CMOS SINGLE-PHOTON AVALANCHE DIODES TOWARDS POSITRON EMISSION TOMOGRAPHY IMAGING APPLICATIONS

CMOS SINGLE-PHOTON AVALANCHE DIODES TOWARDS POSITRON EMISSION TOMOGRAPHY IMAGING APPLICATIONS

By

WEI JIANG

B. Sc. XI'AN JIAOTONG UNIVERSITY, 2006

M. A. Sc. SHANGHAI JIAOTONG UNIVERSITY, 2009

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree of Doctor of Philosophy

McMaster University © Copyright by Wei Jiang, August 2021

McMaster University
Hamilton, Ontario

Doctor of Philosophy (2021)
(School of Biomedical Engineering)

TITLE: CMOS Single-Photon Avalanche Diodes Towards Positron Emission Tomography Imaging Applications

AUTHOR: Wei Jiang,
B. Sc. Xi'an Jiaotong University, Xi'an, China
M. A. Sc. Shanghai Jiaotong University, Shanghai, China

SUPERVISOR: Dr. M. Jamal Deen

NUMBER OF PAGES: xxv, 166

Lay Abstract

Positron emission tomography (PET) imaging is a powerful tool for diagnosis and assessment of cancers and tumors in the clinical field. Due to their capabilities of detecting even a single photon, excellent timing resolution, and their compatibility with magnetic fields to build PET/MRI (magnetic resonance imaging) multimodal imaging systems; single-photon avalanche diodes (SPADs) become the most promising sensor technology for PET imaging applications. SPADs fabricated in standard complementary metal-oxide-semiconductor (CMOS) technologies allow for a lower manufacturing cost and present the potential to integrate with other CMOS circuits to form a complete imaging system. In this thesis, random telegraph signal noise in SPADs is investigated first. Then, the poly gate is used in the design of an n^+/p-well SPAD to improve the noise performance. In addition, a compact and high-speed SPAD pixel is designed and fabricated using an advanced standard CMOS process. Thanks to the fast quench and reset circuit, the SPAD pixel achieves a very short quenching time and a high-count rate. Finally, a differential quench and reset (QR) circuit consisting of two QR circuits on both the cathode and anode to quench and reset the SPAD through both terminals is proposed and studied.

Abstract

Single-photon avalanche diodes' (SPADs) capabilities of detecting–even a single photon with excellent timing resolution and compatibility with strong magnetic fields make them the most promising sensor for positron emission tomography imaging systems. With the advancements of silicon fabrication techniques, SPADs designed in standard planar complementary metal-oxide-semiconductor (CMOS) processes show competitive performance and a lower manufacturing cost. Additionally, CMOS SPADs have the potential for monolithic integration with other CMOS signal conditioning and processing circuits to achieve simple, low-cost, and high-performance imaging solutions. This work targets the design and optimization of SPAD sensors to improve their performance using low-cost standard CMOS technologies.

Firstly, a detailed review on the SPADs in recent literature is presented. Then, the random telegraph signal (RTS) noise is investigated based on n^+/p-well SPADs fabricated in a standard 130 nm CMOS process. Through the measurements and analysis, the RTS noise of a SPAD is found to correlate with its dark count rate and afterpulsing. Next, we design n^+/p-well SPADs with field poly gates to improve the noise performance. Furthermore, a SPAD pixel, consisting of a p^+/n-well SPAD and a compact and high-speed active quench and reset circuit is designed and fabricated in a standard TSMC 65 nm CMOS process. The post-layout simulations show that this pixel achieves a short 0.1 ns quenching time and a 3.35 ns minimum dead time. The measurement results show that the SPAD pixel has a dark count rate of 21 kHz, a peak photon detection probability of 23.8% at a 420 nm wavelength and a timing jitter of 139 ps using a 405 nm pulsed laser when the excess voltage is set to 0.5 V. Due to the short quenching time, almost no afterpulsing is observed even at a low operating temperature of -35 °C. Finally, a new differential quench and reset (QR) circuit consisting of two QR circuits on both the cathode and anode to quench and reset the SPAD through both terminals is proposed to reduce the reset time, to increase the count rate, to reduce the afterpulsing and to reject the common-mode noise.

Acknowledgements

First and foremost, I would like to express my deepest gratitude and appreciation to my supervisor, Dr. M. Jamal Deen, for his constant support, invaluable guidance, and insightful technical discussions throughout my graduate study and this research work. It has been a great honor for me to be a student under the supervision of such an accomplished and outstanding professor, researcher and academician. Dr. Deen's extensive technical knowledge and theoretical expertise in the design of integrated circuits, semiconductor devices and photodetectors guided me to find potential research topics in this field. He has taught me how to do strong research from many aspects such as: how to define the research topic, how to implement the research ideas, how to plan the experiments, and how to present and report the results. He also has spent a lot of time in teaching me skills for technical writing, correcting the mistakes in my manuscripts, and guiding me in responding to reviewers' comments in an effective way. I always feel amazed, and I appreciated his patience and carefulness during the process of manuscript corrections, which helped me to find even very tiny errors. Additionally, I also have learnt a lot from him regarding skills for communications and presentations, such as: how to present your research to others who are not in your research field, how to ask meaningful questions, how to be an active listener, and how to answer the questions from the audience in a precise and concise way. Moreover, as the teaching assistant for Dr. Deen's course (CE 4EK4), I have learnt many teaching skills and strategies like: making teaching plans, supervision of labs and projects, how to communicate with students effectively, and how to prepare exams. Dr. Deen's door is always open for discussion and he is always ready to help with any issues I encountered in not only my research work but also my personal life. I feel incredibly thankful and grateful to work with him.

In addtion, I would like to sincerely thank my committee members, Dr. Shahram Shirani and Dr. Qiyin Fang, for their valuable suggestions and feedback on my research over the past four years and for taking time to review my thesis. I also would like to express

my sincere thanks to Dr. Chih–Hung Chen for his constructive advice and encouragement during the course of my research work.

I also sincerely thank Dr. Deen's former student, Dr. Darek Palubiak, who helped me to get familiar with this research field in the beginning of my graduate study. He took his own time to show me the measurement setup for time-to-digital converters and to answer my questions timely by emails several times. I am very thankful for his kindness and great help.

I am also privileged to work with some great researchers in Dr. Deen's team: Dr. Arif U. Alam, Dr. Si Pan, Dr. Mrwan Alayed, Dr. Mahdi Naghshvarianjahromi, Dr. Sumit Majumder, Dr. Ahmed Elsharabasy, Yamn Chalich, Ryan Scott, Abu Ilius Faisal, Hytham Afifi, Junaid Siddiqui, Mahtab Taheri, and Sophini Subramaniam. I would like to thank them for their valuable suggestions and feedback on my research work during our group meetings. Yamn Chalich and Ryan Scott, these two amazing and talented team members deserve special acknowledgement for our smooth and successful collaborations. They helped me a lot during my research work, and I have learnt a lot from them. I am very lucky and thankful to have them to work with as a team for the chip design projects. I also would like to express my special gratitude to Dr. Sumit Majumder. It was a very pleasant experience to collaborate with him on the patient monitoring system project.

I am also very grateful to the technical staff in the ECE department. Joe Peric, Dan Manolescu and Ron Harwood helped a lot in setting up the environment of our CAD design tools and resolving technology licensing issues. Tyler Ackland assisted in purchasing of electronic components and other research material, and in assembling several printed circuit boards for our measurements. His high working morale, professional skills and hands-on experience with electronics, allowed me to progress my research more effectively and efficiently. I also would like to thank BME Administrative Assistant, Jane Mah, for dealing with all administrative issues I encountered in my graduate study.

I would like to thank CMC Microsystems for the assistance provided by its engineering team during the chip designs and for arranging the fabrications of all chips used in this research work.

Last but not the least, I also sincerely thank my parents and sisters for their support and encouragement during this academic journey. It is difficult for me to find the words to express my gratitude to my dear wife, Fang Ye, for her unconditional support, love and sacrifice these years. I also would like to express my thanks to my daughter and my son. They are my sources of joy after the hard work, and also the comforts when I face obstacles in my life. This thesis is dedicated to them.

Table of Contents

Lay Abstract iii
Abstract iv
Acknowledgements v
Table of Contents viii
List of Figures xii
List of Tables xix
List of Abbreviations xx
List of Symbols xxiii
Declaration of Academic Achievement xxv
Chapter 1 Introduction 1
1.1. Positron Emission Tomography 2
1.1.1. Physics of Positron Emission Tomography 2
1.1.2. Time-of-flight (ToF) PET 5
1.1.3. PET detector 6
1.2. Photosensors for PET Applications 7
1.2.1. Photomultiplier Tubes (PMTs) 7
1.2.2. Avalanche Photodiode (APD) 10
1.2.3. Silicon Photomultiplier (SiPM) 12
1.2.4. Cadmium Zinc Telluride (CZT) Detectors 15
1.3. State-of-the-art PET systems 19
1.3.1. PET/CT systems 19
1.3.2. PET/MRI systems 21
1.4. Research Challenges and Motivation 23
1.4.1. Improve the photon detection efficiency (PDE) 24
1.4.2. Lower the dark noise 25
1.4.3. Improve the timing performance 26

1.5. Research Contributions 29
1.6. Thesis Organization 32
Chapter 2 Performance Parameters for SPAD and Literature Review 34
2.1. Key Performance Parameters 34
2.1.1. Breakdown Voltage 34
2.1.2. Dark Count Rate 36
2.1.3. Afterpulsing 37
2.1.4. Dead Time and Count Rate 38
2.1.5. Photon Detection Efficiency 40
2.1.6. Timing Jitter 41
2.2. CMOS SPAD Design Considerations 42
2.2.1. Junctions 42
2.2.2. Guard Ring Structures 44
2.2.3. Quench and Reset Circuit 48
2.3. Review of State-of-the-art SPADs 54
2.4. Conclusions 55
Chapter 3 Random Telegraph Signal in n^+/p-well CMOS SPADs 58
3.1. Introduction 59
3.2. SPAD Structures and the Breakdown Voltage 61
3.2.1. SPAD Structure 61
3.2.2. Breakdown Voltage 62
3.3. RTS Noise Measurement 63
3.3.1. Measurement Details 63
3.3.2. RTS Noise Versus Biasing Voltage 66
3.3.3. RTS Noise Versus Temperature 67
3.4. DCR And AP Measurement 69
3.4.1. Measurement Details 69
3.4.2. DCR and AP Versus Biasing Voltage 70
3.4.3. DCR and AP Versus Temperature 71

3.5. RTS Noise Correlation with DCR and AP....72
3.5.1. Analytical Model for the Dimension of Defects....72
3.5.2. DCR AND AP comparison....75
3.6. Conclusions....77
Chapter 4 Improved Noise Performance of CMOS Poly Gate SPAD 79
4.1. Introduction....79
4.2. SPAD Design....81
4.2.1. SPAD Structure....81
4.2.2. TCAD Simulation....82
4.3. Experimental Results and Discussions....83
4.3.1. Breakdown Voltage....85
4.3.2. Dark Count Rate (DCR) and Afterpulsing (AP)....85
4.3.3. Performance Summary and Comparison....95
4.4. Conclusions....96
Chapter 5 High-speed SPAD Pixel with Active Quench and Reset Circuit 98
5.1. Introduction....98
5.2. SPAD Design....99
5.2.1. SPAD Structure....99
5.2.2. TCAD Simulation....101
5.2.3. SPAD model parameters....101
5.3. Active Quench and Active Reset Circuit Design....105
5.3.1. Operation Principle....105
5.3.2. Hold-off Time Controlling Circuit....108
5.3.3. Pulse Generation Circuit....110
5.3.4. Monostable Circuit....110
5.3.5. Layout....111
5.4. SPAD Pixel Characterizations and Discussions....112
5.4.1. Breakdown Voltage....112
5.4.2. Dark Count Rate and Afterpulsing....113

5.4.3. Photon Detection Probability 114
5.4.4. Timing Jitter 116
5.4.5. Summary and Comparison 117
5.5. Conclusions 118
Chapter 6 Differential Quench and Reset Circuit for SPADs 120
6.1. Introduction 120
6.2. Operational Principle and Simulations 123
6.2.1. Differential Passive Quench and Reset 123
6.2.2. Simulations 127
6.3. Measurements and Discussions 129
6.3.1. Measurement Setup 129
6.3.2. Waveforms of the SPQR and DPQR Outputs 131
6.3.3. DCR and AP 133
6.3.4. Other Discussions 135
6.4. Conclusions 139
Chapter 7 Conclusions and Future Work 141
7.1. Conclusions 141
7.2. Future Work 145
References 151
Appendix A Copyright Permission 161

List of Figures

Figure 1-1: The basic principle of a positron emission tomography (PET) system: A PET detector ring detects a pair of gamma photons with an energy of 511 keV (red arrows) which results from the annihilation of an electron with a positron emitted by the radiotracer (FDG). 3
Figure 1-2: Detection flow of a positron emission tomography (PET) system. 4
Figure 1-3: Concept of time-of-flight positron emission tomography (ToF PET): (a) illustration of a detector ring detecting pairs of gamma photons from the annihilation events with (green) and without (red) the ToF technique; (b) The probability distribution of the annihilation position along the LoR in ToF PET; (c) The equal probability of the annihilation position along the LoR in non-ToF PET. 5
Figure 1-4: Positron emission tomography (PET) detector: (a) Structure of a PET detector; (b) Detection flow chart of a PET detector. 7
Figure 1-5: Principle of the photomultiplier tube (PMT): (a) Simplified conventional structure; (b) Simplified high-voltage biasing network. 8
Figure 1-6: Illustration of a microchannel plate photomultiplier tube (MCP-PMT) (The microchannels are usually made of highly resistive materials (e.g. lead glass cladding) with the inner wall coated by high secondary emission materials such as MgO and Al_2O_3). 9
Figure 1-7: The impact ionization process. Here, α_n^{-1} is the average distance between each electron multiplication event, while V_{total} is the applied reverse bias plus the built-in potential.. 11
Figure 1-8: Principle of SPAD operation: (a) Avalanche breakdown process in a reverse biased p-n junction; (b) I-V characteristic representation of SPAD operation. 13
Figure 1-9: Readout electronics of a PET detector based on: (a) analog SiPM and (b) digital SiPM. 14

Figure 1-10: Planar illustration of a cadmium zinc telluride (CZT) detector. A negative high voltage is applied on the cathode and the anode is grounded. The e-h pairs are generated by absorbing the energy from the gamma rays. 16
Figure 1-11: Illustration of electrode patterns of a cadmium zinc telluride (CZT) detector: (a) Pixelated pattern; (b) Cross-strip pattern. 17
Figure 2-1: Example of the reversed I-V characteristics of a n^+/n-well SPAD. Note that the *x*-axis is expaned to clearly show V_{BR}. 35
Figure 2-2: Illustration of mechanism for dark noise generation. 36
Figure 2-3: Dark count rate as a function of excess voltage at different temperatures......37
Figure 2-4: Histogram of inter-arrival-time with a double-exponential fit to measure the afterpulsing. 38
Figure 2-5: (a) Schematic diagram of a SPAD with passive quench and reset circuit and the corresponding waveform of ouput showing the dead time. (b) Schematic diagram of a SPAD with active quench and reset circuit and the corresponding waveform of ouput showing the dead time. 39
Figure 2-6: Passivation layer and inter-metal dielectric layers in IBM 130 nm CMOS process [86]. 41
Figure 2-7: Example of the timing jitter measurement on a p^+/n-well SPAD using a 405 nm pulsed laser. 42
Figure 2-8: Cross-sectional view of p-n junctions in the standard TSMC 180 nm CMOS technology: (a) Three diodes at the n^+/p-well, p-well/DNW, and DNW/p-substrate junctions; (b) Two diodes at the p^+/n-well and DNW/p-substrate junctions. 43
Figure 2-9: Cross-sectional view of a n^+/p-well SPAD with (a) STI guard ring, and (b) poly gate placed at the edge of n^+ layer to push the STI away from the depletion region.........45
Figure 2-10: Cross-sectional view of (a) a p^+/n-well SPAD with p-well diffusion guard ring, and (b) a n^+/p-well SPAD with n-well diffusion guard ring. 46
Figure 2-11: Cross-sectional view of a p^+/n-well SPAD with virtual guard ring. 47
Figure 2-12: Cross-sectional view of (a) a n^+/p-well SPAD with its poly gate connected to the cathode and (b) a p^+/n-well SPAD with its poly gate connected to the anode............48

Figure 2-13: Simplified schematic for PQR and the equivalent circuit simulation model. 49
Figure 2-14: Simplified schematic for an active quench-reset (AQR) circuit. 51
Figure 2-15: Simplified schematic for a time-gated (TG) circuit. 52
Figure 2-16: Schematics of the time-gated (TG) pixel with the sizes of MOSFETs in nanometer, and pulse generation circuit and its corresponding layout. 53
Figure 2-17: Post-layout simulation results showing the operation of the time-gated (TG) circuit. 54
Figure 3-1: Cross-sectional view of n^+/p-well SPAD. The deep n-well and p-substrate are connected to the ground. 61
Figure 3-2: (a) I-V characteristic of SPAD at different temperatures. (b) Breakdown voltage changes with temperature...63
Figure 3-3: (a) RTS noise measurement setup. The output of SPAD is connected to a low-noise amplifier through AC coupling. (b) Typical waveform of RTS noise. Three parameters- T_{ON} (time spent in the RTS on-state), T_{OFF} (time spent in the RTS off-state) and I (amplitude of RTS noise current) are used to quantify the RTS noise. 64
Figure 3-4: (a) Histogram of T_{ON} with an exponential fit. (b) Histogram of T_{OFF} with an exponential fit. (c) Histogram of RTS noise amplitude with a Gaussian fit. The measurement was conducted with the V_{BIAS} of 11.42 V at 30 °C. 65
Figure 3-5: (a) T_{ON} and T_{OFF}, (b) Duty cycle and (c) RTS noise amplitude as a function of reverse biasing voltage at 30 °C. T_{ON} and T_{OFF} are on a semi-log scale...........................66
Figure 3-6: (a) RTS noise amplitude as a function of temperature. (b) T_{ON} or T_{OFF} as a function of temperature with an exponential fit when the duty cycle is 50%, (c) The reverse biasing voltage to achieve 50% duty cycle at different temperatures with a linear fit. 68
Figure 3-7: (a) Dark count rate (DCR) and afterpulsing (AP) measurement setup. The SPAD is passively quenched through a 50 kΩ resistor. (b) Illustration of inter-arrival time of the avalanche pulses. Avalanche pulse is a negative pulse in the illustration. (c) Histogram of inter-arrival time with multi-exponential fit. The measurement was conducted with the V_{ex} of 1.2 V at -30 °C. 69

Figure 3-8: Dark count rate (DCR) and afterpulsing (AP) as a function of excess voltage at -30 °C. 70
Figure 3-9: Dark count rate (DCR) and afterpulsing (AP) as a function of temperature at a fixed excess voltage of 0.8 V. The temperature coefficient of the breakdown voltage of SPAD was considered when adjusting the reverse bias voltage to keep V_{ex} fixed...........71
Figure 3-10: (a) Illustration of a cylindrical defect in the depletion region of n^+/p-well junction. (b) Equivalent circuit model of the defect with a SPAD capacitance C_{SPAD}, a serial space charge region resistor R_S, a breakdown voltage V_{BR} and a switch which indicates the “on” and “off” of the avalanche. 73
Figure 3-11: (a) Dark count rate (DCR) as a function of excess voltage at -30°C for 3 SPADs with different calculated size of all traps. (b) Afterpulsing (AP) as a function of excess voltage at -30 °C for 3 SPADs with different calculated trap size. The temperature coefficient of the breakdown voltage of SPAD was considered when adjusting the excess voltage. 77
Figure 4-1: Cross-sectional views of (a) of SPAD_NG (n^+/p-well SPAD without poly gate), (b) SPAD_DG (n^+/p-well SPAD with dummy poly gate) and (c) SPAD_FG (n^+/p-well SPAD with field poly gate). 82
Figure 4-2: TCAD electric field simulations of (a) of SPAD_NG (n+/p-well SPAD without poly gate), (b) SPAD_DG (n+/p-well SPAD with dummy poly gate) and (c) SPAD_FG (n+/p-well SPAD with field poly gate). 84
Figure 4-3: (a) Samples of I-V characteristics for three SPADs at 15 °C. Note that the x-axis is 0.5 V width to clearly show the breakdown voltages. (b) Breakdown voltages as a function of temperature for three SPADs 86
Figure 4-4: (a) Measurement setup of dark count rate (DCR) and afterpulsing (AP) for the passively quenched SPAD. (b) Illustration of inter-arrival time (IAT) of the SPAD output pulses. (c) Illustration of the histogram of IAT to show the components of DCR and AP. 87
Figure 4-5: The histogram of the IAT from the real measurement data to show the DCR and AP of the SPAD_FG with 0.5 V excess voltage at -30 °C. 88

Figure 4-6: Comparison of the DCR of SPAD_NG with (a) SPAD_DG and (b) SPAD_FG at different excess voltages and at different temperatures. 89
Figure 4-7: TCAD electric field simulations at the edge of depletion region of (a) SPAD_DG and (b) SPAD_FG. 91
Figure 4-8: Arrhenius plots of three SPADs with 0.5 V V_{ex}. 92
Figure 4-9: Comparison of the afterpulsing (AP) of SPAD_NG with (a) SPAD_DG and (b) SPAD_FG at different excess voltages and at different temperatures. 93
Figure 5-1: Cross-sectional view of the p^+/n-well SPAD. 100
Figure 5-2: (a) The layout of the p^+/n-well SPAD with the dimensions of the octagon..100
Figure 5-3: (a) TCAD simulation of the p^+/n-well SPAD shows an uniform electric field distribution in the central planar area and a reduced electric field at the edge. 101
Figure 5-4: (a) Equivalent circuit mode for SPADs 102
Figure 5-5: (a) Reverse I-V characteristic of SPAD. (b) Linear fit on the linear part of I-V characteristic where the reverse voltage is larger than the breakdown voltage. 103
Figure 5-6: SPAD junction capacitance and side junction capacitance in the p^+/n-well SPAD. 104
Figure 5-7: Schematic of the active quench and active reset (AQR) SPAD pixel...........106
Figure 5-8: Post-layout transient simulation results for the active quench and active reset (AQR) SPAD pixel with V_C set to 0 V. The quenching time (10% - 90% of the rising edge at node A) is 0.1 ns, and the minimum dead time (DT) is ~3.35 ns. 107
Figure 5-9: Schematic of the hold-off time controlling circuit. 108
Figure 5-10: Simulated waveforms of the hold-off time controlling circuit when the control voltage is set to 0 V. 109
Figure 5-11: Simulated waveforms of the hold-off time controlling circuit when the control voltage is set to 0.7 V. 109
Figure 5-12: Delay as a function of the control voltage. 109
Figure 5-13: Schematic of the pulse generation circuit. 110
Figure 5-14: Simulated waveforms of the pulse generation circuit. 110
Figure 5-15: Schematic of the monostable. 111

Figure 5-16: Simulated waveforms of the monostable. 111
Figure 5-17: Layout of SPAD pixel in a standard 65 nm CMOS process, including the output buffers and monostable for driving a large off-chip capacitive load for measurements. 112
Figure 5-18: Breakdown voltage as a function of temperature for 4 SPAD samples. The averaged temperature coefficient is 4.9 mV/°C. 112
Figure 5-19: Dark Count Rate (DCR) measurement setup. 113
Figure 5-20: Dark Count Rate (DCR) as a function of excess voltage. 114
Figure 5-21: Measurement setup for the photon detection probability (PDP). 115
Figure 5-22: Photon Detection Probability (PDP) as a function of wavelength when the excess voltage is biased at different voltages (0.3V, 0.4V and 0.5 V). 115
Figure 5-23: Timing jitter measurement setup. 116
Figure 5-24: Timing jitters at 0.5 V excess voltages using 405 nm and 685 nm pulsed lasers. 117
Figure 6-1: SPAD with a quench and reset circuit at the (a) cathode and (b) anode. (c) SPAD with a differential quench and reset circuit that contains two quench and reset circuits on both the cathode and anode. 122
Figure 6-2: (a) Equivalent circuit model for the SPAD. (b) Traditional single passive quench and reset circuit with a quench resistor connected to the SPAD's cathode. (c) Differential passive quench and reset circuit consisting of two resistors connected to cathode and anode of the SPAD. 124
Figure 6-3: (a) Simplified SPICE model for the SPAD. (b) Simulation setup for traditional single passive quench and reset circuit with the AC-coupled output. (c) Simulation setup for differential passive quench and reset circuit with AC-coupled differential outputs. . 128
Figure 6-4: Simulated waveforms of traditional single passive quench and reset circuit and differential passive quench and reset circuit. 129
Figure 6-5: Setup to measure the DCR and AP of traditional single passive quench and reset circuit and differential passive quench and reset circuit. 130

Figure 6-6: Measured waveforms of traditional single passive quench and reset circuit and differential passive quench and reset circuit. 132
Figure 6-7: (a) Primary dark count rate (DCR) and (b) afterpulsing (AP) of the SPAD with the traditional single passive quench and reset (SPQR) and differential passive quench and reset (DPQR) as a function of excess voltage (V_{ex}) at -35 °C. 134
Figure 6-8: The ratio of DCR (DCR_{SPQR}/DCR_{DPQR}) and AP (AP_{SPQR}/AP_{DPQR}) for the SPQR and DPQR configurations. 135
Figure 6-9: (a) Layout of a SPAD with a single passive quench and reset circuit consisting of a 100 kΩ poly resistor. (b) Layout of a SPAD with a differential passive quench and reset circuit consisting of two 50 kΩ poly resistors. 136
Figure 6-10: Simulated waveforms of traditional single passive quench and reset circuit and differential passive quench and reset circuit with a common-mode noise spike. 138
Figure 6-11: (a) Traditional silicon photomultiplier (SiPM) with fast output terminal. (b) Proposed SiPM with differential output signals. 138
Figure 7-1: Three stacked junctions: n^+/p-well, p-well/DNW, and DNW/p-substrate junctions. 148
Figure 7-2: Illustration of a 3D pixel structure for a SPAD-based sensor [44]. (© 2021 IEEE) 150

List of Tables

Table 1-1: Comparison of PET performance in the state-of-the-art PET/CT systems 20
Table 1-2: Comparison of PET performance of state-of-the-art PET/ MRI systems 22
Table 2-1: Summary and comparison of SPAD implemented by CMOS technologies. ...56
Table 3-1: The dimension of defects at different temperatures 75
Table 3-2: The effective dimension of defects for 3 SPAD samples................................76
Table 4-1: DCR comparison between SPAD_NG and SPAD_DG: 90
Table 4-2: DCR comparison between SPAD_NG and SPAD_FG:..................................91
Table 4-3: AP comparison between SPAD_NG and SPAD_DG: 94
Table 4-4: AP comparison between SPAD_NG and SPAD_FG:.....................................95
Table 4-5: Comparison of our SPADs to other SPADs in Recent Literature 97
Table 5-1: Comparison of our work to other SPAD pixels with active quench and reset circuits...119

List of Abbreviations

AP	Afterpulsing
APD	Avalanche Photodiode
AQR	Active Quench and Reset
ASIC	Application-Specific Integrated Circuit
BCD	Bipolar-CMOS-DMOS
BSI	Backside Illumination
BTBT	Band-to-Band Tunneling
CB	Conduction Band
CFD	Constant Fraction Discriminator
CIS	CMOS Imaging Sensor
CM	Common-Mode
CMOS	Complementary Metal-Oxide-Semiconductor
CR	Count Rate
CRT	Coincidence Resolving Time
CT	X-Ray Computed Tomography
CZT	Cadmium Zinc Telluride
DCR	Dark Count Rate
DNW	Deep N-Well
DPQR	Differential Passive Quench and Reset
DQR	Differential Quench and Reset
dSiPM	Digital Silicon Photomultiplier
DSM	Deep Submicron
DT	Dead Time
e-h	electron-hole
FDG	^{18}F-Fluorodeoxyglucose
FF	Fill Factor
FPGA	Field-Programmable Gate Array

FWHM	Full-Width at Half-Maximum
GR	Guard Ring
HV	High Voltage
IAT	Inter-Arrival-Time
IMD	Inter-Metal Dielectric
L(Y)SO	Lutetium-(Yttrium) Oxyorthosilicate
LED	Leading-Edge Discriminator
LoR	Line of Response
MCP	Microchannel Plate
MRI	Magnetic Resonance Imaging
OCT	Optical Coherent Tomography
PCB	Printed Circuit Board
PDE	Photon Detection Efficiency
PDP	Photon Detection Probability
PEB	Premature Edge Breakdown
PET	Positron Emission Tomography
PMT	Photomultiplier Tube
PQR	Passive Quench and Reset
PVT	Process, Voltage and Temperature
QE	Quantum Efficiency
RTS	Random Telegraph Signal
SiPM	Silicon Photomultiplier
SNR	Signal-to-Noise Ratio
SPAD	Single-Photon Avalanche Diode
SQR	Single Quench and Reset
SPQR	Single Passive Quench and Reset
SRH	Shockley-Read-Hall
STI	Shallow-Trench Isolation
TAT	Trap-Assisted Tunneling
TCAD	Technology Computer-Aided Design

TDC	Time-to-Digital Converter
TG	Time-Gated
TJ	Timing Jitter
ToF	Time-of-Flight
TSV	Through-Silicon Via
TTR	Transit Time Response
TTS	Transit Time Spread
UI	Ultrasound Imaging
VB	Valence Band

List of Symbols

c	Speed of light [m/s]
C_{SPAD}	Junction capacitance of the SPAD [F]
C_{CS}	Parasitic capacitance from the cathode to the substrate [F]
C_{AS}	Parasitic capacitance from the anode to the substrate [F]
D	Diameter of a PET ring [m]
d_i	Diameter of the *i*-th defect [m]
e	The charge of an electron (= 1.6×10^{-19} C)
E_A	Activation energy [eV]
G	Sensitivity gain from time-of-flight PET measurements
I	Amplitude of the RTS noise current [A]
k_B	Boltzmann's constant (= 1.38×10^{-23} J/K)
m^*	Effective mass of electrons and holes [kg]
N_A	Doping concentration of the p-type silicon [cm^{-3}]
N_D	Doping concentration of the n-type silicon [cm^{-3}]
N_t	Electron trap concentration [cm^{-3}]
$P(t)$	Probability of the SPAD in the "on" or "off" state in RTS noise
P_{AP}	Afterpulsing probability
Q_{QS}	Avalanching charge of single passive quench and reset [C]
Q_{QD}	Avalanching charge of differential passive quench and reset [C]
$R_{s,i}$	Equivalent avalanche resistor of the *i*-th defect [Ω]
R_s	Equivalent resistor of SPAD [Ω]
R_{SPAD}	Avalanching resistance of SPAD [Ω]
R_Q	Quench resistor [Ω]
R_{QC}	Cathode quench resistor [Ω]
R_{QA}	Anode quench resistor [Ω]
SNR_{ToF}	Signal-to-noise ratio for a time-of-flight PET measurement [dB]
$SNR_{Non-ToF}$	Signal-to-noise ratio for a non-time-of-flight PET measurement [dB]

T_{ON}	Time spent in the "on" state of the avalanche [s]
T_{OFF}	Time spent in the "off" state of the avalanche [s]
T	Absolute temperature [K]
V_{BR}	Breakdown voltage [V]
V_{ex}	Excess voltage [V]
V_{BIAS}	Biasing voltage [V]
$V_{BR,i}$	The breakdown voltage of the i-th defect [V].
v_{TH}	Thermal velocity of electrons and holes [m/s]
V_{bi}	Built-in potential barrier [V]
W_e	Effective depletion width [m]
x_A	Width of the avalanche region [m]
$\alpha_{n(p)}$	Impact ionization coefficient for electrons (holes) [1/m]
Δt	Timing difference between anti-parallel photons detected[s]
ε_s	Permittivity of the silicon (= 1.04×10^{-10} F/m)
σ_t	Cross-section of the electron trap [m^2]
τ_{ON}	Time constants for the probability distributions of T_{ON} [s]
τ_{OFF}	Time constants for the probability distributions of T_{OFF} [s]
τ_{QS}	Discharge time constant of single passive quench and reset [s]
τ_{RS}	Reset time constant of single passive quench and reset [s]
τ_{QD}	Discharge time constant of differential passive quench and reset [s]
τ_{RD}	Reset time constant of differential passive quench and reset [s]

Declaration of Academic Achievement

This thesis was written by Wei Jiang under the supervision and guidance of Dr. M. Jamal Deen from McMaster University.

- **Chapters 1 and 2:** I conducted a literature review on sensors for positron emission tomography imaging applications with a focus on single-photon avalanche diode (SPAD) based sensors.
- **Chapter 3:** I investigated the random telegraph signal noise on n^+/p-well SPADs to find its correlation with the dark count rate (DCR) and afterpulsing (AP).
- **Chapter 4:** I designed and measured n^+/p-well SPADs with different poly gate configurations to investigate the function of the poly gate to improve the noise performance. Ryan Scott performed the TCAD simulations for SPADs and Tyler Ackland assembled the printed circuit board needed for the measurements.
- **Chapter 5:** I designed and measured a SPAD pixel with an active quench and reset circuit using a standard TSMC 65 nm CMOS process. In this work, Ryan Scott assisted in the design of the SPAD, schematic and layout. He also participated in the measurement of this pixel. Tyler Ackland assembled the printed circuit board needed for the measurements.
- **Chapter 6:** I proposed and analyzed a differential quench and reset (DQR) circuit for SPADs. I performed the measurement on a prototype of differential passive quench and reset circuit. Ryan Scott participated in some technical discussions about the operation principle of DQR circuit. Tyler Ackland assembled the printed circuit board needed for the measurements.

Chapter 1

Introduction

Medical imaging plays a very important role in the clinical analysis and diagnosis of diseases by providing visual representations of the interior structures of a subject and/or the physiological processes occurring underneath the skin. Medical imaging is used to acquire and archive thousands of anatomical and physiological images in databases. The emergence of these databases becomes a powerful tool for training both doctors and emerging machine-based systems to identify abnormalities. This resource will become more beneficial considering the significant breakthroughs in the rapidly developing fields of big data and machine learning [1].

The field of medical imaging has employed many imaging techniques besides positron emission tomography (PET), the most notable being X-ray computed tomography (CT), magnetic resonance imaging (MRI), ultrasound imaging (UI), and optical coherent tomography (OCT). Among these imaging techniques, PET has become one of the most powerful tools to acquire functional images due to its high sensitivity to differences in the metabolic and biological activities at the molecular level [2]. PET is currently being used in a wide variety of clinical areas, such as oncology for cancer or tumor diagnosis and staging [2], neurology for Alzheimer's disease and movement disorders [3], and cardiology in coronary artery disease and myocardial viability assessment [4].

In recent years, PET has also become an important tool in preclinical applications where animal models are used in place of humans to study disease and experiment with new drug development and treatment strategies. While animal PET imaging is mostly for small rodents, there are also studies done on primates due to the high homology of their genes with humans [5]. PET, as a molecular imaging technique, can offer unique non-invasive and in vivo

----

Adapted from W. Jiang, Y. Chalich, M. J. Deen, Sensors for Positron Emission Tomography Applications, Sensors. 19 (2019). doi:10.3390/s19225019. (Appendix A)

imaging beneficial to the study of the biological and biochemical process of the subjects during the experiments with animal models. As a result, a variety of animal PET scanners have been designed and developed for preclinical applications [6]–[10].

PET has also been integrated with other techniques like CT and MRI to develop multimodal imaging systems that take the advantages of combining both functional and anatomical images for improved diagnostics. While the integration of PET and CT has already seen widespread adoption, the PET/MRI systems took longer to develop owing to the photomultiplier tube (PMT) detector limitations of early PET systems. The integration of PET/MRI requires that the PET system must be compatible with the high magnetic fields from an MRI system. At the same time, the MRI image should not be degraded by the PET system due to the extra noise and electrical interference. Thanks to the advances of photo-sensor technologies in recent years, the PET/MRI system was made commercially available and used in many clinical applications [11]–[17].

In this chapter, first, the physics of PET imaging systems are described. Second, four types of PET detectors: PMTs, avalanche photodiodes (APDs), silicon photomultipliers (SiPMs) and cadmium zinc telluride (CZT) detectors are introduced. Third, the research motivation of this thesis is presented. Fourth, the research contributions are summarized. Finally, a description of the thesis organization is given.

1.1. Positron Emission Tomography

1.1.1. Physics of Positron Emission Tomography

The basic principle for PET is the coincident detection of a pair of gamma rays generated from the annihilation events of the positrons from the radioactive tracer injected into a subject. The radiotracers for PET applications are analogous to common biological molecules such as glucose, peptide and protein in which a radioisotope is used to substitute one of the constituents of the tracer [18], [19]. For example, ^{18}F is used to replace ^{16}O of the glucose to produce ^{18}F-fluorodeoxyglucose (FDG), which is analogous to glucose and can indicate levels of cellular metabolism. Another widely used tracer is ^{11}C-L-methionine, analogous to

the amino acid, which can be used as an indicator of cancer malignancy based on the utilization of the amino acid [18].

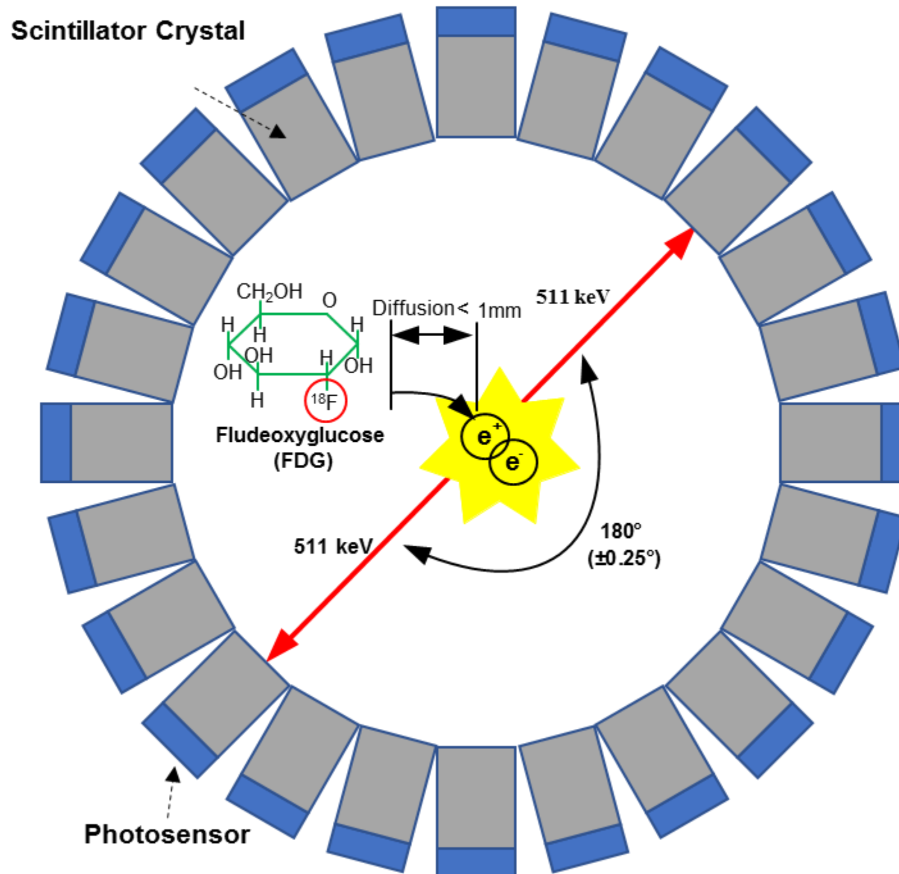


Figure 1-1: The basic principle of a positron emission tomography (PET) system: A PET detector ring detects a pair of gamma photons with an energy of 511 keV (red arrows) which results from the annihilation of an electron with a positron emitted by the radiotracer (FDG).

The illustration of a general PET imaging system with indirect photodetectors is shown in Figure 1-1 [20]. The first step of the PET image scanning is the injection of a radioactive tracer into the subject. The radioactive tracer arrives at the targeted organs or tissues through the circulatory system after a certain amount of time and participates in the metabolic processes of the subject. Since the radioisotopes in the tracer are not stable, they decay (i.e., through β+ decay) during the metabolic process with a specific half-life decay time (around 110 minutes for ^{18}F and 20 minutes for ^{11}C). During the decay process, positrons are generated and travel to collide with the electrons of the neighboring atoms in an annihilation

process. The annihilation generates two gamma rays with an energy of 511 keV and a separation of approximately 180 degrees [21]. In order to detect the gamma rays due to the annihilations, scintillation crystals are used to absorb and convert the high energy gamma rays into low energy visible photons. One of the most commonly used scintillators is lutetium-(yttrium) oxyorthosilicate (L(Y)SO) due to its high density, high light output, fast decay time, and excellent energy resolution. Then, a photosensor like a PMT, APD or SiPMs is used to convert the light signal into an electrical signal. Note that the building blocks of the SiPMs are single-photon avalanche diodes (SPADs). The scintillators, detectors, and the readout electronics are generally assembled together to form a modular detector, which are used to build a detecting ring to record three electronic signals: the time when the gamma ray hits the detector, the position where the gamma ray hits the detector, and the energy of the gamma ray [19]. These electronic signals are then processed through the coincidence unit to get the true events generated by the same annihilation which occur along the line of response (LoR). Finally, the raw data of thousands of LoRs are used to generate the PET image through image corrections and reconstructions. The entire detection flow is shown in Figure 1-2.

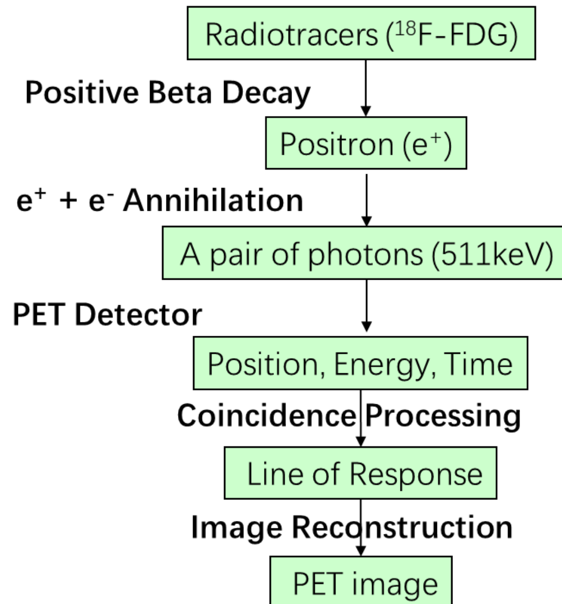


Figure 1-2: Detection flow of a positron emission tomography (PET) system.

