The very high DCD obtained was mainly due to surface leakage component from the sidewalls of the device. A higher responsivity at 1550 nm of 1.08 A/W for the SiN coupled detector was measured, wavelength, compared to the SiON waveguide that demonstrated a value of 0.96 A/W. The measured responsivity, in both cases, was higher than the responsivity for the normal-incidence p-i-n Ge photodetector which was estimated to be 0.45 A/W.

A thin-film-Ge lateral p-i-n photodetector integrated on a Si waveguide was demonstrated by Wang *et al.* in 2008 [121]. The device structure is illustrated in Figure 3.19.

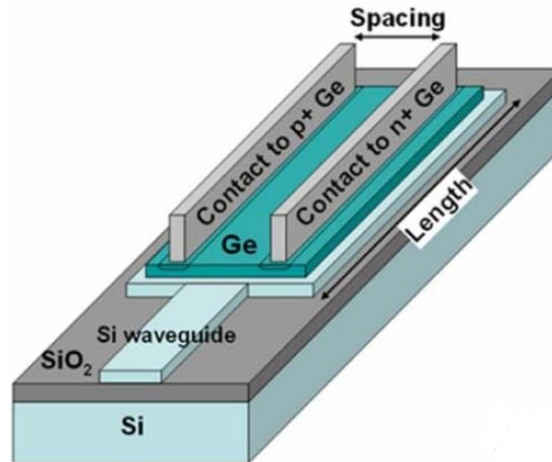


Figure 3.19. Schematic structure of the lateral Ge p-i-n photodetector integrated on a Si waveguide [121].

After defining the Si waveguide on top of a SOI substrate, the Ge was grown in a UHV CVD reactor using the two-step process and a SiGe buffer layer. The detector width was kept constant at 2.4 μm, while its length ranged between 5 and 20 μm. The top Ge surface was implanted with B and P to define the p and n-type contacts, respectively. Different length devices showed a high DCD of ~125 mA/cm^2. The main contribution to the dark current was verified experimentally by low temperature measurements and was mainly due to thermal generation and recombination of carriers in the intrinsic Ge layer. A low responsivity of 0.13 A/W at -1 V and 1550 nm wavelength was measured due to the SiGe buffer layer and intermixing of Si and Ge which resulted in a lower Ge absorption coefficient.

Masini *et al.* in collaboration with Luxtera demonstrated the monolithic integration of a waveguide Ge p-i-n photodetector with CMOS electronics for high-speed optical transceivers [122]. The Ge-on-Si heteroepitaxy was performed using a RPCVD reactor. A high dark current of 3 μA at -1 V was measured due to the defects at the Si/Ge interface. The responsivity at 1550 nm wavelength was 0.85 A/W. The integration with a CMOS receiver also demonstrated a speed up to 20 GHz.

Later in 2009, Vivien *et al.* also demonstrated a p-i-n Ge photodetector integrated in a submicron SOI rib waveguide, which is illustrated in Figure 3.20 [86].

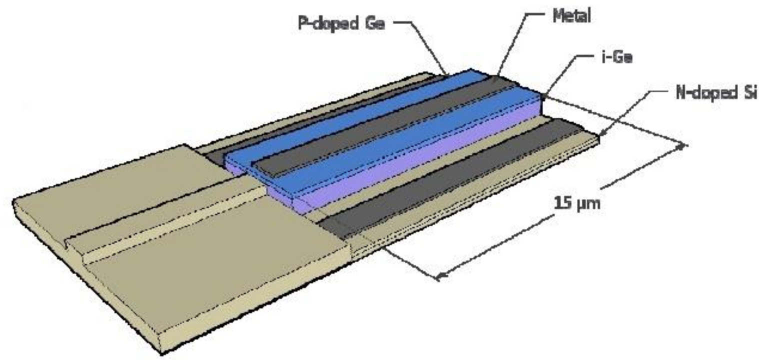


Figure 3.20. Schematic view of p-i-n Ge photodetector integrated in a SOI rib waveguide [86].

A 450 nm-thick Ge layer was grown in a RPCVD reactor using the two temperature process. A 3 µm wide and 15 µm-long mesa detector was defined and passivated with SiO_2. The measured DCD at -1 V was 60 mA/cm^2, while the responsivity at -0.5 V and 1550 nm wavelength was 1 A/W.

In 2009, Feng *et al.* demonstrated a vertical p-i-n Ge photodetector integrated on a Si waveguide [123]. A schematic of the structure is shown in Figure 3.21.

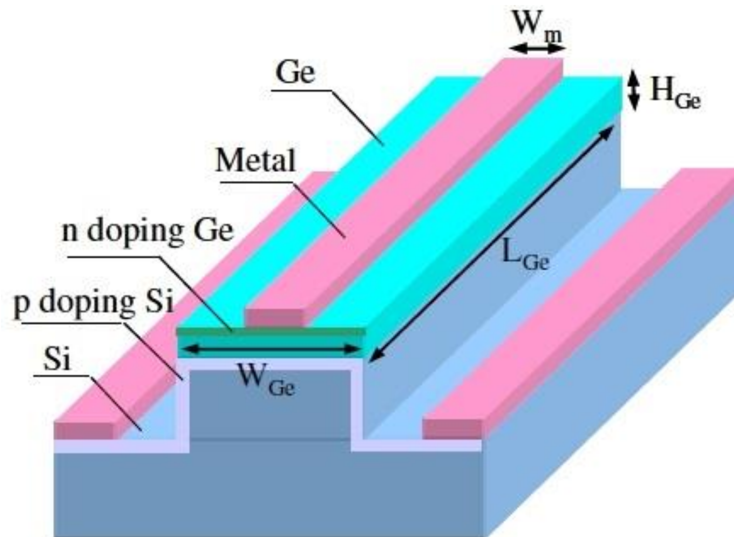


Figure 3.21. Schematic view of a vertical p-i-n Ge waveguide photodetector integrated on top of a large core SOI waveguide [123].

An SOI substrate was used and the single mode waveguide was formed by etching 1.2 µm thick Si with a width of 3 µm. Then the wafer was implanted with boron to define the p-type contact and annealed at 1050 °C to activate the dopant. The 1.2 µm-thick Ge was selectively grown on the Si waveguide using the two-step method. Then it was reduced to 0.92 µm through a chemical mechanical polishing (CMP) step. The top Ge

was implanted with P to form the n-type contact. The measured DCD at -0.5 V for a device 200 μm-long and 3.5 μm wide was 28 mA/cm^2. The responsivity was measured for both TE and TM polarisation, and for the TE coupling a responsivity of 0.5 A/W and 0.72 A/W was measured at 1580 and 1600 nm wavelengths, respectively. It was 0.8 A/W at 1580 nm for TM polarisation.

A lateral Ge p-i-n photodetector fabricated at the end of a Si waveguide was recently reported by Vivien *et al.* [124]. To reduce the TDD due to the lattice mismatch, the Ge was selectively grown in a RPCVD reactor using the two-temperature process. Devices showed a high DCD of 80 A/cm^2 due to dopant diffusion. The measured responsivity was 0.8 A/W at 1550 nm wavelength.

Table X summarises the performance of Ge-on-Si photodetectors integrated on different waveguide geometries.

Structure	**Epitaxial growth technique**	**Responsivity [A/W]**	**Dark Current Density** [mA/cm^2]	**Wavelength [nm]**	**Reference**
Waveguide Ge-on-Si vertical p-i-n	UHV/CVD two-temperature process	1.08 @SiN WG 0.96 @SiON WG	410	1550	[90]
Waveguide Ge-on-Si lateral p-i-n	UHV/CVD two-temperature process with SiGe buffer	0.13 @Si WG	125	1550	[121]
Waveguide Ge-on-Si vertical p-i-n	RPCVD	0.85 @Si WG	NA	1550	[122]
Waveguide Ge-on-Si vertical p-i-n	RPCVD two-temperature process	1 @Si WG	60	1550	[86]
Waveguide Ge-on-Si vertical p-i-n	NA	0.5 @Si WG, TE polarization 0.8 @Si WG, TM polarization	28	1580	[123]
Waveguide Ge-on-Si lateral p-i-n	RPCVD two-temperature process	0.8 @Si WG	80×10^3	1550	[124]

Table X. Summary of various Ge photodetectors integrated on different waveguide geometry.

3.6.3 Ge-on-Si avalanche photodetectors

Avalanche photodiodes (APDs) are high-speed, high sensitivity photodiodes utilising an internal gain mechanism by applying a reverse voltage. Compared to p-i-n photodiodes, APDs can measure even lower light levels and are used in a wide variety of applications requiring high sensitivity such as long-distance optical communications. The operating principles of APDs are described in chapter 2. Although APDs based on the III-V semiconductors for near infrared wavelength have been already commercialised, during the last 5-10 years a lot of effort has been put into researching high-speed and high efficiency APDs based on the Ge-on-Si heteroepitaxial system.

The first Ge-on-Si APD was demonstrated by Kang *et al.* in 2008 [125]. The proposed structure was based on the SACM APD, as shown in Figure 3.22.

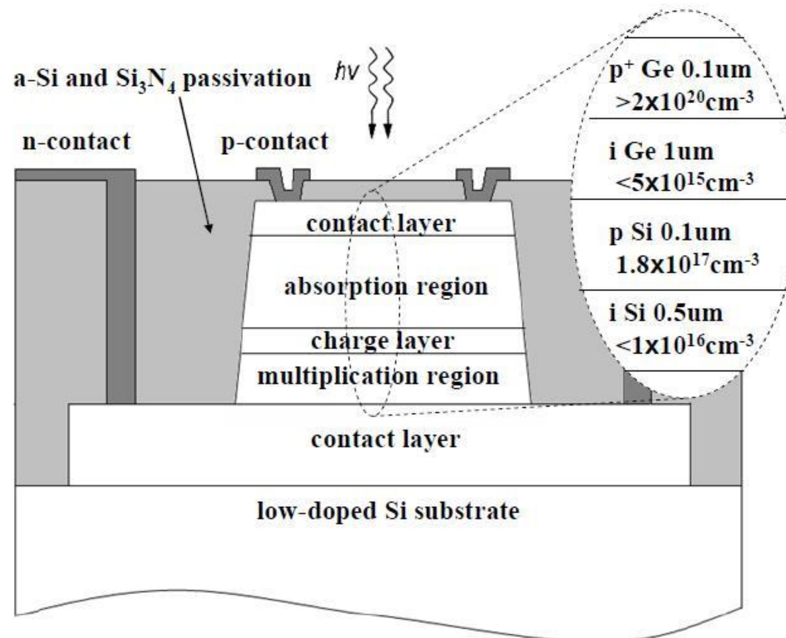


Figure 3.22. Schematic cross-section of the Ge-on-Si SACM APD proposed by Kang et al. [125].

The structure was grown in a CVD reactor by using the two-step process for Ge heteroepitaxy on Si. Circular mesas of diameter ranging between 10 and 200 μm were defined through wet and dry etching for Ge and Si, respectively. Amorphous silicon (a:Si) and silicon nitride were used for passivation. The punch-through voltage V_{PT}, which is the voltage at which the depletion region extends into the Ge absorption layer, occurred between -12 and -20 V. The breakdown voltage V_{BD} (defined at a dark current of 100 μA) was -25.8 V. A high DCD of 237 mA/cm^2 at 90% of V_{BD} was measured. The primary responsivity (at gain = 1) was 0.54 A/W at 1310 nm wavelength. The

measured breakdown voltage thermal coefficient ($\delta = (\Delta V_{BD}/V_{BD})/\Delta T$) was 0.05 %/°C. A gain-bandwidth product of 153 GHz was also obtained. These results were later improved by the same authors using a similar structure but with different processing [126]. A double mesa SACM APD was fabricated, which is shown in Figure 3.23.

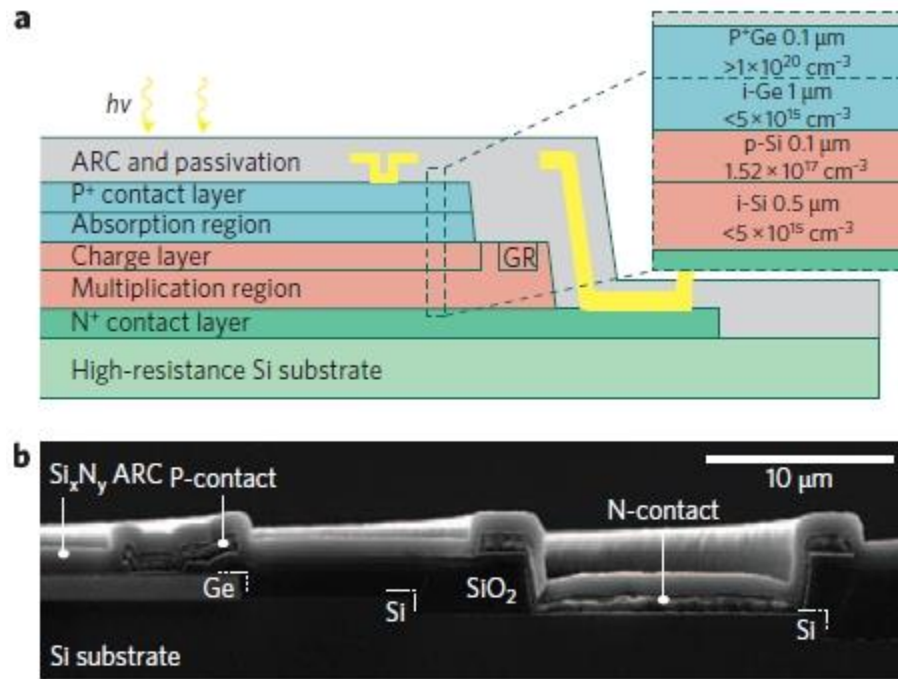


Figure 3.23. (a) Schematic and (b) SEM cross-section of a double mesa Ge-on-Si SACM APD [126].

A floating guard ring (GR), with various distances (1 – 3 μm) between the guard ring and the mesa edge, was also introduced to prevent premature breakdown along the device perimeter. The measured V_{PT} and V_{BD} voltages were -22 and -25 V, respectively. The DCD at 90% of V_{BD} was 175 mA/cm^2. The primary responsivity was 5.88 A/W at 1310 nm wavelength and the highest gain-bandwidth product of 340 GHz was obtained.

A Ge-on-Si SACM APD was demonstrated by Carroll *et al.* [127], as illustrated in Figure 3.24.

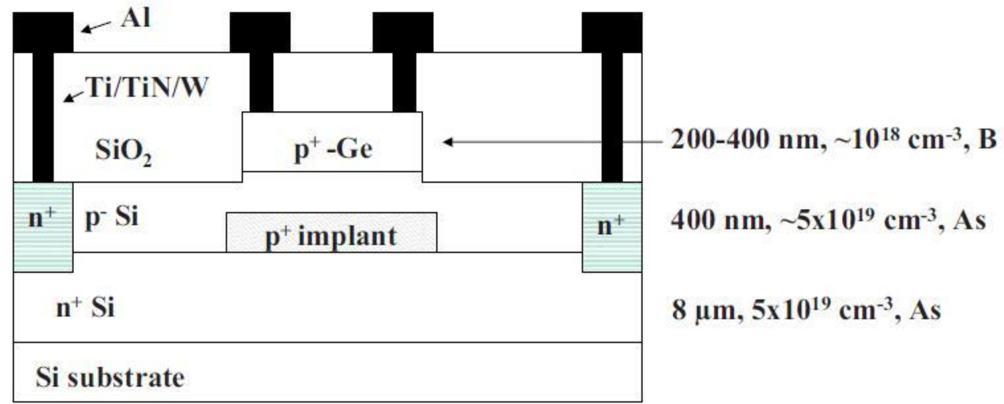


Figure 3.24. Schematic of the Ge-on-Si SACM APD [127].

The structure was grown in a high-density plasma CVD (HDP-CVD) reactor. The Ge epitaxy was performed at low temperature after the Si implant. Because of the high TDD (~5 × 10^{10} cm^{-2}) the Ge was heavily p-type doped to minimise the effect of threading dislocations, which are supposed to be acceptor type [128], and the depletion region width inside it. A V_{PT} of -3V with a V_{BD} equal to -28 V were measured from the I-V characteristic. The calculated DCD at 90% of V_{BD} was 100 mA/cm^2 and the primary responsivities were 3.2 × 10^{-4} and 4.5 × 10^{-5} A/W at 1310 and 1550 nm wavelengths, respectively. The low DCD was obtained because the depletion region was mainly confined in the Si layer causing also low absorption at the near-infrared wavelengths.

A Ge-on-Si SACM APD similar to the structure reported in [125] was demonstrated by Xue *et al.* [129]. Compared to the structure in Figure 3.24, the Si thickness was increased to 700 nm and the charge sheet was obtained by implanting the intrinsic Si with boron. The Ge epitaxy was performed in a UHV CVD reactor using the two temperature process. After the growth, the Ge was implanted to define the p-type contact. SiO_2 and Si_3N_4 were used for the passivation and anti-reflection coating, and circular mesas of diameter ranging between 25 and 70 µm were defined. Devices exhibited a V_{PT} of -29 V and V_{BD} of -39.5 V. The measured DCD at 90% of V_{BD} was 133 mA/cm^2, and the primary responsivity at 1310 nm wavelength was 0.5 A/W.

In 2009, Zhu *et al.* proposed a novel design for a Ge-on-Si SACM APD integrated in a waveguide structure [130], which is shown in Figure 3.25.

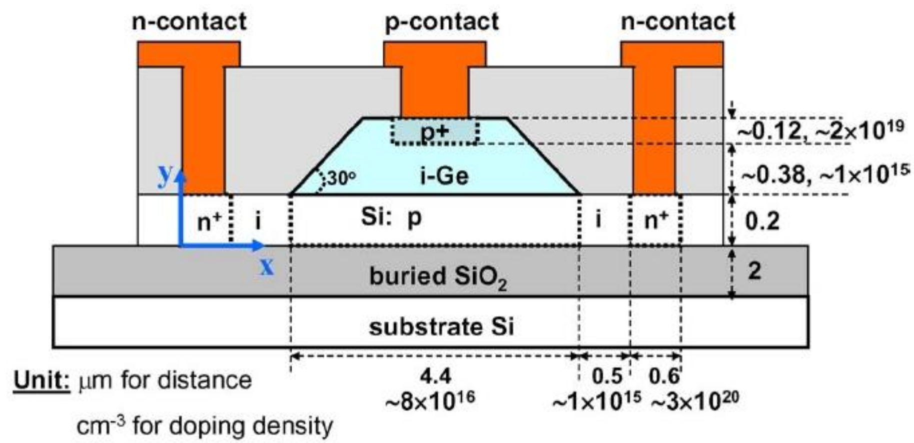


Figure 3.25. Schematic cross-sectional view of the waveguide Ge-on-si SACM APD [130].

Starting from an SOI substrate (200 nm-thick Si device, and 2 μm-thick SOI), the p-Si charge and i-Si multiplication region were located laterally in the Si waveguide whereas the absorbing Ge layer was selectively grown (SEG) on the p-Si charge region using a SiGe buffer layer. The Ge layer was then implanted to define the p-contact region and SiO_2 was used for passivation. Although a new geometry was demonstrated, devices exhibited a high dark current and it was difficult to distinguish the Ge depletion from the Si multiplication process. However, the same research group also reported a normal incidence Ge-on-Si SACM APD [131]. In this structure a thicker Ge layer of 1 μm was grown by SEG. Amorphous Si was used for passivation and Si_3N_4 as anti-reflection coating. For a 60 μm-diameter device, the V_{PT} and V_{BD} voltages measured from the I-V characteristic were -17 and -27 V, respectively. The calculated DCD at 90% of V_{BD} was 22 mA/cm^2 with a primary responsivity of 0.42 A/W at 1310 wavelength. 80 GHz gain-bandwidth product was also calculated.

After Kang *et al.* had reported two different normal incidence SACM APDs in 2008, they also proposed and demonstrated a waveguide integrated Ge-on-Si APD in 2009 [132], as shown in Figure 3.26.

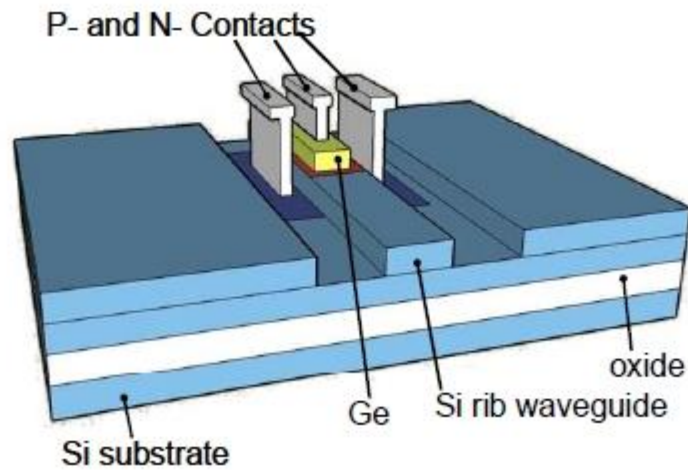


Figure 3.26. Device schematic layout of a WG Ge-on-Si APD [132].

An SOI substrate was used, and the same structure as the vertical APD was grown but with a thinner Ge layer of 600 nm. These WG APDs were 40 to 100 μm long with a junction area 3 to 7 μm wide. These devices were integrated on a multimode Si rib waveguide 6 μm wide and 2 mm long. Devices exhibited a V_{BD} of -24.5 V with primary responsivities of 0.9 and 0.6 A/W at 1310 nm and 1550 nm wavelengths, respectively. The measured DCD at 90% of V_{BD} was around 1 mA/cm^2. They also found that the dark current was not dominated by the bulk component (as in the mesa devices) but that the perimeter played an important role on these small devices.

In 2010, Ang *et al.* reported a waveguide integrated Ge-on-Si SACM APD [133], as illustrated in Figure 3.27.

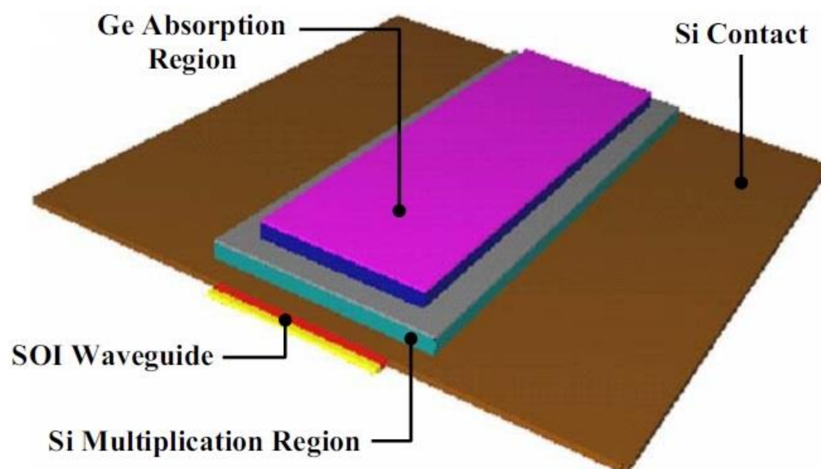


Figure 3.27. Schematic illustration of the WG integrated Ge-on-Si SACM APD proposed in [133].

An SOI substrate (220-nm thick Si device, 2 μm-thick SiO_2) was used and ion implanted to define the n-contact region. A 500 nm-thick intrinsic Si was selectively grown and then implanted with boron to define the p-Si charge layer. 500 nm-thick Ge was selectively grown in a UHV/CVD reactor. For the p contact region 100 nm-thick amorphous Si was grown and implanted with boron. The V_{PT} and V_{BD} voltages extrapolated from the I-V characteristic were -7.5 and -22.5 V, respectively and the DCD at 90% of V_{BD} was 400 mA/cm^2. Finally, a primary responsivity of 0.8 A/W was calculated at 1550 nm wavelength with a gain-bandwidth product of ~105 GHz.

Duan *et al.* demonstrated a vertical incidence Ge-on-Si SACM APD grown on an SOI substrate (220-nm thick Si device, 2 μm-thick SiO_2) in 2012 [134]. A 1 μm-thick Ge absorber layer was selectively grown in a UHV/CVD reactor using the two-temperature process and a SiGe buffer layer and amorphous silicon was deposited for passivation. The V_{PT} and V_{BD} were -10 V and -29.4 V, respectively. The calculated DCD at 90 % of V_{BD} was 282 mA/cm^2 and the measured primary responsivity at 1550 nm wavelength was ~0.3 A/W. A high gain-bandwidth product of 310 GHz was also reported. The same authors have recently reported, in 2013, the same Ge-on-Si SACM APD integrated on a waveguide [135], which is illustrated in Figure 3.28.

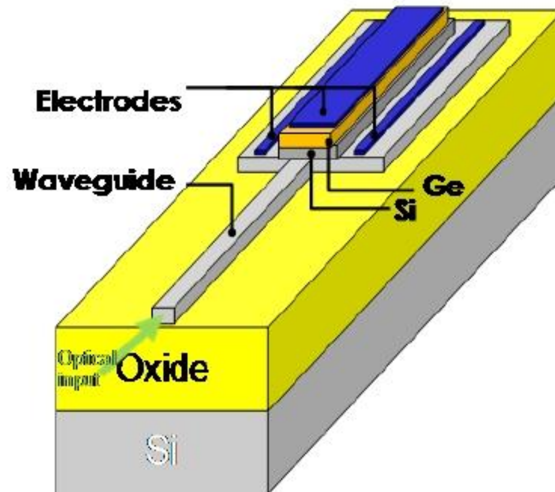


Figure 3.28. Schematic structure of the WG integrated Ge-on-Si SACM APD as reported in [135].

Devices exhibited a V_{PT} and V_{BD} of -27 and -30 V, respectively. A high DCD of 1250 mA/cm^2 was measured at 90% of V_{BD}.

Table XI summarises the performance of vertical-incidence and waveguide-integrated Ge-on-Si avalanche photodetectors.

Structure	Epitaxial growth technique	V_{PT} (V)	V_{BD} (V)	Primary Responsivity [A/W] (at gain = 1)	DCD [mA/cm^2] (at 90% V_{BD})	Gain – Bandwidth product (GHz)	Reference
Ge-on-Si SACM APD	UHV/CVD two-temperature process	-12	-25.8	0.54 @1310 nm	237	153	[125]
Double mesa Ge-on-Si SACM APD	UHV/CVD two-temperature process	-22	-25	5.88 @1310 nm	175	340	[126]
Ge-on-Si SAM APD	HDP-CVD	-3	-28	0.00032 @1310 nm 0.000045 @1550 nm	100		[127]
Ge-on-Si SACM APD	UHV/CVD two-temperature process	-29	-39.5	0.5 @1310 nm	133		[129]
Ge-on-Si SACM APD	SEG with SiGe buffer	-17	-27	0.42 @1310 nm	22	80	[131]
WG Ge-on-Si APD	UHV/CVD		-24.5	0.9 @1310 nm 0.6 @1550 nm	1		[132]
WG Ge-on-Si SACM APD	UHV/CVD SEG	-7.5	-22.5	0.8 @1550 nm	400	105	[133]
Ge-on-Si SACM APD	UHV/CVD two-temperature process with SiGe buffer	-10	-29.4	0.3 @1310 nm	282	310	[134]
WG Ge-on-Si SACM APD	UHV/CVD two-temperature process with SiGe buffer	-27	-30		1250		[135]

Table XI. Summary of vertical-incidence and waveguide-integrated Ge-on-Si avalanche photodetectors.

3.6.4 SPADs operating at infrared wavelengths

The first SPAD proposed in 1980 was Si-based. As shown by Buller and Collins in [136] and by Cova *et al.* in [137], Si SPADs are widely used in a number of photon

counting application areas. These devices, fabricated in a planar geometry, and compatible with standard CMOS technology, have shown good single-photon performance at room temperature, demonstrating SPDEs of 52% and 12% at 550 nm and 850 nm wavelengths, respectively. Low DCR was demonstrated around 300 cs^{-1} and lower, with NEPs reported ranging between 10^{-17} and 10^{-18} $WHz^{-1/2}$ at room temperature and slightly below (~230 K). Although nowadays Si SPADs are commercialised by different companies, such as PerkinElmer or Micro Photon Devices (MPD), their photon-counting application areas are limited in the spectral region between 400 – 1000 nm. This limitation has already been clarified in Figure 3.2, where the absorption coefficient of Si drops substantially at the infrared wavelengths due to its energy gap of 1.1 eV, which makes it unusable for detecting single-photons at these wavelengths.

On the other hand, there are a number of applications that benefit from photon counting techniques in the NIR wavelength bands, like quantum key distribution (QKD) [138], time-of-flight ranging [139], singlet oxygen detection for photodynamic therapy (PDT) dosimetry [140] and many others. Due to their low energy gap, as shown in Figure 3.2, semiconductors usable at the NIR wavelengths are Ge and InGaAs. An intrinsic drawback is the high DCR caused by the higher thermal generation of carriers (which is higher in Ge than InGaAs) than in silicon. For this reason, devices must be cooled well below room temperature to reduce the DCR to a useable level. Germanium APDs have to operate at cryogenic temperatures so that the effect of thermal generation on DCR is negligible, and the sensitivity is impaired by trapping centers and tunnelling. This last effect is negligible in Si SPADs due to the larger energy gap than germanium where it dominates at low temperatures as demonstrated in [141] and [142]. In particular, in the case of Ge, commercially available APDs have been used in Geiger-mode operation to detect single-photons in the infrared regime.

The first demonstration in 1976 by Fichtner and Hacker reported Ge APDs operating in the nanosecond range [143]. First in 1992 and then in 1994, Lacaita *et al.* studied the single-photon performance at 1310 nm wavelength of several commercially available germanium APDs from different suppliers such as Fujitsu (30-μm diameter), Judson (100-μm diameter) and Siemens (50-μm diameter) reporting timing performance below 1 ns [142]. All the tested devices were p^+ - n planar photodiodes, with a typical cross-section as shown in Figure 3.29, and a breakdown voltage varying between -33.5 and -31.5 V at room temperature. Due to the high DCD, devices were cooled and tested at 77

K, verifying a decrease of the breakdown voltage from between -33.5 and -31.5 V to -23.5 and -22.3 V as the temperature decreased from 300 to 77 K. This shift of the breakdown voltage with decreasing temperature of operation is common to every APD due to the reduction of lattice vibrations and therefore enhanced ionisation coefficients. For all tested devices, the authors verified that afterpulsing, caused by carrier trapping in the high field region of the junction, increased the DCR and impaired the achievable sensitivity. The best performance was obtained by the 30-μm diameter device produced by Fujitsu, which demonstrated a SPDE of 3.5 % with a NEP of 7.5×10^{-16} $WHz^{-1/2}$ at 0.2 V of excess bias and 1310 nm wavelength. In terms of timing performance, these devices demonstrated a timing jitter of 100 ps FWHM when operated at 3 V excess bias.

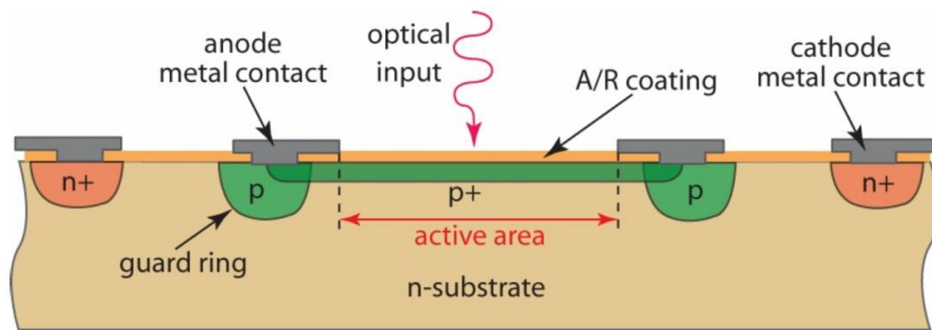


Figure 3.29. Typical p^+n planar Ge APD cross-section [144].

Similar research was also carried out by Fancey in his PhD thesis in 1996 [145]. The Ge APDs characterised in this work were also commercially available planar p^+ - n junctions supplied from different companies such as Fujitsu (30-μm diameter), NEC (30-μm diameter) and GPD Optoelectronics (30-μm diameter). For the same reasons as explained above, characterisation was performed at 77 K. An improvement of the measured SPDE at 1310 nm between 5 and 15 % was obtained, also showing a linear behaviour with excess bias, for small values of over-bias.

The noise performance of the APDs was also evaluated to find out the major noise contribution at the operating temperature. At 77 K the contributions from thermal generation and tunnelling were small, as also demonstrated by Haitz [146], leaving the re-emission from traps of minority carriers as the main noise mechanism. At this temperature the trap state lifetime will increase with decreasing temperature [147]. APDs were tested using an active quenched circuit (AQC) and biased 0.6 V above $V_{BD.}$ The gate repetition rate was varied between 500 Hz and 100 kHz. The lowest noise performance was obtained from the GPD APD at 1 kHz, with mean trap lifetimes calculated between 200 and 500 μs. This device also showed the lowest NEP of ~2 ×

10^{-16} $WHz^{-1/2}$ at the same repetition rate and 1310 nm wavelength. In terms of timing performance, the FWHM were measured at 1310 and 1550 nm wavelengths, and were found to be 310 and 680 ps, respectively. The increase in response time with wavelength was due to the decrease in absorption coefficient, increasing the absorption length of Ge. This led to those carriers photogenerated beyond the depletion layer taking more time to reach the depletion region and consequently increasing the FWHM jitter of the detectors.

A more complete characterisation in terms of single-photon performance on p^+n planar commercially available Ge APDs was carried out by Tosi *et al.* in 2007 [144] on two different Ge APDs from GPD Optoelectronics (40-μm diameter, V_{BD} = 19 V at 77 K) and from Texas Instruments (TI) (350-μm diameter, V_{BD} = 29 V at 77 K). Measurements were performed in gated mode at a gate frequency of 1 kHz with fixed gate-on duration (T_{ON}) of 20 ns and very long gate-off period (T_{OFF}) of 1 ms to avoid afterpulsing effects. The TI APD showed the lower DCR of 700 kcs^{-1} at 0.5 V of excess bias and 77 K and it was further reduced to 250 kcs^{-1} at 50 K. Although the TI APD had a bigger area than the GPD device, its lower DCR showed a better fabrication process that introduced fewer lattice defects than the GPD device. In terms of SPDE these devices were characterised at 1310 and 1550 nm, showing SPDEs of 30 % and 1 % at these wavelengths, respectively. Furthermore, the devices were characterised in terms of afterpulsing by varying the repetition rate between 1 kHz and 250 kHz (with T_{OFF} ranging from 1 ms to 4 μs). This is illustrated in Figure 3.30 for the GPD device.

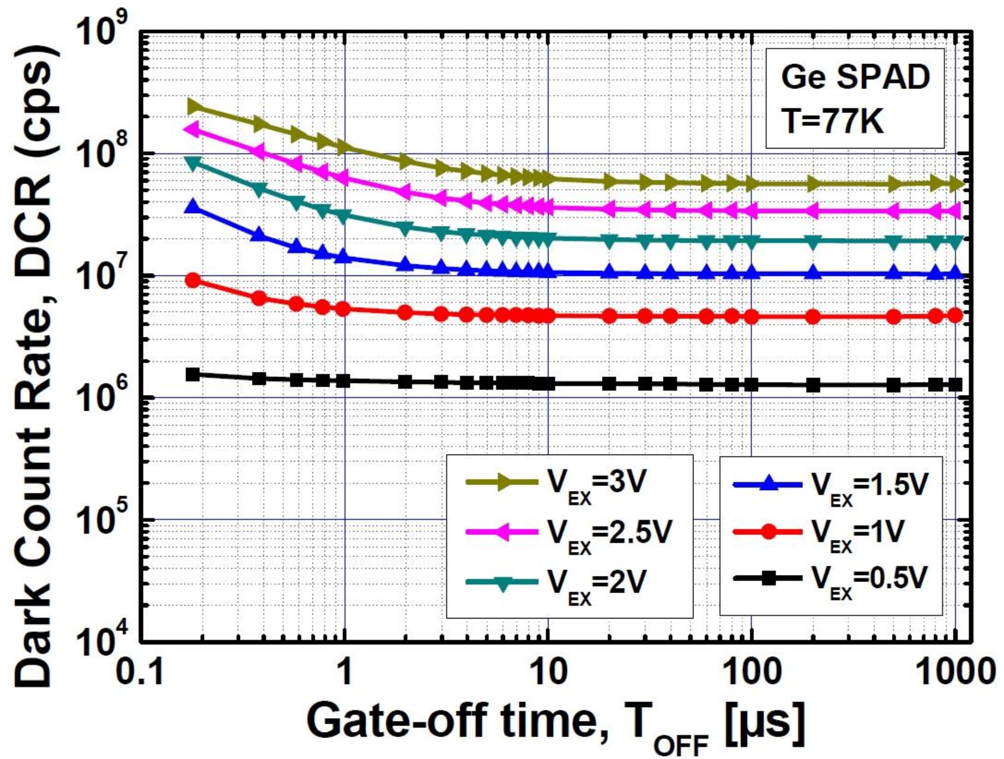


Figure 3.30. DCR of Ge GPD measured at various excess biases and 77 K in gated operation (T_{ON} = 20 ns) as a function of the T_{OFF} time [144].

The primary DCR is measured at sufficiently long T_{OFF} (low repetition rate), when afterpulsing plays a negligible role and the DCR is constant. A similar behaviour was recorded for the TI device. As shown in Figure 3.30, within the measured T_{OFF} range there was no significant increase of the DCR for different excess biases, meaning that the trap time constant was shorter than 4 μs for these devices. As it has already been mentioned and as will be further explained below, this behaviour represents a big advantage when compared to the InGaAs/InP SPAD. The calculated NEPs at 77 K for the GPD APD were ~4.1 × 10^{-15} $WHz^{-1/2}$ and ~3.2 × 10^{-15} $WHz^{-1/2}$ at 0.5 V and 1.5 V of excess bias, respectively. In terms of timing jitter these devices had a similar behaviour as devices tested by Lacaita and Fancey [142], [145].

In the last 20 years, only commercially available Ge APD have been used to detect single-photon in the infrared region, without any proposal for a custom designed Ge SPAD. Perhaps, this was mainly due to their poor performance when compared to their III-V counterparts. An InGaAs/InP SPAD, typically comprising of separate absorption, grading, charge, and multiplication (SAGCM) heterostructure is shown in Figure 3.31. These devices are used for time correlated single-photon counting (TCSPC) typically at 1550 nm, and they can efficiently detect photons up to 1700 nm in wavelength.

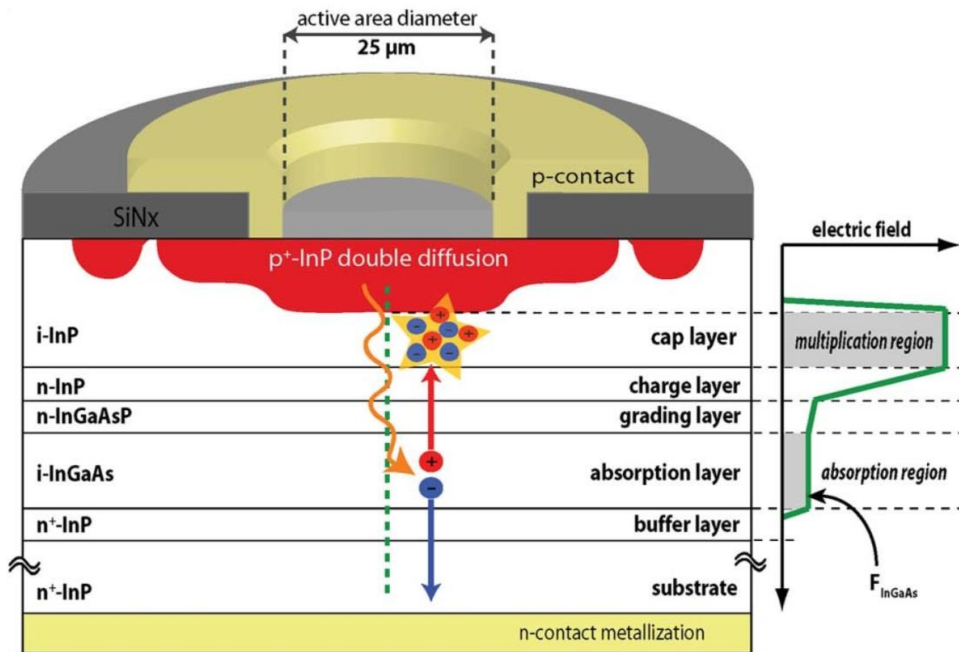


Figure 3.31. Typical planar SAGCM InGaAs/InP SPAD cross-section, with double dopant diffusion and floating guard ring. The electric field along the centre of the active area is also shown [148].

In the last 30 years, their single-photon performance has been extensively studied by many research groups including our group at Heriot-Watt University [149], [150] which worked on the design, modelling and characterisation of these detectors.

Referring to Figure 3.31, the photogenerated electron-hole pair in the narrow band gap $In_{0.53}Ga_{0.47}As$ (E_g ~0.75 eV at 300 K) layer is separated by the electric field, the hole drifts to the InP multiplication region (E_g ~1.35 eV at 300 K), where it can trigger a self-sustaining avalanche. The charge layer between the absorption and multiplication regions, is designed to maintain a low electric field in the narrow bandgap absorber (to avoid tunnelling), while attaining high electric field in the multiplication region (to have impact ionisation). The addition of grading layers between InGaAs and InP is important to reduce carrier (hole) pile-up effects, which result from the valence band offset (~0.4 eV) that arises in an abrupt InGaAs/InP heterojunction [151], [152]. Double p-type (Zn) diffusion confines the high field region and reduces edge effects. The double p-type (Zn) diffusion was first introduced by Liu *et al.* in 1992 [153], and became a standard design for the fabrication of planar InGaAs/InP detector. This novel structure introduced two mechanisms to eliminate the edge breakdown:

1) FGRs that reduced both edge and surface electric fields;
2) A stepped junction edge formed by the double Zn diffusion that enhanced the electric field in the central active region.

Typical current-voltage characteristic for these devices is shown in Figure 3.32.

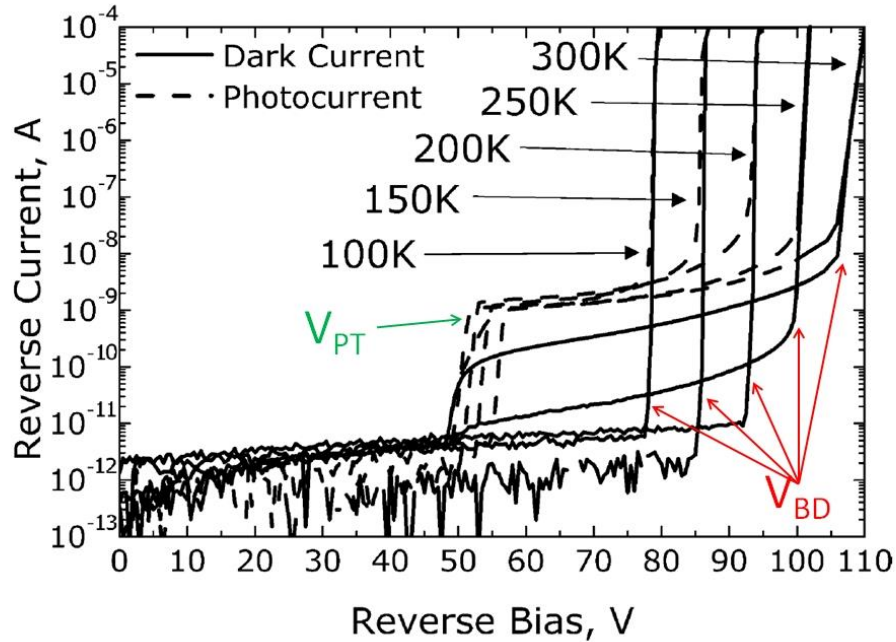


Figure 3.32. Typical dark current (solid lines) and photocurrent (dashed lines) at different temperatures for an InGaAs/InP SPAD [154].

As the temperature decreases the V_{BD} value decreases with a temperature coefficient γ that depends on the semiconductor used to define the multiplication region ($\gamma = 0.17$ V/K for InP). The V_{PT} value is unchanged (~50 V) because it depends only on the doping profile of the junction.

In 1994, Zappa *et al.* reported the single-photon performance of an InGaAs/InP APD (50-μm diameter, V_{BD} = ~90-110 V at 290K) produced by EG&G. Devices were operated in Geiger mode with a gate-on duration of ~250 ns [152]. An exponential dependence of the DCR with the temperature was measured. The measured SPDE at 1310 nm had a maximum of 1 % at 6 V of excess bias and 150 K. Consequently, at the same temperature the minimum NEP was 1×10^{-14} $WHz^{-1/2}$ with a jitter FWHM of ~1 ns. A lower NEP of 2.7×10^{-16} $WHz^{-1/2}$ and improved timing performance (~200 ps) at 2 V of excess bias were later obtained, in 1996, by the same research group by testing a commercially available Fujitsu InGaAs/InP (30-μm diameter) APD at 77 K [155].

Similar results were also demonstrated by Hiskett *et al.* characterising two commercially available Fujitsu InGaAs/InP APDs (80-μm diameter and 30-μm diameter). The NEP was calculated over the range of temperatures from 77 K to 225 K for both devices. The lowest NEP of 4.4×10^{-16} $WHz^{-1/2}$ at 5 V of excess bias was measured on the 80-μm device at 77 K.

Although different authors demonstrated photon-counting with commercially available APDs, there are a number of differences in terms of performance requirements between an APD and a SPAD that must be underlined. Firstly, not all APDs are suitable for photon counting application, but only devices for which the difference $V_{BD} - V_{PT} \geq 30\,V$ at room temperature (which decreases with the temperature, as also shown in Figure 3.32) [149]. If V_{BD} is too close to V_{PT} most of the holes photogenerated in the InGaAs layer fail to cross the heterobarrier and are collected by the guard ring. Secondly, linear mode InGaAs/InP APD should achieve an optimal electric field profile below V_{BD} at modest gain (10-20) since it is within this range of gain that the combination APD and amplifier has the maximum S/N ratio. In contrast, SPADs work above V_{BD} and the electric field profile should be optimised for target overbias. Finally, thinner multiplication regions are desirable in linear mode APDs because they result in more deterministic linear mode avalanche processes and also the APD gain-bandwidth product is inversely proportional to the multiplication region width, as shown in chapter 2. In contrast, SPAD performance can benefit from a much thicker multiplication region (~1 μm) since that breakdown probability increases significantly [156].

In 2006, Pellegrini *et al.* reported the design, fabrication, and performance of planar InGaAs/InP (10 μm diameter) devices specifically developed for single-photon detection at 1550 nm [154]. Device structures were similar to the one shown in Figure 3.31, but two different designs were proposed and grown to smooth the large valence band offset between InGaAs and InP:

1) with a graded region consisting of one quaternary with a bandgap exactly intermediate between InGaAs and InP (SPAD-1Q);
2) with a graded region composed of three sublayers of stepped bandgap (SPAD-3Q).

Devices showed I-V characteristics as reported in Figure 3.32 and were characterised in gated mode with a gate-on duration of 100 ns. The SPAD-3Q showed the best performance demonstrating a SPDE of 10% and a NEP of 6×10^{-16} $WHz^{-1/2}$ at 1550 nm and 200 K. The performance of these devices was comparable with that of commercially available APD from Epitaxx and Fujitsu, showing a lower DCR but also a lower SPDE. Devices were also characterised for afterpulsing as shown in Figure 3.33, highlighting the main disadvantage of this class of detectors represented by the carriers (holes) trapped in the multiplication region, as already explained.

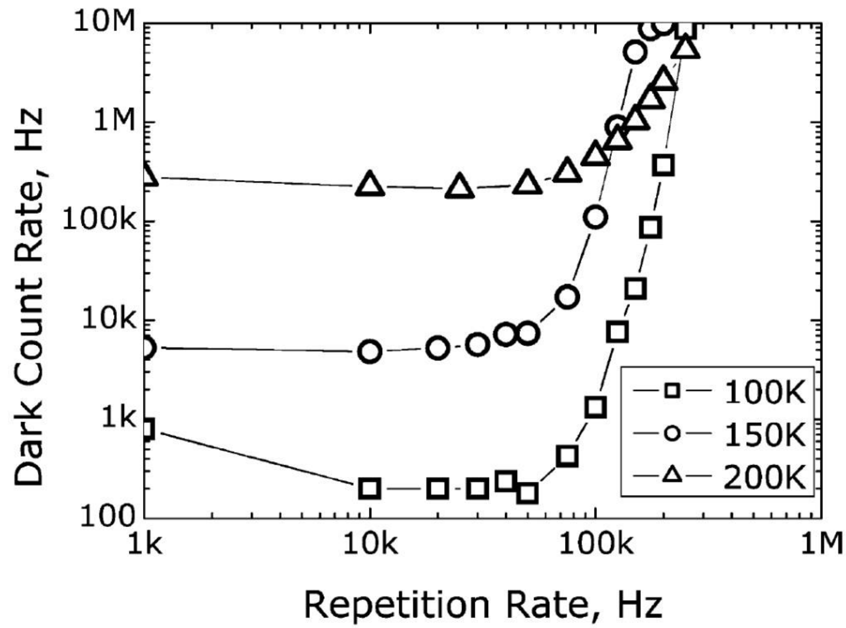


Figure 3.33. DCR as a function of the repetition rate for a 20 μm diameter SPAD-3Q at different temperatures. The DCR increased rapidly with repetition rates higher than 50 kHz [154].

A customised design of InGaAs/InP SPAD was also proposed and demonstrated by Tosi *et al.* with similar result as above [157].

A net improvement in the InGaAs/InP SPAD performance was obtained by Liu *et al.* in 2007 [158]. The tested devices (40 μm diameter) in gated mode (T_{ON} = 4 ns, f_{gate} = 10 kHz) achieved SPDE of 45 % at 1310 nm and DCR of 12 kcs^{-1} at 200 K corresponding to a NEP of 4.5×10^{-17} WHz$^{-1/2}$.

Although InGaAs/InP demonstrated good single-photon performance in terms of high SPDE, low jitter, and operation at temperatures consistent with thermoelectric Peltier cooling, the major drawback of afterpulsing guided researchers toward several different electrical biasing schemes to reduce its detrimental effects. In this direction, the first report of an InGaAs/InP SPAD operating free-running at room temperature was demonstrated by Warburton *et al.* in 2009 [159]. In free-running operation a suitable external circuit senses the onset of the avalanche current, generates a standard output pulse, and quenches the avalanche, by lowering the bias below V_{BD}. However, this mode of operation is possible only with devices that have low DCR. In Ref. [159] a Princeton Lightwave InGaAs/InP SPAD (25 μm device, V_{BD} = 42 V at 270 K) was used in a passive quenching scheme. The value of the series resistor (R_S = 100 kΩ) through which the SPAD is reverse biased was chosen to provide a good compromise in the

trade-off between maximum counting rate and maximum SPDE. Room temperature operation was demonstrated with a DCR of ~1 Mc/s and a NEP of ~8×10^{-15} $WHz^{-1/2}$ at 1550 nm wavelength. With the same device operated in gated mode, room temperature measurements were not possible due to the increased DCR. Results were later improved by using new generation devices from the same company as above that showed DCR of ~40 kc/s and NEP of ~1×10^{-15} $WHz^{-1/2}$ at the same operating conditions [160]. Moreover, a NEP of ~5×10^{-17} $WHz^{-1/2}$ at 1550 nm and 210 K comparable with the one demonstrated from gated mode devices was demonstrated in free-running mode.

InGaAs/InP SPADs represent the state-of-art for detecting single-photons in the infrared region and nowadays are commercialised by different companies such as Princeton Lightwave or MPD. Although the structure of these devices is still similar as the one shown in Figure 3.31, as explained by Acerbi *et al.,* design criteria and growth processes become very important to efficiently detect single-photons in the infrared with low counting rates [148], [161]. Regarding the problem of afterpulsing, it is possible to limit its detrimental effects by various biasing and quenching techniques such as: sine-wave gating [162], self-differencing [163], sinusoidal gating [164], fast AQC [165], etc. However, single-photon detection with InGaAs/InP SPADs is still limited by afterpulsing. On the other hand, Si SPADs demonstrates a negligible afterpulsing, and Si is well known for its good multiplication properties and to be a mature technology, although it cannot efficiently detect photons at wavelengths greater than 1000 nm.

However, Ge APDs used in the Geiger-mode regime can detect single-photons in the infrared and the afterpulsing does not increase as in the InGaAs-based SPADs (although it could be masked from the high level of DCR). As already explained in this chapter, improved approaches in the heteroepitaxy of pure Ge on Si opened an opportunity to design, grow and fabricate a SPAD that can exploit the advantage of both Si and Ge. Although this research is at its early stage, the single-photon performance of these devices will be analysed in chapter 5 and the work done in this thesis represents the first demonstration of a Ge-on-Si SPAD operating at the infrared wavelengths of 1300nm and 1550nm.

For further information, Table XII summarises the single-photon performance of the infrared SPADs analysed in this section, plus the addition of a mesa InGaAs-on-Si APD

fabricated by wafer bonding and characterized in Geiger-mode at 1550 nm by Kang *et al.* in 2004 [63].

Structure	Diameter (μm)	NEP ($WHz^{-1/2}$)	SPDE (%)	FWHW (ps)	Operating temperature	λ (nm)	Reference
Fujitsu p^+n Ge APD	30	7.5×10^{-16}	3.5	100	77 K	1310	[142]
GPD p^+n Ge APD	30	2×10^{-16}	5 - 15	310	77 K	1310	[145]
GPD p^+n Ge APD	40	4×10^{-15}	5 - 30	300	77 K	1310	[144]
EG&G InGaAs/InP APD	50	9×10^{-15}	1	1000	150 K	1310	[152]
Fujitsu InGaAs/InP APD	30	2.7×10^{-16}	5	200	77 K	1310	[155]
Fujitsu InGaAs/InP APD	80	4.4×10^{-16}	15	250	77 K	1550	[166]
InGaAs/InP SPAD	10	1×10^{-16}	10	425	150 K	1550	[154]
InGaAs/InP SPAD	40	4.5×10^{-17}	45	140	200 K	1310	[158]
PLI InGaAs/InP SPAD	25	1×10^{-14}	2	500	290 K	1550	[159]
PLI InGaAs/InP SPAD	25	4.5×10^{-17}	5	450	210 K	1550	[160]
InGaAs on Si APD	150	8×10^{-16}	2.5 - 33		223 K	1550	[63]

Table XII. Summary of performance at the infrared wavelengths of selected Ge, InGaAs/InP, and InGaAs-on-Si single-photon detectors.

3.6.5 Superconducting Nanowire Single-Photon Detector (SNSPD)

As already mentioned in Chapter 2, other technologies exist to detect single-photons such as the superconducting nanowire and the transition edge sensor.

The Niobium Nitride (NbN) SNSPD was introduced by Gol'tsman *et al.* in 1991 [167], and it is single-photon sensitive at visible and infrared wavelengths. Its principle of operation is illustrated in Figure 3.34, where the superconductor is kept at a temperature T below its critical temperature T_C (11.5 K) and direct current biased just below its critical current I_C. The absorption of a photon with energy $\hbar\omega \gg 2\Delta$, where 2Δ is the superconducting energy gap, creates a local nonequilibrium perturbation with a large

number of excited hot electrons (~300 for NbN excited with 790 nm wavelength), and an increase of the average electron temperature above T_C. This results in the formation of a hotspot – a local non-superconducting region of a thermalisation length $2\lambda_T$ (Figure 3.34a). After the initial thermalisation, the resistive hotspot grows (Figure 3.34b) as hot electrons diffuse out of its centre. At the same time, the supercurrent is expelled from the hotspot volume and is concentrated in the "sidewalks" between the hotspot and the edge of the film (Figure 3.34c). If the bias current I_{bias} is sufficient to exceed the critical current in the sidewalks a non-superconducting barrier is formed, causing a resistive strip across the entire width w of the device (Figure 3.34d), which gives rise to a measurable voltage pulse at the detector output. As the wire returns to its operating temperature, the hot spot shrinks, removing the resistive barrier and the wire re-enters its superconducting state.

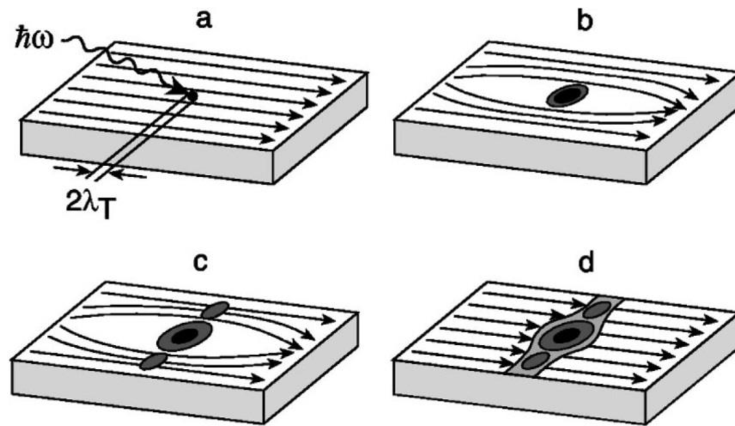


Figure 3.34. Schematics of operation for the NbN SNSPD, which is kept at temperature far below T_C. The arrows indicate direction of the supercurrent flow [168].

SNSPDs have shown very good performance in terms of timing jitter making these detectors very attractive for TCSPC applications. Verevkin *et al.* reported 68 ps FWHM on a 10 × 10 μm^2 meander SNSPD while Pearlman *et al.* reported extremely low time jitter below 18 ps on a similar structure [169]. In order to improve the detection efficiency of these devices, different structures have been proposed, as shown in Figure 3.35. The meander geometry SNSPDs have been developed to improve the coupling efficiency and hence the practical detection efficiency (Figure 3.35a). These NbN meander devices showed system detection efficiency of ~2 – 3 % at 1550 nm, 1 kcs^{-1} DCR at ~3 K [170].

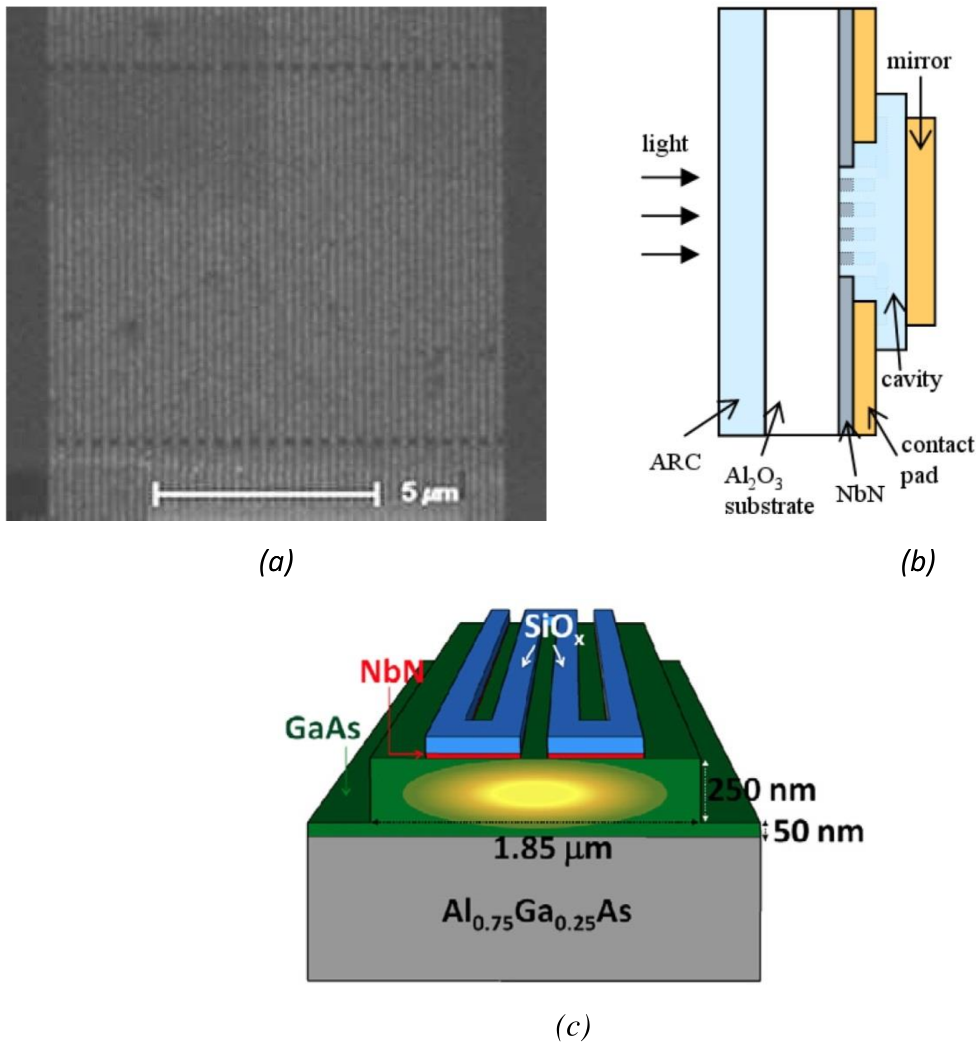


Figure 3.35. Schematic of different SNSPD structures. (a) A scanning electron microscopy (SEM) image of a NbN meander SNSPD covering 10 × 10 µm² area [171]. *(b) A 3 × 3 µm² NbN meander SNSPD integrated with an optical cavity and ARC to reduce loss of photons from reflection and transmission* [172]. *(c) Schematic view of the waveguide integrated SNSPD to improve the optical coupling efficiency* [173].

Cavity and waveguide integrated designs (Figure 3.35b-c) have been employed to boost the absorption efficiency. These devices demonstrated intrinsic detection efficiency of 57 % at 1550 nm and 1.8 K [172]. Miki *et al.* also reported high system detection efficiency of ~24 – 28 % at 1550 nm, and ~23 – 40 % at 1310 nm by using these device configurations [174].

Sprengers *et al.* demonstrated the integration of a NbN SNSPD with a GaAs ridge waveguide (Figure 3.35c). These devices were characterised at 1300 nm showing a

system detection efficiency of ~3.4 % for a 50-μm long device. A DCR of ~100 kcs^{-1} was reported with a 60 ps FWHM jitter.

To improve the detection efficiency, authors have used different superconducting materials. In 2013, Marsili *et al.* reported system detection efficiency over 90 % at 1550 nm by fabricating a SNSPD based on amorphous tungsten silicide ($W_{0.75}Si_{0.25}$, or WSi) superconductor [175]. The WSi SNSPD was embedded in an optical stack designed to enhance absorption at 1550 nm and coupled to single-mode optical fibre. Moreover, it was characterised at 1 K due to its lower critical temperature (3.7 K) than the NbN superconductor. The measured DCR was ~1 kcs^{-1} with a higher timing jitter of ~150 – 200 ps than the NbN counterpart.

In terms of noise performance, SNSPDs showed low DCR, and empirically the DCR rises exponentially as I_{bias} approach I_C, but the origin of this exponential behaviour is poorly understood.

Although at infrared wavelengths SNSPDs have slightly better performance (low DCR and better jitter) than InGaAs/InP, their main drawbacks that limited their use to a laboratory setup are represented from the low temperature of operation and light coupling on a small area (10 μm × 10 μm) device.

3.6.6 Transition Edge Sensor

A superconducting transition-edge sensor (TES), also called a superconducting phase-transition thermometer (SPT), consists of a superconducting film operated in the narrow temperature region between the normal and superconducting state, where the electrical resistance varies between zero and its normal value [176]. A TES thermometer can be used in bolometry (to measure power) or in calorimetry (to measure a pulse of energy). Superconducting TESs, with tungsten (W) as the active device material, are microcalorimeters that have photon-number resolution with negligible dark count. A general microcalorimeter device consists of an absorber for the incident energy, a thermometer to measure the temperature increase resulting from the absorption of energy, and a weak thermal link that enables the cooling of the absorber to its base temperature once the measurement of device temperature is complete [177]. Tungsten is used in the visible and near-infrared wavelengths because of the tunability of its superconducting transition temperature T_C in the ~ 100 mK range and the relatively weak coupling between its electron and phonon systems at these temperature.

The device is cooled below the superconducting transition temperature, and a bias voltage is applied to increase the temperature of the absorber above that of the substrate to a point where a small increase in temperature will result in a large transition in the resistance. A photon is absorbed in the absorber producing a photoelectron which heats the absorber. An increase in the temperature of the absorber results in an increase in the resistance of the device and an increase in resistance leads to increase in temperature. The steep change in resistance versus temperature enables precise measurements of the energy of single photons.

At the early stage of the research, TES detectors were seldom used in practical application because of the difficulty of matching their noise to FET amplifiers (TES normal resistance is typically few ohms or less). In the last ten years, this problem has been largely eliminated by the use of superconducting quantum interference device (SQUID) current amplifiers [178], which are easily impedance-matched to low-resistance TES detectors. TESs are usually fabricated on a Si substrate and demonstrated the highest reported SPDEs of any single-photon detector, up to 95 % at 1550 nm [177], [179] when embedded in an optical cavity design. Moreover, the low DCR of these devices give rise to NEP of the order of ~ 10^{-19} $WHz^{-1/2}$. As the temperature of operation (100s of mK), requires a complex cooling system, TESs are limited by their relative low speed, both in terms of recovery time (in the order of ns) and jitter times (~ 100s of ns). These problems severely constrain the experiments that could benefit from the high efficiencies and photon number resolving capabilities. However, a W-based TES with jitter values of ~ 4 ns obtained by reading out the signal with a low input inductance SQUID has recently been demonstrated [180].

3.7 Conclusion

In this chapter a review of the integration of near-infrared photodetectors on silicon has been presented including a review on different single-photon technologies to detect single-photons at the telecom wavelengths, a vital aspect in a number of emerging application areas.

It is clear how the performance of a photodiode is related to different aspects such as the design, geometry, material, technology and fabrication. This introduction links the choice to the right semiconductor material depending on the wavelength at which the photodetector has to operate. In particular, for operation in the second and third optical telecommunications windows, the choice of the material is mainly between

$In_{0.53}Ga_{0.47}As$ and Ge. Although these two semiconductors have different intrinsic properties, they are both lattice mismatched (8 % and 4.2 %, respectively) with Si substrates. Throughout the chapter the importance of fabricating a near-infrared photodetector on Si has been explained many times. Consequently, it is also important that the lattice mismatch issue is considered. In the last 40 years or more, a lot of effort has been put into overcoming this technological issue, because there could be so many advantages of combining different materials (heterostructure) such as InGaAs and Ge with Si for many applications (mainly, but not only in silicon photonics) which benefit from working in the infrared regime.

The major problems arising from lattice mismatch of materials, such as high surface roughness and high density of threading dislocations, have been addressed and different solutions have been proposed for both material systems, Ge-on-Si and InGaAs-on-Si.

For the Ge-on-Si system different solutions have been proposed starting with the use of SiGe strained layers and SiGe SLS structures. However, research had to face the main disadvantage of the low absorption coefficient (for a Ge content x=0.6, α was ~ 21 cm^{-1} and 2.5 cm^{-1} at 1310 and 1550 nm, respectively) at infrared wavelengths for this class of detectors. The only way to solve this problem was to increase the Ge concentration, but this was in turn limited by the critical thickness and the increase of TDDs (which both increase as the Ge content is increased). Waveguide structures were also proposed but the photoresponse at 1550 nm was still very low. The ternary SiGeC system was also proposed to alleviate the lattice mismatch between Ge and Si, but very little improvement was obtained and further problems arose from the introduction of C in the SiGe system.

It has been shown for the InGaAs-on-Si system, that epitaxial growth is less inconvenient than for Ge-on-Si, due to the increased lattice mismatch (~ 8 %), and due to the contamination problems of InGaAs in Si. Moreover, a comparison between epitaxially grown and wafer bonding InGaAs-on-Si photodetectors has been made showing that the former were still far from good performance, in terms of dark current and responsivity, obtained with the latter technology. However, wafer bonding still suffers from its own intrinsic technological problems.

So, clearly pure Ge represents the best choice to make near-infrared photodetectors on Si. This is also possible thanks to technological progress made in Ge-on-Si heteroepitaxy with the introduction of the two temperature process and post-growth

annealing steps, which permit the reduction of TDD to the order of ~10^7 cm^{-2}. Further improvements to the TDD were obtained with SEG, which demonstrated photodetectors with very low TDD. The effect of TDD on the device performance has been reviewed and this has shown that reducing TDDs helps to reduce the leakage current, but other aspects such as device geometry and surface passivation play an important role, also.

The heteroepitaxy of Ge-on-Si adds an intrinsic strain to the system, due to the different thermal expansion coefficients between the two materials. However, it has been shown that this strain (~ 0.2 %) slightly enhances the absorption in the infrared region (from 840 cm^{-1} for bulk Ge to 3300 cm^{-1} for Ge-on-Si).

A review of Ge-on-Si photodetectors was also detailed. This covered many different device types (p-i-n, waveguide and APD) and explained the advantages and disadvantages of each type given. In particular, p-i-n devices have shown comparable performance with their InGaAs counterparts, with slightly higher DCDs, and different waveguide geometries have been proposed and demonstrated to further enhance the absorption at 1550 nm. APDs rely on the separate, absorption, charge, and multiplication structure that exploits both advantages of good absorption properties of Ge in the infrared and good multiplication properties of Si. In this case different geometries (vertical-incidence and waveguide structure) have been proposed and demonstrated.