1.1.2. Time-of-flight (ToF) PET

Due to improvements in the timing resolution of photosensors, the time-of-flight (ToF) technique can be applied to the conventional PET to significantly increase the sensitivity and signal-to-noise ratio (SNR), thus improving the image quality [18], [19]. Figure 1-3 shows the concept for ToF PET and its comparison to the conventional PET. In the conventional PET, the time difference between a pair of gamma rays generated from the annihilation event can only be used to find the LoR by the coincidence processing unit with the coincidence window usually set to 3 - 5 ns. Once a LoR is determined, we only know that the annihilation occurred along this LoR. The probabilities for all the voxels along this LoR to locate the position of the annihilation event are the same, as shown in Figure 1-3(a) and Figure 1-3(c). Therefore, in the conventional PET, the noise from all voxels will be accumulated to reduce the system SNR, resulting in degraded image contrast and resolution [22], [23]. However, due to the higher timing resolution in ToF PET, the location of the annihilation event can be located by measuring the difference of the arrival time of each photon at two detectors along the LoR. The uncertainty of location for the annihilation point due to the time jitter is limited to a short range along the LoR, as shown in Figure 1-3(b).

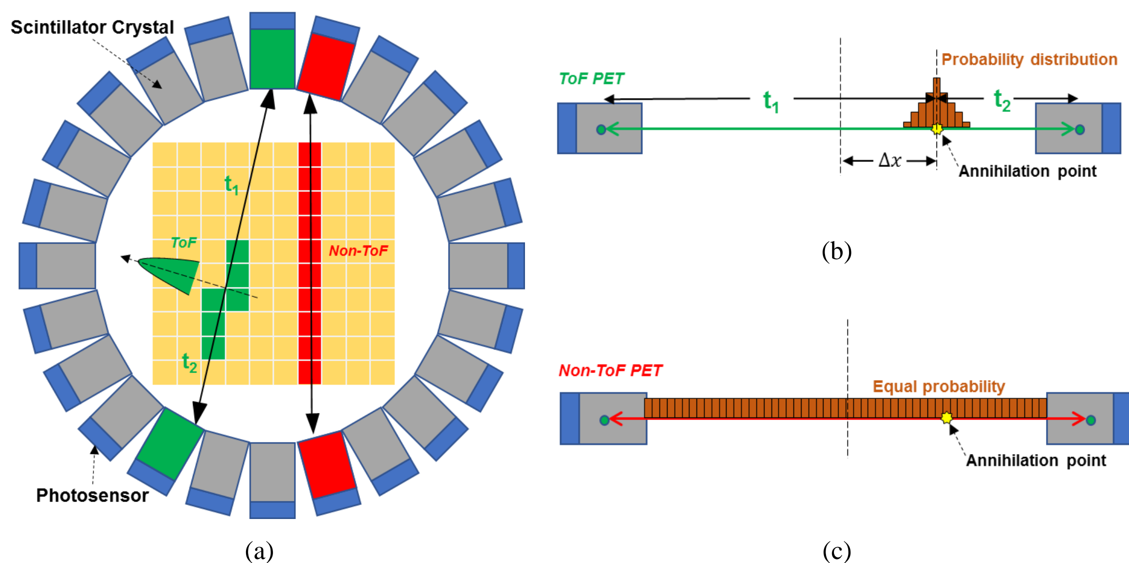


Figure 1-3: Concept of time-of-flight positron emission tomography (ToF PET): (a) illustration of a detector ring detecting pairs of gamma photons from the annihilation events with (green) and without (red) the ToF technique; (b) The probability distribution of the annihilation position along the LoR in ToF PET; (c) The equal probability of the annihilation position along the LoR in non-ToF PET.

Compared to the conventional PET, the improvement of SNR in ToF PET is given by

$$\frac{SNR_{ToF}}{SNR_{Non-ToF}} = \sqrt{\frac{2D}{c \times \Delta t}}, \tag{1-1}$$

and the gain of effective sensitivity G is

$$G = \frac{2D}{c \times \Delta t}, \tag{1-2}$$

where the D is the diameter of the object being scanned, c is the speed of light in vacuum, Δt is the variation of arrival time, defined as the coincidence resolving time (CRT) [24]. For example, if the system is assumed to achieve a CRT of 300 ps, and the diameter of the subject is 40 cm, we can calculate that SNR will increase $\sim$3 times and the effective sensitivity will increase ~ 9 times. The increased SNR and sensitivity will contribute to the improvement of the SNR of the final PET image. One comparison of the image quality between ToF PET and non-ToF PET was reported in [25], which showed three clinical images from a patient with colon cancer by using CT, non-ToF PET and ToF PET. The lesion for colon cancer was easily detected by the ToF PET but could not be observed in the non-ToF PET.

The above analysis shows that the timing jitter performance plays an important role in the spatial resolution of the PET system. As shown in Figure 1-3(b), improvement of the timing jitter allows the PET system using ToF technique to increase the accuracy of detecting the annihilation position, thus providing better spatial resolution when compared to the non-ToF PET system.

1.1.3. PET detector

The basic structure of a PET detector is shown in Figure 1-4(a), which mainly consists of a scintillator block mounted on the photosensor. Note that matching between the scintillator block and the photosensor has a great influence on the efficiency of the detector. To maximize the number of photons that reach the surface of photosensors, the size of the photosensor should match the size of the scintillator block. In addition, optical grease or optical cement are usually applied between the scintillator and the photosensor as light guides to guide the generated photons from the scintillator to the photosensor. The main function of

the PET detector is to acquire three types of information: the position where the gamma photon impacts the scintillator; the time when the output pulse from the photosensor arrives; and the energy of the output pulse. The photon detection flow is illustrated in Figure 1-4(b). As shown in the flow chart, the noise for a PET detector can come from the light illumination process, photosensors and readout electronics. Among these noise sources, the noise from the photosensors is the dominant one, especially when solid state photosensors are used in PET detector. For example, the dark count rate for a SPAD can be as high as several thousand Hz (see details of the noise performance in the literature review of SPADs in Section 2.3).

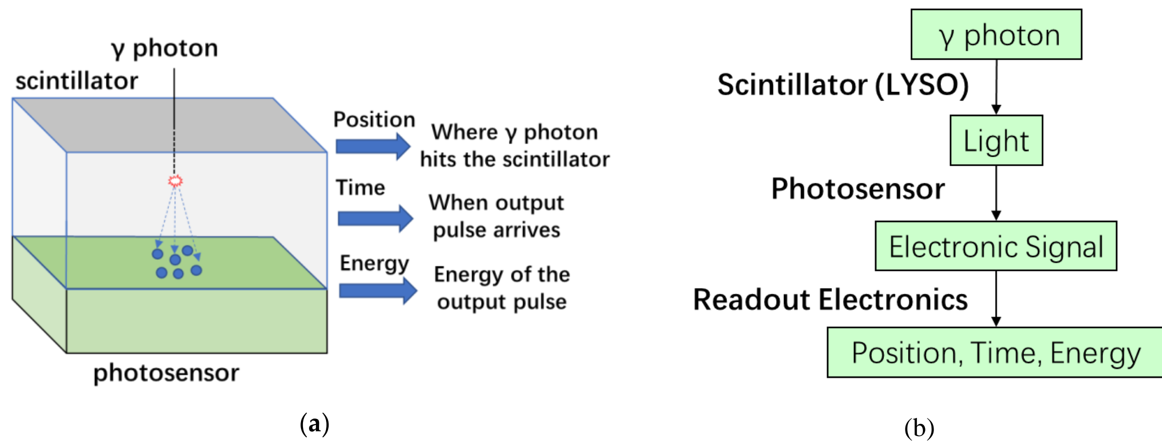


Figure 1-4: Positron emission tomography (PET) detector: (a) Structure of a PET detector; (b) Detection flowchart of a PET detector.

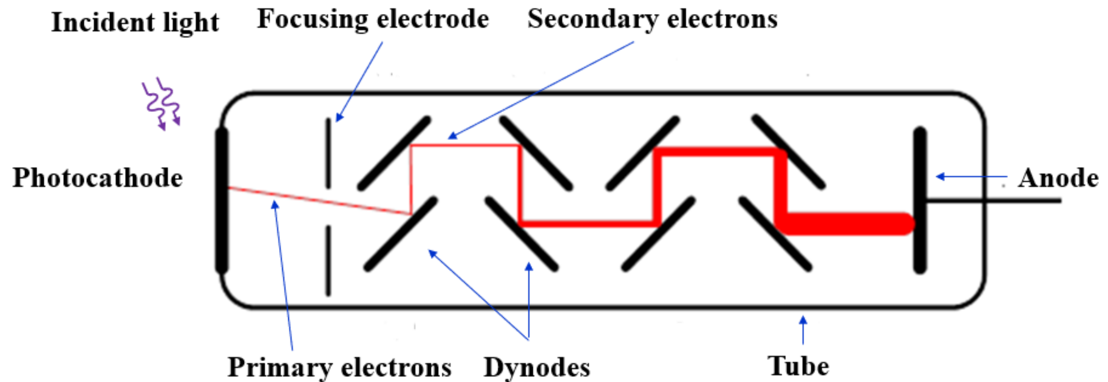
1.2. Photosensors for PET Applications

The photosensor in a PET detector plays a very important role in the system performance in terms of spatial resolution, coincidence timing resolution and energy resolution. There are generally four types of photodetectors being used in research and commercial PET systems: PMTs, APDs, SiPMs and CZT detectors.

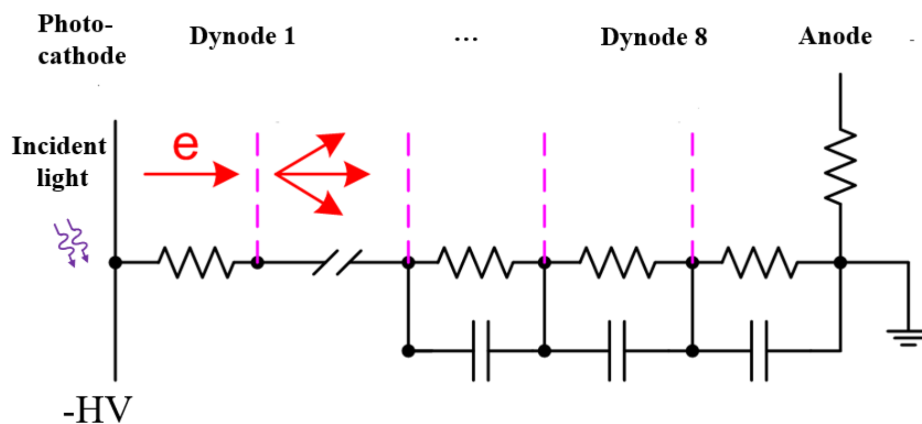
1.2.1. Photomultiplier Tubes (PMTs)

The schematic diagram of a typical PMT structure is illustrated in Figure 1-5 [26]. The primary components of a PMT are a photocathode, a series of electrodes (dynodes) and an anode, which are enclosed in a vacuum glass tube. The different electrodes are biased at progressively higher voltages to form increasingly higher internal electric fields. When the

incident photons hit the photocathode through the glass window, photoelectrons are generated from the photosensitive material on the photocathode due to the photoelectric effect. The high internal electric field between the photocathode and the first dynode significantly increases the kinetic energy of the photoelectrons, allowing them to generate a secondary emission of electrons upon colliding with the dynode. This process is amplified at each progressive dynode as the electrons accelerate between them. Eventually, the anode collects a huge number of electrons to generate a high photo-current pulse, which is significantly larger than the noise current. With proper design of the photocathode and the multiple stages of dynodes, PMTs can easily achieve multiplication gains of 10^6 [27].



(a)



(b)

Figure 1-5: Principle of the photomultiplier tube (PMT): (a) Simplified conventional structure; (b) Simplified high-voltage biasing network.

Also, there exists microchannel plate (MCP) PMTs, shown in Figure 1-6, that utilize the gain effect of the secondary electron emissions. Instead of using a chain of separate dynodes, MCP-PMTs are composed of a large number of micron-sized channels coated with a conductive emissive dynode material. A high electric field along the channel is present when a high voltage is applied. When a high energy photoelectron emitted by the photocathode enters the channel, it is absorbed and causes secondary emissions from the coating material of the internal walls. This process is repeated by subsequent emissions along the channel to result in an exponentially increased number of electrons at the anode to generate a photocurrent distinguishable from the noise. Typical diameters of the channel are 3 - 10 μm.

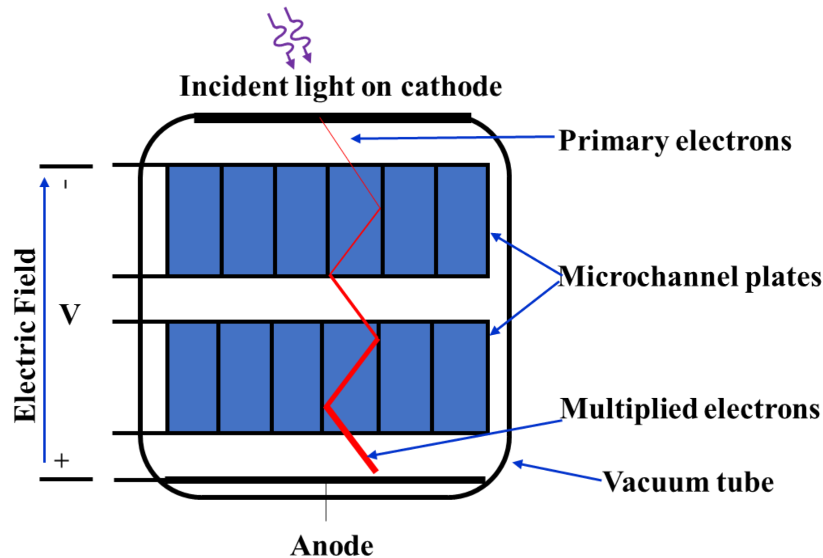


Figure 1-6: Illustration of a microchannel plate photomultiplier tube (MCP-PMT) (The microchannels are usually made of highly resistive materials (e.g. lead glass cladding) with the inner wall coated by high secondary emission materials such as MgO and Al_2O_3).

Higher gains can be achieved by arranging two or three microchannel plates in series. The first microchannel plate can be placed very near to the cathode, resulting in a high performance for the transit-time spread, which can be as short as 25 ps [26].

Ideally, if no single photon is absorbed by the photocathode, there will be no output pulse from the anode of the PMT. However, electrons randomly emitted from the photocathode and dynodes by the process of thermionic emission and/or field emission generate output pulses even if the PMT is kept in the dark. These “dark” pulses per unit time are termed the

dark count rate (DCR) of the PMT. The DCR of the PMT is mainly determined by the cathode material and the design of the dynode chain. Compared to a solid-state sensor, the vacuum environment enables a PMT to achieve a relatively low DCR on the order of tens of counts per second [28].

To describe the timing characteristics of a PMT, the electron transit time response (TTR) and the transit time spread (TTS) are used. The TTR is the average time difference between the arrival of a photon at the photocathode and the output pulse at the anode. The TTS is the standard deviation of the transit time distribution, also known as timing resolution or timing jitter [28]. Furthermore, the TTS is the critical parameter for PET applications since it represents the uncertainty of the photon arrival time and has an impact on CRT.

The photon detection efficiency (PDE) of a PMT is determined by the collection efficiency and the internal quantum efficiency (QE) of the photocathode material, which is used to describe the probability of emission of a photoelectron per incident photon. It is usually measured as a ratio between the number of generated photoelectrons and the number of incident photons. The PMT's quantum efficiency is wavelength dependent and has a typical value of ~ 25% [27]. Due to the limitation of collection efficiency, the PDE is smaller than the QE since not every photoelectron can generate a detectable pulse.

1.2.2. Avalanche Photodiode (APD)

APDs are structurally similar to p-n or p-i-n photodiodes, but the gain mechanism is from avalanche multiplication. They are biased at a large reverse voltage to create a high electric field such that when an incident photon generates a carrier within the depletion (space-charge) region, it is accelerated to become more energetic. With enough energy gained from the electric field, the carrier can generate an electron-hole (e-h) pair through impact ionization when it collides with the lattice, as illustrated in Figure 1-7. This process can be repeated by the initial carrier and subsequently generated carriers through avalanches, resulting in an appreciable increase in the detectable current. The impact ionization coefficient α for the main charge carrier (electrons in the case shown in Figure 1-7 for APDs) determines the multiplication gain and represents the number of e-h pairs generated per unit length by the

carrier. Due to the stochastic nature of the avalanche effect, there exists some variation in the multiplication gain since not every injected or photogenerated carrier leads to the same multiplication. This noise is measured by the excess noise factor F and is dependent on the ratio k of the ionization coefficients for electrons (α_n) and holes (α_p) of the material used. When considering electron injection/multiplication in APDs, $k = \alpha_p/\alpha_n$ (α_n/α_p for hole injection/multiplication) and k should be minimized to reduce this noise. The k value is 0.02 for Si and ~0.5 for Ge and III-V compound semiconductors and is an important reason why the majority of commercial APDs are made of Si [29] to reduce excess noise.

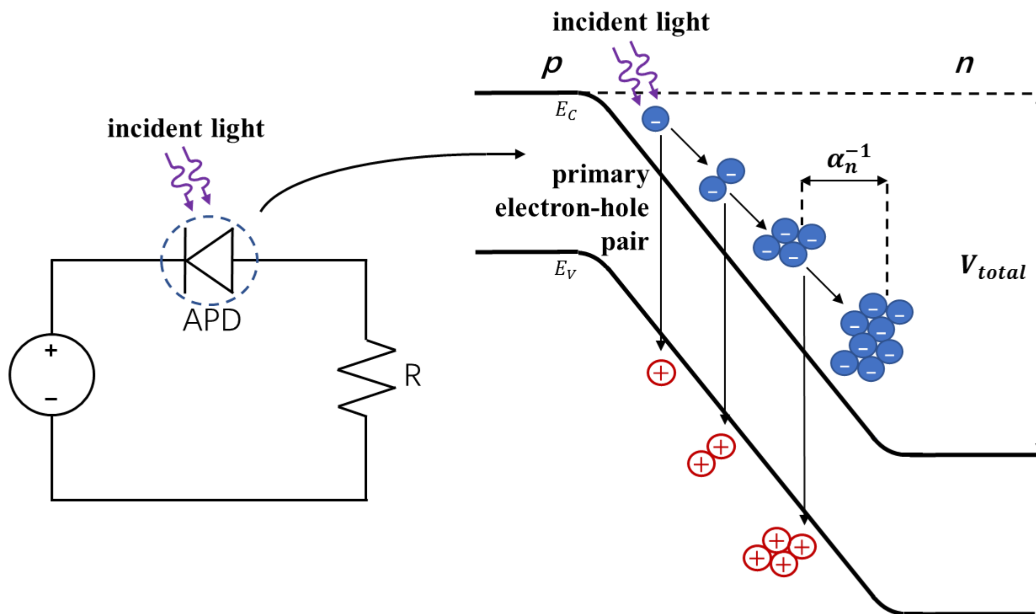


Figure 1-7: The impact ionization process. Here, α_n^{-1} is the average distance between each electron multiplication event, while V_{total} is the applied reverse bias plus the built-in potential.

The performance of an APD is primarily specified in terms of its spectral response, quantum efficiency, gain, and dark current. The spectral response, or responsivity, corresponds to the generated current (A) per unit power (W) of light incident on the APD for different wavelengths. The quantum efficiency is also a function of wavelength and indicates the percentage of photons that reach the depletion region and trigger an avalanche. The quantum efficiency and responsivity are dependent on the light transmitted into the semiconductor, the percentage of photons absorbed and converted to carriers in the multiplication process, and the current collected from non-recombined carriers. The

(multiplication) gain described above through the impact ionization process is ideally sought to be as large and stable as possible to more accurately detect low-level incident light. Multiplication also applies to dark current that primarily occurs from thermally generated e-h pairs, and the random thermal motion and fluctuations of charge carriers. Together, low dark current and high gain allows for distinguishability between the amplified noise and amplified light-triggered events. The dark current and gain vary with reverse bias voltage and temperature such that APDs can be set to function at some optimal operating point to maximize the SNR. While APDs can have gains up to around 1000 [27], Si APDs, such as those from Hamamatsu, typically have an internal gain around 50 at the required 420 nm wavelength for PET applications [30].

1.2.3. Silicon Photomultiplier (SiPM)

An array of SPADs connected in parallel is called an analog SiPM. The output of an analog SiPM is the sum of the individual currents of all SPAD pixel cells. The output amplitude of the current of an analog SiPM is proportional to the number of SPAD breakdowns, be it from photons or noise. As the basic photon-detecting unit, the performance of the single SPAD pixel has great influences on the performance of the array.

A SPAD is a p-n junction that is reverse-biased at an excess voltage Vex above the breakdown voltage (V_{BR}) so as to operate in the Geiger mode (in analogy to a Geiger-Muller detector). Geiger-mode operation creates a large enough electric field across the depletion region of the p-n junction so that the impact ionization process triggered by an initial photogenerated charge carrier is self-sustaining [25], [31]. This avalanching process can generate a large internal current flow from only a single photon shown in Figure 1-8(a). Figure 1-8(b) illustrates the operating states of a SPAD. Initially, when reverse biased above breakdown, the SPAD stays in an OFF state (1) where there is no triggering event to initiate avalanching. When a charge (generated by photon absorption or other means) enters the depletion region and triggers avalanching, the SPAD is brought into its high-current ON state (2). Due to the existence of a large quenching resistor R_Q, the initiated avalanche causes the voltage across the depletion region to be reduced, bringing the SPAD to point

(3), where the voltage is below the breakdown voltage. Here, the avalanche process is no longer self-sustaining and is quenched. The voltage drop can be sensed by a fast discriminator to generate a digital pulse which represents the photon arrival time if the first charge was a photon-generated charge [32], [33]. After quenching the avalanche, the SPAD voltage is recharged back to the initial biasing conditions (1) of V_{SPAD}. The key performances parameters for SPADs include V_{BR}, dark count rate (DCR), afterpulsing (AP), dead time (DT), count rate (CR), photon detection efficiency (PDE) and timing jitter (TJ). The definition of these parameters will be discussed in detail in Chapter 2.

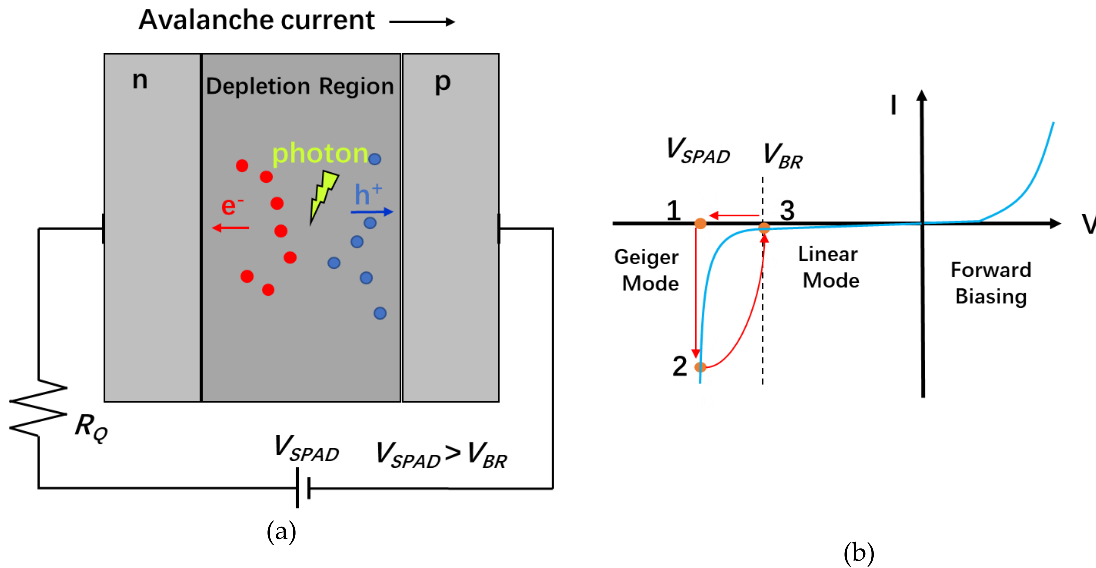
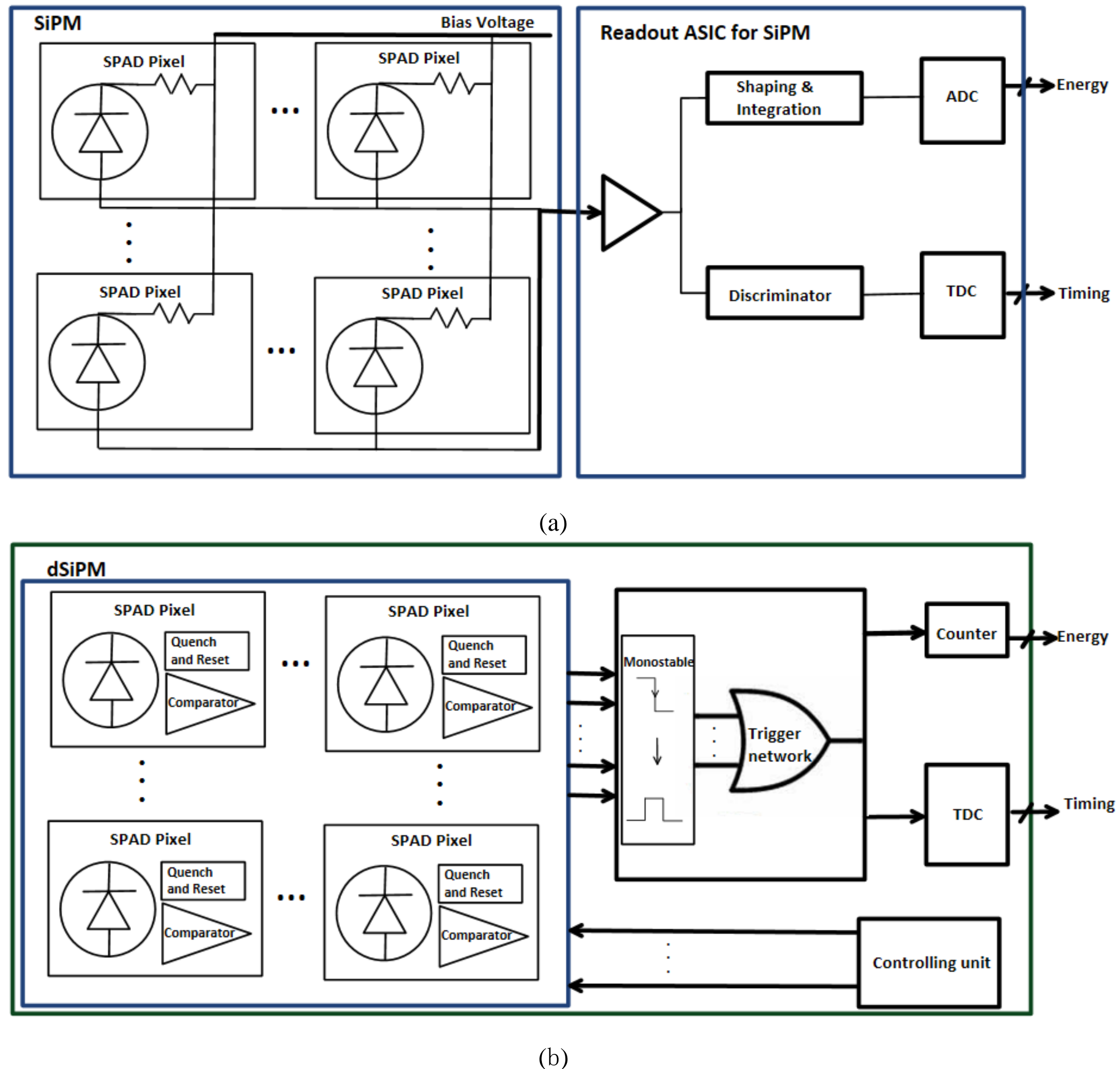


Figure 1-8: Principle of SPAD operation: (a) Avalanche breakdown process in a reverse biased p-n junction; (b) I-V characteristic representation of SPAD operation.

Compared with the single SPAD pixel design, the phenomenon of crosstalk needs to be considered in the applications of SiPM. Crosstalk happens when large SPAD pixel arrays with small pixel pitch are fabricated in one substrate to form a SiPM. When a Geiger pulse is triggered in a pixel cell, there is a finite probability that one or more neighboring pixels may also become triggered due to the electrical and/or optical crosstalk. Electrical crosstalk originates from avalanche-generated carriers diffusing laterally and entering the multiplication region of nearby SPADs. These carriers cause unwanted or spurious Geiger pulses. Electrical crosstalk can be reduced by dielectric and/or junction isolation. Optical crosstalk occurs from photons that are emitted during a Geiger pulse which can be reabsorbed

in neighboring SPADs. The probability of crosstalk increases with increasing excess voltage. Crosstalk will result in the output signal being higher than that triggered by the amount of incident photons [34].



(a)

(b)

Figure 1-9: Readout electronics of a PET detector based on: (a) analog SiPM and (b) digital SiPM.

In the past decade, analog SiPMs was successfully employed in ToF PET systems [35]–[39]. In their application for PET, the detected signal from the analog SiPM needs a readout ASIC (application-specific integrated circuit) to get the timing and energy information of the detected photons, as shown in Figure 1-9(a). The number of the output channels of analog SiPM can be reduced by employing resistive [40] or capacitive [41] multiplexing circuits.

This feature reduces the efforts involved in designing the readout electronics for the analog SiPMs.

With progress in the design and fabrication of SPADs in standard CMOS processes, the idea to integrate the SPAD array with the relevant circuits on one chip was proven and demonstrated, for example in Ref. [42]. This new type of integrated sensor is called a digital silicon photomultiplier (dSiPM). In dSiPMs, not only are active quench and reset (AQR) circuits integrated with the SPAD array, but also other signal conditioning and processing circuits like digital acquisition circuits, time-to-digital converters (TDCs) and counters are also implemented on the same chip. Among these auxiliary circuits, the TDC is the most important one, greatly influencing the time resolution of the whole integrated chip. Detailed reviews of TDCs can be found in [43], [44]. As shown in Figure 1-9(b), the timing and energy information can be directly obtained from a single chip [45]. Due to the fully digital readout and processing circuits being integrated with the sensor on the same chip, the dSiPM offers excellent timing performance and reduces the complexity of the signal conditioning and processing circuits which makes it an attractive sensor for medical imaging systems like PET. However, when dSiPMs are compared to analog SiPMs, there are a large number of electrical signals (i.e., clock, reset, digital data interfaces and some other auxiliary signals) to be connected when integrating the dSiPMs to form a sensor tile. Dealing with such a large number of connections can be challenging because of the limited number of pins of any ASIC or FPGA (field-programmable gate array) that is used.

1.2.4. Cadmium Zinc Telluride (CZT) Detectors

Semiconductor detectors were also used in nuclear radiation detection during the past few decades since they have the ability to convert high-energy rays such as x-rays and gamma rays directly into electronic signals. Unlike the scintillator, the direct conversion of the semiconductor detector can avoid the stochastic effects resulting from the light generation, propagation and conversion into electronic signals, which makes it a favorable alternative to scintillator-based photon detectors [46]. Among the semiconductor detectors, CZT detectors have shown great potential in PET imaging systems. Compared to the L(Y)SO scintillators,

which have a mass attenuation coefficient of 0.117 cm^2/g at 511 keV (Compton scattering mass attenuation coefficient of ~0.038 cm^2/g and photoelectric absorption mass attenuation coefficient of ~0.073 cm^2/g), the CZT detectors show slightly lower performance in the absorption of high-energy gamma rays. The mass attenuation coefficient of CZT detectors is 0.086 cm^2/g at 511 keV. The fraction of Compton scattering is 0.82 and the fraction of photoelectric absorption is 0.18, which translates into a Compton scattering mass attenuation coefficient of ~ 0.071 cm^2/g and a photoelectric absorption mass attenuation coefficient of ~ 0.015 cm^2/g [47]. Despite these low coefficients, the relatively high atomic number (average Z of 49.1 for $Cd_{0.9}Zn_{0.1}T$) allows a good portion of high energy gamma rays be absorbed to generate a large enough signal, which is distinguishable from the noise. The high material density (5.78 g/cm^3) further yields high stopping power. Moreover, the wide band-gap (1.6 eV) allows the CZT detector to operate at room temperature [48], [49], [50].

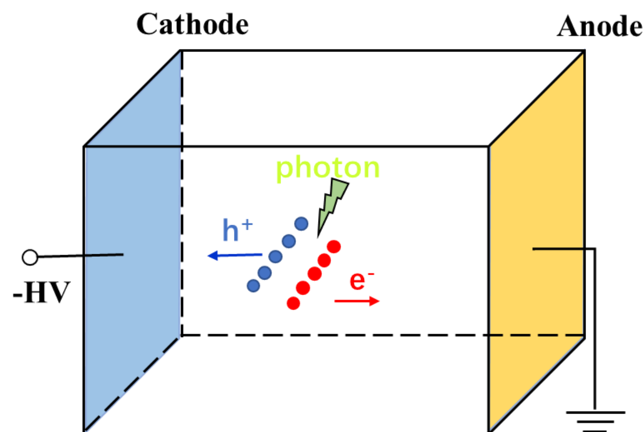


Figure 1-10: Planar illustration of a cadmium zinc telluride (CZT) detector. A negative high voltage is applied on the cathode and the anode is grounded. The e-h pairs are generated by absorbing the energy from the gamma rays.

Figure 1-10 shows the operating principles of a CZT detector. The interaction of the incident gamma ray with the crystal lattice of CZT produces primary electrons. The high-energy primary electrons will undergo impact ionization to generate secondary charges. In semiconductor detectors like CZT, the generated secondary charges are pairs of electrons and holes, with electrons excited from the valence band into the conduction band and the holes left in the valence band. The number of e-h pairs created by impact ionization is proportional to the deposited energy from the incident gamma ray and the energy needed to create an e-h

pair in the semiconductor (the e-h pair creation energy for CZT is 4.64 eV). Due to the high electric field from the applied external voltage between the anode and cathode, the electrons and holes will drift and be collected by the anode and cathode, respectively, giving rise to the signal current. Then, the signal is processed by the readout electronics.

In the CZT detector, both the electrons and holes will drift to the electrodes to generate a current which represents the energy information. However, the mean drift length of electrons is much larger than that of holes in the CZT detectors. This makes the anode the favorable collector for energy information. Two types of electrode patterns are commonly used in the CZT detector as shown in Figure 1-11. Figure 1-11(a) shows the pixelated pattern design, in which only the anode comprises of an array of small pixelated electrodes and the cathode is one shared large electrode. The outputs of the pixelated electrodes produce the position information of interaction. Figure 1-11(b) shows the cross-strip electrode pattern. In this configuration, the same 2D position information can be read from orthogonally crossed anode and cathode strips. Compared to the pixelated configuration, the cross-strip configuration will reduce the numbers of electrodes required to produce the same spatial resolution across the same volume of the detector (from N^2 down to 2N), which greatly reduces the burden for the readout electronics [51].

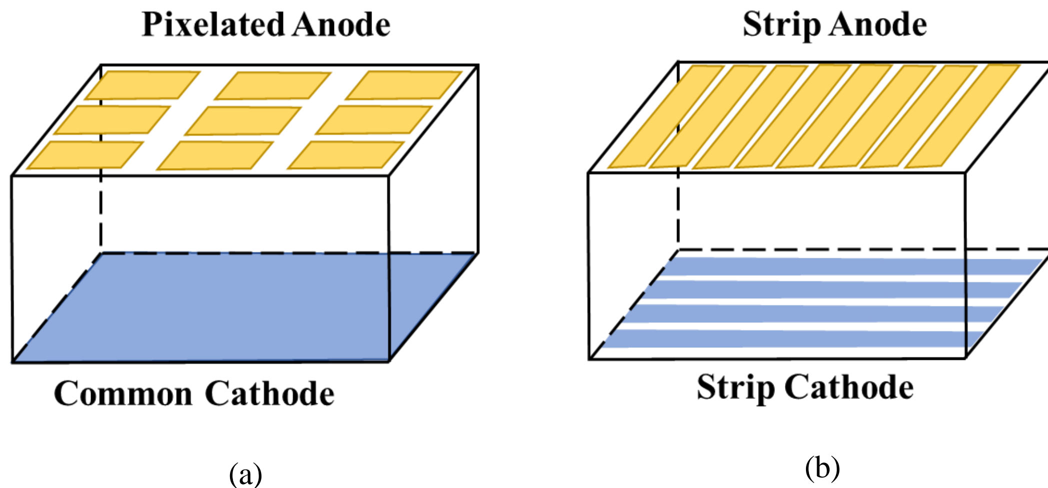


Figure 1-11: Illustration of electrode patterns of a cadmium zinc telluride (CZT) detector: (a) Pixelated pattern; (b) Cross-strip pattern.

In this section, we briefly described four types of sensor technologies employed in the PET applications: PMTs, APDs, SPAD-based detectors and CZT detectors. PMTs were widely used for PET applications in the early days due to their excellent performance metrics of high gain, low noise and fast timing. However, the fragility and bulkiness of the PMT glass tubes, high operating voltage, and sensitivity to magnetic fields ultimately limit this technology for future cost-effective, multimodal systems. As a result, solid-state photodetectors like the APD, SPAD-based and CZT detectors and their applications for PET systems has attracted much research interest, especially owing to the continual advancements in the semiconductor fabrication process.

Unlike the APD or SPAD detectors which need a scintillator to convert the high-energy gamma rays to lower-energy visible light, CZT detectors can generate an electronic signal by directly detecting gamma photons, resulting in the excellent performance in terms of energy resolution and position resolution. However, the timing resolution of CZT detectors lags behind other technologies with a range of nanoseconds to tens of nanoseconds. Although CZT detectors showed excellent position and energy resolutions, their PET applications are limited to small-sized preclinical PET systems due to their poor timing resolution performance. Additionally, the small amplitude of readout signals and a large number of the signal channels of electrodes result in very complex readout electronics for processing the signals from the CZT detectors. Unlike the SPAD-based sensors, where the sensors and the readout electronics can be implemented on the same wafer or on the same package, the CZT detector requires separate readout electronics, thus increasing the difficulties for system integrations.

Compared to APDs, SPAD-based detectors (SiPMs) allow single photon detection that APDs were incapable of due to their limited gain. This high photon detection efficiency together with other benefits such as excellent timing resolution below 1 ns, and ~10^6 intrinsic gain equivalent to that of PMTs, make SPAD-based detectors the most promising and attractive sensors in the PET application. A comparison among PMT, APD, SPAD-based detectors (Analog SiPM and dSiPM) and CZT detectors is given in Table 1-1. As

seen from the table, SPAD-based sensors are the most promising and attractive technology for the PET applications.

1.3. State-of-the-art PET systems

In real clinical applications, PET imaging systems are usually not utilized alone but are combined with other imaging modalities like X-ray CT or MRI. The dual-modality imaging systems like PET/CT and PET/MRI acquire their images simultaneously in the same bed position, which help to reduce errors for image registration and image fusion, provide more diagnosis information and reduce the scanning time of patients. In addition, imaging from one modality can be used for imaging reconstruction and correction in the other imaging modality. PET/CT systems have been widely used in many pre-clinical and clinical areas for decades, while PET/MRI systems were just released within the past several years due to working limitations of PMTs in the magnetic field. Thanks to the technological advancements of solid-state sensors like APD, SiPM and dSiPM, many PET/MRI imaging systems were designed and developed.

1.3.1. PET/CT systems

There are several medical device manufacturers providing PET/CT systems. Table 1-1 lists some state-of-the-art PET/CT systems from these several medical device manufacturers. In this section, only the design and performance of the PET system will be compared and discussed. The system performance parameters such as spatial resolution, timing resolution and energy resolution values are reported as FWHM.

By using a dSiPM (DPC series), the Philips Vereos PET/CT offered the first commercial digital PET system with enhanced ToF performance [52]. The basic detector consisted of DPCs and arrays of LYSO scintillators of size $4\times4\times19$ mm^3 coupled on DPCs through direct 1:1 coupling. Due to the significantly better performance of Philips DPC technology, the overall system achieved a spatial resolution of 4.1 mm, a timing resolution of 325 ps and an energy resolution of 11.1% [53].

Discovery MI PET/CT is the latest PET/CT system from GE. The detection ring of the PET system is made of tens of Lightburst Digital Detectors, which combines a small lutetium-based scintillator (LBS) crystal array with a SiPM [54]. The 4×6 mm^2 SiPM pixels were arranged in a 2×3 array to form a SiPM device with a large active area of 12.6×12.6 mm^2 and a small dead area among the pixels (as low as to 250 μm). The photosensitive area of the SiPM device was then coupled to a 4×3 array of LBS crystals of size 3.95×5.3×25 mm^3. The output signals of the SiPM were processed by a custom ASIC to enhance the integrity of the shape of the SiPM output and lower the power consumption of the module. Based on the modular design concept, the Lightburst Digital Detector can be assembled onto the gantry with different sizes to build a scalable PET system. The system measurement results showed a spatial resolution of 4.57 mm, a timing resolution of 385 ps and an energy resolution of 9.4% [55].

Table 1-1: Comparison of PET performance in the state-of-the-art PET/CT systems

Ref.	[52], [53]	[54], [55]	[56], [57]	[58]
Manufacturer	Philips	GE	Siemens	Canon
Model Name	Vereos Digital PET/CT	Discovery MI PET/CT	Biograph Vision PET/CT	Celesteion™ PUREViSI ON Edition PET/CT
Scintillator material	LYSO	LBS	LSO	LBS
Scintillator size (mm^3)	4×4×19	4.0×5.3×25	3.2×3.2×20	4×4 (length unknown)
Sensor	dSiPM	SiPM	SiPM	PMT
Spatial resolution (mm)	4.1	4.57	3.6	5.1
Timing resolution (ps)	325	385	214	394
Energy resolution (%)	11.1	9.4	9	11.2

The PET system of Biograph Vision PET/CT is the fastest ToF PET currently in the market [56], achieving a CRT of 214 ps [57]. A so-called ultra-dynamic range (UDR) detector was designed for this system. An array of fast LSO scintillator of size 3.2×3.2×20 mm^3 was coupled on the SiPM to form the UDR detector with a 100% coverage. With the smaller scintillator size and the fastest ToF performance, the Biograph Vision PET/CT achieved a spatial resolution as small as 3.6 mm. However, a small scintillator size means the Biograph Vision PET/CT needs more crystals than other systems to achieve the same

FoV, leading to increased complexity in the fabrication and assembly of the scintillators. The energy resolution of the system was measured to be 9% [59].