A review of different single-photon detectors was also given. In particular, this work has been focused on semiconductor SPADs, whose theory and principle of operation (Geiger-mode) was explained in chapter 2. A review of different devices and technologies to detect single-photon in the infrared regime was presented. For SPADs, the performance of commercially available Ge APD and InGaAs/InP was reviewed. The former showed SPDEs of ~ 5 – 30 % at 1310 nm with high DCR (~ Mcs^{-1}), and FWHM of the order of 100s of ps. The lowest reported NEP was ~ 2×10^{-16} $WHz^{-1/2}$. Moreover, these devices were characterised at cryogenic temperature (77 K), due to the high thermal carrier generation, resulting in a very poor SPDEs at 1550 nm. On the other hand, InGaAs/InP SPADs, which are the state-of-art for detecting single-photons in the infrared, showed good SPDE at both 1310 and 1550 nm wavelengths with low DCRs, showing NEP of ~ 10^{-17} $WHz^{-1/2}$ at temperature compatible with Peltier cooling (200 K). However, the main limitation of these devices is the afterpulsing which severely limits their frequency of operation.

Finally, SNSPDs and TESs were reviewed. TESs have an extremely high SPDE up to 95 % and very low DCR, demonstrating NEP in the order of ~ 10^{-19} $WHz^{-1/2}$. However, they are limited by long reset times and slow timing jitter. NbN SNSPDs are fast devices, both in terms of reset time and timing jitter. Moreover, these devices have shown system detection efficiency of ~ 24 – 28 % at 1550 nm, although a very recent research on SNSPD based on WSi have shown system detection efficiencies over 90 % at 1550 nm. However, the main limitation of both these technologies is the low temperature of operation (100s of mK for TES, and ~ 4 K for SNSPD) which restrict their use to a laboratory environment.

3.8 Reference

[1] G. Keiser, *Optical fiber-communication*, 2nd Edition. New York: Mc-Graw-Hill, 1991.

[2] S. K. Ghandhi, *VLSI Fabrication principles*. New York: John Wiley & Sons, Inc., 1994.

[3] J. Michel, J. Liu, and L. C. Kimerling, “High-performance Ge-on-Si photodetectors,” *Nat Photon*, vol. 4, no. 8, pp. 527–534, 2010.

[4] H.-C. Luan, D. R. Lim, K. K. Lee, K. M. Chen, J. G. Sandland, K. Wada, and L. C. Kimerling, “High-quality Ge epilayers on Si with low threading-dislocation densities,” *Appl. Phys. Lett.*, vol. 75, p. 2909, 1999.

[5] C. Claeys and E. Simoen, *Fundamental and Technological Aspects of Extended Defects in Germanium*, vol. 118. Springer, 2009.

[6] L. Colace, P. Ferrara, G. Assanto, D. Fulgoni, and L. Nash, “Low Dark-Current Germanium-on-Silicon Near-Infrared Detectors,” *IEEE Photonics Technol. Lett.*, vol. 19, no. 22, pp. 1813–1815, 2007.

[7] J. W. Matthews, S. Mader, and T. B. Light, “Accommodation of Misfit Across the Interface Between Crystals of Semiconducting Elements or Compounds,” *J. Appl. Phys.*, vol. 41, no. 9, pp. 3800–3804, 1970.

[8] D. L. Dexter and F. Seitz, “Effects of Dislocations on Mobilities in Semiconductors,” *Phys. Rev.*, vol. 86, no. 6, pp. 964–965, 1952.

[9] D. J. Paul, “Si/SiGe heterostructures: from material and physics to devices and circuits,” *Semicond. Sci. Technol.*, vol. 19, no. 10, p. R75, 2004.

[10] H. Kroemer, “Theory of a Wide-Gap Emitter for Transistors,” *Proc. IRE*, vol. 45, no. 11, pp. 1535–1537, 1957.

[11] E. Kasper, H. J. Herzog, and H. Kibbel, “A one-dimensional SiGe superlattice grown by UHV epitaxy,” *Appl. Phys.*, vol. 8, no. 3, pp. 199–205, 1975.

[12] “Chapter 4 - Silicon Germanium Device Application Market Overview,” Amsterdam: Elsevier Science, pp. 73–164, 2002.

[13] S. Luryi, A. Kastalsky, and J. C. Bean, “New infrared detector on a silicon chip,” *IEEE Trans. Electron Devices*, vol. 31, no. 9, pp. 1135– 1139, 1984.

[14] J. H. Van Der Merwe, “Crystal Interfaces. Part II. Finite Overgrowths,” *J. Appl. Phys.*, vol. 34, no. 1, p. 123, 1963.

[15] J. W. Matthews and A. E. Blakeslee, “Defects in epitaxial multilayers: III. Preparation of almost perfect multilayers,” *J. Cryst. Growth*, vol. 32, no. 2, pp. 265–273, 1976.

[16] P. T.P., “Electronic and optical properties of Ge-Si superlattices,” *Prog. Quantum Electron.*, vol. 18, no. 2, pp. 97–152, 1994.

[17] E. A. Fitzgerald, S. B. Samavedam, Y. H. Xie, and L. M. Giovane, “Influence of strain on semiconductor thin film epitaxy,” vol. 15, pp. 1048–1056, 1997.
[18] S. Luryi, T. P. Pearsall, H. Temkin, and J. C. Bean, “Waveguide infrared photodetectors on a silicon chip,” *IEEE Electron Device Lett.*, vol. 7, no. 2, pp. 104–107, 1986.
[19] R. People, “Indirect band gap of coherently strained Ge_{x}Si_{1-x} bulk alloys on 〈001〉 silicon substrates,” *Phys. Rev. B*, vol. 32, no. 2, pp. 1405–1408, 1985.
[20] D. V. Lang, R. People, J. C. Bean, and A. M. Sergent, “Measurement of the band gap of GexSi1−x/Si strained-layer heterostructures,” *Appl. Phys. Lett.*, vol. 47, no. 12, p. 1333, 1985.
[21] J. M. Baribeau, T. E. Jackman, P. Maigne, D. C. Houghton, and M. W. Denhoff, “Heteroepitaxy of Ge on (100) Si substrates,” *J. Vac. Sci. Technol. Vac. Surf. Films*, vol. 5, no. 4, pp. 1898–1902, 1987.
[22] J. M. Baribeau, T. E. Jackman, D. C. Houghton, P. Maigné, and M. W. Denhoff, “Growth and characterization of Si1−xGex and Ge epilayers on (100) Si,” *J. Appl. Phys.*, vol. 63, no. 12, p. 5738, 1988.
[23] J.-M. Baribeau, D. C. Houghton, T. E. Jackman, and J. P. McCaffrey, “Molecular Beam Epitaxy Growth of Ge on (100) Si,” *J. Electrochem. Soc.*, vol. 136, no. 4, pp. 1158–1162, 1989.
[24] H. Temkin, J. C. Bean, T. P. Pearsall, N. A. Olsson, and D. V. Lang, “High photoconductive gain in GexSi1−x/Si strained-layer superlattice detectors operating at λ=1.3 μm,” *Appl. Phys. Lett.*, vol. 49, no. 3, p. 155, 1986.
[25] F. Y. Huang, X. Zhu, M. O. Tanner, and K. L. Wang, “Normal-incidence strained-layer superlattice Ge0.5Si0.5/Si photodiodes near 1.3 μm,” *Appl. Phys. Lett.*, vol. 67, no. 4, p. 566, 1995.
[26] J. Lee, A. L. Gutierrez-Aitken, S. H. Li, and P. K. Bhattacharya, “Responsivity and impact ionization coefficients of Si1-xGex photodiodes,” *IEEE Trans. Electron Devices*, vol. 43, no. 6, pp. 977–981, 1996.
[27] L. M. Giovane, H.-C. Luan, A. M. Agarwal, and L. C. Kimerling, “Correlation between leakage current density and threading dislocation density in SiGe p-i-n diodes grown on relaxed graded buffer layers,” *Appl. Phys. Lett.*, vol. 78, no. 4, pp. 541–543, 2001.
[28] Z. Huang, J. Oh, and J. C. Campbell, “Back-side-illuminated high-speed Ge photodetector fabricated on Si substrate using thin SiGe buffer layers,” *Appl. Phys. Lett.*, vol. 85, no. 15, p. 3286, 2004.
[29] B. Jalali, L. Naval, and A. F. J. Levi, “Si-based receivers for optical data links,” *J. Light. Technol.*, vol. 12, no. 6, pp. 930–935, 1994.
[30] V. P. Kesan, P. G. May, E. Bassous, and S. S. Iyer, “Integrated waveguide-photodetector using Si/SiGe multiple quantum wells for long wavelength applications,” pp. 637–640, 1990.
[31] A. Splett, T. Zinke, K. Petermann, E. Kasper, H. Kibbel, H.-J. Herzog, and H. Presting, “Integration of waveguides and photodetectors in SiGe for 1.3 μm operation,” *IEEE Photonics Technol. Lett.*, vol. 6, no. 1, pp. 59–61, 1994.
[32] B. Li, G. Li, E. Liu, Z. Jiang, J. Qin, and X. Wang, “Monolithic integration of a SiGe/Si modulator and multiple quantum well photodetector for 1.55 μm operation,” *Appl. Phys. Lett.*, vol. 73, no. 24, p. 3504, 1998.
[33] T. P. Pearsall, H. Temkin, J. C. Bean, and S. Luryi, “Avalanche gain in GexSi1-x/Si infrared waveguide detectors,” *IEEE Electron Device Lett.*, vol. 7, no. 5, pp. 330–332, 1986.

[34] H. Temkin, A. Antreasyan, N. A. Olsson, T. P. Pearsall, and J. C. Bean, "Ge0.6Si0.4 rib waveguide avalanche photodetectors for 1.3 μm operation," *Appl. Phys. Lett.*, vol. 49, no. 13, p. 809, 1986.
[35] B. Schuppert, J. Schmidtchen, A. Splett, U. Fischer, T. Zinke, R. Moosburger, and K. Petermann, "Integrated optics in silicon and SiGe-heterostructures," *J. Light. Technol.*, vol. 14, no. 10, pp. 2311–2323, 1996.
[36] A. Y. Loudon, P. A. Hiskett, G. S. Buller, R. T. Carline, D. C. Herbert, W. Y. Leong, and J. G. Rarity, "Enhancement of the infrared detection efficiency of silicon photon-counting avalanche photodiodes by use of silicon germanium absorbing layers," *Opt. Lett.*, vol. 27, no. 4, pp. 219–221, 2002.
[37] M. T. Currie, S. B. Samavedam, T. A. Langdo, C. W. Leitz, and E. A. Fitzgerald, "Controlling threading dislocation densities in Ge on Si using graded SiGe layers and chemical-mechanical polishing," *Appl. Phys. Lett.*, vol. 72, p. 1718, 1998.
[38] S. B. Samavedam, M. T. Currie, T. A. Langdo, and E. A. Fitzgerald, "High-quality germanium photodiodes integrated on silicon substrates using optimized relaxed graded buffers," *Appl. Phys. Lett.*, vol. 73, no. 15, p. 2125, 1998.
[39] F. M. Ross, R. Hull, D. Bahnck, J. C. Bean, L. J. Peticolas, and C. A. King, "Changes in electrical device characteristics during the in situ formation of dislocations," *Appl. Phys. Lett.*, vol. 62, no. 12, p. 1426, 1993.
[40] J. Kolodzey, P. R. Berger, B. A. Orner, D. Hits, F. Chen, A. Khan, X. Shao, M. M. Waite, S. I. Shah, C. P. Swann, and K. M. Unruh, "Optical and electronic properties of SiGeC alloys grown on Si substrates," *J. Cryst. Growth*, vol. 157, no. 1–4, pp. 386–391, 1995.
[41] K. Eberl, S. S. Iyer, S. Zollner, J. C. Tsang, and F. K. LeGoues, "Growth and strain compensation effects in the ternary Si1−x−yGexCy alloy system," *Appl. Phys. Lett.*, vol. 60, no. 24, p. 3033, 1992.
[42] H. J. Osten, E. Bugiel, and P. Zaumseil, "Growth of an inverse tetragonal distorted SiGe layer on Si(001) by adding small amounts of carbon," *Appl. Phys. Lett.*, vol. 64, no. 25, p. 3440, 1994.
[43] J. Kouvetakis, M. Todd, D. Chandrasekhar, and D. J. Smith, "Novel chemical routes to silicon-germanium-carbon materials," *Appl. Phys. Lett.*, vol. 65, no. 23, p. 2960, 1994.
[44] J. Mi, P. Warren, P. Letourneau, M. Judelewicz, M. Gailhanou, M. Dutoit, C. Dubois, and J. C. Dupuy, "High quality Si1−x−yGexCy epitaxial layers grown on (100) Si by rapid thermal chemical vapor deposition using methylsilane," *Appl. Phys. Lett.*, vol. 67, no. 2, p. 259, 1995.
[45] A. St. Amour, C. W. Liu, J. C. Sturm, Y. Lacroix, and M. L. W. Thewalt, "Defect-free band-edge photoluminescence and band gap measurement of pseudomorphic Si1−x−yGexCy alloy layers on Si (100)," *Appl. Phys. Lett.*, vol. 67, no. 26, p. 3915, 1995.
[46] R. A. Soref, Z. Atzman, M. Robinson, R. Westhoff, and F. Shaapur, "Infrared waveguiding in Si1-x-yGexCy upon silicon," *Opt. Lett.*, vol. 21, no. 5, pp. 345–347, 1996.
[47] F. Y. Huang and K. L. Wang, "Normal-incidence epitaxial SiGeC photodetector near 1.3 μm wavelength grown on Si substrate," *Appl. Phys. Lett.*, vol. 69, no. 16, p. 2330, 1996.
[48] F. Y. Huang, K. Sakamoto, K. L. Wang, P. Trinh, and B. Jalali, "Epitaxial SiGeC waveguide photodetector grown on Si substrate with response in the 1.3-1.55 μm wavelength range," *IEEE Photonics Technol. Lett.*, vol. 9, no. 2, pp. 229–231, 1997.

[49] H. J. Osten and E. Bugiel, “Relaxed Si1−xGex/Si1−x−yGexCy buffer structures with low threading dislocation density,” *Appl. Phys. Lett.*, vol. 70, no. 21, p. 2813, 1997.

[50] W. Prost, V. Khorenko, A.-C. Mofor, S. Neumann, A. Poloczek, A. Matiss, A. Bakin, A. Schlachetzki, and F.-J. Tegude, “High performance III/V RTD and PIN diode on a silicon (001) substrate,” *Appl. Phys. A*, vol. 87, no. 3, pp. 539–544, 2007.

[51] Y. Gao, Z. Zhong, S. Feng, Y. Geng, H. Liang, A. W. Poon, and K. M. Lau, “High-Speed Normal-Incidence p-i-n InGaAs Photodetectors Grown on Silicon Substrates by MOCVD,” *IEEE Photonics Technol. Lett.*, vol. 24, no. 4, pp. 237–239, 2012.

[52] A. Bartels, E. Peiner, G. P. Tang, R. Klockenbrink, H.-H. Wehmann, and A. Schlachetzki, “Performance of InGaAs metal-semiconductor-metal photodetectors on Si,” *IEEE Photonics Technol. Lett.*, vol. 8, no. 5, pp. 670–672, 1996.

[53] Z. L. Liau and D. E. Mull, “Wafer fusion: A novel technique for optoelectronic device fabrication and monolithic integration,” *Appl. Phys. Lett.*, vol. 56, no. 8, p. 737, 1990.

[54] F. E. Ejeckam, C. L. Chua, Z. H. Zhu, Y. H. Lo, M. Hong, and R. Bhat, “High performance InGaAs photodetectors on Si and GaAs substrates,” *Appl. Phys. Lett.*, vol. 67, no. 26, p. 3936, 1995.

[55] B. F. Levine, A. R. Hawkins, S. Hiu, B. J. Tseng, C. A. King, L. A. Gruezke, R. W. Johnson, D. R. Zolnowski, and J. E. Bowers, “20 GHz high performance planar Si/InGaAs p-i-n photodetector,” *Appl. Phys. Lett.*, vol. 70, no. 18, pp. 2449–2451, 1997.

[56] B. F. Levine, C. J. Pinzone, S. Hui, C. A. King, R. E. Leibenguth, D. R. Zolnowski, D. V. Lang, H. W. Krautter, and M. Geva, “Ultralow-dark-current wafer-bonded Si/InGaAs photodetectors,” *Appl. Phys. Lett.*, vol. 75, no. 14, p. 2141, 1999.

[57] Y. Kang, P. Mages, A. R. Clawson, S. S. Lau, Y. H. Lo, P. K. L. Yu, A. Pauchard, Z. Zhu, and Y. Zhou, “Wafer-fused p-i-n InGaAs/Si photodiode with photogain,” *Appl. Phys. Lett.*, vol. 79, no. 7, p. 970, 2001.

[58] D. A. Tulchinsky, K. J. Williams, A. Pauchard, M. Bitter, Z. Pan, L. Hodge, S. G. Hummel, and Y. H. Lo, “High-power InGaAs-on-Si pin RF photodiodes,” *Electron. Lett.*, vol. 39, no. 14, pp. 1084–1086, 2003.

[59] A. Pauchard, M. Bitter, Z. Pan, S. Kristjansson, L. A. Hodge, K. J. Williams, D. A. Tulchinsky, S. G. Hummel, and Y.-H. Lo, “Infrared-sensitive InGaAs-on-Si p-i-n photodetectors exhibiting high-power linearity,” *IEEE Photonics Technol. Lett.*, vol. 16, no. 11, pp. 2544–2546, 2004.

[60] G. Roelkens, J. Brouckaert, D. Taillaert, P. Dumon, W. Bogaerts, D. Van Thourhout, R. Baets, R. Notzel, and M. Smit, “Integration of InP/InGaAsP photodetectors onto silicon-on-insulator waveguide circuits,” *Opt. Express*, vol. 13, no. 25, pp. 10102–10108, 2005.

[61] Z. Sheng, L. Liu, J. Brouckaert, S. He, and D. Van Thourhout, “InGaAs PIN photodetectors integrated on silicon-on-insulator waveguides,” *Opt. Express*, vol. 18, no. 2, pp. 1756–1761, 2010.

[62] Y. Kang, P. Mages, A. R. Clawson, P. K. L. Yu, M. Bitter, Z. Pan, A. Pauchard, S. Hummel, and Y.-H. Lo, “Fused InGaAs-Si avalanche photodiodes with low-noise performances,” *IEEE Photonics Technol. Lett.*, vol. 14, no. 11, pp. 1593–1595, 2002.

[63] Y. Kang, Y.-H. Lo, M. Bitter, S. Kristjansson, Z. Pan, and A. Pauchard, "InGaAs-on-Si single photon avalanche photodetectors," *Appl. Phys. Lett.*, vol. 85, no. 10, p. 1668, 2004.
[64] A. R. Hawkins, W. Wu, P. Abraham, K. Streubel, and J. E. Bowers, "High gain-bandwidth-product silicon heterointerface photodetector," *Appl. Phys. Lett.*, vol. 70, no. 3, p. 303, 1997.
[65] J. R. Chelikowsky and M. L. Cohen, "Nonlocal pseudopotential calculations for the electronic structure of eleven diamond and zinc-blende semiconductors," *Phys. Rev. B*, vol. 14, no. 2, pp. 556–582, Jul. 1976.
[66] E. M. Conwell, "Properties of Silicon and Germanium," *Proc. IRE*, vol. 40, no. 11, pp. 1327–1337, 1952.
[67] E. M. Conwell, "Properties of Silicon and Germanium: II," *Proc. IRE*, vol. 46, no. 6, pp. 1281–1300, 1958.
[68] Y. Murakami and T. Shingyouji, "Separation and analysis of diffusion and generation components of pn junction leakage current in various silicon wafers," *J. Appl. Phys.*, vol. 75, no. 7, p. 3548, 1994.
[69] N. A. DiLello, D. K. Johnstone, and J. L. Hoyt, "Characterization of dark current in Ge-on-Si photodiodes," *J. Appl. Phys.*, vol. 112, no. 5, pp. 054506–054506–8, 2012.
[70] A. Rajan, "MBE Growth, Processing and Stability of Zinc blende MgS based Heterostructers," Heriot-Watt university, 2013.
[71] J.-H. Park and T. S. Sudarshan, *Chemical Vapor Deposition*. ASM International, p. 477, 2001.
[72] D. P. Malta, J. B. Posthill, R. J. Markunas, and T. P. Humphreys, "Low-defect-density germanium on silicon obtained by a novel growth phenomenon," *Appl. Phys. Lett.*, vol. 60, no. 7, p. 844, 1992.
[73] J. Liu, H. J. Kim, O. Hul'ko, Y. H. Xie, S. Sahni, P. Bandaru, and E. Yablonovitch, "Ge films grown on Si substrates by molecular-beam epitaxy below 450°C," *J. Appl. Phys.*, vol. 96, no. 1, p. 916, 2004.
[74] Y. Fukuda and Y. Kohama, "Effect of *in situ* Thermal Annealing on Crystalline Quality of Ge Layers Grown by Molecular Beam Epitaxy on Si (100)," *Jpn. J. Appl. Phys.*, vol. 26, no. Part 2, No. 5, pp. L597–L599, 1987.
[75] Y. Fukuda, Y. Kohama, M. Seki, and Y. Ohmachi, "Dislocation Reduction in MBE-Grown Ge on Si(001) by *in situ* Thermal Annealing," *Jpn. J. Appl. Phys.*, vol. 27, no. Part 2, No. 9, pp. L1591–L1593, 1988.
[76] D. p. Malta, J. b. Posthill, P. m. Enquist, R. j. Markunas, T. p. Humphreys, and N. r. Parikh, "Low-Defect-Density Ge on Si for Large-Lattice-Mismatched Semiconductor Integration and Strain-Engineered Devices," *MRS Online Proc. Libr.*, vol. 263, p. null–null,1992.
[77] J. C. C. Fan, B. Y. Tsaur, R. P. Gale, and F. M. Davis, "Reducing dislocations in semiconductors utilizing repeated thermal cycling during multistage epitaxial growth," 1986.
[78] J. W. Lee, H. Shichijo, H.-L. Tsai, and R. J. Matyi, "Defect reduction by thermal annealing of GaAs layers grown by molecular beam epitaxy on Si substrates," *Appl. Phys. Lett.*, vol. 50, no. 1, pp. 31–33, 1987.
[79] M. Yamaguchi, A. Yamamoto, M. Tachikawa, Y. Itoh, and M. Sugo, "Defect reduction effects in GaAs on Si substrates by thermal annealing," *Appl. Phys. Lett.*, vol. 53, no. 23, p. 2293, 1988.
[80] M. Yamaguchi, "Dislocation density reduction in heteroepitaxial III-V compound films on Si substrates for optical devices.," *J. Mater. Res.*, vol. 6, no. 2, pp. 376–384, 1991.

[81] L. Colace, G. Masini, F. Galluzzi, G. Assanto, G. Capellini, L. Di Gaspare, E. Palange, and F. Evangelisti, "Metal–semiconductor–metal near-infrared light detector based on epitaxial Ge/Si," *Appl. Phys. Lett.*, vol. 72, p. 3175, 1998.
[82] L. Colace, G. Masini, F. Galluzzi, G. Assanto, G. Capellini, L. Di Gaspare, E. Palange, and F. Evangelisti, "Metal–Ge-Si heterostructures for near-infrared light detection," *J. Vac. Sci. Technol. B Microelectron. Nanometer Struct.*, vol. 17, no. 2, pp. 465–467, 1999.
[83] H.-C. Luan, D. R. Lim, L. Colace, G. Masini, G. Assanto, K. Wada, and L. C. Kimerling, "Germanium Photodetectors for Silicon Microphotonics by Direct Epitaxy on Silicon," *MRS Online Proc. Libr.*, vol. 607, 1999.
[84] V. A. Shah, A. Dobbie, M. Myronov, and D. R. Leadley, "High quality relaxed Ge layers grown directly on a Si(0 0 1) substrate," *Solid-State Electron.*, vol. 62, no. 1, pp. 189–194, 2011.
[85] J. Hartmann, "Reduced pressure – chemical vapor deposition of Ge thick layers on Si(001) for 1.3–1.55-μm photodetection," *J. Appl. Phys.*, vol. 95, p. 5095, 2004.
[86] L. Vivien, J. Osmond, J. Fédéli, D. Marris-morini, P. Crozat, J. Damlencourt, E. Cassan, Y. Lecunff, and S. Laval, "42 GHz p . i . n Germanium photodetector integrated in a silicon-on-insulator waveguide," *Opt. Express*, vol. 17, no. 8, pp. 6252–6257, 2009.
[87] M. Morse, O. Dosunmu, G. Sarid, and Y. Chetrit, "Performance of Ge-on-Si p-i-n Photodetectors for Standard Receiver Modules," *IEEE Photonics Technol. Lett.*, vol. 18, no. 23, pp. 2442–2444, 2006.
[88] Dongwoo Suh, Sanghoon Kim, Jiho Joo, and Gyungock Kim, "36-GHz High-Responsivity Ge Photodetectors Grown by RPCVD," *IEEE Photonics Technol. Lett.*, vol. 21, no. 10, pp. 672–674, 2009.
[89] G. Dehlinger, S. J. Koester, J. D. Schaub, J. O. Chu, Q. C. Ouyang, and A. Grill, "High-speed Germanium-on-SOI lateral PIN photodiodes," *IEEE Photonics Technol. Lett.*, vol. 16, no. 11, pp. 2547–2549, 2004.
[90] D. Ahn, C. Hong, J. Liu, W. Giziewicz, M. Beals, C. Lionel, J. Michel, J. Chen, and F. X. Kärtner, "High performance , waveguide integrated Ge photodetectors," *Opt. Express*, vol. 15, no. 7, pp. 3916–3921, 2007.
[91] J. Osmond, G. Isella, D. Chrastina, R. Kaufmann, M. Acciarri, and H. von Känel, "Ultralow dark current Ge/Si(100) photodiodes with low thermal budget," *Appl. Phys. Lett.*, vol. 94, p. 201106, 2009.
[92] J. Nakatsuru, H. Date, S. Mashiro, and M. Ikemoto, "Growth of high quality Ge epitaxial layer on Si(100) substrate using ultra thin Si0.5Ge0.5 buffer," *MRS Online Proc. Libr.*, vol. 891, 2005.
[93] T. H. Loh, H. S. Nguyen, C. H. Tung, A. D. Trigg, G. Q. Lo, N. Balasubramanian, D. L. Kwong, and S. Tripathy, "Ultrathin low temperature SiGe buffer for the growth of high quality Ge epilayer on Si(100) by ultrahigh vacuum chemical vapor deposition," *Appl. Phys. Lett.*, vol. 90, no. 9, p. 092108, 2007.
[94] T. A. Langdo, C. W. Leitz, M. T. Currie, E. A. Fitzgerald, A. Lochtefeld, and D. A. Antoniadis, "High quality Ge on Si by epitaxial necking," *Appl. Phys. Lett.*, vol. 76, no. 25, p. 3700, 2000.
[95] J.-S. Park, J. Bai, M. Curtin, B. Adekore, M. Carroll, and A. Lochtefeld, "Defect reduction of selective Ge epitaxy in trenches on Si(001) substrates using aspect ratio trapping," *Appl. Phys. Lett.*, vol. 90, no. 5, p. 052113, 2007.
[96] G. Wang, R. Loo, S. Takeuchi, L. Souriau, J. C. Lin, A. Moussa, H. Bender, B. De Jaeger, P. Ong, W. Lee, M. Meuris, M. Caymax, W. Vandervorst, B. Blanpain, and M. M. Heyns, "Fabrication of high quality Ge virtual substrates by

selective epitaxial growth in shallow trench isolated Si (001) trenches," *Thin Solid Films*, vol. 518, no. 9, pp. 2538–2541, 2010.
[97] A. Sammak, W. de Boer, and L. K. Nanver, "Ge-on-Si: Single-Crystal Selective Epitaxial Growth in a CVD Reactor," *ECS Trans.*, vol. 50, no. 9, pp. 507–512, 2013.
[98] L. Colace, G. Masini, G. Assanto, H.-C. Luan, K. Wada, and L. C. Kimerling, "Efficient high-speed near-infrared Ge photodetectors integrated on Si substrates," *Appl. Phys. Lett.*, vol. 76, no. 10, pp. 1231–1233, 2000.
[99] H. P. Singh, "Determination of thermal expansion of germanium, rhodium and iridium by X-rays," *Acta Crystallogr. Sect. A*, vol. 24, no. 4, pp. 469–471, 1968.
[100] Y. Okada and Y. Tokumaru, "Precise determination of lattice parameter and thermal expansion coefficient of silicon between 300 and 1500 K," *J. Appl. Phys.*, vol. 56, no. 2, p. 314, 1984.
[101] Y. Ishikawa, K. Wada, D. D. Cannon, J. Liu, H.-C. Luan, and L. C. Kimerling, "Strain-induced band gap shrinkage in Ge grown on Si substrate," *Appl. Phys. Lett.*, vol. 82, no. 13, pp. 2044–2046, 2003.
[102] Y. Ishikawa, K. Wada, J. Liu, D. D. Cannon, H.-C. Luan, J. Michel, and L. C. Kimerling, "Strain-induced enhancement of near-infrared absorption in Ge epitaxial layers grown on Si substrate," *J. Appl. Phys.*, vol. 98, no. 1, p. 013501, 2005.
[103] J. Liu, D. D. Cannon, K. Wada, Y. Ishikawa, S. Jongthammanurak, D. T. Danielson, J. Michel, and L. C. Kimerling, "Tensile strained Ge p-i-n photodetectors on Si platform for C and L band telecommunications," *Appl. Phys. Lett.*, vol. 87, p. 011110, 2005.
[104] W. C. Dash and R. Newman, "Intrinsic Optical Absorption in Single-Crystal Germanium and Silicon at 77°K and 300°K," *Phys. Rev.*, vol. 99, no. 4, pp. 1151–1155, 1955.
[105] L. Chen, P. Dong, and M. Lipson, "High performance germanium photodetectors integrated on submicron silicon waveguides by low temperature wafer bonding," *Opt. Express*, vol. 16, no. 15, pp. 11513–11518, 2008.
[106] F. Gity, A. Daly, B. Snyder, F. H. Peters, J. Hayes, C. Colinge, A. P. Morrison, and B. Corbett, "Ge/Si heterojunction photodiodes fabricated by low temperature wafer bonding," *Opt. Express*, vol. 21, no. 14, pp. 17309–17314, 2013.
[107] C. Masini, L. Calace, G. Assanto, Hsin-Chiao Luan, and L. C. Kimerling, "High-performance p-i-n Ge on Si photodetectors for the nearinfrared: from model to demonstration," *IEEE Trans. Electron Devices*, vol. 48, no. 6, pp. 1092–1096, 2001.
[108] S. Famà, L. Colace, G. Masini, G. Assanto, and H.-C. Luan, "High performance germanium-on-silicon detectors for optical communications," *Appl. Phys. Lett.*, vol. 81, p. 586, 2002.
[109] P. R. Bandaru, S. Sahni, E. Yablonovitch, J. Liu, H.-J. Kim, and Y.-H. Xie, "Fabrication and characterization of low temperature (<450 °C) grown p-Ge/n-Si photodetectors for silicon based photonics," *Mater. Sci. Eng. B*, vol. 113, no. 1, pp. 79–84, 2004.
[110] O. I. Dosunmu, D. D. Cannon, M. K. Emsley, L. C. Kimerling, and M. S. Unlu, "High-speed resonant cavity enhanced Ge photodetectors on reflecting Si substrates for 1550-nm operation," *IEEE Photonics Technol. Lett.*, vol. 17, no. 1, pp. 175–177, 2005.
[111] M. Jutzi, M. Berroth, G. Wohl, M. Oehme, and E. Kasper, "Ge-on-Si vertical incidence photodiodes with 39-GHz bandwidth," *IEEE Photonics Technol. Lett.*, vol. 17, no. 7, pp. 1510–1512, 2005.

[112] L. Colace, M. Balbi, G. Masini, G. Assanto, H.-C. Luan, and L. C. Kimerling, "Ge on Si p-i-n photodiodes operating at 10 Gbit/s," *Appl. Phys. Lett.*, vol. 88, no. 10, p. 101111, 2006.
[113] G. Isella, J. Osmond, M. Kummer, R. Kaufmann, and H. von Känel, "Heterojunction photodiodes fabricated from Ge/Si (1 0 0) layers grown by low-energy plasma-enhanced CVD," *Semicond. Sci. Technol.*, vol. 22, pp. S26–S28, 2007.
[114] J. Osmond, G. Isella, D. Chrastina, R. Kaufmann, and H. Von Känel, "Ge / Si (100) heterojunction photodiodes fabricated from material grown by low-energy plasma-enhanced chemical vapour deposition," *Thin Solid Films*, vol. 517, no. 1, pp. 380–382, 2008.
[115] J. Osmond, G. Isella, D. Chrastina, R. Kaufmann, M. Acciarri, and H. von Känel, "Ultralow dark current Ge/Si(100) photodiodes with low thermal budget," *Appl. Phys. Lett.*, vol. 94, no. 20, p. 201106, 2009.
[116] V. Sorianello, L. Colace, N. Armani, F. Rossi, C. Ferrari, L. Lazzarini, and G. Assanto, "Low-temperature germanium thin films on silicon," *Opt. Mater. Express*, vol. 1, no. 5, pp. 856–865, 2011.
[117] V. Sorianello, A. De Iacovo, L. Colace, A. Fabbri, L. Tortora, E. Buffagni, and G. Assanto, "High responsivity near-infrared photodetectors in evaporated Ge-on-Si," *Appl. Phys. Lett.*, vol. 101, no. 8, pp. 081101–081101–4, 2012.
[118] J. R. Weber, A. Janotti, P. Rinke, and C. G. Van de Walle, "Dangling-bond defects and hydrogen passivation in germanium," *Appl. Phys. Lett.*, vol. 91, no. 14, pp. 142101–142101–3, 2007.
[119] D. Suh, S. Kim, J. Joo, and G. Kim, "36-GHz High-Responsivity Ge Photodetectors Grown by RPCVD," *IEEE Photonics Technol. Lett.*, vol. 21, no. 10, pp. 672–674, 2009.
[120] Z. Zhou, J. He, R. Wang, C. Li, and J. Yu, "Normal incidence p–i–n Ge heterojunction photodiodes on Si substrate grown by ultrahigh vacuum chemical vapor deposition," *Opt. Commun.*, vol. 283, no. 18, pp. 3404–3407, 2010.
[121] J. Wang, W.-Y. Loh, K. T. Chua, H. Zang, Y. Z. Xiong, S. M. F. Tan, M. B. Yu, S. J. Lee, G.-Q. Lo, and D.-L. Kwong, "Low-Voltage High-Speed (18 GHz/1 V) Evanescent-Coupled Thin-Film-Ge Lateral PIN Photodetectors Integrated on Si Waveguide," *IEEE Photonics Technol. Lett.*, vol. 20, no. 17, pp. 1485–1487, 2008.
[122] G. Masini, S. Sahni, G. Capellini, J. Witzens, and C. Gunn, "High-Speed Near Infrared Optical Receivers Based on Ge Waveguide Photodetectors Integrated in a CMOS Process," *Adv. Opt. Technol.*, vol. 2008, pp. 1–5, 2008.
[123] N. N. Feng, P. Dong, D. Zheng, S. Liao, H. Liang, R. Shafiiha, D. Feng, G. Li, J. E. Cunningham, A. V. Krishnamoorthy, and others, "Vertical pin germanium photodetector with high external responsivity integrated with large core Si waveguides," *Opt Express*, vol. 18, no. 1, pp. 96–101, 2010.
[124] L. Vivien, A. Polzer, D. Marris-Morini, J. Osmond, J. M. Hartmann, P. Crozat, E. Cassan, C. Kopp, H. Zimmermann, and J. M. Fedeli, "Zero-bias 40Gbit/s germanium waveguide photodetector on silicon," *Opt. Express*, vol. 20, no. 2, pp. 1096–1101, 2012.
[125] Y. Kang, M. Zadka, S. Litski, G. Sarid, M. Morse, M. Paniccia, Y. Kuo, J. Bowers, A. Beling, H. Liu, and others, "Epitaxially-grown Ge/Si avalanche photodiodes for 1.3 μm light detection," *Opt Express*, vol. 16, pp. 9365–9371, 2008.
[126] Y. Kang, H.-D. Liu, M. Morse, M. J. Paniccia, M. Zadka, S. Litski, G. Sarid, A. Pauchard, Y.-H. Kuo, H.-W. Chen, W. S. Zaoui, J. E. Bowers, A. Beling, D. C. McIntosh, X. Zheng, and J. C. Campbell, "Monolithic germanium/silicon

avalanche photodiodes with 340 GHz gain-bandwidth product," *Nat Photon*, vol. 3, no. 1, pp. 59–63, 2009.

[127] M. S. Carroll, K. Childs, R. Jarecki, T. Bauer, and K. Saiz, "Ge–Si separate absorption and multiplication avalanche photodiode for Geiger mode single photon detection," *Appl. Phys. Lett.*, vol. 93, no. 18, pp. 183511–183511–3, 2008.

[128] P. Tsipas and A. Dimoulas, "Modeling of negatively charged states at the Ge surface and interfaces," *Appl. Phys. Lett.*, vol. 94, no. 1, pp. 012114–012114–3, 2009.

[129] Chunlai Xue, Haiyun Xue, Buwen Cheng, Anqi Bai, Weixuan Hu, Yude Yu, and Qiming Wang, "Si/Ge separated absorption charge multiplication avalanche photodetector with low dark current," pp. 178–180, 2009.

[130] S. Zhu, K.-W. Ang, S. C. Rustagi, J. Wang, Y. Z. Xiong, G.-Q. Lo, and D.-L. Kwong, "Waveguided Ge/Si Avalanche Photodiode With Separate Vertical SEG-Ge Absorption, Lateral Si Charge, and Multiplication Configuration," *IEEE Electron Device Lett.*, vol. 30, no. 9, pp. 934–936, 2009.

[131] X. Wang, L. Chen, W. Chen, H. Cui, Y. Hu, P. Cai, R. Yang, C. Y. Hong, D. Pan, K. W. Ang, and others, "80 GHz bandwidth-gain-product Ge/Si avalanche photodetector by selective Ge growth," 2009.

[132] Yimin Kang, M. Morse, M. J. Paniccia, M. Zadka, Y. Saad, G. Sarid, A. Pauchard, W. S. Zaoui, Hui-Wen Chen, Daoxin Dai, J. E. Bowers, Han-Din Liu, D. C. Mcintosh, Xiaoguang Zheng, and J. C. Campbell, "Monolithic Ge/Si avalanche photodiodes," pp. 25–27, 2009.

[133] K. W. Ang, J. W. Ng, A. E. J. Lim, M. B. Yu, G. Q. Lo, and D. L. Kwong, "Waveguide-Integrated Ge/Si Avalanche Photodetector with 105GHz Gain-Bandwidth Product," 2010.

[134] N. Duan, T.-Y. Liow, A. E.-J. Lim, L. Ding, and G. Q. Lo, "310 GHz gain-bandwidth product Ge/Si avalanche photodetector for 1550 nm light detection," *Opt. Express*, vol. 20, no. 10, p. 11031, 2012.

[135] N. Duan, T.-Y. Liow, A. E. Lim, L. Ding, and G. Q. Lo, "High speed waveguide-integrated {Ge/Si} avalanche photodetector," in *Optical Fiber Communication Conference and Exposition and the National Fiber Optic Engineers Conference* pp. 1–3, 2013.

[136] G. S. Buller and R. J. Collins, "Single-photon generation and detection," *Meas. Sci. Technol.*, vol. 21, p. 012002, 2010.

[137] S. Cova, M. Ghioni, A. Lotito, I. Rech, and F. Zappa, "Evolution and prospects for single-photon avalanche diodes and quenching circuits," *J. Mod. Opt.*, vol. 51, no. 9–10, pp. 1267–1288, 2004.

[138] P. J. Clarke, R. J. Collins, P. A. Hiskett, M.-J. García-Martínez, N. J. Krichel, A. McCarthy, M. G. Tanner, J. A. O'Connor, C. M. Natarajan, S. Miki, M. Sasaki, Z. Wang, M. Fujiwara, I. Rech, M. Ghioni, A. Gulinatti, R. H. Hadfield, P. D. Townsend, and G. S. Buller, "Analysis of detector performance in a gigahertz clock rate quantum key distribution system," *New J. Phys.*, vol. 13, no. 7, p. 075008, 2011.

[139] A. McCarthy, X. Ren, A. Della Frera, N. R. Gemmell, N. J. Krichel, C. Scarcella, A. Ruggeri, A. Tosi, and G. S. Buller, "Kilometer-range depth imaging at 1550 nm wavelength using an InGaAs/InP single-photon avalanche diode detector," *Opt. Express*, vol. 21, no. 19, pp. 22098–22113, 2013.

[140] N. R. Gemmell, A. McCarthy, B. Liu, M. G. Tanner, S. D. Dorenbos, V. Zwiller, M. S. Patterson, G. S. Buller, B. C. Wilson, and R. H. Hadfield, "Singlet oxygen luminescence detection with a fiber-coupled superconducting nanowire single-photon detector," *Opt. Express*, vol. 21, no. 4, pp. 5005–5013, 2013.

[141] O. Groezinger and W. Haecker, “Influence of tunneling processes on avalanche breakdown in Ge and Si,” *J. Appl. Phys.*, vol. 44, no. 3, pp. 1307–1310, 1973.
[142] A. Lacaita, P. A. Francese, F. Zappa, and S. Cova, “Single-photon detection beyond 1 μm: performance of commercially available germanium photodiodes,” *Appl. Opt.*, vol. 33, no. 30, pp. 6902–6918, 1994.
[143] W. Fichtner and W. Häcker, “Time resolution of Ge avalanche photodiodes operating as photon counters in delayed coincidence,” *Rev. Sci. Instrum.*, vol. 47, no. 3, pp. 374–377, 1976.
[144] A. Tosi, A. Dalla Mora, F. Zappa, and S. Cova, “Germanium and InGaAs/InP SPADs for single-photon detection in the near-infrared,” in *Proc. SPIE 6771, Advanced Photon Counting Techniques II*, p. 67710P–67710P, 2007.
[145] S. J. Fancey, “Single-photon avalanche diodes for time-resolved photoluminescence measurements in the near infra-red,” Heriot-Watt University, 1996.
[146] R. H. Haitz, “Mechanisms Contributing to the Noise Pulse Rate of Avalanche Diodes,” *J. Appl. Phys.*, vol. 36, no. 10, pp. 3123–3131, 1965.
[147] G. Leising, “Recombination in semiconductors. By Peter T. Landsberg, Cambridge University Press, Cambridge 1991, 595 pp., hardcover, ISBN 0-521-36122-2,” *Adv. Mater.*, vol. 5, no. 1, pp. 67–67, 1993.
[148] F. Acerbi, M. Anti, A. Tosi, and F. Zappa, “Design Criteria for InGaAs/InP Single-Photon Avalanche Diode,” *IEEE Photonics J.*, vol. 5, no. 2, pp. 6800209–6800209, 2013.
[149] S. Pellegrini, “InGaAs/InP SINGLE-PHOTON AVALANCHE DIODES,” Heriot-Watt University, 2005.
[150] R. E. Warburton, “Infrared Time-Correlated Single-Photon Counting,” Heriot-Watt University, 2008.
[151] S. R. Forrest, O. K. Kim, and R. G. Smith, “Optical response time of In0.53Ga0.47As/InP avalanche photodiodes,” *Appl. Phys. Lett.*, vol. 41, no. 1, pp. 95–98, 1982.
[152] F. Zappa, A. Lacaita, S. Cova, and P. Webb, “Nanosecond single-photon timing with InGaAs/InP photodiodes,” *Opt. Lett.*, vol. 19, no. 11, pp. 846–848, 1994.
[153] Y. Liu, S. R. Forrest, J. Hladky, M. J. Lange, G. H. Olsen, and D. E. Ackley, “A planar InP/InGaAs avalanche photodiode with floating guard ring and double diffused junction,” *J. Light. Technol.*, vol. 10, no. 2, pp. 182–193, 1992.
[154] S. Pellegrini, R. E. Warburton, L. J. . Tan, Jo Shien Ng, A. B. Krysa, K. Groom, J. P. . David, S. Cova, M. J. Robertson, and G. S. Buller, “Design and performance of an InGaAs-InP single-photon avalanche diode detector,” *IEEE J. Quantum Electron.*, vol. 42, no. 4, pp. 397– 403, 2006.
[155] A. Lacaita, F. Zappa, S. Cova, and P. Lovati, “Single-photon detection beyond 1 mm: performance of commercially available InGaAs/lnP detectors,” *Appl. Opt.*, vol. 35, no. 16, pp. 2986–2996, 1996.
[156] M. A. Itzler, r. Ben-Michael, C.-F. Hsu, K. Slomkowski, A. Tosi, S. Cova, F. Zappa, and R. Ispasoiu, “Single photon avalanche diodes (SPADs) for 1.5 μm photon counting applications,” *J. Mod. Opt.*, vol. 54, no. 2–3, pp. 283–304, 2007.
[157] A. Tosi, S. Cova, F. Zappa, M. A. Itzler, and R. Ben-Michael, “InGaAs/InP Single Photon Avalanche Diode Design and Characterization,” pp. 335 –338, 2006.
[158] M. Liu, C. Hu, X. Bai, X. Guo, J. C. Campbell, Z. Pan, and M. M. Tashima, “High-Performance InGaAs/InP Single-Photon Avalanche Photodiode,” *IEEE J. Sel. Top. Quantum Electron.*, vol. 13, no. 4, pp. 887–894, 2007.

[159] R. E. Warburton, M. Itzler, and G. S. Buller, “Free-running, room temperature operation of an InGaAs/InP single-photon avalanche diode,” *Appl. Phys. Lett.*, vol. 94, p. 071116, 2009.

[160] R. E. Warburton, M. A. Itzler, and G. S. Buller, “Improved free-running InGaAs/InP single-photon avalanche diode detectors operating at room temperature,” *Electron. Lett.*, vol. 45, no. 19, pp. 996–997, 2009.

[161] F. Acerbi, A. Tosi, and F. Zappa, “Growths and diffusions for InGaAs/InP single-photon avalanche diodes,” *Sens. Actuators Phys.*, vol. 201, pp. 207–213, Oct. 2013.

[162] Y. Nambu, S. Takahashi, K. Yoshino, A. Tanaka, M. Fujiwara, M. Sasaki, A. Tajima, S. Yorozu, and A. Tomita, “Efficient and low-noise single-photon avalanche photodiode for 1.244-GHz clocked quantum key distribution,” *Opt. Express*, vol. 19, no. 21, pp. 20531–20541, 2011.

[163] A. R. Dixon, Z. L. Yuan, J. F. Dynes, A. W. Sharpe, and A. J. Shields, “Continuous operation of high bit rate quantum key distribution,” *Appl. Phys. Lett.*, vol. 96, no. 16, p. 161102, 2010.

[164] Z. Lu, W. Sun, Q. Zhou, J. Campbell, X. Jiang, and M. A. Itzler, “Improved sinusoidal gating with balanced InGaAs/InP Single Photon Avalanche Diodes,” *Opt. Express*, vol. 21, no. 14, pp. 16716–16721, 2013.

[165] F. Acerbi, A. D. Frera, A. Tosi, and F. Zappa, “Fast Active Quenching Circuit for Reducing Avalanche Charge and Afterpulsing in InGaAs/InP Single-Photon Avalanche Diode,” *IEEE J. Quantum Electron.*, vol. 49, no. 7, pp. 563–569, 2013.

[166] P. A. Hiskett, G. S. Buller, A. Y. Loudon, J. M. Smith, I. Gontijo, A. C. Walker, P. D. Townsend, and M. J. Robertson, “Performance and Design of InGaAs /InP Photodiodes for Single-Photon Counting at 1.55 μm,” *Appl. Opt.*, vol. 39, no. 36, pp. 6818–6829, 2000.

[167] G. N. Gol’tsman, A. D. Semenov, Y. P. Gousev, M. A. Zorin, I. G. Godidze, E. M. Gershenzon, P. T. Lang, W. J. Knott, and K. F. Renk, “Sensitive picosecond NbN detector for radiation from millimetre wavelengths to visible light,” *Supercond. Sci. Technol.*, vol. 4, no. 9, p. 453, 1991.

[168] G. N. Gol, O. Okunev, G. Chulkova, A. Lipatov, A. Semenov, and K. Smirnov, “Picosecond superconducting single-photon optical detector,” *Appl. Phys. Lett.*, vol. 705, no. 2001, pp. 98–101, 2011.

[169] A. Pearlman, A. Cross, W. Slysz, J. Zhang, A. Verevkin, M. Currie, A. Korneev, P. Kouminov, K. Smirnov, B. Voronov, G. Gol’tsman, and R. Sobolewski, “Gigahertz counting rates of NbN single-photon detectors for quantum communications,” *IEEE Trans. Appl. Supercond.*, vol. 15, no. 2, pp. 579–582, 2005.

[170] S. Miki, M. Fujiwara, M. Sasaki, B. Baek, A. J. Miller, R. H. Hadfield, S. W. Nam, and Z. Wang, “Large sensitive-area NbN nanowire superconducting single-photon detectors fabricated on single-crystal MgO substrates,” *Appl. Phys. Lett.*, vol. 92, no. 6, p. 061116, 2008.

[171] A. Verevkin, J. Zhang, R. Sobolewski, A. Lipatov, O. Okunev, G. Chulkova, A. Korneev, K. Smirnov, G. N. Gol’tsman, and A. Semenov, “Detection efficiency of large-active-area NbN single-photon superconducting detectors in the ultraviolet to near-infrared range,” *Appl. Phys. Lett.*, vol. 80, no. 25, pp. 4687–4689, 2002.

[172] K. M. Rosfjord, J. K. W. Yang, E. A. Dauler, A. J. Kerman, V. Anant, B. M. Voronov, G. N. Gol’tsman, and K. K. Berggren, “Nanowire single-photon detector with an integrated optical cavity and anti-reflection coating,” *Opt. Express*, vol. 14, no. 2, pp. 527–534, 2006.

[173] J. P. Sprengers, A. Gaggero, D. Sahin, S. Jahanmirinejad, G. Frucci, F. Mattioli, R. Leoni, J. Beetz, M. Lermer, M. Kamp, S. Höfling, R. Sanjines, and A. Fiore, "Waveguide superconducting single-photon detectors for integrated quantum photonic circuits," *Appl. Phys. Lett.*, vol. 99, no. 18, p. 181110, 2011.
[174] S. Miki, T. Yamashita, M. Fujiwara, M. Sasaki, and Z. Wang, "Multichannel SNSPD system with high detection efficiency at telecommunication wavelength," *Opt. Lett.*, vol. 35, no. 13, pp. 2133–2135, 2010.
[175] F. Marsili, V. B. Verma, J. A. Stern, S. Harrington, A. E. Lita, T. Gerrits, I. Vayshenker, B. Baek, M. D. Shaw, R. P. Mirin, and S. W. Nam, "Detecting single infrared photons with 93% system efficiency," *Nat. Photonics*, vol. 7, no. 3, pp. 210–214, 2013.
[176] K. D. Irwin and G. C. Hilton, "Transition-Edge Sensors," C. Enss, Ed. Springer Berlin Heidelberg, pp. 63–150, 2005.
[177] A. E. Lita, A. J. Miller, and S. W. Nam, "Counting near-infrared single-photons with 95% efficiency," *Opt. Express*, vol. 16, no. 5, p. 3032, 2008.
[178] K. D. Irwin, S. W. Nam, B. Cabrera, B. Chugg, G. S. Park, R. P. Welty, and J. M. Martinis, "A self-biasing cryogenic particle detector utilizing electrothermal feedback and a SQUID readout," *IEEE Trans. Appl. Supercond.*, vol. 5, no. 2, pp. 2690–2693, 1995.
[179] A. E. Lita, B. Calkins, L. A. Pellouchoud, A. J. Miller, and S. Nam, "Superconducting transition-edge sensors optimized for high-efficiency photon-number resolving detectors," vol. 7681, p. 76810D–76810D–10, 2010.
[180] A. Lamas-Linares, B. Calkins, N. A. Tomlin, T. Gerrits, A. E. Lita, J. Beyer, R. P. Mirin, and S. Woo Nam, "Nanosecond-scale timing jitter for single photon detection in transition edge sensors," *Appl. Phys. Lett.*, vol. 102, no. 23, p. 231117, 2013.

Chapter 4 - Ge-on-Si SACM structure: Design, modelling and SIMS measurements

4.1 Introduction

In the previous chapters, the operating principles and underlying physical processes of different photodiode structures have been described to underline the main constraints that each device needs to satisfy depending on the application for which it will be used. Typically, it is not possible to build a "*perfect*" device that meets all the requirements, nevertheless it is possible to design a device that satisfies a trade-off between different requirements. This is true for all devices previously described as p-i-n or n-i-p photodiodes, APDs and SPADs. In particular, SPADs are very similar to APDs, but they are designed to meet different performance criteria in terms of efficiency and noise. Si SPADs are mainly divided in two main categories (thick and thin-SPADs) and the same material (Si) is used for both absorption and multiplication. On the other hand, the design of a SPAD capable of detecting single-photons at wavelengths greater than the cut-off threshold of silicon (1100 nm), requires completely different design criteria than a Si SPAD. Typically, two different materials are required for both absorption and multiplication (heterostructure). In terms of all-semiconductor devices, InGaAs/InP SPADs represents the state-of-art for detecting single-photons in the telecom bands and they are mainly based on the Separate Absorption, Charge and Multiplication (SACM) structure, fabricated in a planar technology.

The Ge-on-Si SPADs presented in this work are designed based on a similar structure fabricated in a *mesa* configuration. This chapter starts by describing the design criteria used for this device, and 2D and 1D simulations of dark current and electric field performed using the modelling software SILVACO ATLAS are then given. Furthermore, data on dopant diffusion obtained by using SIMS measurements on the fabricated devices are integrated in the modelling software, and its effect on the device behaviour is analysed.

4.2 Design of a Separate Absorption, Charge and Multiplication Ge-on-Si structure

The main advantage of a SACM structure is that the photon absorption and carrier multiplication layers are spatially separated and can be individually optimised [1], [2].

A schematic SACM structure with the qualitative cross-section electric field profile is shown in Figure 4.1.

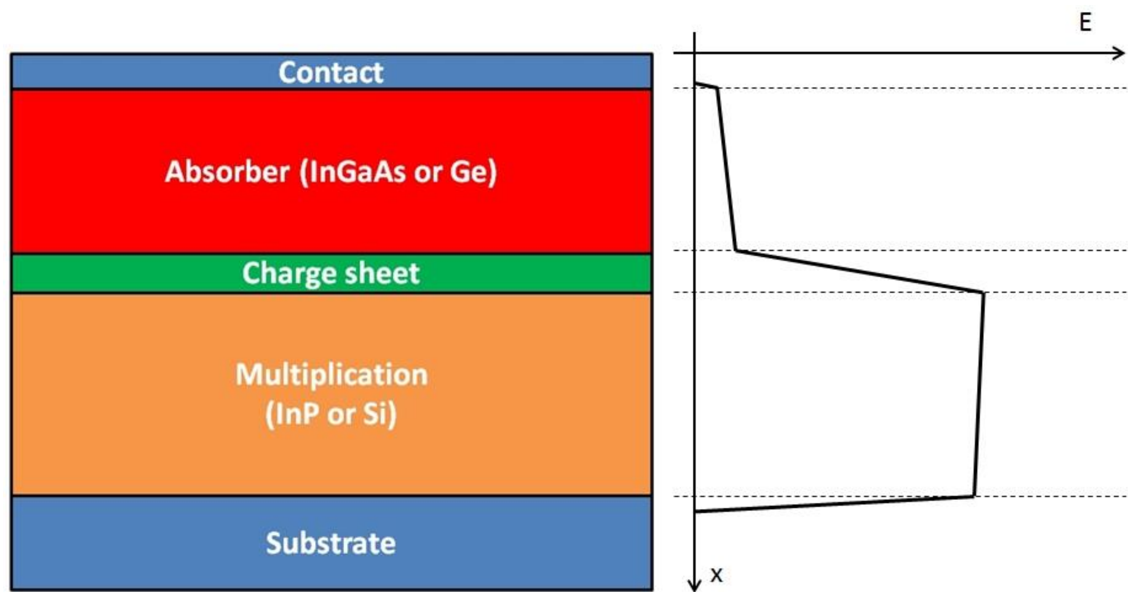


Figure 4.1. Two-dimensional sketch of a typical SACM SPAD showing the three regions in which the device can be subdivided: the absorption region (red) in which light is detected and generation of electron-hole pairs takes place; the charge region (green) that tailors the electric field between the absorption and multiplication region and photo-generated carriers are accelerated and swept into the multiplication region (orange) where the impact ionisation takes place, and the qualitative electric field profile along the device.

The structure is designed in such a way that the infrared light is absorbed in the narrow bandgap material (e.g. InGaAs or Ge), while the multiplication process take places in the wide bandgap material (e.g. InP or Si). The electric field profile along the structure is critical and must be properly designed.

In the absorber layer, the electric field must be kept below the material breakdown field in order to avoid impact ionisation and tunnelling phenomena in this layer, which will be detrimental to device performance, resulting in increased dark currents. This criterion sets a upper limit to the electric field strength in the absorption region because the electric field threshold, at which tunnelling takes place in germanium, is not as well-known as for its InGaAs counterpart [1]. However, for the SACM structure with a depleted absorber, the electric field strength in the absorption region cannot be too low; otherwise carriers in the absorption region cannot reach their saturation velocity, which will adversely affect the performance of the device. Figure 4.2 shows typical carrier characteristics of both electrons and holes in Ge, where the saturation velocity is

reached at ~30 kV/cm [2]. This value represents the lower limit for the electric field strength in the germanium absorber layer, where the doping profile must be as low as possible to ensure a constant and uniform electric field.

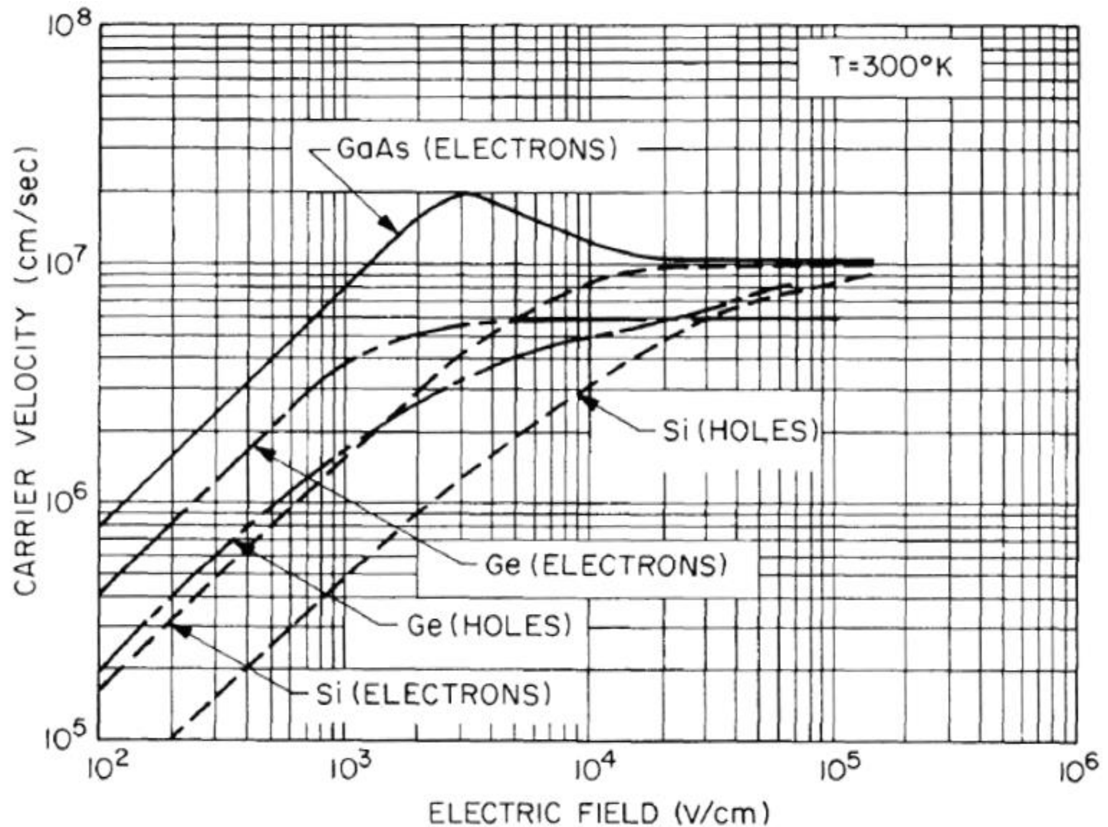


Figure 4.2. Measured carrier velocity vs. electric field for high purity Ge, Si and GaAs. In the high-field region the velocity is independent from the doping concentration [2].

On the other hand, the electric field in the multiplication layer must be kept above the threshold for impact ionisation in order to initiate an avalanche multiplication process. As for the absorber layer, its doping profile must be kept low.

To correctly tune the electric field between these two regions, according to the criteria described above, a moderately doped charge sheet layer is sandwiched between them. Therefore, it is very important to properly design the thickness and doping concentration of the charge sheet. In order to approximate the thickness and the concentration of the charge sheet, a simple linear analysis could be used, considering the breakdown fields of Si and Ge which are 300 kV/cm and 100 kV/cm, respectively. The avalanche breakdown field represents the electric field at which the onset of a self-sustaining avalanche current is observed.

By also making the assumption that the fields in the Ge absorber layer and Si multiplication layer are constant and equal to ~30 kV/cm and 300 kV/cm, respectively, the field variation in the charge sheet layer is ~270 kV/cm. The doping concentration (N) for the charge sheet layer, considering a thickness of 100 nm, can be calculated by solving Poisson's equation:

$$\frac{dE}{dx} = \frac{\rho(x)}{\varepsilon} = \frac{q \cdot N}{\varepsilon_{Si} \cdot \varepsilon_0} \tag{4.1}$$

This calculation gives a value of ~1.7×10^{17} cm^{-3}. For a given thickness of the Si multiplication region, this doping concentration also defines the punch-through voltage, the depletion region edge reaches the absorption layer. The definition of this voltage should be considered as a further design criterion for the SPAD design due to temperature considerations. The punch-through voltage remains almost unchanged when the temperature is reduced, since it will be only affected by the density of ionised donors and acceptors in the device structure. The avalanche breakdown voltage, however, will reduce significantly with decreasing temperature, as explained in Chapter 2 (section 2.3). Since the Ge-on-Si SACM SPAD has an intrinsically high dark current, low operation temperatures (at least down to 200 K) are expected to be necessary, meaning that this voltage difference between punch-through and breakdown voltage should be carefully considered to avoid a device breaking down in fields low enough that the absorber layer is not depleted.

These device design criteria mainly take into account the electric field profile and the doping concentration. However, other parameters can directly affect the performance of a SPAD such as the thickness of both absorption and multiplication regions. It is possible to establish how the thicknesses of both absorption and multiplication layers influence their performance [1], [3] in a relatively mature technology like the InGaAs/InP SPAD. The optimization of these parameters depends on the specific application considered and represents a trade-off between different performance criteria.

In this thesis, the Ge-on-Si SPAD designed is based on a SACM structure similar to that of a InGaAs/InP SPAD [6], [7], but the material system is fundamentally different (as already discussed in Chapter 3). Therefore, the optimisation of these thicknesses must meet different requirements in terms of performance and technology. In addition, InGaAs/InP SPADs are mainly fabricated using a planar geometry and use either dopant diffusion or implantation processes, which require consideration at the design stage [1].

The Ge-on-Si SPAD was designed to be fabricated in a mesa geometry and the doping of different layers was performed *in situ* during the epitaxial growth as described in this chapter. Furthermore, a critical discussion is also given in the next chapter where the experimental results and performance of the first Ge-on-Si SPAD demonstrated in the scientific literature are shown.

4.3 Ge-on-Si SPAD structure

In this section, the early design of Ge-on-Si SPAD structure is illustrated. Based on the above discussion, this design was made mainly according to the following considerations:

a) Electric field profile;
b) Charge sheet layer doping concentration;
c) Operating temperatures.

The cross-section of the designed and simulated Ge-on_Si SPAD structure is shown in Figure 4.3 with the associated doping concentrations and thicknesses for each layer.

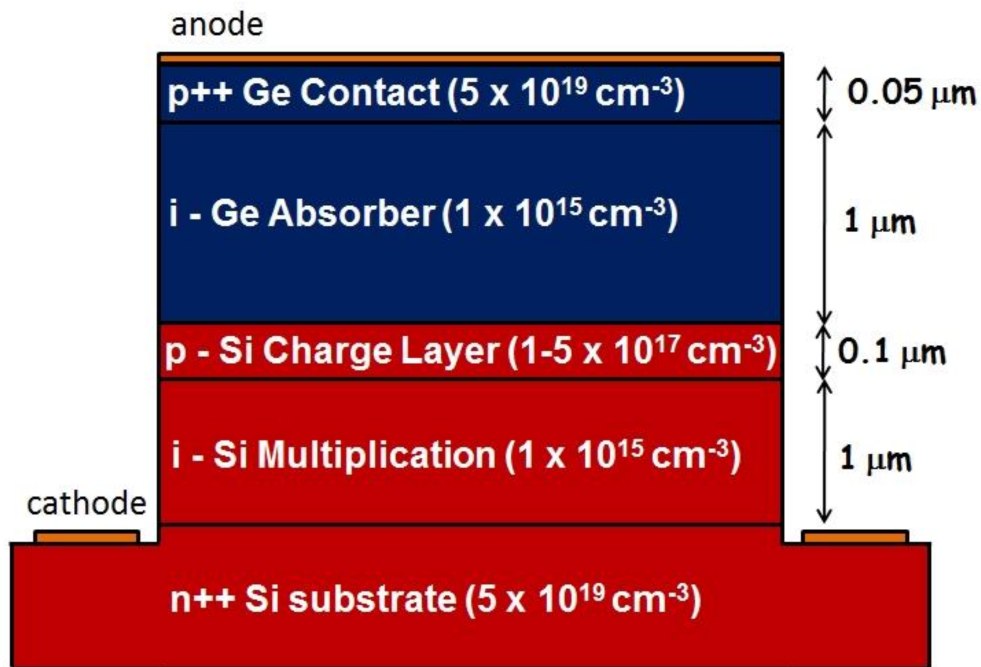


Figure 4.3.Cross-section of the designed mesa geometry Ge-on-Si SPAD structure. For each layer, the thickness is specified with the associated doping concentration (N_a or N_d). In particular, four different doping concentrations were considered for the charge sheet layer (1, 1.5, 2, and 5×10^{17} cm^{-3}).

The structure was designed in such a way that the photon absorption takes place in the Ge absorber layer, while the multiplication occurs only in the Si multiplication layer. This is similar to the APD structure which was designed to operate below the

breakdown voltage by Kang *et al.* [4], [5]. As already explained in Chapter 3, the design of APDs and SPADs needs to satisfy different requirements. Therefore, different doping concentrations in the charge sheet and a thicker multiplication region were considered for the structure proposed in this work, as shown in Figure 4.3, in comparison to the structure shown by Kang *et al.* [4], [5]. The thicker multiplication region was selected in order to increase the probability of a primary carrier triggering a self-sustaining avalanche current, while the charge sheet doping spread was used to tailor the electric field between the Ge absorber and Si multiplication layers and to allow for growth variations.

The commercially available device modelling software Silvaco ATLAS was used as a basis for the simulations [6]. In particular, the software simulates optoelectronic devices such as avalanche photodiodes in linear multiplication mode, giving important information on the detector structure. The software was mainly used to provide information on one- and two-dimensional electric field profile, carrier concentration, punch-through voltage and dark-current voltage characteristics, which were the most fundamental issues to be addressed at this stage. Different geometries were simulated and device parameters were also integrated with the experimental data extrapolated from real devices.

As shown in Figure 4.3, a highly doped p++ Ge contact layer was considered for the structure in order to make an ohmic contact (anode) followed by unintentionally doped Ge absorption layer in which the doping must be kept fairly low (10^{15} cm^{-3}) to guarantee a constant and uniform electric field. The absorption region was separated from the n-doped (10^{15} cm^{-3}) Si multiplication layer by a p-doped Si charge layer. In particular, four doping concentrations (1, 1.5, 2 and 5×10^{17} cm^{-3}) of the charge sheet layer were considered in order to perform simulations around the design parameter (as shown in the previous paragraph) and tune the electric field profile between the absorption and multiplication layer. In addition, an n++ Si substrate was used for the bottom contact (cathode). The modelling did not take into account the band-to-band and trap-assisted tunnelling, as well as traps and defects at the Ge-Si heterointerface which are not well reported in the scientific literature.

When simulating the SPAD device in terms of I-V characteristic, the Drift-Diffusion physical model was implemented, which solves the three equations (Poisson's equation plus the continuity equation for electrons and holes) for charge transport in

semiconductor devices. The Shockley-Read-Hall model was used to take into account recombination and generation in the semiconductors. The values of electron and hole lifetimes used for germanium were the ones calculated by Colace *et al.* [7], [8]. To describe the impact ionisation process, Selberherr's model was used, which includes the strong dependence on the impact ionisation coefficients of the electric field and temperature [9]. In the following section, the modelling of the Ge-on-Si SPAD structure is given.

4.4 Modelling of the Ge-on-Si hetero-structure

The definition of the SPAD structure and the generation of the appropriate mesh constitute the first steps in the simulation process. The specification of the mesh involves a trade-off between the requirements of accuracy and numerical efficiency [10]. Accuracy requires a fine mesh that can resolve all significant features of the solution. Numerical efficiency requires a coarse mesh that minimises the total number of grid points. Figure 4.4 shows the mesh used for the device simulation considering the design parameters which have been previously described in section 4.3.

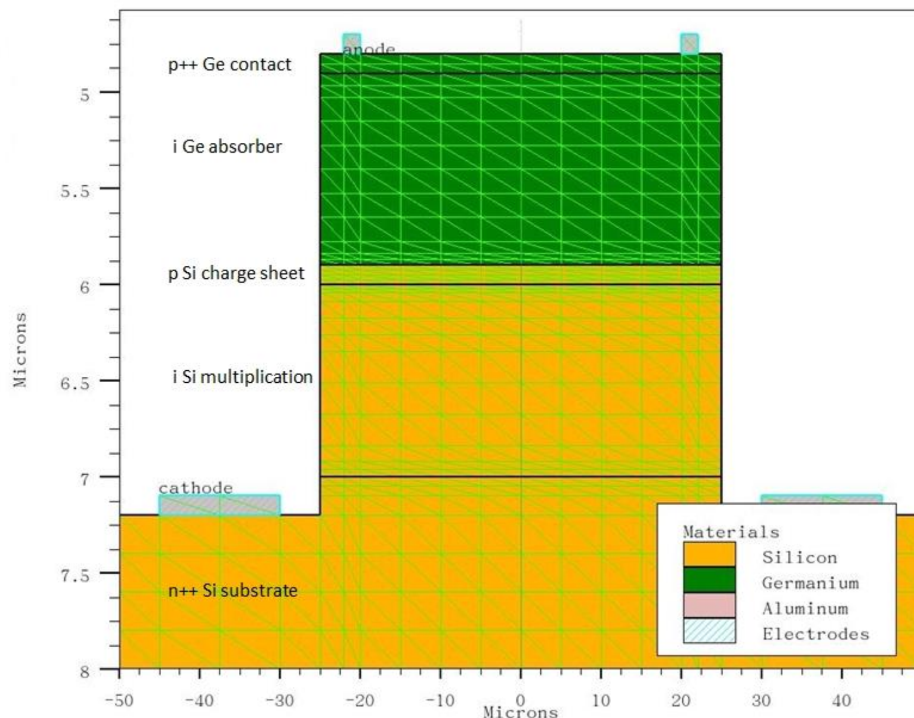


Figure 4.4. Ge-on-Si SPAD structure with the mesh used for the device simulation in Silvaco ATLAS.

A fine mesh was defined at the interface between different materials as well as in regions where a high electric field was expected (Figure 4.4). This mesh definition was a good compromise between accuracy and numerical efficiency for the simulated

structure. The simulated device was cylindrical with a diameter of 50 μm, as shown in Figure 4.4.

The simulated I-V characteristic for the structure (Figure 4.3) and four different doping concentrations of the charge sheet layer are shown in Figure 4.5.

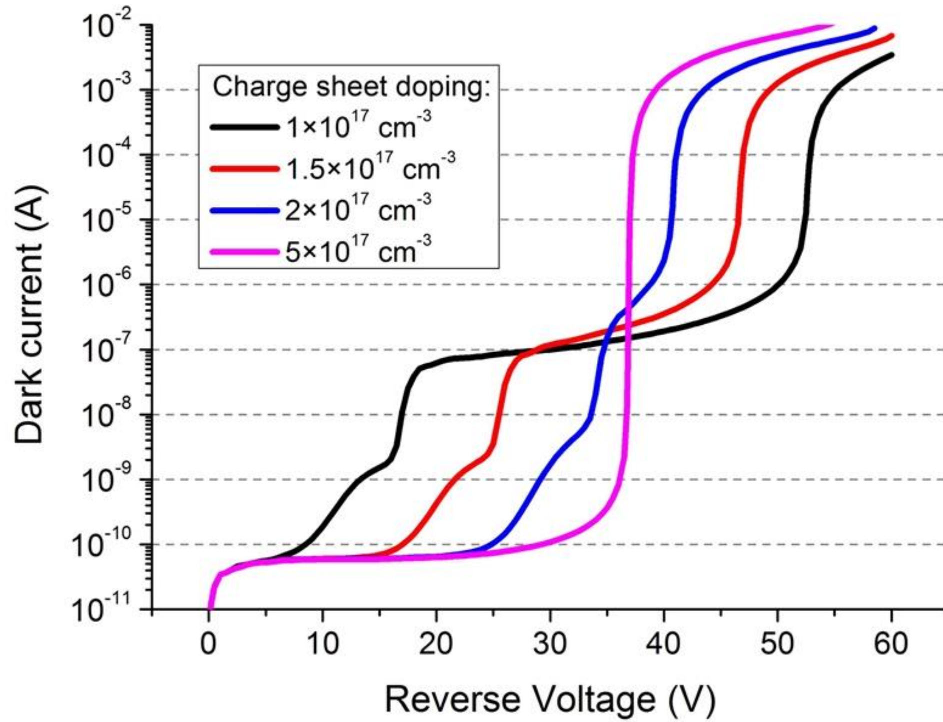


Figure 4.5. Simulated I-V characteristics for the designed structure (reported in Figure 4.3) as a function of the doping concentration of the charge sheet layer.

Depending on the doping concentration of the charge sheet layer, the designed structure had different values of the punch-through (V_{PT}) and breakdown (V_{BD}) voltages. These values are summarised in Table II.

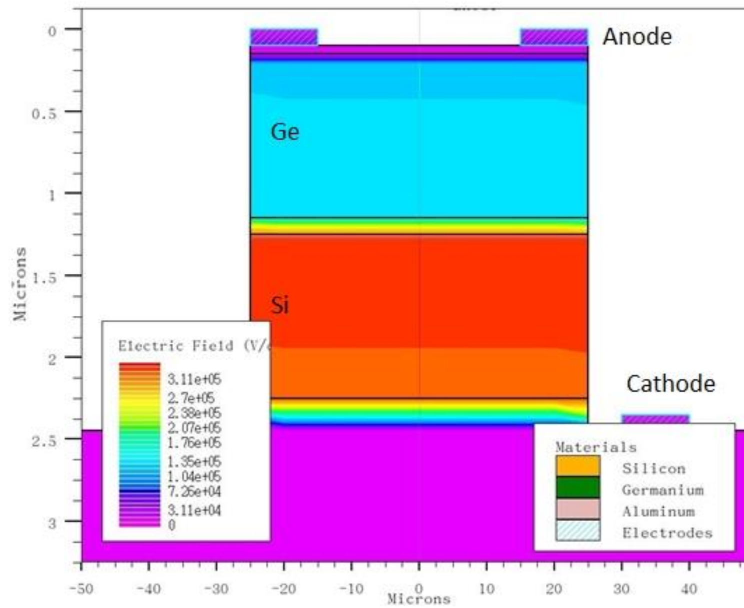
Charge sheet doping concentration (cm^{-3})	V_{PT} (V)	V_{BD} (V)
1×10^{17}	- 18	- 53
1.5×10^{17}	- 27	- 46.5
2×10^{17}	- 35	- 42
5×10^{17}	Undepleted	- 36.5

Table II. Summary of the punch-through and breakdown voltages for the simulated structure (Figure 4.3, for four different doping concentrations of the charge sheet layer.

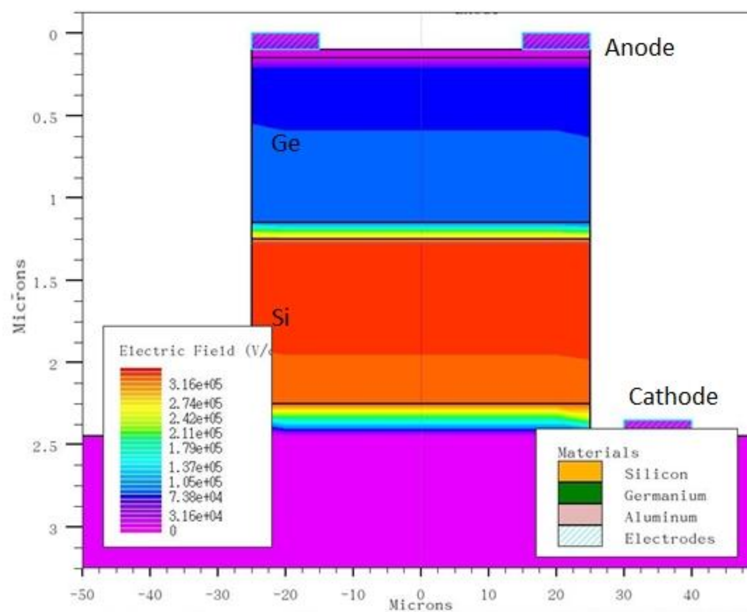
In particular, the I-V characteristic for the highest doping concentration of the charge sheet layer (5×10^{17} cm^{-3}), shown as the magenta line in Figure 4.5, showed the lowest dark current without any punch-through, meaning that the Ge absorber layer was not depleted and the electric field was completely confined in the Si multiplication layer. In fact, this result was further supported by the lowest breakdown voltage obtained, which occurred at ~ -36.5 V.

On the other hand, the structures with a doping concentration of 1, 1.5 and 2×10^{17} cm^{-3} (black, red and blue line as shown in Figure 4.5) respectively, had different V_{PT} and V_{BD} voltages as summarised in Table II. When the depletion region starts to extend into the Ge absorber layer (punch-through), the leakage current increases because of carriers generated through generation-recombination centres. These were taken into account in the simulations through the long (ns) electron and hole lifetimes in the Ge absorber layer and the high intrinsic carrier concentration, as clearly shown in Figure 4.5. Thermal generation is an intrinsic property related to the quality of the material itself and it is unavoidable. Nevertheless, it could be mitigated by using lower operational temperatures.

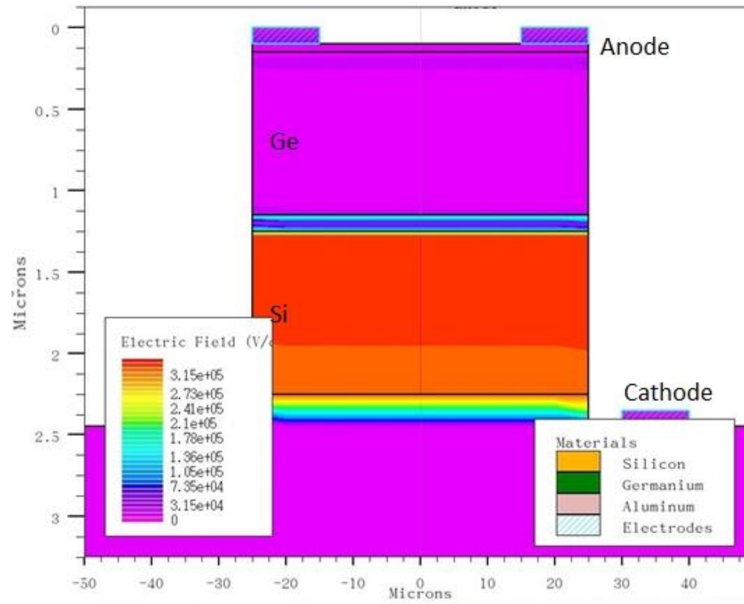
To further understand the effect of the charge sheet doping concentration on the simulated structures, Figure 4.6a-d and 4.7 show the simulated 2D and 1D electric field profile obtained at 95 % of V_{BD} respectively.



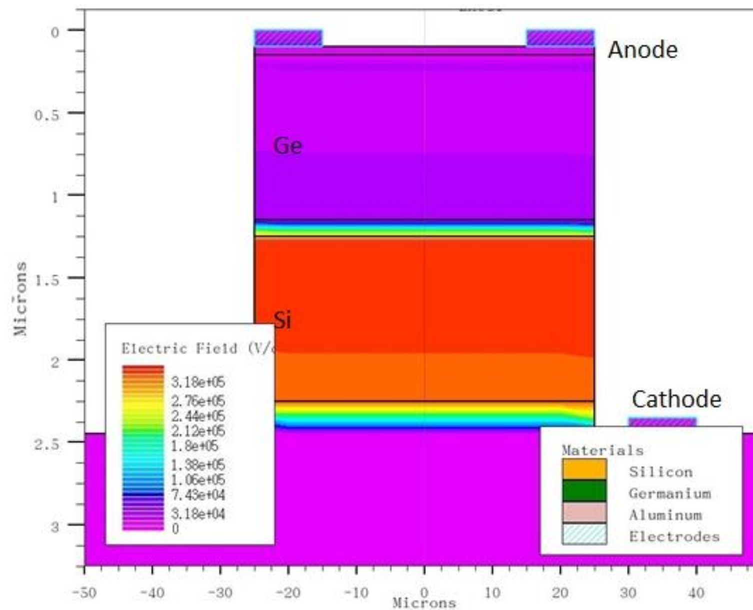
(a)



(b)



(c)



(d)

Figure 4.6. Simulated 2D electric field profile at 95% of V_{BD} for the structure shown in Figure 4.3 and different doping concentration of the charge sheet layer: (a) 1×10^{17} cm^{-3}, (b) 1.5×10^{17} cm^{-3}, (c) 2×10^{17} cm^{-3}, and (d) 5×10^{17} cm^{-3}.

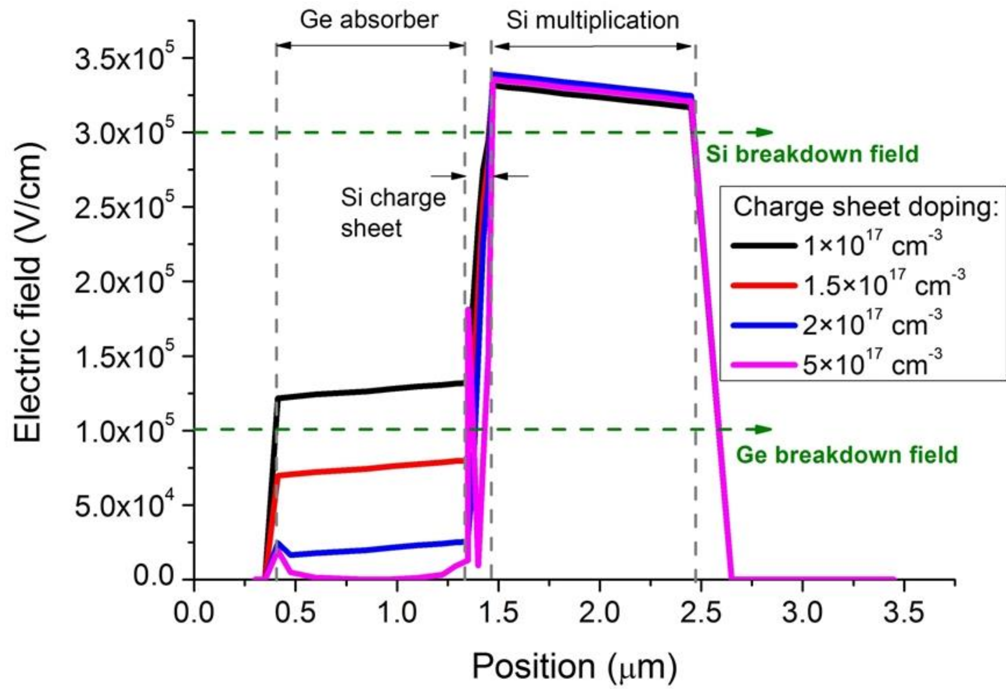


Figure 4.7. Simulated electric field profile at 95% of V_{BD} along section x=0 μm of the device as a function of the doping concentration of the charge sheet layer.

In Figure 4.6a-d, the simulated 2D electric field profile for the four doping concentrations of the charge sheet layer is given. Although this layer plays a different role in tuning the electric field between the Ge absorber and Si multiplication layers, the simulation shows that the electric field is well confined through the active area of the device, without any premature edge breakdown at the sides of the mesa could adversely affect the performance of the device. Furthermore, Figure 4.6a-b shows that both the 1 μm-thick Ge absorber and Si multiplication layers are fully depleted before the breakdown voltage. On the other hand, the simulated 1D electric field profile (Figure 4.7) obtained in the centre of the active area at x=0 μm, showed that both structures (black and red lines in Figure 4.7) meet the electric field requirements in the Si multiplication layer. However, in the Ge absorber layer the simulated device with a charge sheet doping concentration of 1×10^{17} cm^{-3} (black line) had an electric field which was higher than the Ge breakdown field, while for a doping concentration of 1.5 and 2×10^{17} cm^{-3} (red and blue line), the electric field in the absorption layer was high enough to reach the carriers saturation velocity and well below the Ge breakdown field.

Although the plotted electric field profiles have been simulated below the breakdown voltage, it is well known that most of the SPAD parameters are strongly affected by the overvoltage (i.e. the excess bias above the avalanche breakdown) applied to the detector. Since improvements for some of these parameters can be observed as the

excess bias is increased, while other parameters get worse, the choice of the optimum overvoltage that should be applied depends on the specific applications. Silvaco ATLAS does not support simulations of the electric field profile above the avalanche breakdown, but the excess voltage above the avalanche breakdown can be observed as the area in between the two curves (shaded in Figure 4.8), and is given by the variation of the electric field ΔE times the thickness of the space charge region.

Among the different simulated structures, a doping concentration of 1.5 and 2×10^{17} cm^{-3} satisfy all the design criteria in terms of electric field, as described in paragraph 4.2. The charge sheet layer plays an important role in the design, meaning that a small variation of its charge, which is related to thickness and doping concentration, leads to completely different device performance. This is mainly true in real devices for which it is very difficult to accurately control their doping concentration during the epitaxial growth (this is clearly shown later in section 4.9).

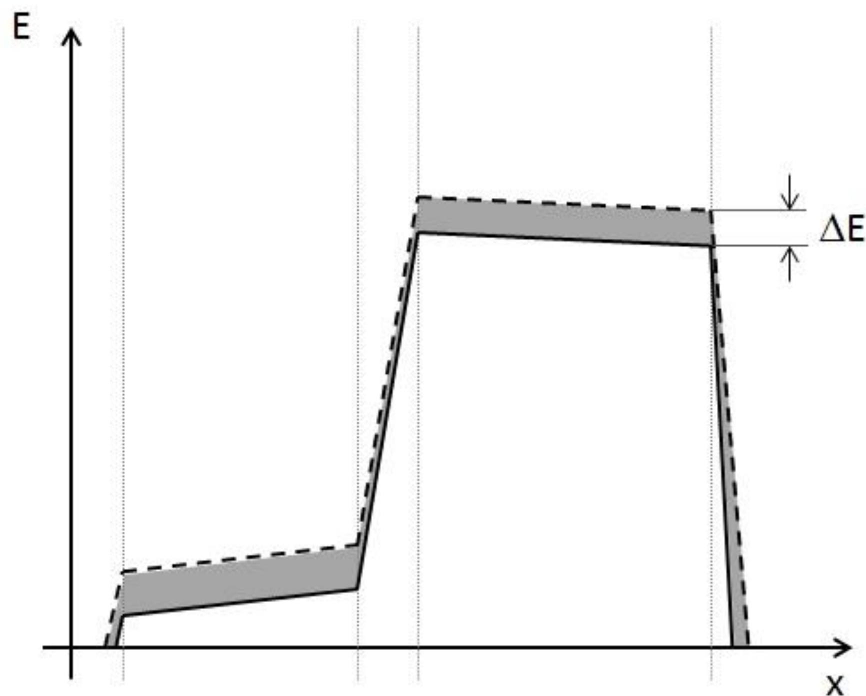


Figure 4.8. Sketch of the electric field for a SACM Ge-on-Si SPAD at the breakdown (solid line) and with an overvoltage applied (dashed line) to the device.

4.5 Effect of temperature on the breakdown voltage

The last criterion considered for the design of the Ge-on-Si structure was the temperature. Simulations at different temperatures were performed to evaluate the variation of the breakdown voltage (shown in Figure 4.9) for the structure with a doping concentration of the charge sheet layer equal to 1.5×10^{17} cm^{-3}. Simulations were

performed above 175 K because of convergence problems related to the algorithm at lower temperature.

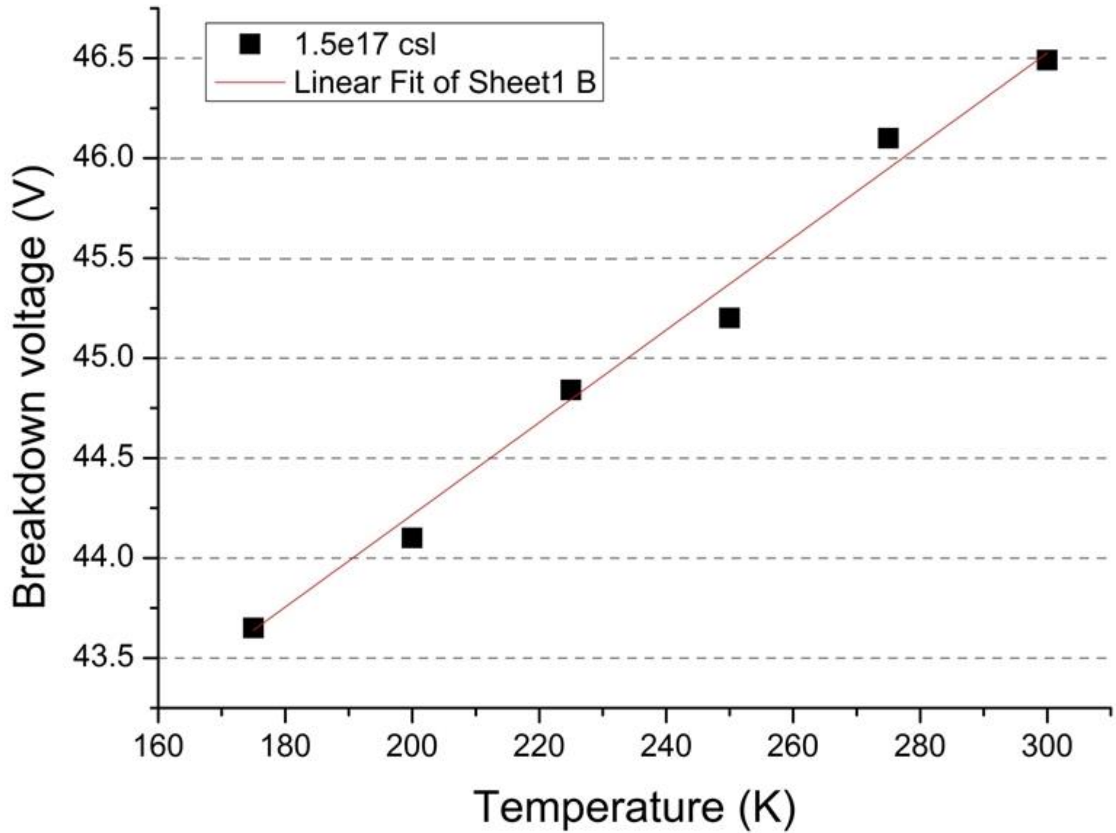


Figure 4.9. Simulated variation of the breakdown voltage with the temperature for the structure with a doping concentration of 1.5 × 10^{17} cm^{-3} in the charge sheet layer.

As the temperature decreases, the lattice vibrational energy is also reduced, which reduces the scattering of electrons and holes. This increases the energy of the carriers and hence their ionisation rates. The temperature dependence of the ionisation rates was taken into account using Selberherr's impact ionisation model for the simulations [9]. As a consequence of the change of the ionisation coefficients, as the temperature decreases, the electric field should be decreased to maintain the breakdown. The breakdown voltage drops with temperature as shown by the following formula:

$$\frac{\Delta V_{BD}}{V_{BD,RT}} = \gamma \Delta T \tag{4.2}$$

where $V_{BD,RT}$ is the breakdown voltage at room temperature, and γ is the temperature coefficient. For the simulated structure, γ is found to be equal to ~ 0.05 %/K, which is in line with the experimental values quoted in the scientific literature for silicon [4], [11]. This means that for the designed structure, with a temperature decrease of 100 K

there is a ~ 2.3 V reduction of the breakdown voltage. Since the device has been designed for temperatures below 150 K due to the high dark current, a good design rule would be to maintain a difference between punch-through and avalanche breakdown voltage at room temperature of at least 5 V, in order to avoid avalanche breakdown of the device before punch-through at lower temperatures. For this reason, it is possible to establish if a Ge-on-Si structure with a doping concentration of the charge sheet layer between 1.5 and 2×10^{17}cm^{-3} satisfies all the design criteria in terms of electric field and temperature, according to values shown in Table II.

4.6 Capacitance-Voltage Measurements on the 1st generation of Ge-on-Si SPAD structures

The simulated structures proposed (Figure 4.3) were grown and fabricated by our collaborators at University of Warwick (growth) and Glasgow (fabrication). Although more details about these two aspects will be given in the next chapter, a summary of the wafer IDs for the 1st generation of Ge-on-Si SPAD structure growth is given in Table III.

Charge sheet doping concentration (cm^{-3})	**Wafer ID**
$\mathbf{1 \times 10^{17}}$	11-167
$\mathbf{2 \times 10^{17}}$	11-141
$\mathbf{5 \times 10^{17}}$	11-142

Table III. Wafer ID related to each doping concentration of the charge sheet layer, as grown.

The SPAD structures have been grown starting from a highly phosphorous doped (~ 5×10^{19} cm^{-3}) Si substrate, while boron was used for the doping of the charge sheet layer as well as the highly doped Ge contact layer (~ 2×10^{19} cm^{-3}). To check the doping concentration of the Ge absorber layer, C-V measurements were performed from our collaborator in Sheffield on two different samples:

- Wafer 11-167;
- Wafer 11-318 shown in Figure 4.10, where 1 μm-thick intrinsic Ge was directly grown on top of a highly doped Si substrate.

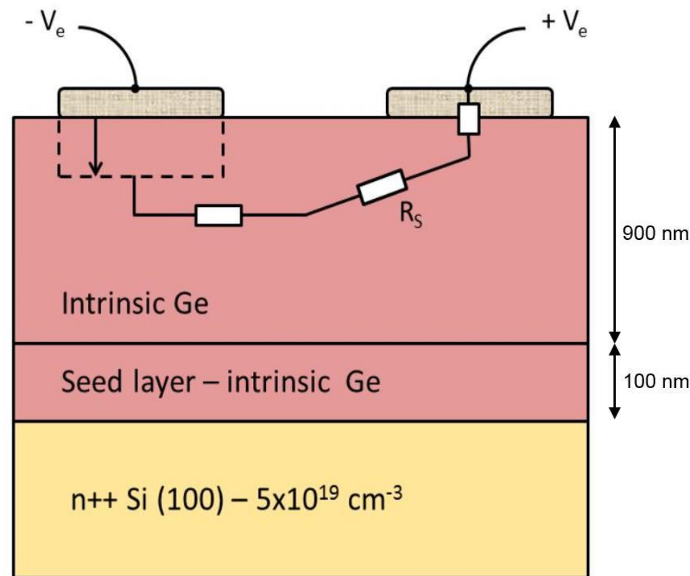


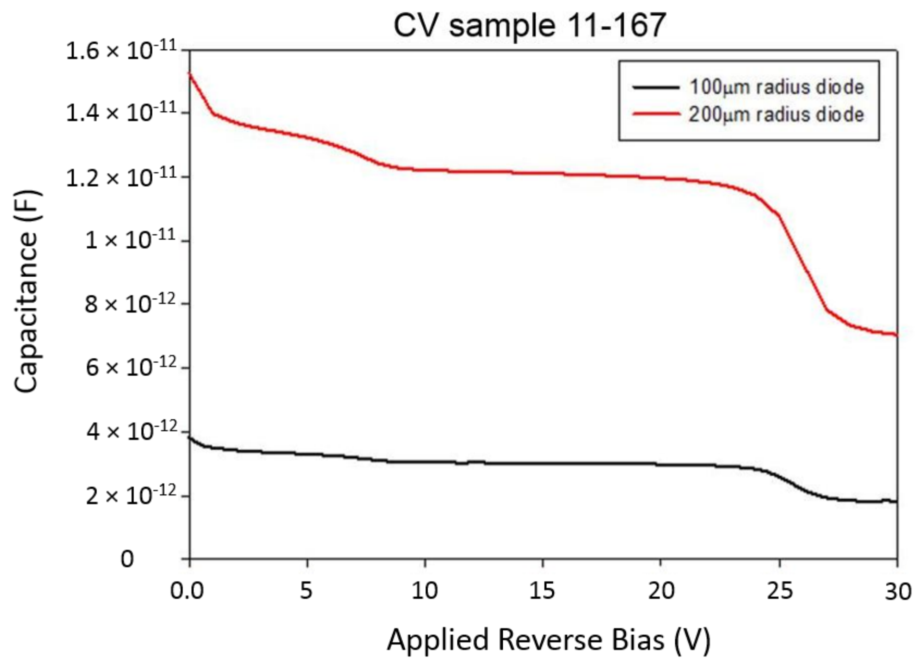
Figure 4.10. Structure of sample 11-318, which was grown to check the doping concentration of the intrinsic germanium by lateral C-V measurements.

Lateral C-V measurements were performed to find the doping level of intrinsic Ge in sample 11-318, as shown in Figure 4.10. There are several advantages to this method:

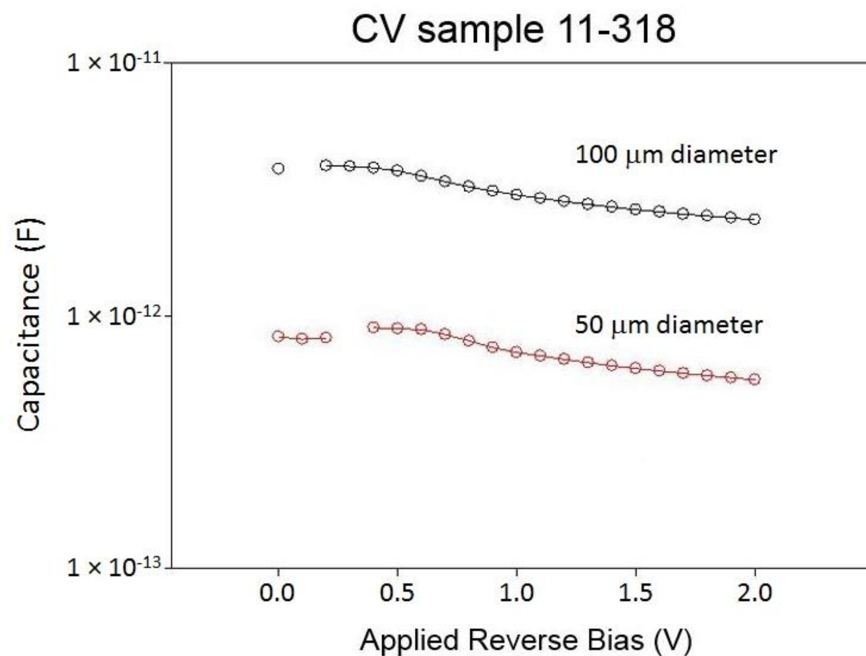
a) The junction formed between Ge and Si is excluded;
b) There is no need to etch the sample into diodes with different sizes;
c) A lateral CV profile is used instead of the standard longitudinal one (as for sample 11-167) in order to avoid the unknown effect of the high density of defects in the seed layer (e.g. to establish if the measured junction is a Si/Ge seed layer or a Ge/contact junction).

However, sample 11-318 required two Schottky contacts, which were grown by vacuum evaporation with composition of 20 nm Titanium/200 nm Gold. The distance between metals was set to 5-10 µm based on the mask alignment, and the contact diameter varied between 50 and 200 µm.

Results of C-V characteristics measured for both samples are presented in Figure 4.10a-b for devices with different diameters.



(a)



(b)

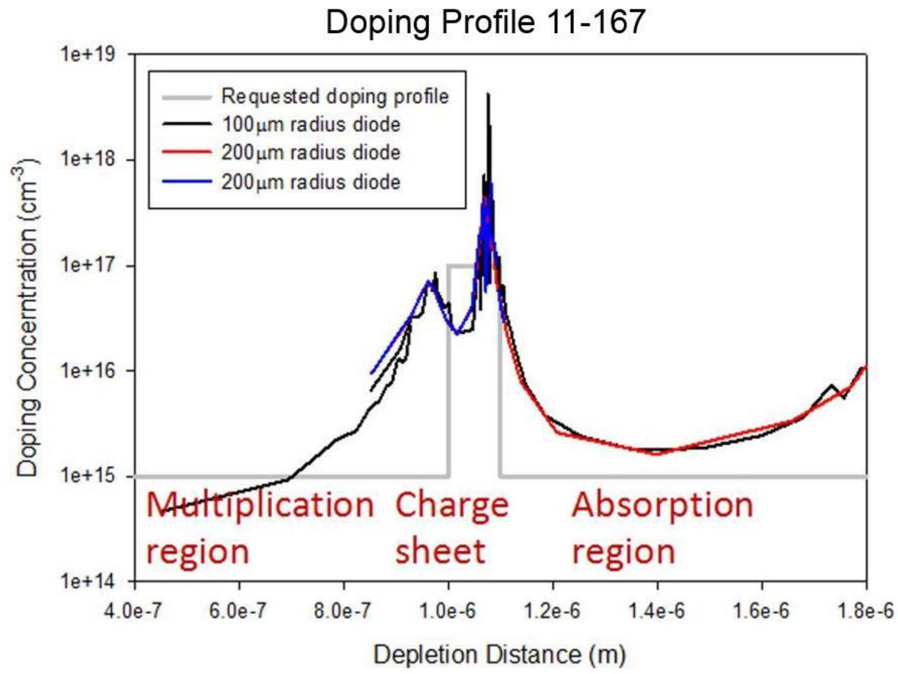
Figure 4.11. C-V measurements performed on sample 11-167 (a) and 11-318 (b) to ascertain the doping concentration of the intrinsic Ge absorber layer.

The C-V measurements performed on sample 11-167 (Figure 4.11a) gives also an indication of the punch-through voltage, which corresponds to the drop in the capacitance value at ~ -25 V. At this voltage, the charge space region extends to the Ge absorber layer and consequently the capacitance of the device decreases. As expected for both samples, the smaller devices exhibited a lower capacitance value.

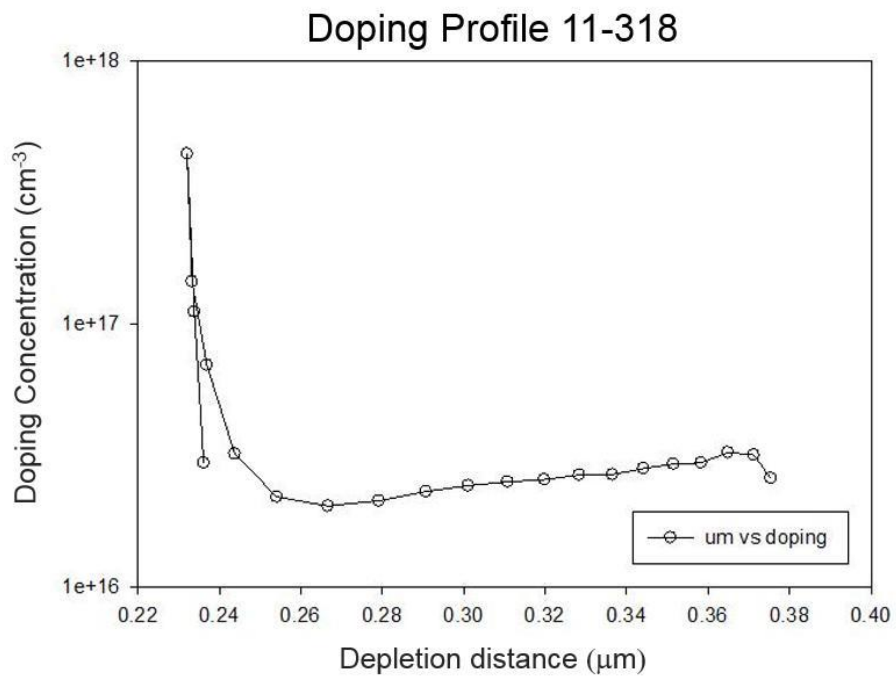
Once the capacitance against the reverse bias has been measured, it is possible to approximate the doping concentration as a function of the depletion width by using the following formula:

$$N_D(W) = \frac{2}{q\epsilon_S A^2 d(1/C^2)/dV} \tag{4.3}$$

Results for both samples are illustrated in Figure 4.11a-b.



(a)



(b)

Figure 4.12. Doping concentration as a function of the depletion distance for (a) a Ge-on-Si device of sample 11-167 and three different size devices: 100 μm (black), 200 μm (red and blue). Grey line indicates the designed doping profile. (b) Doping concentration vs. depletion distance for sample 11-318 (Figure 4.10). The x scale refers to the depletion distance in the Ge absorber layer. .

Two different doping concentrations of the Ge absorber layer were calculated, 5×10^{15} cm^{-3} and 2×10^{16} cm^{-3} for samples 11-167 and 11-138, respectively. It was assumed that for the sample 11-138, the higher doping concentration of the Ge absorber layer is due to the different doping levels of Si on which the germanium was grown. Dopants from the highly doped phosphorous Si substrate in sample 11-318 could have segregated or diffused during the cyclic annealing steps which were performed during the heteroepitaxial growth to reduce the threading dislocation density. This effect could lead to a higher doping level in the absorber layer. This hypothesis was further supported by the fact that n-type dopants diffuse quite rapidly in germanium because defects in germanium are mainly vacancies and behave electrically as p-type dopant by enhancing donor diffusion [12], [13]. Although a low doping concentration of the Ge absorber layer is always desirable, both doping levels, 5×10^{15} and 2×10^{16} cm^{-3}, can be regarded as acceptable values for the designed SPAD structure, as shown in the next section.

As shown in Figure 4.12a, the higher peak at the interface Si charge sheet / Ge seed layer might be an effect due to the threading dislocations in the Ge seed layer (which might cause a capacitive effect due to trapping/de-trapping). Furthermore, we believe that the deeper peak was due to the charge sheet layer. However, the doping concentration of the charge sheet layer (obtained from the performed C-V measurements on sample 11-167) was not very accurate and SIMS measurements were used to ascertain both the doping concentration of the charge sheet and multiplication layer, as shown in the next sections.

4.7 Impact of the Ge absorber layer doping on the SPAD structure

The effect of a higher doping concentration in the Ge has been simulated, as shown in the plotted 1D electric field profile at 95 % of V_{BD} for two different doping concentrations of the charge sheet layer, 1 and 2×10^{17} cm^{-3} (Figure 4.13).

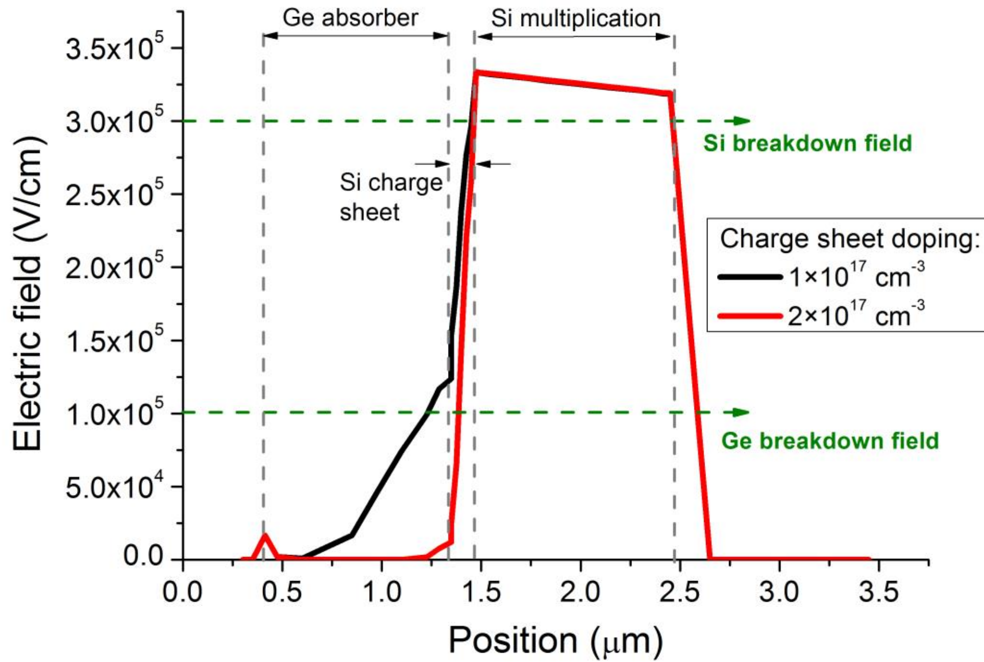


Figure 4.13. Simulated electric field profile at 95% of V_{BD} along section x=0 μm, for the structure in Figure 4.4 taking into account a higher doping concentration in the Ge absorber layer equal to 2 × 10^{16} cm^{-3} and, two different doping concentrations of the charge sheet layer.

As illustrated in Figure 4.13, when a higher doping concentration of the absorber layer (2 × 10^{16} cm^{-3}) was simulated, the electric field was entirely confined in the Si multiplication layer without depleting the Ge absorber layer for a doping concentration of the charge sheet equal to 2 × 10^{17} cm^{-3} (red line), resulting in a device not usable to detect infrared light. Therefore, a lower doping concentration of the charge sheet, 1 × 10^{17} cm^{-3} (black line), was used in the modelling. Results from simulations showed that the Ge absorber layer is partially undepleted, and hence a lower detection efficiency can be expected.

Simulations showed that an order of magnitude higher doping concentration of the absorber layer than the designed value, can have a negative impact on the structure, resulting in a device almost unusable or with deterioration of its performance.

4.8 SIMS measurements on the 2nd generation of SPAD structure

It was not possible to check the doping concentrations of the Si charge sheet and multiplication layer by performing secondary ion mass spectrometry (SIMS)

measurements on the 1st generation SPAD (wafers 11-167, 11-141 and 11-142) structures due to the lack of material which had been mainly used for ohmic contact test, sidewall etch study and device fabrication. However, two structures, one identical to wafer 11-141 (2×10^{17} cm^{-3} charge sheet doping concentration) and an all-Si APD (Figure 4.14), were grown under exactly the same conditions as for the 1st generation SPAD. The wafer IDs for these two structures were 12-027 for the Ge-on-Si SPAD and 12-028 for the all-Si APD. In addition, the n-type doping was made by phosphorous, while the p-type doping was made by using boron, as used for the 1st generation structures.

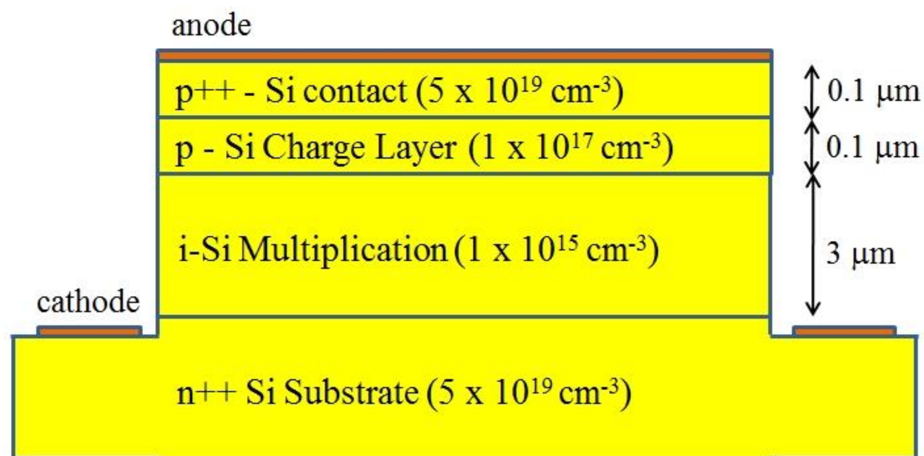
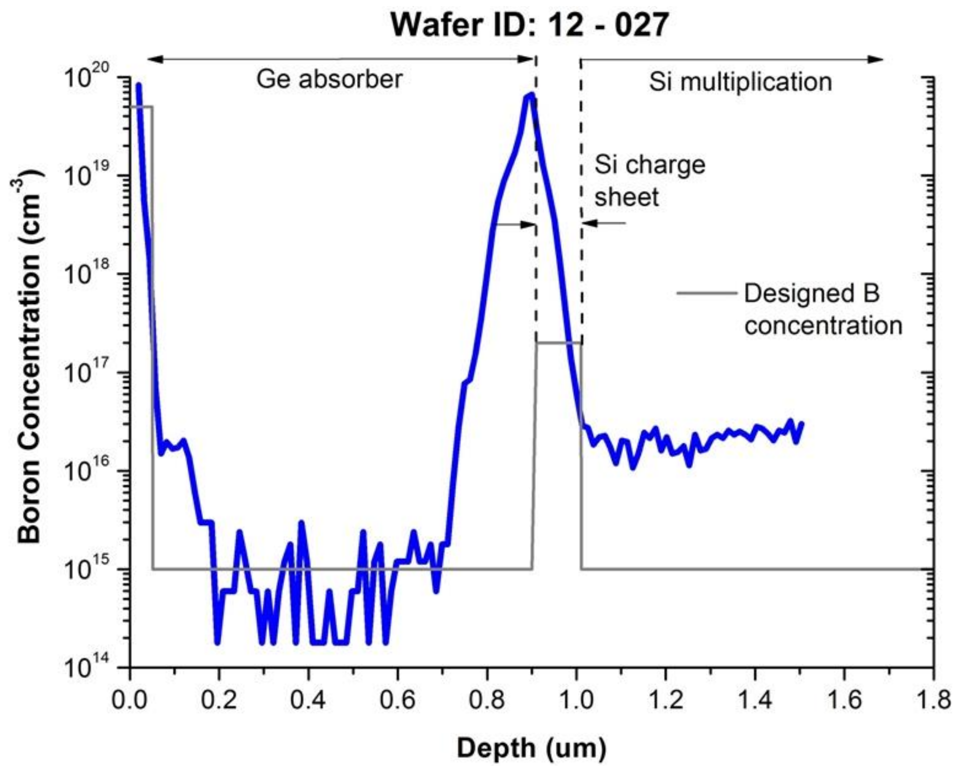


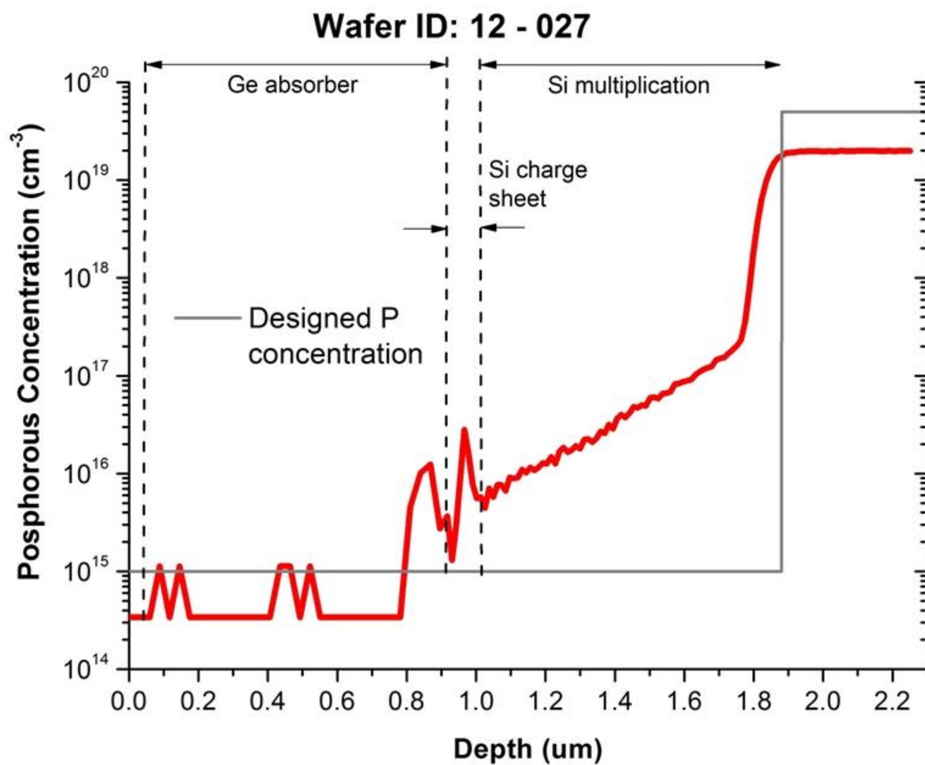
Figure 4.14. Cross-section of the all-Si APD (wafer ID 12-028) which was grown to study the avalanche multiplication properties of Si.

In particular, wafer 12-028 consisted of 3 μm thick intrinsic (1×10^{15} cm^{-3}) Si grown on top of a highly phosphorous doped (5×10^{19} cm^{-3}) Si substrate. A 100 nm-thick boron doped (1×10^{17} cm^{-3}) Si charge layer was then grown, followed by 100 nm-thick highly boron doped (5×10^{19} cm^{-3}) Si contact layer.

Figure 4.15 and 4.16 show the SIMS measurements performed on wafers, 12-027 and 12-028, respectively, by the EAG (Evans Analytic Group) company in USA.

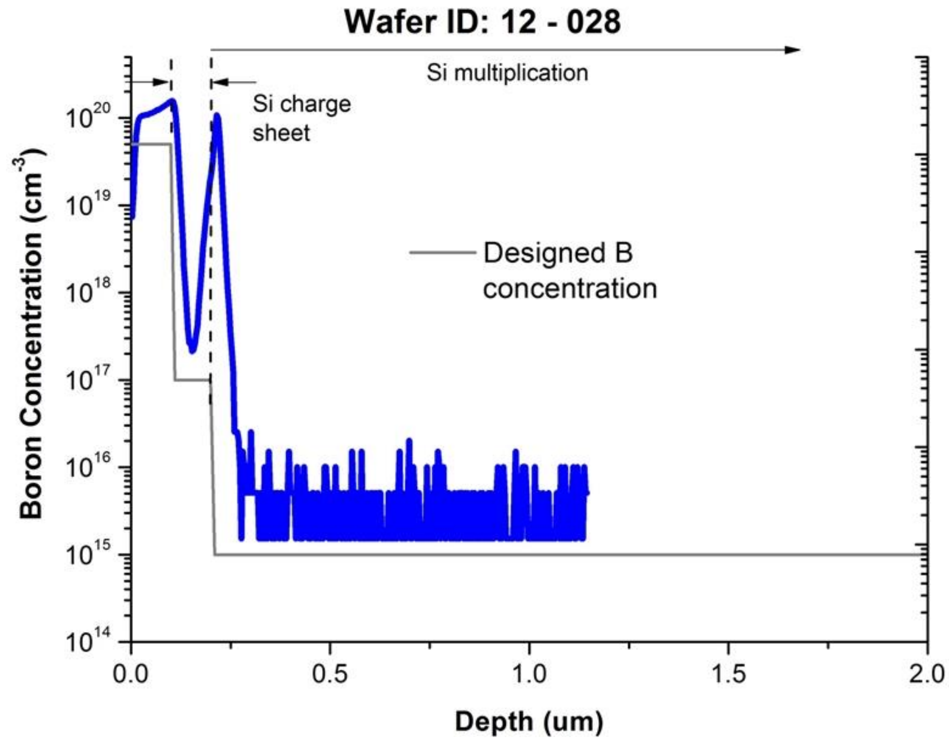


(a)

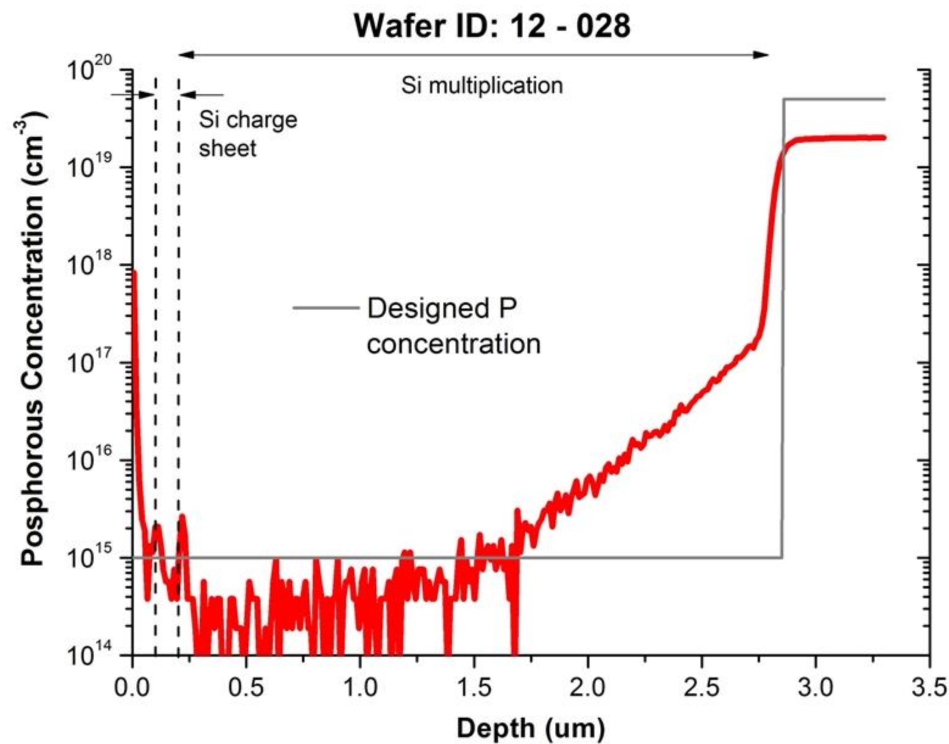


(b)

Figure 4.15. SIMS measurements performed on the wafer labelled 12-027 (Ge-on-Si SPAD) which shows (a) the Boron (blue) and (b) Phosphorous (red) doping profile as a function of the depth across the structure. Grey lines in (a) and (b) show the designed B and P doping concentrations for the structure, respectively.



(a)



(b)

Figure 4.16. SIMS measurements performed on wafer labelled 12-028 (all-Si APD) which shows (a) the Boron (blue) and (b) Phosphorous (red) doping profile as a function of the depth across the structure. Grey line in (a) and (b) shows the designed B and P doping concentration for the structure, respectively.

It is clear from Figure 4.15a that the boron concentration (blue) in the intrinsic Ge absorber layer was quite low (~ 10^{15} cm^{-3}), reaching the SIMS detection limit for boron (~ 10^{15} cm^{-3}). The measured boron doping concentration was similar to the doping concentration of the Ge absorber layer of wafer 11-167 extracted from the C-V profile performed by our collaborator in Sheffield.

On the other hand, various problems related to these wafer can be seen in Figure 4.15a-b and 4.16a-b and are summarised as follows:

1. The doping concentration (boron) of the Si charge sheet layer for both structures (blue line) was much higher (~ 10^{20} cm^{-3}) than the designed values (grey line), 2 $\times$ 10^{17} cm^{-3} and 1 $\times$ 10^{17} cm^{-3} for wafer 12-027 and 12-028, respectively. According to the epitaxial growers, this may have been caused by a problem related to a valve of the growth apparatus which controls the boron flux in the chamber, leading to an unexpected high doping concentration of the charge sheet.
2. The phosphorous concentration (red line) in the intrinsic Si multiplication layer was higher than the designed concentration showing a dopant diffusion tail which slowly decreases away from the Si substrate. Furthermore, sample 12-028, with a 3µm-thick Si intrinsic region, showed the same dopant diffusion tail which reaches the SIMS detection limit (~ 10^{15} cm^{-3}), and hence the design doping concentration after growth of approximately 1.2 µm away from the highly doped Si substrate.
3. Wafers 12-027 and 12-028, which were designed with different thicknesses of the Si intrinsic region, showed similar P doping profiles, as shown in Figure 4.17. It should be underlined that wafer 12-028 did not include any Ge layer. Furthermore, it is important to notice that wafer 12-028 did not go through all the annealing steps which must be performed to reduce the threading dislocations in the Ge-on-Si SPAD structure (wafer 12-027). Although these steps have not been performed, the P doping profile of wafer 12-028 (red line) was identical to the P profile of wafer 12-027 (black line). This behaviour suggested that the main contribution to the doping tail observed in both samples is probably due to the dopant segregation or diffusion from the highly doped Si substrate into the Si intrinsic multiplication region.

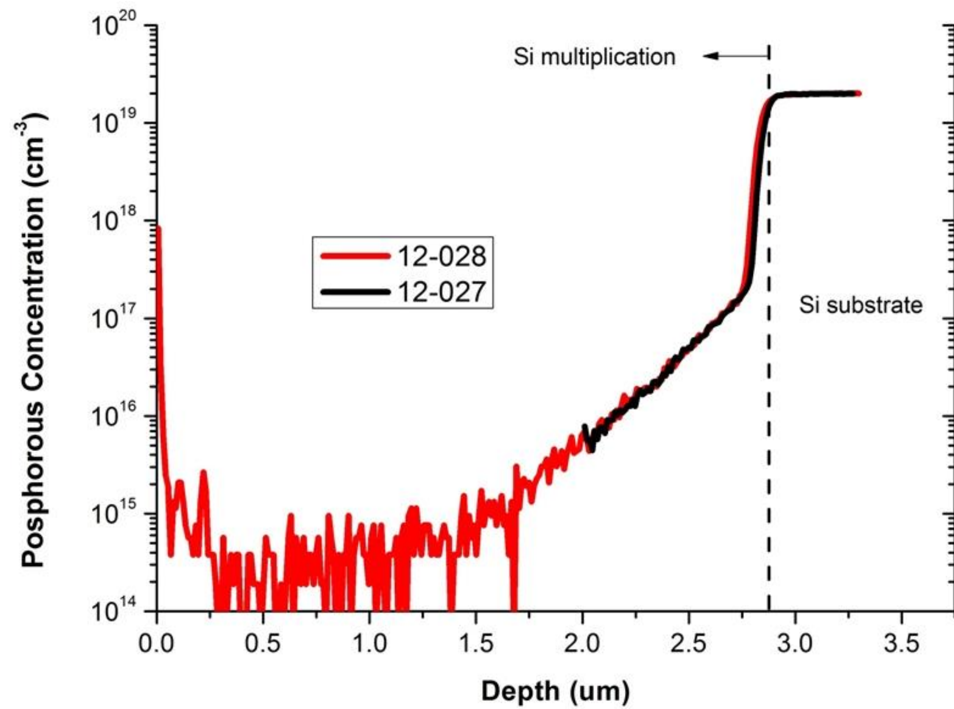
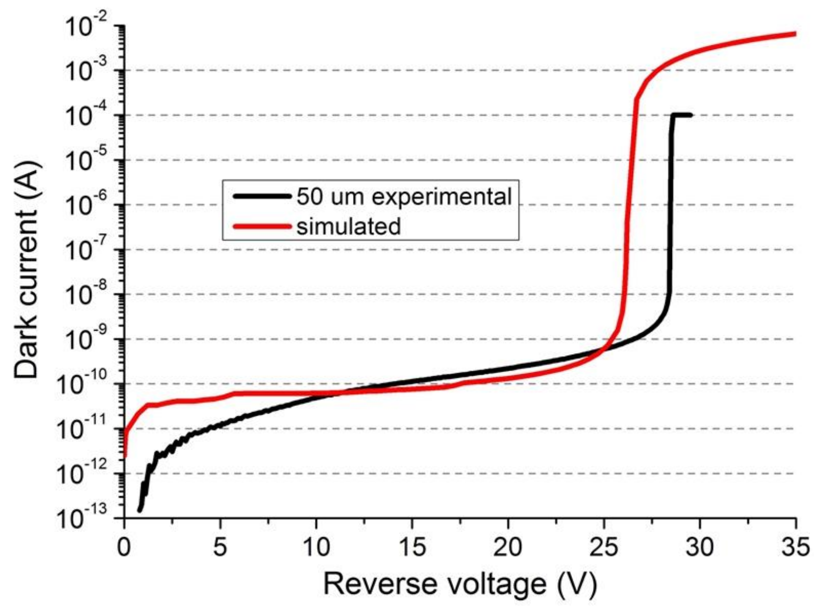


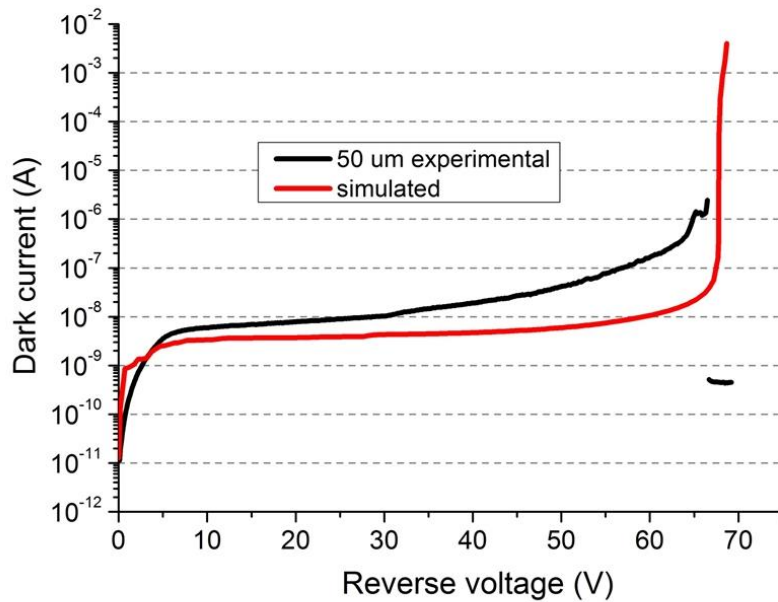
Figure 4.17. A comparison of the SIMS measurements of the Phosphorous doping profile obtained through the Si substrate and multiplication layer for both structures. The black line shows the P profile in the wafer labelled 12-027 (Ge-on-Si SPAD), while the red line is the P doping profile in the wafer labelled 12028 (Si APD).

4.9 Impact of the Si charge sheet and multiplication layer doping profile on the SPAD structure

The effects of both a high doping concentration in the charge sheet layer and a diffusion tail in the Si multiplication layer were evaluated by using the doping profile found in the above SIMS results to model the electric field profile. The simulated I-V characteristic was compared with the experimental I-V measured from devices fabricated for both wafers 12-027 and 12-028 is illustrated in Figure 4.18a-b.



(a)

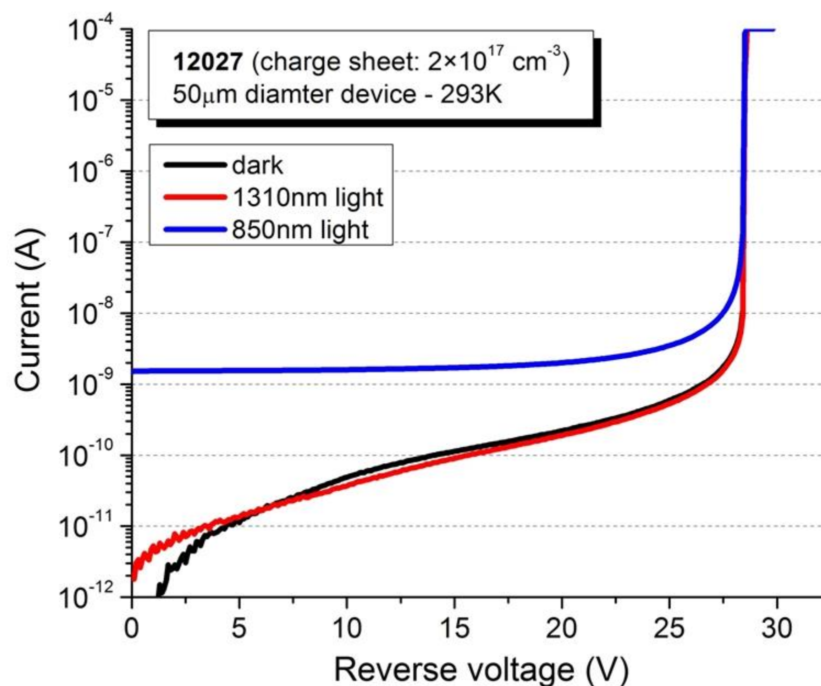


(b)

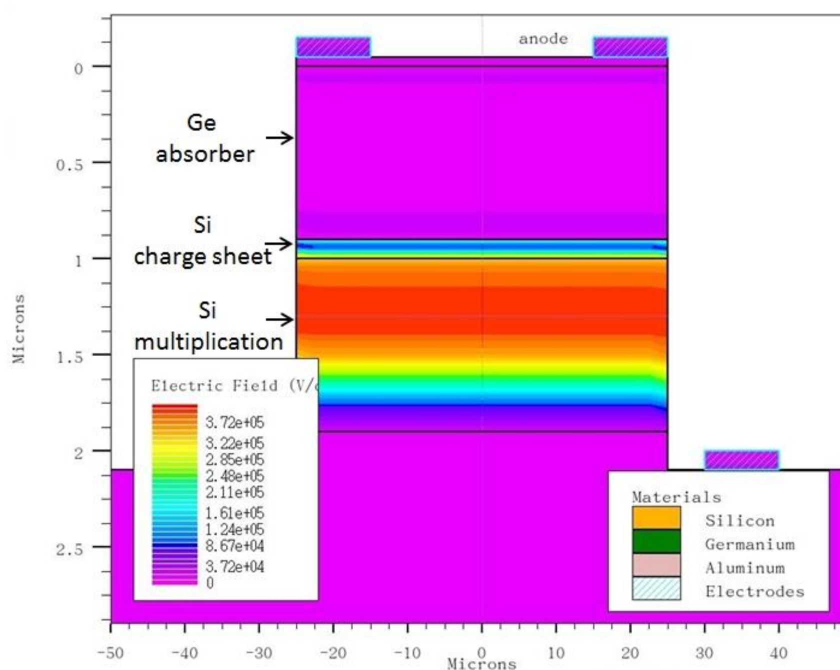
Figure 4.18. Comparison between experimental (black) and simulated (red) I-V characteristics for (a) sample 12-027 (Ge-on-Si SPAD) and (b) 12-028 (whole Si APD). Dopant concentration determined with the SIMS measurements, shown in Fig. 4.14 and 4.15, were integrated in the model.

Due to the high concentration in the charge sheet layer, the structure 12-027 did not show any punch-through, since the electric field was confined in the Si multiplication layer. To further confirm this hypothesis, devices were in turn illuminated by 850 nm and 1310 nm wavelength sources. These devices did not show any photoresponse at 1310 nm because the Ge absorber layer was not depleted. However, photocurrent was measured at 850 nm because of the light absorbed in the Si multiplication layer, as

shown in Figure 4.19a. This was also confirmed by the modelled 2D electric field profile at 95% of V_{BD}, for which all the electric field is confined in the Si multiplication layer (Figure 4.19b). The simulated structure did match well with the measured I-V characteristic, showing a difference of ~2 V in the breakdown voltage. Similar results were obtained for the structure 12-028.



(a)



(b)

Figure 4.19. (a) Measured photocurrent from sample 12-027 using 1310 nm (red) and 850 nm (blue) wavelength illumination. (b) Modelled 2D electric field profile at 95% of V_{BD} for structure 12-027.

Although these two structures (12-027 and 12-028) did not work as originally designed because of the high doping concentration in the Si charge layer, good feedback in terms of the doping concentration of the Si multiplication layer was provided. Since wafers 11-167 and 11-141 worked as single-photon detectors (as shown in the next Chapter) we believe that the doping concentration of the Si charge layer was close to the designed values. However, we expected that the phosphorous diffusion tail in the Si multiplication layer was similar to that of wafer 12-027 and hence modelling was performed to evaluate its effect on the 1st generation of designed structures. The integration of the P doping profile for the Si multiplication layer, without changing any parameters, for the structure shown in Figure 4.3, was modelled with Silvaco. As expected, different values for both V_{PT} and V_{BD} voltages were obtained (Table IV).

Charge sheet doping concentration (cm^{-3})	**V_{PT} (V) - after SIMS**	**V_{PT} (V) – as designed**	**V_{BD} (V) - after SIMS**	**V_{BD} (V) – as designed**
1×10^{17}	- 10	- 18	- 42.5	- 53
1.5×10^{17}	- 12.5	- 27	- 37	- 46.5
2×10^{17}	- 18	- 35	- -31	- 42

Table IV. Summary of the simulated V_{PT} and V_{BD} voltages obtained by using the measured P SIMS profile and the designed P doping concentration for the Si multiplication region in the structure shown in Fig. 4.3.

The phosphorous tail in the multiplication layer had the effect of reducing the thickness of the multiplication layer and hence, depending on the doping concentration of the charge sheet layer the breakdown voltage was shifted to lower voltages (Table IV) than breakdown voltages obtained for an uniform doping concentration of the Si multiplication layer. As a consequence, the punch-through voltage was also shifted to lower voltages.

In addition, the electric field through the centre of the structure was evaluated by integrating the phosphorous doping tail in the model, and then compared with the electric field obtained for the structure with an uniform doping concentration in the Si multiplication layer. This is illustrated in Figure 4.20. For both simulations, the doping

concentration of the charge sheet layer was kept at the same value and equal to 1.5×10^{17} cm^{-3}.