Interestingly, the newest PET/CT system from Toshiba (now part of Canon Group) called the Celesteion™ PUREViSION Edition PET/CT system [58], is still based on PMT technology. The PET detector was designed for a large-bore PET. The detector was made of PMTs with mixed sizes and lutetium-based scintillators (LBS). With optimal use of LBS and a unique mixed PMT configuration, the system achieved a spatial resolution of 5.1 mm, a timing resolution of 394 ps and an energy resolution of 11.2%, which were comparable to PET systems based on solid-state sensor technology like SiPM and dSiPM.

1.3.2. PET/MRI systems

Due to the incompatibility with magnetic fields, PMT-based PET systems can only be combined with MRI systems using extra shielding techniques. Unlike the PMT, solid-state sensors are compatible with the magnetic field. Therefore, PET systems based on solid-state sensors can be directly integrated with the MRI system, which accelerated the development of the hybrid PET/MRI image modality. There are many advantages for this hybrid system. First, the PET/MRI system can take the advantages of both image modalities; PET can provide a functional image to show the metabolic process while MRI shows the excellent soft-tissue contrast. Second, PET and MRI images acquired simultaneously from the same patient position have higher quality when compared to the fused PET and MRI images taken from individual PET and MRI systems at different times. Third, compared to the PET/CT system, the PET/MRI system can achieve a reduced radiation exposure since there is no X-ray involved in the system. Table 1-2 shows a list of state-of-the-art PET/MRI systems in the commercial market. Similar to the previous section, only the design and performance of the PET system will be compared and discussed, and the spatial resolution, timing resolution and energy resolution values are reported as FWHM.

The Ingenuity TF PET/MRI system is a sequential hybrid PET/MRI system for whole-body imaging. It consists of two standalone systems - a Philips Astonish ToF PET and a Philips Achieva 3T MRI system [11]. In this design, the PET system and MRI system are

installed face-to-face separately and a turntable between the PET and MRI is employed to move the patient into the different systems to acquire the images. The PET system in this hybrid system is based on PMT technology. In order to avoid the mutual system interference, especially the magnetic field interference on the PMTs, a magnetic shielding was introduced in the PET gantry design. The array of LYSO of size 4×4×22 mm^3 was coupled on the PMT module to form a basic detector. The overall PET system achieved a spatial resolution of 4.7 mm, a timing resolution of 550 ps and an energy resolution of <13% [12], [60], [61].

Table 1-2: Comparison of PET performance of state-of-the-art PET/ MRI systems

Ref.	[11], [12]	[13]	[14]
Manufacturer	Philips	GE	Siemens
Model Name	Philips-Ingenuity TF PET/MRI system	SIGNA™ PET/MRI system	Biograph mMR System
Scintillator material	LYSO	LBS	LSO
Scintillator size (mm^3)	4×4×22	4.0×5.3×25	4×4×20
Sensor	PMT	SiPM	APD
Spatial resolution (mm)	4.7	4.57	4.6
Timing resolution (ps)	550	385	N/A
Energy resolution (%)	13	9.4	-

The SIGNA™ PET/MRI hybrid system from GE uses the same ToF PET as in the Discovery MI PET/CT [13]. As mentioned in the previous section, the PET system is developed based on SiPM technology, which is compatible with the magnetic field and there is no requirement for a shielding design of the PET system. Compared to the Ingenuity TF PET/MR system, another advantage of the SIGNA™ PET/MR hybrid system is that it is a concurrent system, acquiring the PET image and MRI image at the same patient position at the same time, resulting in a reduced scan time and an improved accuracy for PET and MRI image co-registration and fusion.

The PET system in Siemens Biograph mMR system is based on APD technology [14]. There is no shielding requirement for the PET system since the APD's performance is not affected by the magnetic field. Even though the PET system cannot use the ToF techniques

due to the poor timing performance of APDs, the system still achieved a spatial resolution of 4.6 mm when using LSO scintillators of size 4×4×20 mm^3.

1.4. Research Challenges and Motivation

Positron emission tomography (PET) imaging is an essential tool in clinical applications for the diagnosis of diseases due to its ability to acquire functional images to differentiate among various metabolic and biological activities at the molecular level. One key limiting factor in the development of efficient and accurate PET systems is the sensor technology in the PET detector. As described in Section 1.2, there are generally four types of sensor technologies used for PET applications: PMTs, APDs, SPAD-based (analog SiPMs and dSiPMs) and CZT detectors. The comparison of these four types of sensors in Table 1-3 shows that the SPAD-based sensor is the most promising sensor technology for future PET imaging system.

Table 1-3: Overall comparison of sensors for PET applications

Type	PMT	APD	SPAD-based		CZT
			Analog SiPM	dSiPM	
Conversion Type	Indirect	Indirect	Indirect	Indirect	Direct
Magnetic field compatibility	No	Yes	Yes	Yes	Yes
ToF capability	Limited	No	Yes	Yes	No
Signal readout	Analog	Analog	Analog	Digital	Analog
Operating voltage	High	Low	Low	Low	High
Compactness	Low	Medium	Medium	High	Medium
Readout electronics	Complex	Complex	Complex	Simple	Very complex
Energy Resolution	Medium	Low	Medium	Medium	High
Spatial Resolution	Medium	Low	Medium	Medium	High
Timing Resolution	Medium	Low	Medium	High	Low

1.4.1. Improve the photon detection efficiency (PDE)

The main goal of this work was to design and develop SPADs using standard deep submicron (DSM) CMOS technologies. Although SPADs can achieve good performance in high-voltage (HV) CMOS and CMOS image sensor (CIS) or other customized technologies; SPADs implemented in standard CMOS technology are preferred due to their lower cost and the potential benefits from the integration of SPAD arrays with other signal conditioning and processing circuits on the same chip. This is especially important because thousands of dSiPMs are needed to build a detector ring for a human-sized PET system. However, unlike the CIS technology in which the passivation layer will be removed on the active sensing area, the passivation layer on the top of the sensing area of the SPADs implemented in standard CMOS process still exists after fabrication. Therefore, a large number of incident photons will be reflected or absorbed when passing through the thick passivation layer and inter-metal dielectric (IMD) layers, resulting in only a small portion of the incident photons on the active sensing area of the SPAD. In addition, the depletion region of the p-n junction of SPADs is usually thin, which leads to a lower photon detection probability in the depletion region. Due to these two main reasons, SPADs implemented in the standard CMOS technology suffer from low PDE performance. In order to achieve a better PDE, several techniques were proposed in the literatures.

First, proper post-fabrication techniques can be used to remove or thin the passivation layer to increase light transmission [62]. Further surface treatments like depositing an anti-reflection coating to reduce the surface reflection or using micro-lens arrays to focus the light onto the active area can be used to boost the light transmission. A SPAD with the micro-lens array to improve the PDE was reported in [63]. By depositing micro-lenses on the surface of the SPAD, the highest effective FF was 84% while the native FF was 28%. However, these post-fabrication steps will increase the fabrication cost. Furthermore, in PET applications the use of micro-lens can still be ineffective in dealing with the light emitting from scintillators under wide angles.

Second, a SPAD array with a 3D structure show great potential to achieve high PDE due to their high FF. In the 3D structure, the SPAD array is implemented in the top tier of

the silicon wafer while the other signal conditioning and processing circuits such as quench and reset circuits, counters and TDCs, are implemented in the second tier of the wafer. Signals between the SPAD array on the top tier and the circuits on the second tier are connected by TSVs (through-silicon vias). The TSV bonding technology has been successfully realized in a commercial vertically integrated sensor system. The 6×6 mm^2 SiPM sensors available from SensL (now part of ON Semiconductor) can achieve a FF of greater than 90% [64]. On one hand, SPAD junctions show good performance by employing less-scaled technologies, which have lower doping concentrations resulting in a thicker SCR. This results in a higher PDE and a lower DCR caused by tunneling. On the other hand, the smaller, advanced CMOS technologies are preferred for the signal conditioning and processing circuits like the TDCs and counters so that they can achieve shorter cell delay times, higher operational frequencies and lower power consumptions. By using a 3D structure, the top and bottom tiers can be optimized and fabricated in separate processes. Thus, compromising the performance between the SPADs and the signal conditioning and processing circuits when they are manufactured in the same process technology is not needed, thus achieving the highest performance in the overall system.

1.4.2. Lower the dark noise

Another challenge is to lower the dark noise when designing SPADs in standard CMOS technologies. Several carrier generation mechanisms such as generation-recombination, band-to-band tunneling (BTBT) and trap-assisted tunneling (TAT) are responsible for the dark noise. For SPADs fabricated in a standard CMOS process compared to a customized process, there is a higher probability to have defects in the depletion region. These defects can function as generation-recombination centers to create discrete generation-recombination energy levels in the forbidden energy bandgap, thus trapping and releasing the free carriers according to the Shockley-Read-Hall (SRH) mechanism to generate dark noise. Also, if the thermally generated free carriers occur within the diffusion length of the depletion region of the SPAD, a false dark noise pulse might potentially be triggered.

Tunneling including BTBT and TAT are other important mechanisms for dark noise. Band-to-band tunneling occurs when there is a high electric field across a strongly reverse-biased p-n junction, resulting in a significant flow of electrons from the p-side valence band (VB) to the n-side conduction band (CB). The trap centers within the forbidden band due to defects and impurities assist the tunneling from the VB to CB, giving rise to trap-assisted tunneling. The DCR from tunneling becomes dominant in SPADs fabricated in DSM CMOS technologies due to the decreased depletion width and abrupt doping profiles [65]. Therefore, these defects can function as trap centers, resulting in a large portion of the dark counts being generated from the trap-assisted thermal GR and TAT.

These trap centers may also capture some carriers during the primary avalanche process and release them at random time intervals [66], [67], thus leading to secondary avalanches. The probability of secondary avalanche triggered by the carriers emitted from the trap centers is defined as AP. Therefore, a higher concentration of defects in the depletion region of a SPAD results in a higher DCR and AP.

In order to minimize the AP, the bias for the SPAD should stay below its breakdown voltage for a sufficient amount of time after quenching to ensure that the trapped carriers can be released without triggering after pulses. This is done using an active quench and reset (AQR) circuit. However, an increased hold-off time prolongs the dead time of the SPAD, thus reducing its counting rate. Moreover, when designing the AQR circuit, one should try to minimize the input capacitance of the circuits, which is the load capacitor of the SPAD, thus lowering the charge generation during the avalanche process and ultimately lowering the trapping probability of carriers.

1.4.3. Improve the timing performance

The third major challenge is to improve the timing performance of CMOS SPADs. In PET applications, the timing information for the detectors is processed by the coincidence unit to identify two coincident gamma rays that come from one true annihilation event. The increased accuracy of detecting the time difference between two coincident gamma rays improves the certainty of the event occurring at a specific position along the LoR. Therefore,

higher time resolution means higher SNR of the PET image. With improved time jitter below 500 ps, it is possible to use the ToF technique in PET to obtain great performance enhancements when compared to the conventional PET. A "10 ps challenge" which aims to achieve a CRT of 10 ps FWHM for ToF application was recently set up [68]. With a 10-ps CRT, a high-sensitivity, reconstruction-less PET scanner would achieve reduced scanning time, scanning cost and radiation dose. In order to achieve this challenge, one of the key issues is to improve the timing performance of the photosensor in the PET detector.

The sources of the timing jitter in a SPAD-based detector can be generally categorized into 4 types: jitter from the SPAD junction of the SiPM pixel, jitter from the comparator, jitter from the TDC and jitter from the connections between different blocks on the timing signal chain. For the SPAD junction, there exists an inherent timing jitter due to the variation of delays between photon absorption and the build-up time of the avalanche pulse [69]. Typically, SPADs based on shallow and thin depletion regions can achieve the best timing performance, but they have low PDE. Junctions biased at a high excess voltage, which have a high electric field in the active region, achieve a lower timing jitter because the avalanche build-up time becomes more certain statistically [70]. However, the high excess voltage will increase the DCR.

The amplitude of the output signal of a SPAD-based sensor varies due to the stochastic uncertainty of the avalanche process. If a general leading-edge discriminator (LED) with a fixed threshold is used, the conversion process from the analog output of a SPAD-based sensor to a digital pulse inevitably gives rise to timing jitter. One possible solution to minimize this jitter is to use a constant fraction discriminator (CFD), which keeps a constant timing output even when the amplitude of the SPAD-based sensor's output signals varies. CFD requires more silicon space due to its complexity, thus lowering the FF of the SiPM, but the space requirement might not be an issue for a 3D dSiPM.

The TDC also contributes timing jitter to the whole system due to the different noise sources (e.g., jitter from the delay lines and reference clock) and performance differences between delay cells because of the parasitic mismatch, and process, voltage and temperature (PVT) variations. By down-scaling the CMOS technology, the digital units in the TDC can

achieve less jitter, less delay and high operating frequency. The 3D dSiPM is a good solution for this issue since the TDCs and SiPM can be implemented in their preferred CMOS technologies. However, dealing with a large number of connections between the SiPM layer and circuit layer is challenging. In addition, special placement and routing of delay cells and dummy cells can be used to match the parasitics of the delay cells, and a delay-locked loop can minimize the influence of PVT variations in many TDC structures.

The timing jitter from the connections on the signal chain can be minimized due to the integration of the SPAD array with TDCs in one chip. The one-chip integration can eliminate many off-chip influences such as I/O pads, bonding wires, and on-board matching issues, which exists in a system using separate SPAD and TDC chips.

This research work aims to investigate SPAD dark noise mechanisms, develop new SPAD structures with novel guard rings, and to design innovative quench and reset circuits using standard CMOS processes to enhance the performance of SPAD detectors. The outcomes of this work can push the limits of these major challenges in CMOS SPAD design, thus improving the SPAD's performance such as photon detection efficiency, dark count rate and timing jitter while keeping the fabrication cost low.

New technologies and higher performance in SPAD-based sensors have great potential to improve the PET imaging resolution and to combine PET with MRI systems to obtain fused PET/MRI images. The improved image resolution and hybrid image modalities will allow doctors to diagnosis diseases at an early stage. In addition, the medical costs of diagnosing and treating diseases in their early stages are much lower than in their latter stages. In this sense, research on SPAD-based sensors can also help to relieve the economic burden of the healthcare system. Moreover, all the SPAD designs in this work are based on standard silicon CMOS technologies, which will significantly lower the price of PET systems. This, in turn, lowers the cost for each PET scan, thus making PET scans more available and affordable to the public.

1.5. Research Contributions

The research work conducted in this thesis was focused on the design of advanced CMOS SPADs and their relevant quench and reset circuits, targeting PET applications. The SPADs and circuits were designed in this research work using three standard CMOS technology nodes: IBM 130 nm CMOS process, TSMC 180 nm CMOS process and TSMC 65 nm CMOS process. The major contributions of this work are summarized as follows:

- **An extensive literature review was conducted on sensors for positron emission tomography (PET) applications.** In this review, four types of sensors used in PET applications: photomultiplier tubes (PMTs), avalanche photodiodes (APDs), silicon photomultipliers (SiPMs) and cadmium zinc telluride (CZT) detectors were studied and discussed in terms of their operating principles, key performance parameters, and PET applications. Then, the sensor technologies used in commercially available state-of-the-art PET systems were presented. Finally, the strengths and weaknesses of these four types of sensors were compared and the research challenges for PET sensors were discussed and summarized.
- **Random telegraph signal (RTS) noise in n^+/p-well CMOS SPADs**. In addition to dark count rate (DCR) and afterpulsing (AP), the RTS noise was studied on a n^+/p-well SPAD fabricated in a standard 130 nm CMOS process. An analytical model was developed to extract the effective dimension of the defects in the depletion region from the curve of the RTS noise current amplitude versus biasing voltage. The results of the DCR and AP from three SPAD samples with different defect dimensions derived from the RTS noise properties showed the trend that RTS noise correlates with the DCR and AP: the SPAD with a larger defect effective dimension shows a higher DCR and AP.
- **Improved noise performance of CMOS poly gate SPAD**. The noise performance of three types of n^+/p-well single-photon avalanche diodes (SPADs) fabricated in a standard 180 nm CMOS technology was studied. The SPADs had different poly gate configurations: no poly gate (SPAD_NG), a dummy floating poly gate

(SPAD_DG), and a field poly gate connected to the n^+ cathode (SPAD_FG). The measurements of dark count rate and afterpulsing showed that the SPAD_DG had better noise performance compared to the SPAD_NG. The dummy poly gate pushed the shallow trench isolation away from the active region of the SPAD, thus reducing the dark noise generated from the Si-SiO_2 interface. The measurements also revealed that the noise performance can be further improved by connecting the poly gate to the n^+ cathode. The voltage on the poly gate in SPAD_FG reduced the electric field in the n-well guard ring (GR) region, thus reducing the number of carriers from the GR region entering the active region of SPADs and initiating dark counts.

- **High-speed SPAD pixel with active quench and reset (AQR) circuit.** A compact, high-speed AQR circuit was designed and fabricated in a standard 65 nm CMOS technology. To investigate the performance of the AQR circuit, it was integrated with a p^+/n-well SPAD to form a SPAD pixel. The post-layout simulations showed that the quenching time for this AQR circuit was only 0.1 ns, and the smallest dead time is 3.35 ns (corresponding to a maximum count rate of ~300 Mcps). The SPAD pixel achieved a dark count rate of 21 kHz, a peak photon detection probability of 23.8% at a 420 nm wavelength, and a timing jitter of 139 ps (using a 405 nm pulsed laser) at 0.5 V excess voltage.
- **Differential quench and reset circuit for SPADs**. A differential quench and reset (QR) configuration consisting of two QR circuits to quench and reset through both cathode and anode terminals was proposed and investigated. Using a SPAD fabricated in a 65 nm CMOS process, the measurement results of the traditional single passive quench and reset (SPQR) circuit and the proposed differential passive quench and reset (DPQR) circuit showed that the SPAD with the DPQR circuit had a reduced reset time, an increased count rate and a decreased afterpulsing. In addition, the differential output pair from the DPQR circuit has the potential to increase the signal-to-noise ratio of the detection system due to its ability to reject common-mode noise and electrical interferences.

Publications:

Published:

1. W. Jiang, Y. Chalich, M. J. Deen, Sensors for Positron Emission Tomography Applications, *Sensors*. 19 (2019). doi:10.3390/s19225019.
2. W. Jiang, Y. Chalich, R. Scott, and M. J. Deen, "Time-Gated and Multi-Junction SPADs in Standard 65 nm CMOS Technology," *IEEE Sens. J.*, pp. 1–1, 2021, doi: 10.1109/JSEN.2021.3063319.
3. W. Jiang and M. J. Deen, "Random Telegraph Signal in n^+/p-Well CMOS Single-Photon Avalanche Diodes," *IEEE Trans. Electron Devices*, pp. 1–6, 2021, doi: 10.1109/TED.2021.3070557.
4. W. Jiang*, S. Majumder*, S. Kumar, S. Subramaniam, X. Li, R. Khedri, T. Mondal, M. Abolghasemian, I. Satia, M. J. Deen, "A Wearable Tele-Health System towards Monitoring COVID-19 and Chronic Diseases," *IEEE Rev. Biomed. Eng.*, 2021, doi: 10.1109/RBME.2021.3069815. (*equal contribution)
5. R. Scott, W. Jiang, and M. J. Deen, "CMOS Time-to-Digital Converters for Biomedical Imaging Applications," *IEEE Rev. Biomed. Eng.,* 2021, doi: 10.1109/RBME.2021.3092197
6. W. Jiang, R. Scott, and M. J. Deen, "Differential Quench and Reset Circuit for Single-Photon Avalanche Diodes," *Journal of Lightwave Technology*, 2021, doi: 10.1109/JLT.2021.3111119

Submitted:

1. W. Jiang, R. Scott, and M. J. Deen, "Improved Noise Performance of CMOS Poly Gate Single-Photon Avalanche Diodes," *IEEE Photonics J.* (Submitted on July 22, 2021)
2. W. Jiang, R. Scott, and M. J. Deen, "High-speed Active Quench and Reset Circuit for SPAD in a Standard 65 nm CMOS Technology," *IEEE Photonics Technology Letters* (Submitted on July 13, 2021)

US Patent:

1. W. Jiang and M. J. Deen, "Differential quench and reset circuits for avalanche photodiodes." (Provisional application, Nov.12, 2020)
2. S. Majumder, W. Jiang, R. Khedri, T. Mondal, I. Satia, M. J. Deen, "A wearable tele-health monitoring system," (Provisional application, Jan. 7, 2021)

1.6. Thesis Organization

In Chapter 1, an overview of PET applications is described, including: background on PET physics, ToF techniques, and PET detectors. Then, 4 types of photosensor technologies used in PET applications were introduced and compared. Next, we described the motivation of this research. Lastly, we provided a summary of the major contributions of this research and the organization of this thesis.

In Chapter 2, the key performance parameters for SPADs are described in detail. Then, three key points for CMOS SPAD design: junctions, guard ring structures, and quench and reset circuits are discussed. Next, the performance of state-of-the-art SPADs in the recent literature is summarized in a table. This table functions as a reference of the performance for SPADs designed and optimized in this research work.

In Chapter 3, the properties of RTS noise in SPADs is investigated by using n^+/p-well SPADs fabricated in a standard 130 nm CMOS process. The RTS noise of a SPAD is found to be correlated with its DCR and AP.

In Chapter 4, three types of n^+/p-well SPADs with different poly gate configurations: no poly gate (SPAD_NG), a dummy floating poly gate (SPAD_DG), and a field poly gate connected to the n^+ cathode (SPAD_FG) were fabricated in a standard 180 nm CMOS technology. The simulations and measurements are performed to investigate the function of a poly gate as a guard ring for n^+/p-well SPADs.

A high-speed and compact SPAD pixel with an active quench and reset (AQR) circuit is proposed in Chapter 5. The SPAD pixel consists of a p^+/n-well SPAD with a p-well guard ring and an AQR circuit. This pixel was designed and fabricated using a standard 65 nm CMOS process. The operational principle is described using the general schematic of the

AQR circuit and the post-layout simulations. The layout of the SPAD pixel is also presented. The SPAD pixel was fully characterized to obtain its performance parameters such as breakdown voltage, DCR, PDP and timing jitter.

In Chapter 6, a differential quench and reset (DQR) configuration is proposed for SPADs. Compared to the traditional single quench and reset, the DQR configuration consists of two QR circuits to quench and reset through both the cathode and anode terminals, thus resulting in a reduced reset time and AP, and improved SNR due to the potential to reject the common-mode noise.

In Chapter 7, a summary of this thesis is presented on the important findings of this research. Based on the achieved results, we provide recommendations for the future work on the design of CMOS SPADs and their related circuits.

Chapter 2
Performance Parameters for SPAD and Literature Review

In this chapter, we will describe the key performance parameters of the SPAD, which include breakdown voltage (V_{BR}), dark count rate (DCR), afterpulsing (AP), dead time (DT), count rate (CR), photon detection efficiency (PDE) and timing jitter (TJ). For each parameter, the definition is first introduced. Then, a brief review explaining the importance of each performance parameter is given. Some representative measurement results are also provided.

After the introduction of the performance parameters, we discuss three key considerations in the design of CMOS SPADs. We first discuss the types of junctions that are available in CMOS processes. Then, different techniques in the design of guard ring structures are investigated. Third, the quench and reset (QR) circuits including passive QR circuits, active QR circuits and time-gated circuits are described and compared. We also provide a table that summarizes the performance of state-of-the-art SPADs reported in recent literature.

2.1. Key Performance Parameters

2.1.1. Breakdown Voltage

In order to ensure Geiger-mode operation, the breakdown voltage of the SPAD junction should first be measured. The breakdown voltage depends on both the p-n junctions biased

----
adapted from
W. Jiang, Y. Chalich, M. J. Deen, Sensors for Positron Emission Tomography Applications, Sensors. 19 (2019). doi:10.3390/s19225019. (Appendix A)
W. Jiang, Y. Chalich, R. Scott, and M. J. Deen, "Time-Gated and Multi-Junction SPADs in Standard 65 nm CMOS Technology," IEEE Sens. J., pp. 1–1, 2021, doi: 10.1109/JSEN.2021.3063319. (Appendix A)

for photon detection and the fabrication details of the CMOS process used for the design. It is generally known that the breakdown voltage of SPADs decreases with the scaling down of CMOS technology because of the increased doping concentrations and reduced junction widths that come with each new generation. Another important parameter for evaluating the performance of SPADs is the temperature coefficient of the breakdown voltage. The breakdown voltage increases with temperature because the increased phonon scattering at higher temperatures make it more difficult for the electrons or holes to reach the energy threshold for avalanche [71], [72]. The breakdown voltage temperature coefficient can vary from a few to tens of millivolts per degree Celsius depending on the materials used, doping concentration and diode structure. For example, the SPAD (p^+/n-well junction) designed in a 180 nm CMOS process had a measured breakdown voltage temperature coefficient of 7.14 mV/°C [73]. Another SPAD using the n^+/p-well junction based on a 130 nm CMOS process showed a similar breakdown voltage temperature coefficient of 7.22 mV/°C [74], [75]. SPADs based on the p-well/buried deep n-well (DNW) junction were reported to have a higher breakdown voltage at room temperature of 20 V (180 nm) in [76] and 12.4 V (130 nm) in [77], with higher temperature coefficients of 40 mV/°C and 20 mV/°C respectively. The breakdown voltage of a SPAD can be measured through the reversed I-V (current-voltage) characteristic. The maximum derivative of the I-V curve corresponds the breakdown voltage. Figure 2-1 shows an example of the I-V characteristic of a n^+/p-well SPAD in which the breakdown voltage is 11.354 V.

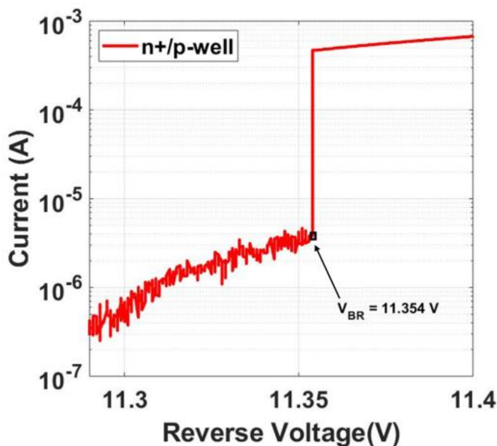


Figure 2-1: Example of the reversed I-V characteristics of a n^+/n-well SPAD. Note that the *x*-axis is expaned to clearly show V_{BR}.

2.1.2. Dark Count Rate

Since a SPAD's gain is theoretically infinite, it operates digitally by counting incident photons instead of outputting an amplified current such as with APDs. Therefore, like PMTs, the SPAD's noise performance is characterized by a DCR, defined as the avalanche pulses per second when there is no illumination. As shown in Figure 2-2, there are several carrier generation mechanisms responsible for the DCR, such as thermal generation-recombination (case 2), trap-assisted thermal generation (case 3), band-to-band tunneling (case 5) and trap-assisted tunneling (case 4). In addition, thermal generation can also occur in the neutral region, and when it is near the depletion region, carriers can diffuse into the depletion region to cause dark counts (case 1).

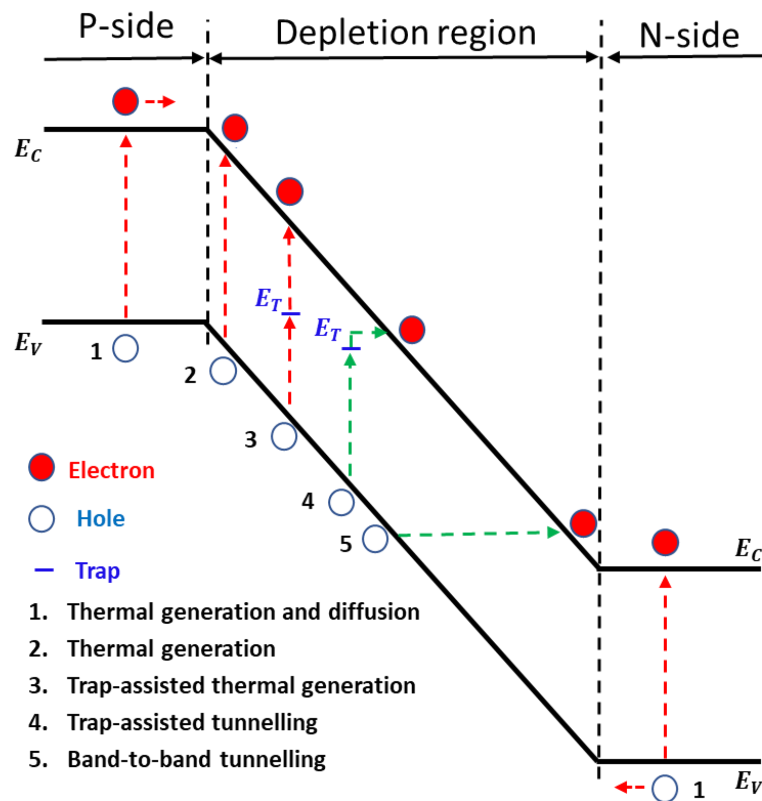


Figure 2-2: Illustration of mechanism for dark noise generation.

Among these noise generation mechanisms, the main mechanism at room temperature is generation-recombination, where thermally generated free carriers are created within the

depletion region or a diffusion length of the depletion region of the SPAD [78], [79]. Therefore, the DCR will increase with the operating temperature. Moreover, for a SPAD fabricated in a standard CMOS process, there exists higher level of impurities and defects which induce many energy levels in the forbidden band. This increases the probability of tunneling and thus leads to a higher DCR. Higher excess bias voltages also increase DCR since a higher electric field across the SPAD junction not only increases the likelihood of avalanches from photon generation, but also from noise sources. Figure 2-3 shows an example of the DCR dependence on excess voltage (V_{ex}) at different temperatures. The DCR for current SPADs fabricated in CMOS technology can be optimized down to tens of Hz at room temperature [80], [81].

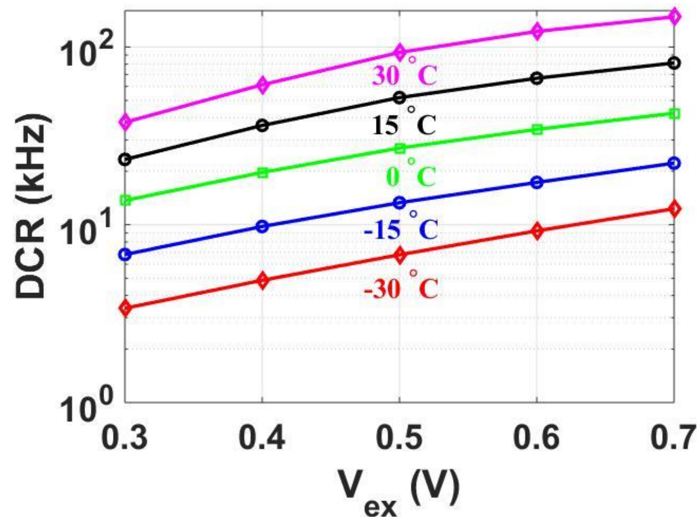


Figure 2-3: Dark count rate as a function of excess voltage at different temperatures.

2.1.3. Afterpulsing

A performance parameter associated with DCR is the afterpulsing probability (AP); that is, the probability of secondary avalanches caused by the emission of trapped carriers during the primary avalanche. The defects and impurities in the depletion region can trap some carriers generated during the avalanche process, and these trapped carriers can be released later at random time intervals [66], [67]. If the biasing voltage of the SPAD is completely recharged before all the trapped carriers in the trap centers have been released, a secondary

avalanche might occur. The most efficient way to lower the AP is to optimize the SPAD front-end circuits.

Typically, there are two main types of front-end circuits: passive quench and reset (PQR) and active quench and reset (AQR). A time-gating technique in [74] is also used to reduce the AP of a SPAD implemented in a 130 nm CMOS process. It was shown that with the proper design of quenching and reset circuits, an AP of <1% can be achieved [82], [83]. The AP is usually measured using the inter-arrival-time (IAT) method in which the time differences between many consecutive SPAD's output pulses are recorded to form a histogram. Then, the primary DCR and AP can be derived from the histogram of the IAT data by a multi-exponential fit. The details of how to use IAT method to extract the AP was described in [84]. Figure 2-4 shows one example of AP measurement results using the IAT method. In this example, the primary DCR is 1.35 kHz and the AP is 6.32%.

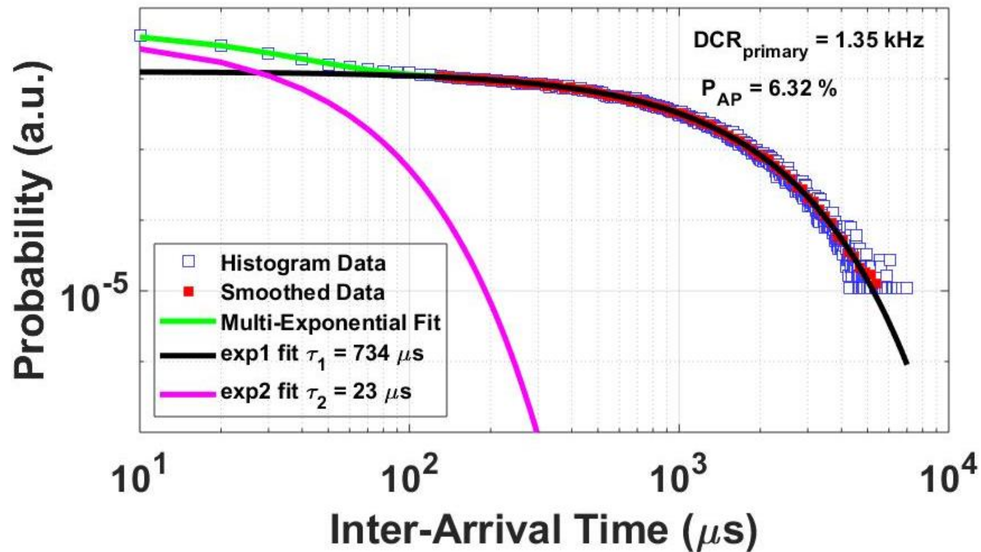


Figure 2-4: Histogram of inter-arrival-time with a double-exponential fit to measure the afterpulsing.

2.1.4. Dead Time and Count Rate

As illustrated in Figure 1-8, the SPAD is biased above the breakdown voltage to operate in the Geiger mode. Under this condition, when an avalanche is triggered, a quench circuit is needed to stop the self-sustaining avalanching process. Then, a reset circuit is required to bring the quenched SPAD back to its initial biasing condition for the detection of next photon.

The time for completing a detection cycle (i.e. from the initiating time of the avalanche to the time when the SPAD is reset) is defined as the dead time (DT). Due to the large quench resistor (tens to hundreds of kΩ), SPADs with passive quench and reset (PQR) circuits usually have a long dead time in the range of hundreds of nanoseconds to microseconds. While the quench and reset times are controllable in the active quench and reset (AQR) circuit, the DT can be designed down to several nanoseconds. Figure 2-5 shows an illustration

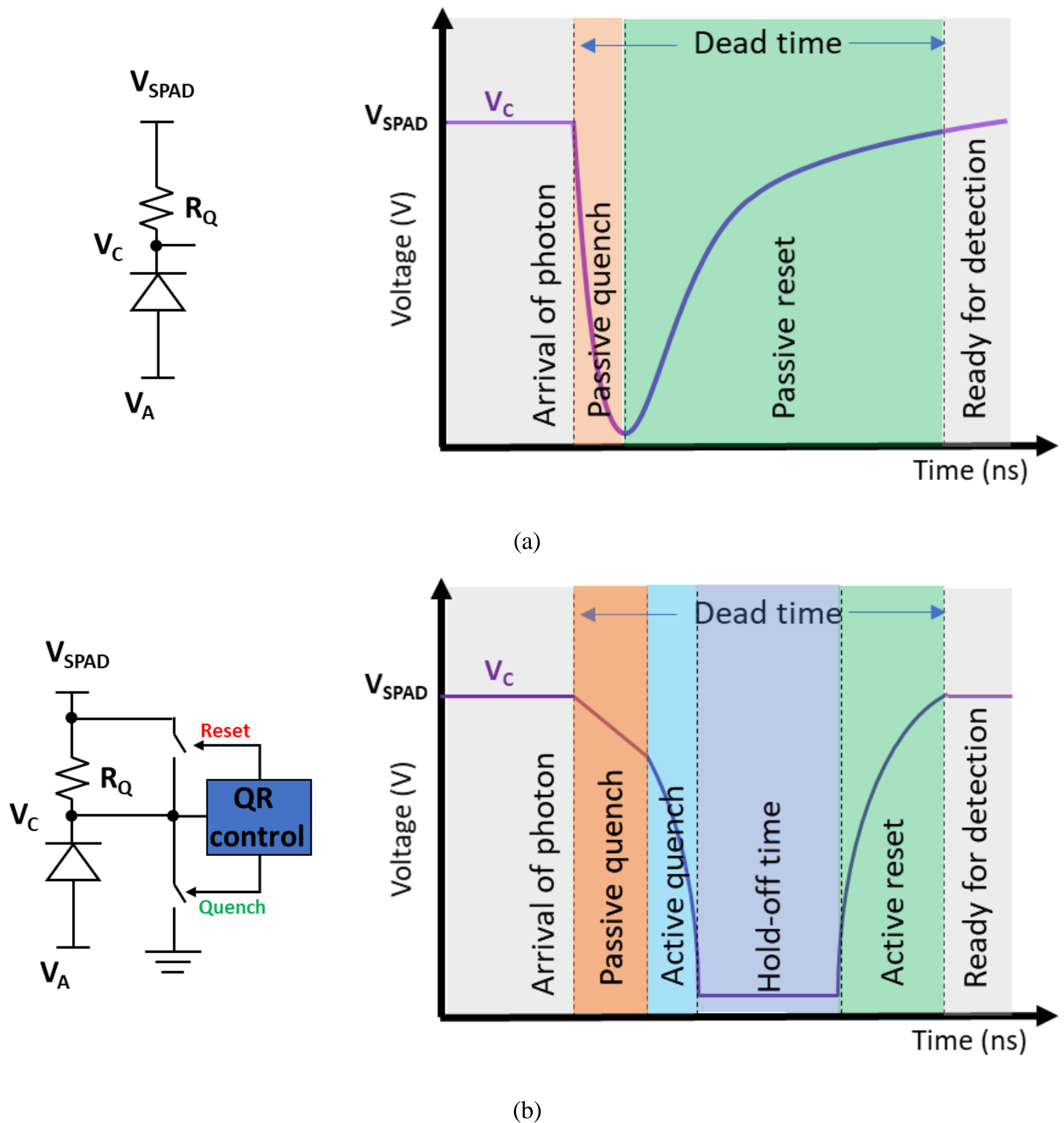


Figure 2-5: (a) Schematic diagram of a SPAD with passive quench and reset circuit and the corresponding waveform of ouput showing the dead time. (b) Schematic diagram of a SPAD with active quench and reset circuit and the corresponding waveform of ouput showing the dead time.

of the DT of SPADs with PQR and AQR circuits. In the PQR configuration, when the SPAD is triggered by an arriving photon or a dark carrier, the SPAD will undergo passive quench first, and then passive reset. The DT in this case is the sum of the passive quench and reset times. In the AQR configuration, the SPAD will first be quenched through the quench resistor passively, and then by the active quench switch once the active quench control signal is enabled. After quench, the SPAD will usually stay below the breakdown voltage for some time (known as hold-off time) in order to lower the AP. Then, after the active reset control signal is enabled, the SPAD will reset back to its initial biasing condition and be ready to detect the next photon. The DT in the AQR configuration is the sum of the passive quench time, active quench time, hold-off time and active reset time. Note that a SPAD can only respond to the first triggering photon and cannot distinguish two or more photons arriving within its DT since the SPAD does not finish the detecting cycle for the first triggering photon. This means that the dead time determines the maximum frequency of operation of the SPAD, which is known as count rate (CR). It is worth noting that the hold-off time in the AQR circuit can be set long enough to make sure the majority of trapped carriers from the trap centers are released before the SPAD is reset. This helps to reduce the AP. However, the prolonged hold-off time means a long DT and reduced CR. Therefore, the trade-off between the AP and the CR needs to be taken into consideration in the design of the quench and reset circuit [31].

2.1.5. Photon Detection Efficiency

The photosensitivity of a SPAD is defined by its PDE, which is the ratio of the number of detected photons to the number of total incident photons. The PDE is the product of the geometric fill-factor (FF) (the ratio of photo-sensitive area to total pixel area), absorption probability (wavelength dependent), and avalanche triggering probability [85]. The SPADs implemented in standard silicon CMOS technology usually suffer from a low PDE due to two main reasons: the photon reflection and absorption of a thick passivation layer and the internal dielectric oxide insulation layers, and a thin depletion region. Figure 2-6 shows a

cross-sectional view of the structure of the IBM 130 nm CMOS technology, which shows 8 inter-metal dielectric (IMD) layers and one passivation layer [86].

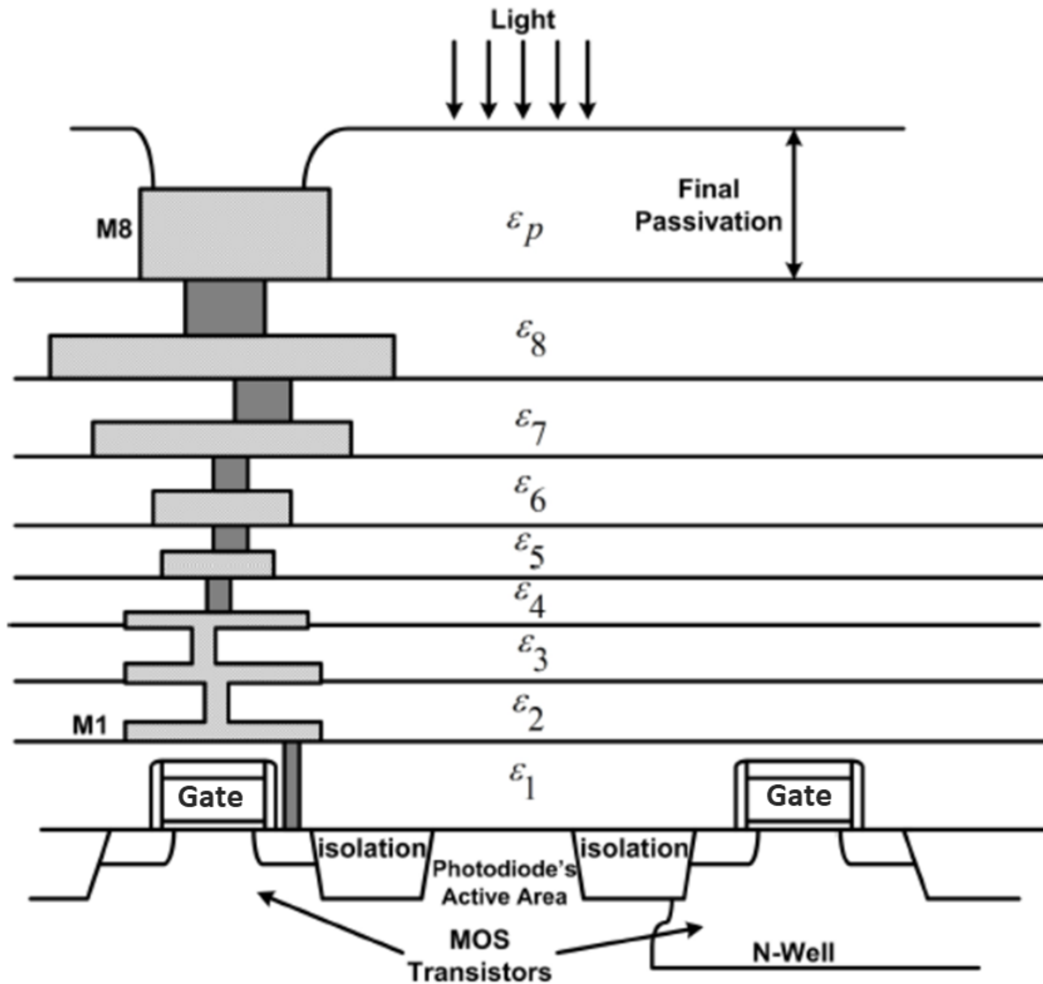


Figure 2-6: Passivation layer and inter-metal dielectric layers in IBM 130 nm CMOS process [86].