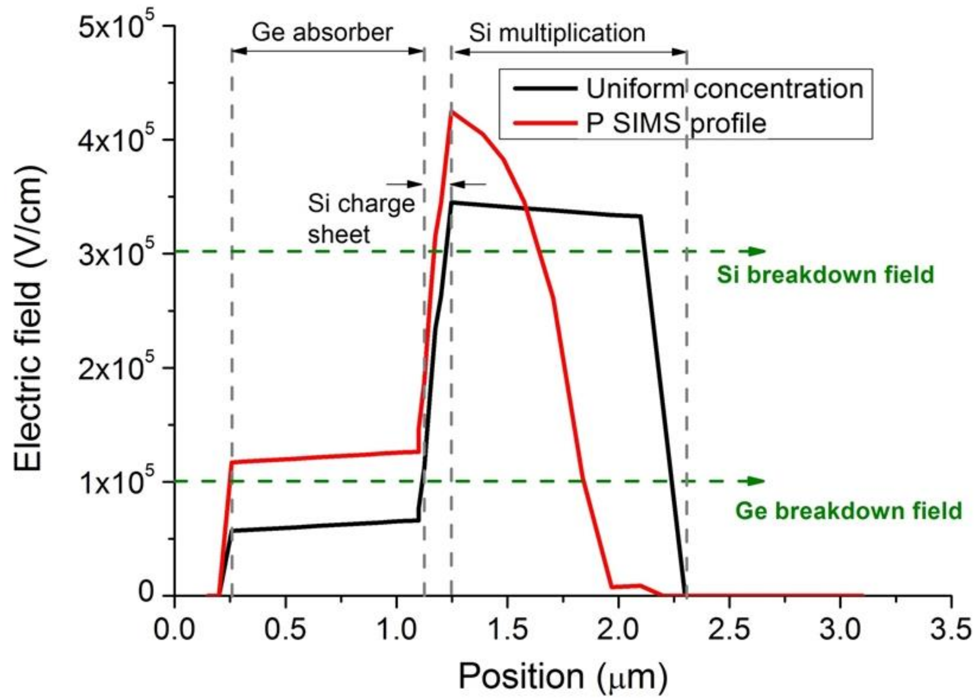


Figure 4.20. Comparison of the simulated electric field profile through section x=0 μm for structure shown in Fig. 4.3. The black line shows the electric field when a uniform doping concentration in the 1 μm thick Si multiplication layer, equal to 1×10^{15} cm^{-3} is used. The red line shows the electric field obtained when the phosphorous doping profile is added to the model. The charge sheet doping concentration was 1.5×10^{17} cm^{-3}.

The simulated electric field for the structure integrating the phosphorous doping profile (red line) highlighted two main problems:

1. For the same doping concentration of the charge sheet layer, the electric field in the Ge absorber layer was higher than the Ge breakdown field, and hence the dark current was expected to be higher than the value obtained for the same structure with an uniform doping;
2. There was a high electric field at the interface between the charge layer and the multiplication layer which was partially undepleted. As a consequence, the electric field across the multiplication layer was not uniform. This result suggested that the avalanche triggering probability might be smaller than the one for the same structure with a uniform doping concentration.

4.10 3rd Generation of Ge-on-Si SPAD structure: SIMS measurements and modelling

To further improve the doping concentration in the Si multiplication layer, a third generation of Ge-on-Si SPAD structures were grown. These structures were identical to the 1st generation of SPAD structures, as shown in Fig. 4.3, with the exception that the thickness of the Si multiplication layer was increased from 1 μm to 1.5 μm. This thickness was chosen in order to counteract the effect of the expected doping diffusion tail in the design and have a bigger portion of the Si multiplication layer with low background doping and uniform electric field. In addition, the highly doped substrate was also changed and arsenic was used instead of phosphorous. The reason for this change was because arsenic has a lower diffusivity than phosphorous in silicon from available scientific literature [14] and hence might allow a more uniform electric field profile in the multiplication region. Five Ge-on-Si SPAD structures were grown with a variation of the charge sheet layer from 5×10^{16} cm^{-3} to 5×10^{17} cm^{-3}. This distribution of values of nominal charge sheet layer doping density was chosen to take into account possible calibration errors of the doping concentration of this layer during the heteroepitaxial growth.

Simulations were performed by using the new design parameters and including the measured phosphorous diffusion tail from the previous SIMS measurements. While an arsenic doped substrate was used during the heteroepitaxial growth, these simulations were performed to qualitatively ascertain the value of punch-through and breakdown voltages. Figure 4.21a-b illustrates the simulated I-V characteristics, while Table V summarises the value of V_{PT} and V_{BD} voltages.

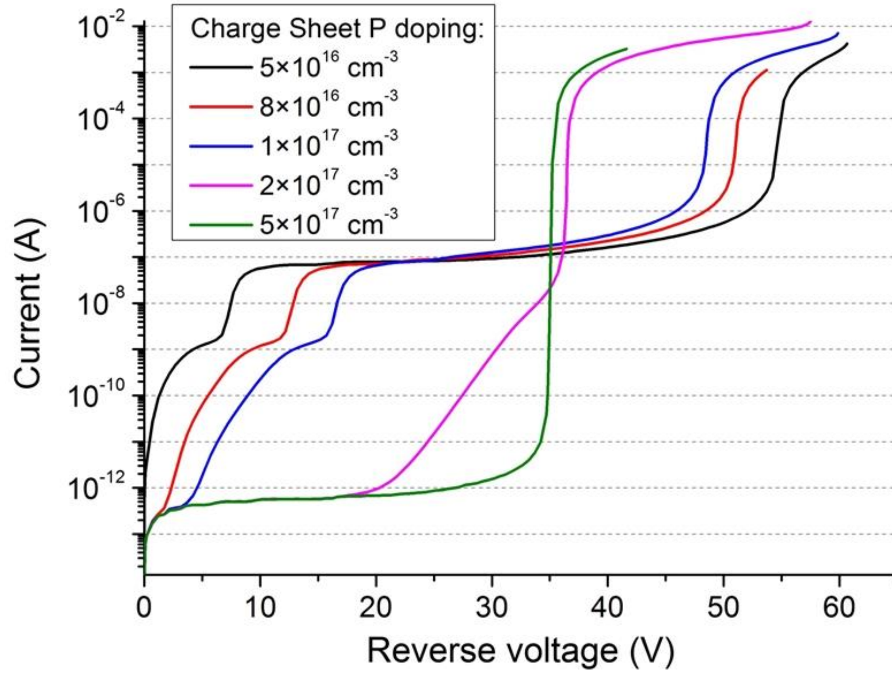


Figure 4.21. Simulated I-V characteristics for the 3rd Generation of Ge-on-Si SPAD structures. These structures were similar to the structures shown in Figure 4.3, but the thickness of the Si multiplication layer was increased from 1 μm to 1.5 μm. The structures were simulated for different doping concentrations of the charge sheet layer: 5 × 10^16 cm^-3 (black), 8 × 10^16 cm^-3 (red), 1 × 10^17 cm^-3 (blue), 2 × 10^17 cm^-3 (magenta), and 5 × 10^17 cm^-3 (green).

Charge sheet doping concentration (cm^{-3})	**V_{PT} (V)**	**V_{BD} (V)**
5×10^{16}	- 9.5	- 54
8×10^{16}	- 15.4	- 50.9
1×10^{17}	- 19.3	- 48.37
2×10^{17}	Partially depleted	- 36.5
5×10^{17}	Undepleted	- 35.2

Table V. Summary of the V_{PT} and V_{BD} voltages for the simulated 3rd Generation of Ge-on-Si structures which I-V characteristics are reported in Figure 4.21.

Results showed that the Ge absorber layer was fully depleted when a doping concentration of the charge sheet layer in the range from 5×10^{16} cm^{-3} to 1×10^{17} cm^{-3}

was used (black, red and blue lines of Figure 4.21) (Table V). The last two doping concentrations showed either incomplete or no depletion of the absorber layer with most of the electric field confined within the Si multiplication layer, and hence lower breakdown voltages (magenta and green line) than the previous structures were observed.

Five wafers were grown by our collaborator in Warwick, and the wafer IDs are reported in Table VI.

Charge sheet doping concentration (cm^{-3})	**Wafer ID**
$\mathbf{5 \times 10^{17}}$	13-312
$\mathbf{2 \times 10^{17}}$	13-313
$\mathbf{1 \times 10^{17}}$	13-314
$\mathbf{8 \times 10^{16}}$	13-315
$\mathbf{5 \times 10^{16}}$	13-316

Table VI. Wafer ID related to each doping concentration of the charge sheet layer, as grown by our collaborator at University of Warwick.

From these wafers, mesa geometry devices were fabricated by our collaborators at the University of Glasgow. Figure 4.22 illustrates the experimental I-V characteristics measured for devices of each wafer.

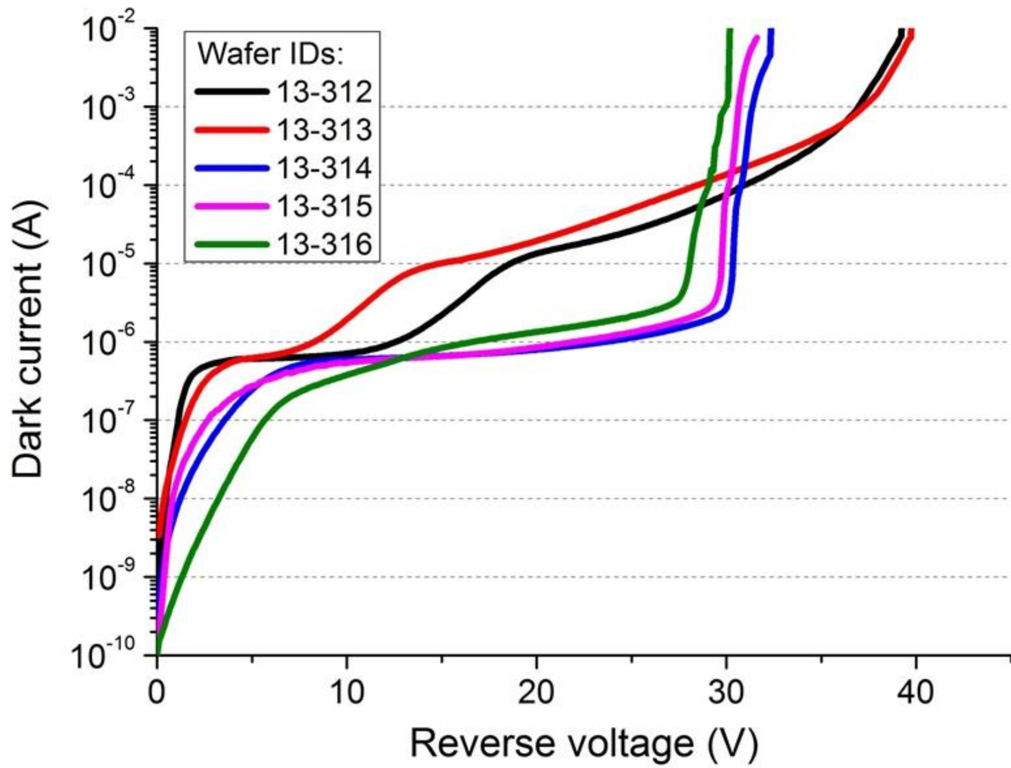
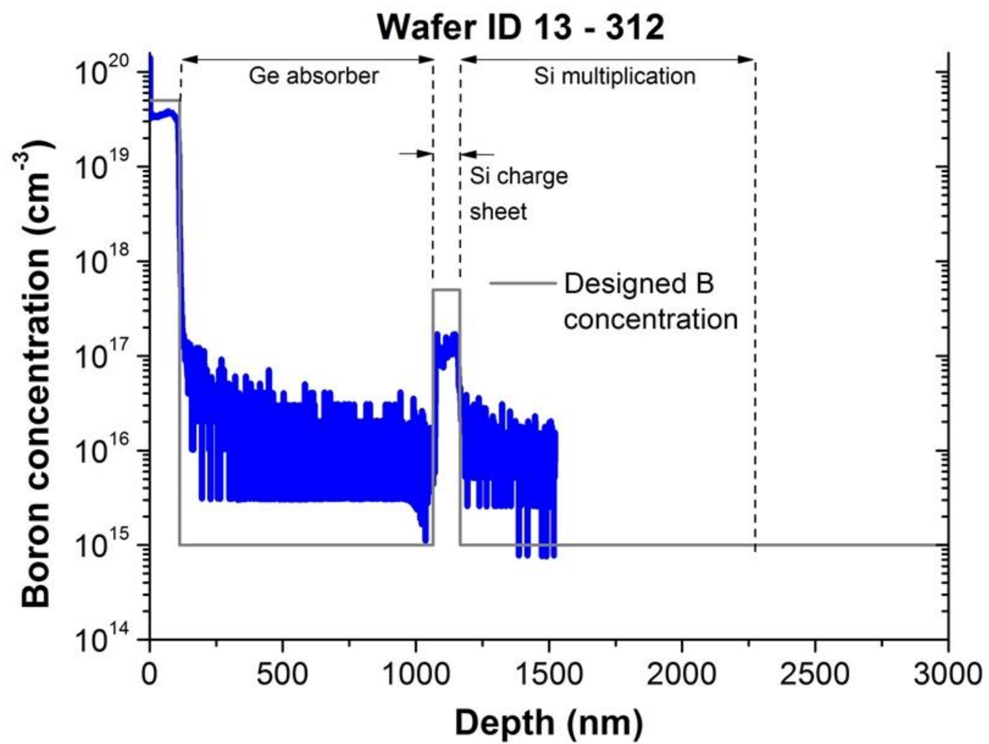
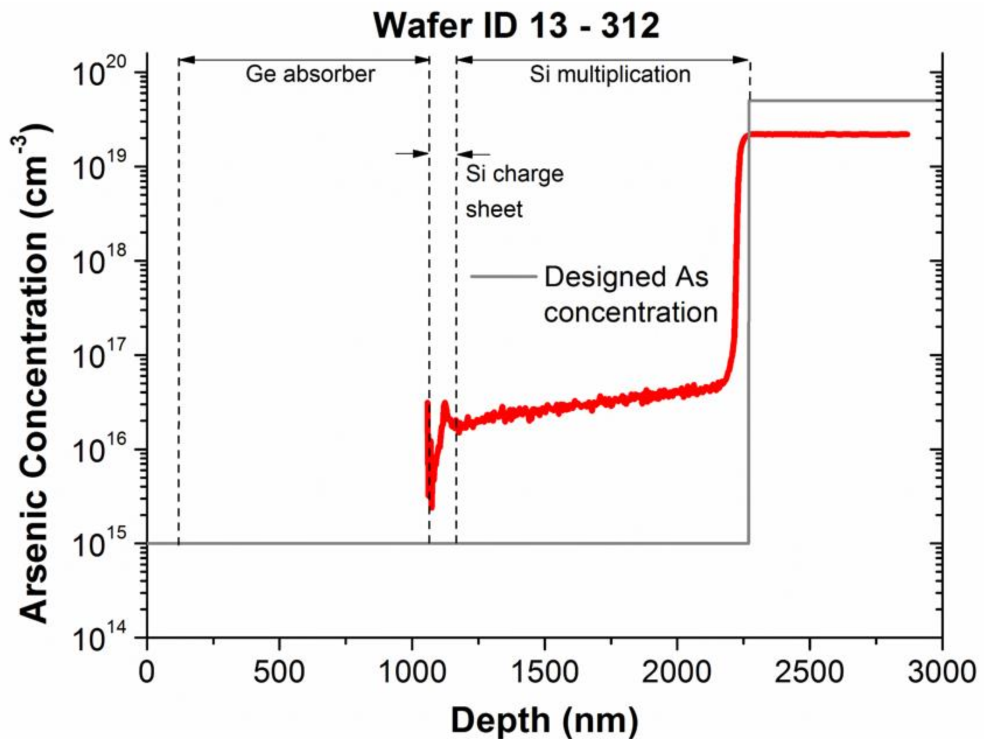


Figure 4.22. Experimental I-V characteristics for devices fabricated from the wafer reported in table V. Different colours refer to different doping concentrations of the Si charge layer: 5 × 10^{17} cm^{-3} (black), 2 × 10^{17} cm^{-3} (red), 1 × 10^{17} cm^{-3} (blue), 8 × 10^{16} cm^{-3} (magenta), and 8 × 10^{16} cm^{-3} (green).

Devices from wafer 13-312 and 13-313 showed a punch-through voltage at approximately -10 and -15V with a breakdown voltage of ~ -42V, respectively, as shown in Fig. 4.21. The remaining wafers show a low level of dark current without any punch-through voltage suggesting that the Ge absorber layer was not depleted and, consequently, the doping value either of the Si charge region or the Ge absorber was much higher than the designed value. Therefore, SIMS measurements were performed on wafers 13-312 and 13-316, and are illustrated in Figure 4.23 and 4.24, to measure the doping concentration of each structure layer.

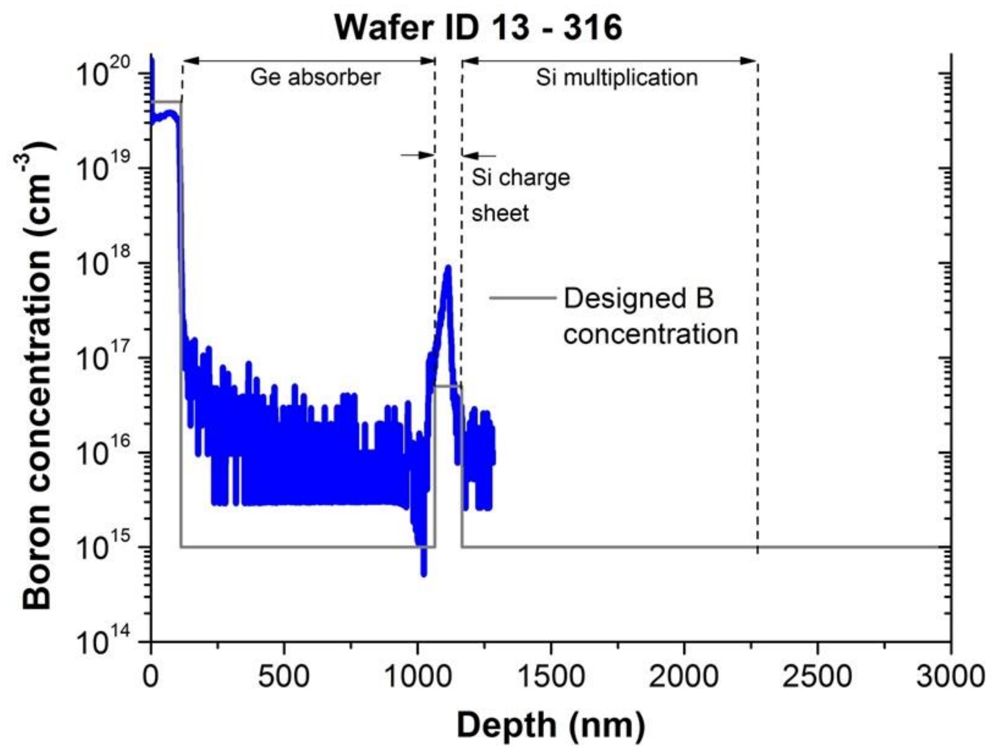


(a)

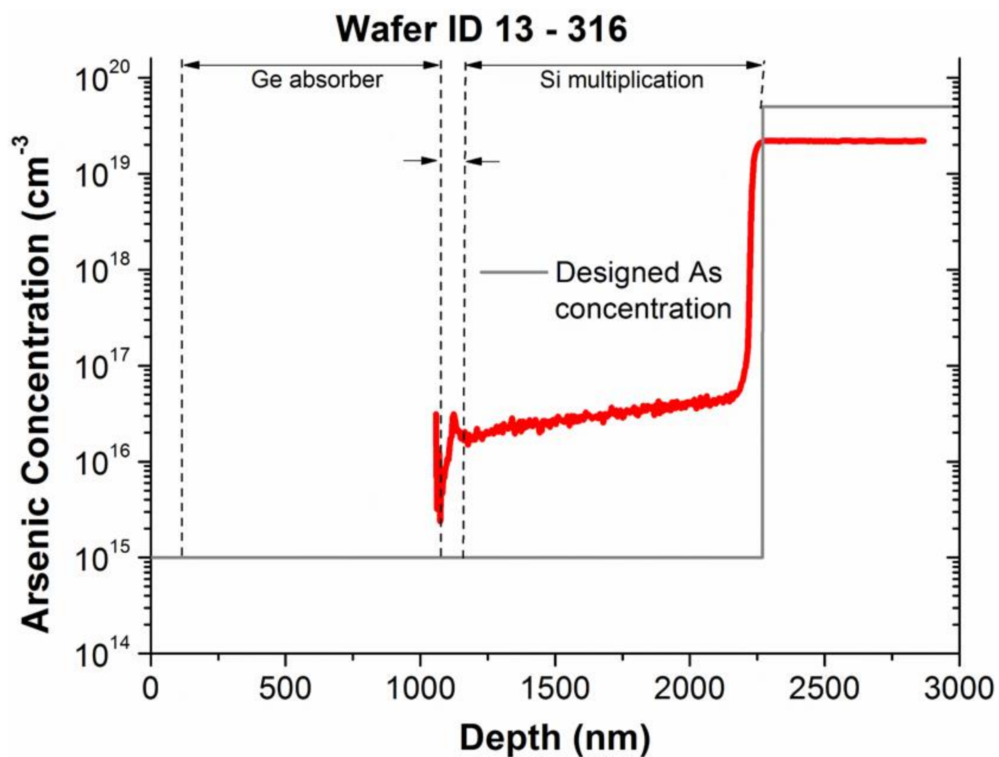


(b)

Figure 4.23. SIMS measurements performed on wafer labelled 13-312 (Ge-on-Si SPAD) which shows (a) the Boron (blue) and (b) Arsenic (red) doping profile as a function of the depth across the structure. Grey lines in (a) and (b) show the designed B and As doping concentrations for the structure, respectively.



(a)



(b)

Figure 4.24. SIMS measurements performed on wafer labelled 13-316 (Ge-on-Si SPAD) which shows (a) the Boron (blue) and (b) Arsenic (red) doping profile as a function of the depth across the structure. Grey lines in (a) and (b) show the designed B and As doping concentrations for the structure, respectively.

As obtained for the samples analysed previously, the doping level of the Ge absorber layer was quite low and uniform through the whole layer. However, for both wafers, the boron doping concentration (blue) in the Si charge sheet layer did not correspond to the designed value (grey lines in Figure 4.23 and 4.24). In particular, for wafer 13-312, the measured B doping level in the charge sheet was estimated as ~1.5×10^{17} cm^{-3}, lower than the design value of 5×10^{17} cm^{-3}. Although the estimated value was not as designed, its value was in the range of the designed doping concentrations for the Si charge sheet, and led to a working device, with a depleted absorber layer, as shown in Figure 4.20 (black line).

For wafer 13-316, the measured B concentration was much higher and estimated as being approximately ~1×10^{18} cm^{-3} , very different to the design value of 5×10^{16} cm^{-3}. Because of this high doping level, the electric field drops across the Si multiplication layer without depleting the Ge absorber layer, as shown in Figure 4.20 (green line), resulting in a low level of dark current and a low breakdown voltage. Both SIMS measurements confirmed the difficulty to accurately control the doping level in the Si charge sheet layer during the hetero-epitaxial growth.

In addition, for both analysed wafers, SIMS revealed that the arsenic level present due to diffusion (red) in the Si multiplication region was ~2 orders of magnitude higher than the designed value (1×10^{15} cm^{-3}). This doping level was also higher than the phosphorous level measured in wafers 12-027 and 12-028. This is illustrated in Figure 4.25 which shows a comparison between the As and P concentrations in the Si multiplication layer.

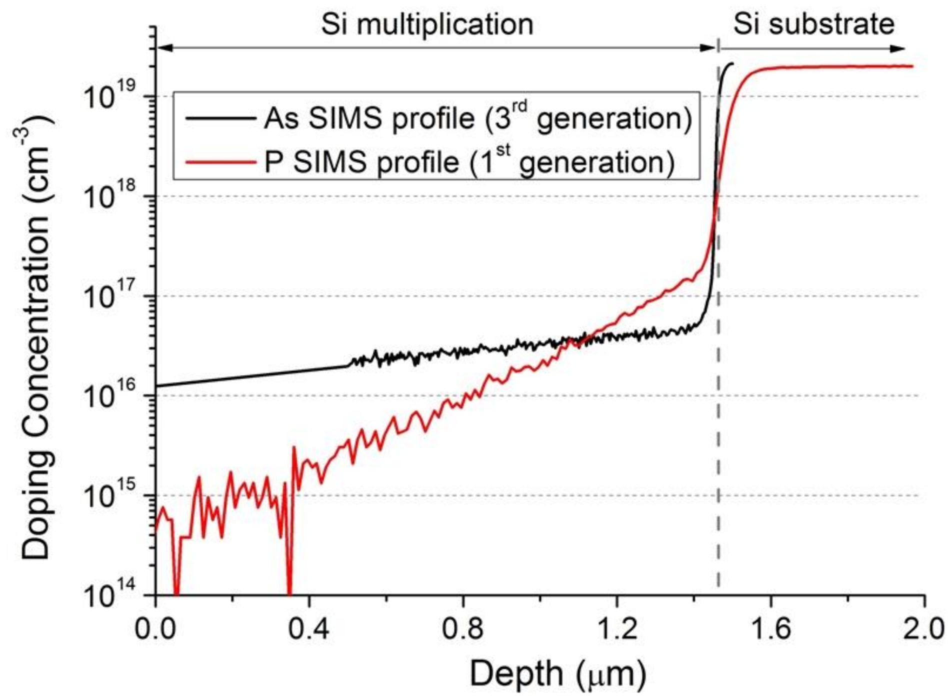


Figure 4.25. Comparison of the doping profile between the P profile of the 1st Generation (red) and the As profile of the 3rd Generation (black) of Ge-on-Si structures which were obtained by using SIMS measurements.

The impact of the arsenic diffusion tail was analysed by integrating the SIMS measurements into the modelled structure. A boron doping concentration of 1.5×10^{17} cm^{-3} for the charge sheet layer was considered to match the boron level measured in the charge sheet layer with SIMS for wafer 13-312. Figure 4.26 shows the simulated I-V characteristic.

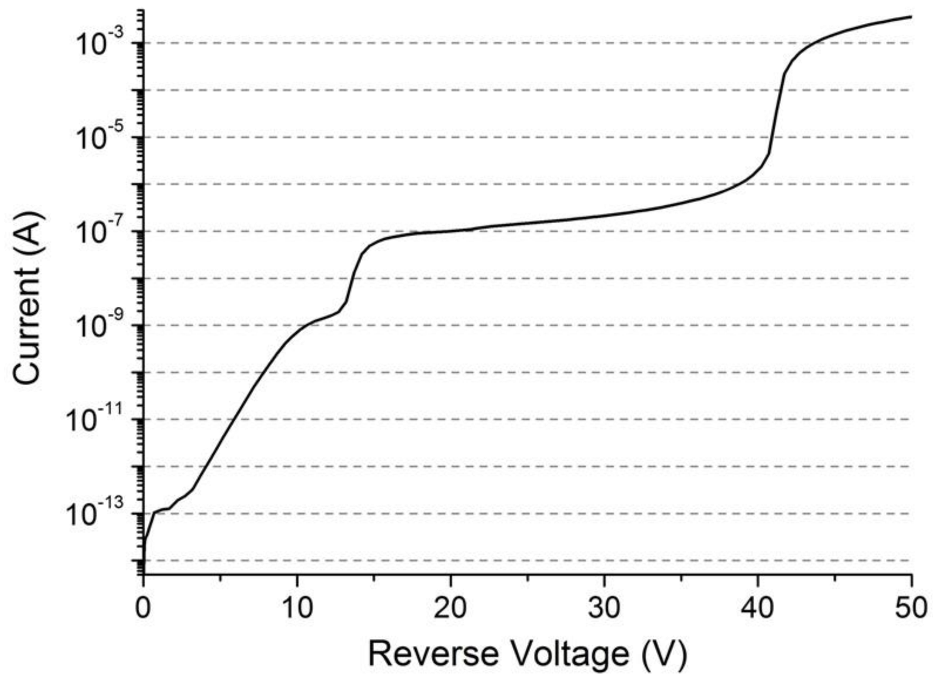


Figure 4.26. Simulated I-V characteristic for the 3rd Generation of Ge-on-Si structure obtained by integrating the SIMS measured As doping profile (Figure 4.23 and 4.24) in the Si multiplication layer and for a doping concentration of the charge sheet layer equal to 1.5×10^{17} cm^{-3}.

The simulated V_{PT} and V_{BD} were found to be equal to ~ -15 V and ~ -42 V, respectively and are comparable with the experimental value shown in Figure 4.22 for wafers 13-312 and 13-313. To further understand the effect of the As diffusion tail in the multiplication region, the 1D electric field profile at 95% of V_{BD} was extracted from the simulation and compared with the simulated 1D electric field profile at 95% of V_{BD} obtained from an identical structure integrating the P diffusion tail in the multiplication layer. This result is shown in Figure 4.27.

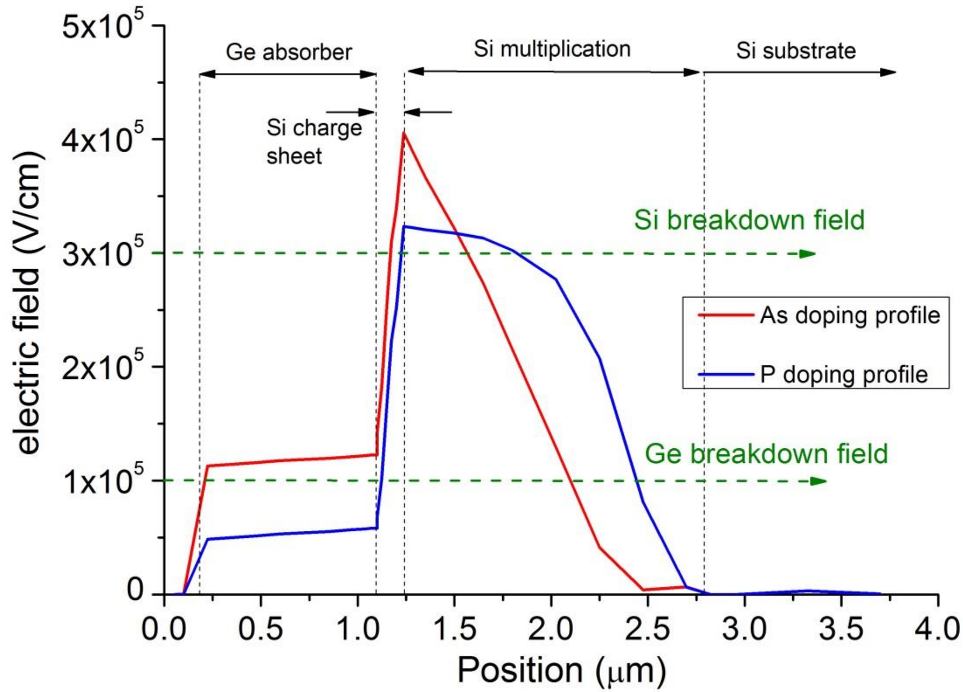


Figure 4.27. Comparison of the simulated electric field profile through section x=0 μm of structure in Fig. 4.3, for the 3rd Generation of Ge-on-Si structures. The red line shows the electric field obtained when the As doping profile measured by SIMS (Figure 4.23 and 4.24) in the Si multiplication layer was used. Blue line shows the electric field obtained when the P doping profile (Figure 4.17) was added to the model. Both structures were simulated for the same doping concentration of the charge sheet layer equal to 1.5×10^{17} cm^{-3}.

The simulated electric field obtained by modelling the measured SIMS results for the arsenic diffusion tail (red line) in the Si multiplication led to a different set of problems:

1. A triangular shape of the electric field near the charge sheet layer. This uneven electric field distribution means that avalanche breakdown is more confined to a small part of the multiplication layer leading to a small portion of the Si multiplication layer where carriers have a high probability of creating an avalanche of carriers. As a consequence, a lower photon detection efficiency was expected. In addition, the observed spike of the electric field at the interface between the Si charge layer and multiplication could enhance some microplasma effect in a small portion of the Si multiplication layer. This problem negatively affected the reliability and yield of the fabricated devices from these wafers which showed a short circuit behaviour in the current voltage characteristic after repeating the I-V characteristic more than one time.

2. The Si multiplication layer was partially undepleted. This problem might lead to a much slower device, since the carriers may reach the back contact by diffusion and not drift, affecting the SPAD performance (e.g. leading to a much longer timing jitter).

The simulated electric field obtained by using the phosphorous doping concentration (blue line) in the modelling software, for a thickness of the Si multiplication layer equal to 1.5 μm, showed a better uniformity of the electric field in the Si multiplication layer with a smaller portion of this layer undepleted. Furthermore, both structures were simulated for the same doping concentration of the charge sheet layer equal to 1.5×10^{17} cm^{-3}, but the electric field in the Ge absorber layer was higher for the structure integrating the As doping (red line), due to the resulting smaller multiplication layer thickness than the structure integrating the P doping profile.

4.11 Future development

Both the doping concentrations of the charge sheet and multiplication layer have been found to be critical aspects of the structure design. Regarding the charge sheet region, it was very difficult to accurately control its doping concentration due to different problems in the heteroepitaxial growth apparatus. SIMS measurements showed a much higher doping concentration than the designed value, and this in turn led to a device which cannot be used to detect infrared light, as demonstrated by experimental measurements and simulations. Different solutions have been proposed to overcome this problem:

1. To balance the high doping concentration of this layer, its thickness should be reduced to have the same integrated charge as an identical layer with lower doping and greater thickness. However, only small adjustments can be made by reducing the thickness of the charge layer. It is not possible to use a doping concentration of the order of 10^{18} cm^{-3} because the thickness of the charge sheet layer will be reduced to less than 10 nm. This might cause further problems such as tunnelling through this thin layer which could also be changed to a conductive layer due to the high doping concentration. This might result in a structure unable to detect infrared light.
2. It could be possible to introduce an intrinsic spacer layer between the Si charge sheet and the Ge absorber layer to perform a step graded dopant diffusion to obtain the designed doping concentration of $\sim 1.5 \times 10^{17}$ cm^{-3}. However, this

should be carefully calibrated during the growth and might add further complexity to the structure.

3. To better control the doping concentration of the charge sheet layer, ion implantation could be used. However, this process could possibly introduce further defects at the hetero-interface with the Ge absorber layer and hence this might enhance the threading dislocation density.

Regarding the multiplication layer, several efforts have been performed by the grower to reduce the doping diffusion tail. In particular, using an intrinsic silicon substrate instead of a highly doped substrate for future generations of Ge-on-Si SPAD device was suggested. Then, a highly doped Si layer, with a well-established thickness, was epitaxially grown on top of the intrinsic substrate followed by a 1.5 μm-thick intrinsic multiplication layer. The wafer ID, for this test sample was 14-253. This approach demonstrated a reduction of the Si diffusion tail into the intrinsic Si layer (multiplication region), as shown by a SIMS measurement performed on a test sample (Figure 4.28).

Another solutions to this problem was proposed by IQE (a company based in UK performing epitaxial growth) who suggested to use a highly antimony (Sb) doped (~10^{18} cm^{-3}) Si substrate. A test sample was also made by growing a 1 μm-thick intrinsic multiplication layer on top of the substrate. The wafer ID, for this test sample, was 22-003. SIMS measurements for both wafers are also shown in Figure 4.28.

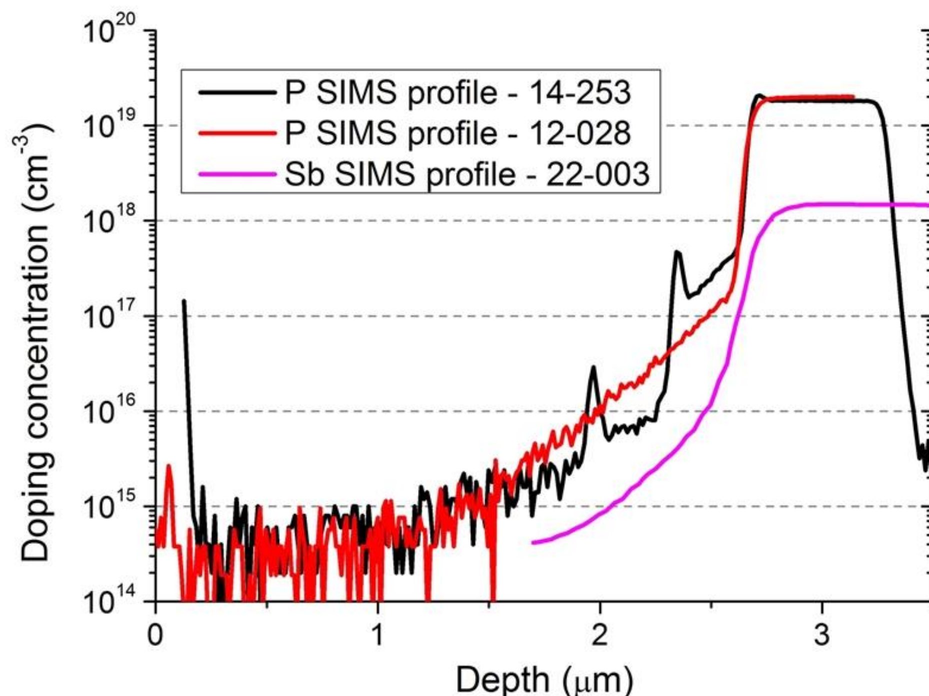


Figure 4.28. Comparison of the SIMS measured doping profile for structure 14-253 (black), 12-028 (red), and 22-003 (magenta).

Results from wafers 14-253 (black) and 22-003 (magenta), shown in Figure 4.28, were also compared with the P doping profile from wafer 12-028 (red), as illustrated in the previous sections. The doping profile from wafer 14-253 showed a slight improvement compared to the previous result obtained from wafer 12-028. The Sb doping profile showed a very small diffusion tail, and the designed level for the intrinsic region (~10^{15} cm^{-3}) was obtained after ~300 nm from the substrate (Figure 4.28).

Electric field profile simulations, at 95% of V_{BD}, were carried out with the modelling software to better evaluate the impact of these three different doping concentrations on the designed Ge-on-Si structure. The charge sheet doping concentration was kept fixed at 1.5×10^{17} cm^{-3} for all simulations. This is illustrated in Figure 4.29 where the colour code used is the same as in Figure 4.28 (black: P profile 14-253; red: P profile 12-028; magenta: Sb profile 22-003).

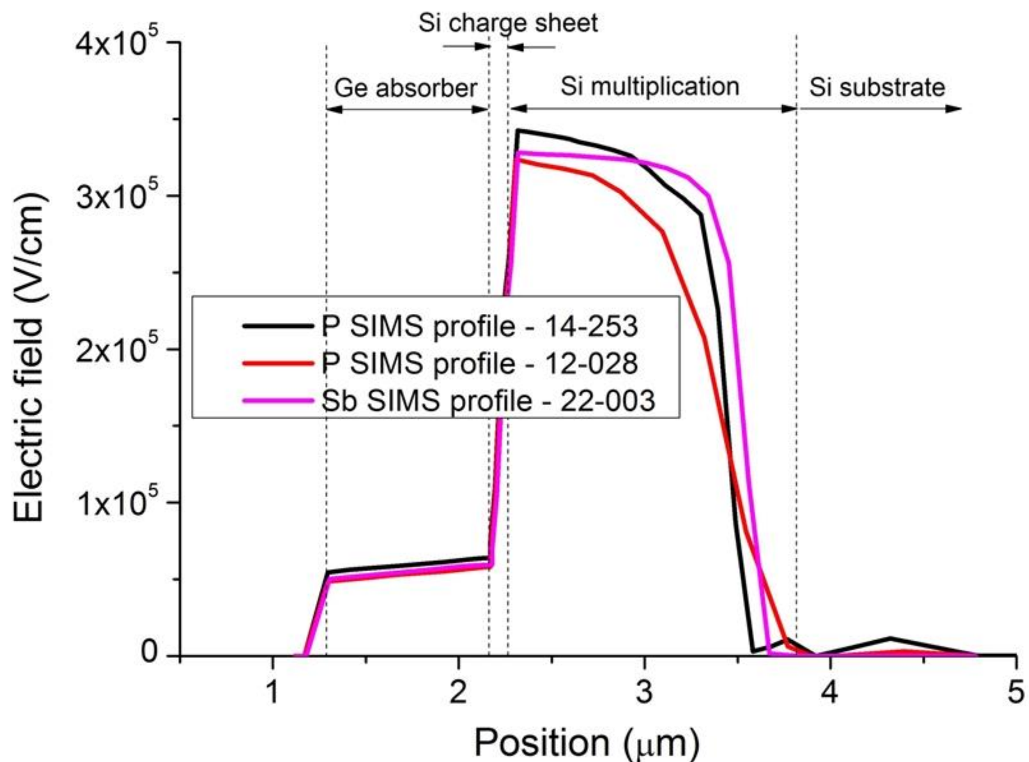


Figure 4.29. Simulated electric field profile for a 3rd Generation Ge-on-Si structure. In particular, the SIMS measured doping profile reported in Figure 4.28 were used for Si multiplication layer: P profile 14-253 (black), P profile 12-028 (red), and Sb profile 22-003 (magenta). The doping concentration of the charge sheet layer was kept fixed at 1.5×10^{17} cm^{-3} for all the simulated structures.

It is clear from Figure 4.29 that the Sb doping profile (magenta line) produced the most uniform electric field profile across the multiplication layer. This improvement was more evident when this profile is compared with the simulated electric field profile

obtained by using the P doping profile of wafer 12-028 (red). Regarding to the electric field profile obtained by using the P doping profile of wafer 14-253 (black), this could be considered as acceptable for the designed structure.

Finally, the choice of the right doping profile for the Si multiplication layer must consider the reproducibility of the doping profile obtainable during the epitaxial growth.

4.12 Conclusion

In this chapter, the main advantages of a SACM structure have been described. The Ge-on-Si SPAD structure proposed in this work is similar to the structure of an InGaAs/InP SPAD, but the material system is completely different, thereby different design rules have been introduced.

The early design of the Ge-on-Si SPAD was based on three main aspects: electric field, doping concentration and temperature. Simulations have been performed to introduce the concepts of punch-through and breakdown voltage. In addition, the effects of the charge sheet doping concentration on the simulated structure were considered to evaluate the simulated electric field profile at 95% of the breakdown voltage. It has been underlined how is important to tailor the electric field between the Ge absorber layer and the Si multiplication layer to guarantee the detection of the infrared light with a high probability of triggering an avalanche of carriers.

Different SPAD generations have been analysed. Experimental data obtained from capacitance-voltage and SIMS measurements have been used in the modelling software to evaluate their impact on the designed structure. Therefore, comparisons between experimental and simulated data were evaluated. Although different problems with the epitaxial growth apparatus were experienced, their impact on the designed structure was helpful to better understand the behaviour on the structure itself. In particular, the high boron doping concentration of the charge sheet layer resulted in a device which could not be used to detect infrared light. However, the designed structure has been adjusted to take into account the doping diffusion tail from the Si substrate in the intrinsic Si multiplication layer. Furthermore, different dopants (P, As, and Sb) were studied and then introduced in the model to evaluate their impact on the electric field profile. Results of the measured Sb doping profile were found to show the most uniform electric field profile across the multiplication region, and this result suggested that this design

solution might be used in future research the growth of the next Ge-on-Si SPAD generations.

4.13 Reference

[1] F. Acerbi, A. Tosi, and F. Zappa, "Growths and diffusions for InGaAs/InP single-photon avalanche diodes," *Sens. Actuators Phys.*, vol. 201, pp. 207–213, Oct. 2013.

[2] M. A. Itzler, Ben-Michael, C.-F. Hsu, K. Slomkowski, A. Tosi, S. Cova, F. Zappa, and R. Ispasoiu, "Single photon avalanche diodes (SPADs) for 1.5 μm photon counting applications," *J. Mod. Opt.*, vol. 54, no. 2–3, pp. 283–304, 2007.

[3] F. Acerbi, M. Anti, A. Tosi, and F. Zappa, "Design Criteria for InGaAs/InP Single-Photon Avalanche Diode," *IEEE Photonics J.*, vol. 5, no. 2, pp. 6800209–6800209, 2013.

[4] S. M. Sze and K. K. Ng, *Physics of semiconductor devices*. John Wiley & Sons, 2006.

[5] M. A. Itzler, M. Entwistle, M. Owens, K. Patel, X. Jiang, K. Slomkowski, S. Rangwala, P. F. Zalud, T. Senko, J. Tower, and J. Ferraro, "Design and performance of single photon APD focal plane arrays for 3-D LADAR imaging," 2010, vol. 7780, p. 77801M–77801M–15.

[6] Y. Liu, S. R. Forrest, J. Hladky, M. J. Lange, G. H. Olsen, and D. E. Ackley, "A planar InP/InGaAs avalanche photodiode with floating guard ring and double diffused junction," *J. Light. Technol.*, vol. 10, no. 2, pp. 182–193, 1992.

[7] A. Lacaita, F. Zappa, S. Cova, and P. Lovati, "Single-photon detection beyond 1 mm: performance of commercially available InGaAs/lnP detectors," *Appl. Opt.*, vol. 35, no. 16, pp. 2986–2996, 1996.

[8] Y. Kang, H.-D. Liu, M. Morse, M. J. Paniccia, M. Zadka, S. Litski, G. Sarid, A. Pauchard, Y.-H. Kuo, H.-W. Chen, W. Sfar Zaoui, J. E. Bowers, A. Beling, D. C. McIntosh, X. Zheng, and J. C. Campbell, "Epitaxially-grown germanium/silicon avalanche photodiodes for near infrared light detection," 2009, pp. 733906–733906–8.

[9] Y. Kang, H.-D. Liu, M. Morse, M. J. Paniccia, M. Zadka, S. Litski, G. Sarid, A. Pauchard, Y.-H. Kuo, H.-W. Chen, W. S. Zaoui, J. E. Bowers, A. Beling, D. C. McIntosh, X. Zheng, and J. C. Campbell, "Monolithic germanium/silicon avalanche photodiodes with 340 GHz gain-bandwidth product," *Nat Photon*, vol. 3, no. 1, pp. 59–63, 2009.

[10] A. U. Manual, "Silvaco," *St. Clara CA*, 2008.

[11] W. Shockley and W. T. Read, "Statistics of the Recombinations of Holes and Electrons," *Phys. Rev.*, vol. 87, no. 5, pp. 835–842, 1952.

[12] L. Colace, G. Masini, and G. Assanto, "Ge-on-Si approaches to the detection of near-infrared light," *IEEE J. Quantum Electron.*, vol. 35, no. 12, pp. 1843 –1852, 1999.

[13] S. Selberherr, *Analysis and simulation of semiconductor devices*. Springer Wien; New York, 1984.

[14] A. U. Manual, "Silvaco DevEit," *St. Clara CA*, 2008.

[15] Y. Kang, M. Zadka, S. Litski, G. Sarid, M. Morse, M. Paniccia, Y. Kuo, J. Bowers, A. Beling, H. Liu, and others, "Epitaxially-grown Ge/Si avalanche photodiodes for 1.3mm light detection," *Opt Express*, vol. 16, pp. 9365–9371, 2008.

[16] S. Brotzmann and H. Bracht, "Intrinsic and extrinsic diffusion of phosphorus, arsenic, and antimony in germanium," *J. Appl. Phys.*, vol. 103, no. 3, p. 033508, 2008.
[17] A. Chroneos, R. W. Grimes, B. P. Uberuaga, S. Brotzmann, and H. Bracht, "Vacancy-arsenic clusters in germanium," *Appl. Phys. Lett.*, vol. 91, no. 19, p. 192106, 2007.
[18] F. F. Y. Wang, *Impurity Doping Processes in Silicon*. Elsevier, 2012.

Chapter 5 - Ge-on-Si Single-Photon Avalanche Diodes: Device fabrication, Characterisation and Performance

5.1 Introduction

In the previous chapter a description of the Ge-on-Si SACM SPAD was presented together with the simulations of different structures. Additionally, results from simulations were integrated with experimental doping concentration data obtained by SIMS, and devices fabricated from different generations of Ge-on-Si wafers were simulated. In this chapter, a brief introduction to the Ge-on-Si heteroepitaxial growth performed by our collaborators at the University of Warwick is described. In particular, the main focus will be given to the 1st generation of SPAD devices (wafer 11-167 and 11-141), since the other generations suffered from different growth and fabrication issues (as described in the previous chapter), and it was not possible to perform any single-photon characterisation. The fabrication of the 1st generation of Ge-on-Si devices was performed independently by collaborators at the McMaster University and the University of Glasgow.

Therefore, two different processing procedures were performed at McMaster University: in the first one, large active area mesa devices (500 and 250 μm diameter) were fabricated without using any passivation layer, while the second processing procedure included small active area mesa devices (from 20 to 100 μm diameter) and SiO_2 was used to passivate the mesa sidewalls. Large active area devices were not characterised as single-photon detectors due to their high leakage current, whilst small active area devices were irreversibly damaged when the temperature was reduced below 200 K. However, data collected from these devices were used to perform a different analysis in terms of dark current, as shown in this chapter.

Mesa devices (ranging from 25 to 500 μm diameter), using a Si_3N_4 passivation layer fabricated at the University of Glasgow, showed a good yield and reproducibility of the current-voltage characteristics, even at lower temperatures. For this reason, their fabrication process is described in the work presented here. Dark current analysis of these devices is described to understand the main leakage mechanisms. Small active area devices (25 and 50 μm diameter) were characterised in terms of their single-photon performance, and a comparison with homojunction Ge APDs, InGaAs/InP SPADs as

well as other Ge-on-Si SPADs reported in the scientific literature, is discussed in this Chapter.

5.2 Ge-on-Si SPAD growth

A full description of the Ge-on-Si SACM structures is given in Chapter 4, and different structures were proposed, designed and fabricated. In particular, 1st Generation of Ge-on-Si SPAD structures were characterised as single-photon avalanche detectors. The typical structure of the device is shown in Figure 5.1.

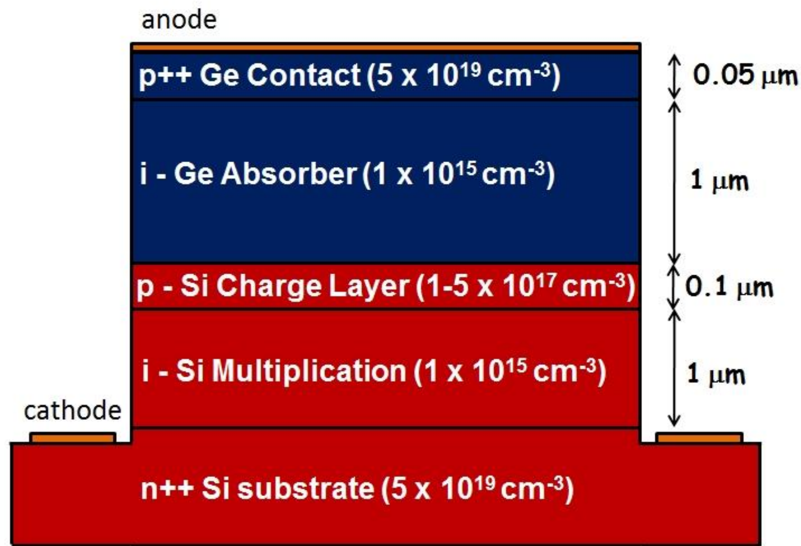


Figure 5.1.Cross-section of the designed mesa geometry Ge-on-Si SPAD structure. For each layer is specified the thickness with the associated doping concentration (N_a or N_d).

Three structures were grown by collaborators at Warwick University, each containing a different dopant concentration of the charge sheet layer as explained in Chapter 4. A summary of the wafer IDs related to each doping concentration of the charge sheet layer for the 1st generation of Ge-on-Si SPAD structure growth is given in Table I.

Charge sheet doping concentration (cm^{-3})	Wafer ID
1×10^{17}	11-167
2×10^{17}	11-141
5×10^{17}	11-142

Table I. Wafer ID related to each doping concentration of the charge sheet layer, as grown.

The structures were grown by RP-CVD starting from a highly P-doped (~5×10^{19} cm^{-3}) Si substrate. 1 μm thickness of intrinsic Si (typically ~10^{15} cm^{-3} or lower) was epitaxially grown in order to form the multiplication region, followed by 100 nm of B-doped (see Table I) Si for the charge sheet region. After this layer, a low temperature (~350-450 °C) Ge seed layer of ~50 nm was grown followed by a high temperature (~650 °C) growth of ~1 μm-thick intrinsic Ge layer (designed to be ~10^{15} cm^{-3}) for the absorber region. The structure was then annealed (800-900 °C, several cycles) to reduce the TDD typically ~$10^6 – 10^7$ cm^{-2} [1]. Finally, 50 nm of a highly B-doped (~2×10^{19} cm^{-3}) Ge layer was epitaxially grown.

5.3 Ge-on-Si SPAD fabrication

The Ge-on-Si SPAD structures were fabricated by collaborators at Glasgow University. The devices were fabricated in a mesa geometry which was required to confine the electric field within the active region of the device. Before the fabrication process, the Ge-on-Si wafers were cleaved by a diamond saw into 1 cm^2 chips. In order to protect the surface of the wafer from contaminants created during cleaving, a protective polymer was applied that could later be dissolved in acetone. The minimum device size was limited by the laser spot that could be focussed in the single-photon characterisation setup, meaning that fabricated devices were greater than 20 μm in diameter. Photolithography was performed to transfer the desired pattern onto the substrate by using a Karl Suss Microtec MA6 with 0.5 μm accuracy. Figure 5.2 illustrates the photo-mask used for the photolithography where each colour represents a different lithography stage. Metal or dry etched markers were incorporated in the photo-mask to achieve alignment (1 μm) between the photolithography stages.

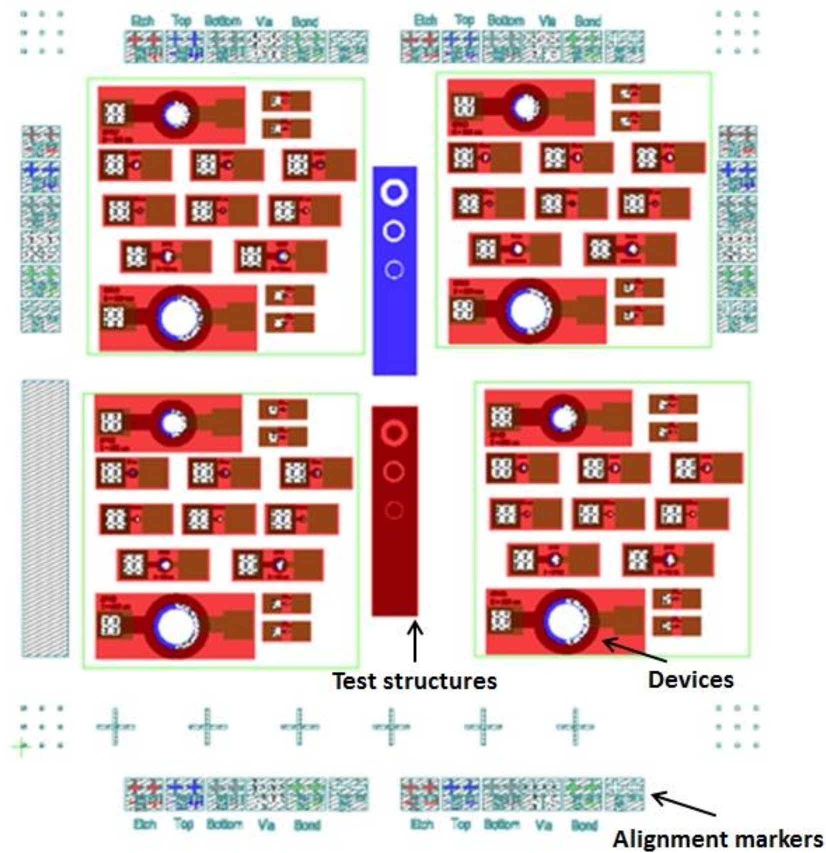


Figure 5.2. Photolithography mask set used for the fabrication of the Ge-on-Si SPADs by photolithography. Each colour represents a different lithography stage.

Therefore, cylindrical mesas, ranging from 25 to 500 µm in diameter, were defined and etched anisotropically (dry etch) down to the highly doped Si substrate, by an inductively coupled plasma reactive ion etching (ICP-RIE) tool using fluorine-based chemistry (SF_6/C_4F_8), resulting in near-vertical sidewalls [2].

Ni was chosen for the contacts (anode and cathode), as it is known to form the lowest electrical resistivity phases for silicides and germanides [3], [4]. A 20 nm layer of Ni was deposited to form the top contact and then annealed at 325 °C. Structures were then planarised and passivated by using a plasma enhanced chemical vapour deposition (PECVD) of Si_3N_4. Via holes were dry etched to allow interconnects to the contacts and then bond pads of 1.2 µm of Al were sputtered. Figure 5.3 illustrated a scanning electron microscope (SEM) image of an Al top bond pad sputtered on a Ge-on-Si SPAD.

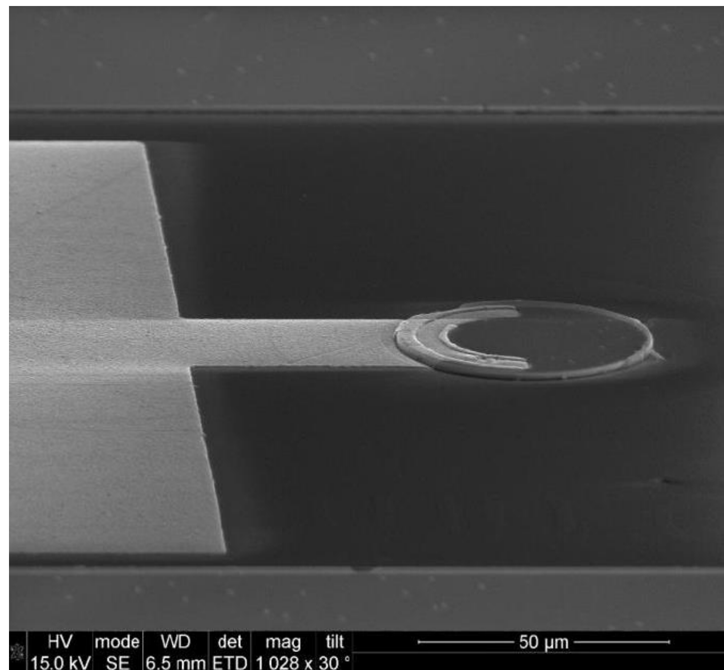
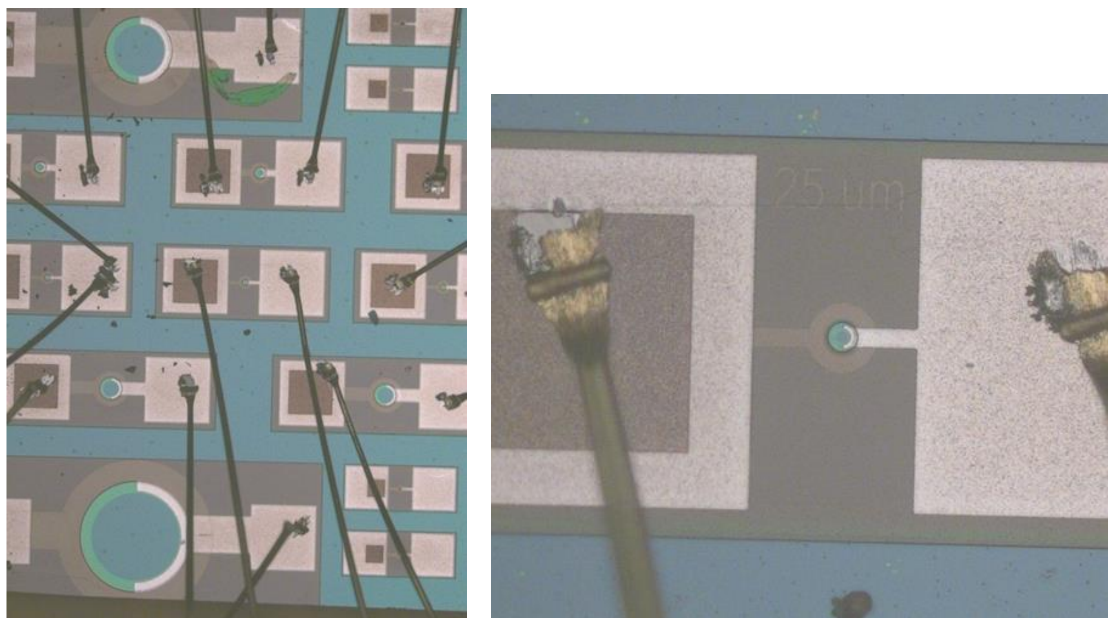


Figure 5.3. A scanning electron microscope image of a sputtered Al top bond pad on the Ge-on-Si SPAD.

Finally, the sample was cleaved into 3 × 3 mm pieces which were firstly mounted on a header package by using conductive silver paint, and then wire bonded with Al. Figure 5.4a-b shows a microscope image of the whole Ge-on-Si chip.



(a) (b)

Figure 5.4. Microscope images of the fabricated (a) Ge-on-Si SPAD chip, and (b) a 25 µm diameter device.

5.4 The Current-Voltage Characteristic

The first step in the characterisation of a device is to measure its current-voltage (I-V) characteristic, which provides information about the breakdown voltage (V_{BD}), the punch-through voltage (V_{PT}) and the dark current (I_d) at 95% of V_{BD}. Measurements were undertaken with a semiconductor parameter analyser (HP 4145B), capable of an accuracy of several picoamperes.

There are four principal mechanisms that can contribute to the reverse bias leakage current of a SACM structure. The first two mechanisms have been described in Chapter 2. Diffusion current from minority carriers in the quasi-neutral region of the device shows a dependence on the temperature proportional to $e^{-E_g/kT}$, where E_g is the bandgap of the material. This temperature dependence is due to the n_i^2 term (see equation 2.3, Chapter 2) where n_i is the intrinsic carrier concentration.

The second form of leakage current is due to the generation of carriers in the depletion region defined by the junction. The extent to which the current increases can be defined by considering how the depletion region grows, which in turn is dependent on the doping profile. The temperature dependence of this term is proportional to $e^{-E_g/2kT}$ and also related to the n_i term (see equation 2.4, Chapter 2).

The third form of leakage current is the generation of carriers at the surface of the semiconductor due to electrically active states at the semiconductor/insulator surface. The value of this current ($J_{surface}$) can expressed by:

$$J_{surface} = qS_0 n_i \left[\exp\left(\frac{qV}{2kT}\right) - 1\right] \quad (5.1)$$

where S_0 is the surface recombination velocity and is proportional to the interface trap concentration. In reverse bias, this current has no dependence on voltage. As a function of temperature, however, devices dominated by this mechanism will show a dominant exponential dependence on ~$E_g/2$.

The last current mechanism is the tunnelling (J_{tunnel}) and field assisted emission which can be expressed as:

$$J_{tunnel} \propto \exp(E^x) \quad (5.2)$$

where E is the electric field and x is a fitting parameter. This kind of mechanism is very difficult to study, and can affect both the bulk and the surface of a device.

5.4.1 I-V characteristic of sample 11-167

Figure 5.5 shows the I-V characteristic measured for a 50 µm diameter device of wafer 11-167 (1×10^{17} cm^{-3} designed charge sheet doping concentration). This characterisation was performed in dark conditions (solid line) and by illuminating the device at a wavelength of 1310 nm (dashed line). These measurements were also performed over a range of different operating temperatures: 293 K (black), 200 K (red), 150 K (blue), 125 K (magenta), and 100 K (green).

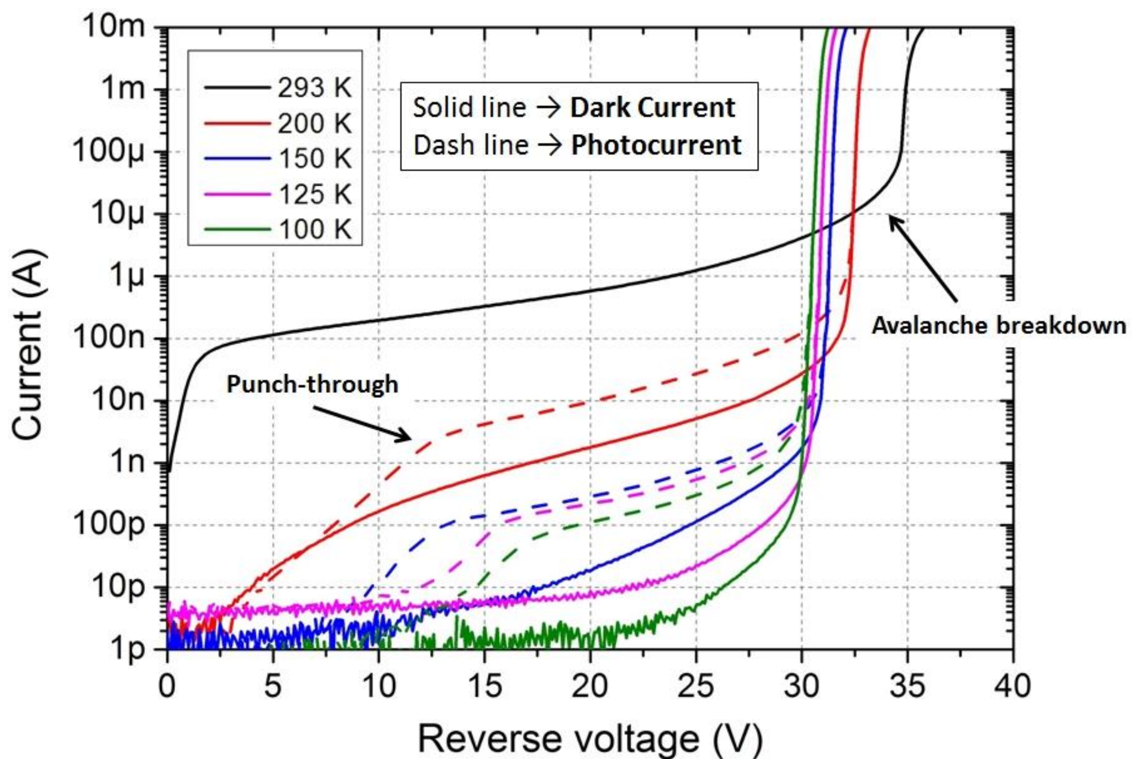


Figure 5.5. I-V characteristic measured for a 50 µm diameter device of sample 11-167 (1×10^{17} cm^{-3} designed charge sheet doping concentration). In particular, dark- (solid line) and photo-currents at λ = 1310 nm (dashed line) were measured over a range of different operating temperatures: 293 K (black), 200 K (red), 150 K (blue), 125 K (magenta), and 100 K (green).

When a reverse bias is applied to the device, its depletion region starts to extend into the Si multiplication region, and then expand in to the Ge absorption region, when the bias reaches the punch-through level. Once the device is fully depleted, the increase in bias generates an increase of the electric field, until the field is so high in the multiplication region, impact ionisation is possible and the avalanche breakdown occurs at a voltage

level equal to V_{BD}. The photocurrent (dashed line) clearly shows the punch-through voltage (~ -12 V): when the depletion region extends to the Ge absorber layers, the current increases due to the photo-generated carriers. For the device in Figure 5.5, the breakdown at room temperature (293 K) occurs at -34 V and the dark current at 95 % of V_{BD} is 15 μA. The dark current measured at 95 % of V_{BD} represents a reasonable indicator of the magnitude of the dark count rate to be expected when the device is operated in Geiger mode, although the precise physical relationship between these quantities is not well established [5], [6]. Because of the high level of dark current at room temperature, Ge-on-Si SPADs operate at lower temperatures. For that reason, I-V measurements were acquired at lower temperatures (Figure 5.5).

When the temperature decreases, the V_{BD} decreased due to the increased electron mean free path in the multiplication region, as already described in Chapter 2. Figure 5.6 illustrates the variation of the breakdown voltage in relation to the temperature. A temperature coefficient $\gamma \approx 0.043\ \%/K$ was calculated which is similar to the γ value of ~0.05 %/K obtained from simulations (section 4.5 of Chapter 4), confirming that the avalanche breakdown was consistent with Si breakdown. In contrast, the punch-through voltage V_{PT} remained relatively unchanged, considering its dependence on the doping profiles of the junction and the layers thickness. In addition, as the temperature decreases, the dark current also decreases, since the thermal generation-recombination effect is reduced.