2.1.6. Timing Jitter

Timing jitter is the variation in the delay between the output pulse of the SPAD relative to the actual photon absorption time. The timing resolution of a SPAD (as well as an APD) is limited by both the avalanche build-up time and the transit time across the device [29], [87]. The exponential current increase in Geiger-mode operation allows the SPAD to have a superior timing performance compared to the APD. Although the thin depletion region of a SPAD can hinder its PDE performance, especially as silicon technology nodes scale down,

it offers the benefit of reduced transit time since the charge carriers will take less time to be collected at the edges of the depletion region. Also, with the increase of biasing voltage of the SPAD, the turn-on transient of the avalanche current becomes faster, typically lasting tens of picoseconds [88]. A timing jitter in the range of 10 - 100 ps can be achieved in a SPAD. A typical timing jitter measurement result is shown in Figure 2-7, in which the full-width at half-maximum (FWHM) of the timing jitter is 148 ps. There is a Gaussian peak and a diffusion tail in the figure. The Gaussian peak is from photons that are absorbed in the depletion region. The fast response triggered by these photons has a timing jitter which is only determined by the statistical fluctuations of the build-up time of the avalanche, thus presenting a Gaussian peak. However, photons that are absorbed in the neutral region have to diffuse to the depletion region to trigger an avalanche, thus resulting in a diffusion tail in the histogram of the timing jitter [89].

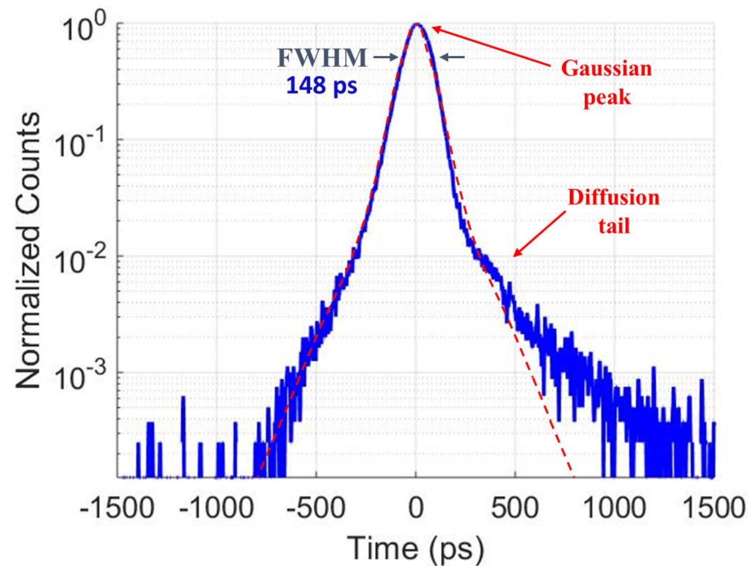


Figure 2-7: Example of the timing jitter measurement on a p^+/n-well SPAD using a 405 nm pulsed laser.

2.2. CMOS SPAD Design Considerations

2.2.1. Junctions

The most common scintillation crystal used in PET applications is L(Y)SO. This crystal converts the high energy gamma rays into visible light with a peak wavelength at 420 nm. In

order to achieve the highest detection efficiency, the PDE peak of the SPAD should ideally be at the same wavelength. This peak wavelength is determined by the depth of the p-n junction depletion region since the absorption of certain wavelengths of light in silicon varies as a function of depth. However, the depletion region depth is determined by the doping concentrations and profiles which are fixed by the foundry in a given standard CMOS process. In order to achieve a peak PDE at a 420 nm wavelength for PET applications, different p-n junctions with varying depths in standard CMOS processes need to be investigated. Figure 2-8 shows the available options for p-n junctions in the standard TSMC 180 nm CMOS technology, which include n^+/p-well, p-well/DNW, n-well/p-substrate p^+/n-well, and DNW/p-substrate. Comparing the PDE versus wavelength of different junctions, an optimal junction can be found for a given standard CMOS process.

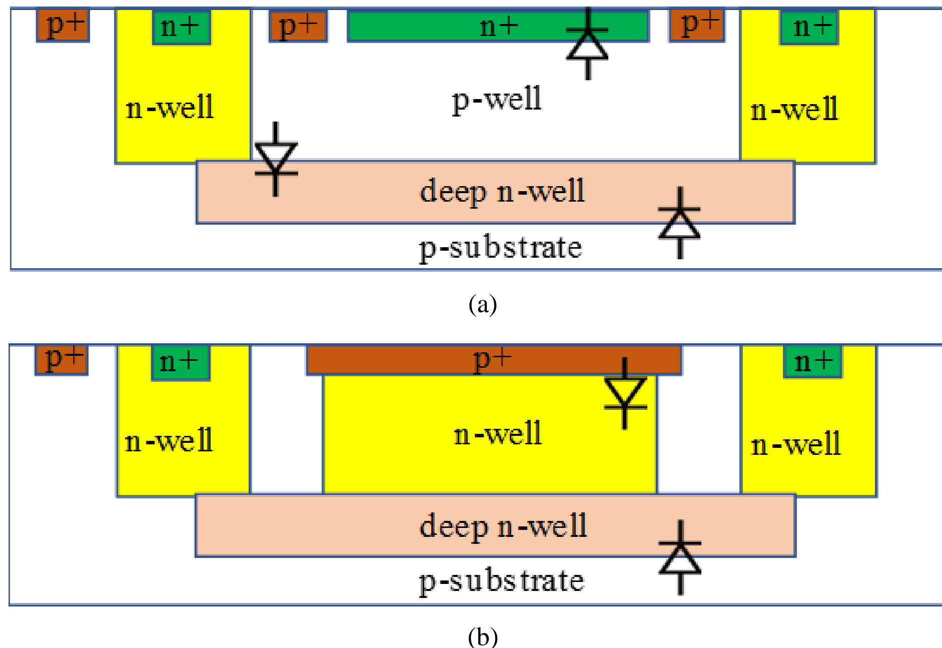


Figure 2-8: Cross-sectional view of p-n junctions in the standard TSMC 180 nm CMOS technology: (a) Three diodes at the n^+/p-well, p-well/DNW, and DNW/p-substrate junctions; (b) Two diodes at the p^+/n-well and DNW/p-substrate junctions.

In most SPAD applications, only one p-n junction is used for photon detection. However, multi-junction structures are also a possible solution for some applications [90]–[92]. The principle of this structure is that several p-n junctions are designed and stacked vertically, with all the junctions independently biased to ensure simultaneously operation. Compared to

the one-junction structure, the extra junctions in this structure will increase the PDE. In [91], a dual-junction (p^+/n-well and n-well/p-substrate) SPAD implemented in 250 nm CMOS process achieved a peak PDE of 19% for the shallow diode at 400 nm and 29% for the deep diode at 500 nm. Another dual-junction (p-well/DNW and DNW/p-substrate) SPAD implemented in 130 nm CMOS process achieved the highest PDE of 31% at 460 nm for the shallow junction, 30% at 660 nm for the deep junction and 38% at 640 nm when both SPADs were working [92]. Besides dual-junctions, the triple-junction SPADs are also an available option for standard CMOS technology [93]. However, the multi-junction structures require complicated biasing circuits to ensure that the multiple junctions work in the Geiger mode as well as an ac-coupled output strategy in order to feed the outputs of the SPADs into the following signal conditioning and processing circuits.

2.2.2. Guard Ring Structures

In SPADs, there usually exists a higher chance for early avalanche breakdown at the sharp edges of the depletion region depending on its layout and shape. This is known as premature edge breakdown (PEB). With PEB, the SPAD cannot be biased above its breakdown voltage while maintaining a uniform electric field across its entire areal depletion region, which means that the SPAD cannot function correctly. To overcome this issue, guard rings (GRs) are used in many SPAD designs. One important GR structure is shallow-trench isolation (STI).

In deep submicron (DSM) CMOS technologies, STI is traditionally used for providing the isolation between transistors and to also improve the density of transistors. In a SPAD design, STI can be used to eliminate PEB in p^+/n-well or n^+/p-well junctions since the edges of the depletion region are confined by the oxide trench to prevent the formation of sharp edges. Figure 2-9(a) shows one example of the cross-sectional view of a n^+/p-well SPAD with STI GR.

The first STI-bounded SPAD implemented in a DSM non-high-voltage 180 nm CMOS technology was reported in [94]. The silicon-dioxide trench guard ring structures was able to withstand a significantly higher electric field, resulting in reduced SPAD size, reduced space

between pixels and thus improved FF. The electrical characterization of this STI-bounded SPAD showed a reliable operation over 5×10^{10} cycles at room temperature [94]. However, the STI is known to result in many defects at the Si-SiO_2 interface, which function as carrier generation-recombination centers. Therefore, if the STI directly contacts with the active region of the SPAD, the carriers generated at the interface can easily enter the depletion region of the diode, which results in an increase in the DCR and AP. To mitigate this issue, the authors in [95] showed a guard ring structure which combined the STI and a special p-well passivation in a 130 nm CMOS imaging sensor (CIS) process. Compared to the traditional STI-bounded SPAD, the DCR of the SPAD with this new STI guard ring was reduced by more than an order of magnitude. However, this p-well passivation implant is only available in a CIS process, which is more expensive than a standard CMOS process. In addition, masks (for example, the poly gate layer) blocking the formation of STI near the active region were proven to be an effective method to alleviate this issue [77], [96]–[99]. In Figure 2-9(b), a poly gate is placed at the edge of the n^+ layer to push the STI away from the depletion region of the SPAD.

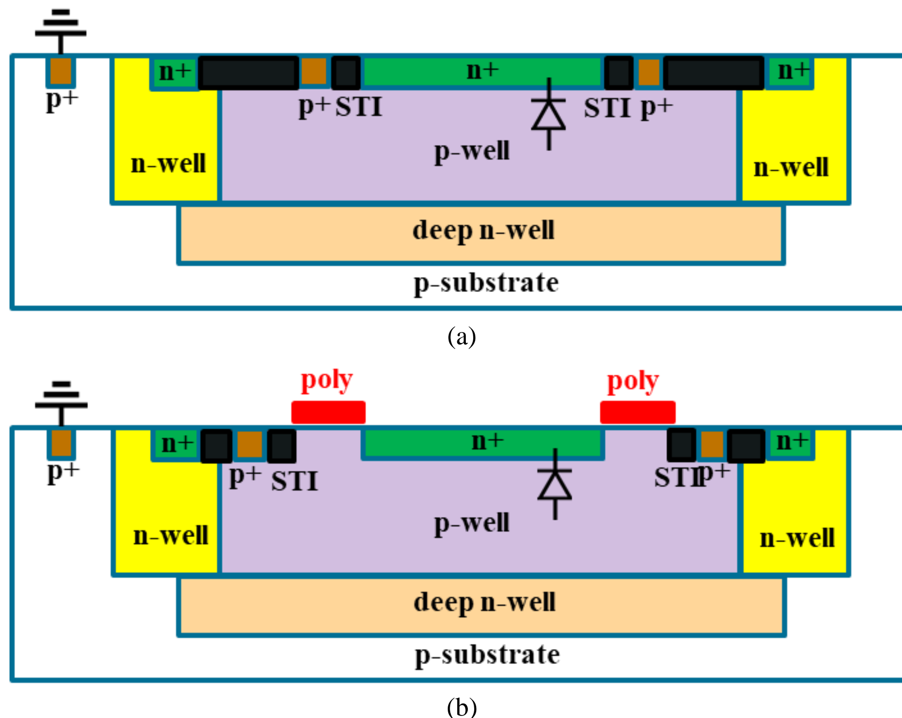


Figure 2-9: Cross-sectional view of a n^+/p-well SPAD with (a) STI guard ring, and (b) poly gate placed at the edge of n^+ layer to push the STI away from the depletion region.

A diffusion GR structure is another structure widely used in SPAD design, such as the p-well diffusion GR for the p^+/n-well junction and the n-well diffusion GR for the n^+/p-well junction, as shown in Figure 2-10. In Figure 2-10(a), a SPAD with diffusion p-well implants as a GR was implemented in a standard CMOS technology. The p^+ implantation was enclosed by the p-well implants in an n-well. Due to the low doping profile in the diffusion p-well, the electric field at the device edges was reduced to ensure the initiation of the avalanche was only within the planar depletion region. The same technique was also employed in SPAD designs in DSM standard CMOS processes [100]–[102]. A diffusion n-well guard ring structure was demonstrated in [31], as shown in Figure 2-10(b). The n^+ implantation was surrounded by the n-well in a p-well. In order to reduce side effects of the STI, the n^+ implantation extended into the n-well by 0.75 μm, resulting in the STI being positioned at a far enough distance away from the active region. In this way, carriers generated at the Si-SiO_2 interface are more likely to undergo recombination rather than diffuse into the active region, thus reducing the DCR.

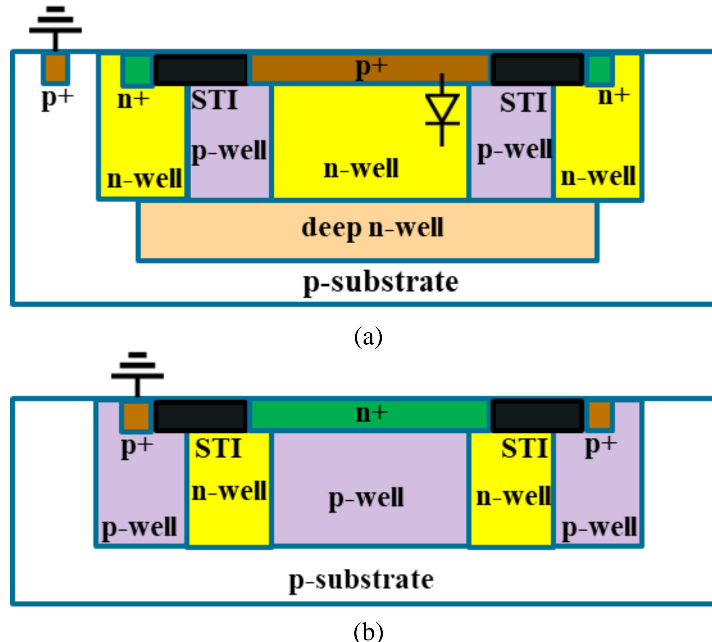


Figure 2-10: Cross-sectional view of (a) a p^+/n-well SPAD with p-well diffusion guard ring, and (b) a n^+/p-well SPAD with n-well diffusion guard ring.

A novel virtual GR in the 130 nm CMOS process was introduced in [77]. The virtual guard ring was implemented by using the retrograde doping profile of the deep buried n-well,

which is an available feature in the triple-well CMOS technology (< 250 nm). Due to the lower doping profile of the DNW to the surface, PEB can be prevented at the periphery of the p-n junction. Compared to a SPAD with a diffusion guard ring, the virtual guard ring allowed the active area of SPAD to be scaled down to a much smaller size [80]. With the use of a virtual guard ring, the SPAD size can be reduced to 2 μm with improved DCR, timing jitter and yield in the DSM technology because of the lower probability of having defects in a smaller active region [77]. Another virtual GR for a p-well/DNW junction was implemented using the lower doping concentration resulting from the diffusion between DNWs [103]. The TCAD simulation and measurement results proved the effectiveness of this virtual GR. Figure 2-11 shows one example of a virtual GR implemented by the diffusion between n-wells.

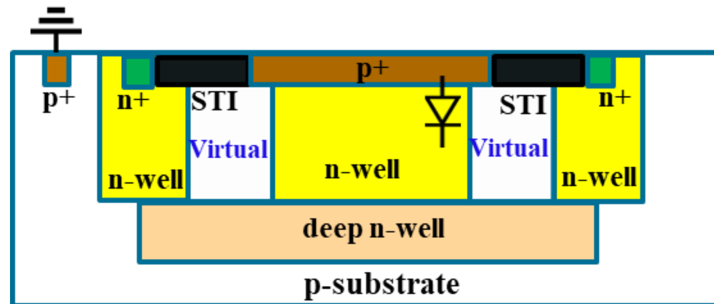


Figure 2-11: Cross-sectional view of a p^+/n-well SPAD with virtual guard ring.

The poly gate design, shown in Figure 2-9(b), can be used to push the STI away from the depletion region to reduce the dark noise. In addition, a voltage can be applied to the poly gate to improve the performance of the GR. A perimeter-gated p^+/n-well SPAD fabricated in a 0.5 μm CMOS process was reported in [104]. This SPAD had a lateral diffusion n-well guard ring with a perimeter gate over the edge of active region. When a negative voltage is applied to the perimeter gate, the TCAD simulation showed a reduced electric field distribution at the edge of the SPAD, thus proving the prevention of PEB. This perimeter-gated SPAD also showed a reduced DCR because of the voltage applied on the gate. However, an extra pin was needed to bias the gate of the SPAD. To solve this issue, a p^+/n-well SPAD with a field poly gate connected to its p^+ anode was proposed in [105] using a 180 nm BCD (BIPOLAR-CMOS-DMOS) process. In addition to this field gate SPAD, two other types of SPADs: one common p^+/n-well SPAD with a n-well diffusion GR, and one p^+/n-well SPAD

with a dummy floating gate (no voltage applied on the gate), were fabricated to perform a performance comparison. The measurements showed that noise performance of the SPAD can be improved by using the dummy poly gate, and can be further improved by connecting the poly gate to the anode. Figure 2-12(a) and (b) show the cross-sectional view of n^+/p-well and p^+/n-well SPADs with field gate GRs, respectively.

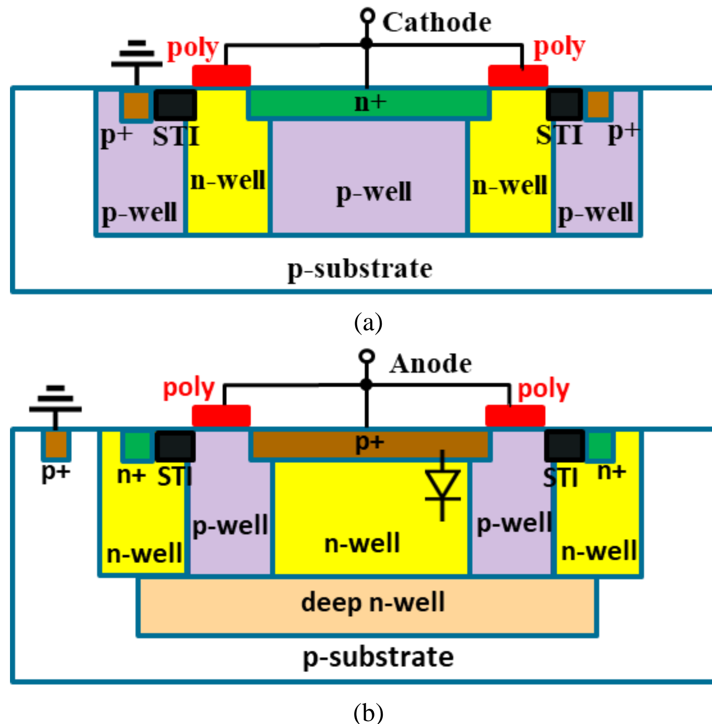


Figure 2-12: Cross-sectional view of (a) a n^+/p-well SPAD with its poly gate connected to the cathode and (b) a p^+/n-well SPAD with its poly gate connected to the anode.

2.2.3. Quench and Reset Circuit

For proper operation, SPADs require a quenching circuit to stop the avalanche process (triggered either by a photon or noise), and a reset circuit to recharge the SPAD and bring it back to Geiger mode for the next photon detection. The time between the start and stop of the avalanche is called the "quenching time". The reset time is the time needed to bring the SPAD back to the initial detection state by recharging the biasing voltage back to ($V_{ex} + V_{BR}$). In general, the quench and reset circuits can be divided into two categories: passive quench and reset (PQR) and active quench and reset (AQR).

Passive Quench and Reset:

The PQR circuits are generally implemented by using a large quench resistor (R_Q). Figure 2-13 shows a simplified schematic of a PQR circuit and its equivalent model. As shown in the model, when an avalanche occurs, the charges on the C_{SPAD} (capacitor of the SPAD) will discharge through the parallelly connected R_D (SPAD resistance) and R_Q (quenching resistor), thus resulting in an exponential decrease of the excess voltage with a final current approximately dependent on the bias voltage and the quench resistor. If this current is large enough, the avalanche is self-sustaining since enough charges are present in the depletion region, but below a certain value there exists a high probability that no carriers remain in the region after a random time, resulting in a quenched avalanche. Even though the value of this threshold is not well determined, a value of 100 μA is usually used in many SPAD analytical calculations and simulation models [32], [106]–[108]. The threshold of the quenching current determines the minimum value for the quenching resistor, which usually ranges from 50 to 500 kΩ for thin junction SPADs [109]. An approach by using the measured I-V characteristics curve from the SPAD to determine the optimum value for the quench resistor was shown in [110]. PQR circuits have been widely used in SPAD pixel design since they are simple and only occupy a small area, resulting in a higher FF and decreased parasitics [111]. The PQR circuits are even employed in commercially available analog SiPMs [112], [113]. However, PQR does have some disadvantages.

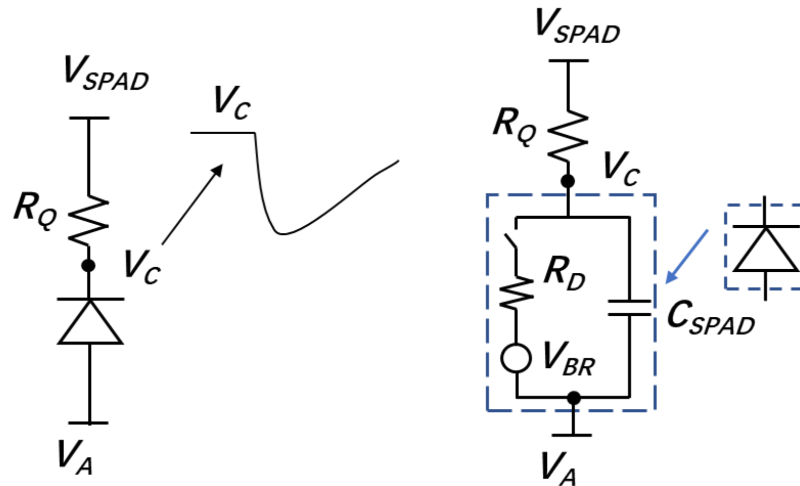


Figure 2-13: Simplified schematic for PQR and the equivalent circuit simulation model.

In PQR, the value of the quenching resistor should be large enough to ensure a fully quenched avalanche process, thus leading to a long recharge RC time constant. This time constant means a long dead time (DT) and a reduced count rate for the SPAD. A typical DT for PQR is several hundred nanoseconds. For example, a SPAD with a PQR circuit implemented in 130 nm CMOS process had a DT of 450 ns [114]. In addition, a long reset time can give rise to an early triggered avalanche during the voltage reset process, resulting in an undetectable low output or paralyzed DT behavior [115]. SPADs with a long reset time cannot function properly at a high-count rate since the excess voltage is unable to recover to the originally biased value before the arrival of the next photon. In order to overcome these drawbacks, AQR circuits were intensively investigated for CMOS SPADs.

Active Quench and Reset:

AQR circuits have been widely employed to achieve a better performance in terms of DT [115], AP [74], [116], [117] and time-gating capabilities [74], [118]. Figure 2-14 shows the basic concept of the active quench and reset. When the avalanche is detected, the output of the SPAD will be used to generate the quench pulse to reduce the biasing voltage of the SPAD below the breakdown voltage, and then the reset pulse to bring the excess voltage back to the initial state. Generally, after quenching, the biasing voltage will stay below breakdown voltage for a predetermined time (hold-off time) before recharging. The carriers released from the trap centers within the hold-off time will not cause secondary pulses since the biasing voltage is below breakdown. A longer hold-off time leads to a reduced probability of AP, but also a longer DT, thus reducing the count rate of the SPAD. A typical value of the hold-off time for AQR is tens of nanoseconds, which is generally much lower than the DT of PQR. The hold-off time of a SPAD with fully integrated AQR circuits (implemented in a HV 180 nm process) could be set as low as 10 ns [119]. Another drawback of AQR circuits is that they require additional area to be implemented and are usually placed inside the pixel, ultimately reducing the pixel's fill factor (FF). SPADs with AQR fabricated in early sub-micron CMOS processes had small FFs due to the large size of transistors and design rule constraints. However, this limitation is reduced with the dimension shrinking of CMOS technology. A trend that FF increased with the scaling down of technologies was shown in

[120]. The FF is around 1% in 800 nm CMOS technology, 9% in 350 nm CMOS technology, 25% in 130 nm CMOS technology and 35% in 65 nm CMOS technology. However, compared to the advanced CMOS processes (e.g. 45 nm CMOS or 65 nm CMOS), SPAD pixel's light detection performance is better in less scaled technologies such as 180 nm CMOS and 350 nm CMOS, which have lower doping concentrations resulting in a thicker depletion region. As a result, higher photon detection probability (PDP) and a lower DCR can be expected. In order to deal with this dilemma, SPADs with a 3D structure were proposed, in which the top tier of the semiconductor only has the SPAD array with improved photodetection capabilities while the quenching and reset circuits can be implemented in the bottom tier usually with a more advanced technology node to benefit from their increased speed and lower power. Signals between the SPAD array on the top tier and the circuits on the second tier are connected by through-silicon vias (TSVs). A back-illuminated 3-D stacked SPAD was reported in [120]. The SPAD arrays, implemented on the top tier by using 45 nm CIS technology, were stacked on top of a bottom tier containing the quench and reset circuits implemented in 65 nm standard CMOS technology. This design achieved a high FF up to 60.5%. According to the achieved results, a SPAD with optimized design parameters can achieve an FF higher than 70%. However, the fabrication, assembly and packing are more costly for the 3D SPAD structure.

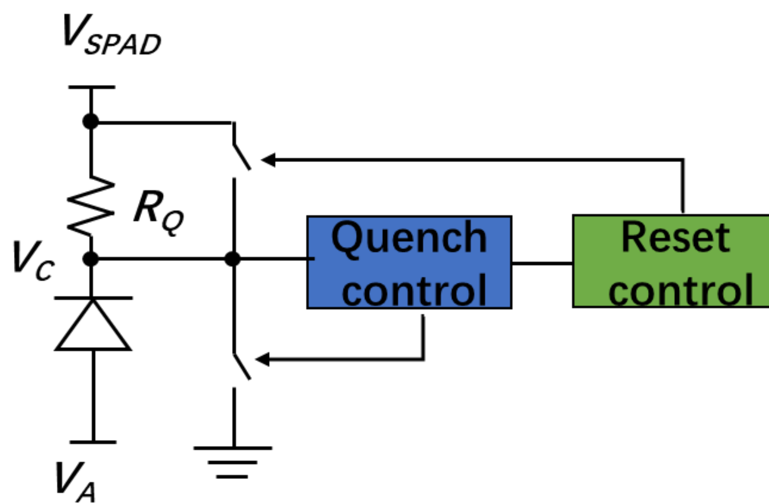


Figure 2-14: Simplified schematic for an active quench-reset (AQR) circuit.

Time-gated Circuit:

SPADs with time-gated (TG) circuits can run in the so-called TG mode in which the SPADs only operate within a specified gate window. This helps to improve the noise performance of the SPAD since only the dark noise within the gate window can influence its signal-to-noise ratio (SNR). The TG circuit can be regarded as a special type of AQR circuit since the SPAD is actively quenched and reset at the start and end of the gate window. Figure 2-15 shows a simplified schematic of a TG circuit. Only when the gate is enabled to close the gate switch, will the SPAD detect photons.

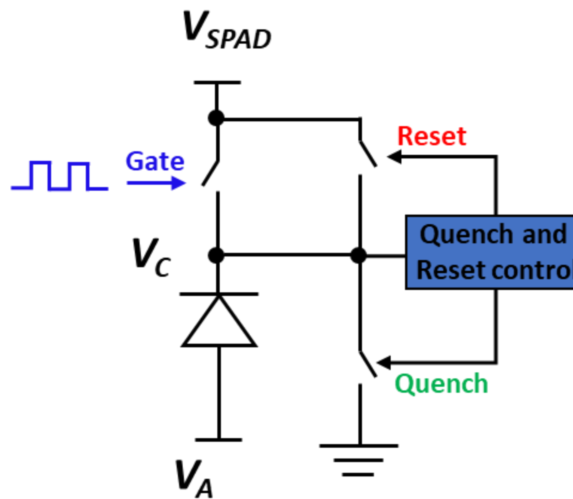


Figure 2-15: Simplified schematic for a time-gated (TG) circuit.

A design example [121] of a TG SPAD pixel that was implemented in a standard 65 nm CMOS process is shown in Figure 2-16. Its post-layout simulation results with a 100 MHz input clock are shown in Figure 2-17. The TG circuit uses only 5 transistors and requires 3 pulses (P1, P2, and P3) to operate in the TG mode. As shown in Figure 2-17, P1 is initially set high to open PMOS M1 to disconnect the cathode from the biasing voltage (V_{SPAD}) and hold the cathode voltage (V_C). In the meanwhile, P2 and P3 are also held high to close NMOS M2 and M5, and open PMOS M3 to keep the output (V_{OUT}) at the ground level (*GND)*. When the operation initiates, P1 and P2 become low simultaneously to open M2 to disconnect the cathode from *GND* and to close M1 to connect the cathode to V_{SPAD}. P1 only lasts ~500 ps to charge the cathode voltage to V_{SPAD} while P2 keeps low for the

entire cycle, thus holding the cathode voltage to arm the SPAD. After P1 goes high, P3 immediately becomes low to enable readout circuit to detect the avalanche if the SPAD is

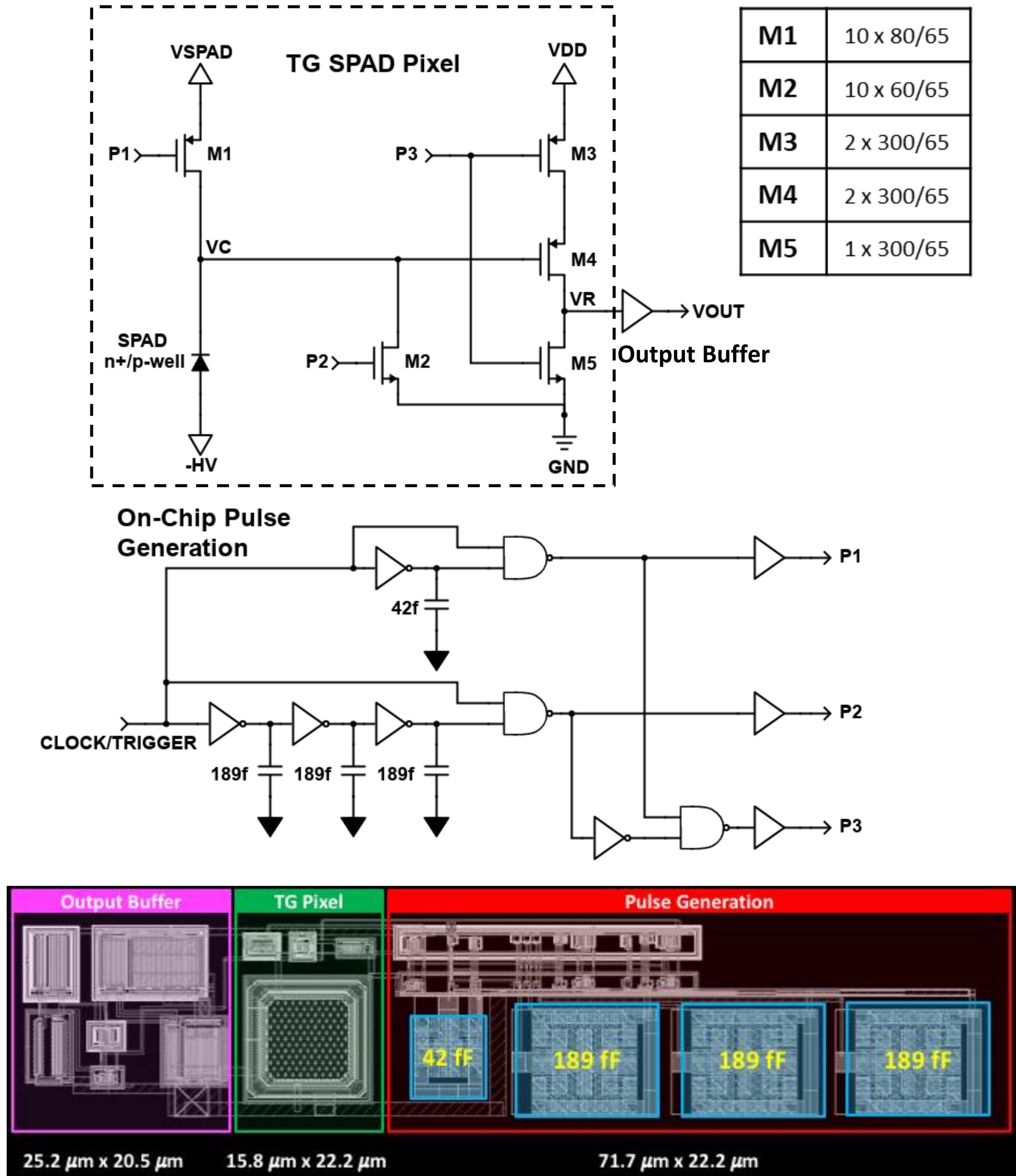


Figure 2-16: Schematics of the time-gated (TG) pixel with the sizes of MOSFETs in nanometer, and pulse generation circuit and its corresponding layout.

triggered during the gate window. Finally, P2 and P3 become high again to end the gate window and bring the cathode voltage to *GND* to quench the SPAD. The operation cycle is then repeated for the next gate window. In this TG circuit, the gate window is the time between the rising edges of P1 and P3, which was designed to be 3 ns.

Besides the TG circuits, an on-chip pulse generator was designed to generate P1, P2 and P3. The pulse generator was implemented using NAND gates and delay units. Additionally, a sized-up output buffer was placed at the end of output in order to drive the large probe capacitance for the measurements.

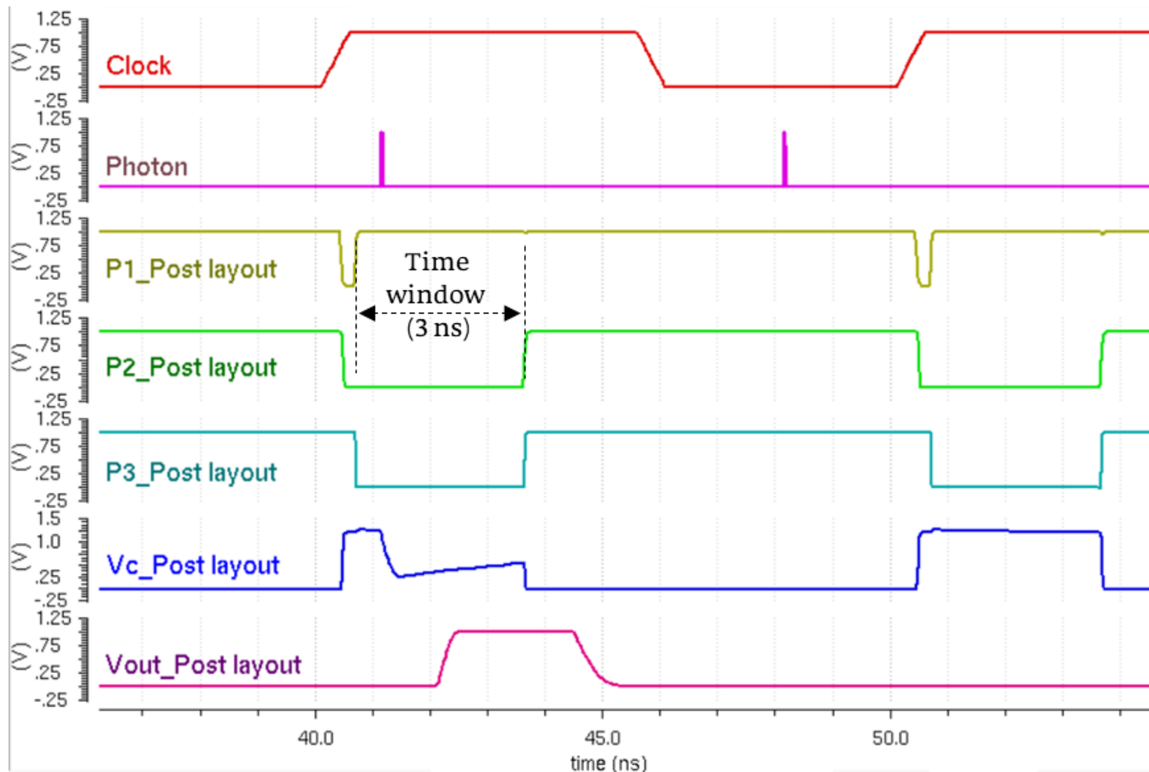


Figure 2-17: Post-layout simulation results showing the operation of the time-gated (TG) circuit.

2.3. Review of State-of-the-art SPADs

Table 2-1 (placed at the end of this chapter) lists the specifications of SPADs in the recent publications. From this table, we can see that the SPADs are implemented using various junctions (p-well/DNW, p^+/n-well, p^+/DNW, p-well/Niso, etc.) in different scaled CMOS technologies. Most the shapes of the active area of SPADs are circular in order to prevent

PEB and the diameters of the active area are between 8 μm to 20 μm to ensure a reasonable DCR. The breakdown voltage is between 10 V and 30 V, which means that the biasing voltage for SPAD in CMOS technologies is very easy to implement and safer when compared to the high operating voltage of PMTs. As the technology scales down, the breakdown voltage has a trend of decreasing due to higher doping profiles and thinner depletion of the junction. A PDE around 40% can be achieved with a moderate excess voltage at room temperature. The time resolutions in CMOS SPADs are mostly less than 100 ps, which make SPADs favorable sensors for ToF PET applications.

2.4. Conclusions

In this chapter, we described the key performance parameters for SPADs including breakdown voltage, dark count rate, afterpulsing, dead time, count rate, photon detection efficiency and timing jitter.

We also discussed three major design considerations for CMOS SPADs. First, we discussed the availability of SPAD junctions when using CMOS technologies. We used cross-sectional views of SPADs based on the standard TSMC 180 nm CMOS process to show different p-n junctions that can be used. Second, we discussed different types of guard ring structures for CMOS SPAD designs. These guard ring structures contain shallow trench isolation, p-well or n-well diffusion guard rings, virtual guard rings and field poly gate guard rings. Third, we described three types of the quench and reset circuits: passive quench and reset circuits, active quench and reset circuits and time-gated circuits. We provided a real design example of a time-gated (TG) pixel design using a standard 65 nm CMOS process. The operation principle was explained using the schematic of the TG pixel and its post-layout simulation results.

Finally, we performed a detailed review of the SPAD designs in the recent literature. The performance parameters were summarized and compared in a table (Table 2-1). This table can be a very useful reference when designing SPADs using CMOS technologies.

Table 2-1: Summary and comparison of SPAD implemented by CMOS technologies.

Ref. Year	**Tech. (nm)**	**Junction**	**V_{BR} (V)**	**Active area (μm) [a]**	**PDE (%)**	**Median DCR (Hz)**	**DT (ns)**	**After pulse (%)**	**Time resolution (ps)**
[82] 2017	CIS 65 (BSI) [b]	–	–	- Square	21.9 @ 660 nm V_{ex} = 4.4 V	–	8	0.08 @ V_{ex} = 4.4 V DT = 8 ns	95 @ V_{ex} = 4.4 V (700 nm)
[122] 2012	CIS 90	p-well/ DNW	14.9	6.4 Circular	44 @ 690 nm V_{ex} = 3.5 V	~100 @ low excess voltage	8	0.375 @ V_{ex} = 0.36 V DT = 15 ns	82 @ V_{ex} = 0.36 V (470 nm) 53 @ V_{ex} = 0.46 V (470 nm) 51 @ V_{ex} = 0.56 V (470 nm)
[80] 2009	CMOS 130	p-well/ DNW	14.4	8 Circular	28 @ 500 nm V_{ex} = 1.4 V	25 @ V_{ex} = 1.4 V 20℃	–	–	–
[123] 2011	CMOS 150	p-well/ Niso	23.1	10 Circular	31 @ 470 nm V_{ex} = 5 V	~230 @ V_{ex} = 5 V 25℃	30	2.1 @ V_{ex} = 5 V DT = 30 ns	170 @ V_{ex} = 5 V (470 nm)
[123] 2011	CMOS 150	p^+/n-well	16.1	10 Circular	26 @ 470 nm V_{ex} = 3.5 V	~160 @ V_{ex} = 3.5 V 25℃	30	1.3 @ V_{ex} = 5 V DT = 30 ns	60 @ V_{ex} = 3.5 V (470 nm)
[83] 2017	CMOS 150	p^+/n-well	18.0	10 Square	31 @ 450 nm V_{ex} = 5 V	3.9k @ V_{ex} = 3 V [c]	–	0.85 @ V_{ex} = 3 V DT = 150 ns	52 @ V_{ex} = 4 V (468 nm) 42 @ V_{ex} = 4 V (831 nm)
[81] 2018	CMOS 180	p^+/shallow n-well	16.8	12.08 Square	55 @ 480 nm V_{ex} = 4 V	28 @ V_{ex} = 1 V 217 @ V_{ex} = 4 V	–	–	260 @ V_{ex} = 4 V (640 nm)
[124] 2010	CMOS HV 180	p^+/DNW	20.3	8 Octagon	20 @ 470 nm V_{ex} = 3.5 V	180 @ V_{ex} = 3.5 V	6	0 @ DT = 6 ns	80 @ V_{ex} = 3.5 V (470 nm)

Table 2-1. *Cont.*

Ref., Year	Tech. (nm)	Junction	V_{BR} (V)	Active area (µm)	PDE (%)	Median DCR (Hz)	DT (ns)	After pulse (%)	Time resolution (ps)
[125], 2014	CMOS 180	p-well/ DNW	23.5	12	40 @ 440 - 620 nm V_{ex} = 10 V	17 @ V_{ex} = 2 V 1.45k @ V_{ex} = 10 V	300	0.3 @ V_{ex} = 10 V DT = 300 ns	70 @ V_{ex} = 10 V (405 nm) 86 @ V_{ex} = 10 V (637 nm)
[126], 2015	CMOS 180	p^+/n-well/ DNW	14.64	12 Circular	40 @ 440 - 580 nm V_{ex} = 4 V	31 @ V_{ex} = 1 V 1.8k @ V_{ex} = 4 V 25℃	300	0.2 @ V_{ex} = 4 V DT = 300 ns	95 @ V_{ex} = 4 V (405 nm) 141 @ V_{ex} = 4 V (637 nm)
[127], 2016	CMOS 180	p-well-epi-BN (Shallow PW)	36.5	–	27.8 @ 490 nm V_{ex} = 4 V	2.195k @ V_{ex} = 4 V 25℃ [d]	–	0.34 @ V_{ex} = 4 V DT = 300 ns	427 @ V_{ex} = 2 V (637 nm) 223 @ V_{ex} = 4 V (637 nm) 243 @ V_{ex} = 2 V (405 nm) 141 @ V_{ex} = 4 V (405 nm)
[127], 2016	CMOS 180	p-well-epi-BN	25.46	–	>40 @ 460 - 600 nm V_{ex} = 11 V	40 @ V_{ex} = 4 V 25℃ [d]	–	7 @ V_{ex} = 11 V DT = 300 ns	139.5 @ V_{ex} = 3 V (637 nm) 100.8 @ V_{ex} = 11 V (637 nm) 133 @ V_{ex} = 3 V (405 nm) 97.2 @ V_{ex} = 11 V (405 nm)
[128], 2013	CMOS 350	–	25	20	28 @ 570 nm V_{ex} = 6 V [e]	25 @ V_{ex} = 6 V Room Temp	20	1.3 @ V_{ex} = 6 V DT = 20 ns	–
[129], 2009	CMOS 350	p^+/DNW	27.5 - 28.4	–	–	4k @ V_{ex} = 4 V Room Temp	–	4.5 @ V_{ex} = 5 V DT = 500 ns	–

[a] The active area stands for the diameter for circular shapes or the length for square and octagonal shapes.