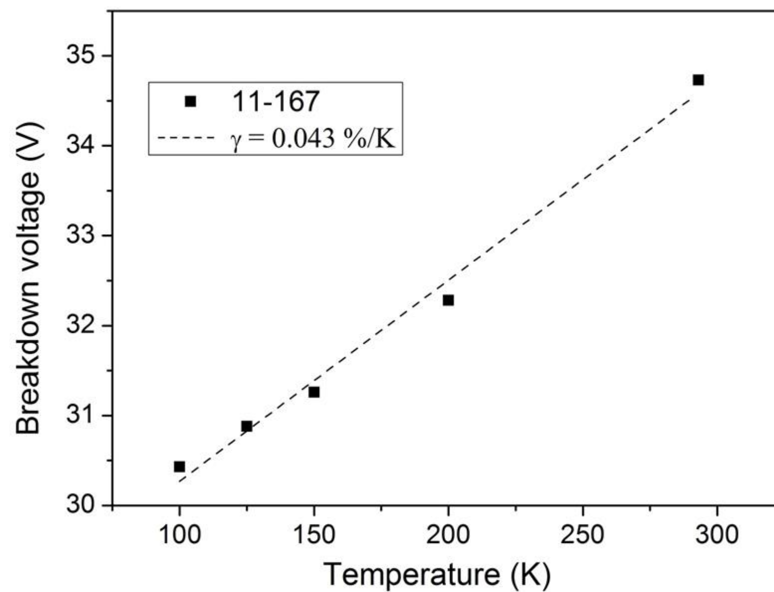
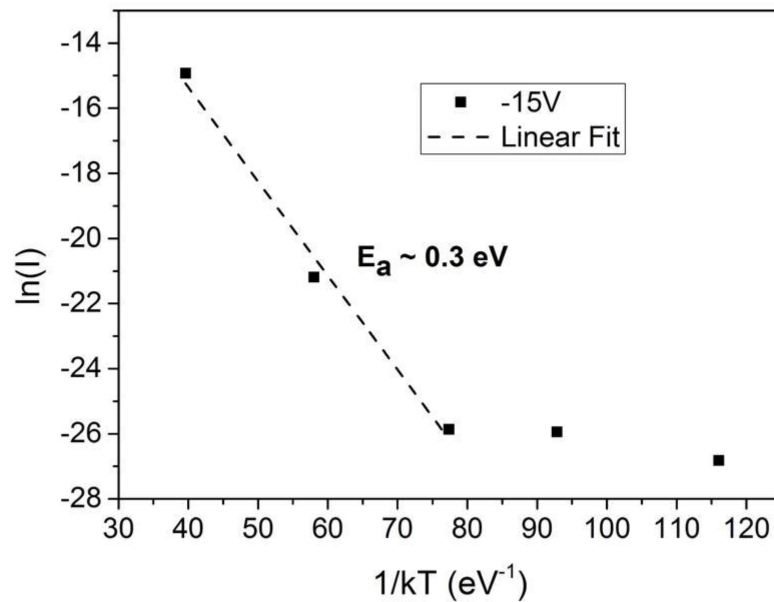
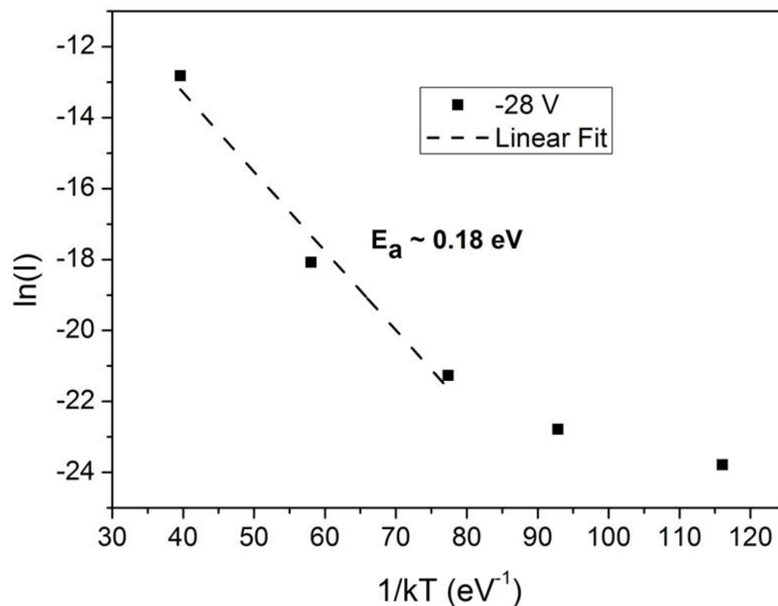


Figure 5.6. Breakdown voltage as function of the temperature for a 50 μm diameter device of sample 11-167 (1 × 10^{17} cm^{-3} charge sheet doping concentration). A temperature coefficient γ equal to 0.043 %/K was calculated.

To investigate the main mechanism contributing to the leakage current, Figure 5.7a-b shows the Arrhenius plot at two different voltages, -15 V and -28 V (i.e. reverse current (logarithm scale) versus the reciprocal of temperature to determine the activation energy (E_a)). The activation energy is close to the energy gap ($E_a \approx E_g$) when reverse current is dominated by a diffusion current, close to half of the energy gap ($E_a \approx E_g/2$) when dominated by a generation-recombination current, and less of half of the energy gap ($E_a \ll E_g/2$) when dominated by tunnelling or field-assisted emission.



(a)



(b)

Figure 5.7. Arrhenius plot of the leakage current for the 50 μm diameter device of wafer 11-167 at two different voltages: (a) -15 V, and (b) -28 V. For each voltage the extracted activation energy (E_a) is given.

The extracted activation energies at a reverse bias of -15 V and -28 V are ~0.3 eV and ~0.18 eV, respectively are shown in Figure 5.7a-b where, in both graphs, the activation energy was calculated by considering the leakage current between 293 K and 150 K. This was chosen because it is evident that, a different effect is contributing to the leakage current at lower temperatures than 150 K. At a reverse bias of -15 V, the activation energy (~0.3 eV) was half of the bandgap of Ge (Figure 5.7a), and this value might suggest that the generation-recombination in the absorber region is the main contribution to the leakage current at this reverse bias. Further contributions can also be that, at this reverse bias the Ge absorber layer is fully depleted (above punch-through voltage), as shown from the photocurrent in Figure 5.5. At a reverse bias of -30 V, activation energy of ~0.18 eV might suggest that the main mechanism contributing to the leakage current could be related to tunnelling or field-assisted emission. This hypothesis is supported by the following considerations:

a) At -30 V a high electric field can exist through the device, since this voltage is very close to the avalanche breakdown voltage;
b) Dark current measurements at different temperatures (Figure 5.5) showed that the dark current increases with an exponential behaviour before breaking down at high voltages near the avalanche breakdown. However, a constant behaviour of the dark current can be observed for InGaAs/InP or Si SPADs when the voltage is increased, and the device therefore breaks down with a steep profile without showing any exponential behaviour of the leakage current;
c) Wafer 11-167 has a charge sheet doping concentration equal or close to 1×10^{17} cm^{-3}. According to the electric field simulations shown in Chapter 4 (Figures 4.7 and 4.19), this charge sheet doping concentration gives rise to a high electric field in the absorber layer (above the Ge breakdown field) which might be responsible for the low activation energy extracted at this high reverse bias;
d) It was not possible to perform any single-photon characterisation even at the lowest temperature of operation and lowest excess bias, because the dark count rate was too high in order to make credible measurements;
e) Devices from this wafer (11-167) showed a low yield and reproducibility when an AC gate at a frequency of 10 kHz was applied. This caused the device to be irreversibly damaged. Probably due to the high electric field which can contribute to microplasma effects at the Ge/Si interface where a high concentration of defects was located.

5.4.2 I-V characteristic of sample 11-141

Figure 5.8 shows a comparison of the dark I-V characteristic measured for 50 μm (solid line) and 25 μm (dash line) diameter devices from wafer 11-141 (2×10^{17} cm^{-3} designed charge sheet doping concentration). These measurements were also performed over a range of different operating temperatures: 293 K (black), 200 K (red), 150 K (blue), 125 K (magenta), and 100 K (green).

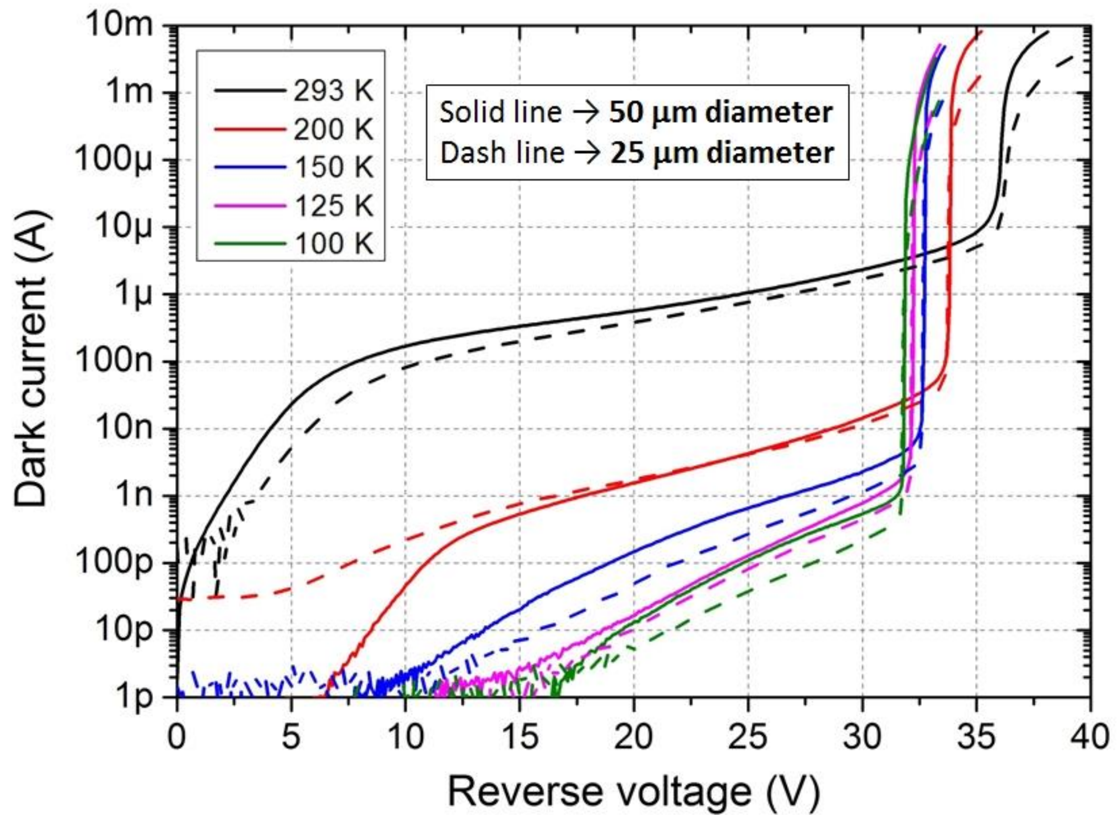


Figure 5.8. Comparison of the dark I-V characteristic measured for 50 μm (solid line) and 25 μm (dashed line) diameter devices of sample 11-141 (2×10^{17} cm^{-3} designed charge sheet doping concentration). In particular, dark-current was measured over a range of different operating temperatures: 293 K (black), 200 K (red), 150 K (blue), 125 K (magenta), and 100 K (green).

The dark current measured for the 25 μm diameter device was slightly lower than that from the device with 50 μm diameter (Figure 5.8). The breakdown at room temperature (293 K) occurred at ~ -36 V and the dark current at 95 % of V_{BD} was ~ 40 μA. A temperature coefficient $\gamma \approx 0.059\ \%/K$ was calculated (Figure 5.9).

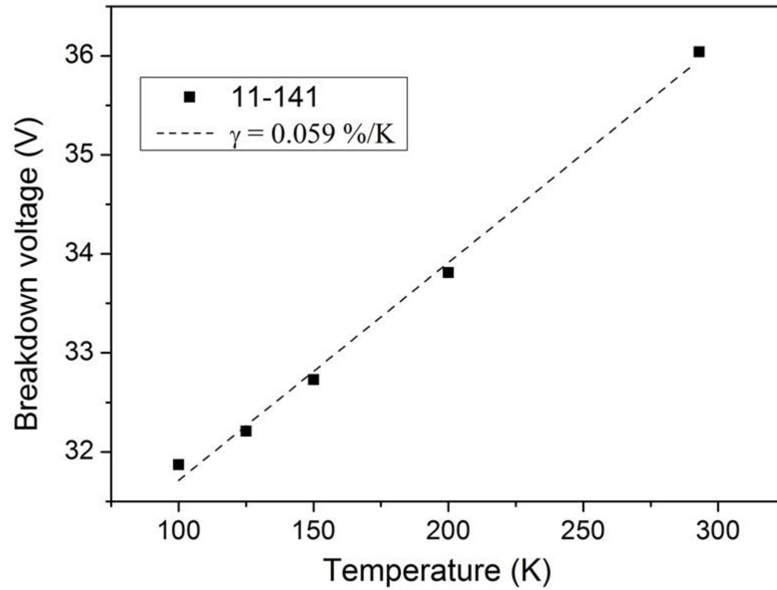
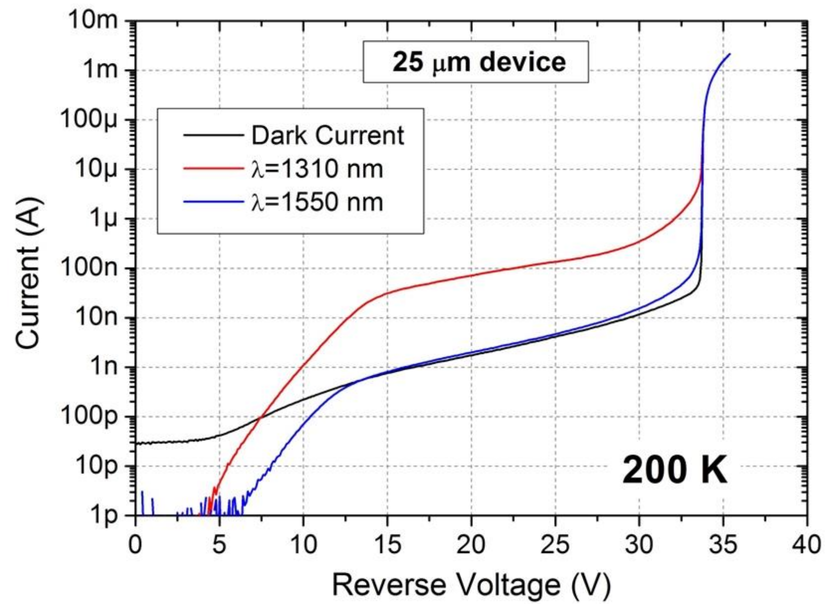
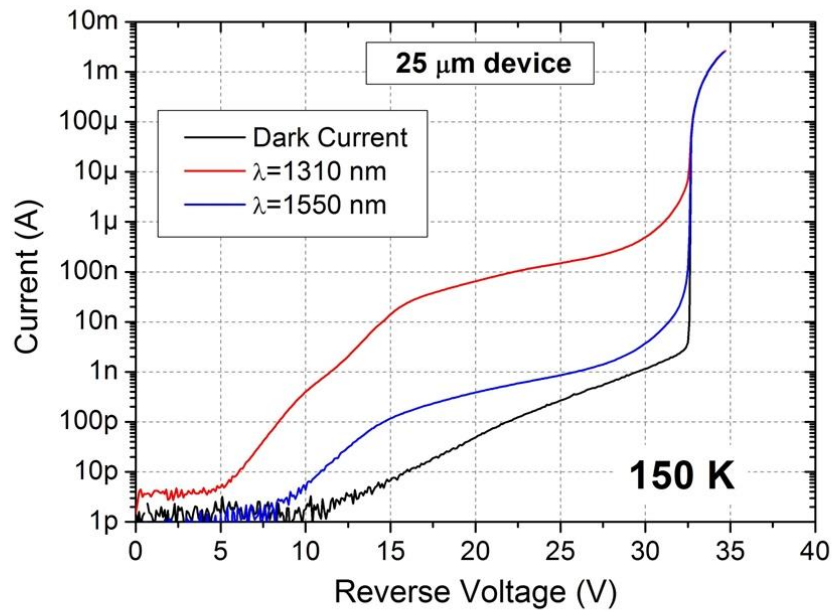


Figure 5.9. Avalanche breakdown voltage as function of the temperature for a 25 μm diameter device from sample 11-141 (2×10^{17} cm^{-3} designed charge sheet doping concentration). A temperature coefficient γ equal to 0.043 %/K was calculated.

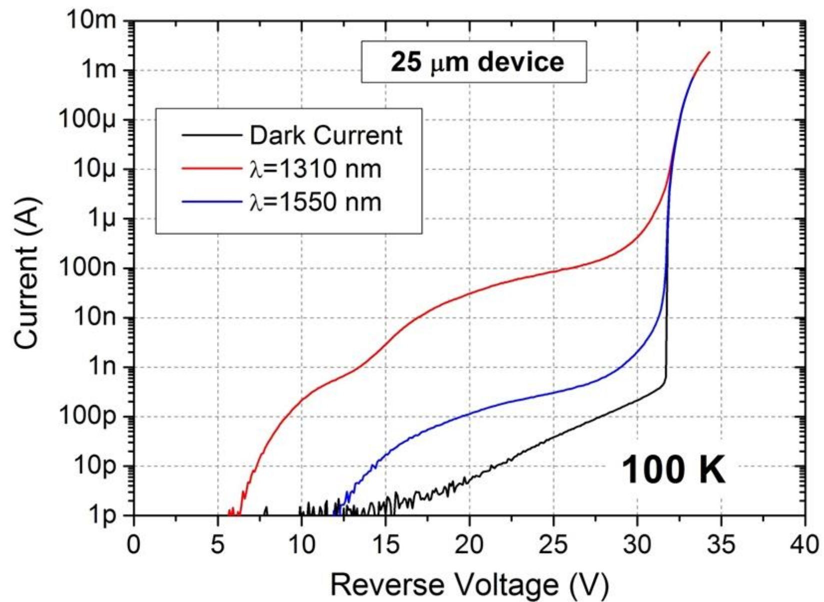
Figure 5.10a-c illustrates the dark- (black) and photo-current measured at a wavelength of 1310 nm (red) and 1550 nm (blue) at temperatures of 200 K, 150 K, and 100 K.



(a)



(b)



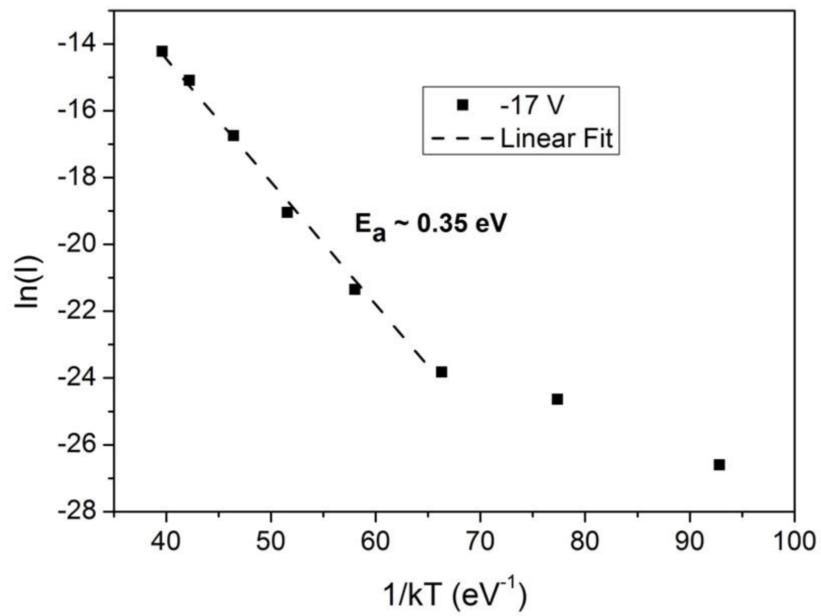
(c)

Figure 5.10. I-V characteristic measured for a 25 μm diameter device from wafer 11-141 (2 × 10^{17} cm^{-3} designed charge sheet doping concentration) in dark condition (black) and by illuminating the device at λ=1310 nm (red) and λ=1550 nm (blue) at different temperatures: (a) 200 K, (b) 150 K, and (c) 100 K.

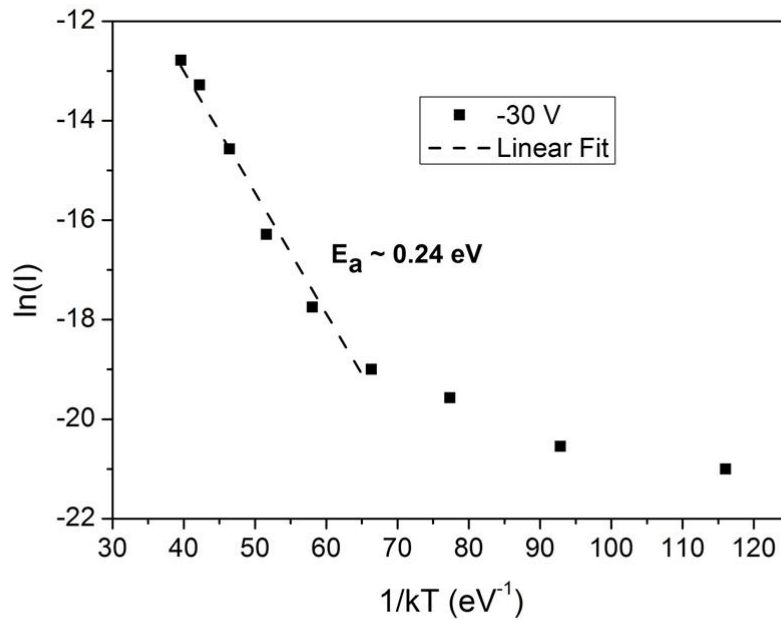
The punch-through voltage was measured to be ~ -15 V. Considering the photocurrent measured at 1310 nm and 1550 nm, it was evident that a decrease of the photocurrent at the longer wavelength was due to the absorption coefficient of Ge (which decreases at longer wavelengths, please see Figure 3.16 of Chapter 3 for details) which was ~10^4 cm^{-1} at λ=1310 nm and one order of magnitude lower at λ=1550 nm at room temperature.

In addition, when the temperature decreased to 100 K, the Ge absorption coefficient remained almost unchanged at λ=1310 nm, while dramatically decreasing to less than 10 cm^{-1} at λ=1550 nm.

As shown in Figure 5.11a-b, the activation energies extracted from the Arrhenius plot were ~0.35 eV and ~0.24 eV, at -17 V and -30 V, respectively. As observed for wafer 11-167, the activation energy at -17 V (after punch-through) suggested that the main mechanism contributing to the leakage current is the generation-recombination in the bulk Ge. At -30 V (before the avalanche breakdown), the activation energy was less than half of the bandgap, suggesting also in this case that tunnelling or field-assisted emission are contributing to the leakage current. However, this value (~ 0.24 eV), in comparison to the activation energy calculated for sample 11-167 was slightly higher, indicating that this effect might be less pronounced in sample 11-141. To further support this assumption, the dark current of sample 11-141 showed a steep behaviour at the breakdown (as clearly shown in Figure 5.10a-c). Devices from wafer 11-141 were characterised in terms of their single-photon performance, as is illustrated in the next sections.



(a)



(b)

Figure 5.11. Arrhenius plots of the leakage current for the 25 μm diameter device of wafer 11-141 at two different voltages: (a) -17 V, and (b) -30 V. For each voltage the extracted activation energy (E_a) is given.

5.4.3 Comparison between devices from samples 11-167 and 11-141

Figure 5.12 compares the dark I-V characteristics for 50 μm diameter devices of samples 11-167 (solid lines) and 11-141 (dash lines) at different operating temperatures: 293 K (black), 200 K (red), 150K (blue), and 100 K (magenta).

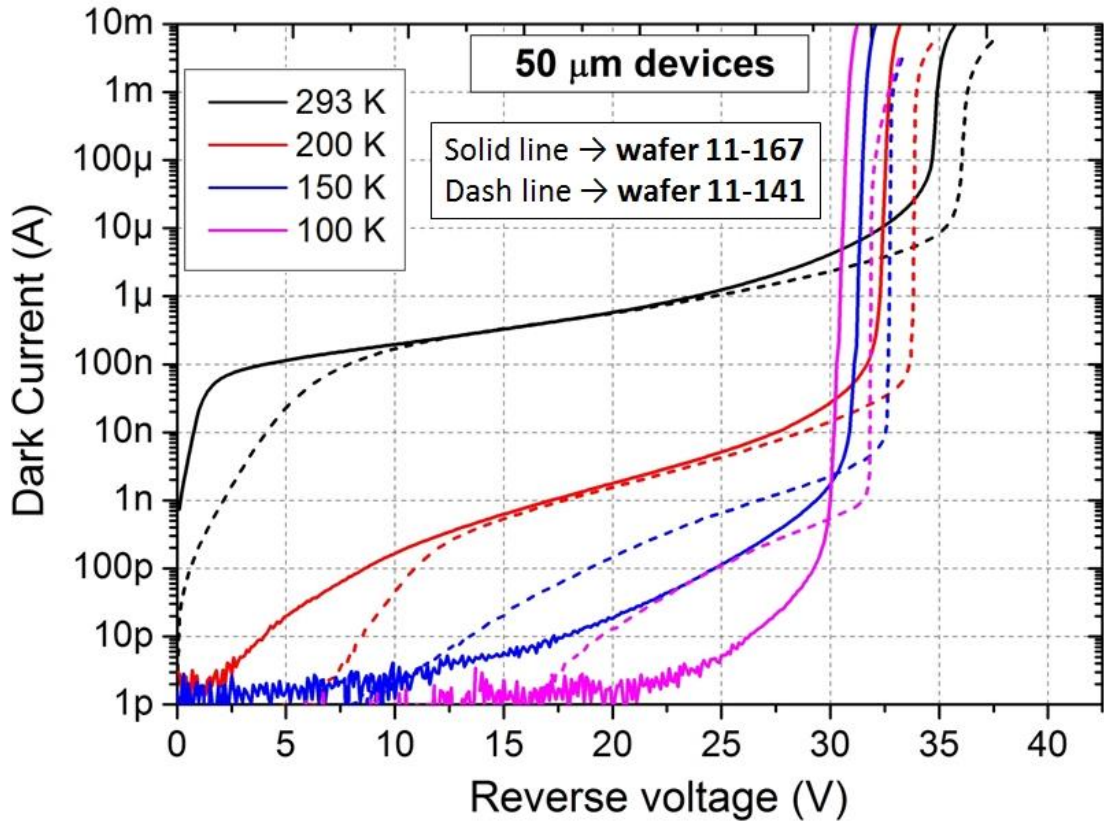


Figure 5.12. Comparison of the dark I-V characteristic measured for a 50 μm diameter device of wafer 11-167 (solid line) and wafer 11-141 (dash line) at different operating temperatures: 293 K (black), 200 K (red), 150 K (blue), 100 K (magenta).

As shown in Figure 5.12, a voltage difference of ~2 V was observed for the breakdown voltages (~ -36 V and ~ -34 V for device of wafer 11-141 and 11-167, respectively) at room temperature of the two wafers. A similar voltage difference was also measured for the punch-through voltage, corresponding to ~ -15 V and ~ -12 V for the two wafers (details are described in the previous sections). The two devices showed similar dark currents at room temperature and 200 K, while it was slightly lower (between -15 V and -28 V) for the device from wafer 11-167 at lower temperatures (150 K and 100 K).

There was a clear difference avalanche breakdown for the two devices. Devices from wafer 11-141 showed a steep curve before breaking down, while devices from wafer 11-167 showed an exponential behaviour before breaking down. Although the two wafers had the same structures (Figure 5.1), these were designed to have different doping concentrations of the charge sheet, 1 and 2 × 10^{17} cm^{-3} for wafers 11-167 and 11-141, respectively. Simulations performed in Chapter 4 (Table IV) showed that these two different doping concentrations of the charge sheet led to a difference in the breakdown

voltage of ~ 10 V and in the punch-through voltage of ~ 8 V. This voltage spread was larger than the ones measured experimentally in Figure 5.12.

In addition, simulations also showed that devices with doping concentrations of the charge sheet of 1×10^{17} cm^{-3} had a higher electric field (at 95% of V_{BD}) in the Ge absorber layer, well above the Ge breakdown field, unlike devices from wafer 11-141 (2×10^{17} cm^{-3} charge sheet doping concentration). As already described, devices from wafer 11-167 showed a behaviour of the dark current before the breakdown voltage compatible with a current mechanism (tunnelling or field-assisted emission), and this was caused by the high electric field. Furthermore, the small difference between breakdown and punch-through voltages suggests that the two wafers had a similar doping concentration of the charge sheet layer. Simulations also showed that a device with a charge sheet doping concentration of 1.5×10^{17} cm^{-3} had a breakdown voltage of ~ -37 V and a punch-through voltage of ~ -12 V. These values are similar to the values measured for samples 11-167 and 11-141.

Additionally, the two pieces of wafer (one for wafer 11-167 and one of wafer 11-141) processed at the University of Glasgow were the last two pieces left from the edge of the wafer, and this also suggests that the thicknesses of the different layers could be thinner than the designed values due to non-uniformity of the temperature at the edge of the wafers during the heteroepitaxial growth. Therefore, the two wafers, 11-167 and 11-141, might have similar doping concentrations in the charge sheet layer but different thicknesses of the Si multiplication layer (smaller for wafer 11-167) leading to the observed differences in the breakdown and punch-through voltages and different behaviour of the dark current at the breakdown voltage. As shown in Chapter 4, the breakdown and punch-through voltages depend on different parameters such as the doping concentration and thicknesses of different layers. It was not possible to measure the doping concentrations and thicknesses of different layers to confirm these hypotheses because most of the material was used for testing different metals for ohmic contacts and fabrication of devices performed at the University of Southampton.

5.5 Single-Photon Characterisation

In order to characterise the devices as single-photon detectors, the gated quenching mode was used (as described in Chapter 2). The gated mode allows a precise characterisation of the device.

5.5.1 TCSPC technique

The characterisation of the Ge-on-Si SPAD devices in terms of their single-photon performance was carried out using a time-correlated single-photon counting (TCSPC) technique which is described in detail in reference [7]. Individual photon detection can be treated as independent events that follow a random temporal distribution. As a result, photon counting is a classical Poisson process, and the number of photons *N* measured by a given detector over a time interval t is described by the discrete probability distribution:

$$P(N = k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!} \tag{5.3}$$

where λ is the expected number of photons per unit time interval, which is proportional to the incident light irradiance, and the rate parameter λt corresponds to the expected incident photon count. Figure 5.13 shows the probability distribution of a Poisson process for different λt. The uncertainty described by this distribution is known as photon noise. Since the incident photon count follows a Poisson distribution, it has the property that its variance is equal to its expectation λt. Therefore, the photon noise is signal dependent, and its standard deviation grows with the square root of the signal. Since photon noise is derived from the nature of the signal itself, it provides a lower bound on the uncertainty of measuring light. In general, the only way to reduce the effect of photon noise is to capture more signal, and hence photons, by using longer exposure times. The ratio of signal to photon noise grows with the square root of the number of photons captured $\sqrt{\lambda t}$. As already shown in Chapter 2, dark current noise also follows a Poisson distribution due to the thermally generated carriers which cause the detector to release electrons at random.

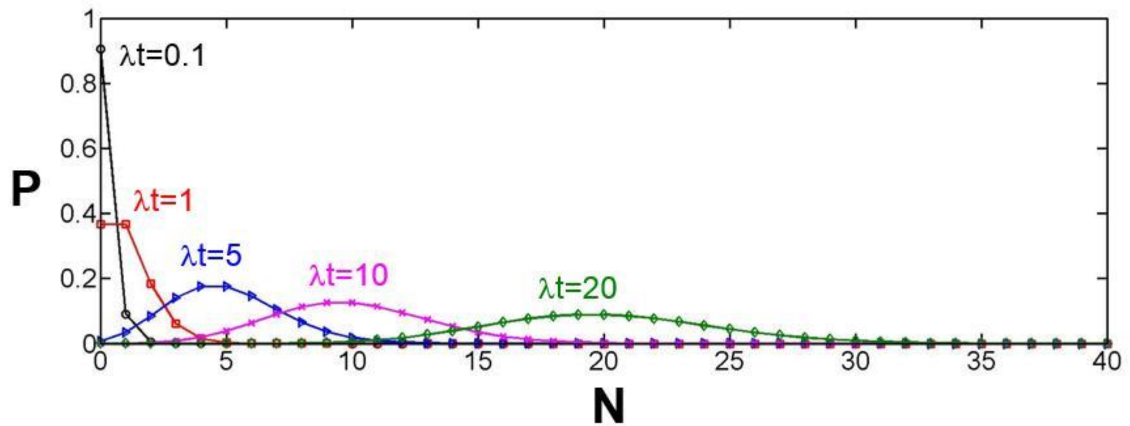


Figure 5.13. The Poisson probability distribution for various values of the expected incident photon count (λt).

The TCSPC technique relies on measuring a sequence of electrical pulses generated by the detection of a stream of single-photons. Each photon in the stream is recorded with respect to its arrival time relative to a fixed reference signal, usually provided by an external clock, as shown in Fig. 5.14.

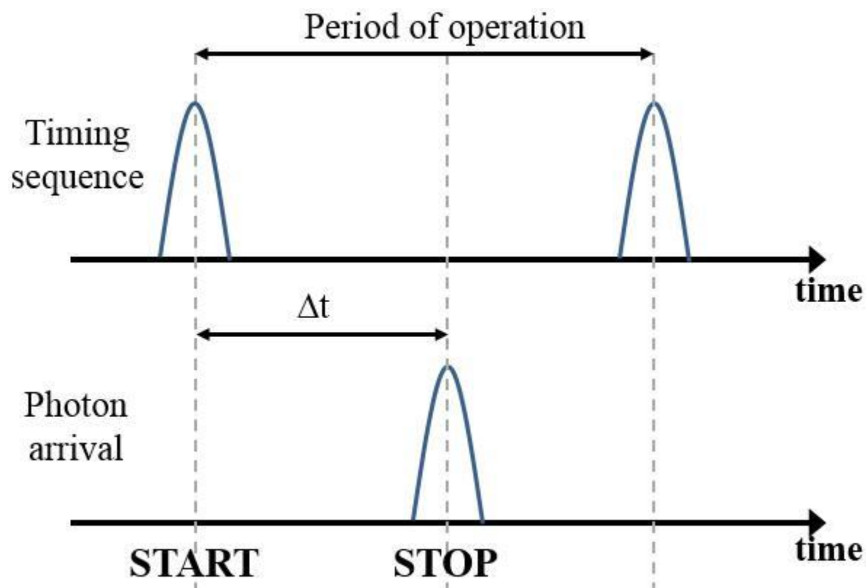


Figure 5.14. Timing diagram for a TCSPC operation showing the start signal, usually provided by an external clock, the stop pulse from the SPAD detector after the detection either of a single-photon or dark count. Δt represents the time difference between the start and stop pulse.

SPADs used in TCSPC mode produce an electrical output pulse for each detected single-photon. The amplitude of the output signal is not dependent on the intensity of the incoming photon flux. These events are often recorded by a photon-counting card which produces a histogram as shown in Figure 5.15.

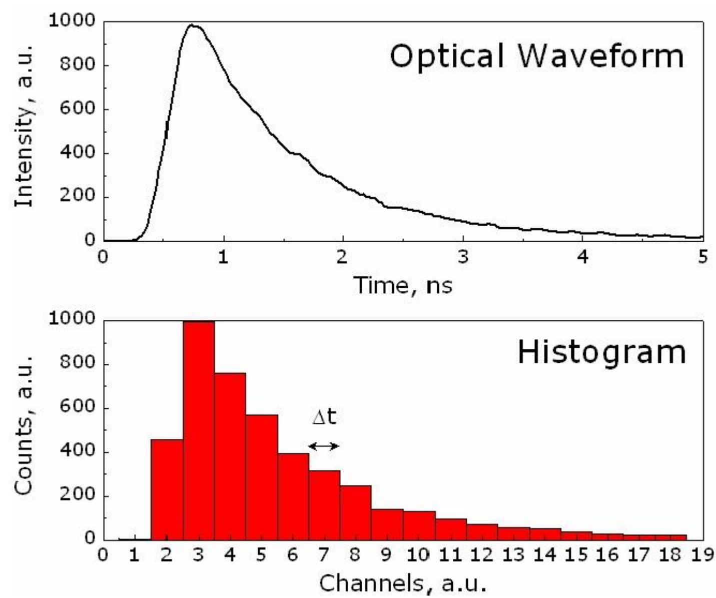


Figure 5.15. Histogram of photon counts recorded according to arrival time obtained with TCSPC. Each time a photon event is measured a count is added to the appropriate time bin [5].

The histogram bin width can be chosen by the user before recording the histogram depending on the required accuracy.

The accuracy of the time measurements itself is not limited by the width of the detector pulse. TCSPC is a thresholding technique where only the first part of the voltage pulse from the detector is used to indicate the stop of the timing process by using a constant fraction discriminator (CFD). A constant voltage threshold does not fully exploit the high timing resolution that is achievable using TCSPC as shown in Figure 5.16 [8]. The voltage discriminator level is set to a value which enables all output pulses of the device to be counted, if the voltage pulse does not exceed this level, the pulse will not be counted. Since the voltage pulses from the detector may differ slightly in magnitude, a threshold voltage might be achieved at a slightly different time which increases the inaccuracy of the timing.

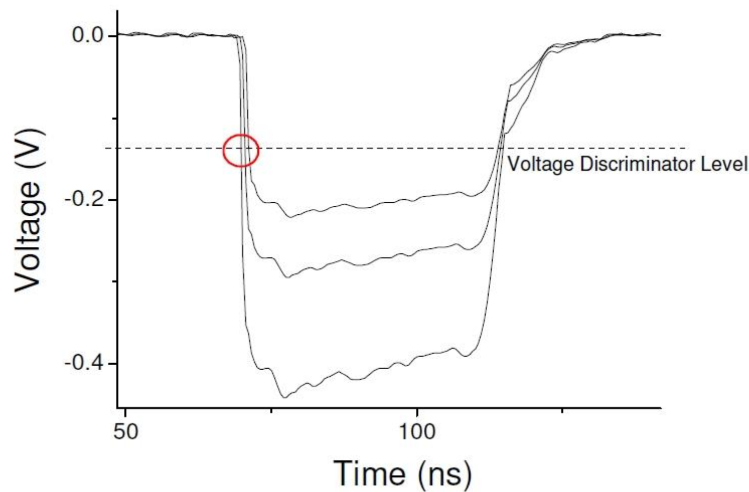


Figure 5.16. Single-photon detector temporal output as seen on a GHz oscilloscope. The voltage discriminator level (dashed line) is set to a value which enables all output pulses of the device to be counted. However, the voltage pulses pass through the discriminator level at different times (red circle) relative to the start of the individual measurement. This variance in the rise-time results in the jitter seen in the histogram of the time difference between the start and stop pulses [8].

The CFD method, as suggested by the name, uses a constant fraction point on the leading edge of the pulse to trigger the timing stop pulse, resulting in a constant timing point for pulses with similar shapes yet varying amplitudes. A second CFD is also used on the reference signal (start) to prevent possible amplitude fluctuations which may contribute to the overall time jitter. The output pulses of the CFDs are used as start and stop pulses of a time-to-amplitude converter (TAC). The TAC within a photon counting card generates an output signal proportional to the time between the start and stop pulses, and it can only register one stop for each start event. The TAC then has a certain reset, or "dead-time", during which time no stop pulses can be counted. Therefore, the count rate of a TCSPC system must not exceed ~ 10 % of the overall system repetition rate (the "start" rate) (i.e. the probability of detecting a photon per clock period is much less than unity). This avoids a problem known as "pulse pile-up" whereby there are many photons, but only one start-stop pair per clock-cycle, so only early events are recorded in the TAC window. This leads to a skewing of the recorded histogram towards the start of the timing window and results in an unreliable measurement, since the likelihood of observing any photon counts in the latter part of the window is greatly reduced. For the measurements described throughout this Chapter, the count rate was always kept to less than 10 % to avoid pulse pile-up.

5.5.2 *Experimental Setup and Methodology*

In order to characterize the devices in terms of single photon detection efficiency (SPDE), dark count rate (DCR), noise equivalent power (NEP), and jitter at full-width half-maximum (FWHM) the setup in Figure 5.17 was used.

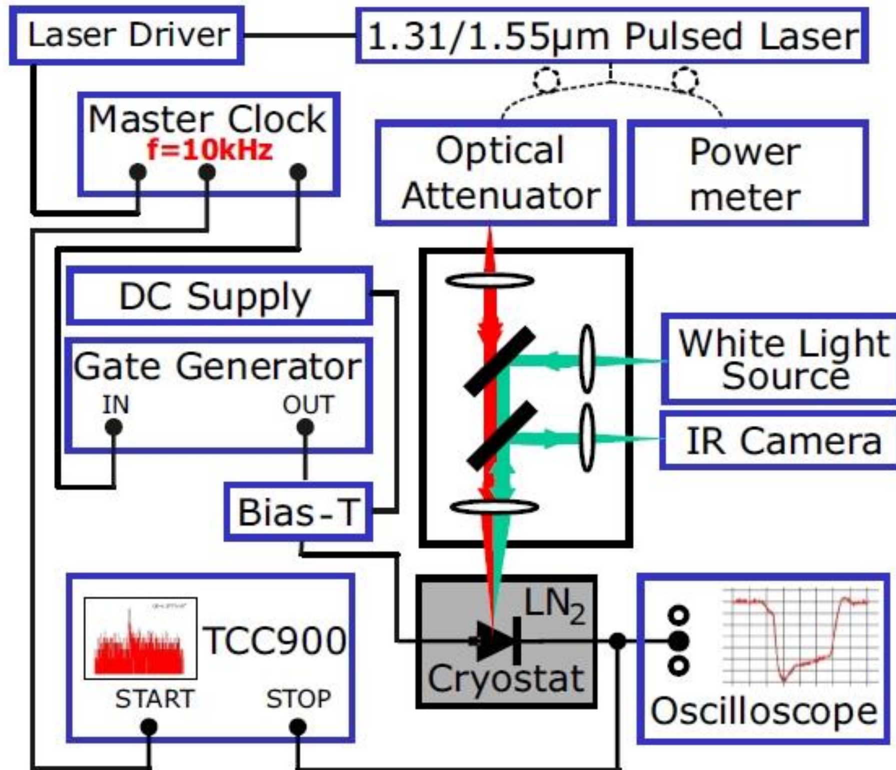


Figure 5.17. Single-photon characterisation setup. The gated quenching mode was used to characterise the devices in terms of SPDE, DCR, NEP, and jitter at FWHM. The connections denoted by black lines are electrical connections. Those with dotted lines are fibre-optic connections, and those with coloured lines are free-space optical connections.

Devices were mounted in a liquid nitrogen cryostat with free-space optical access (Oxford Optostat DN2), enabling accurate and stable temperature tuning between 77 K and 300 K. Two pulsed semiconductor laser modules manufactured by PicoQuant were used to test the detection efficiency at wavelengths of 1310 nm and 1550 nm respectively. The output from these lasers was coupled into single-mode fibre (SMF-28) and into a 50/50 splitter. One arm of the splitter was used to constantly monitor the optical power level, whilst the other passed through a calibrated optical attenuator (HP 8156A) with up to 60 dB of attenuation available. Prior to characterisation, the optical power level arriving at the device under test was measured and compared with the reading on a calibrated fibre-coupled power meter. This enabled direct monitoring of

the power level arriving at the device, once the optical pulse had passed through the optical attenuator, and thereby ensuring that a photon flux of <0.1 photons per pulse (on average) was incident on the device to properly demonstrate the single photon counting abilities of the detectors.

The average number of photons *n* in a single pulse is calculated according to:

$$n = \frac{P}{h\nu f} \tag{5.4}$$

Where P is the optical power and f is the laser repetition rate. The distribution of photons in a laser pulse follows Poisson statistics (see section 5.5.1). With an average number of photons per pulse of 0.1, there is a ~10% of probability that a given pulse contains a photon, while the probability to have 2-photons per pulse is less than 0.5%. Therefore, if the repetition rate of the ac pulse is 10 kHz, a one second integration time allows for a maximum of 10^4 photon counts. However, the value of 0.1 represents the average number of photons, and hence there is a certain probability that a pulse will contain a certain number of photons for a given average photon number. As the average photon number per pulse is increased, then the probability to have more than one photon per pulse also increased. In rigorous single-photon characterisation the average photon number must be kept less than 1 photon per pulse.

The fibre output of the optical attenuator was collimated with a microscope objective, and directed at the sample. A white-light channel was also present in the system and coupled through the use of a partially reflecting pellicle. The image was relayed to an InGaAs camera (Hamamatsu C10633) through the use of another pellicle to enable imaging of the device, and the laser spot on the sample, to ensure accurate alignment and focusing. The devices were electronically addressed with GHz bandwidth sub-miniature coax cables to minimise any deterioration of the electric pulses to, and from, the devices. For characterisation purposes, the devices were operated in gated-mode [9]. A dc bias of few Volts below the breakdown voltage V_{BD} was applied, and an electrical pulse was superimposed through the use of a bias-tee in order to bias the device above V_{BD} into the so-called Geiger mode of operation. When an avalanche was initiated (by either a dark count or photo-generated event) the avalanche current persisted until the end of the gate when the voltage was brought back below V_{BD}. The output pulse from the device was split to enable oscilloscope traces to be recorded while simultaneously providing the stop signal for the photon counting card (Edinburgh

Instruments TCC900). The start signal was provided by a master clock that has three outputs: one for the TCC900, another for the laser driver, and the third output for the gate generator. Various delays were inserted into the start and stop paths and between the triggering signals and the laser driver/ gate generator to synchronize all signals and enable a histogram of photon/dark count arrival time relative to the start signal to be recorded by the TCC900. For each operating condition, two photon-counting histograms were recorded: one in completely dark conditions, and one with an attenuated laser pulse coincident with the gate on the detector.

The gated mode operation was used for all the experimental results reported in the following sections. This consisted of applying a constant DC bias to the detector at ~1 V below V_{BD}, then an AC gate (~50ns duration gate) was superimposed at amplitude corresponding to a relative excess bias (V_{ex}) of 10% of V_{BD} which was calculated as follows:

$$V_{ex}(\%) = \frac{(V_{DC} + V_{gate}) - V_{BD}}{V_{BD}} \times 100 \tag{5.5}$$

5.5.3 Dark Count Rate

The mechanisms that contribute to the dark count rate of a SPAD have been described in Chapter 2. The dark count probability P_D can be calculated as

$$P_D = \frac{N_D}{f \cdot t_{acq}} \tag{5.6}$$

Where N_D is the total number of dark counts per measurements, f is the repetition rate of the gate pulse, and t_{acq} is the total acquisition time for the measurement. The dark count rate (DCR) can be calculated by dividing the dark count probability by the gate duration (t_g), yielding the number of dark counts per one second. The DCR of the characterised Ge-on-Si devices was so high that the detectors could not be used and characterised at room temperature. However, by cooling down the device, the dark count rate was reduced almost exponentially and DCR measurements were performed between temperatures from 100 K to 150 K.

Figure 5.18 illustrates the DCR as a function of the relative excess bias for two 25 µm diameter devices of wafer 11-141.

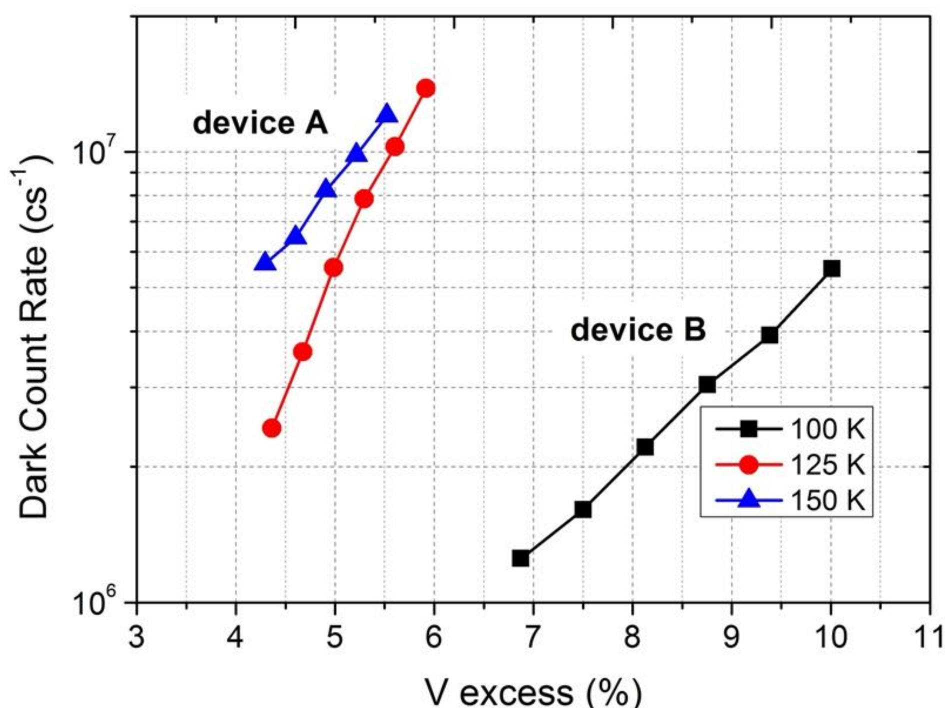


Figure 5.18. Dark count rate as a function of the relative excess bias for two 25 μm diameter devices of wafer 11-141. In particular, the DCR for the device A were measured at two different temperatures, 150 K (blue) and 125 K (red), while the DCR of device B was only measured at 100 K (black).

As shown in Figure 5.18, the DCR of device A was measured at two different temperatures, 150 K (blue) and 125 K (red), while device B was only measured at 100 K (black). It was not possible to characterise the DCR of a single device at different temperatures, since devices of wafer 11-141 could only be thermally cycled a few times prior to irreversible damage. It was not clear if this damage was related to the Geiger mode characterisation, the device fabrication or the material itself.

However, both devices exhibited a high DCR which in turn limited the operating temperatures and restricted the maximum excess bias applied. It is evident that the DCR decreased as the temperature was decreased. The lowest DCR ranging from 1 to 8 $\times$ 10^6cs^{-1} was measured at 100 K on device B at different relative excess biases. Although the DCR on the device B was measured at a lower temperature, the DCR of device A increased quite rapidly at higher temperatures, even if it was measured at lower relative excess bias than the one used on device B (Figure 5.18).

The DCR of two different size devices, 50 μm (black) and 25 μm (red) is compared in Figure 5.19. It was not possible to measure the DCR of the 50 μm diameter devices at temperatures higher than 100 K, since these devices demonstrated a higher dark current

near the breakdown than smaller size devices. At a temperature of 100 K, the dark currents at 95% of V_{BD} were 0.7 nA and 0.22 nA for the 50 and 25 μm diameter devices.

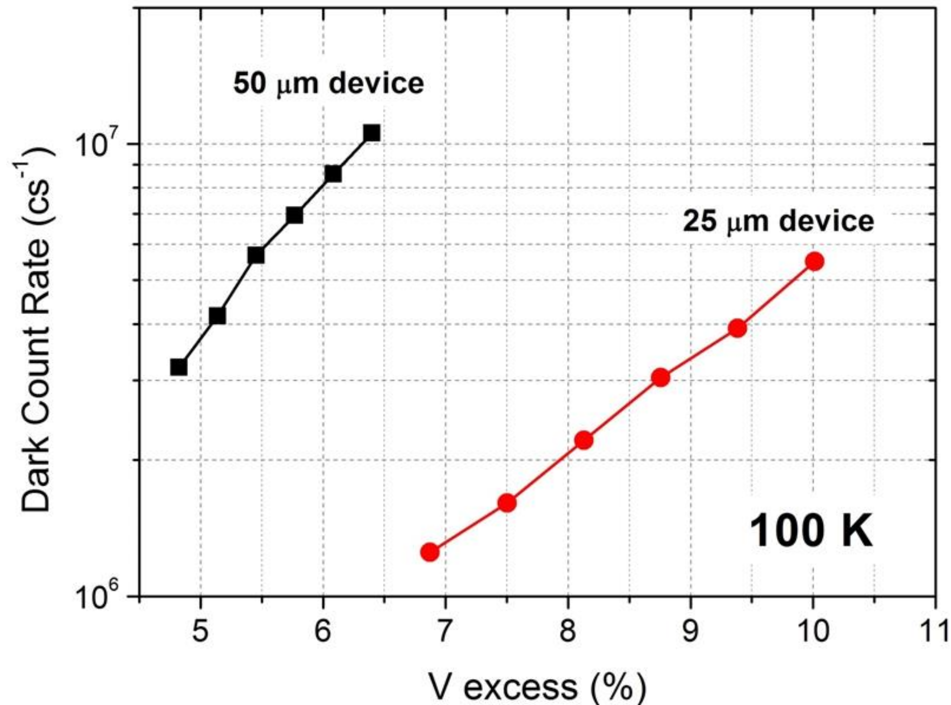


Figure 5.19. Dark count rate as a function of the relative excess bias for two different size devices: 50 μm (black) and 25 μm (black) diameter devices of wafer 11-141.

At 100 K, the DCR of the 50 μm diameter device was higher than the DCR of the 25 μm diameter device, and between 3 - 12 Mcs^{-1}, even if it was measured at lower relative excess bias.

All the DCR measurements, performed at different temperatures and with different size devices, demonstrated an exponential dependence of DCR on excess bias. There were not enough data for a single device at different temperatures to give an indication of the dependence of the DCR on the temperature. Results from the dark current suggested that the main current mechanism near the breakdown voltage was probably due to tunnelling or field-assisted emission (section 5.4.2). Considering that both these effects could be more pronounced in Geiger mode of operation, this suggests that these effects might be the main contribution to the DCR for the analysed structures. Furthermore, in future work, a more detailed analysis of DCR behaviour (i.e. an Arrhenius plot of DCR for different excess bias) would be needed to ascertain the main carrier mechanisms that are contributing to the DCR.

The DCR measured for the Ge-on-Si SPAD is similar to those obtained by using commercial Ge APD. In reference [10], two commercial Ge homojunction APDs fabricated in planar technology from GPD Optoelectronics (40 μm diameter) and Texas Instruments (350 μm diameter) were characterised in terms of their single-photon performance. At 100 K, the GPD device demonstrated a DCR of ~2 Mcs^{-1} at the lowest relative excess bias of 2.6%, while it increased to ~40 Mcs^{-1} at 13% of excess bias. Although, the Texas Instruments device had a large active area (350 μm diameter), this demonstrated a better fabrication process with less lattice defects considering that the measured DCR was between 2 Mcs^{-1} and 10 Mcs^{-1}, at the same temperatures. In fact, the authors stated that if a 40 μm diameter device had been fabricated with the same technology, this would have resulted in a Ge SPAD with a low DCR at 77 K, comparable to the commercially available InGaAs/InP SPAD at 200 K.

Compared to the planar InGaAs/InP SPAD, the DCR is several orders of magnitude higher as DCRs of $10^2 – 10^3$ cs^{-1} are achievable at these temperatures [6], [9].

However, the characterised Ge-on-Si SPAD device was fabricated in a mesa geometry, whereas Ge APDs and InGaAs/InP SPAD were fabricated as planar devices. It is well known that mesa geometry devices suffer from deleterious effects on dark count rate and dark current caused by the high density of surface states at the sidewalls of the mesa. This is mainly true for germanium, since surface passivation is still an open question in the scientific literature. In particular, DiLello *et al.* [11] demonstrated that the dark current of Ge-on-Si pin photodiodes is dominated by the generation of carriers at the germanium surface mainly in small device between 5 – 25 μm, while larger active area devices, 100 – 500 μm, are dominated by bulk carrier generations. A similar analysis was also performed in this work suggesting that the surface effect dominates the dark current, and hence the DCR in this sample set.

5.5.4 Single-Photon Detection efficiency

As explained in Chapter 2, the single-photon detection efficiency (SPDE) depends mainly on three different factors: the absorption efficiency, the transport efficiency and the triggering probability. In particular, the triggering probability depends on photo-generated carriers triggering an avalanche. This is strongly dependent on the excess bias, the depth at which the incident photon is absorbed, and the impact ionisation coefficients of both electrons and holes.

Experimentally, the SPDE was calculated using the formula (5.7) when the detector was operated in gated mode:

$$SPDE = \frac{Total\ measured\ counts - dark\ counts}{number\ of\ incident\ photons} \cdot \frac{1}{t_{acq} \cdot f} \quad (5.7)$$

Where t_{acq} is the total acquisition time for the measurements and f is the repetition rate.

For devices of sample 11-141, the highest SPDE was measured for a 25 μm diameter device at a wavelength of 1310 nm and a temperature of 100 K, as shown in Figure 5.20. The average number of photons incident on the device was kept less than 0.1 photons per pulse. A SPDE ranging from 2.2 % to 4 % was measured between 6.5 – 10 % of excess bias. Results showed that the SPDE increases linearly with the excess bias due to the increase of the field within the device.

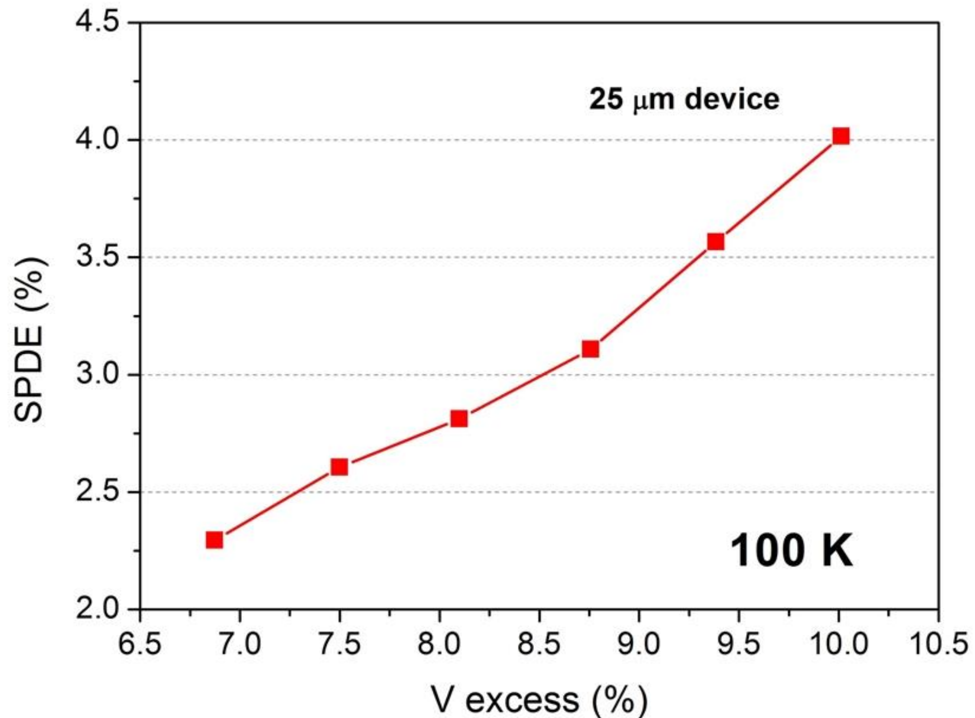


Figure 5.20. Single-photon detection efficiency as a function of the relative excess bias measured in Geiger mode at λ=1310 nm and T=100 K for a 25 μm diameter device from wafer 11-141. The average number of photons incident on the device was kept <0.1 photons per pulse.

Figure 5.21 shows a comparison between the SPDE measured at a wavelength of 1310 nm and a temperature of 100 K for both 25 μm and 50 μm diameter devices.

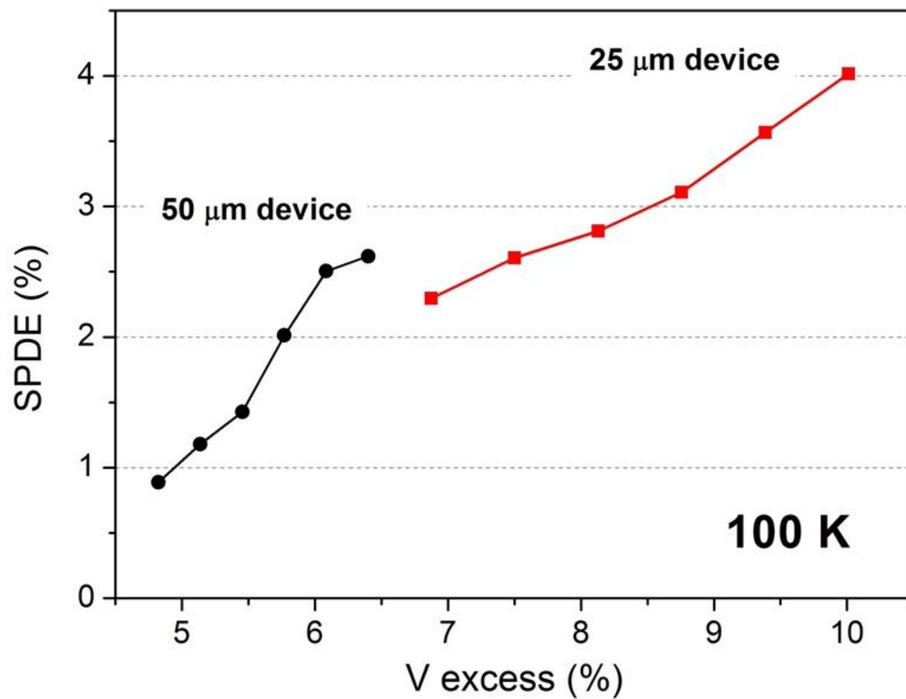


Figure 5.21. Comparison of the single-photon detection efficiency as a function of the relative excess bias measured in Geiger mode at λ=1310 nm and T=100 K for both 25 μm and 50 μm diameter device of wafer 11-141. The average number of photons incident on the devices was kept <0.1 photons per pulse.

The SPDE measured on the 50 μm diameter device (black line) was between 1% and 3% at a relative excess bias ranging from 4.5% to 6.5%. This relative excess bias was lower than the one used on the 25 μm diameter device (red line), and it was not possible to increase the excess bias on the 50 μm diameter devices due to the high DCR.

The SPDE at the longer wavelength of 1550 nm was measured on a 25 μm diameter device at 125 K, as shown in Figure 5.22. SPDE of ~0.09 % at 6 % excess bias was measured. As expected, the SPDE was quite low at this wavelength since the Ge bandgap increased as the Ge was cooled, and hence the absorption coefficient at longer wavelengths decreases rapidly (the direct bandgap at 125 K is 0.84 eV, and hence the 1550 nm (0.8 eV) photons lay outside the main absorption edge) [12].

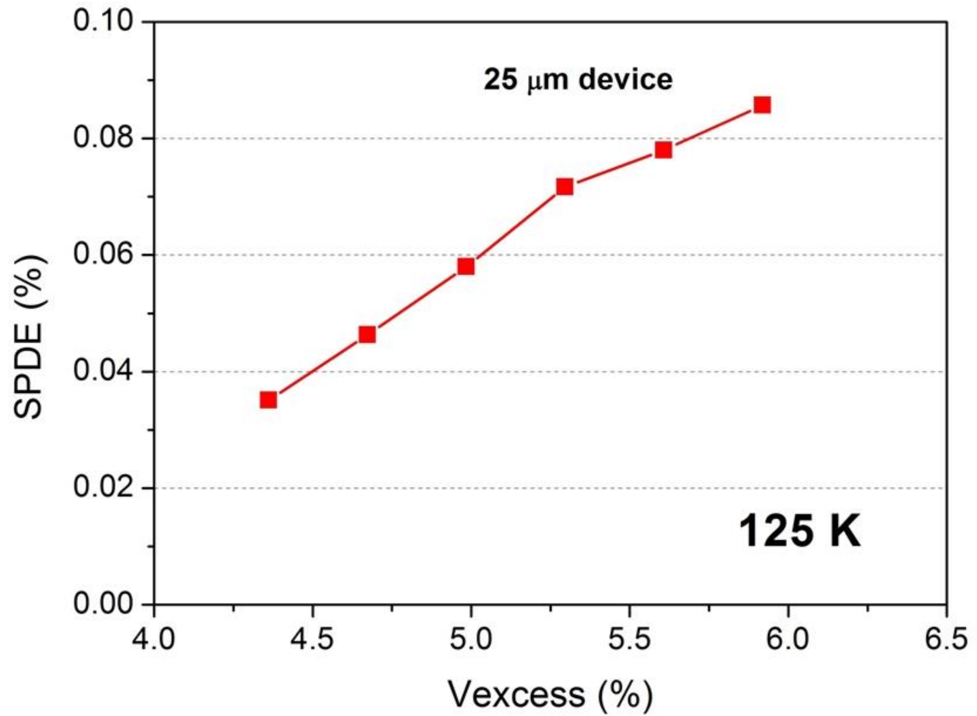


Figure 5.22. Single-photon detection efficiency as a function of the relative excess bias measured in Geiger mode at λ=1550 nm and T=125 K for a 25 μm diameter device of wafer 11-141. The average number of photons incident on the device was kept <0.1 photons per pulse.

The commercial Ge homojunction APD demonstrated a SPDE ranging from 4 % to 30 % at a wavelength of 1310 nm and a temperature of 77 K, while it decreased to 0.1 – 1 % at the longer wavelength of 1550 nm and same temperature [10]. These results were in good agreement with the SPDE measured on the Ge-on-Si SPAD devices from wafer 11-141.

In section 4.8 of Chapter 4, SIMS results performed on the 2nd generation of Ge-on-Si SPAD devices were shown. Although the Si multiplication layer was designed to have a very low background doping, these measurements revealed that there was a phosphorous diffusion tail which slowly decreased as a function of distance from the doped Si substrate. Simulations demonstrated that this dopant diffusion tail led to a non-uniform electric field in the Si multiplication layer, which led to a reduced effective thickness of the multiplication layer. This, in turn, led to an increased electric field in the Ge absorber layer (Figure 4.19 of Chapter 4). In particular, the first two problems suggested a low avalanche triggering probability which would lead to a lower SPDE. Since 1st generation of Ge-on-Si SPAD were grown under the same conditions as 2nd

generation devices, this problem might have affected the SPDE of the measured devices.

Another mechanism that might have affected the SPDE of the characterised SPADs is the threading dislocations. As already shown in Chapter 3 (section 3.4.2), Colace *et al.* demonstrated that the threading dislocations can increase the leakage current and also reduce the responsivity of Ge-on-Si pin diodes, since they act as generation-recombination centres [13]. Therefore, a more detailed analysis of the impact of the threading dislocation on the SPDE is needed to help the understanding of its contribution to the SPAD performance.

5.5.5 Noise Equivalent Power

The noise equivalent power (NEP) is an useful figure of merit for a SPAD, since it takes into account both the SPDE and DCR, as shown in Chapter 2 (section 2.4.2.1). The more sensitive the detector is to a given incident wavelength, then the lower the NEP.

The lowest NEP of ~ 1×10^{-14} $WHz^{-1/2}$ at a wavelength of 1310 nm was measured on a 25 μm diameter device at a temperature of 100 K. The NEP was similar across the range of the relative excess voltages measured (Figure 5.23).

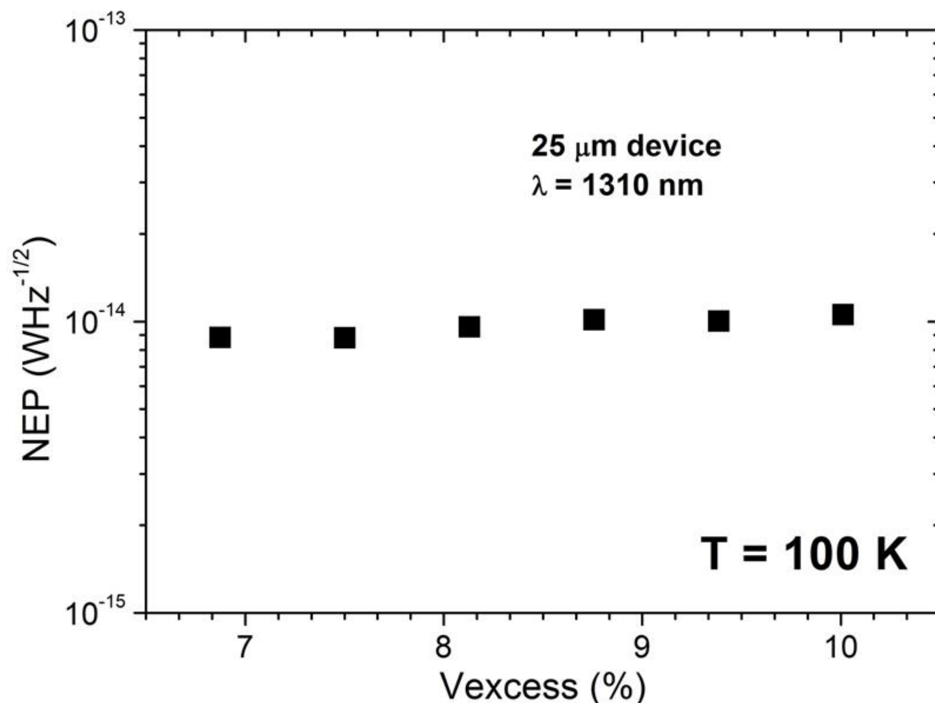


Figure 5.23. Noise equivalent power at a wavelength of 1310 nm measured on a 25 μm diameter device of wafer 11-141 at a temperature of 100 K.

Furthermore, at the same temperature and wavelength, the NEP measured for the 50 μm diameter device was in the range of ~ 2 - 4 × 10^{-14} $WHz^{-1/2}$, as shown in Figure 5.24.