[b] The SPAD is implemented using a 3D structure. The SPAD is using back-side illumination CMOS imaging 65 nm technology and the circuits are implemented in standard CMOS 40 nm technology.

[c] Calculation of total DCR based on the value per µm.

[d] The values are read from the curves.

[e] The highest PDE is 55%, but the measurement setup is not revealed in the publication.

Chapter 3

Random Telegraph Signal in n^+/p-well CMOS SPADs

Single-photon avalanche diodes (SPADs) fabricated in standard complementary metal-oxide-semiconductor (CMOS) processes become favorable photodetectors in many applications such as biomedical imaging, light detection and ranging and communication. These CMOS SPADs have low fabrication costs and the potential to be integrated with other signal conditioning and processing circuits to form an integrated detection system without any extra mask layers or post-processing after fabrication. However, SPADs implemented in standard CMOS process have low photon detection efficiency and high dark noise because of the limitations of the fabrication process. Therefore, noise performance optimization is very important to achieve a good signal-to-noise ratio in SPADs' applications. In this chapter, in addition to two commonly known noise parameters, dark count rate (DCR) and afterpulsing (AP), we explore another interesting phenomenon, random telegraph signal (RTS) noise during the transitional phase of the avalanching process. We present the properties of the RTS noise and their dependence on the biasing voltage and temperature. An analytical model is used to extract the effective dimension of the defects in the depletion region from the variation of the RTS noise current amplitude with biasing voltage. By comparing the DCR and AP of SPAD samples with different effective defect dimensions derived from the RTS noise properties, a trend that the SPAD with a larger defect effective dimension shows a higher DCR and AP is found.

Adapted from W. Jiang and M. J. Deen, "Random Telegraph Signal in n+/p-Well CMOS Single-Photon Avalanche Diodes," *IEEE Trans. Electron Devices*, pp. 1–6, 2021, doi: 10.1109/TED.2021.3070557. (Appendix A)

3.1. Introduction

SPADs are capable of detecting even a single photon, thus showing promise in a wide range of applications such as Raman spectroscopy [118], fluorescence lifetime imaging microscopy (FLIM) [130], positron emission tomography (PET) [131], diffuse optical imaging (DOI) [132], light detection and ranging (LiDAR) [133] and optical wireless communication (OWC) [134]. Further, the advances and progress in silicon fabrication technologies have stimulated much research on SPADs designed and manufactured in a standard complementary metal-oxide-semiconductor (CMOS) process. These SPADs have low fabrication costs and the potential to be more easily integrated with other signal conditioning and processing circuits such as trans-impedance amplifiers, discriminators, counters and time-to-digital converters to form a complete detection system on the same chip. Unlike the CMOS imaging sensor, high-voltage or another customized process, SPADs fabricated in a standard CMOS process usually have a lower photon detection efficiency mainly because of reflections from the passivation layer on the top of the chip and the thin depletion region of the SPAD [131]. Therefore, the noise performance of the SPAD is very important for a good signal-to-noise ratio in photon detection.

When a SPAD is biased above its breakdown voltage, an avalanche pulse will occur if there is a carrier generated in the depletion region. However, the generated carrier is not limited to photo-generated electron-hole pairs; it can also result from the dark noise. Therefore, even when a SPAD is kept in the dark, avalanche pulses can still be generated due to dark carriers. The count rate of such avalanche pulses is called dark count rate (DCR). Several mechanisms of carrier generation such as Shockley-Read-Hall (SRH) thermal generation-recombination, trap-assisted thermal generation-recombination , band-to-band tunneling (BTBT) and trap-assisted tunneling (TAT) [135], [136], [137], contribute to dark noise. Another important source of the dark noise is afterpulsing (AP). For SPADs fabricated in a standard CMOS process, compared to a customized process, there is a higher probability to have defects in the depletion region. These defects can function as trap centers, which can cause dark counts generated from the trap-assisted thermal generation-recombination and TAT. These trap centers may also capture some carriers during the

primary avalanche process and release them at random time intervals [66], [67], thus leading to secondary avalanches. The probability of secondary avalanches triggered by carriers emitted from the trap centers is defined as AP. Therefore, a higher concentration of defects in the depletion region of a SPAD results in a higher DCR and AP.

Recently, another phenomenon called Random Telegraph Signal (RTS) was studied in SPADs. RTS behavior was observed in sub-micrometer sized MOSFETs [138], polysilicon emitter bipolar transistors [139], [140], memory cells [141], charge-coupled devices [142] and active-pixel imaging sensors [143]. The RTS noise was shown to have a negative impact on the reliability and performance of these devices [141], [142], [143]. Recently, RTS was also observed in SPADs when they operated in Geiger mode [144], [145], [83] or are in the transitional phase of breakdown [146], [147]. In the early days, this noise was also called bistable noise and categorized into two types: burst noise and microplasma noise [148], [149]. It is believed that the RTS noise is caused by defects in the depletion region of SPADs [150].

In Geiger mode, the RTS noise can result in fluctuations of the DCR [145], [83]. That is, the mean DCR of the SPAD often switches between two different discrete values. Such an effect can be enhanced by the proton irradiation which introduces defects in silicon [151]. The mean DCR of a SPAD irradiated by protons can even switch between more than two levels [152], [153]. In the transitional phase of avalanche when the SPAD is biased around its breakdown voltage, the SPAD switches between “on” and “off” states, which also shows the RTS behavior [146], [147]. In this case, the RTS noise amplitude increases with the excess voltage (V_{ex}). The average time spent in the “on” state (T_{ON}) also increases with V_{ex} while the average time spent in the “off” state (T_{OFF}) decreases with V_{ex}. This is because the carriers in the depletion can more easily trigger avalanches due to their higher energies when V_{ex} increases. Therefore, the SPAD is more likely to stay in the “on” state rather than the “off” state. As described in Chapter 1, V_{ex} is defined as the voltage difference between the biasing voltage and the breakdown voltage which is known to depend on temperature. Therefore, temperature is also expected to play an important role on the

properties of RTS noise of the SPAD in the transitional phase of the avalanche, which is not discussed in [147].

In this work, we conducted RTS noise measurements at different temperatures during the transitional period of the avalanche on a n^+/p-well SPAD fabricated in a standard 130 nm CMOS process. An analytic model was used to derive the effective dimensions of the defects based on the properties of the RTS noise. Additionally, because the defects in the depletion region have a significant impact on the performance of the DCR and AP of a SPAD, the RTS noise which originates from the defects is expected to show some correlations with these two noise parameters. We measured the RTS noise of 3 SPAD samples and then extracted the effective defect dimensions based on their properties. We also measured and compared the DCR and AP of these SPAD samples, and investigated their correlations with the RTS noise.

3.2. SPAD Structures and the Breakdown Voltage

3.2.1. SPAD Structure

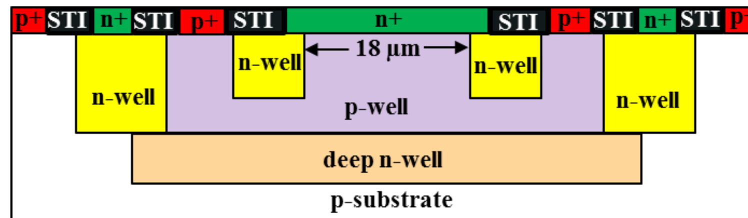


Figure 3-1: Cross-sectional view of n^+/p-well SPAD. The deep n-well and p-substrate are connected to the ground.

Our SPAD structure is an n^+/p-well junction fabricated in a standard 130 nm CMOS process, and has a similar structure as shown in [154], [75]. The cross-sectional view of the SPAD is shown in Figure 3-1. The active area of the square-shaped SPAD is a (18 μm)^2. An n-well ring surrounding the active area functions as a guard ring for the n^+/p-well junction to reduce the electric field at the edge of the junction, thus preventing the premature edge breakdown. The outer n-well guard ring is electrically connected to the grounded p-substrate by connecting the n^+ ohmic contact of the outer n-well to the p^+ ohmic

contact of the p-substrate. The deep n-well (DNW) under the p-well contacting with the outer n-well ring forms an outer guard ring to isolate the active region of the SPAD from the p-substrate. When a negative high voltage is applied on the anode (p-well) of the SPAD, the depleted DNW will prevent the injection of the minority carriers to the active region of the SPAD to reduce the dark noise. It is worth noting that the n^+ layer is extended by 0.75 μm into n-well to separate the shallow trench isolation (STI) from the active region. This is because many defects can occur at the interface of SiO_2-Si during the formation of the STI structure. These defects can function as generation-recombination centers of carriers. Therefore, a high DCR and AP can be expected if the STI has a direct contact to the active region of the SPAD.

3.2.2. Breakdown Voltage

To measure the RTS noise and study its dependence on temperature, the breakdown voltage V_{BR} of the SPAD is needed to apply the proper biasing voltage. The V_{BR} is obtained from its current-voltage (I-V) characteristics. An Agilent B1500A Semiconductor Device Analyzer was used to measure the I-V characteristics of the SPAD. As shown in Figure 3-1, besides the n^+/p-well junction, there are two other parasitic diodes – p-well/DNW junction and DNW/p-substrate junction. The reverse biasing voltages for these two parasitic diodes must be kept below their breakdown voltages to ensure the normal operation of the n^+/p-well SPAD. Since the DNW is grounded by design, the DNW/p-substrate junction will not influence measurements of the n^+/p-well SPAD. Another parasitic diode (p-well/DNW) was found to have a breakdown voltage around 9.5 V at room temperature. Therefore, the voltage applied on the p-well anode was set to a fixed negative voltage -6.5 V, which kept the voltage across the parasitic p-well/DNW diode below its breakdown voltage.

To measure the breakdown voltage of the n^+/p-well junction, the voltage applied on the n^+ cathode of the SPAD was swept from 4.2 V to 5.1 V with a step of 1 mV. The temperature coefficient of the breakdown voltage was also measured by placing the SPAD in a thermal chamber to adjust the operational temperature. Figure 3-2(a) shows the I-V

characteristics of the SPAD in dark and at different temperatures. The results show that the n^+/p-well SPAD has a breakdown voltage of 11.467 V at 30 °C, 11.370 V at 15 °C, 11.272 V at 0 °C, 11.121 V at -15 °C and 11.018 V at -30 °C, which gives a temperature coefficient of 7.6 mV/°C, as shown in Figure 3-2(b). These results match our previous measurements [73], [106], [154].

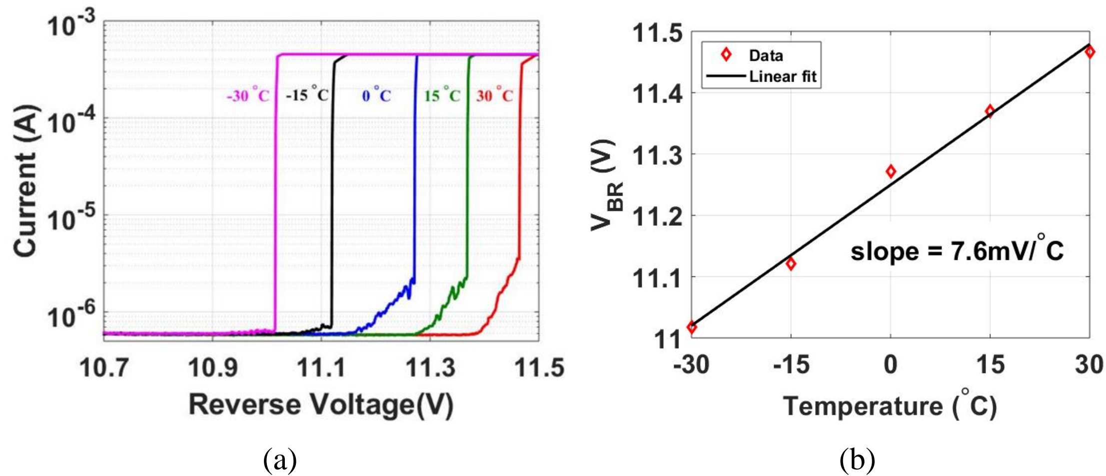


Figure 3-2: (a) I-V characteristic of SPAD at different temperatures. (b) Breakdown voltage changes with temperature.

3.3. RTS Noise Measurement

3.3.1. Measurement Details

Figure 3-3(a) illustrates the measurement setup for the RTS noise. An adjustable DC voltage source (V_C) was connected to the cathode and a 50 Ω resistor to the anode as a resistive load. Another DC voltage source providing a negative voltage was connected in series with the load resistor. The output signal from the anode has a small amplitude, so it is amplified with a low noise amplifier (Parc 113) through an AC-coupled interface. A high-speed digital oscilloscope (LeCroy WaveRunner 625Zi) acquires and processes the amplified signals of the RTS noise. Figure 3-3(b) shows a typical waveform of the RTS noise, in which the current flowing the SPAD junction is not constant and switches between the “on” and “off” states. As defined in [146], [147], the RTS noise phenomena in the SPAD can be quantified by three parameters:

- T_{ON}, which is the time spent in the "on" state of the avalanche.
- T_{OFF}, which is the time spent in the "off" state of the avalanche.
- *I*, which denotes the amplitude of the RTS noise current and can be defined as the current difference between the averaged current in the "on" state and the averaged current in the "off" state of the avalanche.

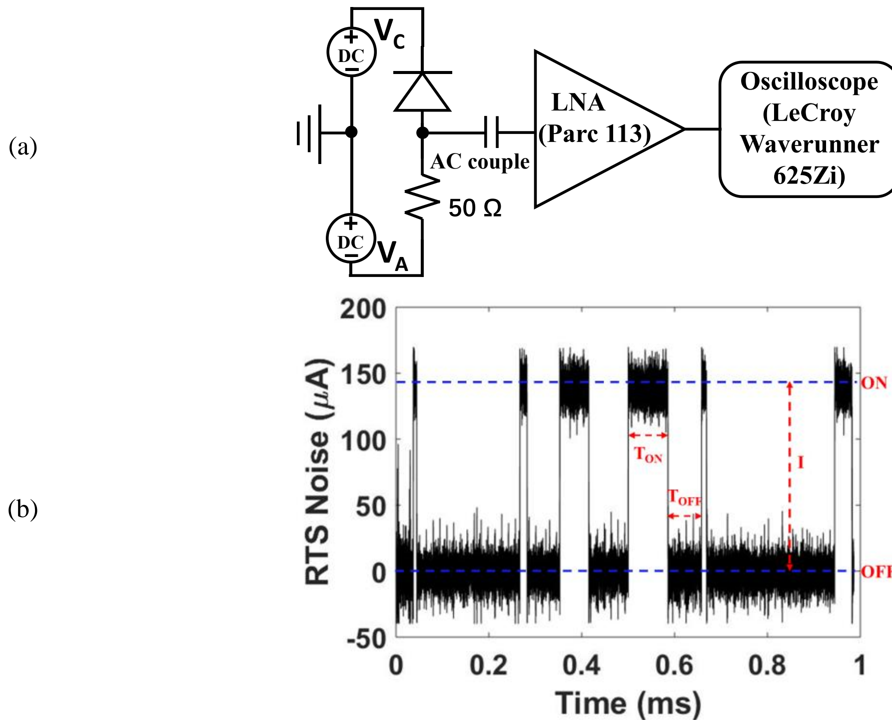


Figure 3-3: (a) RTS noise measurement setup. The output of SPAD is connected to a low-noise amplifier through AC coupling. (b) Typical waveform of RTS noise. Three parameters- T_{ON} (time spent in the RTS on-state), T_{OFF} (time spent in the RTS off-state) and I (amplitude of RTS noise current) are used to quantify the RTS noise.

For RTS noise in a SPAD, the time for the avalanche in the "on" or "off" states are random. Therefore, statistical properties of the three parameters are usually used to analyze the distribution of RTS noise. An example of statistical histograms of the RTS on-time (T_{ON}), off-time (T_{OFF}) and noise amplitude (*I*) when the SPAD is biased at 11.42 V at room temperature is shown in Figure 3-4. Exponential fits of the histograms (0.1 μs bin size) of T_{ON} and T_{OFF} are shown in Figures 3-4(a) and (b), respectively. The avalanche "on" and "off" states are independent (i.e. memory-less) random switching events. Therefore the switching of RTS noise is in accordance with a Poisson distribution [155]. As a result, the

T_{ON} and T_{OFF} are exponentially distributed, which is similar to the results in [145], [146], [152], [156]. The probability $P(t)$ of the SPAD in the "on" or "off" state is given by [145]

$$P(t) \cong \frac{1}{\tau}\exp(-t/\tau), \tag{3-1}$$

where t and τ represent T_{ON} and τ_{ON} in Figure 3-4(a), T_{OFF} and τ_{OFF} in Figure 3-4(b), respectively.

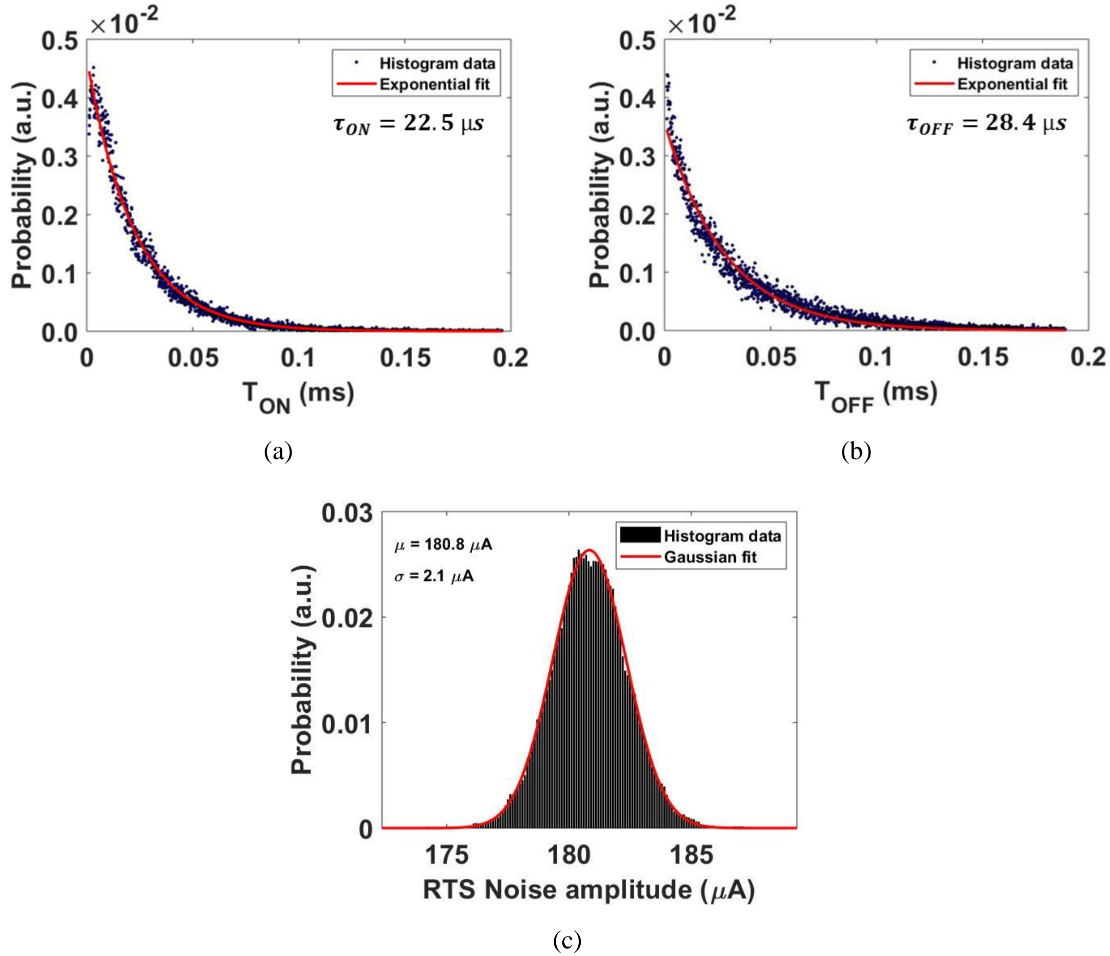


Figure 3-4: (a) Histogram of T_{ON} with an exponential fit. (b) Histogram of T_{OFF} with an exponential fit. (c) Histogram of RTS noise amplitude with a Gaussian fit. The measurement was conducted with the V_{BIAS} of 11.42 V at 30 °C.

The time constants for the probability distributions of T_{ON} and T_{OFF} were calculated to be 22.5 μs for τ_{ON} and 28.4 μs for τ_{OFF}, respectively. The histogram of the RTS noise amplitude with a bin size of 0.1 μA is shown in Figure 3-4(c), which follows a normal

distribution. The Gaussian fit gives a mean value of the RTS noise amplitude of 180.8 μA with a standard deviation of 2.1 μA.

3.3.2. RTS Noise Versus Biasing Voltage

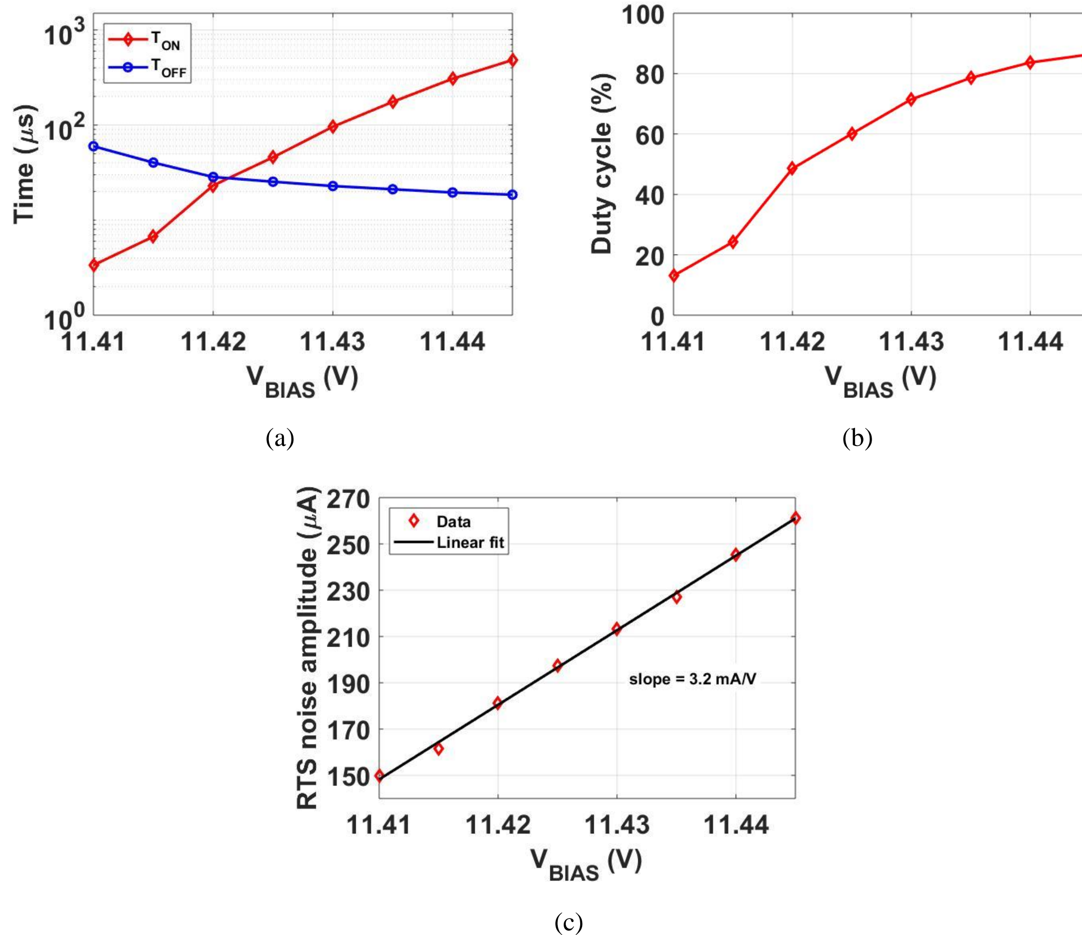


Figure 3-5: (a) T_{ON} and T_{OFF}, (b) Duty cycle and (c) RTS noise amplitude as a function of reverse biasing voltage at 30 °C. T_{ON} and T_{OFF} are on a semi-log scale.

The biasing voltage of the SPAD (V_{BIAS}) influences its avalanche process and performance. Since RTS noise was observed at the transitional period of the avalanche, its performance should also depend on the biasing voltage. When V_{BIAS} increases, carriers in the depletion can more easily trigger avalanches due to their higher energies. Therefore, the SPAD is more likely to stay in the “on” state rather than the “off” state. The biasing voltage dependence of the mean values T_{ON} and T_{OFF} is shown in Figure 3-5(a). As expected, when

V_{BIAS} increases, the mean value of T_{ON} increases while the mean value T_{OFF} decreases. The duty cycle of the RTS noise is used to indicate the fraction of the time when SPAD is in the "on" state, which can be calculated with $T_{ON}/(T_{ON} + T_{OFF})$. The duty cycle also increases with V_{BIAS}, as shown in Figure 3-5(b). Figure 3-5(c) shows the RTS noise amplitude as a function of V_{BIAS}. The measurement results show that the RTS noise amplitude increases with V_{BIAS} almost linearly. This means that the avalanching resistance is almost constant with our measurement conditions. The reason is that the width of the depletion region is almost the same due to the small difference (35 mV) of the V_{BIAS} (11.41 – 11.445 V) during the measurement. A similar result of the relationship between the RTS noise amplitude and biasing voltage was shown in [147].

3.3.3. RTS Noise Versus Temperature

Figures 3-6(a) and (b) show the RTS noise dependence on temperature. As shown in Figure 3-2(a), the RTS noise amplitude depends on the breakdown voltage, which shows a positive temperature coefficient of 7.6 mV/°C. To eliminate the influences resulting from the changes of breakdown voltage at different temperatures on the RTS noise measurement, the biasing voltage V_{BIAS} was set for a 50% duty cycle of the RTS noise at different temperatures. Figure 3-6(c) shows that the V_{BIAS} to achieve 50% duty cycle is a linear function of temperature. The slope of the linear fit is 7.55 mV/°C, which almost matches the measured breakdown voltage temperature coefficient of 7.6 mV/°C shown in Figure 3-2(b). Figure 3-6(a) shows the RTS noise amplitude as a function of temperature. The results indicate that, after compensating the changes of the breakdown voltage with temperature, the RTS noise amplitude is a quite constant value with a mean of 219 μA and a standard deviation of 3.92 μA in the temperature range of -30 to +30 °C. Therefore, the RTS noise amplitude only depends on the absolute excess voltage ($V_{BIAS} - V_{BR}$) across the SPAD in the measured temperature range (-30 to +30 °C). Since the V_{BIAS} was set to achieve a 50% duty cycle, the mean values of T_{ON} and T_{OFF} are the same. Their dependence on temperature was shown in Figure 3-6(b), in which the same value of T_{ON} and T_{OFF} is denoted as the lifetime. Generation-recombination (GR) defects in the depletion region of the SPAD function as trap

centers. The "on" and "off" states of RTS noise correspond to the carriers "trapping" and "detrapping" processes of the GR centers. Therefore, the mean value of T_{ON} of the RTS noise corresponds to the carrier lifetime during the avalanche period, which strongly depends on temperature. The exponential fit in Figure 3-6(b) shows that the lifetime of carriers decreases exponentially when the temperature increases. This trend of the lifetime is also found in other RTS publications [145], [152], which showed that the lifetime varied with temperature according to the Arrhenius relationship. That is the reason the lifetime shows an exponential trend with temperature.

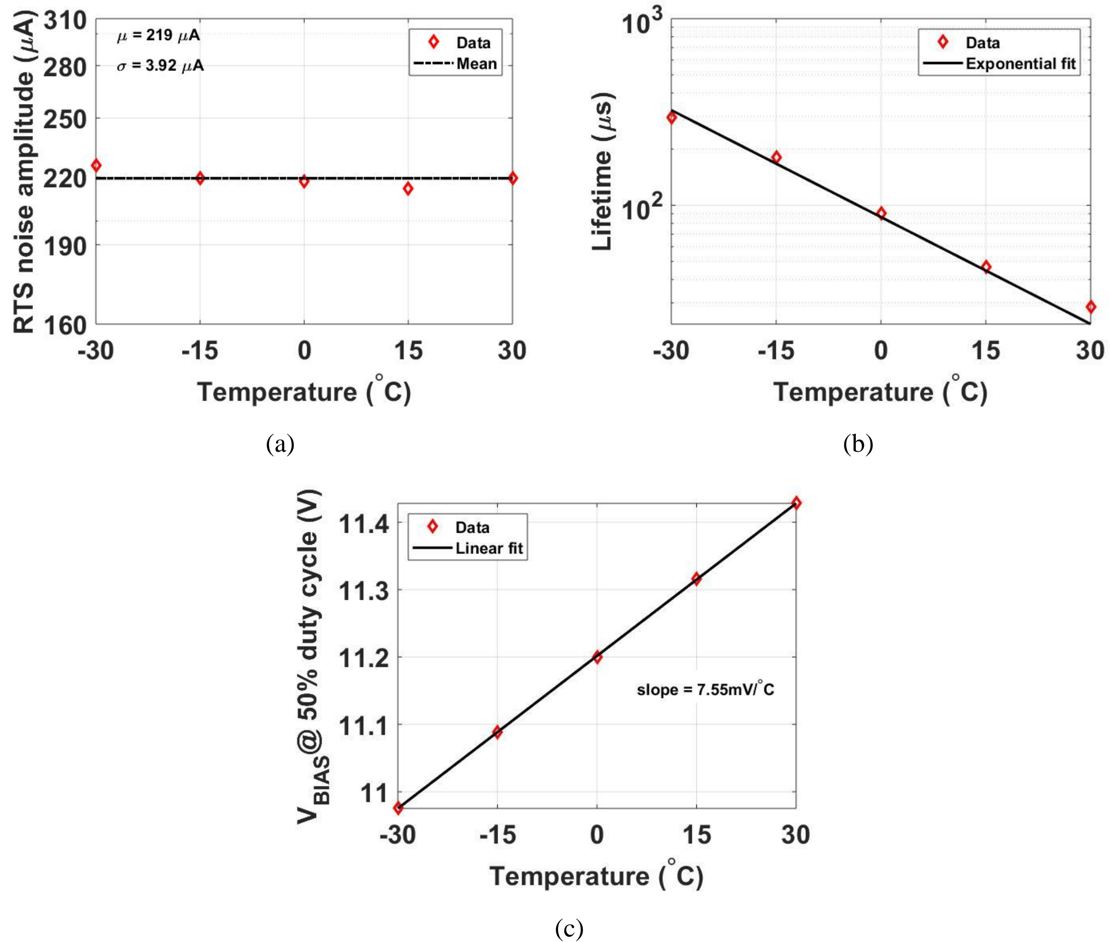


Figure 3-6: (a) RTS noise amplitude as a function of temperature. (b) T_{ON} or T_{OFF} as a function of temperature with an exponential fit when the duty cycle is 50%, (c) The reverse biasing voltage to achieve 50% duty cycle at different temperatures with a linear fit.

3.4. DCR And AP Measurement

3.4.1. Measurement Details

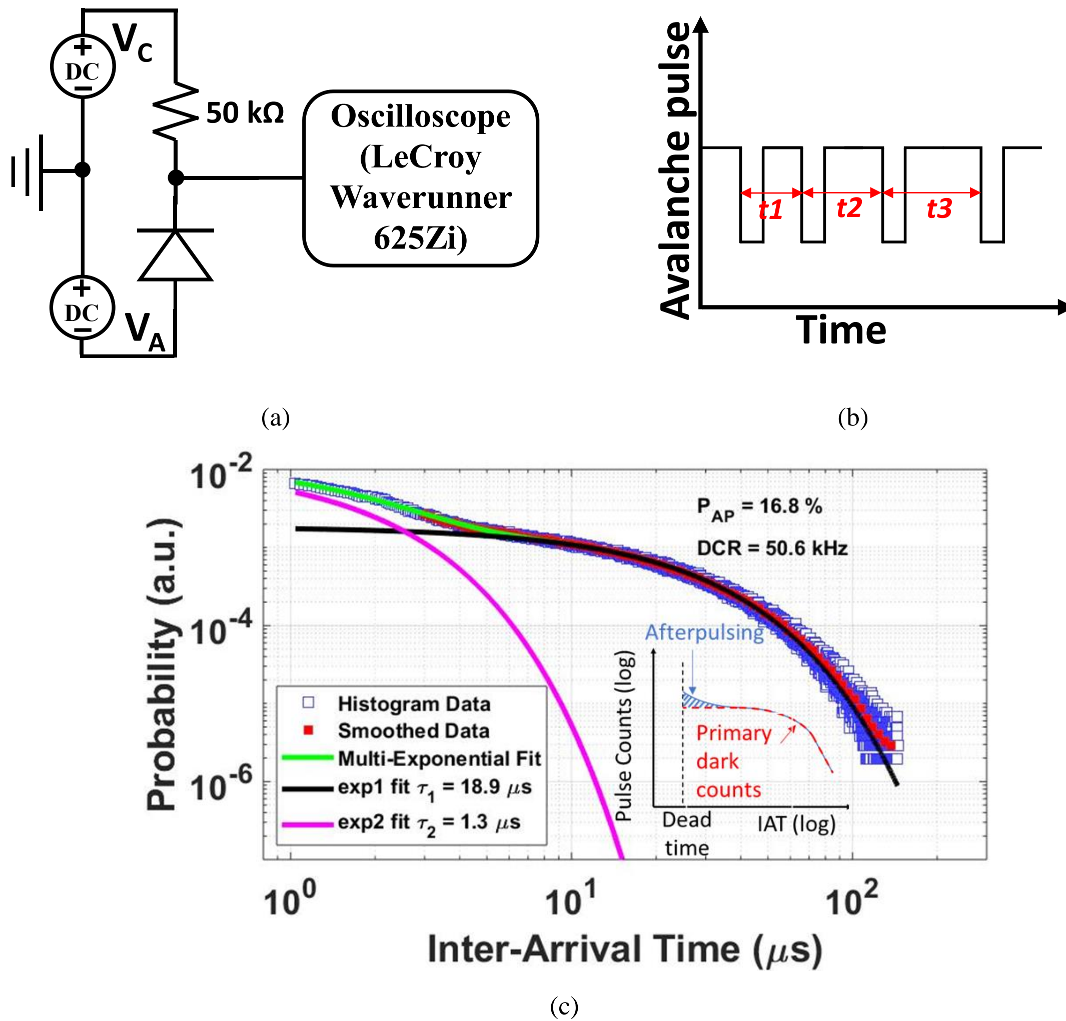


Figure 3-7: (a) Dark count rate (DCR) and afterpulsing (AP) measurement setup. The SPAD is passively quenched through a 50 kΩ resistor. (b) Illustration of inter-arrival time of the avalanche pulses. Avalanche pulse is a negative pulse in the illustration. (c) Histogram of inter-arrival time with multi-exponential fit. The measurement was conducted with the V_{ex} of 1.2 V at -30 °C.

Figure 3-7(a) shows the setup for the DCR and AP measurements. A 50 kΩ passive quench resistor was connected the cathode of the SPAD. Two DC voltage sources provided biasing voltages to the SPAD. The LeCroy oscilloscope connected to the SPAD cathode output was used to record the time intervals between two consecutive avalanche pulses as

illustrated in Figure 3-7(b). This inter-arrival-time (IAT) measurement has been used to characterize the DCR and AP in many SPADs [154], [157], [158]. A detailed description of the IAT method can be found in [84]. The primary DCR and AP can be derived from the histogram of the IAT data by applying a multi-exponential fit. An example of the DCR and AP measurements is presented in Figure 3-7(c). In order to show a clear example of AP, the measurement was conducted at a low temperature of -30 °C when the SPAD is biased at a high excess voltage of 1.2 V. Under this condition, the multi-exponential fit of the histogram of the IAT gives a DCR of 50.6 kHz and an AP of 16.8%.

3.4.2. DCR and AP Versus Biasing Voltage

As the excess voltage increases, the electric field across the depletion region of SPAD will also increase. Therefore, the carriers in the depletion region will have a higher avalanche probability due to their increased energies, thus leading to a higher DCR and AP. Figure 3-8 shows the DCR and AP as a function of the excess voltage. The measurements were conducted at a low temperature (-30 °C) specifically to present a high value for AP. With the consideration of the temperature coefficient of the breakdown voltage, the excess voltages were adjusted from 0.4 V to 1.2 V in steps of 0.2 V.

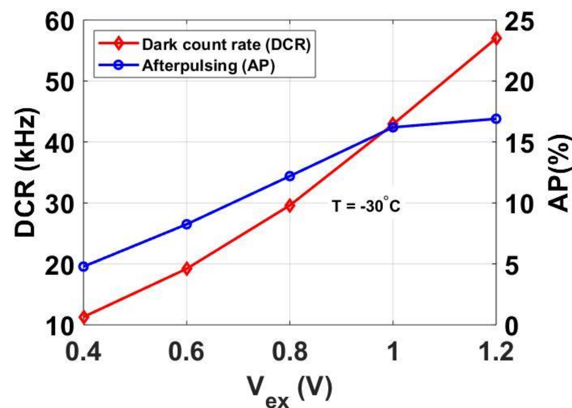


Figure 3-8: Dark count rate (DCR) and afterpulsing (AP) as a function of excess voltage at -30 °C.

3.4.3. DCR and AP Versus Temperature

Figure 3-9 shows the temperature dependence of the DCR and AP when the SPAD was biased at a fixed excess voltage at 0.8 V. Similarly, the temperature dependence of the breakdown voltage was considered to keep a fixed excess bias at different temperatures. The results show that DCR increases with temperature while AP decreases with temperature. The SRH theory indicates that the rate of thermal generation-recombination increases exponentially with temperature while the tunneling noise shows a relative independence on temperature [75], [159], [160]. Additionally, the lifetime of the traps decreases exponentially with temperature. The exponential fit of the DCR indicates that SRH thermal generation noise dominates in the measured -30 to 10 °C temperature range. It is worth noting that the SPAD with an excess voltage of 0.8 V did not show any AP when the temperature is above 0 °C. This implies that, under this condition, the lifetime of the traps is short enough so that almost all the carriers captured by the trap centers are released within the dead time of the SPAD and do not trigger a secondary avalanche event.

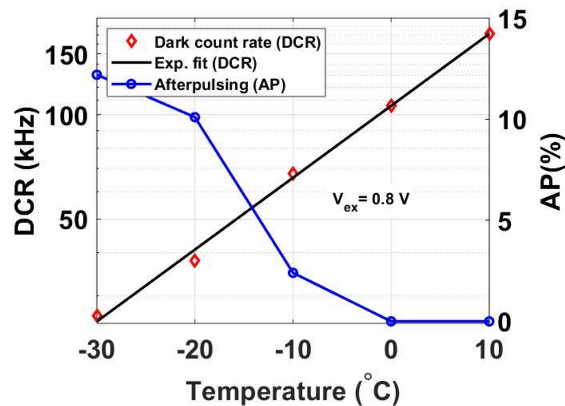


Figure 3-9: Dark count rate (DCR) and afterpulsing (AP) as a function of temperature at a fixed excess voltage of 0.8 V. The temperature coefficient of the breakdown voltage of SPAD was considered when adjusting the reverse bias voltage to keep V_{ex} fixed.

3.5. RTS Noise Correlation with DCR and AP

3.5.1. Analytical Model for the Dimension of Defects

A cylindrical defect model with a diameter "*d*" and a width "*w*" is located in the depletion region in the n^+/p-well junction [147], as illustrated in Figure 3-10(a). The equivalent electrical circuit model is shown in Figure 3-10(b), where C_{SPAD} is the junction capacitance, R_s is the equivalent resistor when the SPAD is avalanching, V_{BR} is used to represent the breakdown voltage, and a switch is used to represent the "on" and "off" states of the avalanche. When an avalanche is triggered at the initial time (denoted as $t = 0$), the switch is closed. The current $I(t)$ will flow from the cathode (n^+ terminal) to the anode (p-well terminal) through the equivalent resistor R_s. The peak current is the initial current when the switch closes, which is given by

$$I_{\text{peak}} = I(t = 0) = \frac{V_{\text{BIAS}} - V_{BR}}{R_s}. \tag{3-2}$$

where V_{BIAS} is the reverse biasing voltage. The RTS noise shows a constant amplitude in our measurements, which means that the initial avalanching current I_{peak} can be substituted by the RTS noise amplitude current I in Eq. 3-2. To simplify the analysis, the breakdown voltage is considered constant if there is no change in the operating temperature. Then, the equivalent resistor R_s can be derived from Eq. 3-2, as,

$$\frac{1}{R_s} = \frac{dI_{\text{peak}}}{d(V_{\text{BIAS}} - V_{BR})} = \frac{dI}{dV_{\text{BIAS}}}. \tag{3-3}$$

Therefore, the equivalent resistor can be derived from the curve of RTS noise amplitude versus biasing voltage.