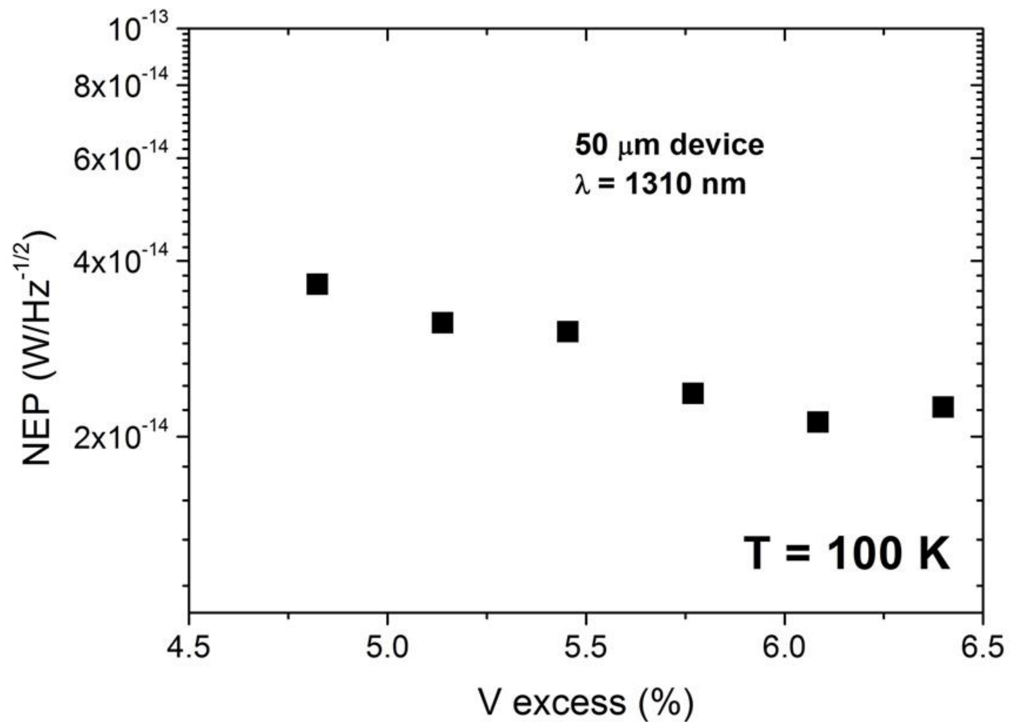


Figure 5.24. Noise equivalent power at a wavelength of 1310 nm measured on a 50 μm diameter device of wafer 11-141 a temperature of 100 K.

The NEP at the longer wavelength of 1550 nm was measured on a 25 μm diameter device at 125 K, as shown in Figure 5.25. An NEP of ~ 1 × 10^{-12} $WHz^{-1/2}$ over a range of excess biases was measured.

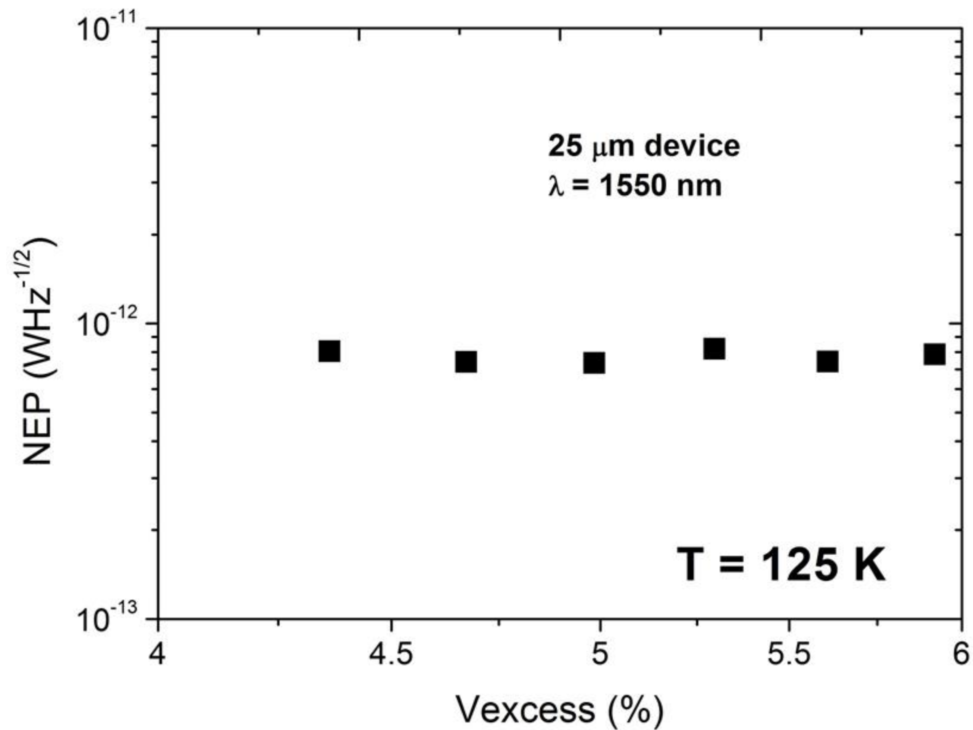


Figure 5.25. Noise equivalent power at a wavelength of 1550 nm measured on a 25 μm diameter devices of wafer 11-141 a temperature of 125 K.

These values of NEP (Figure 5.23) compared well with those obtained from previous works, where NEPs of ~ 1.6×10^{-14} $WHz^{-1/2}$ and ~ 4×10^{-15} $WHz^{-1/2}$ were reported using commercially available planar all-Ge APDs operated in Geiger mode at a temperature of 77 K [10], [14]. In addition, the measured NEP of ~ 1×10^{-14} $WHz^{-1/2}$ represents the lowest NEP demonstrated for any Ge-on-Si SPAD reported in the scientific literature, as shown in the next sections.

However, there still remains a significant performance gap between these results obtained for Ge-on-Si SPADs and those for InGaAs/InP SPADs which have previously demonstrated NEPs of less than 1×10^{-17} $WHz^{-1/2}$ at a wavelength of 1550 nm and a temperature of 193 K.

5.5.6 Timing jitter

Timing jitter at full-width at half maximum (FWHM) was investigated at various excess bias levels. The measured jitter was a convolution of the laser pulse width (~50 ps), the detector response, and the contribution from the rest of the acquisition system.

The time response for a 25 μm diameter device at a relative excess bias of 10 % and a temperature of 100 K is shown in Figure 5.26. The time response is given by the value

of the FMHM of the histogram measured when the detector is illuminated with a highly attenuated pulsed laser (<0.1 photons per pulse).

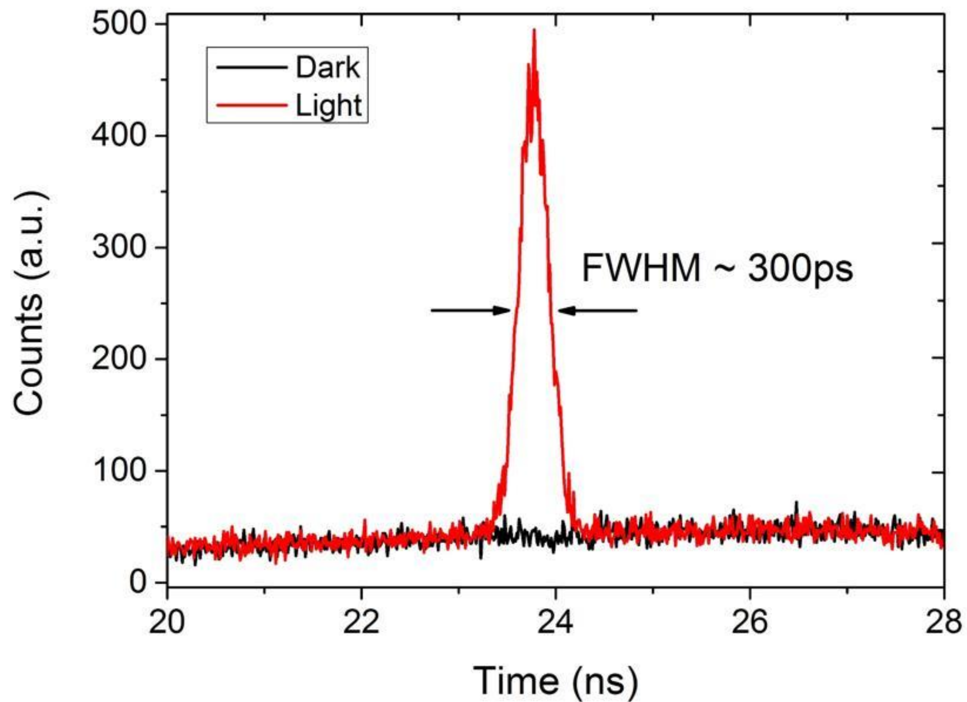


Figure 5.26. Time response histograms, dark (black) and light (red), at 10 % relative excess bias measured in a time-correlated single photon counting setup (~50 ns gate width, repetition rate of 10 kHz) on a 25 μm diameter device from wafer 11-141 at a wavelength of 1310 nm (<0.1 ppp), and a temperature of 100 K.

The use of histograms for the characterisation provided some useful information on afterpulsing, as well as jitter. If the background levels are the same for both dark and light measurements (as in Figure 5.26), this indicates that the detector is operating in a regime with negligible after-pulsing.

Figure 5.27 illustrated the measured jitter at FMHM on a 25 μm diameter device at different relative excess biases and a temperature of 100 K.

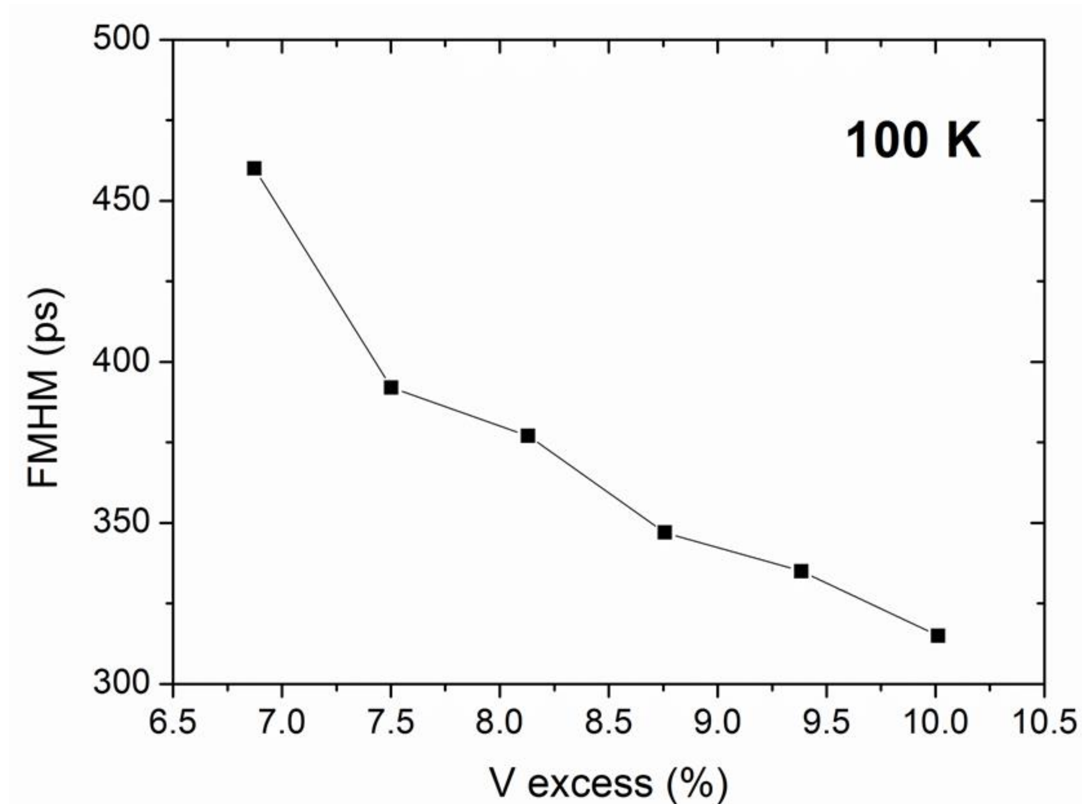
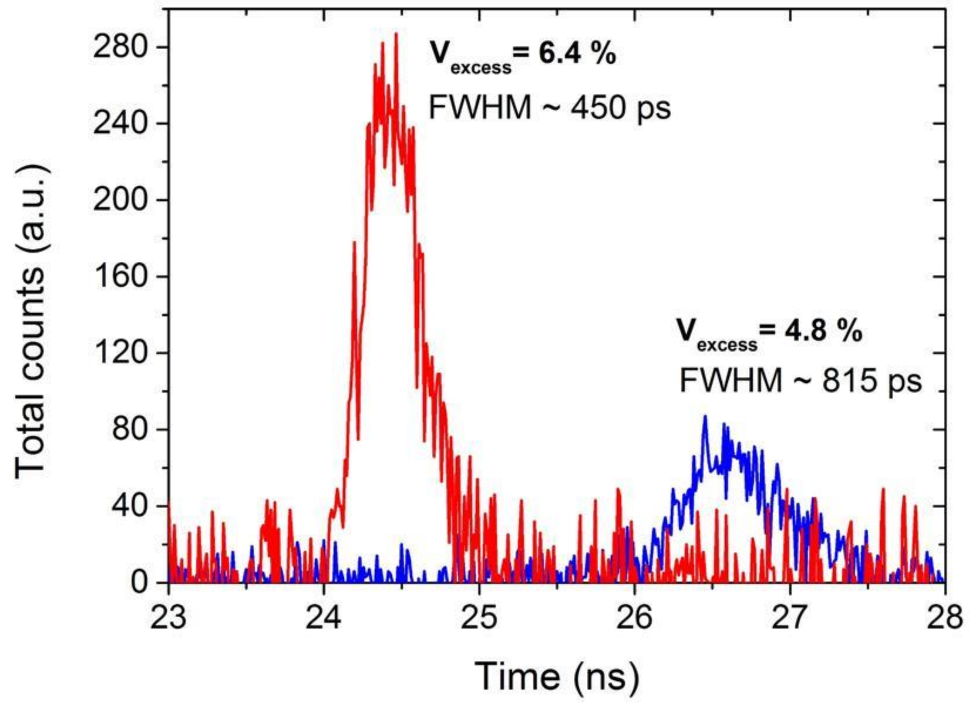


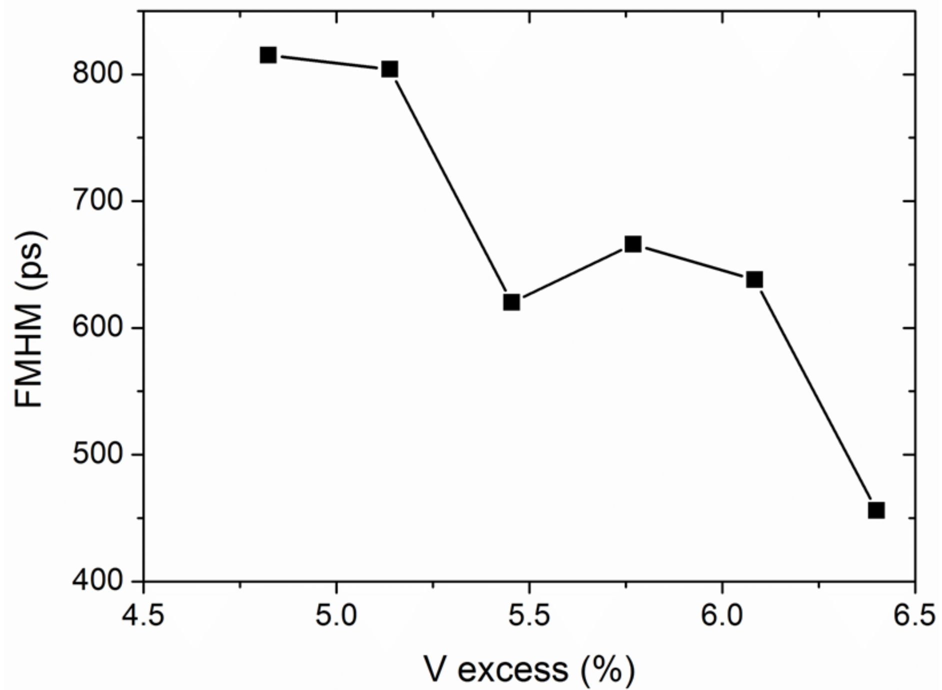
Figure 5.27. Time jitter at FMHW measured at different relative excess biases on a 25 μm diameter device from wafer 11-141 at a wavelength of 1310 nm, and a temperature of 100 K.

The timing jitter of the detector decreased as the excess bias increased, as expected. The drawback is that the DCR also increases. The lowest value of ~300 ps was measured at 10 % relative excess bias. With the same experimental setup, timing jitters of less than 80 ps were measured with alternative low-jitter all-Si SPAD detectors, and hence it is reasonable to assume that the overall measured jitter is dominated by the detector contribution.

At the same temperature (100 K) and wavelength (1310 nm) the 50 μm diameter device demonstrated a response with a FMHM in the range of 450 – 815 ps. This is shown in Figure 5.28a, where the time response histograms obtained by subtracting the dark from the light counts (total numbers of counts) are reported for the minimum and maximum relative excess bias (4.8 % (blue) and 6.4 % (red), respectively). Conversely, Figure 5.28b illustrates the FMHM as a function of the relative excess bias.



(a)



(b)

Figure 5.28. (a) Time response histograms of a 50 μm diameter device from wafer 11-141 obtained by subtracting the dark from light counts (total counts) for two different relative excess bias, 4.8 % (blue) and 6.4 % (red), respectively. (b) FMHM as a function of the relative excess bias. Timing jitter measurements were performed at T=100 K and λ=1310 nm (~50 ns gate width, repetition rate of 10 kHz).

The measured timing jitter values showed good agreement with the timing jitter measured previously (~100 – 350 ps, λ = 1310 nm, laser pulse width = 40 ps) for homojunction Ge APDs operated in Geiger mode [10].

At the longer wavelength of 1550 nm, the minimum time response of ~420 ps was measured at 6 % of relative excess bias on a 25 μm diameter device at a temperature of 125 K, as shown in Figure 5.29.

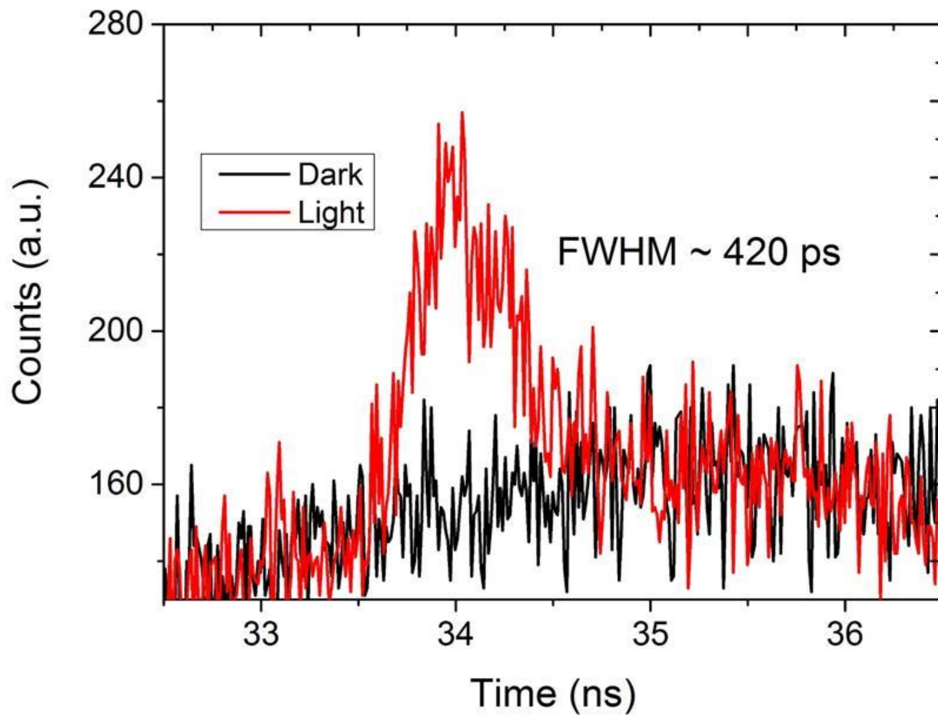


Figure 5.29. Time response histograms, dark (black) and light (red), at 6 % of relative excess bias measured in a time-correlated single photon counting setup (~50 ns gate width, repetition rate of 10 kHz) on a 25 μm diameter device from wafer 11-141 at a wavelength of 1550 nm (<0.1 photon per pulse), and a temperature of 125 K.

5.5.7 DCR versus frequency (afterpulsing)

Although the main contribution to the DCR is given by thermally generated carriers, this is not the only mechanism responsible for DCR. There is another effect known as afterpulsing which is essentially a memory effect due to the large amount of carriers flowing through the device during the avalanche process associated with any individual photon event or a dark count. These carriers can be trapped and then released in the following gate periods, hence contributing to the overall DCR. This contribution is only negligible at very low frequencies of detector operation, when there is enough time between two gate pulses to release all the trapped carriers. Since these SPADs are

operated in gated mode, the probability of observing a dark count within a certain window increases as the repetition rate is increased (the frequency at which the AC gate is superimposed on the SPAD to take the device beyond breakdown and into Geiger mode).

Figure 5.30 illustrates the impact of the repetition rate on the DCR on a 25 μm diameter Ge-on-Si SPAD at two different temperatures, 150 K (black) and 100 K (red), respectively.

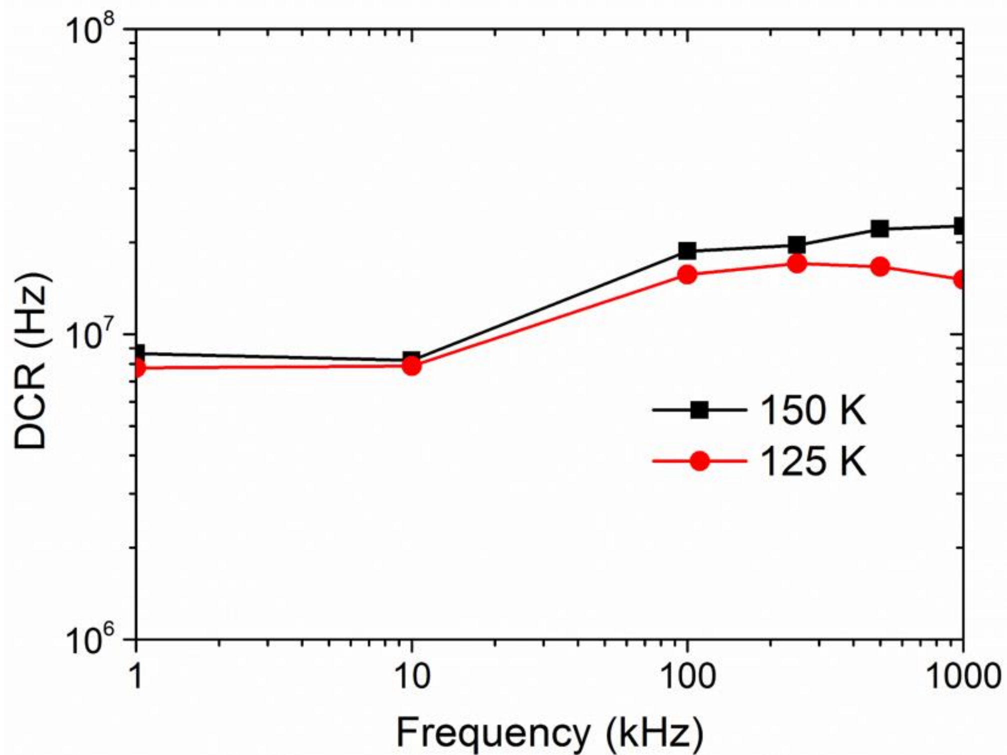


Figure 5.30. DCR versus gating frequency for a 25 μm diameter Ge-on-Si SPAD at two temperatures, 150 K (black) and 100 K (red), respectively.

A slight increase (by nearly a factor of 2) in the DCR was observed when increasing the gating frequency from 1 kHz to 1 MHz. Similar behaviour was also observed using two different Ge homojunction APDs operated in Geiger mode [10].

Although InGaAs/InP SPADs demonstrated a lower DCR at this temperature, such devices demonstrated a rapid increase in the DCR due to afterpulsing, as already discussed in Chapter 3. This was the main problem affecting InGaAs/InP SPADs, considering that it significantly reduced the frequency of operation of the devices below 100 kHz. In these devices, the traps contributing to afterpulsing are located within the high-field InP multiplication region [15]. Buller *et al.* demonstrated experimentally that

only traps within the high field region are of concern, since traps within the low field InGaAs layer do not initiate an after-pulse [16].

For the Ge-on-Si SPADs the avalanche is initiated from electrons, and if a photon is absorbed in the Ge, the photogenerated hole will then drift to the top contact adjacent to the Ge layers. When the photogenerated electrons pass into the multiplication region, both electrons and holes will undergo impact ionisation. The holes will drift out of the Si layers towards the Ge absorber and to the top contact, while the electrons will drift to the Si substrate contact. If holes are trapped within the Ge layer and are later released, they cannot contribute to after-pulsing since they drift directly to the top contact and do not enter into the multiplication region. Hence, electrons and/or holes trapped in the Si multiplication region are the only carriers that can probably initiate an after-pulse.

Generally, the after-pulsing within all-Si SPADs is very low at higher temperatures [17] and much lower than exhibited in all-Ge SPADs [18]. Therefore, by using Si for the multiplication region, the Ge-on-Si SPAD might prove to be a superior device in terms of afterpulsing than alternative SPADs operating at this wavelength band. In future, a more comprehensive analysis of afterpulsing should be performed by using time-correlated carrier counting methods in order to better understand its impact on these structures.

5.5.8 Ge-on-Si SPADs in the scientific literature

In Chapter 3 different Ge-on-Si APD structures (section 3.6.4) have been described but none of them were characterised in Geiger mode. Very few reports on Geiger mode characterisation of Ge-on-Si SPAD devices have been reported in the scientific literature. In the work of Lu *et al.*, a 30 μm diameter Ge-on-Si SPAD was characterised in Geiger mode. The wafer layer microstructure was similar to that used in this work with the exception of the thickness of the Si multiplication layer which was 0.5 μm thick. An SPDE of up to 14 % was reported at a wavelength of 1310 nm. However, the SPDE was measured with an incident photon flux of 1 photon per pulse, thus giving a high probability of multi-photon pulses incident on the device (as shown in section 5.5.1), potentially leading to an overestimation of SPDE.

In addition, these devices demonstrated high DCR (from 100 MHz to 500 MHz at different excess bias), meaning it is likely that the device had insufficient time to recharge before another dark count was triggered, resulting in an underestimate of the

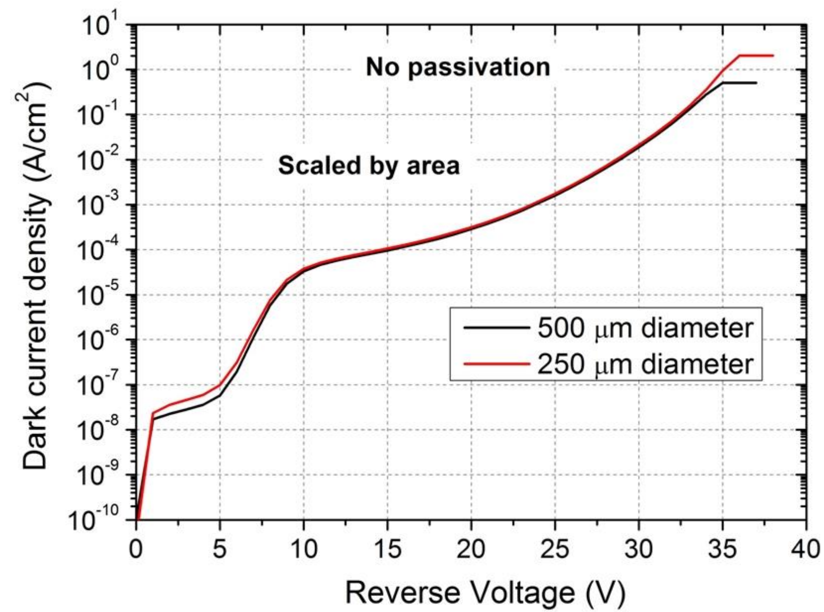
DCR for a given excess bias. Further evidence of this recharge issue was observed as the gating frequency was increased from 1 kHz to 1 MHz where there was a notable decrease in the DCR- exactly opposite to that expected with afterpulsing (as shown in the previous section), but entirely consistent with lack of voltage recovery.

Another report of a Ge-on-Si device claiming single-photon sensitivity was published by Aminian *et al.* [19], although the detection efficiency in the Geiger mode was measured only through the analysis of the photocurrent (at a wavelength of only 1100nm) above breakdown, and this cannot be regarded as a robust single-photon counting characterization method.

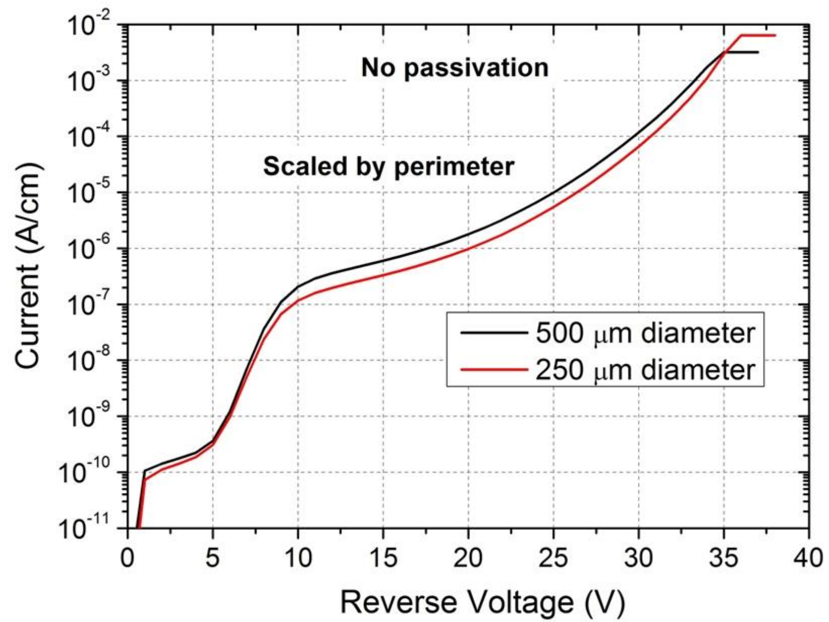
5.6 Dark current analysis

Ge-on-Si wafers from the 1st generation (wafer 11-167 and 11-141) were also fabricated by our collaborators at the McMaster University. In particular, two different fabrications were performed: the first consisted of large active area devices of 500 μm and 250 μm diameter, which were simply etched, and Al top and bottom contacts were then deposited without passivating the mesa sidewalls; in the second, different size devices ranging from 20 μm to 100 μm diameter were fabricated and SiO_2 was used for the mesa sidewall passivation.

Investigating how the dark current scales with different mesa diameters was a first step used to understand the main mechanisms contributing to the dark current. Figure 5.31a-b shows the dark current of large active area devices (with no passivation) of wafer 11-141 scaled by both perimeter and area of the devices. The same analysis was performed on both small active area devices with SiO_2 passivation from the McMaster University devices (Figure 5.32a-b) and also devices with Si_3N_4 passivation fabricated at the University of Glasgow (Figure 5.33a-b). All measurements were performed at room temperature.

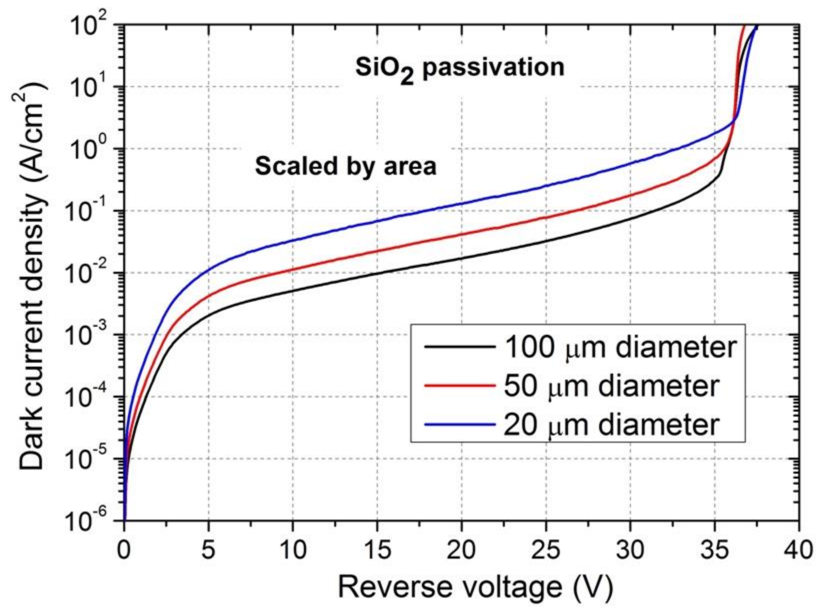


(a)

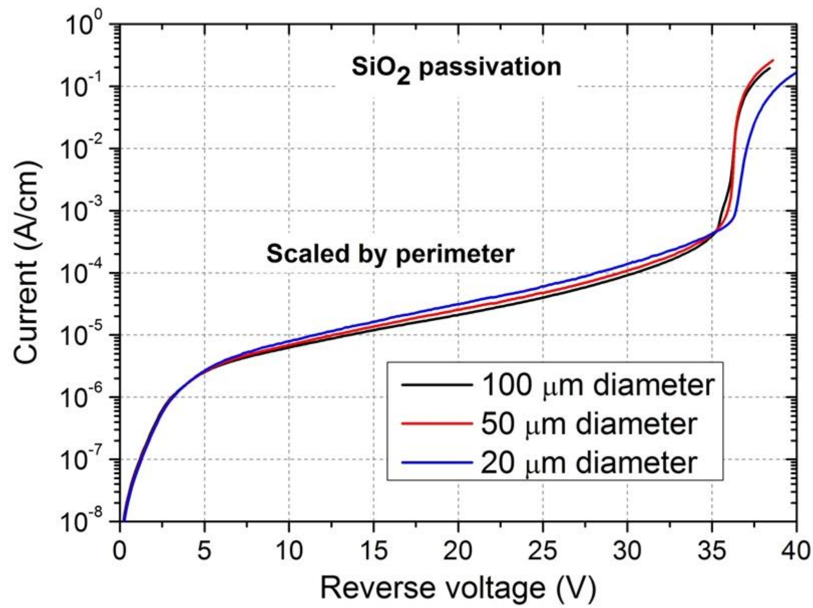


(b)

Figure 5.31. I-V characteristic of large active area devices (with no passivation), 500 μm (black) and 250 μm (red) diameter, fabricated at McMaster University scaled by (a) area and (b) perimeter.

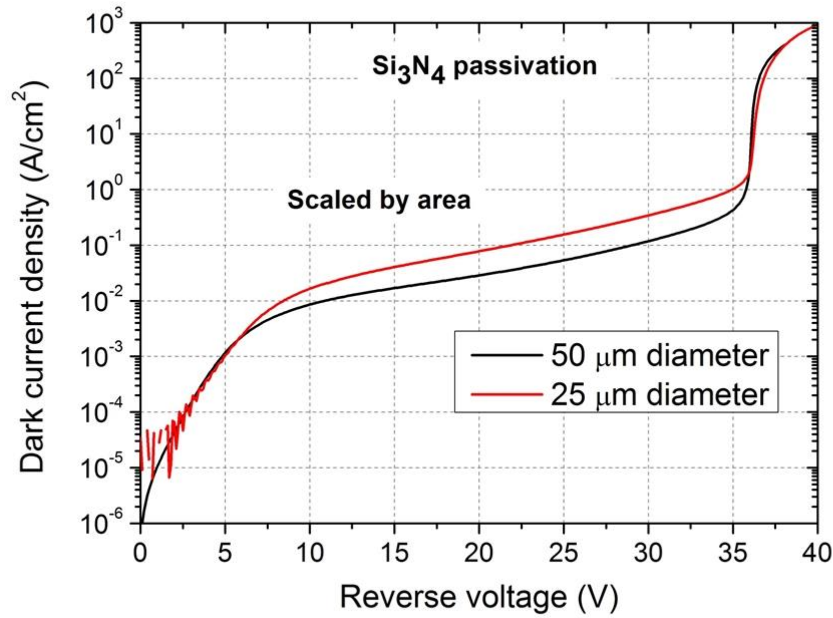


(a)

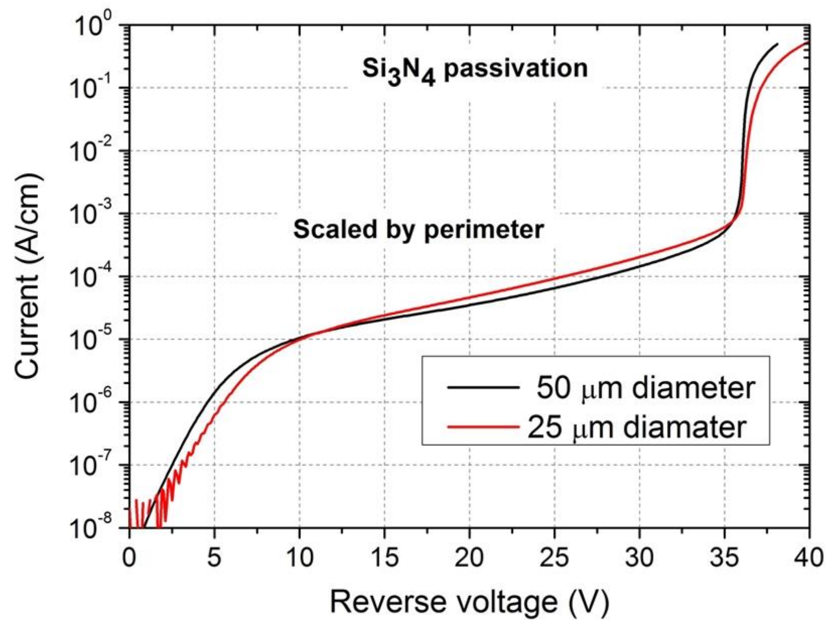


(b)

Figure 5.32. I-V characteristic of small active area devices (with SiO_2 passivation), 100 μm (black), 50 μm (red), and 20 μm (blue) diameter, fabricated at McMaster University scaled by (a) area and (b) perimeter.



(a)



(b)

Figure 5.33. I-V characteristic of small active area devices (with Si_3N_4 passivation), 50 μm (black), and 25 μm (red) diameter, fabricated at University of Glasgow scaled by (a) area and (b) perimeter.

Large active area devices (with no passivation) scaled by area (Figure 5.31a), and hence this result suggested that a bulk effect is dominating the dark current, while small active area devices, with either SiO_2 or Si_3N_4 passivation, scaled by perimeter (Figures 5.32b and 5.33b) suggesting that a surface effect related to the perimeter might be the main mechanism.

To further understand the perimeter and area dependencies of the dark current, we can express the dark current as:

$$I_D = J_P \times Perimeter + J_A \times Area \quad (5.8)$$

Where J_P is the current density due to the perimeter of the device and J_A is the current density due to the area of the device. Dividing equation 5.8 by the area of the device:

$$J_D = J_P \times \frac{Perimeter}{Area} + J_A \quad (5.9)$$

Therefore, the slope of the line corresponds to the perimeter component of the leakage current and the y-intercept corresponds to the area component when the dark current is plotted against the perimeter/area ratio for a range of different size devices. By performing this analysis at different reverse voltages, the perimeter component of the dark current for the three different processes was extracted as shown in Figure 5.34. Regarding to the area component of the dark current, it was found that J_A was negative for the small active area devices (fabricated from both McMaster and Glasgow Universities) indicating perhaps that the dark current did not fit this model very well and another mechanism dominated the dark current.

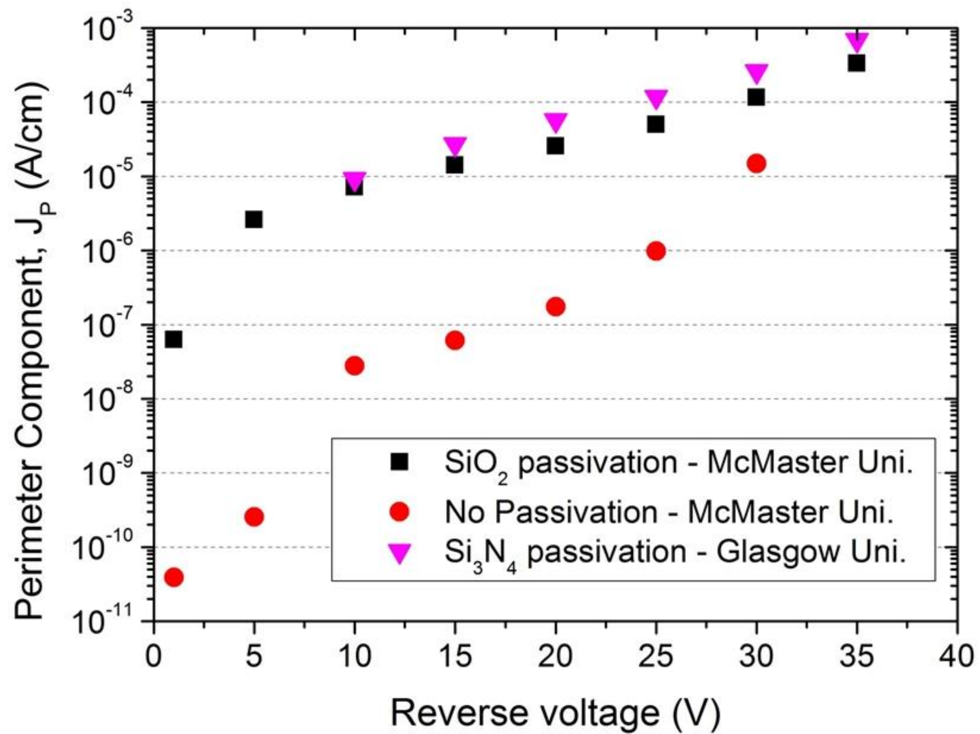


Figure 5.34. Perimeter component of the dark current versus the reverse voltage calculated for different sizes devices: large active area device without passivation fabricated at McMaster University (red circle), small active area device with SiO_2 passivation fabricated at McMaster University, and small active area device with Si_3N_4 passivation fabricated at University of Glasgow.

The perimeter component has an exponential dependence on the applied reverse voltage. Large active area devices (red circle) showed the lowest perimeter component since these devices have small perimeter/area ratios and are dominated by effects related to the area. Small active area devices (which have a large perimeter/area ratios) fabricated by both McMaster (black square) and Glasgow (magenta triangle) Universities showed a similar perimeter component, even if two different materials were used for the sidewalls passivation, SiO_2 and Si_3N_4 respectively. Although this analysis give an indication of the main mechanisms contributing to the leakage current of different sizes device, further research would be needed to understand how different passivation materials can help to reduce the perimeter component, and hence the dark current, of small sized devices which are the preferable candidates for single-photon detectors. This analysis was performed only on devices fabricated from wafer 11-141, since there were not enough data to perform the same analysis on wafer 11-167. In future, it will be interesting to evaluate how the perimeter component of small sized devices is related to different electric field profiles through the Ge-on-Si structures.

5.7 Future work

Although the characterised Ge-on-Si SPADs reported in this work demonstrated the lowest NEP of Ge-on-Si SPADs as reported in the scientific literature, different problems were found, such as:

a) Lack of reproducibility and reliability of I-V characteristic;
b) The requirement for low operational temperatures;
c) Poor single photon performance (high DCR and low SPDE).

For these reasons, future development of these devices should consider these aspects. In terms of I-V characteristics, devices showed a low reproducibility, since the electrical behaviour of the devices changed when the I-V characteristic was swept more than one time. Although, this problem could be related either to the quality of the material itself or to the fabrication process, different device fabrication processes demonstrated different levels of reproducibility. The same material (1st generation of Ge-on-Si SPAD) processed by various collaborators such as the McMaster University and the University of Glasgow, showed different behaviour. Large active area devices (with no passivation) processed at the McMaster University showed a good reproducibility in terms of I-V characteristics with almost 100 % yield. An attempt has been made to characterise these devices in terms of single-photon performance, but the dark count rate was too high (due to the large active area) in order to undertake reliable measurements.

A second batch from the same material was then processed, and small active area devices ranging from 20 to 100 μm diameter (with SiO_2 passivation) were fabricated. Although, these devices showed 100 % yield at room temperature, they were irreversible damaged when the temperature was decreased to 200 K. Furthermore, devices processed at the University of Glasgow demonstrated good reproducibility in terms of I-V characteristics, and they were characterised in terms of their single-photon performance, as shown in previous sections. However, large active area devices (100 μm and 200 μm diameter) showed very high dark current at low reverse bias voltages (-5 V) which made them unusable, while small active area devices (25 μm and 50 μm diameter) demonstrated a consistent behaviour even at the lowest temperatures (as shown in the previous sections). Therefore, different device processing had different impact on the electrical behaviour of these detectors, and hence a robust fabrication process is required for future generations of devices.

In terms of low temperature operation, devices needed to be cooled down to reduce the leakage current, as expected. The small active area devices (25 μm diameter) demonstrated the lowest dark current, and analysis suggested that the main contributions to the leakage current are related to the surface, tunnelling and field-assisted emission. In terms of surface leakage component, our future work aims at investigating different passivation techniques for mesa sidewall surface states. For example, recent research work investigated the impact of a GeO_2 surface passivation layer on the dark current of all-Ge p-n diodes, and pointed out a reduction in terms of surface leakage current [20]. However, a better understanding would be needed to evaluate if the main surface contribution to the dark current is related to the Ge absorber layer or the Si multiplication layer corresponding also to the high field region. Future developments should consider different etch steps of the mesa in order to evaluate different contributions to the leakage current. In order to examine the tunnelling component of the dark current, it will be important to evaluate different electric field profile through the structures. The doping concentration of the charge sheet layer plays an important role, and hence it will be necessary to obtain a good dopant growth reproducibility, as this represents one of the limiting factors during heteroepitaxial growth (as shown in Chapter 4).
In terms of single-photon performance, the 1st generation of Ge-on-Si SPADs demonstrated a high DCR and low SPDE at low temperatures of operation. In terms of DCR, it will be important to analyse different size devices to understand the main contributions to the DCR. Leakage current reduction could be used to improve the DCR of these devices as well as the investigation of a planar geometry.

To further understand the contribution of threading dislocations on the DCR, it will be crucial to fabricate these devices using selective area growth (section 3.4.1.2 of Chapter 3). This technique has already been demonstrated by different research groups involved in the fabrication of p-n or p-i-n Ge-on-Si detectors showing that a lower density of defects helps to improve the performance of these devices [21], [22], but its impact on the performance of a Ge-on-Si SPAD has not been reported in the scientific literature.

Regarding the low SPDE measured at a wavelength of 1310 nm, we believe that the Si multiplication layer played an important role. In Chapter 4, SIMS measurements on the 2nd generation of Ge-on-Si devices (section 4.8 of Chapter 4) showed a pronounced phosphorous diffusion tail from the Si substrate in the Si multiplication layer. This effect led to a thinner multiplication layer and a non-uniform electric field across this

layer. These problems might have had the effect of reducing the avalanche triggering probability, and hence reducing the SPDE. As shown in Chapter 4, the optimisation of the doping in the Si multiplication layer has already been achieved by using Sb doped silicon substrate, and this should therefore useful to increase the SPDE. Additionally, higher temperatures of operation combined with a reduced dark current, could help improve the SPDE at the longer wavelength of 1550 nm. Finally, a waveguide integrated Ge-on-Si SPAD design is under investigation, since it could give the opportunity to improve the device efficiency and integrate these devices with other Si photonics components.

5.8 Conclusion

In this Chapter, the growth, fabrication and characterisation of Ge-on-Si SPADs have been described. Current-voltage characteristics of Ge-on-Si devices fabricated from wafer 11-167 and 11-141 have been reported and analysed at different temperatures. Devices from wafer 11-141 were characterised in terms of their single-photon performance. The characterisation was performed at temperatures ranging from 100 K to 150 K, since the dark current of these devices was too high to make credible measurements at higher temperatures. The best single-photon performances were achieved on a 25 μm diameter device at a temperature of 100 K. At this temperature, a DCR of 1 – 6 Mcs^{-1} and a SPDE of 2.2 – 4 % was measured at different excess biases. Although these values were in line with the values of DCR and SPDE measured for commercially available all-Ge APDs, there still exists an important performance gap with InGaAs/InP SPADs. Nonetheless, the lowest NEP of Ge-on-Si SPADs reported in the scientific literature was illustrated.

A potential major advantage of these devices compared to the InGaAs/InP SPADs could be that of reduced afterpulsing. Only a slight increase (by nearly a factor of 2) in the DCR was observed when increasing the gating frequency from 1 kHz to 1 MHz. Further detailed analysis will be carried out for the next generations of devices by using a time-correlated carrier counting method in order to better understand its impact on these structures. Finally, I-V analysis of devices fabricated using different processing techniques suggested that the main current component of the dark current of small active area devices is related to a surface effect. Therefore, the next immediate step in the fabrication of these devices will be the study and evaluation of different passivation techniques to reduce the leakage current and DCR. New detectors will be also fabricated

with an optimised doping concentration of the Si multiplication layer and better reproducibility of the charge sheet doping concentration. While the first element should help improve the SPDE, the second will help in view of studying the single-photon properties of these devices at different electric field profile across the structures and hence understanding the impact of high field effects (i.e. tunnelling) on device performance. SEG could also give a further opportunity to evaluate the impact of threading dislocations on the single-photon performance, since no research has been reported in the scientific literature to date.

5.9 Reference

[1] V. A. Shah, A. Dobbie, M. Myronov, and D. R. Leadley, "Effect of layer thickness on structural quality of Ge epilayers grown directly on Si(001)," *Thin Solid Films*, vol. 519, no. 22, pp. 7911–7917, 2011.

[2] M. M. Mirza, H. Zhou, P. Velha, X. Li, K. E. Docherty, A. Samarelli, G. Ternent, and D. J. Paul, "Nanofabrication of high aspect ratio (50:1) sub-10 nm silicon nanowires using inductively coupled plasma etching," *J. Vac. Sci. Technol. B Microelectron. Nanometer Struct.*, vol. 30, no. 6, p. 06FF02, 2012.

[3] K. Gallacher, P. Velha, D. J. Paul, I. MacLaren, M. Myronov, and D. R. Leadley, "Ohmic contacts to n-type germanium with low specific contact resistivity," *Appl. Phys. Lett.*, vol. 100, no. 2, pp. 022113–022113–3, 2012.

[4] J. P. Gambino and E. G. Colgan, "Silicides and ohmic contacts," *Mater. Chem. Phys.*, vol. 52, no. 2, pp. 99–146, 1998.

[5] S. Pellegrini, "InGaAs/InP single-photon avalanche diodes," Ph.D., Heriot-Watt University, 2005.

[6] M. A. Itzler, r. Ben-Michael, C.-F. Hsu, K. Slomkowski, A. Tosi, S. Cova, F. Zappa, and R. Ispasoiu, "Single photon avalanche diodes (SPADs) for 1.5 μm photon counting applications," *J. Mod. Opt.*, vol. 54, no. 2–3, pp. 283–304, 2007.

[7] W. Becker, *Advanced Time-Correlated Single Photon Counting Techniques*, vol. 81. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005.

[8] R. E. Warburton, "Infrared Time-Correlated Single-Photon Counting," Ph.D., Heriot-Watt University, 2008.

[9] S. Pellegrini, R. E. Warburton, L. J. . Tan, Jo Shien Ng, A. B. Krysa, K. Groom, J. P. . David, S. Cova, M. J. Robertson, and G. S. Buller, "Design and performance of an InGaAs-InP single-photon avalanche diode detector," *IEEE J. Quantum Electron.*, vol. 42, no. 4, pp. 397– 403, 2006.

[10] A. Tosi, A. Dalla Mora, F. Zappa, and S. Cova, "Germanium and InGaAs/InP SPADs for single-photon detection in the near-infrared," p. 67710P–67710P, 2007.

[11] N. A. DiLello, D. K. Johnstone, and J. L. Hoyt, "Characterization of dark current in Ge-on-Si photodiodes," *J. Appl. Phys.*, vol. 112, no. 5, pp. 054506–054506–8, 2012.

[12] W. C. Dash and R. Newman, "Intrinsic Optical Absorption in Single-Crystal Germanium and Silicon at 77°K and 300°K," *Phys. Rev.*, vol. 99, no. 4, pp. 1151–1155, 1955.

[13] L. Colace, G. Masini, G. Assanto, H.-C. Luan, K. Wada, and L. C. Kimerling, "Efficient high-speed near-infrared Ge photodetectors integrated on Si substrates," *Appl. Phys. Lett.*, vol. 76, no. 10, pp. 1231–1233, 2000.

[14] G. S. Buller, S. J. Fancey, J. S. Massa, A. C. Walker, S. Cova, and A. Lacaita, "Time-resolved photoluminescence measurements of InGaAs/ InP multiple-quantum-well structures at 1.3μm wavelengths by use of germanium single-photon avalanche photodiodes," *Appl. Opt.*, vol. 35, no. 6, pp. 916–921, 1996.
[15] S. Cova, A. Tosi, A. Gulinatti, F. Zappa, and M. Ghioni, "Avalanche diodes and circuits for infrared photon counting and timing: retrospect and prospect," *IEEE LEOS Newsletter*, vol. 20, no. 25, pp. 25–28, 2006.
[16] G. S. Buller, R. E. Warburton, S. Pellegrini, J. S. Ng, J. P. R. David, L. J. J. Tan, A. B. Krysa, and S. Cova, "Single-photon avalanche diode detectors for quantum key distribution," *IET Optoelectron.*, vol. 1, no. 6, pp. 249–254, 2007.
[17] S. Cova, M. Ghioni, A. Lotito, I. Rech, and F. Zappa, "Evolution and prospects for single-photon avalanche diodes and quenching circuits," *J. Mod. Opt.*, vol. 51, no. 9–10, pp. 1267–1288, 2004.
[18] A. Lacaita, P. A. Francese, F. Zappa, and S. Cova, "Single-photon detection beyond 1 mm: performance of commercially available germanium photodiodes," *Appl. Opt.*, vol. 33, no. 30, pp. 6902–6918, 1994.
[19] M. Aminian, A. Sammak, L. Qi, L. K. Nanver, and E. Charbon, "A Ge-on-Si single-photon avalanche diode operating in Geiger mode at infrared wavelengths," p. 83750Q–83750Q–10, 2012.
[20] M. Takenaka, K. Morii, M. Sugiyama, Y. Nakano, and S. Takagi, "Dark current reduction of Ge photodetector by GeO2 surface passivation and gas-phase doping," *Opt. Express*, vol. 20, no. 8, pp. 8718–8725, 2012.
[21] S. Park, T. Tsuchizawa, T. Watanabe, H. Shinojima, H. Nishi, K. Yamada, Y. Ishikawa, K. Wada, and S. Itabashi, "Monolithic integration and synchronous operation of germanium photodetectors and silicon variable optical attenuators," *Opt. Express*, vol. 18, no. 8, p. 8412, 2010.
[22] S. A. Claussen, K. C. Balram, E. T. Fei, T. I. Kamins, J. S. Harris, and D. A. B. Miller, "Selective area growth of germanium and germanium/silicon-germanium quantum wells in silicon waveguides for on-chip optical interconnect applications," *Opt. Mater. Express*, vol. 2, no. 10, pp. 1336–1342, 2012.

Chapter 6 - Fill-factor recovery of Si CMOS SPAD array by using microlens integration: improvement factor and spatial uniformity characterisation

6.1 Introduction

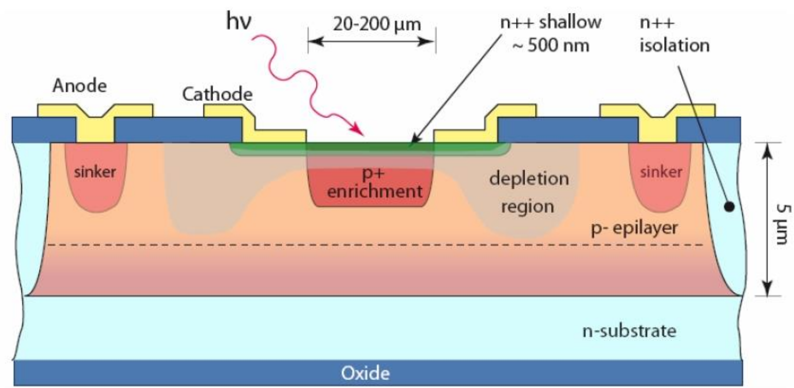
The detection of single photons in the visible and near-infrared (NIR) spectral range below wavelengths of 1 μm is mostly covered by low-noise Si SPADs which now represent a commercially available technology. Si SPAD detectors can be divided in two groups: custom-processed and CMOS compatible. The former devices are based on tailored processing with the aim of obtaining the best performance from the photodiode when operated in Geiger-mode. The latter devices rely on standard CMOS fabrication processes from the microelectronic industry which introduces some advantages, such as the routine on-chip integration with the electronics required for photon counting and photon timing techniques. This efficient integration allows the possibility of fabrication of two-dimensional SPAD-based focal plane arrays. However, CMOS Si SPADs typically have a detection performance which is not as good as optimised, custom processed Si SPADs. In addition, generally the in-pixel circuitry takes up a great deal of space in a CMOS SPAD array, meaning the detector fill-factor is necessarily low in CMOS SPAD arrays [1], [2], thus reducing the effective detection efficiency.

Different solutions have been proposed and attempted in order to recover the loss of fill-factor, these include the use of microlens arrays and 3D integration of the circuitry and detector plane. The former solution was used in this work and different microlens arrays were integrated on top of a 32 × 32 Si CMOS SPAD array [3]. Microlens integration has been demonstrated by other research groups [4]–[6], but a full characterization of Si CMOS SPAD array integrating microlenses in the spectral range of interest (500-900 nm) and at various f-numbers has not been reported. The research covered in this chapter aims to fill this gap in the scientific literature and suggest ways forward to improve detector fill-factor. Two main parameters, the improvement factor given by microlens array integration and the array uniformity, are presented, analyzed and discussed in the following sections.

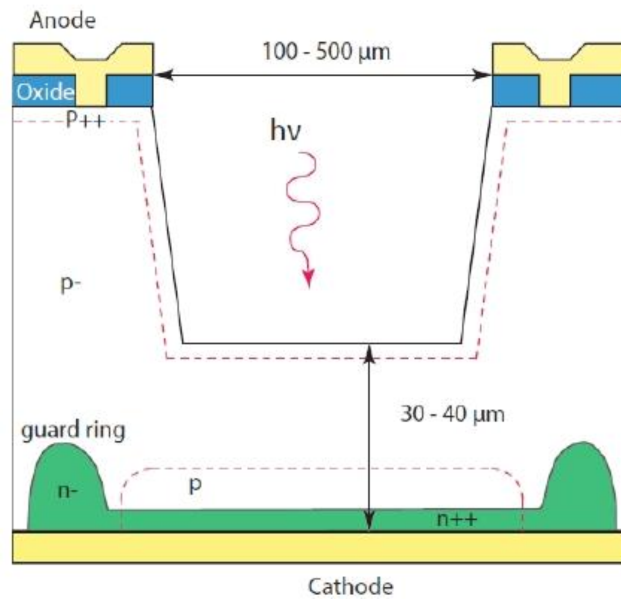
6.2 Si SPAD: custom processing and CMOS fabrication

When Si SPADs are fabricated by using custom processing, it is possible to define two main groups according to the thickness of the depletion layer of the p-n junction [7]:

1. Thin junction SPADs, with depletion layers typically 1-5 μm thick (Figure 6.1a). These devices demonstrate fairly good single-photon detection efficiency in the visible range, (about 45% at 500 nm wavelength). This efficiency reduces to 32% at 630 nm wavelength and 15% at 730 nm, whilst in the NIR it is about 10% at 830 nm and a few 0.1% at 1064 nm. Thin Si SPADs require low breakdown voltages of 20 – 50V, and small active area devices, with diameters ranging from 20 μm to 200 μm, can be fabricated [8]–[10];
2. Thick junction SPADs have depletion layer thicknesses varying between 20 μm to 150 μm (Figure 6.1b). These devices operate at high breakdown voltages, between 200 – 500 V, and have fairly wide active areas, with diameters ranging from 100 μm to 500 μm. The single-photon detection efficiency is very high in the visible range due to the thicker depletion region, and is higher than 50% over all the range of 540 – 850 nm, and reduces in the NIR, to around 3% single-photon detection efficiency at a wavelength of 1064 nm [11], [12].



(a)



(b)

Figure 6.1.(a) Cross section of a thin junction Si SPAD [13], *(b) and a thick junction Si SPAD* [14].

Thick Si SPADs show high dissipation power (up to 10W) because of the high reverse bias required for their operation and the large avalanche current (tens of mA). Generally, the devices have relatively poor timing resolution with a jitter FWHM of several hundreds of picoseconds. Furthermore, these devices are fabricated with custom fabrication technologies which are not compatible with arrayed device technology.

On the other hand, thin junction Si SPADs have correspondingly lower breakdown voltages and exhibit very low power consumption, and a very good timing resolution of less than 50 ps FWHM. They can be fabricated by using planar silicon technology which makes these devices particularly suitable for monolithic integration of arrays. Furthermore, the same design could be adapted to be fabricated using CMOS

technology, leading to several advantages such as increased levels of miniaturization, and lower cost, power and size per pixel compared to a custom technology.

From the early 2000s, several research groups have explored the design of a monolithically integrated single-photon imaging system in high-voltage (HV) and standard deep-sub-micron (DSM) CMOS technologies [15]–[19]. However, CMOS SPADs generally show drawbacks in terms of detector performance compared to devices made from custom fabrication technology. This is due to the device designers being unable to modify the CMOS fabrication processes which have to face all the problems related to the CMOS SPAD fabrication such as premature edge breakdown, tunnelling effect, electric field non uniformity and junction depth.

As reported in references [13], and [14], CMOS SPAD performance has improved as the technology scales, whilst introduces the benefit of lower excess bias voltages (without impairing the SPDE) and the miniaturisation of the SPAD and electronics. These two last conditions help in reducing the DCR, afterpulsing probability of the fabricated device, and improving the timing performances, since the avalanche build-up times have a narrower statistical spread on smaller devices. In particular, Si SPADs fabricated in 0.35 μm HV-CMOS technology, as reported in [21], demonstrated performance in terms of DCR (which was only 120 counts/s for a 30 μm diameter SPAD at room temperature) and afterpulsing comparable to those of the best in class custom Si SPADs (for details refer to Table 2 in reference [21]).

Although technology scaling guarantees some advantages in terms of DCR, afterpulsing and timing resolution, another crucial parameter that must be considered is the detection efficiency of a Si CMOS SPAD array, as explained in the following section.

6.3 Single-photon Detection Efficiency and Fill-Factor recovery

The SPDE of a SPAD is given by the product of the absorption efficiency and the trigger probability, as pointed out in Chapter 3. The SPDE of a SPAD fabricated in standard CMOS process can also affected by technological parameters which play a significant role. For example, the incident photons must pass through a thick layer of passivation placed on top of the chip during the final step of the fabrication to protect the device from external contamination. The light must also pass through several dielectric layers with different refractive indices. While these processes can be optimised in a custom CMOS imaging technology, standard CMOS technologies do not readily offer such options [20].

In addition, when in-pixel circuitry is included as in a CMOS Si SPAD array, we can refer to a pixel as a *smart* pixel [22]. Therefore, another parameter must be considered to evaluate the SPDE of a *smart* pixel: this is the geometric fill-factor which is the ratio of photo-sensitive area to total pixel area, usually expressed as a percentage. When a CMOS SPAD array is illuminated at a specific wavelength, not all the optical power that is incident on the sensor strikes the photosensitive area of the detector due the large proportion of the surrounding area being dedicated to the electronics surrounding the Si SPAD. Therefore a reduction in fill-factor causes a deterioration of the SPDE, since a larger fraction of incident photons are not detected.

Although small active area SPADs demonstrate lower noise performance, better yield (since there is less probability of having a defect in the active area), and lower power consumption, these devices also give a lower fill-factor compared to the values obtained for a large active area SPAD. For example, in reference [1] a SPAD array fabricated in 0.35 μm CMOS technology achieved a very high resolution of 10ps but a very low fill-factor of 0.5%. In reference [23], a Si SPAD array fabricated in 130nm CMOS technology with 119ps time resolution and 2.3% fill-factor was reported. The same CMOS technology was also used in [2] to fabricate a Si SPAD array with a time resolution of 55ps and fill-factor of 1%. In reference [17], a 128 x 128 SPAD array was fabricated in 0.35 μm CMOS technology demonstrating a fill-factor of around 6% by using a 7 μm diameter Si SPAD and a multiplexed architecture. The fill-factor was further increased in reference [15] to 20% and 35% by using a larger active area SPAD of 50 μm and 100μm respectively.

Different solutions have been proposed and attempted to recover the loss of sensitivity due to the low fill-factor of CMOS SPAD array, such as the use of a 3-D integration technology or an array of micro-optical concentrators. In the first approach, the in-pixel electronics no longer surround the detector, but are placed on a separate silicon wafer which is either wafer-to-wafer bonded to another wafer containing an array of SPADs or is connected to the SPAD chip using through-silicon-vias (TSVs) [24], [25]. In Ref. [26], the Si SPAD array was fabricated in the silicon-on-insulator (SOI) wafer which was flip-chip bonded to the electronics wafer and then etched so that only the high-quality SOI film on which the SPAD pixels are fabricated, could be back-side illuminated through the buried oxide layer. These techniques have also been mainly demonstrated for InGaAsP SPAD arrays (not compatible with the CMOS technology)

which are flip-chip bonded to the CMOS readout integrated circuit, as shown in Ref. [27].

In the second approach, each pixel has its own micro-optical concentrator which collects light from the objective plane in the focal plane and directs it to the sensitive area of each pixel in the detector plane, and hence enhancing the fill factor of a SPAD array. The downside of this approach is the increase of the fabrication complexity and cost. The performance of a SPAD array integrating microlens is evaluated by considering two main parameters: the improvement factor resulting from microlens integration and the spatial uniformity of detection.

A number of approaches have been attempted, including the use of non-imaging concentrators such reflective-based structures [28], however these are difficult to fabricate on the scale required. Other approaches utilising molded refractive microlenses have been used [4], [5] and large improvements in fill factor were reported under limited conditions for a single detector at low numerical apertures (ie <0.05, corresponding to an f-number of greater than f/10). More recently, Pavia *et al.* also reported the fabrication and integration of 128 x 128 refractive microlenses on a SPAD array fabricated in 0.35 μm HV-CMOS technology with 6 μm diameter SPADs, and 25 μm pitch [6]. This resulted in a fill-factor of 5%. Different microlenses with different heights (defined as the distance from the base of the microlenses to the photosensitive area of the chip), of 30 μm and 70 μm, were fabricated in order to evaluate their performance. For both microlenses, a maximum IF of ~9 (30 μm height) and ~7 (70 μm height) was measured at f/22.

The uniformity of the improvement factor was also measured and the mean value and standard deviation were reported at different f-numbers. From these values, it is possible to calculate the coefficient of variation (CV) which is the ratio between the standard deviation and the mean. A CV of 5.6% at f/2.8 was obtained, while it increased to 18.2% at f/22. The authors stated that the IF was less uniform at higher f-number due to either a telecentric error or a misalignment between pixels and microlenses, since both problems had the same effect of decentring the focal spot from the pixel active area. However, the uniformity of the IF represents a relative measurement, since it is calculated as the ratio between the intensity light profile of two different SPAD arrays, the one integrating the microlenses, and the bare chip with no microlenses. Therefore, it should be more consistent to report the uniformity value of the two SPAD arrays

separately in order to ascertain if the microlenses introduce any degradation to the uniformity of the array which was not stated in ref. [6].

The same approach of microlens integration was used in this research work to recover the fill-factor of a 32x32 Si CMOS SPAD array. In comparison to previous work described above, two different sets of diffractive microlens were fabricated, integrated and characterised. In particular, a completely new approach based on a double telecentric imaging system was used for the microlens characterisation in terms of IF and uniformity. Furthermore, a full characterisation of a SPAD array in the spectral range between 500 nm and 900 nm is reported for the first time in the scientific literature, as shown in the following sections.

6.4 HV-CMOS Si SPAD

The 32 x 32 Si CMOS SPAD detectors arrays used in this research work were developed under the European Commission funded project called MiSPIA (Microelectronic Single-Photon 3D Imaging Array for low-light high-speed Safety and Security Applications). These arrays were designed by our collaborator at Politecnico di Milano and fabricated using a 0.35 µm HV-CMOS technology process at the Fraunhofer foundry IMS [19]. This array was designed to be used in applications such as time-of-flight 3D ranging and microscopy. The 32 x 32 array had 150 x 150 µm pixels and comprised a 30 µm active area diameter SPAD and its associated circuitry for counting, timing and quenching. These smart pixel dimensions resulted in a fill-factor of 3.14% for the chip without microlenses. A block diagram of the pixel is shown in Figure 6.2.

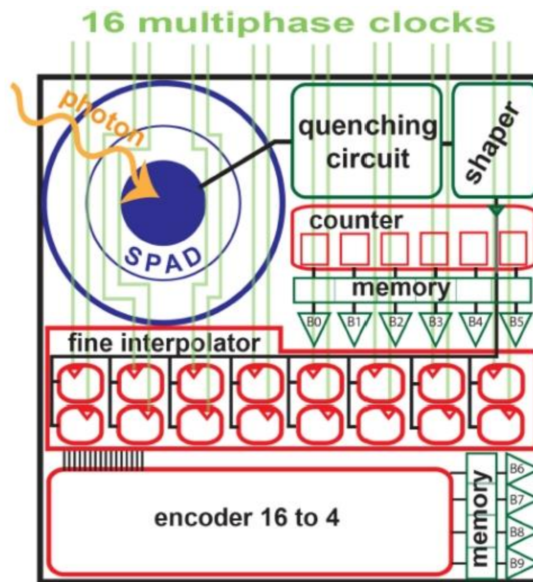


Figure 6.2. Smart pixel block diagram with 30 µm active area SPAD and its associated circuitry for counting, timing and quenching [19].

Each smart pixel of the array is able to detect a single-photon (in the 300-900 nm wavelength range), time-stamp its arrival time (with 312 ps resolution) for acquiring waveforms and time-resolved maps, and count photons for providing photon-number (e.g. intensity) resolved 2-D videos [21]. The designed and fabricated SPAD has a structure similar to the one reported in Figure 6.2. The 30 µm diameter SPADs were characterised by our collaborators at the Politecnico di Milano in terms of SPDE, DCR, time jitter, afterpulsing. Their performance is reported in Ref. [21]. These devices demonstrated a peak of the SPDE of 55% at a wavelength of 450 nm, while it was 45% at 400 and 500 nm, and still 20% at 300 nm in the NUV and at 650 nm, and 5% at 850 nm in the NIR ends of the silicon sensitivity (Figure 6.3). The measured DCR at room temperature for a 30 µm diameter SPAD was 120 counts/s at 6V excess bias (Figure 6.4), while the time jitter measured at FWHM was 85 ps at the peak wavelength (450 nm).

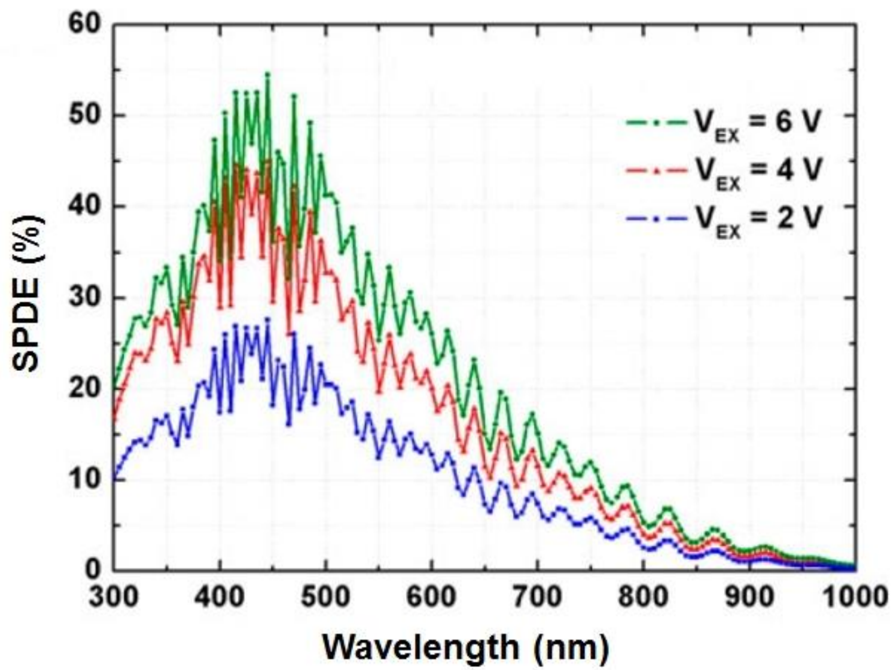


Figure 6.3. Single Photon detection efficiency vs. wavelength at different excess bias voltages: 2V (blue), 4V (red), and 6V (green). The peak SPDE is about 55% at λ = 450 nm, and still 5 % at λ = 850 nm [21].

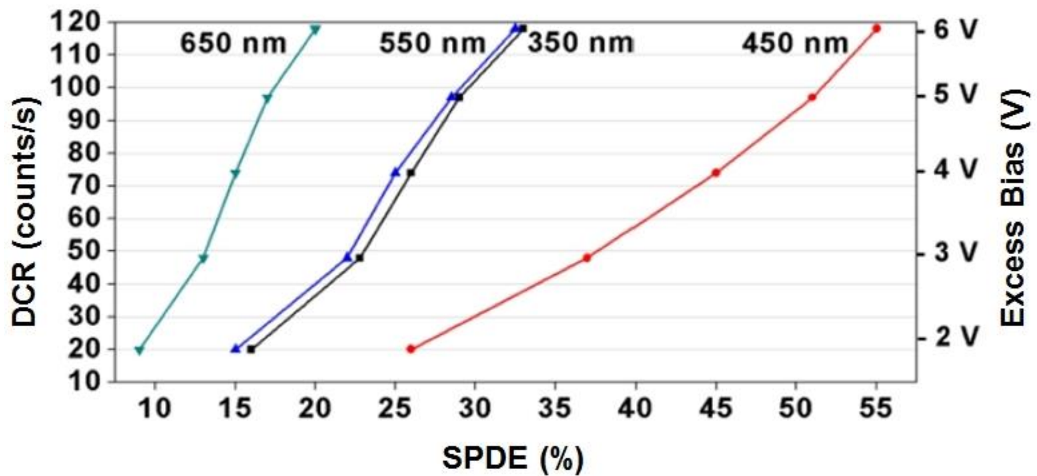


Figure 6.4. Dependence of DCR versus SPDE at different excess bias and four different wavelengths: 450 nm (red), 350 nm (black), 550 nm (blue), and 650 nm (cyan) [19].

The in-pixel electronics (Figure 6.2) included a quenching circuit with active reset for fast avalanche sensing and quenching, pulse shaping electronics for proper synchronisation, a 10 bit time to digital converter (TDC), an 8 bit counter, and an 8 bit memory register. A 10 bit memory latch stores the results of the conversion, and the B0:B9 output buffers drive the readout data buses. By using a 1024 in-pixel memory, a new frame can be acquired while performing the readout of the previous one in a global shutter technique (all pixels start and stop counting at the same time instant): which

avoids problems connected with image smearing of full-frame CCDs. The array is designed to work in two modes of operation: photon-timing and photon-counting. The photon counting mode was used for all the measurements reported in this chapter, and it is explained in the following section.

6.4.1 Photon-counting mode of operation

In this mode of operation, the beginning of a frame is marked by a Start signal which resets the in-pixel counter (Figure 6.5a).

When a photon is absorbed by the SPAD active area, this can initiate the avalanche process and produces a macroscopic current signal which is sensed by the electronics (Figure 6.5b). The avalanche process is quenched and the electronics provide a pulse to the counter which increments the number of events detected (Figure 6.5c). The electronics hold the SPAD off (quenched) for a pre-defined hold-off time during which photons cannot be detected (Figure 6.5d). Finally, the SPAD is reset back to Geiger mode (above breakdown), so that the pixel is ready to detect other photons (Figure 6.5e.). This procedure (from Figure 6.5b to Figure 6.5e) is performed when a photon is incident and triggers the detector during the whole time duration of a frame.

At the end of every frame, an external synchronisation pulse forces the on-chip global electronics to generate a Stop pulse which latches the number of photons collected by the counter into a temporary storage register (Figure 6.5f).

After few nanoseconds, a new frame acquisition will start (Figure 6.5a).

At the end of a frame the acquired values are copied and stored in a local memory, so that a new acquisition can start immediately without waiting for the readout which is performed simultaneously during the next acquisition frame. The readout is achieved by following a row-by-column scheme, and the readout of each pixel is performed in 10 ns, meaning that the readout of the whole array is carried out in 10 μs, corresponding to a maximum frame rate of 100 kframe/s [19].

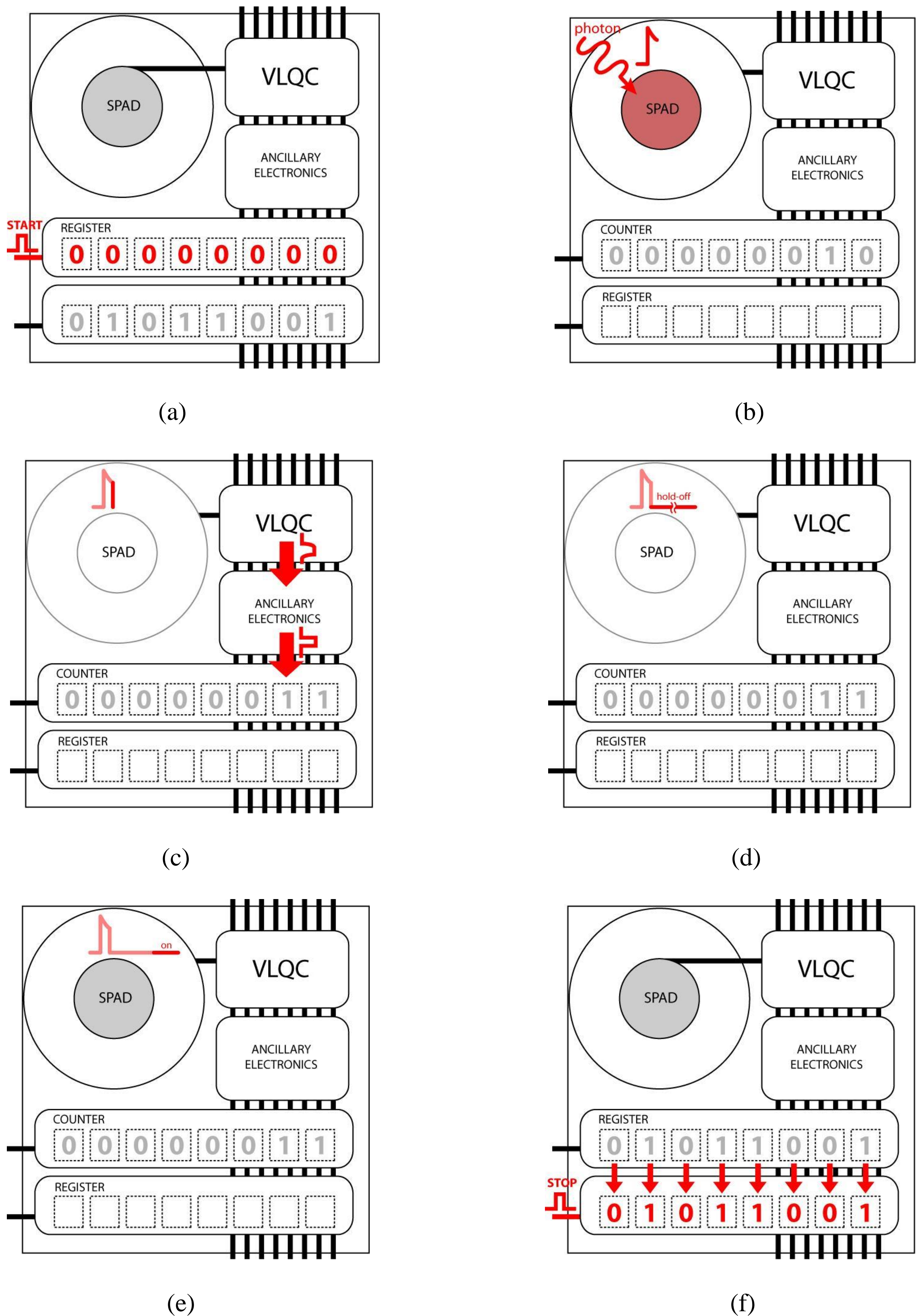


Figure 6.5. Photon-counting mode of operation of the CMOS Si SPAD array used for all the measurements reported in this chapter. (a) The beginning of a frame is marked by a Start signal, which resets the in-pixel counter. (b) When a photon is incident on the SPAD, this initiates an avalanche which is sensed by the electronics. (c) The SPAD is quenched and a pulse signal is provided to the

counter which increments the number of events detected. (d) The SPAD is held-off for a well-established time during which photons cannot be detected. (e) The SPAD is reset back to Geiger mode in order that other photons can be detected before the end of a frame. (f) At the end of the frame, an external synchronisation pulse forces the on-chip electronics to generate a Stop signal which latches the number of photons collected by the counter into a temporary storage register.

6.5 Diffractive Microlens Array Design

The phase function of a lens can be implemented as a refractive or a diffractive element. While the former element generates the phase distribution by varying the optical path length through a phase plate, in the latter element the phase function is mainly generated by the position and the grating period of a local grating. A diffraction lens is based on near field diffraction at a Fresnel zone plate (an amplitude pattern which is formed by a series of concentric rings), so the shape of the grating determines the diffraction efficiency of the element, which is the amount of light that goes into a particular diffraction order. When a Fresnel zone plate (FZP) is illuminated by a plane monochromatic wave of wavelength λ, a multitude of diverging and converging spherical waves can be observed behind the FZP. Each wave represents one diffraction order whose amplitude and focal length are determined by the FZP pattern. The undiffracted light which passes through the FZP forms the zeroth order [29].
In order to realise a diffractive element starting from a refractive lens profile, the phase function is wrapped to an interval between 0 and an integer multiple of 2π. Therefore, a refractive phase profile can be approximated in a diffractive lens by slicing the phase profile into 2π width layers, or phase shift (modulo 2π lens) for the considered wavelength for a maximum diffraction efficiency, as shown in Figure 6.6.

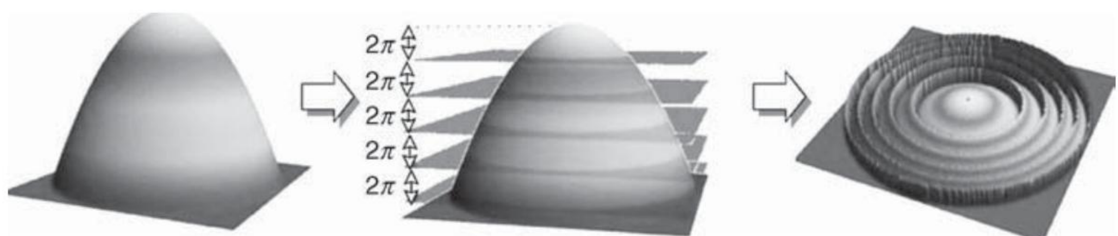


Figure 6.6. From the initial refractive general phase profile to the diffractive optical element fringes [30].

The position and widths of the successive zones of a spherical Fresnel lens are then computed by using CAD tools or analytic expressions. Figure 6.7 shows how the fringe

positions are computed (integer numbers of waves departing from the desired focal point).

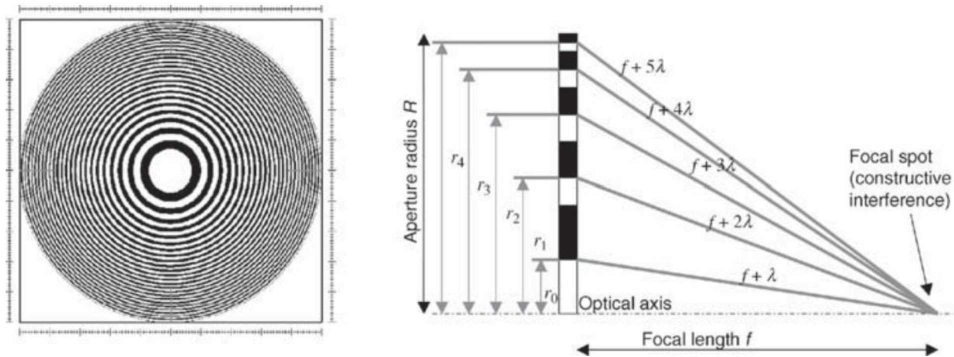


Figure 6.7. Fresnel zone microlens [30].

As shown in Figure 6.7, the zone radii are defined so that the distance from the edge of each zone (*m*) to the focal point is a multiple of the designed wavelength.

Diffractive microlenses belong to the category of digital diffractive optics since the performance of the diffractive optical elements can only be estimated numerically and fabricated through well-established lithographic techniques (using binary masks, as done for digital electronics), as shown in Figure 6.8.

Figure 6.8. Process steps required for the fabrication of a diffractive microlens.

The microlens profile used in this work had a staircase-like phase profile, as shown in Figure 6.9, and was approximated by power-of-two (N) discrete phase levels which can be fabricated through applications of $\log_2(N)$ binary amplitude masks in sequence [31]. The diffractive microlens arrays used in this Chapter consisted of 16 phase levels, and hence required four successive binary masks ($16 = 2^4$) to be generated. These four masks were then used in multilevel photolithography to generate the 16 phase levels.

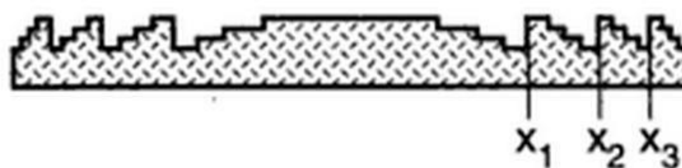


Figure 6.9. Schematic of a diffractive microlens with a multilevel phase profile [32].

The diffraction efficiency (η) of staircase gratings depends on the number of phase levels, and is given by [32]:

$$\eta_m^N = \left| \frac{\sin(\frac{m\pi}{N})}{\frac{m\pi}{N}} \right|^2 \quad (6.1)$$

where m is the diffraction order and N is the maximum number of phase level. N can be related to the first-order diffraction angle θ from the following equation:

$$N = \frac{\lambda}{sin\theta \cdot m_f} \quad (6.2)$$

where m_f is the minimum feature size that can be fabricated with the available technology and λ is the wavelength. Therefore, by considering equation 6.2 the diffraction efficiency is given by:

$$\eta(\theta) = \left(\frac{sin\left[\left(\pi m_f sin\theta\right)/\lambda\right]}{\left(\pi m_f sin\theta\right)/\lambda} \right) \quad (6.3)$$

For a diffractive microlens, the angle θ varies from the centre ($\theta = 0$) up to the rim ($\theta = \theta_{max}$). The maximum angle is determined by the numerical aperture of the lens (NA = sin θ_{max}). The efficiency versus the diffraction angle is shown in Figure 6.10. Although equation 6.3 gives the maximum efficiency (solid line) for a constant sample space (= m_f), in practise the efficiency is lower because the phase values are quantised. This is also shown in Figure 6.10 where the efficiency profiles are plotted as dashed lines when 8, 4, 2 phase level are used.

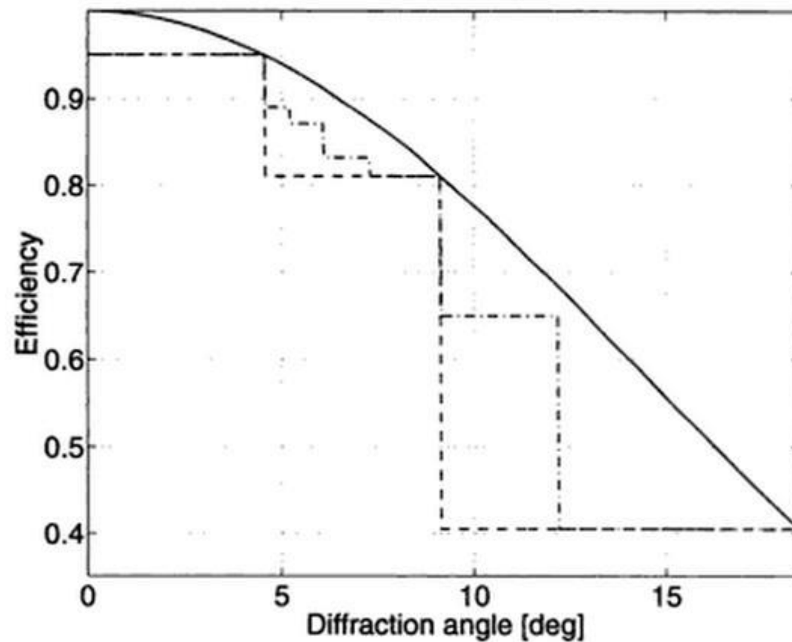


Figure 6.10. Optimum diffraction efficiency profile (solid line) and efficiency profiles when 8, 4, and 2 phase levels are used (dashed line). The fabrication limit is assumed to be 1 μm (m_f) and the operating wavelength is 632.8 nm [32].

For the diffractive microlens used in this work, consisting of 16 phase levels, the theoretical maximum diffraction efficiency is 98.7%. It is interesting to note that when the number of phase levels grows to 16 or more, the diffractive element can be easily considered as a quasi-analog surface-relief element. In fact, it does not make sense to fabricate a diffractive element with more than 16 levels using conventional multilevel masking techniques, since the successive systematic lateral misalignment errors and cascade etching depth errors would reduce the diffraction efficiency dramatically [29]. In particular, the fabrication is limited by the lateral size of the smallest structure that can be etched into the final substrate [30]. As the fringe width of a diffractive lens decreases radially, the maximum number of phase levels approximating the analog fringe relief surface also decreases (Figure 6.11).

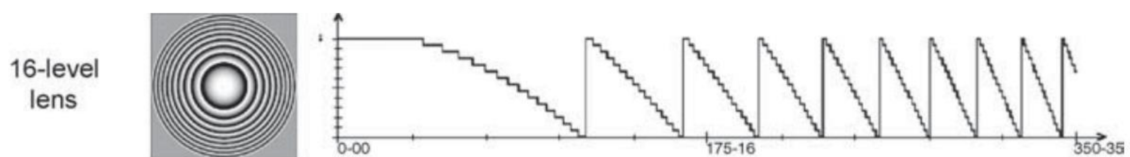


Figure 6.11. Diffractive optical element encoded over 16 phase levels [30].

The designed microlens surface was approximated by a modulo 2π zone plate representation, resulting in a fill-factor of the available area of ~100% (no dead zone

between microlenses) for a square pixel with total focusing efficiency in the 75-85% range and diffraction limited performance [29].

However, diffractive microlenses are very sensitive to changes in operational wavelength and show a strong spectral dispersion, since the focal lens (f_d) of a diffractive lens is inversely proportional to the wavelength λ given by [32]:

$$f_d(\lambda) = f_0 \frac{\lambda_0}{\lambda} \tag{6.4}$$

Where λ_0 is the design wavelength and f_0 is the design focal length. However, diffractive microlenses can be optimised to work over a wider range of wavelengths. The diffractive microlenses used in this study were designed to work at a wavelength of 808 nm, but simulations of the overall coupling efficiency (defined as the percentage of the incident light landing within a 30 μm diameter circle on the focal plane of the lens) as a function of the wavelength for three different lens design showed that these microlenses can produce a focal spot within the SPAD active area over a significant range of wavelengths, as illustrated in Figure 6.12. Experimental results of the IF, also confirmed that a high IF was also obtained at different wavelengths rather than at the designed one (as shown in detail in following sections).

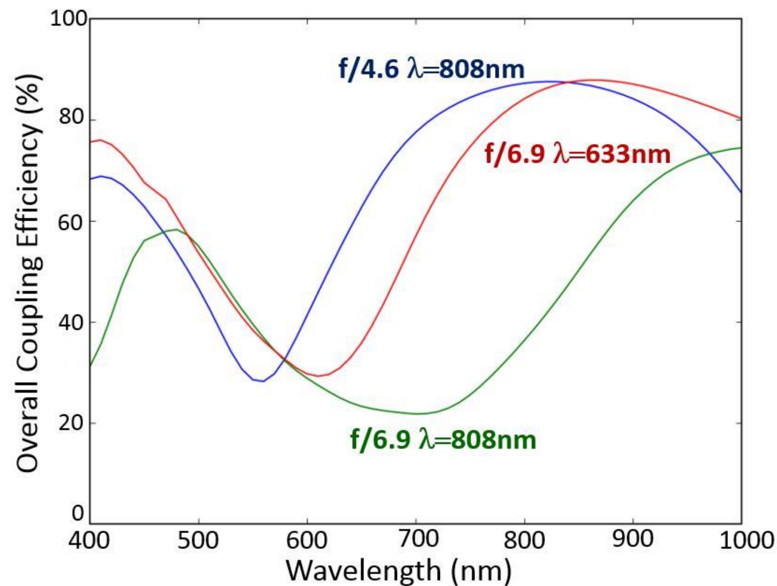


Figure 6.12. Overall coupling efficiency versus wavelengths for three different diffractive microlens designs.

Due to the fabrication limitations on the minimum attainable feature size (~1.8μm), the f-number of the microlenses was constrained to be greater than f/2.9 in order to ensure

100% fill-factor. To improve the mechanical handling capabilities of the microlens arrays, it was decided to use substrate thicknesses of 1.0mm for the microlens arrays. These substrate thicknesses gave lenses with f-numbers equal to f/4.6.

In particular, two different sets of 32 x 32 plano-convex infinite and finite conjugate diffractive microlenses arrays were designed and fabricated on a fused silica substrate. While the infinite conjugate microlenses were designed to image the light from a source placed at infinity, the finite conjugate microlenses were designed to image an object placed a finite distance from the microlens. A schematic representation of both microlenses is shown in Figure 6.13a-b. In addition, both microlenses focused the light on the back surface so that they could be mechanically bonded on the SPAD array (see section 6.5.2).

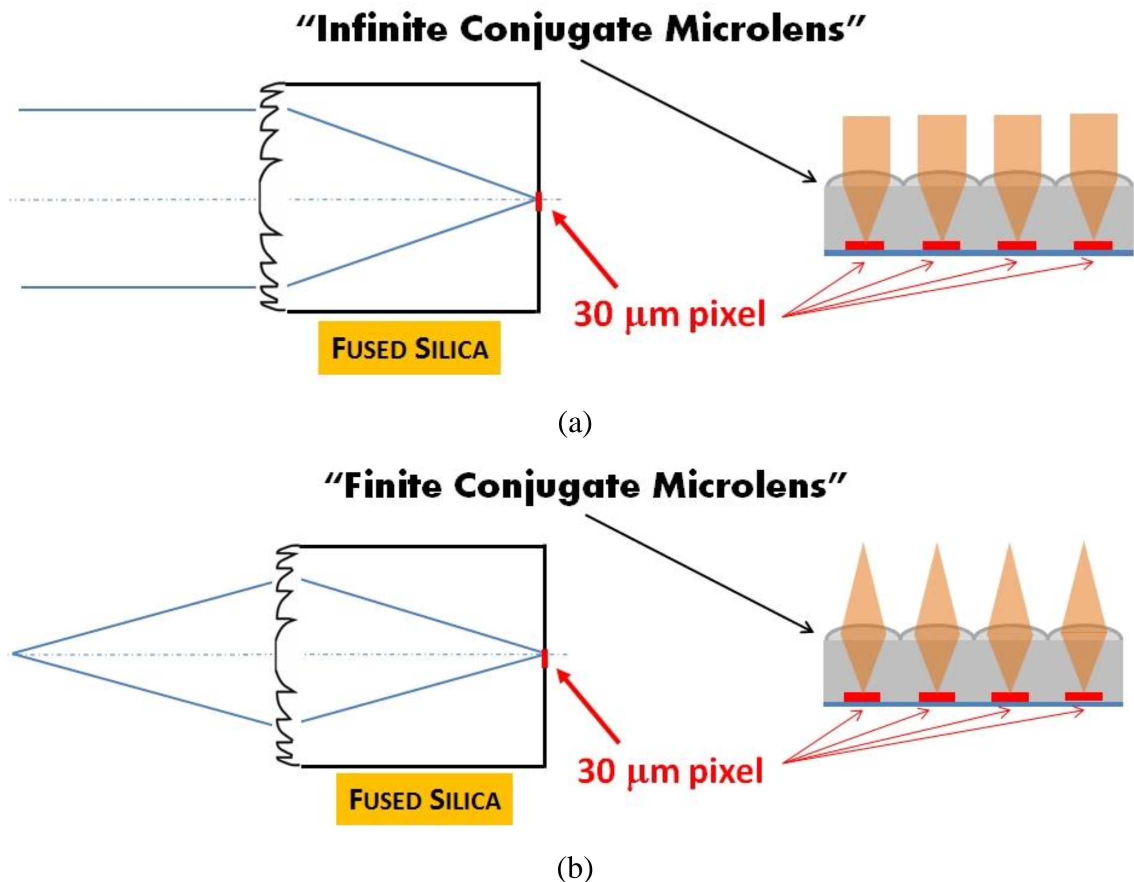


Figure 6.13. Schematic representation of the designed microlenses: (a) infinite conjugate microlenses were designed to image the light from an object placed at infinity; and (b) finite conjugate microlenses were designed for practical applications since an object at a finite distance from the microlens is imaged onto the detector.

Figure 6.14 shows simulations, performed by Dr. Andrew Waddie at Heriot-Watt University, of the theoretical IF as a function of the wavelength for the designed two

different sets of diffractive microlenses, infinite (red) and finite (blue) conjugate, respectively.

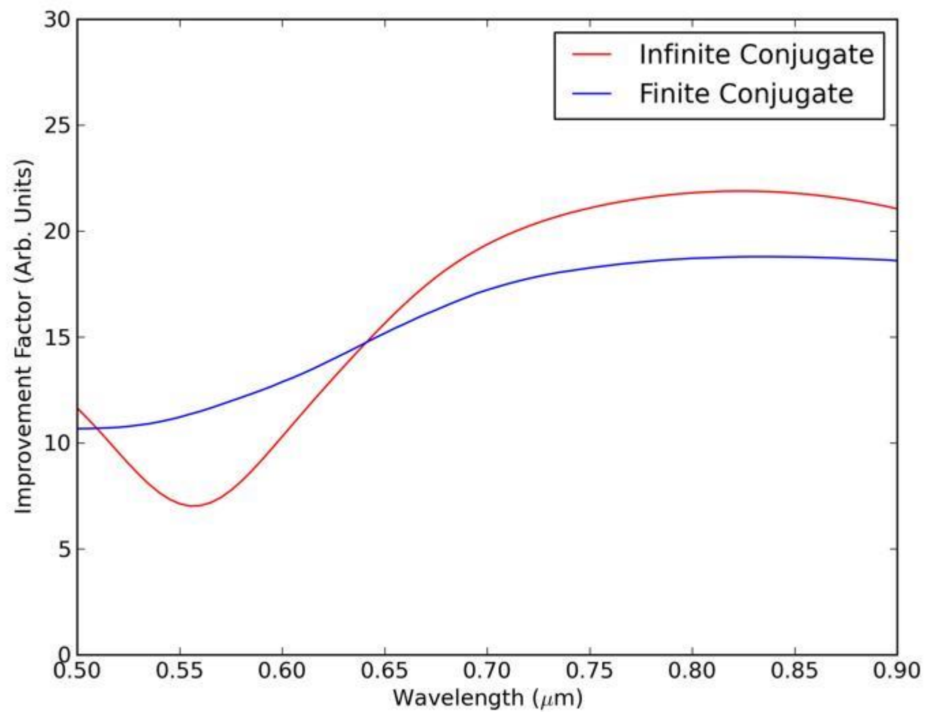


Figure 6.14. Simulations of the theoretical IF as a function of the wavelength for the designed two different sets of diffractive microlens, infinite (red) and finite (blue) conjugate respectively.

6.5.1 Microlens performance

Figure 6.15 shows an image of a part of the 32 x 32 diffractive microlens array designed to operate at a wavelength of 808 nm.

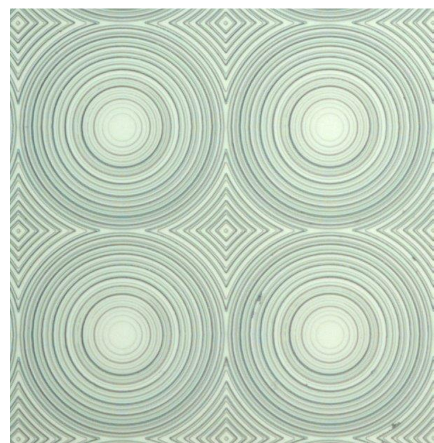


Figure 6.15. An image showing part of 32 x 32 microlens array designed for 808nm illumination.

The focal lengths of the fabricated lens were tested by Dr Andrew Waddie at Heriot-Watt. Due to the lack of an 808 nm wavelength source at that time, the testing was performed using a 780 nm wavelength source. However, this change in operational wavelength could be accounted for. For the f/4.6 diffractive microlens, the measured focal length at a wavelength of 780 nm was equal to 0.703 mm ± 0.5 mm. This value was in line with the focal length of 0.715 mm predicted by theory. The diffraction limited spot size predicted by theory was 9 μm, while the estimated focal spot size (Figure 6.16) was 7.6 μm ± 2 μm.

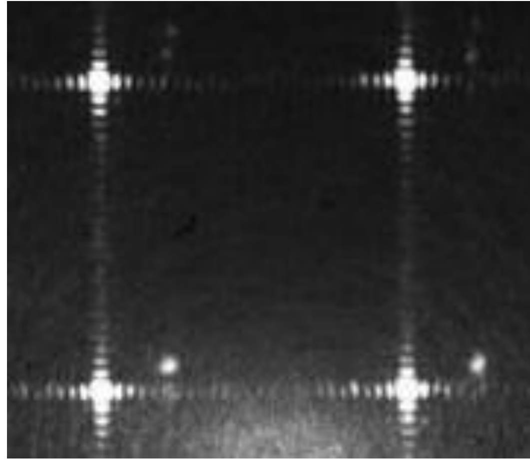


Figure 6.16. Diffracted limited focal spot size estimated for the designed diffractive microlens obtained by imaging the microlens on a CCD camera using a ×10 microscope objective.

6.5.2 Microlens integration

The microlenses were assembled to build the final 32 × 32 arrays (~5.6 × 5.6 mm total dimensions) which were mounted on top of the final SPAD imager chips (~9 × 9 mm). To perform the bonding operation, three metal (chrome) alignment marks or fiducials (Figure 6.17) were deposited on the same side as the microlenses. The same fiducials were also placed on the SPAD array imager during its fabrication.

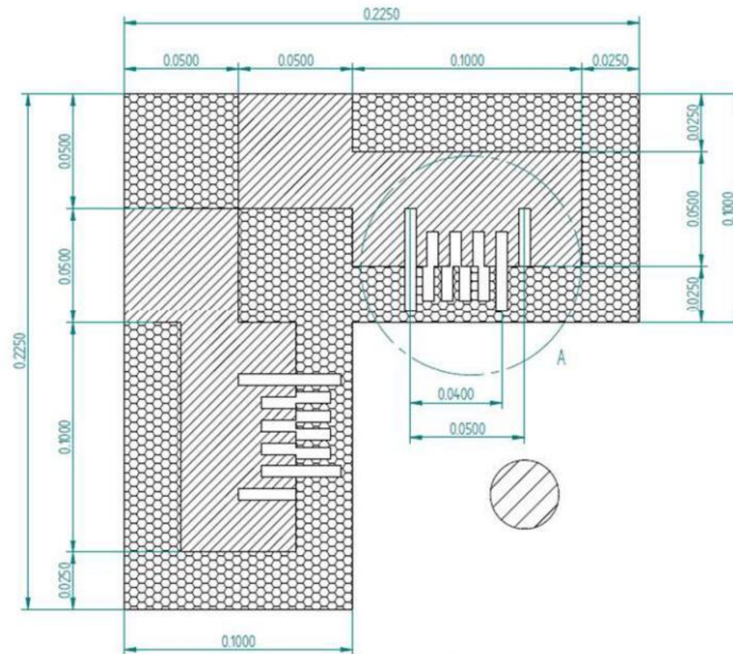


Figure 6.17. Technical drawing of fiducials mark which was deposited on the microlens substrate and at the corners of the SPAD array chips in order to perform the alignment and bonding operation between the microlens array and the SPAD sensor. All the dimension are reported in μm.

The two arrays, microlens and SPAD imager, were mounted (held in place with vacuum chucks) in a flip-chip bonder (Karl Suss FC6) which guaranteed a precise alignment with an accuracy of less than 5 μm. The fiducials on both arrays were first aligned, the appropriate bonding parameters were then configured, and the bonding sequence was initiated. After a post bond inspection of the fiducial alignment by using a microscope, the wire bonding was carried out. Figure 6.18 shows a photograph of a fully assembled 32 × 32 Si CMOS SPAD sensors integrating microlens.

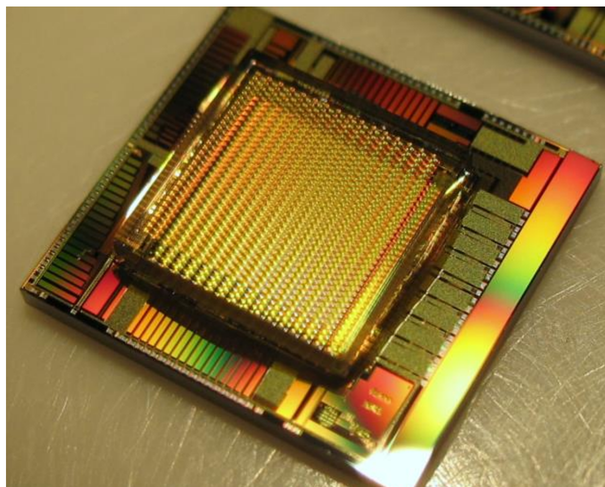


Figure 6.18. Photograph of a fully assembled 32 × 32 Si CMOS SPAD sensors integrating microlens.

6.6 Characterisation setup

The optical setup used for the characterisation of the SPAD arrays integrating the two different sets of diffractive microlenses arrays (described in section 6.5) is illustrated in Figure 6.19.

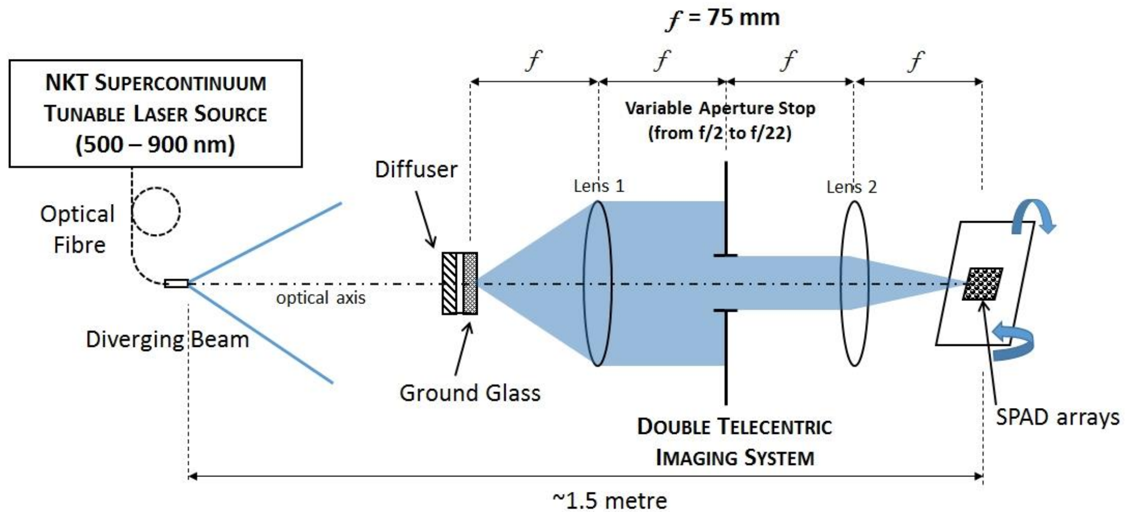


Figure 6.19. Experimental setup used to evaluate the improvement factor resulting from microlens integration at varying f-numbers (from f/2 to f/22 with one-stop increment) and different wavelengths between 500 nm and 900 nm. This setup was also used to evaluate the uniformity of SPAD arrays.

A tunable NKT Supercontinuum laser source (SuperK Extreme EX-W6) was used to make measurements of the IF and uniformity in the spectral range between 500 – 900 nm. Light coupled in a single-mode fibre (5 µm diameter core) then diverges upon exiting the fibre and propagates along the optical bench over a length of ~1.5 metre where it was incident on a diffuser. This diffuser ensured that the near-collimated light is diffused uniformly throughout a large cone angle (~ 50º), as illustrated in Figure 6.20.

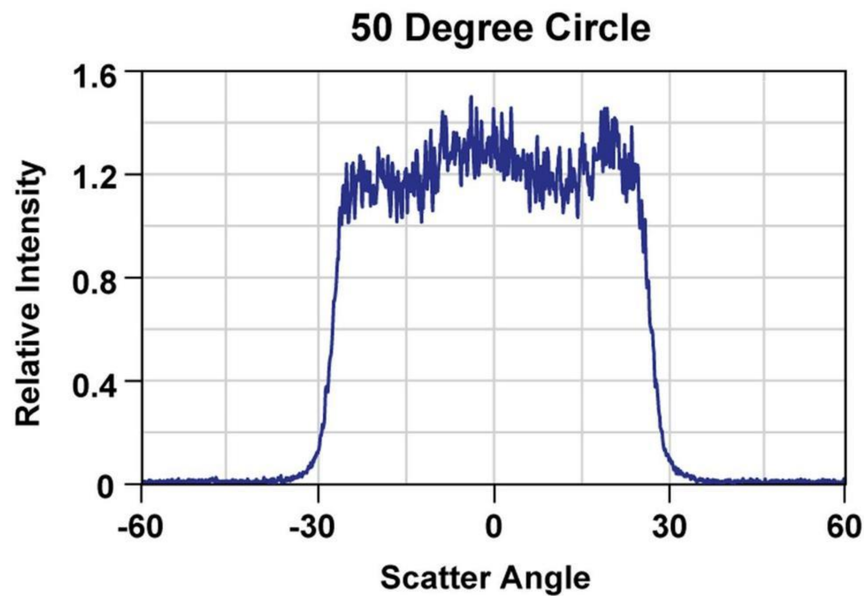


Figure 6.20. Output light intensity from the diffuser which was used in the characterisation setup (Figure 6.19) expressed as a function of the scattered angle.

After the diffuser, a ground glass plate (220 grit) was inserted as the surface to be imaged onto the SPAD arrays by means of a double telecentric image system. The telecentric system offers advantages such as a constant magnification over a defined depth of field (the setup in Figure 6.19 had a fixed magnification of 1:1), no perspective error, very low geometrical distortion, and uniform illumination. This system was composed of two identical achromatic doublet plano-convex lenses (focal length, $f = 75$ mm), with an aperture stop placed at the common focal point. In particular, the aperture stop was placed at a distance f from the first lens of the double telecentric system by using the optical setup illustrated in Figure 6.21. A diverging beam at a wavelength of 808 nm (which was the designed wavelength for both sets of microlens arrays) was incident on the aperture stop which was imaged through lens 1 and the concave mirror (with a focal length equal to $f_m = 560$ mm) on a CCD camera. The camera was precisely placed at a distance f_m from the mirror. By adjusting the distance between the aperture stop and lens 1, the image of the aperture stop on the CCD camera reached its maximum sharpness when the lens 1 was placed f away from the aperture stop. The same technique was also used to precisely place in position the other components of the double telecentric imaging system showed in Figure 6.19.

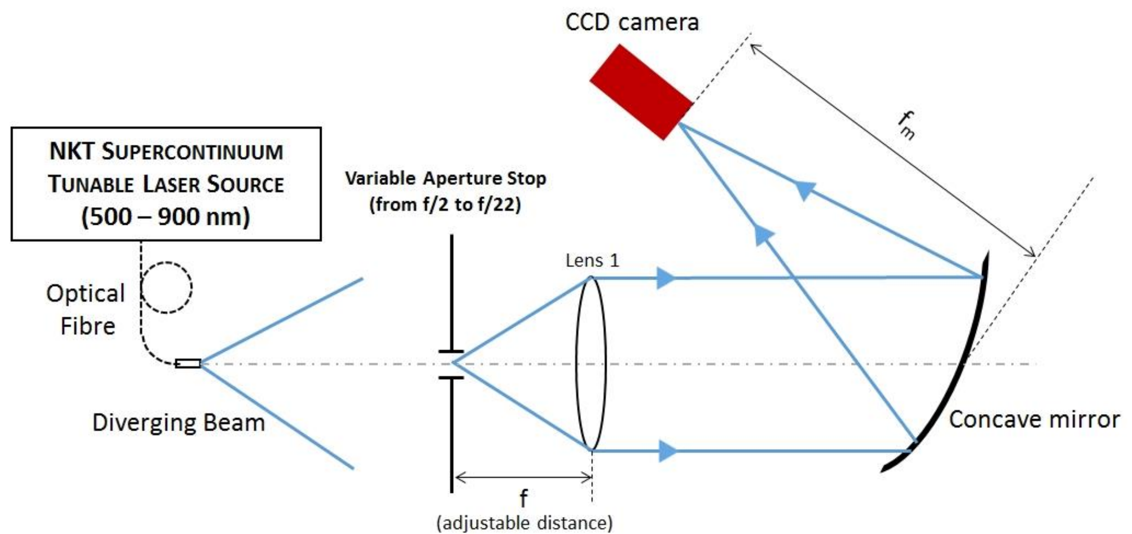


Figure 6.21.Optical setup used to precisely place in position the variable aperture stop at a distance f (f=75mm) from the first lens of the double telecentric imaging system. A diverging beam at a wavelength of 808 nm was incident on the aperture stop which was imaged through the lens 1 and the concave mirror (with a focal length equal to f_m) on a CCD camera. The camera was precisely placed at a distance f_m from the mirror. The image of the aperture stop on the CCD camera reached its maximum sharpness when lens 1 was placed f away from the aperture stop.

The SPAD array chips (namely the bare chip, infinite conjugate microlens and finite conjugate microlens SPAD arrays) were mounted, in turn, on a micrometer six-axis translation stage which was placed at a distance *f* from the rear lens of the telecentric system. The SPAD array chip, with or without microlenses, was mounted to be orthogonal to the incident beam coming out from the double telecentric imaging system. For this purpose, a reference beam was used to ensure that the incident light was normally incident on the SPAD array chip. A photograph of the custom made double-telecentric imaging system and the SPAD array is shown in Figure 6.22.

The optical setup illustrated in Figure 6.19 was used to characterize the SPAD arrays integrating the two different diffractive microlens arrays in terms of two main parameters: the IF and spatial uniformity. Measurements of both parameters were performed at different f-numbers (from f/2 to f/22 with one-stop increments), and different wavelengths between 500 and 900 nm. In addition, all the reported measurements were performed at room temperature and in completely dark conditions.

The excess bias voltage on SPAD arrays under test, with and without microlens arrays, was kept fixed at a value of 3 V.

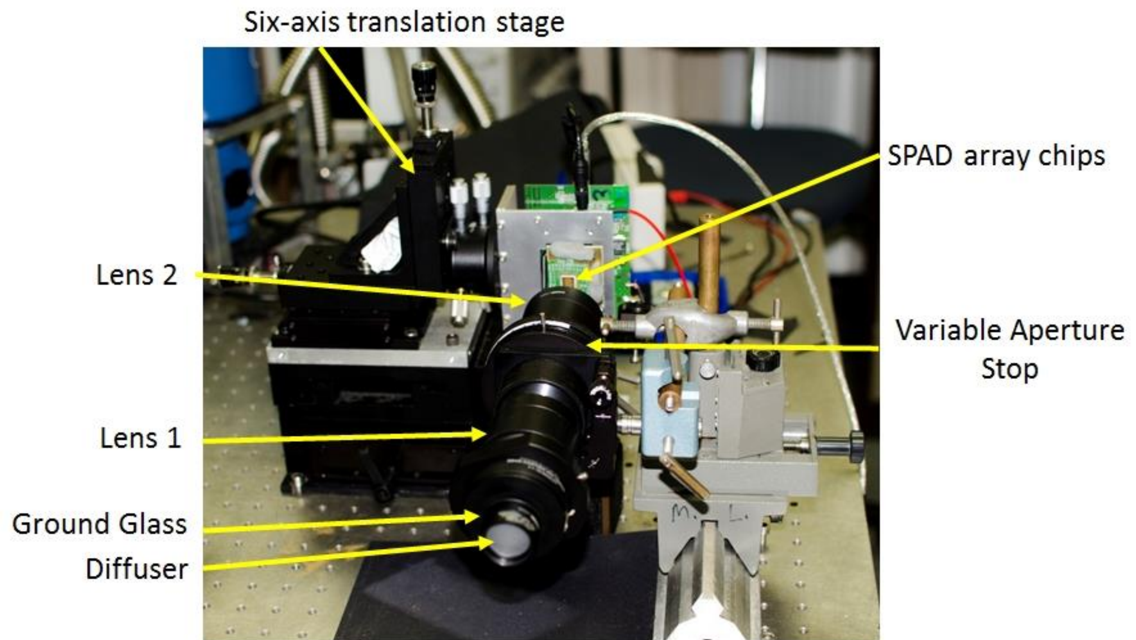


Figure 6.22. Photograph of the custom made double-telecentric imaging system and the SPAD array used in the setup showed in Figure 6.19.

The above optical system (Figure 6.19) worked in a similar manner to a camera system, such that when the f-number is incremented by one-stop, the light intensity at the camera sensors is halved, and hence a linear relationship exists between the f-number and the light intensity. Therefore, the robustness of the characterisation setup was demonstrated by measuring this linear relationship between the total numbers of counts (light intensity), for the SPAD sensor without microlenses, as a function of the aperture area (f-number) between f/2 and f/22. The total numbers of counts for the SPAD sensor were obtained by acquiring an image of the SPAD array in completely dark conditions (background). The SPAD sensor was then illuminated with light at a specific wavelength and another image was acquired (signal). Finally, the two images, light and dark, were subtracted to obtain the light intensity on the bare chip. The optical system guaranteed a linear relationship between aperture area and count rate across the measured range (from f/2 to f/22) and at different wavelengths (600, 700, and 808 nm), as shown in Figure 6.23.

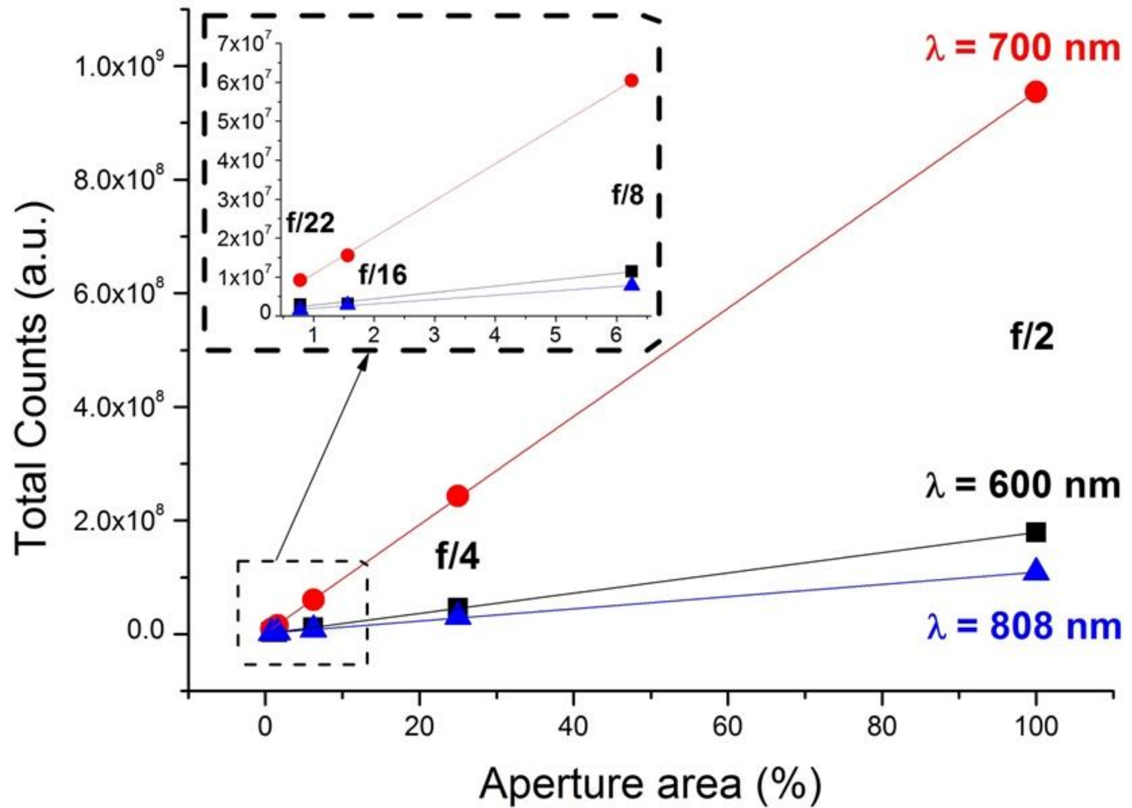


Figure 6.23. Linearity test of the total number of counts collected, in 10 frames, from the SPAD array without microlenses as a function of the aperture area (f-number). This test was performed at a constant laser power and three different wavelengths, 600 nm (black), 700 nm (red), and 808 nm (blue). Linear fitting is also shown. The inset in the figure also shows the linearity at the highest f-numbers, f/8, f/16, and f/22 respectively.

From Figure 6.23, it can also be observed that the magnitude of the total number of counts varied depending on the wavelength. Although these measurements were taken at a constant power level, this variation was due to the different detection efficiency of the SPAD array at different wavelengths (which is lower at longer wavelengths due to the lower Si absorption coefficient, see Figure 6.3 and ref. [21]) and the different power levels at different wavelength of the tuneable laser source used in our experiment. However, it is clear that as the f-number was decreased by one-stop, the light intensity (or total counts) was halved from the previous stop at any given wavelength. This linear relationship was demonstrated at three different wavelengths, 600 nm (black), 700 nm (red), and 808 nm (blue), respectively. In addition, these measurements (Figure 6.23) also demonstrated the robustness of our optical system even at the highest f-numbers, where the linear relationship between total counts (signal minus background) vs. aperture area was still satisfied (as shown from the inset in Figure 6.23).

6.7 Improvement Factor (IF) measurements

6.7.1 Methodology

To evaluate the infinite and finite conjugate diffractive microlenses performance resulting from the integration on the 32 × 32 SPAD array, the concept of improvement (or concentration) factor proposed in ref. [28] was used:

$$IF = \frac{E_0}{E_i} \tag{6.5}$$

where E_i is the input irradiance (optical power per unit area) at the microlens surface, and E_0 is the output irradiance or the irradiance at the photosensitive area of the pixel.

Empirically, the improvement factor was obtained by operating the SPAD array in photon counting mode (section 6.4.1) and dividing the detected photon event profile (which is the difference of the light and background signal) on the SPAD array with integrated microlens arrays by the detected photon profile measured by the SPAD array with no integrated microlens under the same illumination conditions. As mentioned above, the IF was measured in the spectral range between 500 and 900 nm. Hence, the improvement factor at a given f-number was the results of two measurements which were performed for each sensor, with and without microlenses, one illuminated (at a specific wavelength) and the other in complete darkness. Each measurement consisted of 10 frames which were acquired, and the counts for each pixel of the SPAD array in each frame were summed. Before calculating the IF, the hot pixels were removed from both the detected photon profile measured for both SPAD arrays, with and without microlenses, as shown in the next section.

6.7.2 Hot pixel removal

Hot pixels are defined as those SPADs within the array with a DCR much higher than the average value of those SPADs with a lower DCR [21]. Hot pixels within an array are mainly due to non-uniformities during the fabrication process, therefore their evaluation is a signature of the reliability and reproducibility of the fabrication. Although there is no technical definition in the scientific literature for a hot pixel, a pixel with a DCR higher than 2.5 times the average DCR was considered as a hot pixel for all measurements reported in this chapter. In addition, it is important to distinguish between hot pixels and pixels which could be affected by some kind of noise (optical or electronics). While the former pixels have the same behaviour (high DCR) during all

frames acquisitions, the latter pixels might show a high DCR only during some frames due to some noise. Since the hot pixel definition is based on the DCR, Figure 6.24 illustrated a background image (consisting of 10 frames) acquired in completely dark condition for the SPAD array without microlenses at f/5.6 to illustrate this concept further.

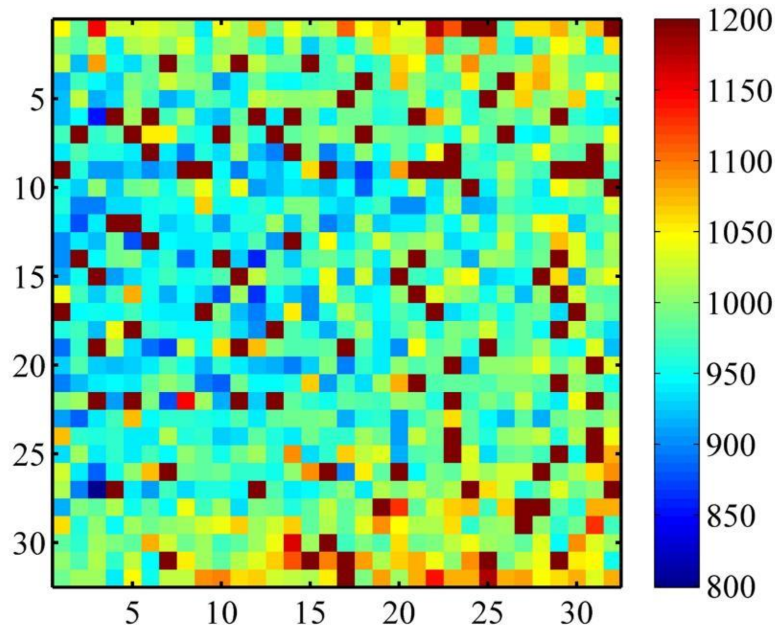


Figure 6.24. Background image (consisting of 10 frames) acquired in completely dark condition for the SPAD array without microlenses at f/5.6. The colour code scale indicates the dark count intensity for each pixel.

The colour code scale in Figure 6.24 indicates the dark count intensity for each pixel within the array. The majority of pixels had very similar value in the range of 950 – 1000 (similar colour range, between light blue and light green), while there were a certain number of pixels which reached and exceed the maximum value of the scale of 1200 (dark red). This dark count difference can be clearly observed in Figure 6.25 which shows the mean value of counts for each pixel in 10 frames for the same image reported in Figure 6.24.

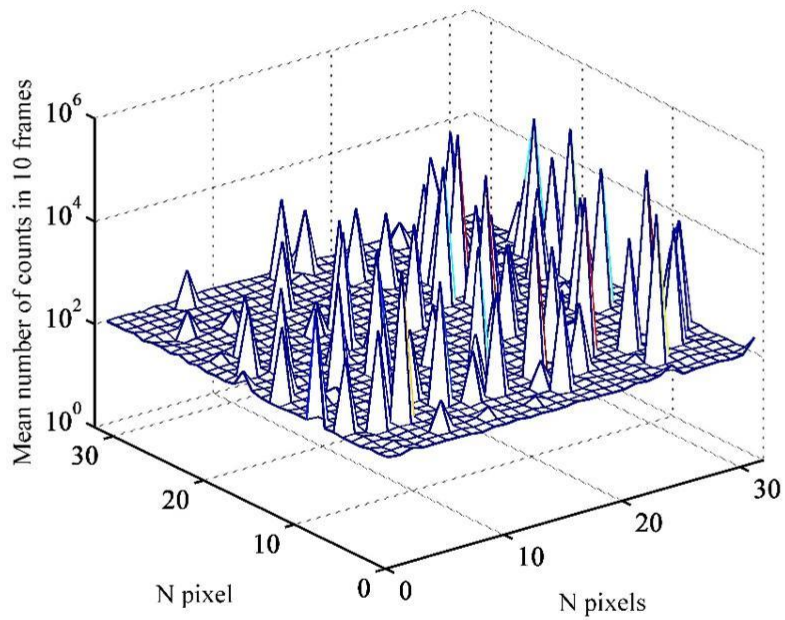


Figure 6.25. Mean number of counts in 10 frames for each pixel within the SPAD array without microlenses for the background image illustrated in Figure 6.24.

All the pixels that had a dark count intensity higher than 2.5 times the average value of the dark count of the majority of the pixels can be considered as hot pixels. Figure 6.26 clearly shows the number of hot pixels for the SPAD array under test for the image in Figure 6.24, where all the pixels with similar average dark count intensity were coloured in blue, while the hot pixels were coloured in dark red (Figure 6.26).

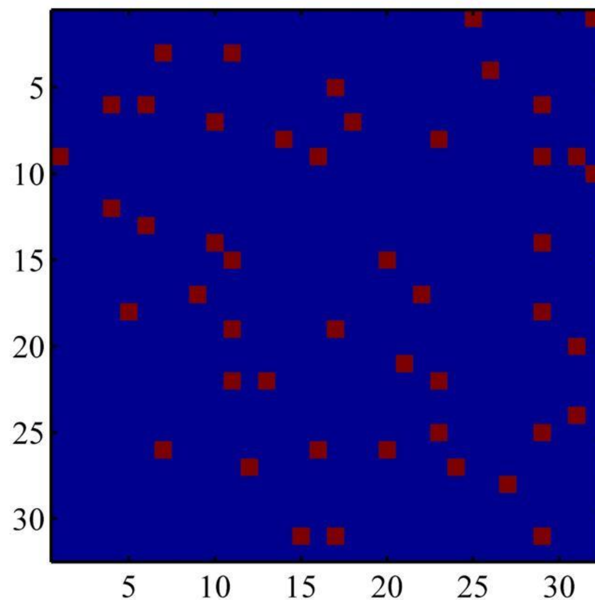


Figure 6.26. Number of hot pixels for the SPAD array without microlenses and the background image reported in Figure 6.24. All the pixels with a similar average value of the dark count intensity are displayed in blue, while the hot pixels are displayed in dark red.

Therefore the number of hot pixels for the SPAD array without microlenses was less than 5 % of the total pixels.

The same analysis was also performed for the SPAD array integrating the infinite or finite conjugate diffractive microlenses, and similar numbers of hot pixels (< 5% of total pixels) was measured for both chips.

Hot pixels were removed by using a median filter of rank of 3 × 3. This process of removing the hot pixels is illustrated in Figure 6.27. By considering a box of 9 pixels (3 × 3) with the hot pixel (HP) in the middle of the box, the value of the HP was then substituted by the median value (M) calculated for the neighbouring 8 pixels.

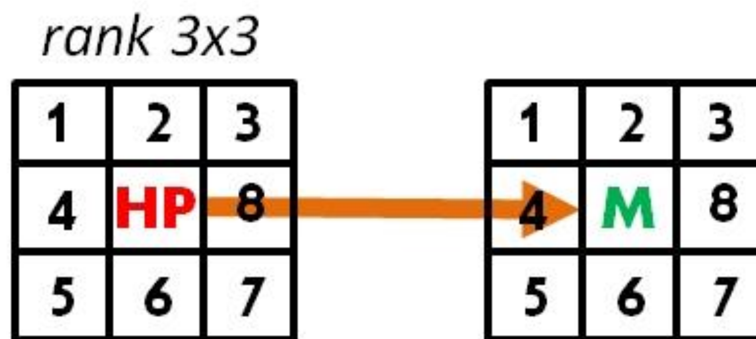


Figure 6.27. Hot pixel removal process by using a median filter of rank 3 × 3. The value of a hot pixel (HP) is substituted by the calculated median value for neighbouring 8 pixels.

6.7.3 Results of the Improvement Factor

After removing the hot pixels from the photon intensity profile for both chips, the SPAD array integrating the infinite or finite conjugate diffractive microlenses and the bare chip, the average IF was calculated as the ratio between the intensity profiles of the SPAD sensor with microlens and the bare chip. Figure 6.28 shows the IF measured for the SPAD array integrating the infinite conjugate diffractive microlenses as a function of the f-number (between f/2 and f/22 with one-stop increment) and different wavelengths (between 500 and 900 nm). Results of the IF are also summarised in Table I.

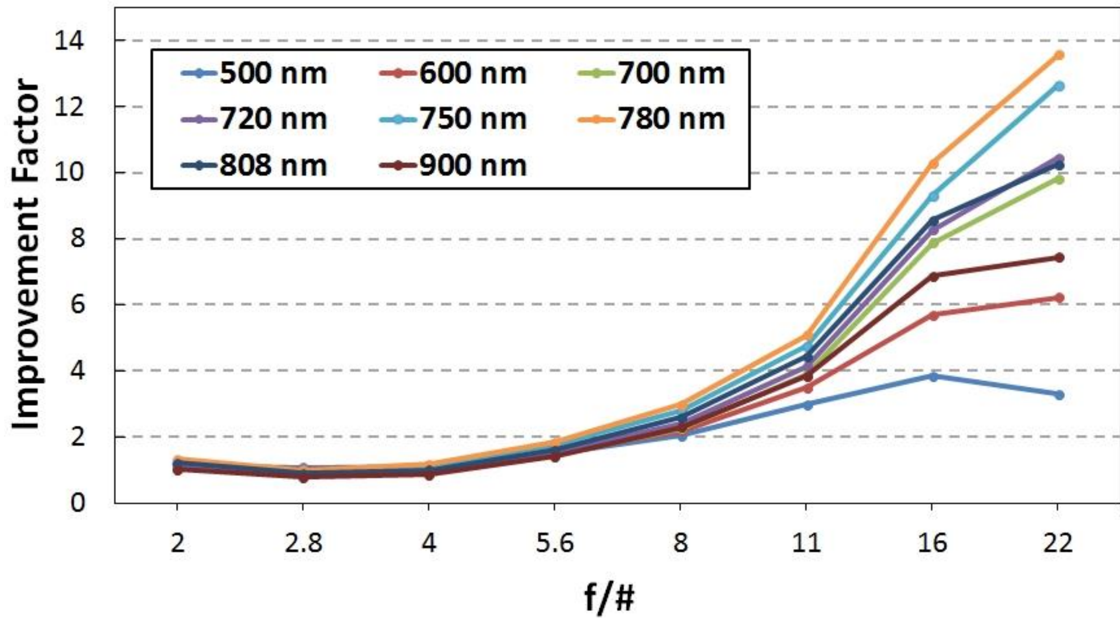


Figure 6.28. Measured improvement factor as a function of the f-number (from f/2 to f/22 with one-stop increment), and different wavelengths (500 – 900 nm) for the SPAD array integrating the infinite conjugate diffractive microlenses.

IMPROVEMENT FACTOR								
	Wavelength							
f/#	**500 nm**	**600 nm**	**700 nm**	**720 nm**	**750 nm**	**780 nm**	**808 nm**	**900 nm**
2	1.16	1.16	1.05	1.09	1.22	1.33	1.21	1.02
2.8	1.07	1.02	0.83	0.86	0.96	0.99	0.89	0.78
4	1.09	1.09	0.95	0.99	1.12	1.18	1	0.85
5.6	1.5	1.53	1.47	1.55	1.76	1.85	1.6	1.4
8	2.03	2.17	2.26	2.39	2.78	2.98	2.58	2.3
11	2.98	3.49	3.9	4.14	4.76	5.08	4.44	3.85
16	3.84	5.7	7.88	8.27	9.31	10.3	8.56	6.87
22	3.29	6.22	9.82	10.46	12.65	13.57	10.25	7.45

Table I. Summary of the improvement factor results which were measured for the SPAD array integrating the infinite conjugate diffractive microlenses.

Although the infinite conjugate microlenses were designed for a wavelength of 808 nm, as already discussed in section 6.5, these showed an IF across the whole wavelength range under investigation (500 – 900 nm), as shown in Figure 6.28. In particular, the

highest IF of ~14 was measured at λ = 780 nm and f/22. At the designed wavelength (808 nm) the IF was lower (but still high) and reached a maximum value of ~10 at f/22. The observed wavelength shift, from 808 nm to 780 nm, could be due to some tolerance error during the microlens array fabrication. This problem could also represent the reason of the wavelength shift for the measured IF for the SPAD array integrating the finite conjugate diffractive microlenses, as shown in Figure 6.29 and Table II. The highest IF of ~15.5 was measured at the peak wavelength of 750 nm and f/16, while it was lower and equal to ~7 at the designed wavelength of 808 nm and f/16.

The measured values of the IF for the SPAD arrays integrating the two sets of diffractive microlenses, infinite (Figure 6.28) and finite (Figure 6.29), represented the highest values of the IF measured and reported in the scientific literature across the whole spectral range between 500 and 900 nm. Therefore, these results demonstrated that diffractive microlenses can compensate substantially for very low fill-factor devices such as the SPAD sensors used in this work.

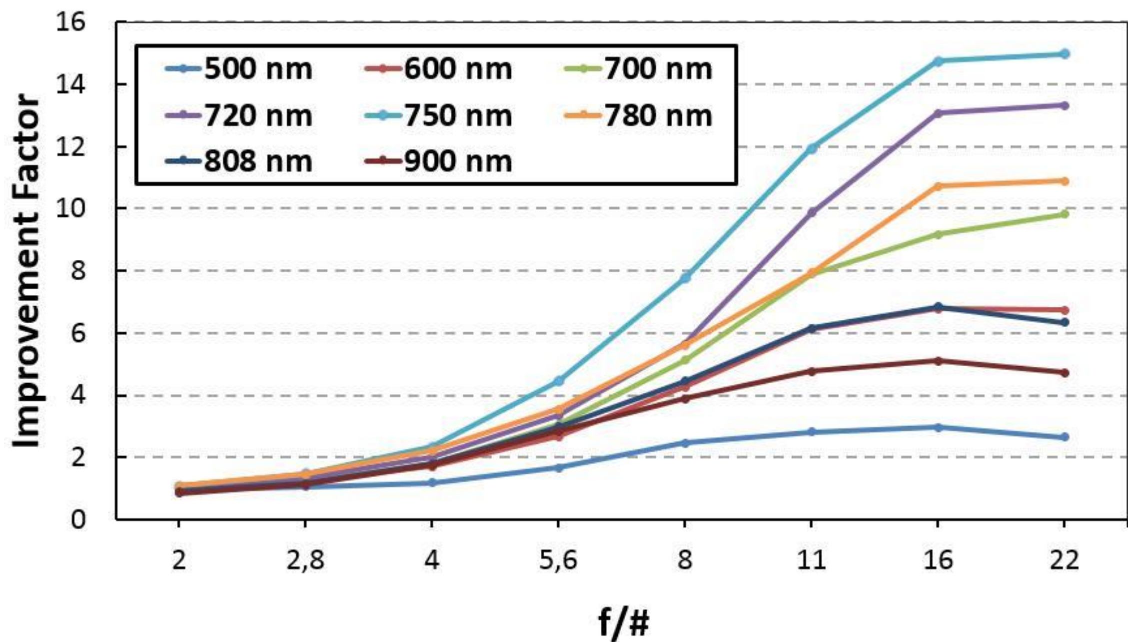


Figure 6.29. Measured improvement factor as a function of the f-number (from f/2 to f/22 with one-stop increment), and different wavelengths (500 – 900 nm) for the SPAD array integrating the finite conjugate diffractive microlenses.