In a real SPAD, many localized defects rather than a single defect could exist in its depletion region. The measured RTS noise current is a sum of all currents originating from the "trapping" and "detrapping" processes of all defects. Therefore, the measured RTS noise amplitude can be expressed as [147]

$$I = \sum_{i=1}^{n} \frac{V_{\text{BIAS}} - V_{BR,i}}{R_{s,i}}, \tag{3-4}$$

where $R_{s,i}$ is the equivalent avalanche resistor of the i-th defect and $V_{BR,i}$ is the breakdown voltage of the i-th defect.

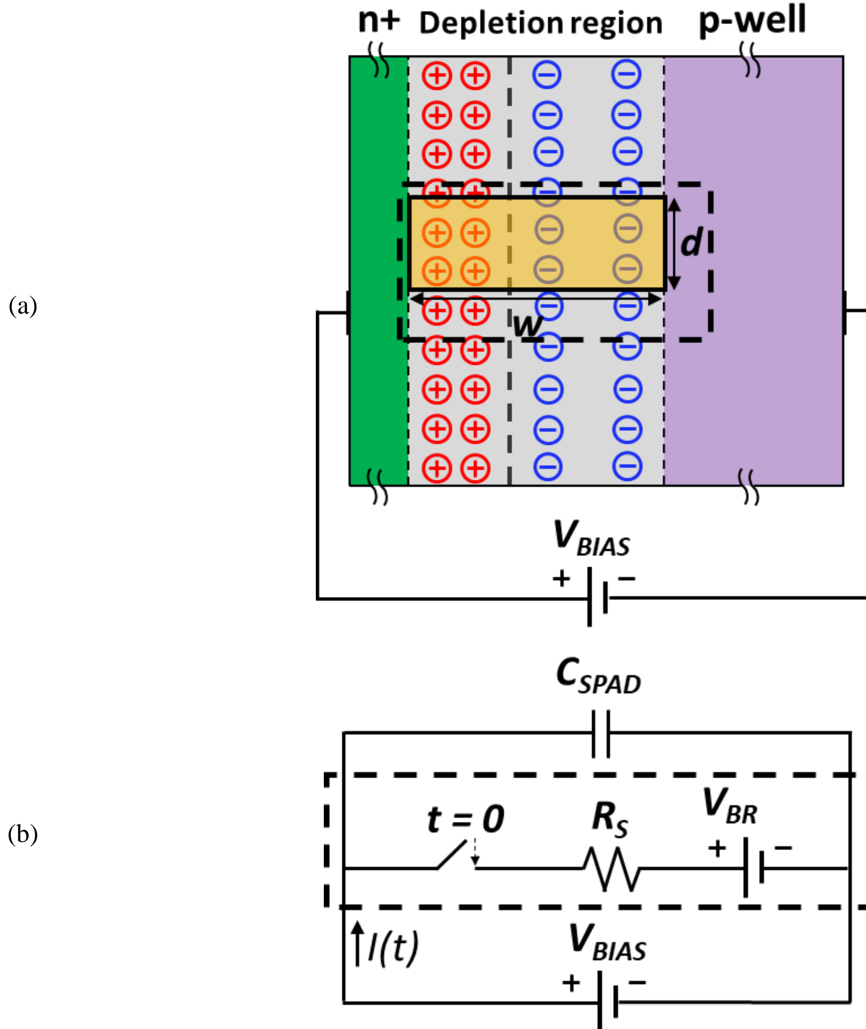


Figure 3-10: (a) Illustration of a cylindrical defect in the depletion region of n+/p-well junction. (b) Equivalent circuit model of the defect with a SPAD capacitance C_{SPAD}, a serial space charge region resistor R_S, a breakdown voltage V_{BR} and a switch which indicates the “on” and “off” of the avalanche.

By substituting Eq. 3-4 into Eq. 3-3, the equivalent R_s can be rewritten as

$$\frac{1}{R_s} = \sum_{i=1}^{n} \frac{1}{R_{s,i}}, \tag{3-5}$$

which shows that equivalent resistor R_s can be considered as equivalent resistors from all defects connected in parallel.

The equivalent resistor $R_{s,i}$ for the i-th defect depends on the dimension of the i-th defect and the physical properties of the SPAD. In the abrupt n$^+$/p-well junction, the $R_{s,i}$ can be expressed as [160]

$$\frac{1}{R_{s,i}} = \frac{\pi \varepsilon_s v_{TH} {d_i}^2}{2(w - x_A)^2}, \tag{3-6}$$

where ε_s (= 1.04 × 10^{-10} F/m) is the permittivity of the silicon, d_i is the diameter of the i-th defect, w is the width of the depletion region of the SPAD, x_A is the width of avalanche region, which is around 0.3w for an abrupt single-sided junction [149], and v_{TH} is the thermal velocity of electrons and holes.

Substituting x_A with 0.3w and rearranging Eq. 3-6, the diameter of the i-th defect d_i can be described by

$${d_i}^2 = \frac{2(w - x_A)^2}{\pi \varepsilon_s v_{th} R_{s,i}} = \frac{0.98w^2}{\pi \varepsilon_s v_{TH}} \times \frac{1}{R_{s,i}}. \tag{3-7}$$

Therefore, the effective value of the diameter d_t representing all defects in the SPAD can be expressed as

$$d_t = \sqrt{\sum_{i=1}^{n} {d_i}^2} = \sqrt{\frac{0.98w^2}{\pi \varepsilon_s v_{TH}} \times \frac{dI}{dV_{BIAS}}}, \tag{3-8}$$

from which the effective size (d_t) for the sum of defects can be obtained from the slope of the characteristics of RTS noise amplitude versus biasing voltage shown in Figure 3-5 (c).

In Eq. 3-8, the carrier thermal velocity v_{TH} is given by Eq. 3-9,

$$v_{TH} = \sqrt{\frac{3kT}{m^*}}, \tag{3-9}$$

in which k (= 1.38 × 10^{-23} J/K) is the Boltzmann's constant, T is the absolute temperature in Kelvin and m^* is the effective mass of electrons and holes. To simplify the analysis,

the effective mass of electrons and holes are regarded to have the same value, which is $1.08m_0$. m_0 is free electron rest mass and is equal to 9.11×10^{-31} kg.

A simple expression for the depletion region width w is

$$w = \sqrt{\frac{2\varepsilon_s(V_{\text{BIAS}} + V_{bi})}{e} \times \frac{N_A+N_D}{N_A N_D}}, \quad (3\text{-}10)$$

where V_{BIAS} is the biasing voltage across the junction, V_{bi} is the built-in potential barrier, e (= 1.6×10^{-19} C) is the charge of an electron, N_A is the doping concentration of the p-well (8×10^{17} cm^{-3}) and N_D is the doping concentration of the n$^+$ layer (2×10^{20} cm^{-3}) in this 130 nm CMOS process [161].

Table 3-1 shows some results of the defect dimensions of SPAD1 by using the model described above. The characteristics of the RTS noise amplitude versus the biasing voltage were measured at different temperatures. The linear fit was applied to the data to get the slope ($dI/\,dV_{BIAS}$), and then the dimension was calculated using Eq. 3-8. The effective defect dimension of the SPAD under measurement presents a mean value of 1.38 µm with a standard deviation of 0.05 µm in the temperature range of -30 to +30 °C.

Table 3-1: The dimension of defects at different temperatures

Temp. (°C)	**Slope (dI/dV_{BIAS}) (µA/V)**	**R_S (Ω)**	**d_t (µm)**
-30	3680.9	271.7	1.45
-15	3468.3	288.3	1.39
0	3432.5	291.3	1.38
15	3468.4	288.3	1.38
30	3222.5	310.3	1.32
Mean	3454.3	290.0	1.38
S.D.	162.7	13.7	0.05

3.5.2. DCR AND AP comparison

As explained in Section I, the performance of DCR and AP of a SPAD fabricated in a CMOS process are significantly influenced by the properties of defects in the depletion region. Considering the fact that RTS noise of a SPAD also originates from the defects in

the depletion region, it is expected that RTS noise may show some correlations with the DCR and AP.

Table 3-2: The effective dimension of defects for 3 SPAD samples

SPAD#	**Slope (dI/dV_{BIAS}) (μA/V)**	**R_S (Ω)**	**d_t (μm)**
SPAD1	3680.9	271.7	1.45
SPAD2	3030.2	330.0	1.31
SPAD3	2764.7	361.7	1.25

The RTS noise of 3 SPAD samples with identical dimensions (18 μm × 18 μm square) were measured to get the curves of RTS noise amplitude with biasing voltage, as shown in Figure 3-5(c). Then, the diameter for the total defects (effective defect diameter) was calculated for each SPAD using Eq. 3-8 in the defect model, which is 1.45 μm for SPAD1, 1.31 μm for SPAD2 and 1.25 μm for SPAD3, as summarized in Table 3-2 (SPAD1 is the SPAD used in the measurement for Figures 3-2 to 3-9). The DCR and AP of 3 SPADs with different effective defect dimensions were measured at different excess voltages from 0.4 to 1.2 V in steps of 0.2 V. To emphasize the values of AP, the measurements were performed at a low temperature of -30 °C. The DCR and AP at different excess voltages for 3 SPAD samples are shown in Figure 3-11. The results indicate that the SPAD with a larger calculated defect dimension presents a higher DCR and AP. It can be seen from Figure 3-11(a), the differences of the DCR between different SPAD samples are increasing with the excess voltage. For example, the DCR difference between SPAD1 and SPAD2 is 0.3 kHz at an excess voltage of 0.4 V while the difference increases to 9.5 kHz at an excess voltage of 1.2 V. The difference of the AP between SPADs with different defect dimensions shows a similar trend, which is shown in Figure 3-11(b). It is worth noting that the ratios of the DCRs between different SPADs and the ratios of APs between different SPADs do not exactly follow the ratios of the calculated defect dimensions of different SPADs. This may be because of the differences of the defect distributions in the depletion region as well as small differences in the defect energy levels in silicon. Overall, higher DCR and AP are expected in the SPAD with a larger effective defect dimension. As stated in the previous section regarding the analytical model for the SPAD, the defect dimension can be extracted from the

RTS noise properties. Therefore, RTS noise of a SPAD shows some correlations with its DCR and AP. A large slope in the curve of RTS noise amplitude with biasing voltage indicates a larger dimension of total defects, thus leading to a higher DCR and AP.

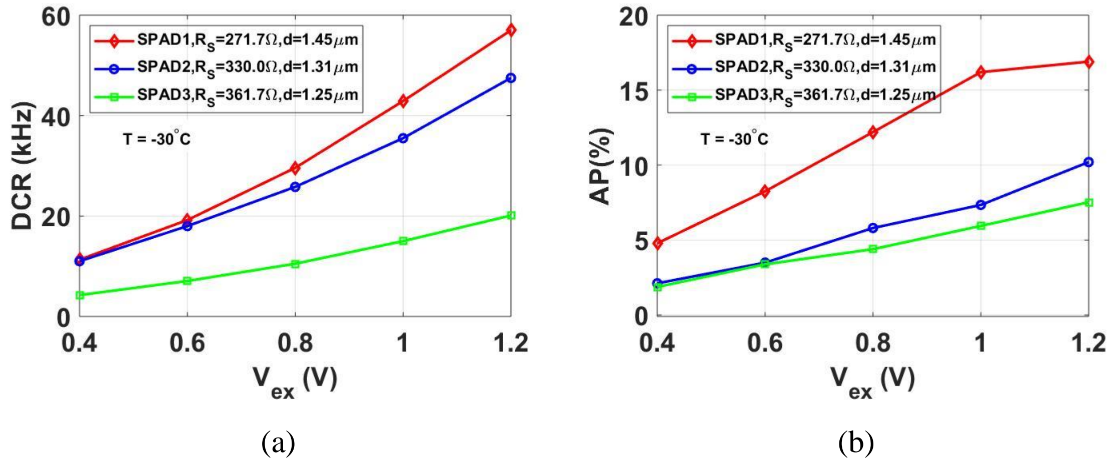


Figure 3-11: (a) Dark count rate (DCR) as a function of excess voltage at -30°C for 3 SPADs with different calculated size of all traps. (b) Afterpulsing (AP) as a function of excess voltage at -30 °C for 3 SPADs with different calculated trap size. The temperature coefficient of the breakdown voltage of SPAD was considered when adjusting the excess voltage.

3.6. Conclusions

In this chapter, we investigated the random telegraph signal (RTS) phenomenon during the transitional period of the avalanche of single-photon avalanche diodes (SPADs). The SPADs under measurement are n^+/p-well SPADs fabricated in a standard 130 nm CMOS process. Three parameters are used to quantify the RTS noise: T_{ON}, the time spent in the "on" state of the avalanche; T_{OFF}, which is the time spent in the "off" state of the avalanche; and I, the amplitude of the RTS noise current. The experimental results of biasing voltage (V_{BIAS}) dependence of RTS noise demonstrate that the mean value of T_{ON} increases with V_{BIAS} while the mean value T_{OFF} decreases with V_{BIAS}. The amplitude of RTS noise increases almost linearly with V_{BIAS}. The biasing voltage dependence of the amplitude and time constants are consistent with the measurement results of p^+/n-well SPADs reported in [147]. The temperature dependent measurements show that the RTS noise amplitude does not change with temperature and is almost constant as long as the absolute excess voltage

across the SPAD is kept the same. The mean values of T_{ON} and T_{OFF} decrease exponentially as the temperature increases.

The RTS noise is believed to originate from the defects which exist in the depletion region of the SPAD. An analytical model was used to describe the defects which cause RTS noise during the transitional period of the avalanche. With this model, the dimension of the total defects can be estimated based on the curve of RTS noise amplitude with biasing voltage. Additionally, the defects in the depletion region can function as trap centers which play a significant role on the performance of the dark count rate (DCR) and afterpulsing (AP). Therefore, the RTS noise of a SPAD which results from the defects in the depletion region is expected to show some correlations with its DCR and AP. Three SPAD samples with different dimensions of defects were used to measure and compare their values of DCR and AP at a low temperature of -30 °C. The experimental results revealed a trend that the SPAD with a larger effective dimension of defects has a higher DCR and AP.

Chapter 4

Improved Noise Performance of CMOS Poly Gate SPAD

In this chapter, the noise performance of three types of n^+/p-well single-photon avalanche diodes (SPADs) fabricated in a standard 180 nm CMOS technology is studied. The SPADs had different poly gate configurations: no poly gate (SPAD_NG), a dummy floating poly gate (SPAD_DG), and a field poly gate connected to the n^+ cathode (SPAD_FG). The measurement results of dark count rate and afterpulsing showed that the SPAD_DG had better noise performance compared to the SPAD_NG. The dummy poly gate pushed the shallow trench isolation away from the active region of the SPAD, thus reducing the dark noise generated from the Si-SiO_2 interface. The measurement results also revealed that the noise performance can be further improved by connecting the poly gate to the n^+ cathode. The voltage on the poly gate in SPAD_FG reduced the electric field in the n-well guard ring (GR) region, thus reducing the carriers from the GR region that can enter the active region of SPADs and initiate dark counts.

4.1. Introduction

CMOS single-photon avalanche diodes (SPADs) have lower cost and can be easily integrated with other CMOS circuits to form complete and compact systems compared to a custom technology. For Geiger-mode operation, the SPAD is biased in excess of its breakdown voltage, so guard rings (GRs) are generally used to prevent premature edge breakdown (PEB). However, GRs can increase the total area of a SPAD, thus reducing its

----

Adapted from W. Jiang, R. Scott, and M. J. Deen, "Improved Noise Performance of CMOS Poly Gate Single-Photon Avalanche Diodes," *IEEE Photonics J.* (Submitted on July 22, 2021)

fill factor. More importantly, the GR structure can significantly affect the SPAD's noise performance.

There are several options for GRs for SPADs in CMOS processes to prevent PEB. One solution uses shallow trench isolation (STI) to confine the edges of the depletion region [94]. However, such SPADs suffer from high dark noise due to the high density of defects at the Si-SiO_2 interface during the formation of STI. To mitigate this issue, diffusion GR structures are preferred. For example, n-well diffusion GRs were used for n^+/p-well junctions [31], [162] and p-well diffusion GRs were used for p^+/n-well junctions [125], [126]. The lower doping in the diffusion GRs widens the depletion regions at the edges of SPADs to locally decrease the electric field, thus preventing PEB. Another alternative are virtual GRs which rely on: the retrograde doping profile that is created from the deep n-well (DNW) to the surface during the formation of the DNW [80]; or the lower doping concentration due to the diffusion between DNWs [103]. The results in [80], [103] showed that the virtual GRs can effectively prevent PEB.

Another interesting approach for GRs uses the polysilicon layer, typically used for the poly gates of MOSFETs. It is known that the formation of the STI introduces a high defect density at the Si-SiO_2 interface, which usually leads to higher dark noise [94]. Therefore, the floating poly gate was used as a stop mask in the SPAD design to prevent the formation of STI underneath. That is, the poly gate helps to push the STI away from the depletion region of the SPAD to reduce the dark noise. The function of the floating poly gate has been investigated in both p^+/n-well and n^+/p -well SPADs [114], [163]. Recently, it was found that the noise performance can be further improved by applying negative voltages on the poly gates of p^+/n-well SPADs [104], [105]. A perimeter-gated p^+/n-well SPAD in a 0.5 μm CMOS process designed with a lateral diffusion n-well guard ring and a perimeter gate over the edge of active region was reported in [104]. By applying a negative voltage on the perimeter gate, the electric field at the edge of the SPAD is reduced, thus preventing PEB. The DCR is also reduced in this perimeter-gated SPAD due to the voltage on the gate. However, the operation of this perimeter-gated structure requires an extra pin to bias the gate. In addition to a common p^+/n-well SPAD with a n-well diffusion GR, the researchers in [105]

presented two other types of SPADs in a 180 nm BCD (BIPOLAR-CMOS-DMOS) process: one p^+/n-well SPAD with a dummy floating gate, and a p^+/n-well SPAD with a field poly gate connected to its p+ anode. The measurements showed that the noise performance of SPAD was improved by adding the dummy poly gate, and was further improved by connecting the poly gate to the anode. The current research on the function of poly gate in SPADs are all based on p^+/n-well SPAD. However, n^+/p SPADs become favorable sensors in the automotive markets because of their improved photon detection efficiency in the near-infrared range if they are back-illuminated [164].

To achieve a better signal-to-noise ratio for n^+/p SPADs, one option is to reduce their dark noise. Based on a careful review, we believe that this is the first report on the design and dark noise improvements of n^+/p poly gate SPADs. Specifically, we designed and measured three types of n^+/p-well SPADs with different configurations of the poly gate. Compared to the poly gate SPADs in the expensive BCD process [105], the SPADs designed in this work use a low-cost standard 180 nm CMOS process, thus decreasing the fabrication cost. In future, these CMOS SPADs can be easily integrated with CMOS peripheral circuits for further signal processing. After comparing the measured DCR and afterpulsing (AP) of these three SPADs at different excess voltages and at different temperatures, we found that the dummy poly gate reduces the DCR up to 48.7% and AP up to 40% when compared to the SPAD without a poly gate. The DCR and AP can be further reduced in the field gated structure.

4.2. SPAD Design

4.2.1. SPAD Structure

Figure 4-1 shows cross-sectional views of the three types of SPADs studied in this work. Figure 4-1(a) shows the design of a traditional n^+/p-well SPAD with only a n-well diffusion GR. The n^+ layer is designed to extend into the n-well diffusion GR, pushing the shallow trench isolation (STI) away from the active region to reduce the dark noise. This SPAD with no poly gate layer is named SPAD_NG. The SPAD shown in Figure 4-1(b) has the same structure except a dummy poly gate that is placed on top of the n-well diffusion region. Because the poly gate is floating during the operation of the SPAD, it is named

SPAD_DG (i.e., dummy gate). Compared to SPAD_DG, the third type of SPAD shown in Figure 4-1(c) has the same structure but the poly gate is connected to the n^+ cathode. In this structure, the poly gate in the SPAD functions like a field gate in a MOSFET which can change the electric field under the poly gate, thus leading to its given name: SPAD_FG. All SPADs have a circular shape with a diameter of 12.4 µm and were fabricated in a standard 180 nm CMOS process.

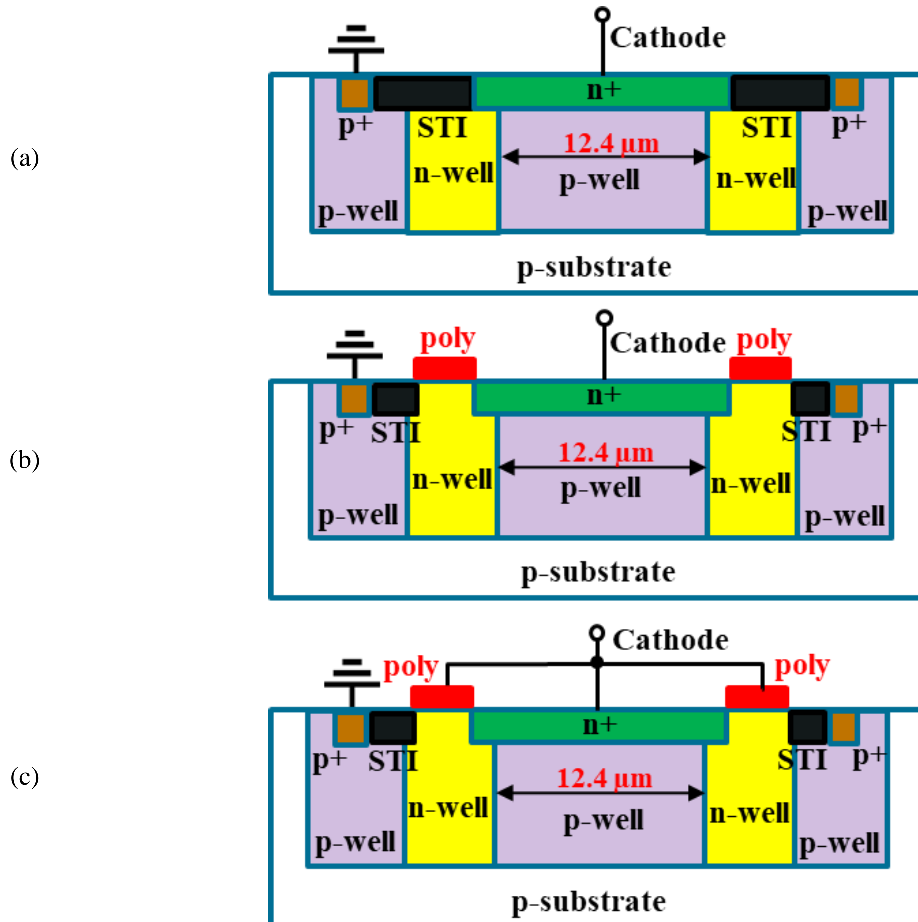


Figure 4-1: Cross-sectional views of (a) of SPAD_NG (n^+/p-well SPAD without poly gate), (b) SPAD_DG (n^+/p-well SPAD with dummy poly gate) and (c) SPAD_FG (n^+/p-well SPAD with field poly gate).

4.2.2. TCAD Simulation

Technology computer-aided design (TCAD) simulations were performed using Sentaurus Device to show differences in the electric field distributions of three types of

SPADs. The doping concentration parameters used in the simulation are based on Ref. [165], and are 5×10^{19} cm^{-3} for the n^+ and p^+ layers, 1.3×10^{17} cm^{-3} for the n-well, 2×10^{17} cm^{-3} for the p-well and 1.3×10^{15} cm^{-3} for the p-substrate. For the simulation, silicon dioxide (SiO_2) is used to form the STI. The simulation results are shown in Figure 4-2 when the SPAD is biased at the measured breakdown voltage of 12.3 V. Since all three SPADs have the same core junction structure, a n^+/p-well junction with a n-well diffusion guard ring, the simulation results showed that the electric field distributions for the three types of SPADs are nearly identical in the central depletion region. However, the electric fields near the edges are different due to the different configurations of their poly gates (see detailed analysis in Section 4.3). As shown in Figure 4-2, the lower electric fields at the edges of the depletion regions demonstrates the effectiveness of the n-well guard ring in preventing PEB for all three SPADs.

4.3. Experimental Results and Discussions

The DCR of 5 samples of chips were measured at room temperature with the same excess voltage (0.5 V). In each group of 5, the SPAD structure with a median value of DCR was chosen to perform detailed measurements to compare their performance. As stated in Section 4.2.1 (SPAD Structure), SPAD_DG and SPAD_FG have the same structure except for their poly gate connection. To eliminate the effect of process variations when comparing the performance of SPAD_DG with SPAD_FG, the same SPAD is used for all measurements. The poly gate is floating when used as SPAD_DG, but is connected to the n^+ cathode externally when used as SPAD_FG.

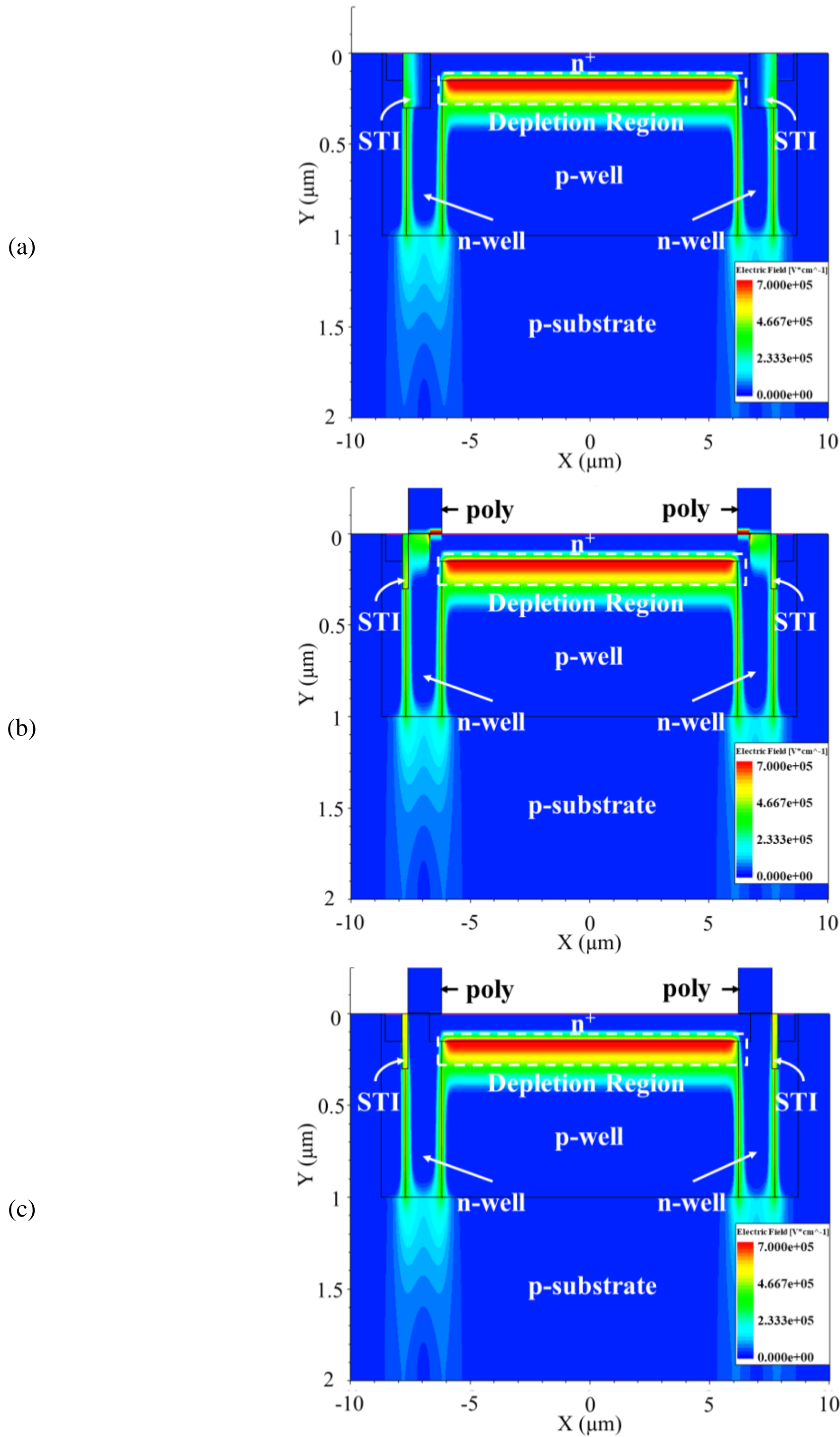


Figure 4-2: TCAD electric field simulations of (a) of SPAD_NG (n+/p-well SPAD without poly gate), (b) SPAD_DG (n+/p-well SPAD with dummy poly gate) and (c) SPAD_FG (n+/p-well SPAD with field poly gate).

4.3.1. Breakdown Voltage

The breakdown voltage (V_{BR}) of a SPAD is defined as the reverse bias voltage that leads the electric field across the junction to reach a critical level capable of causing self-sustaining avalanches. It can be measured through the reverse current-voltage (I-V) characteristic. The voltage at which the onset of the large increasing current is measured corresponds to the breakdown voltage, which is the same as the voltage at a current of 1 μA in our SPADs.

The $V_{BR}s$ of the three types of SPADs were first measured to ensure proper Geiger-mode operation. A Semiconductor Device Analyzer (Agilent B1500A) was used to measure their I-V characteristics to get V_{BR}. To the determine temperature dependence of the SPAD's performance, the measurements were performed from -30 °C to 30 °C in steps of 15 °C. Since the V_{BR} for SPAD_DG and SPAD_FG were measured on the same n^+/p-well junction, they showed almost the same $V_{BR}s$. SPAD_NG shows slightly higher breakdown voltages than the other two SPADs due to the process variations on different samples. Samples of the I-V characteristics for these three SPADs at 15 °C are shown in Figure 4-3(a). At 15 °C, V_{BR} is 12.18 V for SPAD_NG while V_{BR} is 12.14 V for SPAD_DG and SPAD_FG. The temperature dependences of the V_{BR} are shown in Fig. 4-3(b). The temperature coefficients of the breakdown voltages were obtained from the slopes of linear fits of breakdown voltages as a function of temperature [71], [87], which are 12.3 mV/°C for SPAD_DG, 11.6 mV/°C for SPAD_DG and SPAD_FG.

4.3.2. Dark Count Rate (DCR) and Afterpulsing (AP)

The measurement setup for DCR and AP is shown in Figure 4-4(a). The SPAD is passively quenched through an on-chip 50 kΩ resistor connected to the cathode. The other terminal of the quench resistor is connected to a DC voltage source (Agilent E3646A). The anode of the SPAD is grounded. Then, the output signal from the cathode is sent to a high-speed oscilloscope (LeCroy WaveRunner 625Zi) to measure the inter-arrival time (IAT) between the avalanching pulses. Figure 4-4(b) presents an illustration of the IAT measurement. Note that the avalanche pulses in Figure 4-4(b) are negative pulses. In the

illustration, t_n represents the time difference between n^{th} and $(n+1)^{th}$ pulse. For example, t_1 stands for the time between the 1st pulse and the 2nd pulse. After obtaining the histogram of IAT, the DCR and AP can be derived through the exponential fits. The first-order exponential fit represents the primary dark counts and the area deviating from the first-order exponential fit represents the AP, as shown in Figure 4-4 (c). Other details of the IAT method for the AP measurement can be found in [84]. Figure 4-5 shows an example of the experimental data that was measured on SPAD_FG with an excess voltage (V_{ex}) of 0.5 V excess voltage. The measurement was performed at -30 °C to show obvious AP. As shown in Figure 4-5, a multi-exponential fit is applied on the IAT histogram. The primary term of

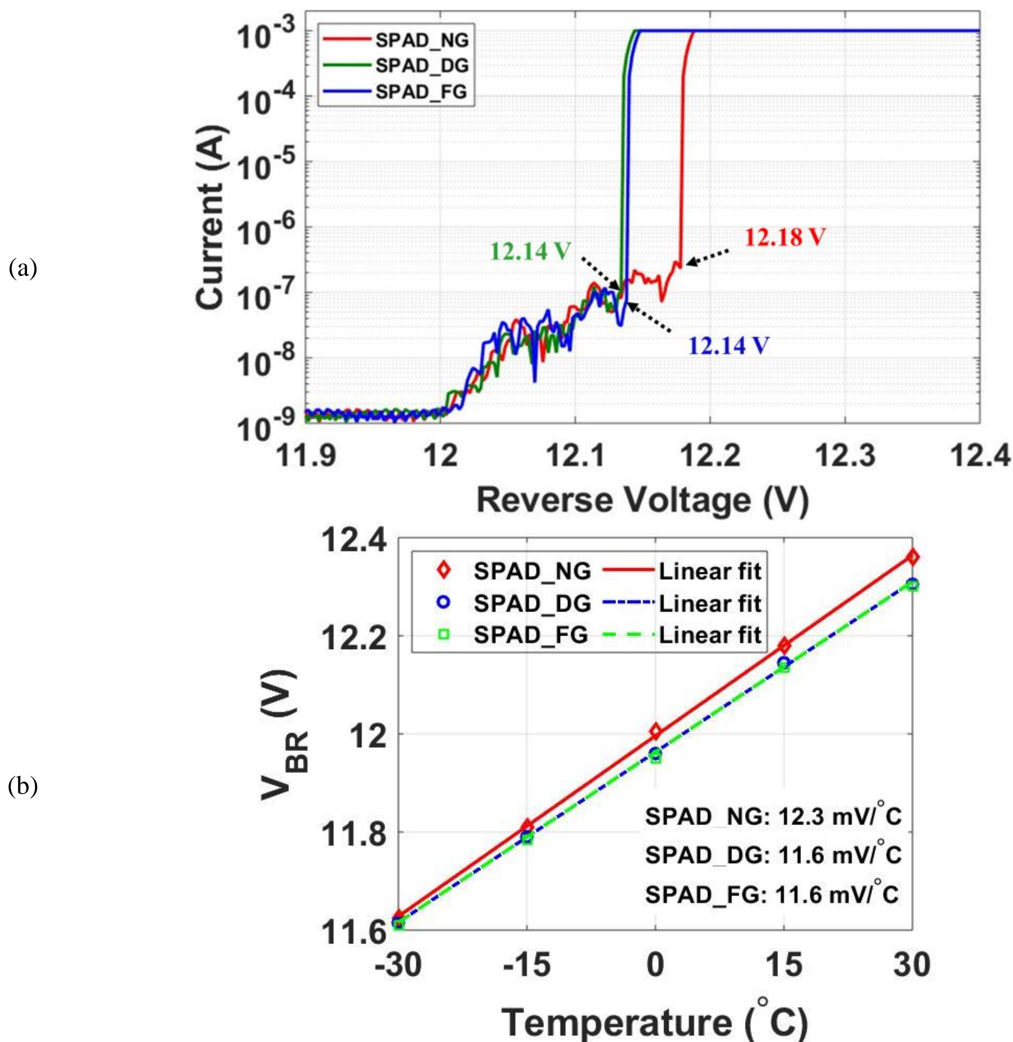


Figure 4-3: (a) Samples of I-V characteristics for three SPADs at 15 °C. Note that the x-axis is 0.5 V width to clearly show the breakdown voltages. (b) Breakdown voltages as a function of temperature for three SPADs

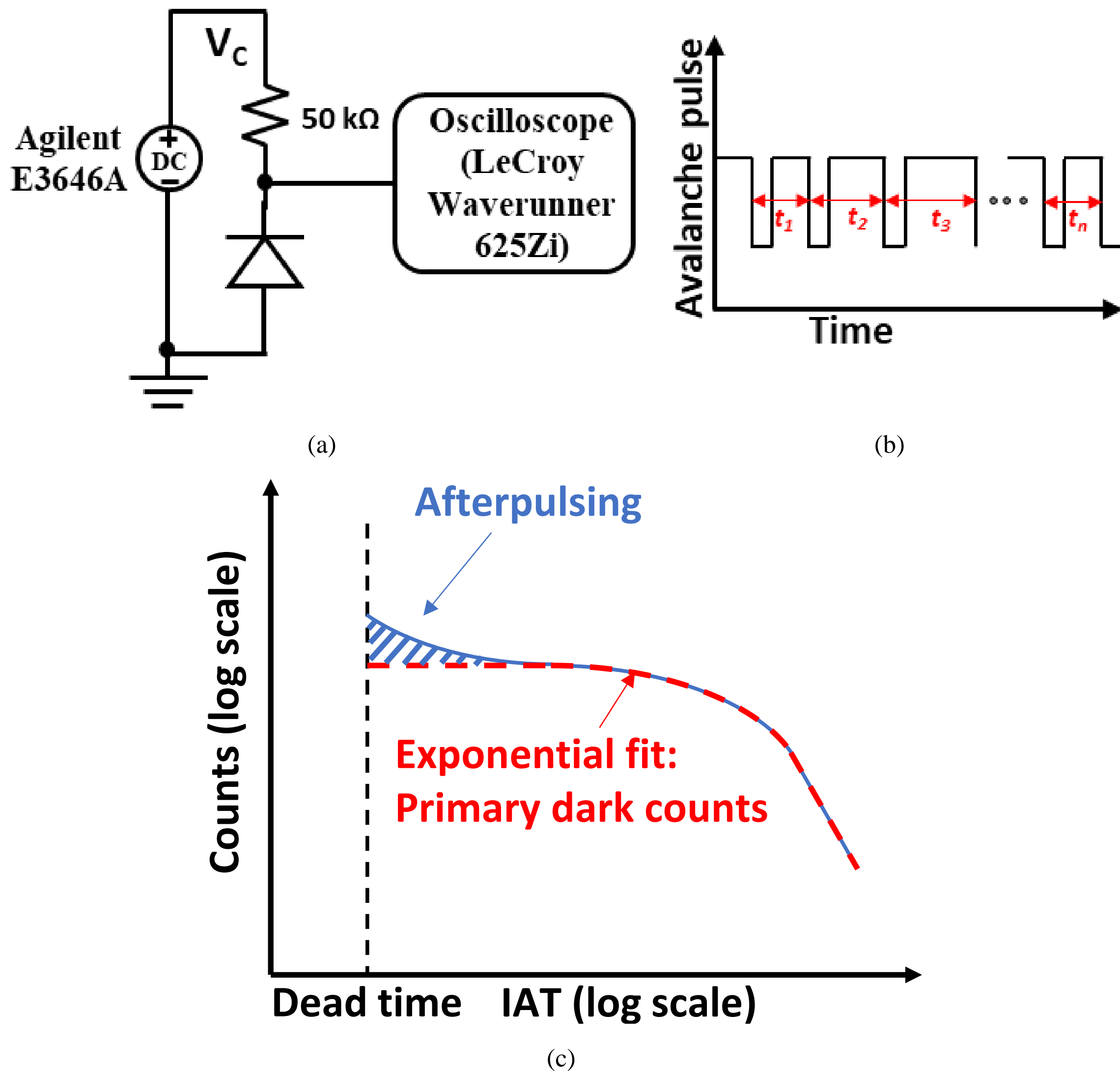


Figure 4-4: (a) Measurement setup of dark count rate (DCR) and afterpulsing (AP) for the passively quenched SPAD. (b) Illustration of inter-arrival time (IAT) of the SPAD output pulses. (c) Illustration of the histogram of IAT to show the components of DCR and AP.

the exponential fit gives the DCR of 2.17 kHz and the AP has a large value of 17.2 % due to the low operating temperature.

The DCR and AP measurements were done at different excess voltages (0.3 V to 0.7 V with a step size of 0.1 V) and different temperatures (-30 °C to 30 °C with a step size of 15 °C). Note that the breakdown voltage temperature coefficient was considered to ensure a constant excess voltage at different temperatures. The measurement results of DCR as a

function of V_{ex} at three different temperatures (15 °C, 0 °C and -15 °C) for the three SPADs are shown in Figure 4-6. Using the DCR of SPAD_NG as the baseline, Figure 4-6(a) shows the DCR comparison of SPAD_DG with SPAD_NG while Figure 4-6(b) shows the DCR comparison of SPAD_FG with SPAD_NG. They all show similar trends: DCR increases with excess voltages and operating temperatures. However, it is seen that SPAD_DG and SPAD_FG can significantly reduce the dark noise when compared to SPAD_NG.

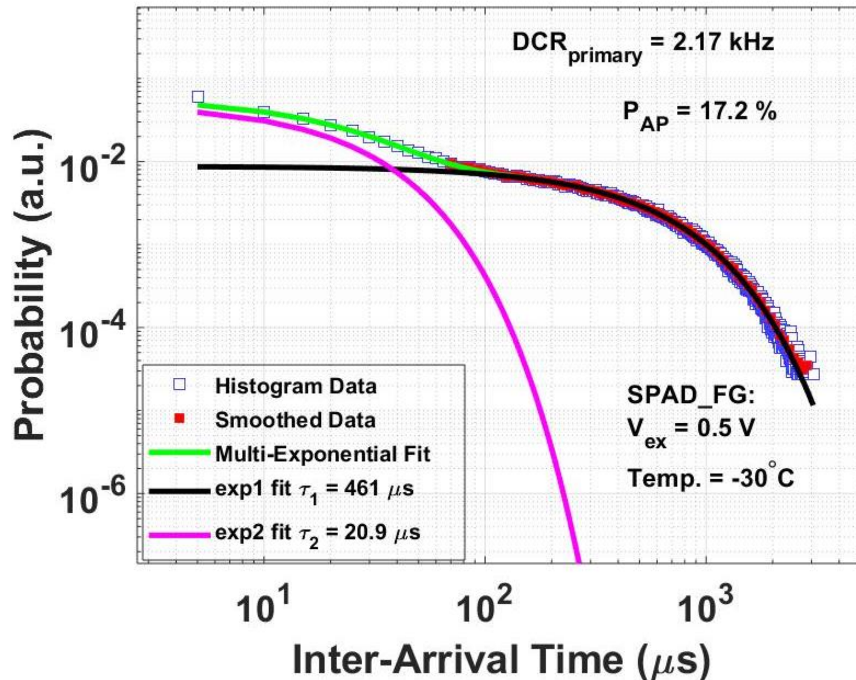


Figure 4-5: The histogram of the IAT from the real measurement data to show the DCR and AP of the SPAD_FG with 0.5 V excess voltage at -30 °C.