IMPROVEMENT FACTOR								
	Wavelength							
f/#	**500 nm**	**600 nm**	**700 nm**	**720 nm**	**750 nm**	**780 nm**	**808 nm**	**900 nm**
2	0.96	1.08	0.94	0.97	1.07	1.08	0.9	0.84
2.8	1.06	1.26	1.2	1.32	1.48	1.45	1.17	1.13
4	1.19	1.7	1.82	2	2.33	2.22	1.78	1.75
5.6	1.66	2.67	3.05	3.34	4.45	3.55	2.98	2.84
8	2.46	4.26	5.12	5.65	7.76	5.6	4.46	3.89
11	2.82	6.14	7.9	9.87	11.94	7.94	6.16	4.76
16	2.96	6.78	9.17	13.07	14.75	10.73	6.85	5.11
22	2.64	6.74	9.83	13.33	14.97	10.9	6.33	4.73

Table II. Summary of the improvement factor results which were measured for the SPAD array integrating the finite conjugate diffractive microlenses.

In addition, it can be noted that the IF increased as the f-number was increased for both sets of microlens arrays (Figure 6.28 and Figure 6.29), reaching the highest values at the highest f-numbers. This trend of the IF at higher f-numbers is schematically explained in Figure 6.30 (using a single object point for clarity) where the dimension of the microlens array has been exaggerated in comparison to the double telecentric system. At low f-numbers between f/2 and f/5.6 (solid line), the aperture stop diameter is relatively large and the light coming from the double telecentric imaging system has a large cone angle (e.g. at f/4 is ~15°), while at higher f-numbers (dash line) the aperture diameter is smaller and, therefore, the light cone angle is reduced (e.g. at f/16 is ~3.57°). Hence, at low f-number this cone angle does not match the numerical aperture of the microlens configuration, while it does at the higher f-numbers. As a consequence, at low f-number the microlenses do not focus the light on to the SPAD active area as efficiently as they do at the higher f-numbers.

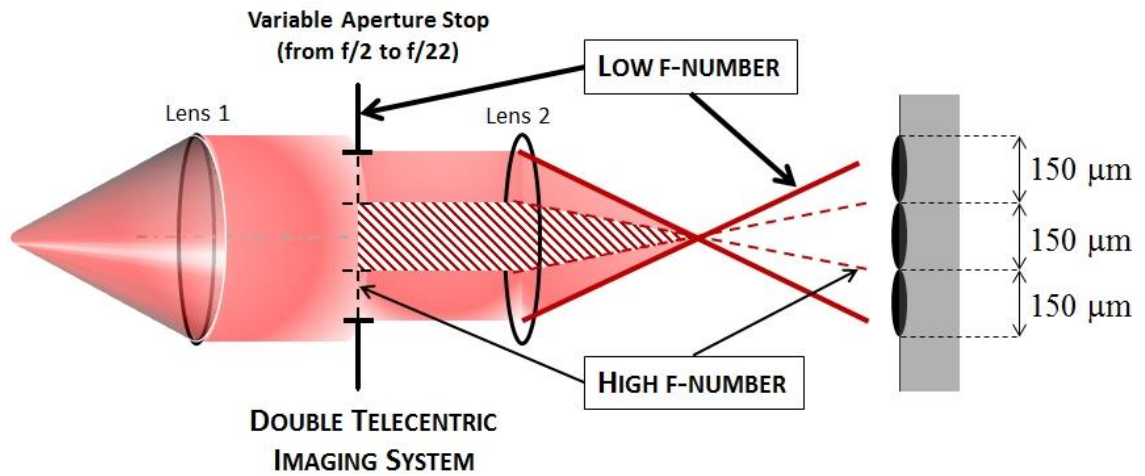
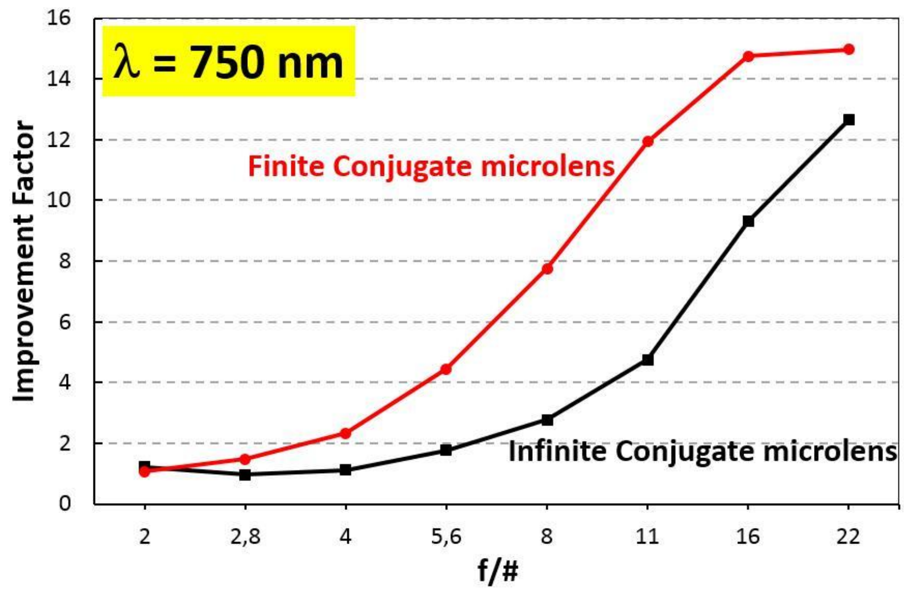


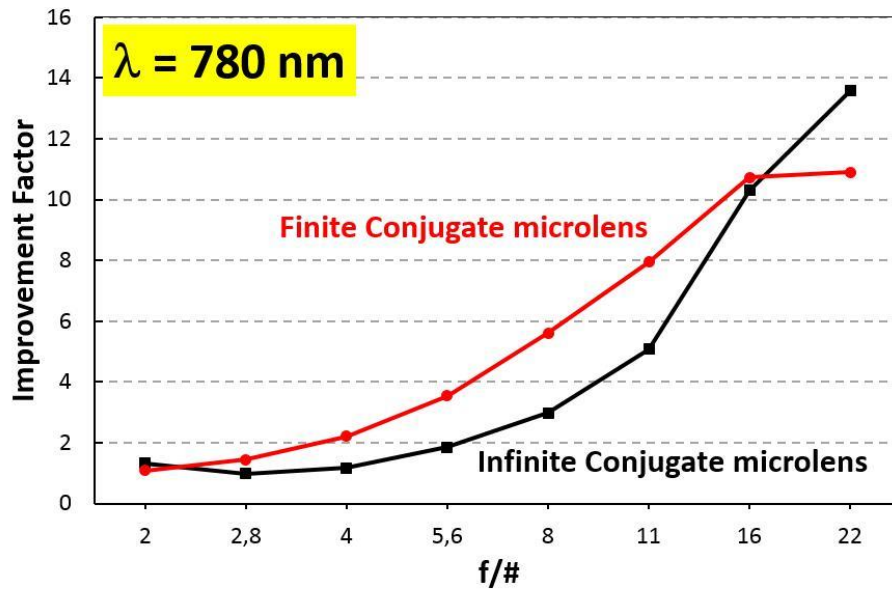
Figure 6.30. Behavioural model of the characterisation setup at different f-numbers. At low f-numbers (solid line) the aperture stop diameter is relatively large and the light coming from the double telecentric imaging system has a large cone angle, but at high f-numbers (dash line) the aperture diameter is smaller and the cone angle is reduced. The dimension of the microlens array has been exaggerated in comparison to the double telecentric imaging system.

6.7.4 Improvement Factor comparison

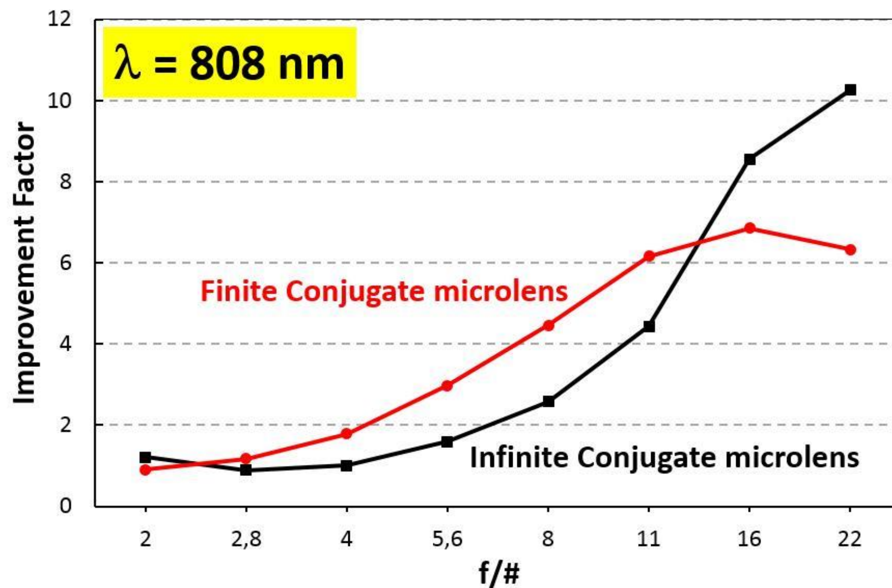
To further understand the behaviour of the IF at different f-numbers, for both microlenses, it is important to analyse a comparison of the IF between the two SPAD arrays integrating the diffractive microlenses. Figure 6.31a-c shows a comparison of the measured IF for both SPAD array integrating microlenses at their respective peak wavelengths (780 and 750 nm for the infinite and finite conjugate microlenses, respectively) and at the designed wavelength (808 nm).



(a)



(b)



(c)

Figure 6.31. Comparison of the experimentally measured IF at different wavelengths: (a) λ = 750 nm, (b) λ = 780 nm, and (c) λ = 808 nm for the SPAD arrays integrating the two different sets of diffractive microlenses, infinite (black) and finite (red) conjugate, respectively.

Figure 6.31a-c shows that the behaviour of both diffractive microlenses can be divided in two main regions depending on the f-number:

1. Low f-number operation (between f/2 and f/8). In this f-number range, it is evident that the finite conjugate microlenses (red line) demonstrated a higher IF than the infinite conjugate microlenses (black line) for all the IF comparisons (Figure 6.31a-c). The latter microlenses showed either no or only a slight improvement in this f-number range. In addition, this behaviour was also confirmed at a wavelength of 780 nm (Figure 6.31b) which correspond to the peak wavelength for the SPAD array integrating the infinite conjugate microlenses. At this wavelength the finite conjugate microlenses also demonstrated a higher value of the IF. This could be explained by the fact that the optical setup was mainly designed for the characterisation of a finite conjugate system since the light from the double telecentric imaging system was focused on the SPAD array. This is schematically shown in Figure 6.32.

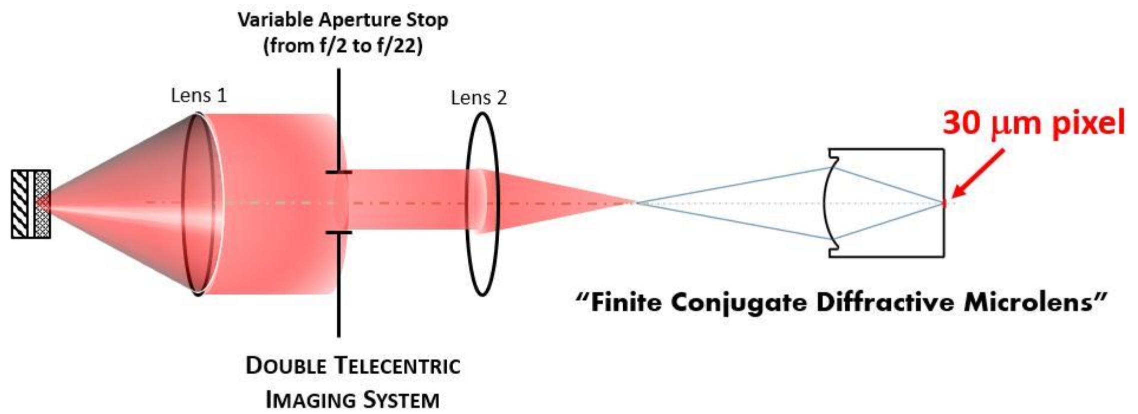


Figure 6.32. Schematic drawing of the optical setup used for the characterisation of the SPAD arrays integrating the diffractive microlenses in low f-number operation between f/2 and f/8.

2. High f-number operation (f/16 and f/22). In this f-number range, it is clear from Figure 6.31a-c that the IF of the SPAD array integrating the finite conjugate microlenses (red line) reached its maximum value at f/16 and, then it remained fairly constant at f/22. On the other hand, the SPAD array integrating the infinite conjugate microlenses showed an increase of the IF at f/16 and it reached its maximum value at the highest f-number (f/22). This behaviour could be explained by the fact that the aperture is very small in this f-number range, and therefore the light is coming from the double telecentric imaging system with a very small cone angle. This can be approximated to a flat wavefront which represents the optimum condition for the characterisation of the infinite conjugate microlenses (see Figure 6.13). This effect is schematically shown in Figure 6.33. These results also suggested that the IF of the infinite conjugate microlenses might further increase at a higher f-number than f/22. However, this could not be demonstrated, due to the lack of light at these extreme conditions.

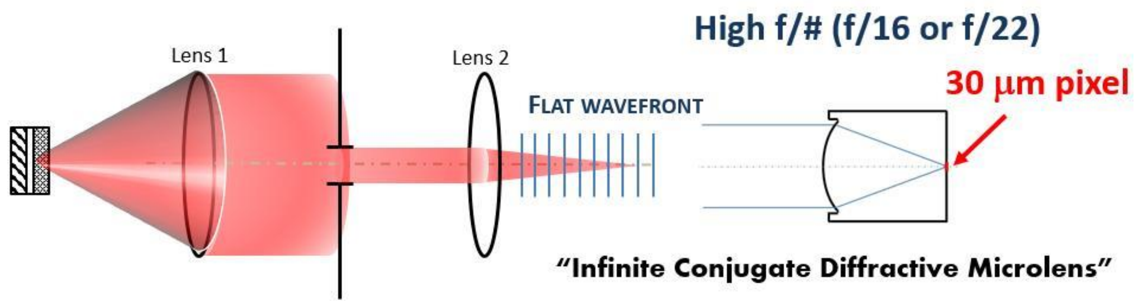


Figure 6.33. Schematic drawing of the optical setup used for the characterisation of the SPAD arrays integrating the diffractive microlenses at high f-number operation (f/16 and f/22).

It is also important to show a comparison of the IF as a function of the wavelength for both SPAD arrays integrating the two different diffractive microlenses array and at fixed f-number (f/16), as shown in Figure 6.34.

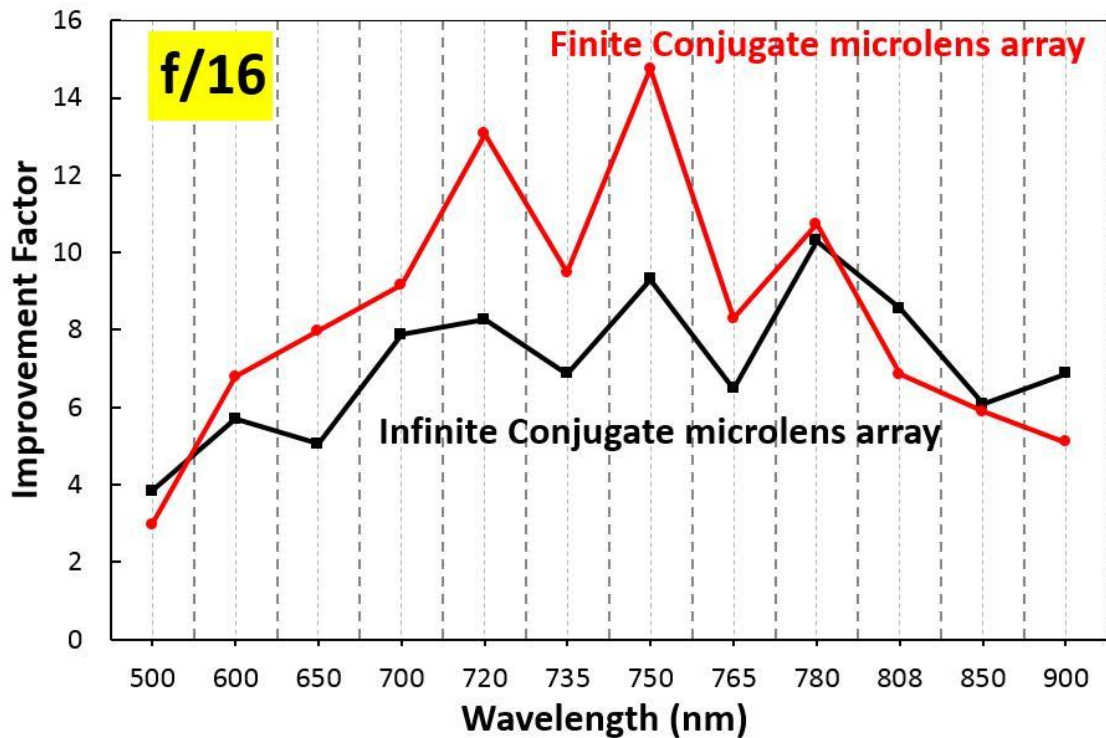


Figure 6.34. Comparison of the measured IF at a fixed f-number (f/16) for both SPAD array integrating the two different sets of diffractive microlenses, infinite (black) and finite (red) conjugate.

The peak wavelength at which the IF reached its maximum value for the two sets of microlenses, infinite (black line) and finite (red line) can be easily located in Figure 6.34. However, pronounced fringe behaviour (at different wavelengths) of the IF was observed for both microlens arrays. This observed IF trend could be explained by considering the microscopic structure of the composite microlens/SPAD assembly

which included the Si wafer (where SPAD devices are fabricated), a passivation (SiO_2) and an anti-reflection coating (Si_3N_4) layer, plus the optical cement layer (used for the bonding of the microlenses on to the SPAD array chip) and the microlens substrate (fused silica). By considering this structure which is schematically represented in Figure 6.35, simulations of the IF as a function of wavelength were performed by Dr. Andrew Waddie at Heriot-Watt University. Results of these simulations for both microlens arrays, infinite (red) and finite (blue) conjugate, were compared with experimental results (red and blue points for the infinite and finite conjugate) of the IF (Figure 6.36). In addition, simulations were carried out by taking into account the dispersion (change of refractive index with wavelength) model for all layers, with the exception of the optical cement for which the dispersion model was not known.

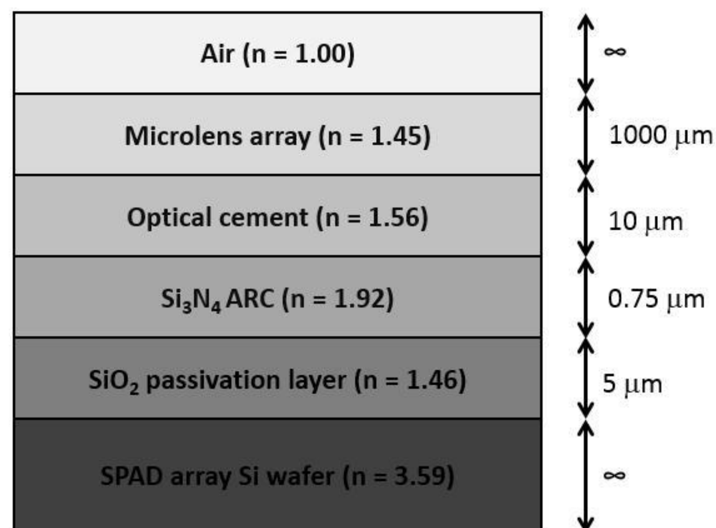


Figure 6.35. Schematic representation of the microscopic structure of the composite microlens/SPAD array. Starting from the bottom, the structure includes a Si layer where the 32 × 32 smart pixel are fabricated, then two layers of SiO_2 passivation and Si_3N_4 anti reflection coating (ARC) were grown during the CMOS process. On top of the SPAD array an optical cement for the bonding is placed and then the microlens array. For each layer, the thickness and refractive index used for the simulation are given.

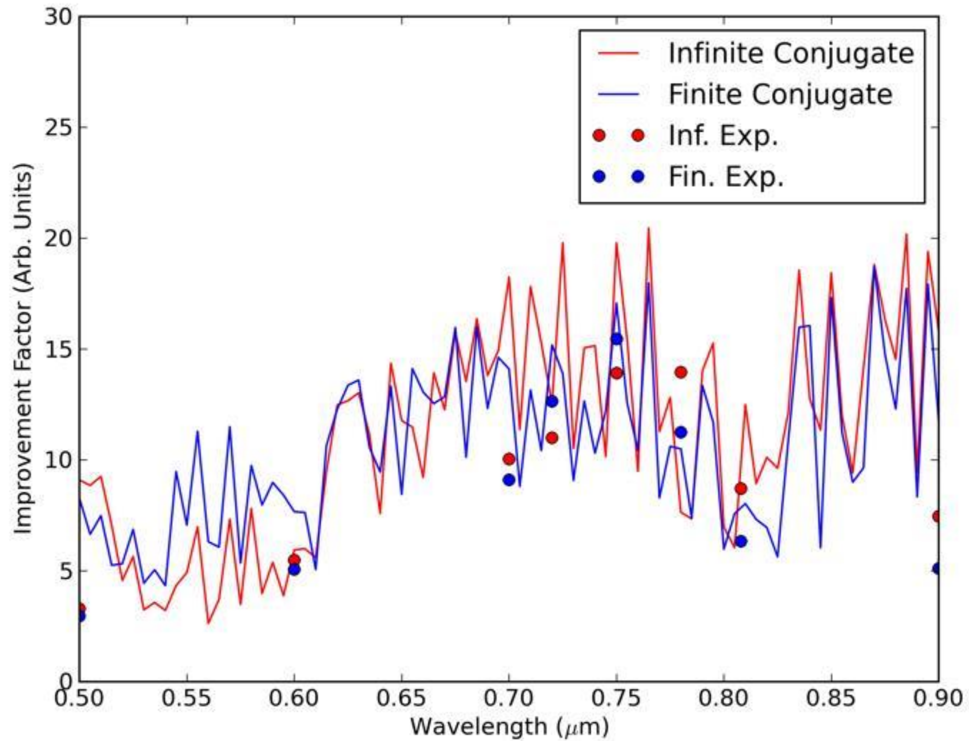


Figure 6.36. Comparison between the theoretical IF as a function of the wavelength for both microlens arrays, infinite (red) and finite (black), and the experimental values of the IF for the infinite (red dots) and finite (blue dots) conjugate. The simulations were integrated by considering the microscopic structure of the composite microlens/SPAD array shown in Figure 6.35.

It can be observed that the simulations match qualitatively, if not quantitatively, the fringe structure observed in the experimental results illustrated in Figure 6.36. Although this fringe behaviour of the IF could be due to the optical interference inside this multilayer structure, further analysis would be needed to understand its impact on the measured IF.

6.8 Spatial Uniformity of the SPAD arrays

Another important parameter to be evaluated during the characterisation of the SPAD arrays is the spatial uniformity of detection, in order to ascertain any degradation due to integration of the microlenses arrays. This parameter was evaluated by considering the Coefficient of Variation (CV) defined as the ratio of the standard deviation (σ) and the mean (μ) expressed as a percentage. For the measurements under investigation, the calculation of the CV aimed to describe the variability, and hence the uniformity, of the measured light intensity and the IF across the whole sensor.

Figures 6.37a-b - 6.39a-b show, as an example, images obtained for all three SPAD arrays (bare chip, infinite and finite conjugate microlenses) at the designed wavelength (808 nm) and f/16. All the images were taken under the same lighting conditions. SPAD arrays were operated in photon counting mode (see section 6.4.1) and two acquisition were performed for each sensor, one with the sensor in complete darkness and another illuminating the sensor at a specific wavelength. The two images were then subtracted to obtain the light intensity profile for each sensor and hot pixels were removed.

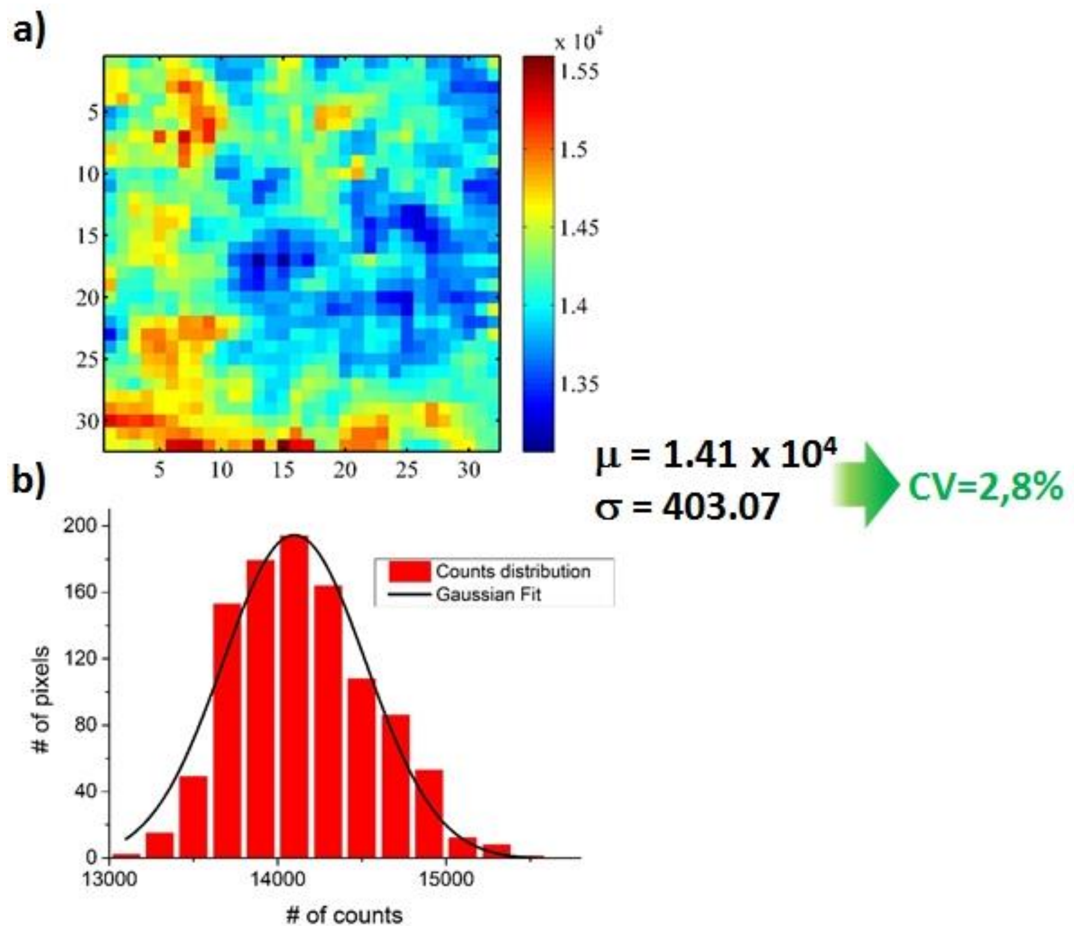


Figure 6.37. (a) Light intensity image for the SPAD array without microlenses acquired at the designed wavelength (808 nm) and f/16. The hot pixels were removed from the image by using a median filter. (b) Counts (or light intensity) distribution across the image shown in a) with the Gaussian fit (black line). The value of mean and standard deviation are also reported, demonstrating a variation across the chip of 2.8%.

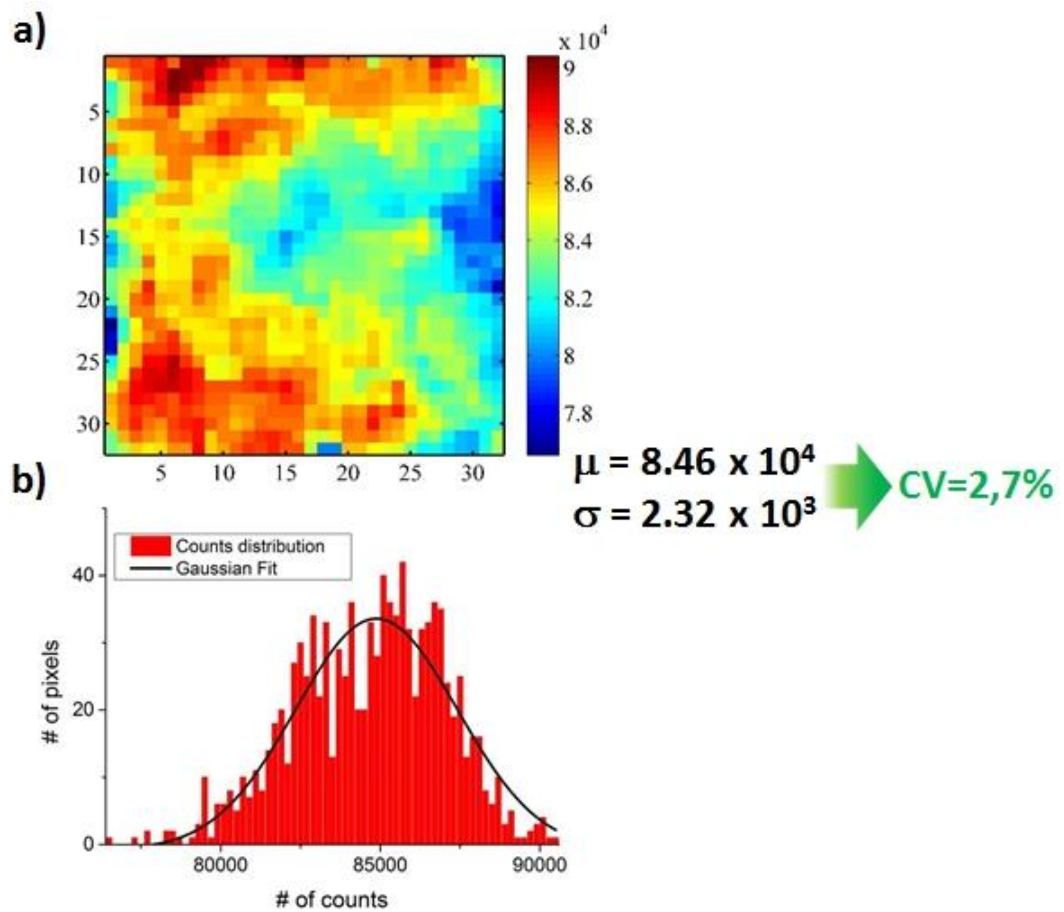


Figure 6.38. (a) Light intensity image for the SPAD array integrating the infinite conjugate diffractive microlenses acquired at the designed wavelength (808 nm) and f/16. The hot pixels were removed from the image by using a median filter. (b) Counts (or light intensity) distribution across the image shown in a) with the Gaussian fit (black line). The value of mean and standard deviation are also reported, demonstrating a variation across the chip of 2.7%.

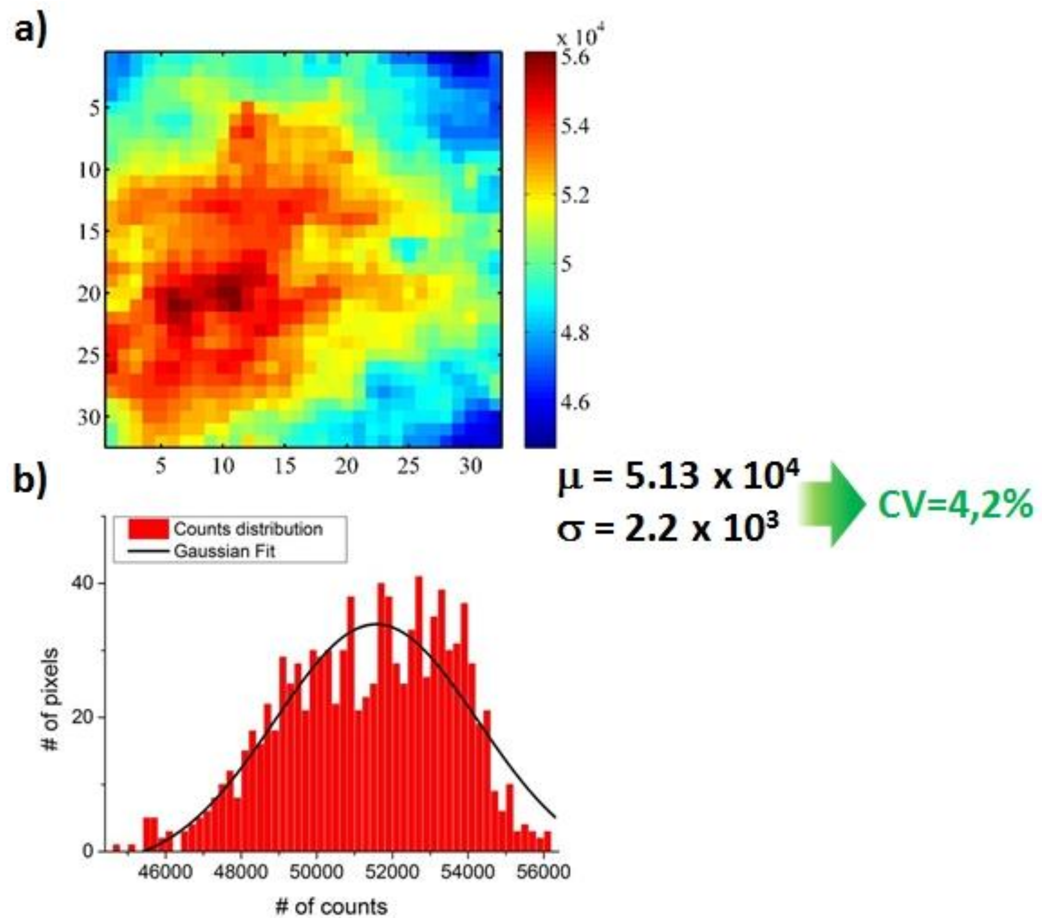
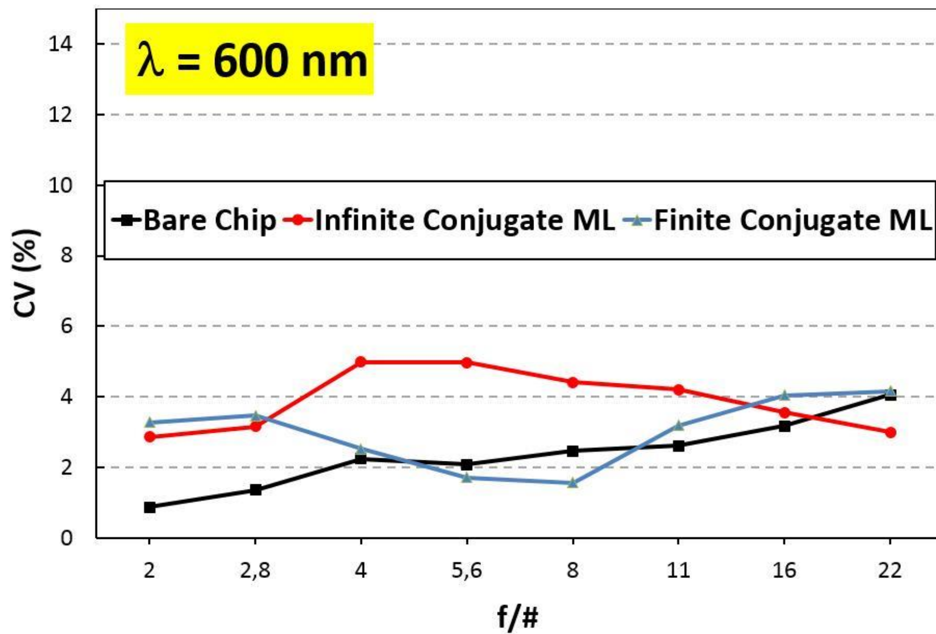


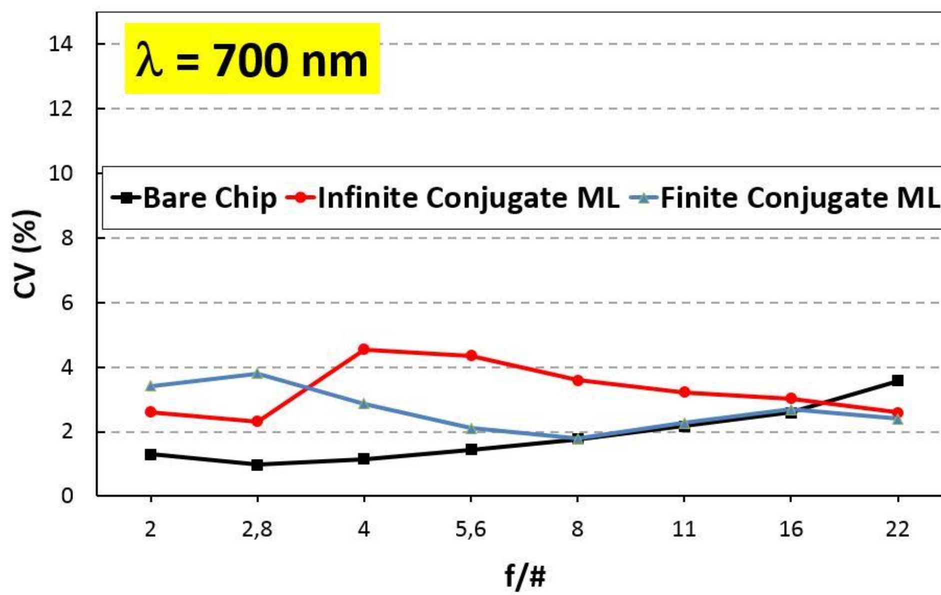
Figure 6.39. (a) Light intensity image for the SPAD array integrating the finite conjugate diffractive microlenses acquired at the designed wavelength (808 nm) and f/16. The hot pixels were removed from the image by using a median filter. (b) Counts (or light intensity) distribution across the image shown in a) with the Gaussian fit (black line). The value of mean and standard deviation are also reported, demonstrating a variation across the chip of 4.2%.

For each image shown in Figure 6.37a – 6.39a, the scale intensity for each array was adapted to enhance the visibility of the images because of the different improvement factors of each array, and hence higher number of counts. In fact, the increase in light intensity due to the microlens integration is clearly visible. The counts distribution (Figure 6.37b – 6.39b) as well as the value of mean and standard deviation are also reported, demonstrating a CV of 2.8%, 2.7% and 4.2% for the SPAD array without microlenses (bare chip), with infinite conjugate microlenses and with finite conjugate microlenses array, respectively at the designed wavelength and f/16.

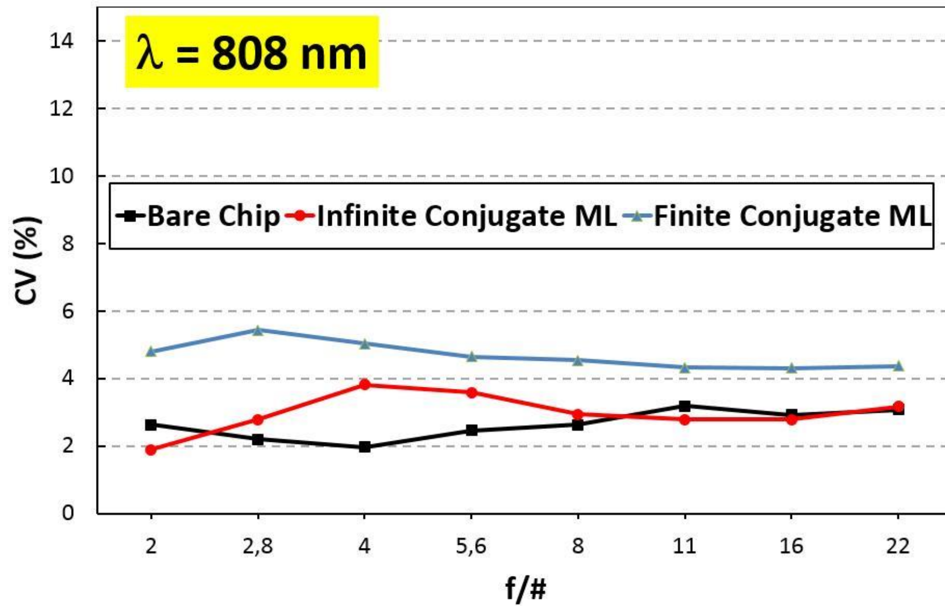
As in the case of the IF, the uniformity of all three SPAD arrays was measured at different f-numbers and three different wavelengths, 600 nm, 700 nm, and 808 nm. These results are illustrated in Figure 6.40a-c.



(a)



(b)



(c)

Figure 6.40. Spatial uniformity of detection measured as the Coefficient of Variation (CV) as a function of the f-number (from f/2 to f/22 with one-stop increment) for three different wavelengths, (a) 600 nm, (b) 700 nm, and (c) 808 nm for the bare chip (black), the infinite (red) and finite (light blue) conjugate diffractive microlenses array.

In general, all three SPAD arrays demonstrated a good uniformity, since the variation of the light intensity across the chip was less than 6% at all f-numbers for the three different wavelengths (600 nm, 700 nm, and 808 nm), as shown in Figure 6.40a-c. These results also demonstrated that microlens integration does not introduce any measurable degradation of the uniformity of the SPAD array. In addition, the robustness of the optical setup used for the array characterization was further confirmed since the CV only varied slightly in the f-number range between f/2 and f/22.

6.9 Future Outlook

Micro-optical elements can be refractive, diffractive or hybrid (refractive/diffractive). The right choice of the optical element for a specific optical problem depends on many parameters such as the spectrum of the light source, the optical task (beam shaping, imaging), the efficiency required, the application, etc. As already explained in this chapter, diffractive microlenses guarantee a 100% fill-factor for a square pixel, and can be fabricated by using well-established lithography techniques. In theory, it is possible to have diffractive lenses 100% efficient, but in practise the efficiency results around

90%. However, diffractive microlenses show a strong dispersion effects (section 6.5), and have a low tolerance angle for the incident beam.

A refractive microlens is schematically shown in Figure 6.43.

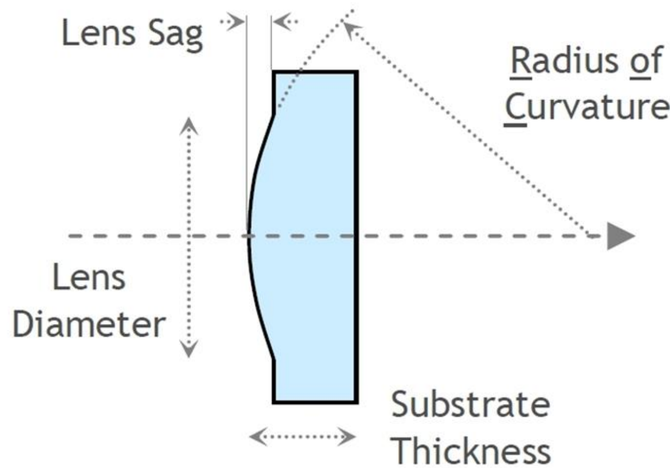


Figure 6.43. Schematic drawing of a refractive microlens with its most important parameters.

In general this lens shows achromatic behaviour, since the variation of its focal length (f_r) with respect to the wavelength is small and depends only on the material dispersion according to the following equation [32]:

$$f_r(\lambda) = \frac{R_C}{n(\lambda) - 1} \qquad (6.6)$$

where $n(\lambda)$ and R_C are the refractive index and radius of curvature, respectively. As a result, the IF should be less dependent on wavelength when compared to its diffractive counterpart. In addition, these microlenses show a higher efficiency (almost 100%) since there is no reduction in efficiency when changing wavelength and they do not produce unwanted orders. In addition, they are less dependent on the angle of the incident beam when compared to their diffractive counterpart. However, refractive microlenses cannot guarantee a 100% fill-factor (due to geometrical reasons), and are more difficult to fabricate with a precise focal length or aspheric shapes.

In terms of applications, a high IF is desirable mainly at low f-numbers of operation. As shown in section 6.7, the IF reached its maximum value at the highest f-numbers (f/16 and f/22) for both infinite and finite conjugate diffractive microlens. As already explained, diffractive microlenses are limited to f-number greater than f/3, since the maximum attainable f-number depends on the minimum attainable feature size in the lithography process. On the other hand, refractive microlenses can demonstrate f-

numbers above f/1 through careful design and fabrication of the lens sag. This faster behaviour can help to improve the IF at low f-number (see section 6.7.3, Figure 6.30) for which diffractive microlenses demonstrate some limitations.

6.10 Conclusions

In this chapter, the two main classes of Si SPAD (thin and thick junction) have been described underlining the main performance advantages and disadvantages. In particular, thin junction Si SPADs can be fabricated using a planar technology which is compatible with standard CMOS fabrication used in the IC industry. CMOS technology allows the integration of Si SPADs with the electronics required for photon-timing and photon-counting and hence SPAD array fabrication. However, CMOS SPAD arrays are limited in terms of fill-factor, since a fraction of the optical power, which is incident on the array, will be incident on the surrounding electronics.

In this work, an array of 32 × 32 SPADs with a 3.14 % fill-factor was used. Diffractive microlenses were used to concentrate the incoming light onto the active area of each pixel and hence to improve on the low fill-factor. In particular, two different sets of diffractive microlens arrays, infinite and finite conjugate, were designed for 808 nm, fabricated in a fused silica substrate, and integrated on the top surface of the 32 × 32 SPAD array. The characterisation of both microlenses arrays, in terms of improvement factor and spatial uniformity, was performed in a large spectral range (500 – 900 nm) at different f-numbers (from f/2 to f/22) by using a completely new approach based on a double telecentric imaging system.

The highest value of the IF of ~16 was demonstrated for a SPAD array with an integrated microlens. However, a shift was observed between the peak wavelength (780 nm and 750 nm for the infinite and finite conjugate microlenses, respectively) and the designed wavelength of 808 nm, and this might be due to tolerance errors during the microlens fabrication. A comparison of both microlens arrays was evaluated and the behaviour in the low and high f-number regime was explained.

Finally, the SPAD arrays, with and without microlenses, were evaluated in terms of the spatial uniformity of detection, and a variation between 2% and 6% was measured at different f-numbers. These results demonstrated that the microlenses did not add any degradation to the SPAD array and hence the good quality of both the fabrication and integration process was confirmed. Furthermore, a good uniformity was also an

indication of the robustness of the characterisation setup. These results provide a relevant and useful contribution to future research for all photon-counting applications which require very high detection efficiency combined with a very high frame-rate and picosecond timing resolution such as time-of-flight ranging, biomedical science (fluorescence lifetime imaging, positron emission tomography), and LIDAR.

6.11 Acknowledgements

I would like to thank my supervisor, prof. Gerald Buller, and my colleagues Dr. Aongus McCarthy and Dr. Ryan Warburton for their help and guidance on the experimental setup as well as for their valuable advice and discussion on the array characterisation.

Reference

[1] B. Markovic, S. Tisa, F. A. Villa, A. Tosi, and F. Zappa, “A High-Linearity, 17 ps Precision Time-to-Digital Converter Based on a Single-Stage Vernier Delay Loop Fine Interpolation,” *IEEE Trans. Circuits Syst. Regul. Pap.*, vol. 60, no. 3, pp. 557–569, Mar. 2013.

[2] C. Veerappan, J. Richardson, R. Walker, D.-U. Li, M. W. Fishburn, Y. Maruyama, D. Stoppa, F. Borghetti, M. Gersbach, R. K. Henderson, and E. Charbon, “A 160 x 128 single-photon image sensor with on-pixel 55ps 10b time-to-digital converter,” in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2011 IEEE International*, 2011, pp. 312–314.

[3] G. Intermite, R. E. Warburton, A. McCarthy, X. Ren, F. Villa, A. J. Waddie, M. R. Taghizadeh, Y. Zou, F. Zappa, A. Tosi, and G. S. Buller, “Enhancing the fill-factor of CMOS SPAD arrays using microlens integration,” 2015, vol. 9504, p. 95040J–95040J–12.

[4] E. Randone, G. Martini, M. Fathi, and S. Donati, “SPAD-array photoresponse is increased by a factor 35 by use of a microlens array concentrator,” in *IEEE LEOS Annual Meeting Conference Proceedings, 2009. LEOS ’09*, 2009, pp. 324–325.

[5] S. Donati, E. Randone, M. Fathi, J.-H. Lee, and E. Charbon, “Uniformity of concentration factor and back focal length in molded polymer microlens arrays,” in *2010 Conference on Lasers and Electro-Optics (CLEO) and Quantum Electronics and Laser Science Conference (QELS)*, 2010, pp. 1–2.

[6] J. M. Pavia, M. Wolf, and E. Charbon, “Measurement and modeling of microlenses fabricated on single-photon avalanche diode arrays for fill factor recovery,” *Opt. Express*, vol. 22, no. 4, pp. 4202–4213, Feb. 2014.

[7] F. Zappa, S. Tisa, A. Tosi, and S. Cova, “Principles and features of single-photon avalanche diode arrays,” *Sens. Actuators Phys.*, vol. 140, no. 1, pp. 103–112, Oct. 2007.

[8] M. Ghioni, S. Cova, A. Lacaita, and G. Ripamonti, “Epitaxial silicon single photon detector with reduced carrier diffusion effect and picosecond resolution,” in *Electron Devices Meeting, 1987 International*, 1987, vol. 33, pp. 452–455.

[9] M. Ghioni, S. Cova, A. Lacaita, and G. Ripamonti, “New silicon epitaxial avalanche diode for single-photon timing at room temperature,” *Electron. Lett.*, vol. 24, no. 24, p. 1476, 1988.

[10] A. Lacaita, M. Ghioni, and S. Cova, “Double epitaxy improves single-photon avalanche diode performance,” *Electron. Lett.*, vol. 25, no. 13, pp. 841–843, Jun. 1989.

[11] H. Dautet, P. Deschamps, B. Dion, A. D. MacGregor, D. MacSween, R. J. McIntyre, C. Trottier, and P. P. Webb, “Photon counting techniques with silicon avalanche photodiodes,” *Appl. Opt.*, vol. 32, no. 21, pp. 3894–3900, Jul. 1993.
[12] R. G. W. Brown, K. D. Ridley, and J. G. Rarity, “Characterization of silicon avalanche photodiodes for photon correlation measurements. 1: Passive quenching,” *Appl. Opt.*, vol. 25, no. 22, pp. 4122–4126, Nov. 1986.
[13] A. Gulinatti, I. Rech, F. Panzeri, C. Cammi, P. Maccagnani, M. Ghioni, and S. Cova, “New silicon SPAD technology for enhanced red-sensitivity, high-resolution timing and system integration,” *J. Mod. Opt.*, vol. 59, no. 17, pp. 1489–1499, Oct. 2012.
[14] G. S. Buller and R. J. Collins, “Single-photon generation and detection,” *Meas. Sci. Technol.*, vol. 21, p. 012002, Jan. 2010.
[15] C. Scarcella, A. Tosi, F. Villa, S. Tisa, and F. Zappa, “Low-noise low-jitter 32-pixels CMOS single-photon avalanche diodes array for single-photon counting from 300 nm to 900 nm,” *Rev. Sci. Instrum.*, vol. 84, no. 12, p. 123112, Dec. 2013.
[16] D. Bronzi, F. Villa, S. Tisa, A. Tosi, F. Zappa, D. Durini, S. Weyers, and W. Brockherde, “100000 Frames/s 64 x 32 Single-Photon Detector Array for 2-D Imaging and 3-D Ranging,” *IEEE J. Sel. Top. Quantum Electron.*, vol. 20, no. 6, pp. 354–363, Nov. 2014.
[17] C. Niclass, C. Favi, T. Kluter, M. Gersbach, and E. Charbon, “A 128x128 Single-Photon Image Sensor with Column-Level 10-bit Time-to-Digital Converter Array,” *J. Solid-State Circuits*, vol. 43, no. 12, pp. 2977–2989, 2008.
[18] M. Gersbach, Y. Maruyama, R. Trimananda, M. W. Fishburn, D. Stoppa, J. A. Richardson, R. Walker, R. Henderson, and E. Charbon, “A Time-Resolved, Low-Noise Single-Photon Image Sensor Fabricated in Deep-Submicron CMOS Technology,” *IEEE J. Solid-State Circuits*, vol. 47, no. 6, pp. 1394–1407, Jun. 2012.
[19] F. Villa, R. Lussana, D. Bronzi, S. Tisa, A. Tosi, F. Zappa, A. Dalla Mora, D. Contini, D. Durini, S. Weyers, and W. Brockherde, “CMOS Imager With 1024 SPADs and TDCs for Single-Photon Timing and 3-D Time-of-Flight,” *IEEE J. Sel. Top. Quantum Electron.*, vol. 20, no. 6, pp. 364–373, Nov. 2014.
[20] D. Bronzi, F. Villa, S. Bellisai, S. Tisa, G. Ripamonti, and A. Tosi, “Figures of merit for CMOS SPADs and arrays,” 2013, vol. 8773, pp. 877304–877304–7.
[21] F. Villa, D. Bronzi, Y. Zou, C. Scarcella, G. Boso, S. Tisa, A. Tosi, F. Zappa, D. Durini, S. Weyers, U. Paschen, and W. Brockherde, “CMOS SPADs with up to 500 μm diameter and 55% detection efficiency at 420 nm,” *J. Mod. Opt.*, vol. 61, no. 2, pp. 102–115, Jan. 2014.
[22] F. Villa, B. Markovic, S. Bellisai, D. Bronzi, A. Tosi, F. Zappa, S. Tisa, D. Durini, S. Weyers, U. Paschen, and W. Brockherde, “SPAD Smart Pixel for Time-of-Flight and Time-Correlated Single-Photon Counting Measurements,” *IEEE Photonics J.*, vol. 4, no. 3, pp. 795–804, Jun. 2012.
[23] M. Gersbach, Y. Maruyama, E. Labonne, J. Richardson, R. Walker, L. Grant, R. Henderson, F. Borghetti, D. Stoppa, and E. Charbon, “A parallel 32x32 time-to-digital converter array fabricated in a 130 nm imaging CMOS technology,” in *Proceedings of ESSCIRC, 2009. ESSCIRC ’09*, 2009, pp. 196–199.
[24] J. Dukovic, S. Ramaswami, S. Pamarthy, R. Yalamanchili, N. Rajagopalan, K. Sapre, Z. Cao, T. Ritzdorf, Y. Wang, B. Eaton, R. Ding, M. Hernandez, M. Naik, D. Mao, J. Tseng, D. Cui, G. Mori, P. Fulmer, K. Sirajuddin, J. Hua, S. Xia, D. Erickson, R. Beica, E. Young, P. Kusler, R. Kulzer, S. Oemardani, H. Dai, X. Xu, M. Okazaki, K. Dotan, C. Yu, C. Lazik, J. Tran, and L. Luo, “Through-silicon-via technology for 3D integration,” in *Memory Workshop (IMW), 2010 IEEE International*, 2010, pp. 1–2.

[25] S. J. Koester, A. M. Young, R. R. Yu, S. Purushothaman, K. N. Chen, D. C. La Tulipe, N. Rana, L. Shi, M. R. Wordeman, and E. J. Sprogis, “Wafer-level 3D integration technology,” *IBM J. Res. Dev.*, vol. 52, no. 6, pp. 583–597, Nov. 2008.
[26] F. Zappa and A. Tosi, “MiSPIA: microelectronic single-photon 3D imaging arrays for low-light high-speed safety and security applications,” 2013, p. 87270L.
[27] M. A. Itzler, M. Entwistle, M. Owens, K. Patel, X. Jiang, K. Slomkowski, S. Rangwala, P. F. Zalud, T. Senko, J. Tower, and J. Ferraro, “Design and performance of single photon APD focal plane arrays for 3-D LADAR imaging,” 2010, vol. 7780, p. 77801M–77801M–15.
[28] S. Donati, G. Martini, and M. Norgia, “Microconcentrators to recover fill-factor in image photodetectors with pixel on-boardprocessing circuits,” *Opt. Express*, vol. 15, no. 26, pp. 18066–18075, Dec. 2007.
[29] J. Jahns and S. J. Walker, “Two-dimensional array of diffractive microlenses fabricated by thin film deposition,” *Appl. Opt.*, vol. 29, no. 7, pp. 931–936, Mar. 1990.
[30] B. C. Kress and P. Meyrueis, *Applied Digital Optics: From Micro-optics to Nanophotonics*. John Wiley & Sons, 2009.
[31] M. R. Taghizadeh, P. Blair, B. Layet, I. M. Barton, A. J. Waddie, and N. Ross, “Design and fabrication of diffractive optical elements,” *Microelectron. Eng.*, vol. 34, no. 3–4, pp. 219–242, Dec. 1997.
[32] H. P. Herzig, *Micro-Optics: Elements, Systems And Applications*. CRC Press, 1997.

Chapter 7 - Conclusions and Future work

Single-photon counting is used in a range of emerging applications, with the range of these applications expanding as detector and data acquisition hardware improves. Si SPADs are increasingly becoming the standard single-photon detectors at wavelengths below 1 μm. At longer wavelengths, the most mature technology in terms of semiconductor devices is currently InGaAs/InP SPADs. Although both material systems are available commercially from a number of sources, Si SPADs can also be fabricated by using the standard CMOS fabrication processes used in the microelectronic industry. The same is not true for the InGaAs/InP SPAD which requires a dedicated fabrication technology which is not compatible with Si fabrication processes.

Near-infrared detection is of interest in many applications. For example, in data communication because of the low attenuation windows of standard optical fibres, or in laser ranging where these spectral regions provide lower solar background noise, lower atmospheric attenuation, as well as relaxed eye-safety thresholds. As a consequence, there is an increasing demand for detectors working in this spectral range. If this demand can be used in conjunction with those from the silicon photonics community, then the combination of silicon and germanium might be adopted effectively for near-infrared detection.

In terms of photon-counting, InGaAs/InP SPADs have the major drawback of afterpulsing that affects the maximum count rate possible and the overall noise level. On the other hand, Si is often the best material choice for photodiodes using avalanche multiplication, and afterpulsing is not a major problem in high purity silicon structures. The investigation of new materials for near-infrared photon counting that can take advantage of silicon for its good avalanche multiplication properties, low afterpulsing, CMOS compatibility, leads directly to the choice of germanium as an absorbing layer to extend the operational wavelength of silicon-based devices.

The goal of this thesis was to design and demonstrate a SPAD working in the near-infrared wavelength region by using the Ge-on-Si heteroepitaxial system. The device design was based on a separate absorption, charge and multiplication (SACM) structure, since the photon absorption and carrier multiplication were spatially separated and could be individually optimised. Device simulations suggested that charge doping concentrations of 1.5 and 2×10^{17} cm^{-3} satisfy the design criteria in terms of electric

field and temperature. Different Ge-on-Si device generations were grown by RP-CVD at the University of Warwick, and mesa geometry devices were fabricated at the University of Glasgow. The single-photon performance of 25 μm and 50 μm diameter Ge-on-Si SPAD devices were evaluated at temperatures ranging from 100 K to 150 K. The lowest DCR ranging between 1 - 6 Mcs^{-1} was measured on a 25 μm diameter device at 100 K and different relative excess biases. Under the same conditions, the highest SPDE ranging between 2 % to 4 % was also measured at a wavelength of 1310 nm. These values led to the lowest NEP of $\sim 1 \times 10^{-14}$ $WHz^{-1/2}$ of Ge-on-Si SPADs reported in the scientific literature. Furthermore, a potential major advantage of these devices compared to the InGaAs/InP SPADs could be the reduced afterpulsing. To investigate afterpulsing, the gating frequency was increased from 1 kHz to 1 MHz and only a slight increase (by nearly a factor of 2) in the DCR was observed. Further detailed analysis will be carried out for the next generation of devices by using a time-correlated carrier counting method in order to better understand the effect of afterpulsing on these structures.

Although a prototype Ge-on-Si SPAD device has been demonstrated, future developments should consider the ultimate goal of obtaining an integrated Ge-on-Si SPAD with low dark count rate, higher operating temperatures, improved single-photon detection efficiency, high reliability and compatibility with industrial Si microelectronic processes. To achieve these requirements, efforts are required in several key aspects of these detectors:

a) A uniform electric field profile in the Si multiplication layer is highly desirable with the aim of improving the single-photon performance of these devices. In view of this, a promising approach is the use of an antimony (Sb) doped Si substrate where initial trials have demonstrated a very low background doping in the multiplication layer (in line with the designed doping concentration of 1×10^{15} cm^{-3}). Fabrication of devices grown using this approach is now ongoing.

b) Several different research groups have shown that the main contribution to the leakage current of small active area Ge-on-Si devices can be attributed to the current perimeter component. It was possible to draw a similar conclusion for the dark current of the mesa geometry Ge-on-Si SPAD devices designed and characterised in this work. Different passivation techniques such as SiO_2, GeO_2, and atomic layer deposition (ALD) are currently under investigation with the

aim of mitigating the current contribution from the sidewalls of the devices. However, it is challenging to optimally passivate each of the two different materials, Ge and Si, and hence evaluate which passivation can give the best trade-off in terms of performance. Additionally, there is a different electric field profile through the Ge-on-Si structure, and the high electric field region is in the Si multiplication layer. In view of reducing the contribution from the high-field Si layer, a partially etched device (Figure 7.1) has been proposed. By analysing the dark current and performance of this new structure geometry, it will be possible to further understand the impact of the Si multiplication layer on the overall device characteristics. The fabrication of devices based on the proposed geometry is ongoing at the University of Glasgow.

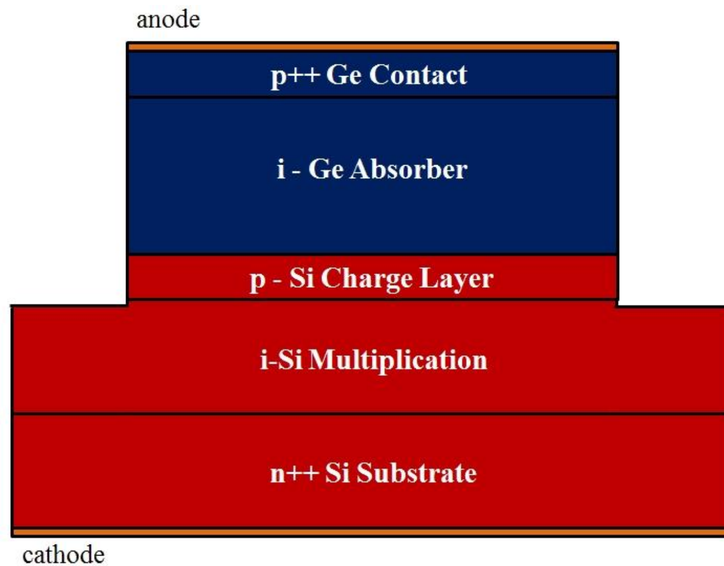


Figure 7.1. Partially etched Ge-on-Si SPAD structure which will be fabricated at the University of Glasgow. This structure will be used to evaluate the contribution of the high electric field Si multiplication region on the device performance.

c) A reduction of threading dislocation density at the Si-Ge interface can be achieved by combining selective epitaxial growth (SEG) and cyclic thermal annealing. When SEG is applied to a small area, threading arms of misfit dislocations propagate at 45° angle from the Si substrate, and terminate at the edge of the growth area (oxide sidewalls) [1]. This process, combined with cyclic thermal annealing, reduces the overall threading dislocation density and leaves a defect-free Ge surface, as pointed out by Wang *et al.* and Langdo *et al.* [2], [3] (refer to section 3.4.1.2 for details). This decrease should help to reduce

the leakage current of the device, and may also reduce the DCR of the device and increase the SPDE. Furthermore, SEG also gives the possibility of selectively introducing Ge on CMOS technology for large scale integration with other Si photonics components.

d) InGaAs/InP SPADs as well as Si SPADs are typically fabricated in a planar geometry, with the p-n junction formed by post-growth dopant diffusion or ion implantation. Conversely, the Ge-on-Si SPAD devices illustrated in this thesis are based on mesa geometry. There are several advantages of using a planar technology, with the main and most immediate issue being that the sidewall contribution to the dark current and the problems related to the effective sidewall passivation of Si and Ge are considerably reduced.

e) In view of a CMOS compatible Ge-on-Si SPAD, the thermal budget required for the growth of these devices including the high temperatures needed during the two-step Ge growth and subsequent cyclic annealing, are far beyond the thermal budget of CMOS. As a consequence, further improvements in the hetero-epitaxial growth and a reduction of the annealing temperatures are therefore required for the development of low temperature Ge growth technique with low dislocation density Ge films.

f) In order to increase the detection efficiency of the Ge-on-Si SPAD in the near-infrared wavelength range, the use of waveguide geometry Ge-on-Si SPADs is under investigation. Such a geometry will provide direct integration with Si-based waveguide circuitry, for example as used in linear optical computing [4]. Two different designs have been proposed for future fabrication and epitaxial growth. In the first approach, a waveguide on top of the Ge-on-Si SPAD is considered. Both Si_3N_4 or SiO_xN_y waveguides can be used in this geometry. Although the structure requires a further planarization step, it can be used to give a direct comparison with a normal-incidence Ge-on-Si SPAD in terms of detection efficiency. Additionally, further heteroepitaxial growth is not required since the existing wafers could be used for device fabrication. The second geometry relies on a Si-on-insulator (SOI) wafer using a Si waveguide which is fabricated on the device layer of the SOI wafer. Modelling of the waveguide integrated Ge-on-Si SPAD is being performed to evaluate the overall coupling efficiency between the Si waveguide and Ge absorber layer.

Si SPADs are a mature technology for detecting single-photons in the visible and near-infrared spectral range. The main advantages introduced by the fabrication of these devices using the standard CMOS technology were underlined. CMOS technology allowed the integration of Si SPADs with the electronics required for photon-timing and photon-counting and hence SPAD array fabrication. The CMOS-based Si SPAD arrays used in this work suffered from a low fill-factor of 3.14% of the detector photo-sensitive areas compared to the overall detector pixel area. To recover this loss of sensitivity, two different diffractive optical microlens arrays, infinite and finite conjugate, were integrated onto 32×32 SPAD arrays, fabricated using a 0.35 μm CMOS technology process. To the best of the author's knowledge, a full characterisation of SPAD arrays integrating microlenses over a large spectral range between 500 nm and 900 nm and using different f-numbers from f/2 to f/22, in terms of improvement factor and spatial uniformity of detection was demonstrated for the first time. Both microlens arrays showed high IFs across the whole wavelengths range, with its maximum value of ~16 at a wavelength of 750 nm and ~14 at a wavelength of 780 nm for the finite and infinite conjugate microlenses array, respectively. These values of the IF represented the highest values ever measured for a CMOS-based Si SPAD array integrating diffractive microlens arrays. In terms of spatial uniformity of detection, a variation between 2 % and 6 % was measured for all three SPAD arrays tested (bare chip, SPAD arrays integrating the infinite and finite conjugate microlens arrays) at different wavelengths and across the whole f-numbers range. Future work on CMOS SPAD arrays integrating microlens is currently ongoing, and will consider the use of refractive microlens arrays to recover the low fill-factor. In theory, it is possible to fabricate refractive microlens with very low f-numbers through a careful design and fabrication. This faster behaviour may contribute to further enhance the IF in the low f-number region (between f/2 and f/8) which is desirable in many practical applications.

7.1 Reference

[1] C. T. DeRose, D. C. Trotter, W. A. Zortman, A. L. Starbuck, M. Fisher, M. R. Watts, and P. S. Davids, "Ultra compact 45 GHz CMOS compatible Germanium waveguide photodiode with low dark current," *Opt. Express*, vol. 19, no. 25, p. 24897, 2011.

[2] G. Wang, R. Loo, S. Takeuchi, L. Souriau, J. C. Lin, A. Moussa, H. Bender, B. De Jaeger, P. Ong, W. Lee, M. Meuris, M. Caymax, W. Vandervorst, B. Blanpain, and M. M. Heyns, "Fabrication of high quality Ge virtual substrates by selective epitaxial growth in shallow trench isolated Si (001) trenches," *Thin Solid Films*, vol. 518, no. 9, pp. 2538–2541, 2010.

[3] T. A. Langdo, C. W. Leitz, M. T. Currie, E. A. Fitzgerald, A. Lochtefeld, and D. A. Antoniadis, “High quality Ge on Si by epitaxial necking,” *Appl. Phys. Lett.*, vol. 76, no. 25, p. 3700, 2000.

[4] M. G. Thompson, A. Politi, J. C. F. Matthews, and J. L. O’Brien, “Integrated waveguide circuits for optical quantum computing,” *IET Circuits Devices Syst.*, vol. 5, no. 2, pp. 94–102, 2011.