Table 4-1 shows the DCR comparison between SPAD_NG and SPAD_DG at different excess voltages and temperatures. The values in Table 4-1 are calculated using $[\mathrm{DCR(SPAD_NG)} - \mathrm{DCR(SPAD_DG)}]/\mathrm{DCR(SPAD_NG)}$. The calculation results show that the DCR of SPAD_DG is reduced by almost two times at -30 °C when compared to SPAD_NG. This is because the dummy poly gate layer on top of the n-well guard ring pushed the STI away from the n^+ layer. Therefore, the contribution to the DCR from the high density of defects at the Si-SiO_2 interface is reduced. It is seen that as the temperature increases, the DCR improvement of SPAD_DG compared to SPAD_NG is reduced. For example, the DCR reductions at -30 °C are more than 40% while DCR reductions at 30 °C

show much smaller values (< 20%). This result implies that at higher temperatures, the contribution from thermally generated carriers in the planar depletion region to the total dark noise is larger than carriers diffusing from the guard ring.

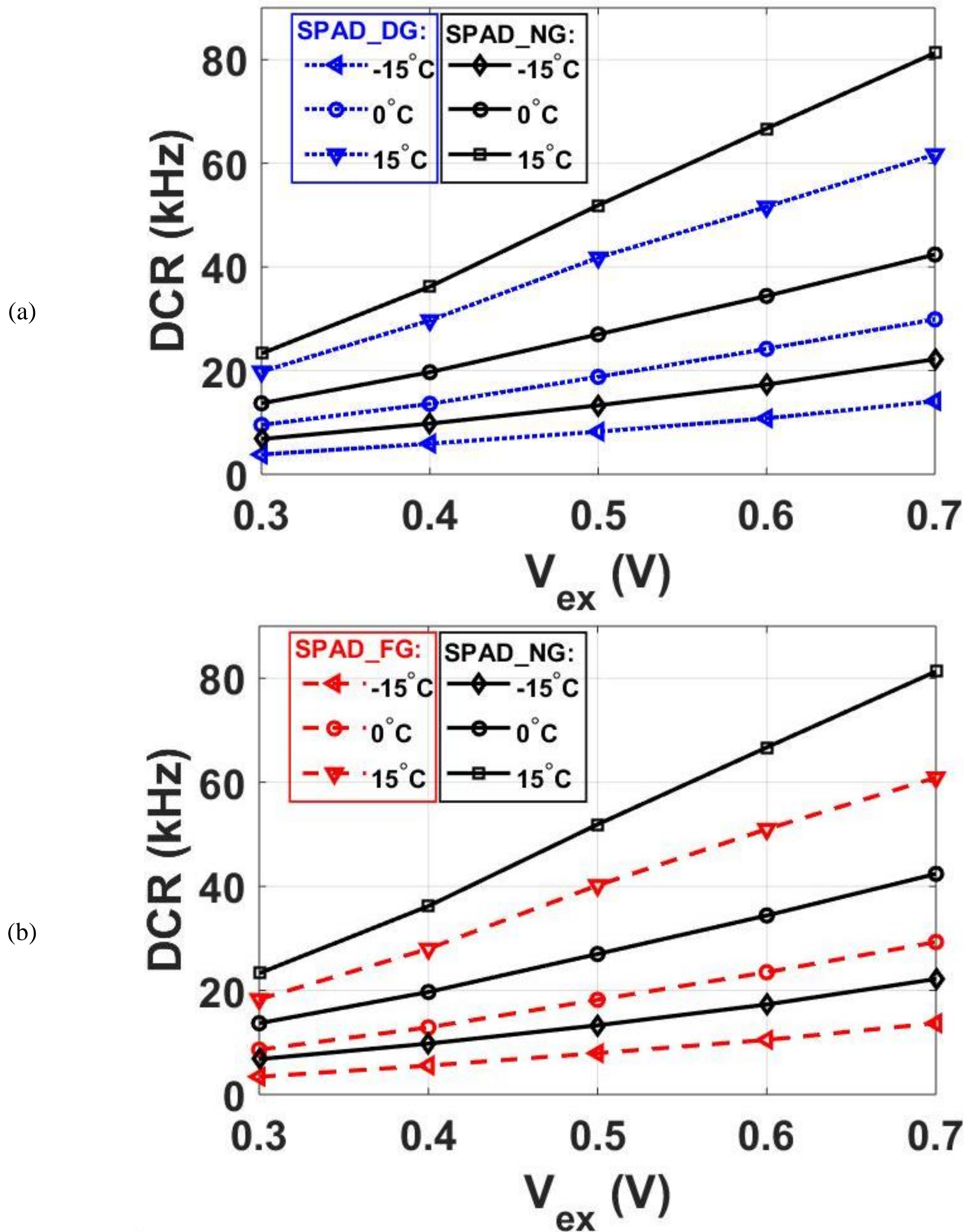


Figure 4-6: Comparison of the DCR of SPAD_NG with (a) SPAD_DG and (b) SPAD_FG at different excess voltages and at different temperatures.

Table 4-1: DCR comparison between SPAD_NG and SPAD_DG: [DCR(SPAD_NG)-DCR(SPAD_DG)]/DCR(SPAD_NG) × 100 (%)

Temp. (°C)	**Excess Voltage (V)**				
	0.3	**0.4**	**0.5**	**0.6**	**0.7**
-30	48.7	47.8	44.5	44.4	40.7
-15	43.8	39.4	38.2	37.6	36.5
0	30.5	31.0	30.4	29.7	29.5
15	15.0	18.0	19.5	22.6	24.1
30	3.4	7.7	13.8	14.5	16.6

Table 4-2 shows the DCR comparison between SPAD_NG and SPAD_FG at different excess voltages and at different temperatures. The values in this table are calculated using $[\mathrm{DCR(SPAD_NG)} - \mathrm{DCR(SPAD_FG)}]/\mathrm{DCR(SPAD_NG)}$. Compared to Table 4-1, the values in Table 4-2 are better (~2 - 10%), which indicates that SPAD_FG shows further reduction in DCR. Similar to SPAD_DG, the poly gate pushes the STI away from the active region of SPAD. In addition, the poly gate connected to the cathode of SPAD_FG reduces the electric field distribution near the n-well guard ring.

Figure 4-7 shows TCAD simulations focussed on the edges between the n-well guard ring and depletion region for SPAD_DG and SPAD_FG when the SPAD is biased at the measured breakdown voltage of 12.3 V. The simulation results show that the field poly gate in SPAD_FG helps to suppress the electric field in the region of n-well guard ring when compared to SPAD_DG since it ensures that there is no voltage difference between the poly gate and the semiconductor located underneath. The reduced electric field at the n-well guard ring of SPAD_FG improves the effectiveness of the guard ring, and lowers the probability that carriers in the guard ring region can enter the depletion region, thus reducing the DCR. The implementation of a field gate in the SPAD structure does not require extra customized masks in a standard CMOS process, or an additional pin to bias the field gate since it is connected to the cathode. Therefore, to reduce DCR, it is beneficial to implement the SPAD_FG structure when designing n+/p-well SPADs with a poly gate.

Table 4-2: DCR comparison between SPAD_NG and SPAD_FG:
[DCR(SPAD_NG)-DCR(SPAD_FG)]/DCR(SPAD_NG) × 100 (%)

Temp. (°C)	**Excess Voltage (V)**				
	0.3	**0.4**	**0.5**	**0.6**	**0.7**
-30	52.8	50.7	46.9	45.3	41.5
-15	49.7	43.1	40.3	39.3	38.3
0	37.1	34.5	32.6	31.7	30.9
15	21.5	22.7	22.5	23.5	25.1
30	11.6	12.4	16.1	16.2	18.3

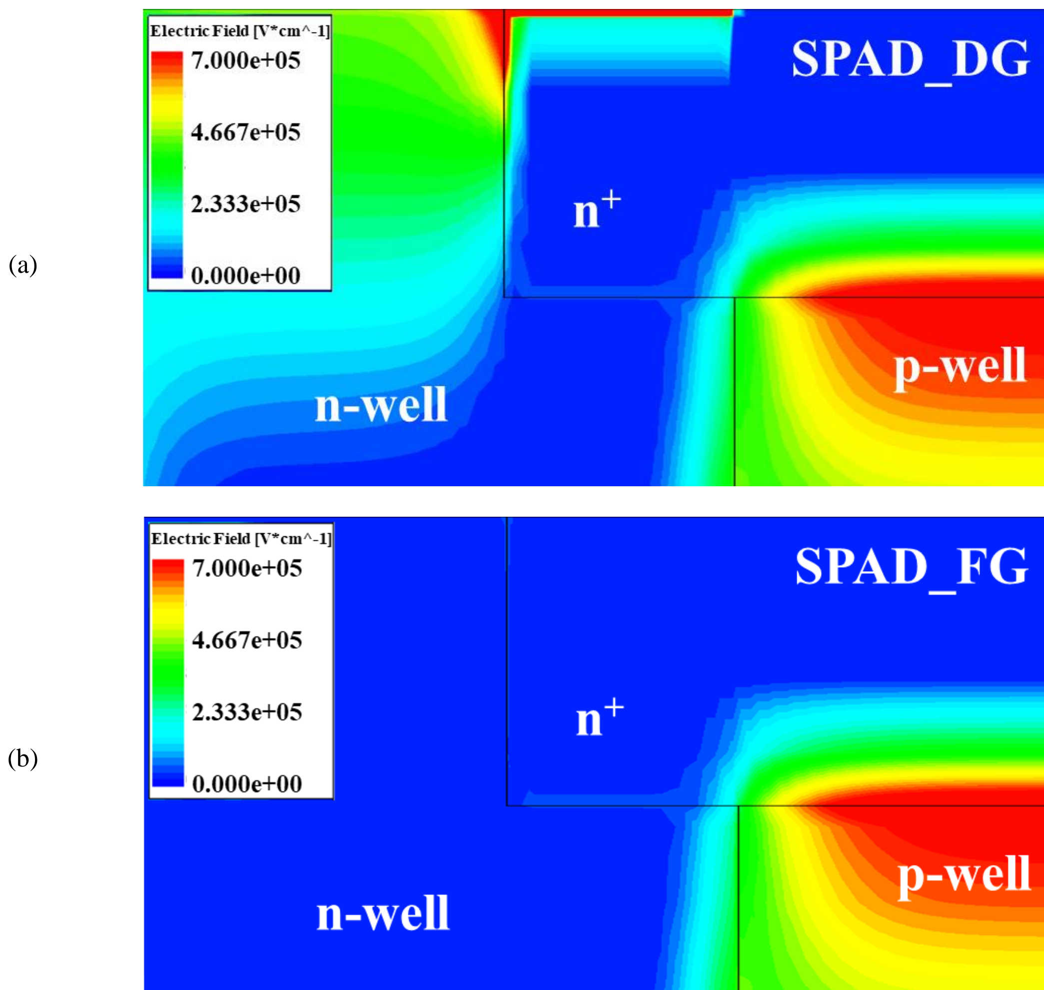


Figure 4-7: TCAD electric field simulations at the edge of depletion region of (a) SPAD_DG and (b) SPAD_FG.

The DCR of SPADs can be expressed as a function of temperature using the Arrhenius equation, given by [121]

$$\mathrm{DCR} \propto T^2 \exp\left(-\frac{E_A}{k_B T}\right) \tag{4-1}$$

where E_A is the activation energy, k_B is Boltzmann's constant, and T is the absolute temperature in Kelvin. By taking the natural logarithm on Eq. 4-1, we get that $ln(\mathrm{DCR}/T^2)$ has a linear relationship to $1/k_BT$. The slope of this linear fit gives the activation energy E_A. The Arrhenius plots of the three SPADs at 0.5 V excess voltage are shown in Figure 4-8, and the E_A is 0.24 eV for SPAD_NG, and 0.29 eV for both SPAD_DG and SPAD_FG. The small values of the activation energy in all three SPADs indicate that the trap-assisted tunneling noise is likely the dominant source for the dark noise in the measured temperature range (-30 to 30 °C).

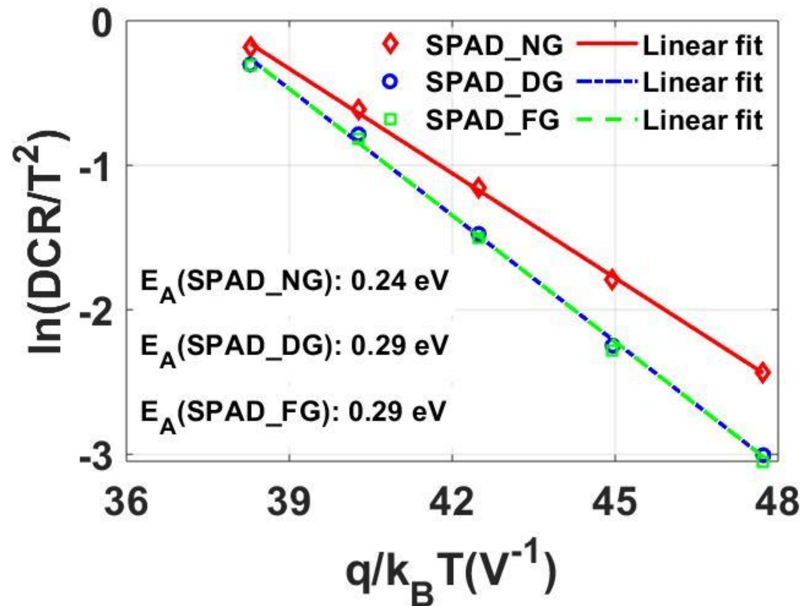


Figure 4-8: Arrhenius plots of three SPADs with 0.5 V V_{ex}.

The measurement results of AP as a function of V_{ex} at three different temperatures (15 °C, 0 °C and -15 °C) for the three SPADs are shown in Figure 4-9. Using the AP of the SPAD_NG as the baseline, Figure 4-9(a) shows the AP comparison of SPAD_DG with SPAD_NG while Figure 4-9(b) shows the AP comparison of SPAD_FG with SPAD_NG.

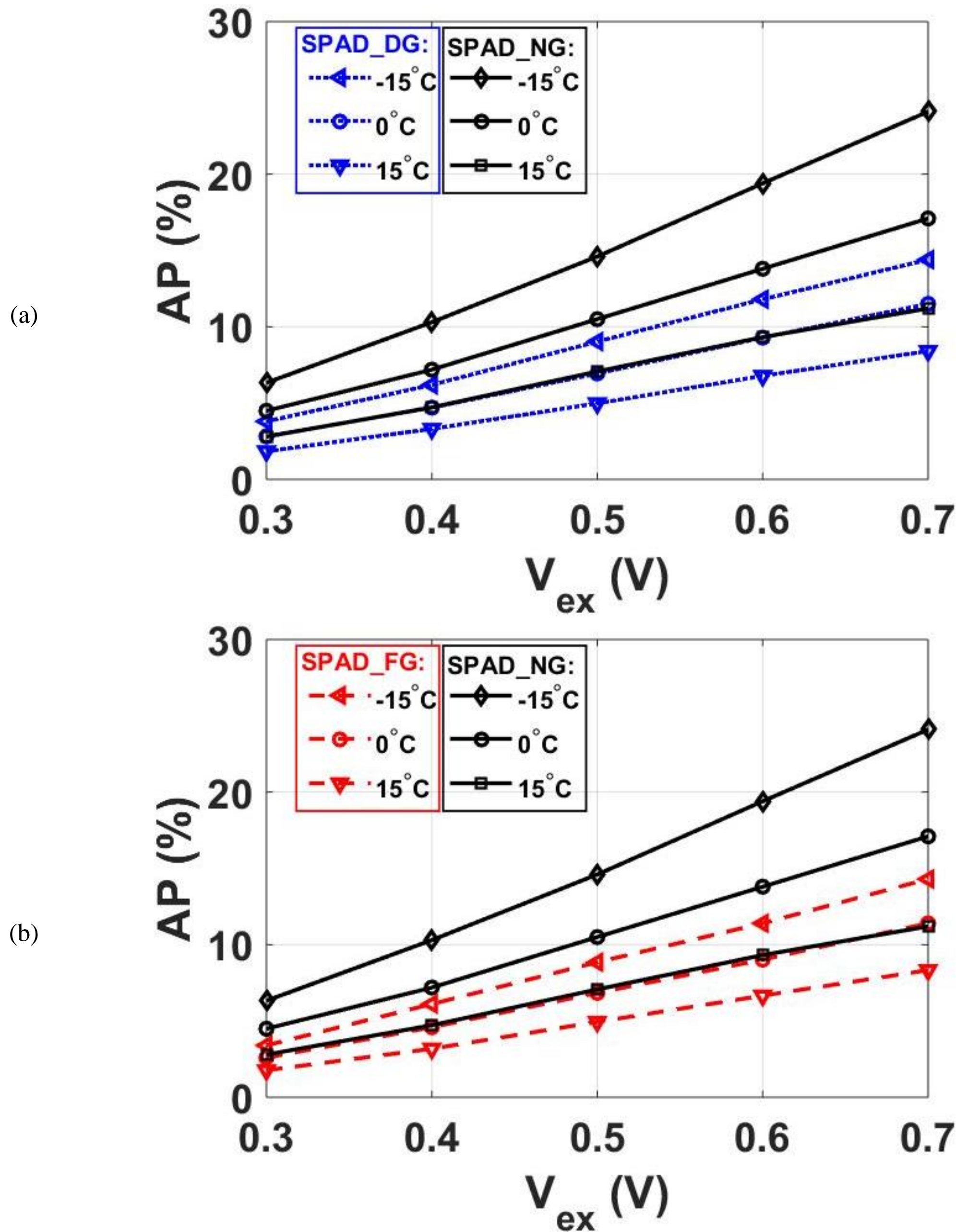


Figure 4-9: Comparison of the afterpulsing (AP) of SPAD_NG with (a) SPAD_DG and (b) SPAD_FG at different excess voltages and at different temperatures.

It is seen that AP increases as temperature decreases. This is because the lifetime of trapped carriers in the trap centers increases exponentially as temperature decreases, thus resulting in a higher probability for carriers released from the trap centers to generate a secondary pulse. As for the dependence on excess voltage, the measurement results indicate

that the AP has a linear relationship to the excess voltages at different temperatures for all three SPADs. This linear relationship can be interpreted using the model described in [67] as

$$P_{AP} \approx (C N_t \sigma_t W_e V_{ex})/q \quad (4\text{-}2)$$

where P_{AP} is the afterpulsing probability, C is the junction capacitance of the SPAD, N_t is the electron trap concentration, σ_t is the cross-section of the electron trap, and W_e is the effective depletion width. Note that SPAD_DG and SPAD_FG do not show the any AP at 30 °C. This means that lifetime of captured carriers by the trap centers of SPAD_DG and SPAD_FG at 30 °C is small enough so there is negligible generation of secondary pulses.

The detailed comparison of AP between the three SPADs are shown in Tables 4-3 and 4-4, which demonstrates similar trends as DCR. Compared to SPAD_NG, SPAD_DG showed improved AP performance as shown in Table 4-3. For example, the AP reduction is as high as 40% at 0.3 V V_{ex} and -15 °C. Compared to Table 4-3, the values in Table 4-4 are better (more reduction), which indicates AP performance of SPAD_FG was further improved when compared to SPAD_DG. As stated in the section on DCR, the field gate in SPAD_FG not only pushes the STI away from the active region of SPAD, but it also reduces the electric field at the edge of the depletion region. This leads to a reduction in carriers diffusing from the guard ring region into the depletion region, thus resulting in the best AP performance among the three SPADs.

Table 4-3: AP comparison between SPAD_NG and SPAD_DG: [AP(SPAD_NG)-AP(SPAD_DG)]/AP(SPAD_NG) × 100 (%)

Temp. (°C)	**Excess Voltage (V)**				
	0.3	**0.4**	**0.5**	**0.6**	**0.6**
-30	17.7	13.2	12.3	12.8	13.9
-15	40	39.9	38.2	39.2	40.2
0	37.2	34.6	33.9	32.7	32.7
15	34.3	29.9	29.6	27.3	24.9

Table 4-4: AP comparison between SPAD_NG and SPAD_FG:
[AP(SPAD_NG)-AP(SPAD_FG)]/AP(SPAD_NG) × 100 (%)

Temp. (°C)	**Excess Voltage (V)**				
	0.3	**0.4**	**0.5**	**0.6**	**0.6**
-30	24.3	17.1	15.7	16.3	14.6
-15	46.3	40.7	39.4	41.2	40.7
0	42.1	36.2	34.9	34.6	33.3
15	36.8	32.6	30.1	28.5	25.6

4.3.3. Performance Summary and Comparison

In Table 4-5, a summary of the properties and noise performance of the three SPAD designs (last row) and a comparison to other SPADs in the recent literature is given. However, note that only [73] used a process similar to ours. The n^+/p-well SPADs in this work show better noise performance than other SPADs fabricated in standard CMOS processes. However, the DCR of these three SPADs are higher than SPADs in other non-standard 180 nm processes. Several reasons could be responsible for the higher dark noise. First, the three SPADs in this work are implemented in a fully standard CMOS process and no special layers such as buried-N layer in [125], [126] and [105], or deep p-well layer [81] are available to isolate the active region of SPADs from the noisy bulk. Second, even though the DNW is available in this 180 nm process, it cannot be used in this n^+/p-well SPAD design since the bottom of the n-well reaches the DNW in this process. As shown in Figure 4-1, if the DNW was placed underneath the p-well to isolate the active region of the SPAD from the p-substrate, the n-well guard ring will touch the DNW, thus making the p-well anode of the SPAD inaccessible. Third, compared to breakdown voltages of the SPADs with a very low DCR (23.5 V in [125], 16.8 V in [81], 25.46 V in [166]), the three n^+/p-well SPADs in this work present a relatively low breakdown voltage (~12.1 V). This means that the doping concentrations of the n^+ layer and p-well in this 180 nm standard CMOS process are higher, thus leading to a thinner depletion region and a higher

probability for tunneling noise, which is in agreement with the small activation energy values for three SPADs shown in Figure 4-8.

The comparison between the three SPADs in this work proves the effectiveness of a field gate in n^+/p-well SPADs to prevent PEB and to reduce the primary DCR and AP. Also, since these SPADs are designed in a standard 180 nm process, they can be easily integrated with other CMOS signal processing and conditioning circuits to form a complete imaging system with lower costs and very good overall performance.

4.4. Conclusions

We designed and fabricated three types of n^+/p-well SPADs with different poly gate configurations: no poly gate (SPAD_NG), a dummy floating poly gate (SPAD_DG), and a field poly gate connected to the n^+ cathode (SPAD_FG) in a low-cost standard 180 nm CMOS process. Based on these three types of SPADs, the TCAD simulations showed the effectiveness of the guard rings to prevent premature edge breakdown (PEB). The dark count rate (DCR) and afterpulsing (AP) of these three SPADs operating at different excess voltages and at different temperatures were measured and compared. The comparison of measurement results between SPAD_NG and SPAD_DG or SPAD_FG reveal that the poly gate greatly helps to reduce the DCR and AP. This is because the presence of the poly gate over the n-well guard ring pushes the noisy Si-SiO_2 interface of the STI away from the active region of the SPAD. The best results were obtained from SPAD_FG because the voltage on the poly gate reduces the magnitude of the electric field distribution in the guard ring region, as shown from the TCAD simulation. The reduced electric field decreases the likelihood that carriers in the guard ring can enter the active region of the SPAD and initiate dark counts, resulting in improved the noise performance. In addition, SPAD_FG in this work is implemented in a standard CMOS process, which does not require extra mask layers in the fabrication nor an additional pin for biasing the SPAD, thus reducing the effort for integration of SPADs with other CMOS signal conditioning and processing circuits. Overall, this work verifies that the n^+/p-well field gate SPAD is an effective structure for improving the noise performance of CMOS SPADs.

Table 4-5: Comparison of our SPADs to other SPADs in Recent Literature

Ref., Year	Tech. (nm)	Junction	Active area (μm)	Guard ring	V_{BR} (V)	DCR (cps/μm^2) @ V_{ex}, T	Afterpulsing (%) @ V_{ex}, T
Non-standard CMOS Process							
[125], 2014	180 nm CMOS	p^+/DNW	12 (Circular)	p-well	23.5	0.15 @ 3 V, 25°C 12.84 @ 10 V, 25°C	0.03 @ 2 V, 25°C 0.3 @ 10 V, 25°C
[126], 2015	180 nm CMOS	p^+/n-well	12 (Circular)	p-well	14.64	16 @ 4 V, 25°C	0.2 @ 4 V, 25°C
[81], 2018	180 nm HV CMOS	p^+/shallow n-well	12.08 (Square)	shallow p-well	16.8	0.19 @ 1 V, 25°C 1.49 @ 4 V, 25°C	-
[166], 2018	180 nm CMOS	p-well/p-epi/BN	16.38 † (Circular)	PW lateral diffusion and low doping p-epi	25.46	0.26 @ 4 V, 25°C	-
[105], 2020	180 nm BCD	p^+/n-well	10 (Circular)	p-well	11.55	6.34 @ 1 V, 22°C	-
				p-well and poly gate (floating)	11.85	4.82 @ 1 V, 22°C	-
				p-well and poly gate (connected to p^+ anode)	11.4	2.37 @ 1 V, 22°C	-
Standard CMOS Process							
[99], 2013	65 nm standard CMOS	n+/p-well	8 (Octagonal)	n-well and poly gate (floating)	9.1	15.6 k @ 0.4 V, 20℃	1 @ 0.4 V, 25℃
[167], 2018	65 nm standard CMOS	p+/n-well	20 (32-sided polygon)	p-well	9.9	2.8 k @ 1.5 V, 25 ℃	<10 @ 1.5 V, 25 ℃
[121], 2021	65 nm standard CMOS	DNW/p-sub	10 (Square *)	-	12.35	550 @ 0.3 V, 25℃	<1 @ 0.4 V, 25℃
[73], 2008	180 nm standard CMOS	p^+/n-well	10 (Octagonal)	p-well	10.2	764 @ 0.5 V, 20°C	-
This work	180 nm standard CMOS	n^+/p-well	12.4 (Circular)	n-well	12.18	430 @ 0.5 V, 15°C	7.07 @ 0.5 V, 15°C
				n-well and poly gate (floating)	12.14	346 @ 0.5 V, 15°C	4.98 @ 0.5 V, 15°C
				n-well and poly gate (connected to n^+	12.14	333 @ 0.5 V, 15°C	4.94 @ 0.5 V, 15°C

* Square with 45° corners. † The active area was not reported. The value of 16.38 μm is the pixel pitch in a 512 × 512 SPAD array.

Chapter 5
High-speed SPAD Pixel with Active Quench and Reset Circuit

5.1. Introduction

Single-photon avalanche diodes (SPADs) fabricated in standard complementary metal-oxide-semiconductor (CMOS) technologies have a lower cost and can be easily integrated with other CMOS front-end and signal processing circuits to achieve strong detector performance at the system level. To ensure the Geiger mode operation, a quench circuit is needed to quench the avalanching process when it is triggered. Also, a reset circuit is required to reset the SPAD back to its initial biasing condition to detect the next photon. One simple method is the passive quench and reset (PQR) circuit that uses a large resistor connected to either the cathode or anode of the SPAD. The PQR circuits are simple and can be easily implemented in a small layout area in standard CMOS processes. However, the long reset time due to the large quench resistor limits their applications in high-speed detection systems. To address this issue, active quench and reset (AQR) circuits are used in many SPAD pixel designs.

A fast AQR circuit using 0.35 µm CMOS technology was reported in [128]. This fast front-end circuitry consisting of 18 transistors achieved a short active quenching time (QT) of ~1 ns and a short reset time of 3 ns when the SPAD was biased at 6 V excess voltage (V_{ex}). Due to the fast QT, the SPAD achieved a low afterpulsing (AP) of 1.3% with 20 ns hold-off time. Researchers in [168] reported a SPAD integrated with a cascaded quenching circuit that was fabricated in a 0.35 µm CMOS technology. The active QT was only 0.48

Adapted from W. Jiang, R. Scott, and M. J. Deen, "High-speed Active Quench and Reset Circuit for SPAD in a Standard 65 nm CMOS Technology," *IEEE Photonics Technology Letters* (submitted on July 13, 2021)

ns, and the dead time (DT) could be adjusted from 9.5 to 17 ns. The AP was only 0.9% with its highest V_{ex} (6.6 V) and longest DT (17 ns). An AQR circuit in a 0.18 μm CMOS process in [169] used a positive feedback loop to facilitate the quenching process. The simulation and measurements showed that the QT was 0.7 ns with a minimum DT of 5 ns (corresponding a maximum count rate (CR) of 200 Mcps). The AP was 0.75% with a 4 ns hold-off time.

In summary, several key points need to be considered in the design of an AQR circuit. First, the SPAD's load capacitance from the AQR circuit should be minimized to reduce the avalanching charge that may be trapped to later initiate afterpulses. Second, the avalanching process should be quenched by the AQR circuit as soon as possible to minimize the avalanching current. That is, the QT should be kept as short as possible. Third, the pulse width for the reset control signal should be set long enough to ensure that the SPAD can be fully reset to its initial biasing condition. Fourth, an adjustable hold-off time may be needed to deal with the trade-off between the CR and AP. Fifth, the area of the AQR circuit should be kept small to avoid too much reduction of the pixel fill factor (FF).

In this work, a compact high-speed AQR circuit is designed and fabricated using a 65 nm standard CMOS technology. To investigate the performance of the AQR circuit, it is integrated with a p^+/n-well SPAD. The results show that the AQR circuit achieves a fast QT due to the use of a positive feedback loop, a short DT and a small layout area because of the advanced technology node.

5.2. SPAD Design

5.2.1. SPAD Structure

The cross-sectional view of the p^+/n-well SPAD is shown in Figure 5-1. The diameter of the active area was 10 μm and it had a silicide blocking layer to improve the light absorption. To prevent premature edge breakdown (PEB), an octagonal shape and a p-well guard ring were used. The layout of this SPAD is shown in Figure 5-2.

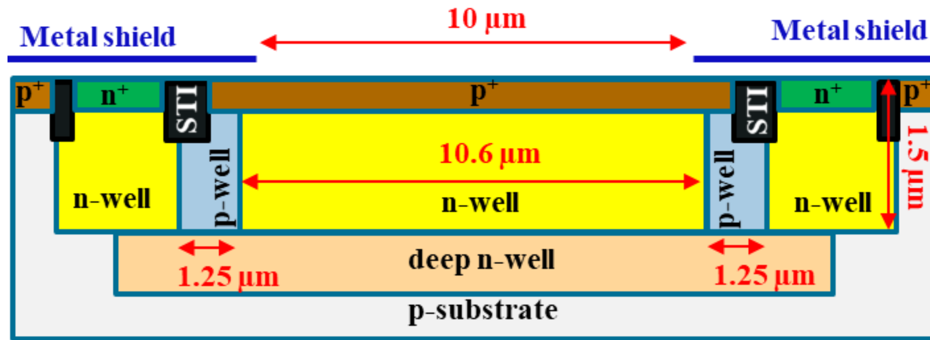


Figure 5-1: Cross-sectional view of the p+/n-well SPAD.

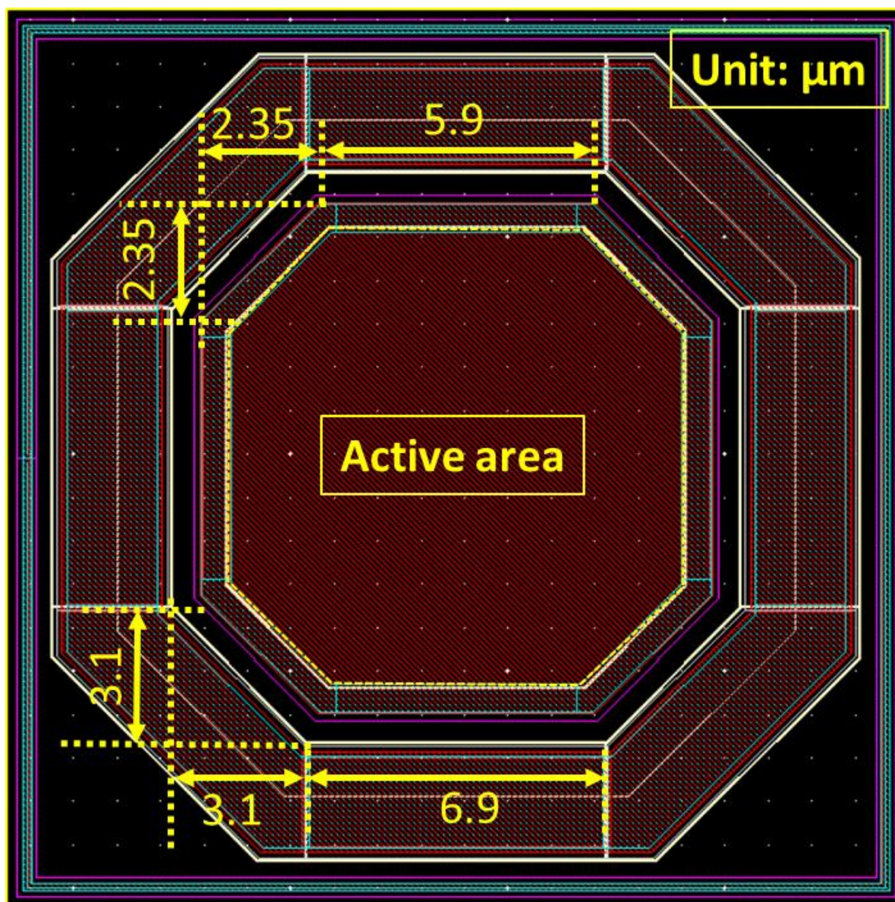


Figure 5-2: (a) The layout of the p+/n-well SPAD with the dimensions of the octagon.

5.2.2. TCAD Simulation

The uniformity of the electric field distribution was verified by the Technology Computer-Aided Design (TCAD) simulation shown in Figure 5-3. The doping profiles used in the TCAD simulation are as follows. The doping concentration for n^+ and p^+ layers is 5×10^{19} cm^{-3}, the doping concentration for n-well and p-well is 2×10^{18} cm^{-3}, the doping concentration for DNW (deep n-well) is 2×10^{18} cm^{-3} and the doping concentration for the p-substrate is 1×10^{16} cm^{-3}. While the anode terminal had a direct contact to the p^+ region, the presence of the p-well guard ring requires the cathode to be connected through the deep n-well. The p^+ region was extended outside of the active area and into the p-well to push the shallow trench isolation (STI) away from the depletion region of the SPAD to reduce the dark counts caused by the defects at the interface of the STI.

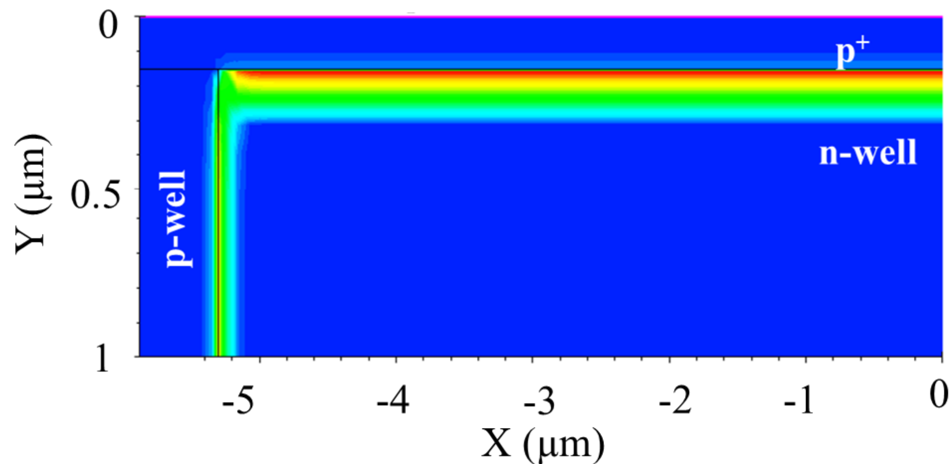


Figure 5-3: (a) TCAD simulation of the p^+/n-well SPAD shows a uniform electric field distribution in the central planar area and a reduced electric field at the edge.

5.2.3. SPAD model parameters

Figure 5-4 shows an equivalent circuit model for SPADs, in which there are 3 key parameters: SPAD junction capacitance (C_{SPAD}), avalanching resistance (R_{SPAD}) and the breakdown voltage (V_{BR}). With these three parameters, a SPICE (Simulation Program with Integrated Circuit Emphasis) model can be developed for circuit simulation [108].

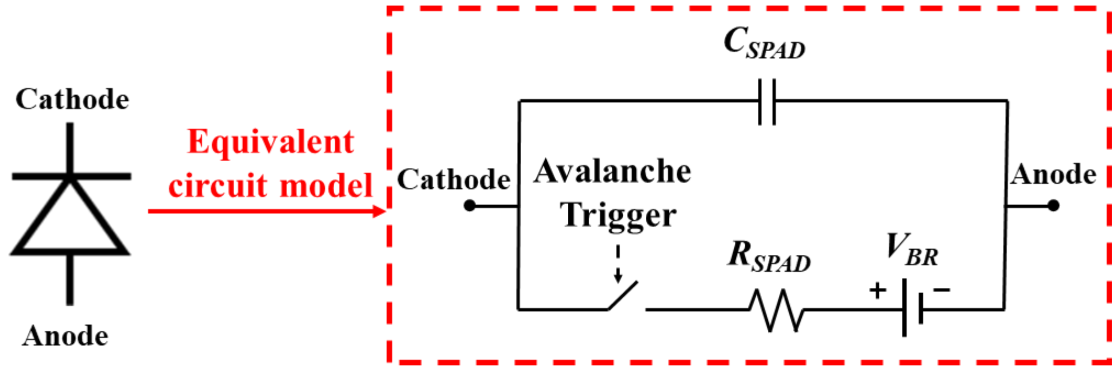


Figure 5-4: (a) Equivalent circuit mode for SPADs

The breakdown voltage V_{BR} can be measured through the reverse I-V characteristic of the SPAD (see details in Section 5.4.1). R_{SPAD} can also be extracted from the I-V characteristics. Figure 5-5(a) shows a typical example of the I-V characteristics at room temperature. After the reverse voltage reaches the breakdown voltage (9.87 V), the SPAD begins avalanching and the current shows a linear relationship with the reverse voltage. R_{SPAD} is the inverse of the slope of linear fit on the linear part of the I-V characteristics. As shown in Figure 5-5(b), the slope of the linear fit is 0.007, giving a resistance of 142.8 Ω (1/0.007) for R_{SPAD}.

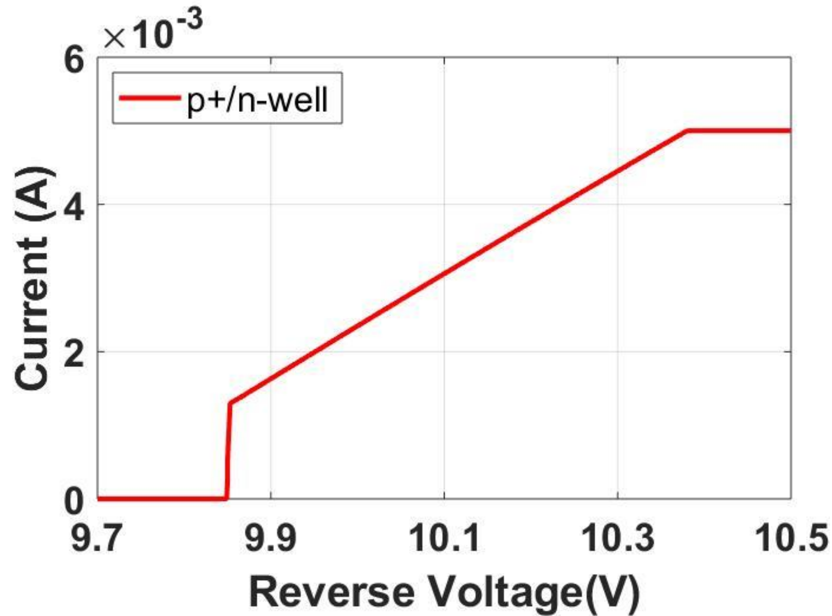
C_{SPAD} can be estimated through theoretical calculations. The capacitance for a p-n abrupt junction can be expressed as [160]

$$C = \sqrt{\frac{N_D N_A}{N_D + N_A} \times \frac{e\varepsilon_s}{2(V_{bi} + V_{BIAS})}}, \tag{5-1}$$

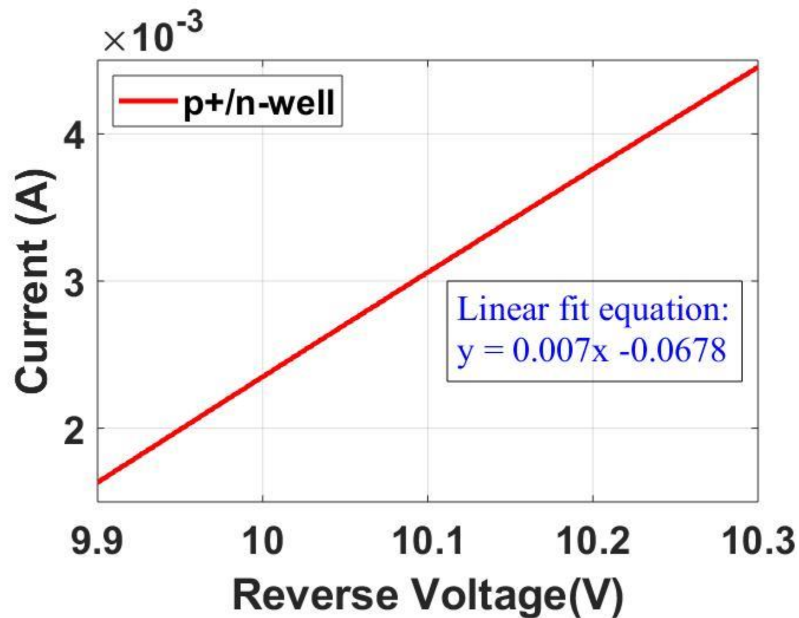
where ε_s (= 1.04 × 10^{-10} F/m) is the permittivity of silicon, e (= 1.6 × 10^{-19} C) is the charge of an electron, V_{BIAS} is the biasing voltage across the junction, V_{bi} is the built-in potential barrier, N_A is the doping concentration of the p-type silicon and N_D is the doping concentration of the n-type silicon.

Since the SPAD output is generated from the anode in the SPAD pixel design (see details in Section 5.3.1), its anode capacitance has great influence on its performance, while its cathode capacitance is connected to an AC ground. From the cross-sectional view of the SPAD, the capacitance at the anode of the SPAD contains two parts: the vertical p^+/n-well SPAD junction capacitance and the p-well/n-well or p-well/DNW side junction capacitance.

Figure 5-6 indicates the location of these two types of junctions. In order to clearly show these two capacitances, the same cross-sectional view of the SPAD shown in Figure 5-1 is used in Figure 5-6, but without the original texts.



(a)



(b)

Figure 5-5: (a) Reverse I-V characteristic of SPAD. (b) Linear fit on the linear part of I-V characteristic where the reverse voltage is larger than the breakdown voltage.

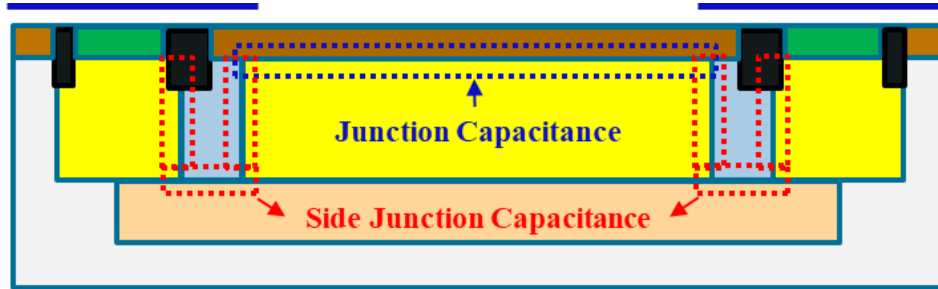


Figure 5-6: SPAD junction capacitance and side junction capacitance in the p+/n-well SPAD.

The p+/n-well junction capacitance can be calculated using Eq. 5-1:

$$C_{p+/nw} = \sqrt{\frac{(2\times10^{18})\times(5\times10^{19})}{(2\times10^{18})+5\times10^{19}}\times10^{6}\times\frac{(1.6\times10^{-19})\times(1.04\times10^{-10})}{2\times(0.7+9.9)}} = 1.2\times10^{-3}\ \mathrm{F/m^2} \quad (5\text{-}2)$$

Since the n-well and DNW are assumed to have the same doping concentration, we use $C_{pw/nw}$ to denote both p-well/n-well and p-well/n-well side junctions. Similarly, the side junction capacitance can also be calculated using Eq. 5-1:

$$C_{pw/nw} = \sqrt{\frac{(2\times10^{18})\times(2\times10^{18})}{(2\times10^{18})+2\times10^{18}}\times10^{6}\times\frac{(1.6\times10^{-19})\times(1.04\times10^{-10})}{2\times(0.7+9.9)}} = 0.85\times10^{-3}\ \mathrm{F/m^2} \quad (5\text{-}3)$$

Using the dimensions shown in Figure 5-2, the area of the p+/n-well junction can be calculated to be 101.3 μm² as shown in Eq. 5-4.

$$A_{p+/nw} = 10.6^2 - 2\times2.35^2 = 101.3\ \mu\mathrm{m}^2 \quad (5\text{-}4)$$

The area for the side p-well/n-well junction is calculated to be 174.1 μm² as shown in Eq. 5-5.

$$A_{pw/nw} = 1.5\ \times[(5.9+3.32)\times4\ +(6.9+4.38)\times4\,]+[(13.1^2-2\times3.1^2) - 101.3] = 55.3+67.7+51.1 = 174.1\ \mu\mathrm{m}^2 \quad (5\text{-}5)$$

Then, the total capacitance at the anode of the SPAD is calculated to be 270 fF.

$$C_{Total} = C_{p+/nw} \times A_{p+/nw} + C_{pw/nw} \times A_{pw/nw}$$
$$= (1.2 \times 10^{-3}) \times (101.3 \times 10^{-12}) + (0.85 \times 10^{-3}) \times (174.1 \times 10^{-12}) \quad (5\text{-}6)$$
$$= 121.6 + 148 = 270 \text{ fF}$$

5.3. Active Quench and Active Reset Circuit Design

5.3.1. Operation Principle

The schematic of the proposed AQR circuit is shown in Figure 5-7. In general, this AQR circuit can be divided into 4 parts:

- AQR core circuit to perform quench and reset for the SPAD,
- a voltage-controlled delay circuit to adjust the hold-off time,
- a pulse generation circuit to generate reset pulses,
- and a monostable to widen the output pulse so it can be measured by the oscilloscope probes with large input capacitances.

A sample of the post-layout simulation results when the delay control voltage (V_C) is set to 0 V is presented in Figure 5-8. The operation principle is described based on the simplified schematic together with post-layout simulated waveforms. Initially, node *A* is at 0 V in a high-impedance state since the transistors M_2 and M_3 are off. The SPAD is biased above breakdown by the large positive voltage V_{HV} applied to the cathode. When either a photon or dark carrier initiates a self-sustaining avalanche in the SPAD, the voltage at node *A* will begin to increase towards (V_{HV} - V_{BR}) by a passive quench process through transistor M_3 which provides a large resistance in the off state. The skewed inverter formed by transistors M_4 and M_5 is designed to have a low switching threshold such that the avalanche is detected early, causing the voltage at node *B* to fall to 0 V. As the inverter formed by M_4 and M_5 is cross coupled to transistor M_2, then M_2 turns on and creates a positive feedback path that will quickly charge node *A* through M_1 and M_2. This reduces the charge flowing through the SPAD during an avalanche, which in turn reduces the carriers trapped in SPAD's defects which are later released to cause AP.

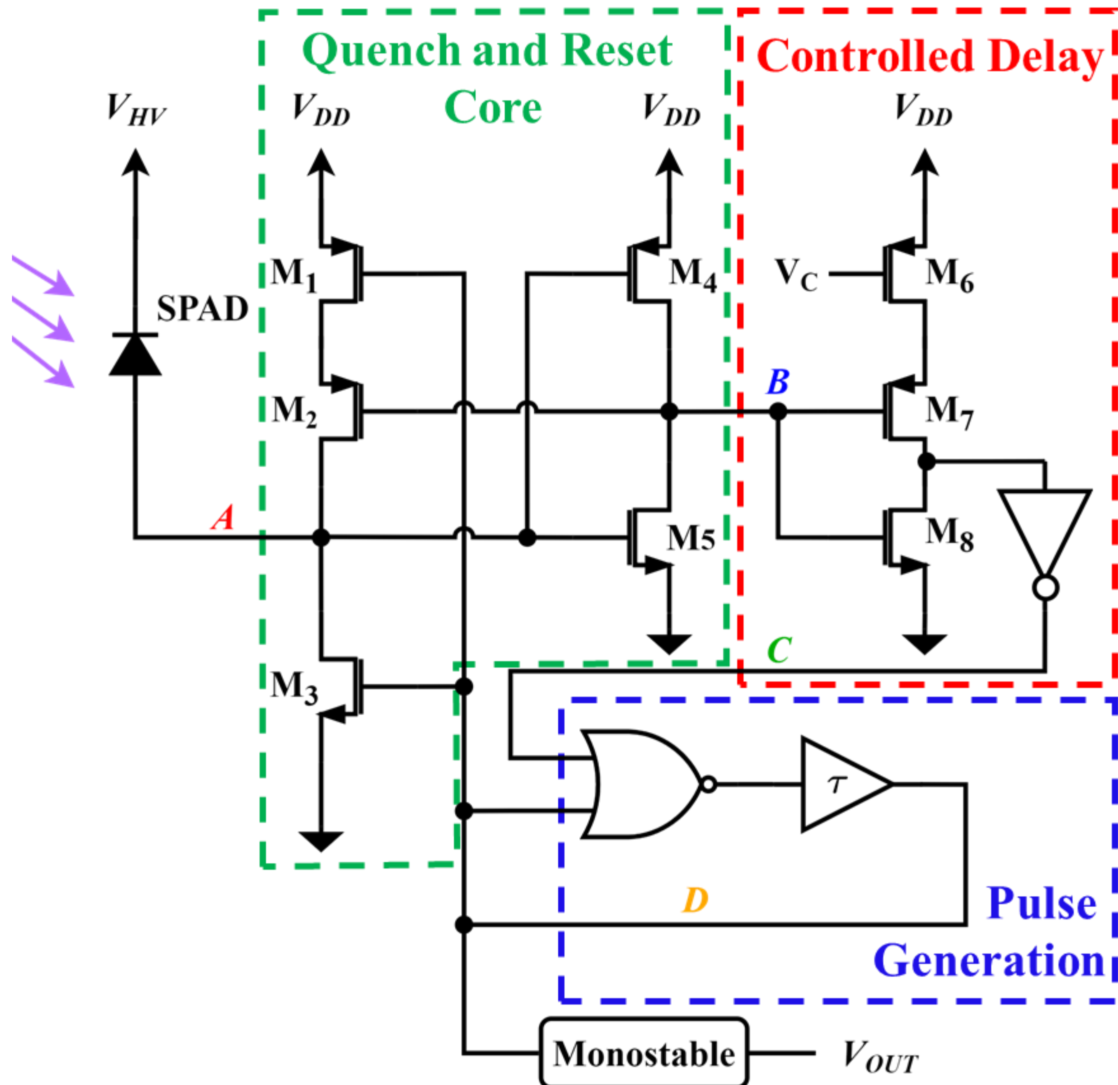


Figure 5-7: Schematic of the active quench and active reset (AQR) SPAD pixel.

Following the quench process, transistors $M_6 – M_8$ and the following inverter form a voltage-controlled delay buffer from the nodes *B* to *C*. The voltage V_C (at the gate of M_6) limits the switching current of the inverter formed by M_7 and M_8 during its rising transition so as to increase the delay from nodes *B* to *C* through current starving. This delay controls the hold-off time of the pixel, which can be increased to allow trapped charges to be released while the SPAD's is biased below breakdown to limit the AP. When the falling edge of node *C* occurs, the following NOR gate and a 1.5 ns fixed-delay buffer are used to generate the reset pulse at node *D*. This reset pulse will turn off M_1 and turn on M_3 in order to reset node *A* to 0 V and restore the SPAD excess bias. This short reset pulse would

additionally be used as the SPADs output pulse if the SPAD did not need to be directly connected off-chip, resulting in a short DT for the SPAD pixel.

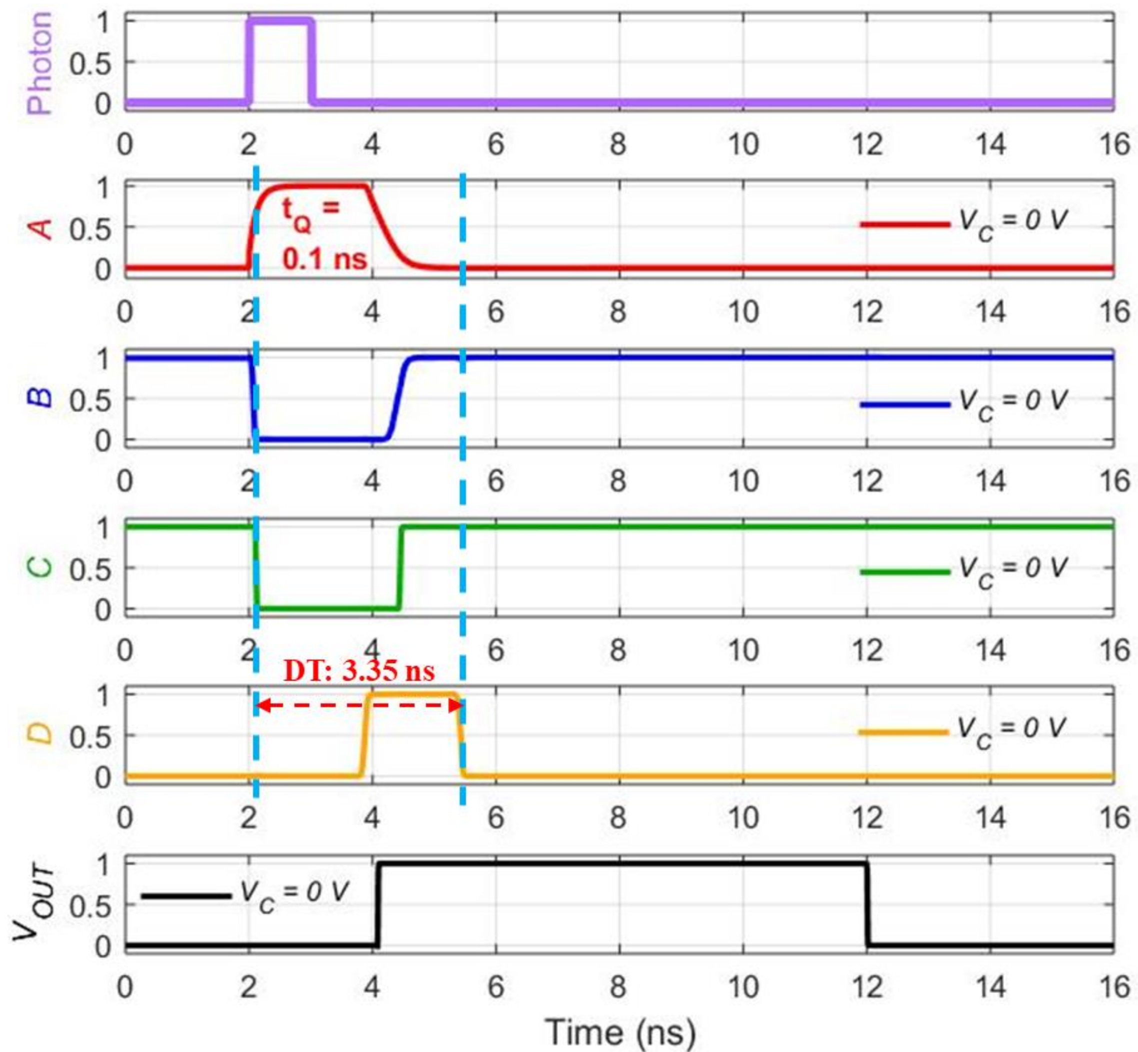


Figure 5-8: Post-layout transient simulation results for the active quench and active reset (AQR) SPAD pixel with V_C set to 0 V. The quenching time (10% - 90% of the rising edge at node A) is 0.1 ns, and the minimum dead time (DT) is ~3.35 ns.

The measurement setup to test a single SPAD pixel required the SPAD output to drive the relatively large input capacitance of an oscilloscope probe, so a monostable was used to widen the pulse widths that are sent off chip. Without the monostable, the DT of the AQR circuit can be measured from the 50% rising edge of the voltage at node A to the 50% falling edge of the reset signal voltage at node D. The minimum DT of the SPAD pixel from the post-layout simulation with V_C set to 0 V is 3.35 ns, corresponding to a maximum

CR of ~ 300 MHz.

In the next sub-section, the details of the design and the function of the voltage-controlled delay circuit, pulse generation circuit and monostable will be discussed.

5.3.2. Hold-off Time Controlling Circuit

Figure 5-9 shows the schematic of the hold-off time controlling circuit, which consists of a current-starved inverter and a standard inverter. The switching current for the rising transition of the inverter formed by M_7 and M_8 is controlled by the voltage (V_C) applied on the gate of M_6. When V_C is set to 0 V, M_6 is fully on, and thus presents a very low channel resistance. Therefore, the delay is approximately the delay induced by two standard inverters, which is only few picoseconds (~ 10 ps). As the controlling voltage V_C increases, the rising time of the voltage-controlled inverter will become longer due to the limited current available for switching in the rising transition. Therefore, the time for the output voltage of the voltage-controlled inverter to reach the threshold of the standard inverter is delayed, giving a delayed falling edge for the output. Figures 5-10 and 5-11 show the post-layout simulation examples when V_C is set to 0 V and 0.7 V, in which the delay (50% - 50% between the falling edges of the input and output pulses) is 1.6 ns with V_C at 0.7 V. The delay as a function of the control voltage is shown in Figure 5-12. The delay presents an approximately exponential relationship with the control voltage, which indicates that the control voltage needs to be precisely adjusted when a long delay time is required.

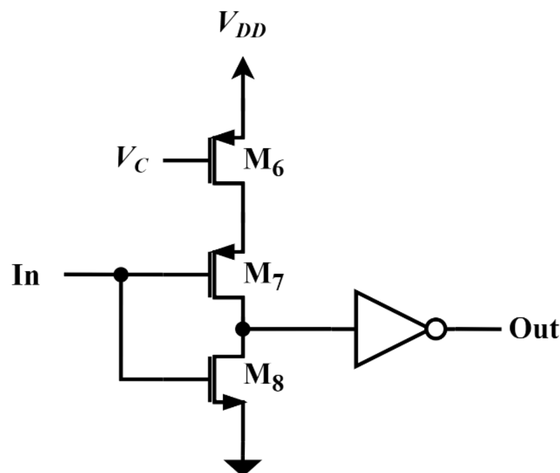


Figure 5-9: Schematic of the hold-off time controlling circuit.

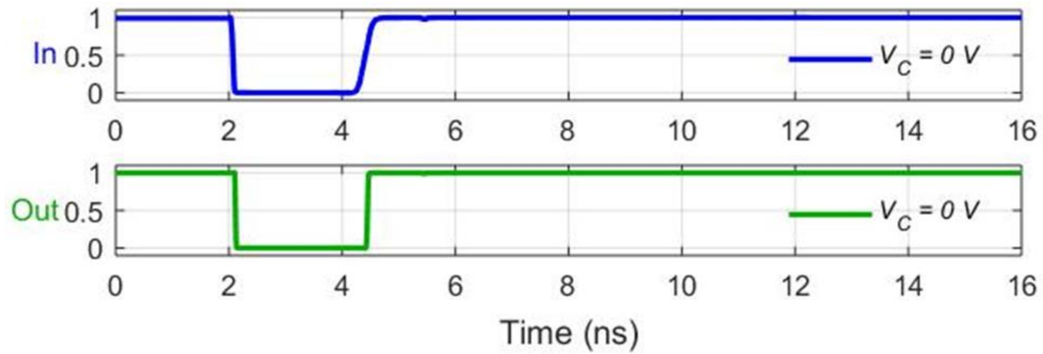


Figure 5-10: Simulated waveforms of the hold-off time controlling circuit when the control voltage is set to 0 V.

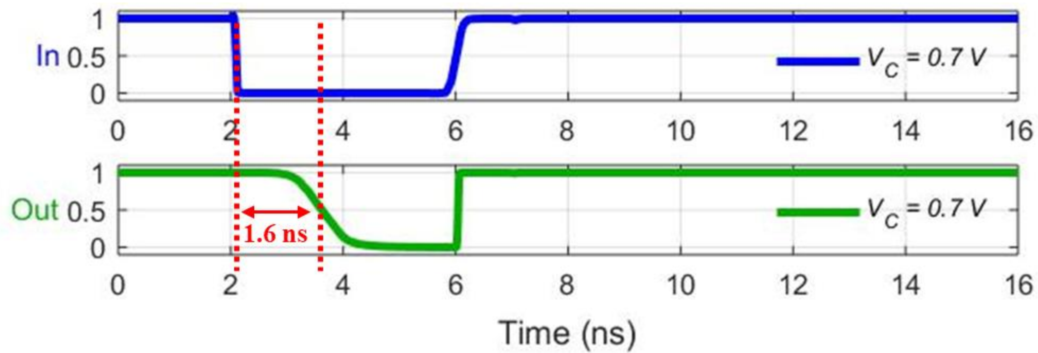
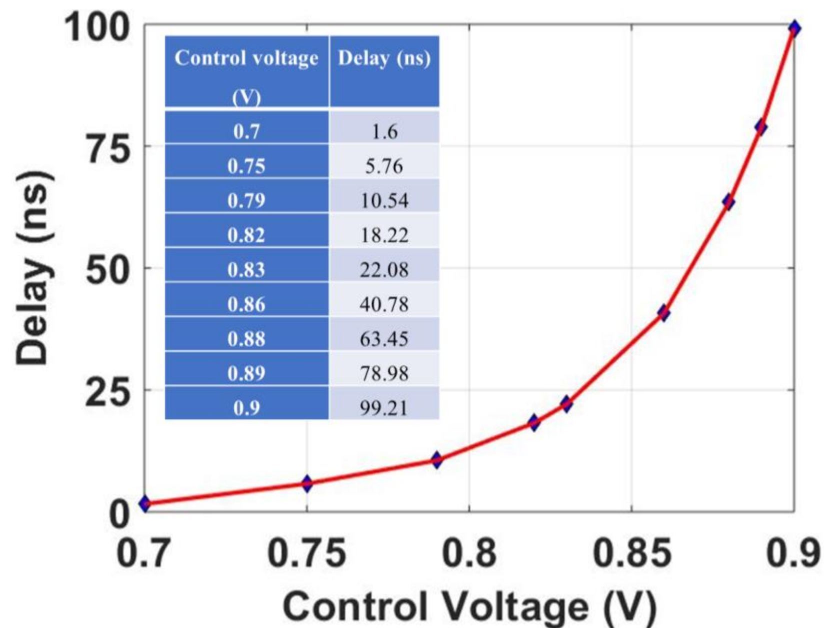


Figure 5-11: Simulated waveforms of the hold-off time controlling circuit when the control voltage is set to 0.7 V.



Control voltage (V)	Delay (ns)
0.7	1.6
0.75	5.76
0.79	10.54
0.82	18.22
0.83	22.08
0.86	40.78
0.88	63.45
0.89	78.98
0.9	99.21

Figure 5-12: Delay as a function of the control voltage.

5.3.3. Pulse Generation Circuit

The pulse generation circuit consists of one NOR gate and a fixed 1.5 ns delay buffer, as shown in Figure 5-13. The 1.5 ns delay buffer has two functions: first, it will delay the input signal for 1.5 ns which is the minimum hold-off time; second, it will make the pulse width of output signal to be 1.5 ns so that the SPAD can be fully reset to its initial condition for detection of the next photon. Figure 5-14 shows the simulated waveforms.

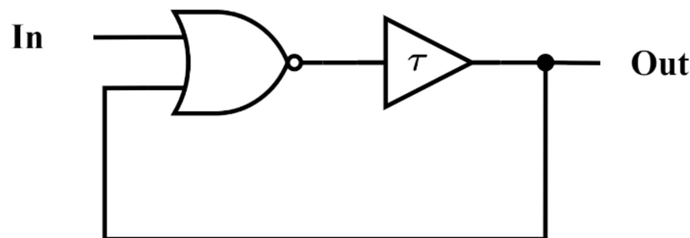


Figure 5-13: Schematic of the pulse generation circuit.

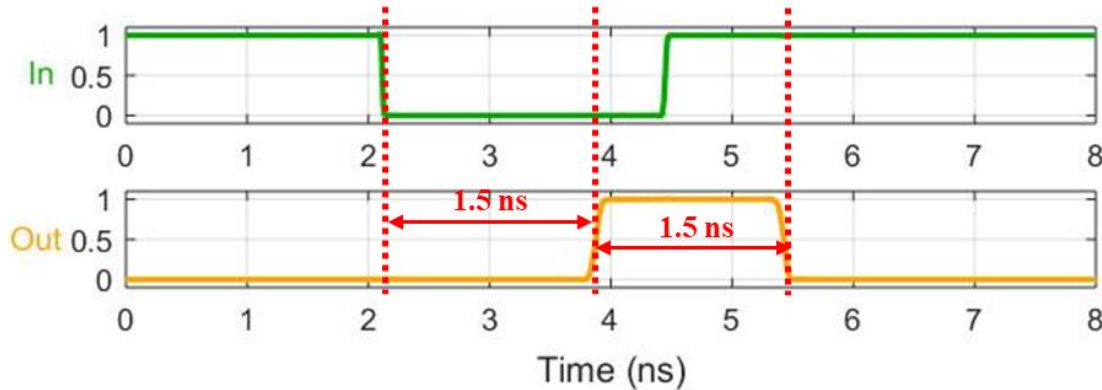


Figure 5-14: Simulated waveforms of the pulse generation circuit.

5.3.4. Monostable Circuit

The monostable shown in Figure 5-15 is placed at the end of the signal path to widen the output pulse width so that it can be measured by the oscilloscope probes with large input capacitances. The monostable consists of a NOR gate, an AC-coupling capacitor (200 fF), a pull-up resistor (70 kΩ) and an inverter. The output is sent back to the second terminal of the NOR gate. Note that this monostable circuit is not needed if the output of the AQR circuit does not need to be sent off-chip directly and can be processed by internal circuits

on the same chip. As shown in Figure 5-16, with a large pull-up resistor and an AC-coupling capacitor, this monostable generates an output pulse with the width of 8 ns.

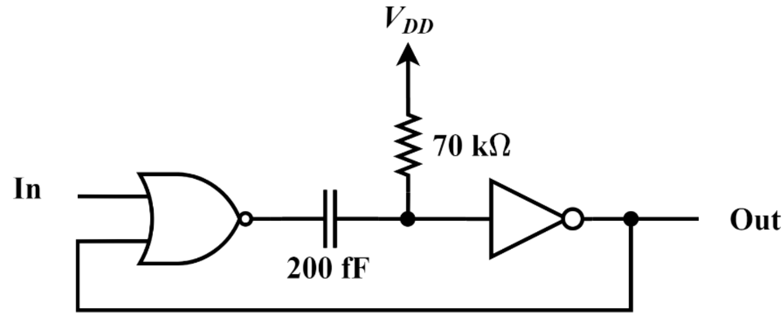


Figure 5-15: Schematic of the monostable.

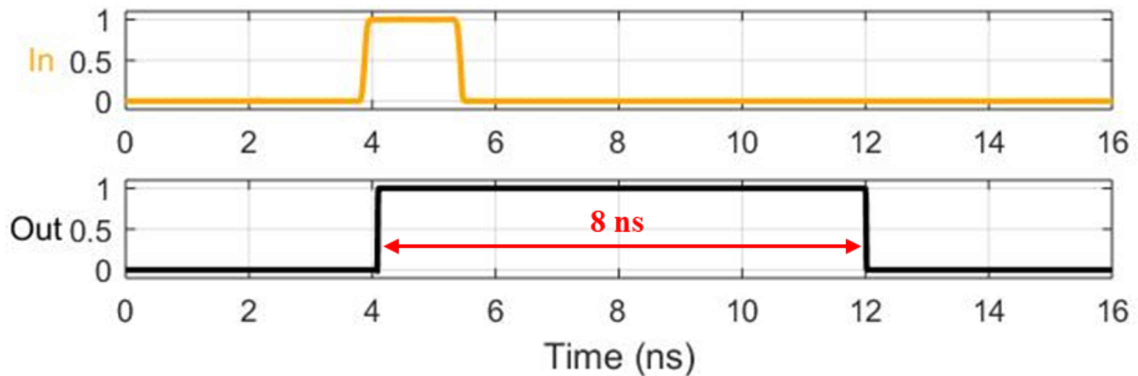


Figure 5-16: Simulated waveforms of the monostable.

5.3.5. Layout

The layout of the AQR SPAD pixel including the monostable and output buffers is shown in Figure 5-17. The size of the SPAD is 20.2 µm × 20.2 µm with a 90 µm^2 active area, resulting in an FF of 22% for the SPAD itself. The lower FF of the SPAD results from the use of a wide p-well guard ring to avoid PEB and could be improved in a further optimized implementation. The size of the AQR front-end circuitry is only 2.1 µm × 10.4 µm. In a fully integrated system, the output buffer and monostable would not be required and the SPAD output would only need to drive low capacitive loads such as counters or time-to-digital converters on chip. As such, the resulting pixel pitch would be 20.2 µm × 23.2 µm for an achievable FF of 19%.

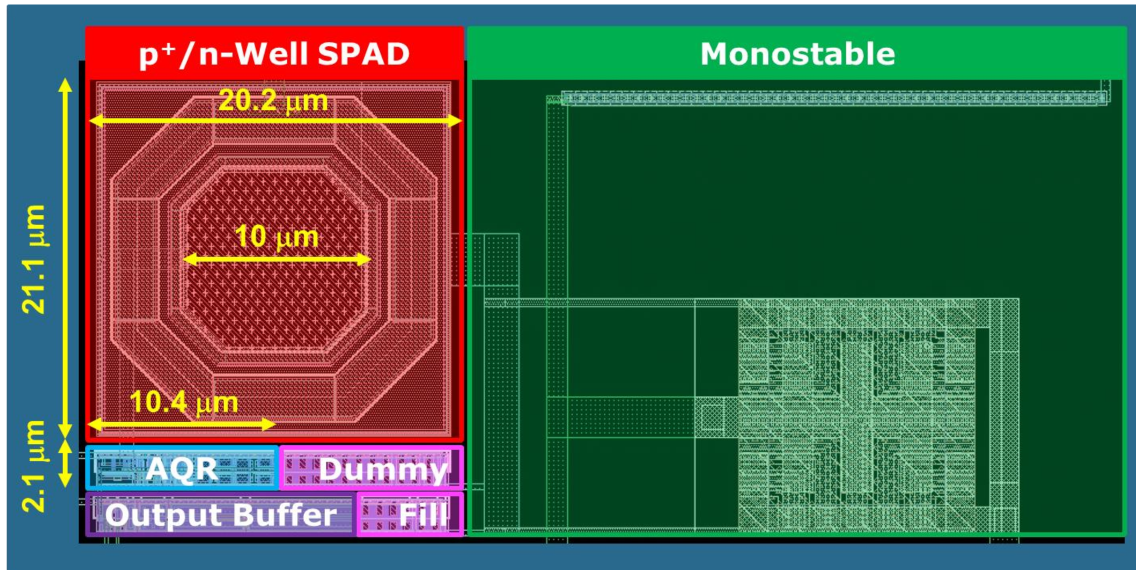


Figure 5-17: Layout of SPAD pixel in a standard 65 nm CMOS process, including the output buffers and monostable for driving a large off-chip capacitive load for measurements.

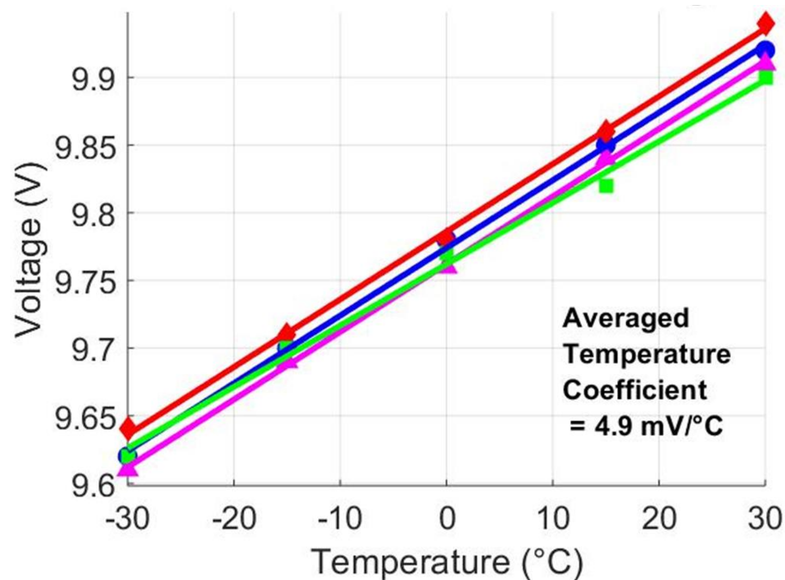


Figure 5-18: Breakdown voltage as a function of temperature for 4 SPAD samples. The averaged temperature coefficient is 4.9 mV/°C.

5.4. SPAD Pixel Characterizations and Discussions

5.4.1. Breakdown Voltage

The breakdown voltage (V_{BR}) was measured were for 4 different SPADs with the same structure as in the pixel design. Figure 5-18 shows extracted V_{BR} from 4 SPAD samples as

a function of temperature. The average temperature coefficient for V_{BR} is 4.9 mV/°C. The average of the V_{BR} for the 4 samples is 9.9 V at room temperature (25 °C).

5.4.2. Dark Count Rate and Afterpulsing

The inter-arrival times (IATs) of the pixel output were recorded in the dark with a digital oscilloscope (LeCroy WaveRunner 625Zi). The dark count rate (DCR) and AP can be derived from the histogram of the IAT using exponential fitting. The measurement setup is shown in Figure 5-19. One example of the exponential fit on the histogram of the IAT is also presented. Figure 5-20 shows the DCR as a function of V_{ex} at room temperature. DCR increases with V_{ex} because enhanced electric field in the depletion region increases the avalanching probability of the dark carriers. It is known that the AP increases as temperature decreases. That is because the lifetime for the carriers released from the trap

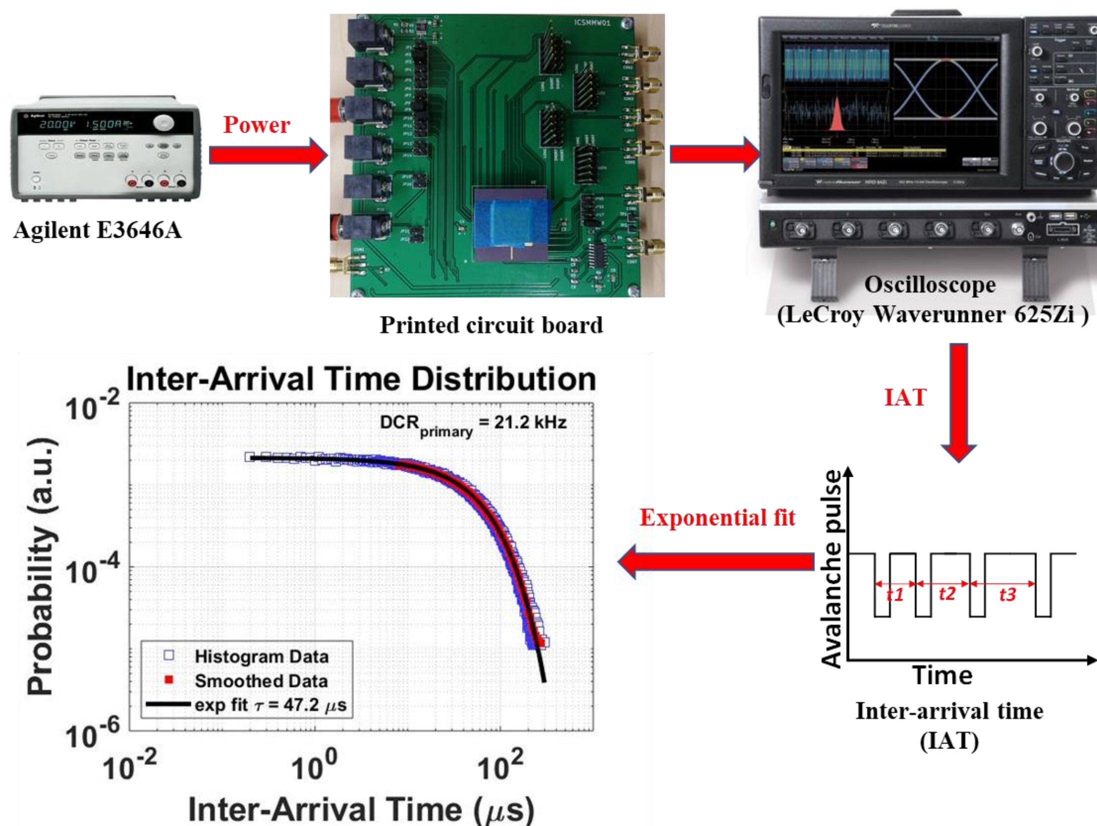


Figure 5-19: Dark Count Rate (DCR) measurement setup.

centers increases exponentially when temperature decreases, thus enhancing the AP probability. However, negligible AP was observed on our SPAD pixel even at a low temperature of -35 °C. This can be explained by the reduced avalanching charges due to the short QT (0.1 ns) as shown in Fig. 5-8. Additionally, a monostable is added at the end of the signal path to widen the output pulse width so it can be measured by the oscilloscope. The measured width of the output pulse is 11.5 ns, which indicates that only pulses with the IAT larger than 11.5 ns can be recorded. This could be another reason for no AP being measured in the SPAD pixel.

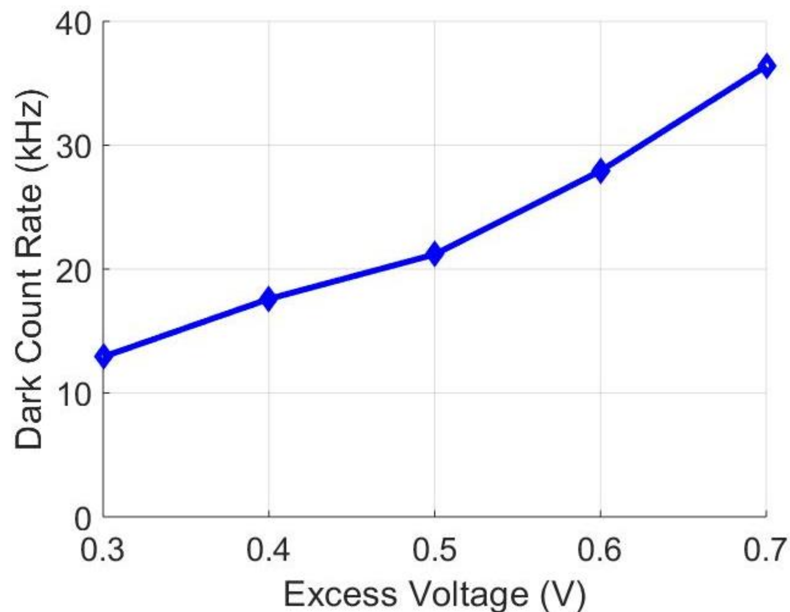


Figure 5-20: Dark Count Rate (DCR) as a function of excess voltage.

5.4.3. Photon Detection Probability

Figure 5-21 shows the measurement setup for the photon detection probability (PDP) of the SPAD pixel. A Xenon lamp provided a continuous light source for the PDP measurement. Optical bandpass filters are used to select the wavelength while neutral density filters together with the adjustment of lamp current are used to control the intensity of light to avoid saturation. The PDP is defined as the ratio of the photon count rate detected by the SPAD to the incident photon rate. The photon count rate was measured using the LeCroy digital oscilloscope and the incident photon rate was calculated from the optical

power measured by optical power meter (Newport 1830-C) with a calibrated silicon photodetector (Newport 818-SL).

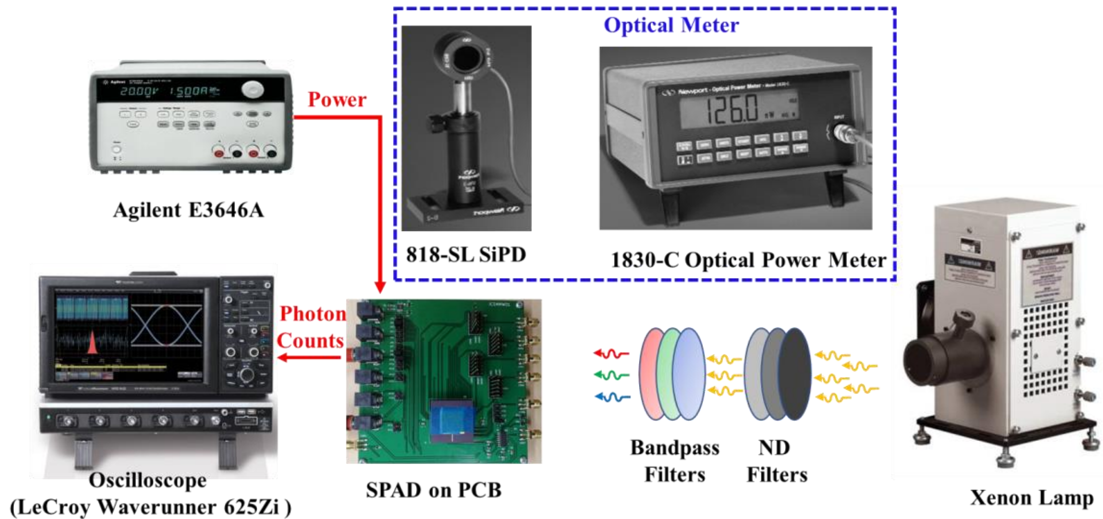


Figure 5-21: Measurement setup for the photon detection probability (PDP).

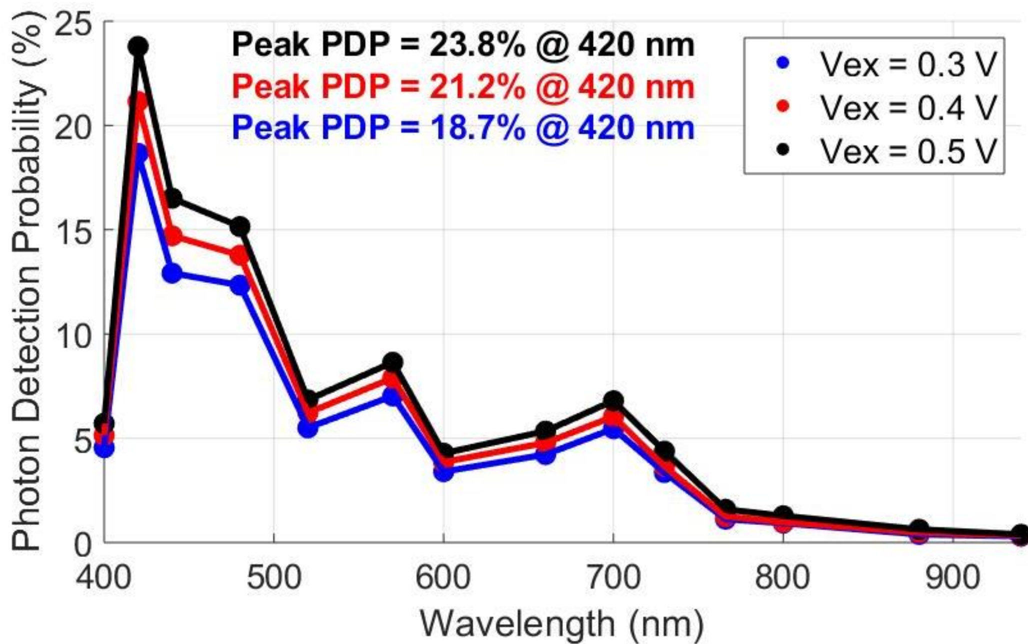


Figure 5-22: Photon Detection Probability (PDP) as a function of wavelength when the excess voltage is biased at different voltages (0.3V, 0.4V and 0.5 V).

Figure 5-22 shows the PDP as a function of wavelength when V_{ex} is 0.3 V, 0.4V and 0.5 V. It can be seen that the PDP increases with the excess voltage. This is because the

increased V_{ex} enhances the electric field in the depletion region of the SPAD, thus leading to a higher avalanche triggering probability. The measurement shows that the peak PDP is 18.7%, 21.2% and 23.8% at the wavelength of 420 nm with V_{ex} of 0.3 V, 0.4 V and 0.5 V, respectively. The peak response at a short wavelength of 420 nm is because of the shallow depth of p^+/n-well junction in this advanced 65 nm CMOS technology node. The fluctuation of PDP at the longer wavelengths can be explained by the Fabry-Perot resonator formed by the stack of passivation layer and inter-metal dielectric layers in a standard CMOS process [170].

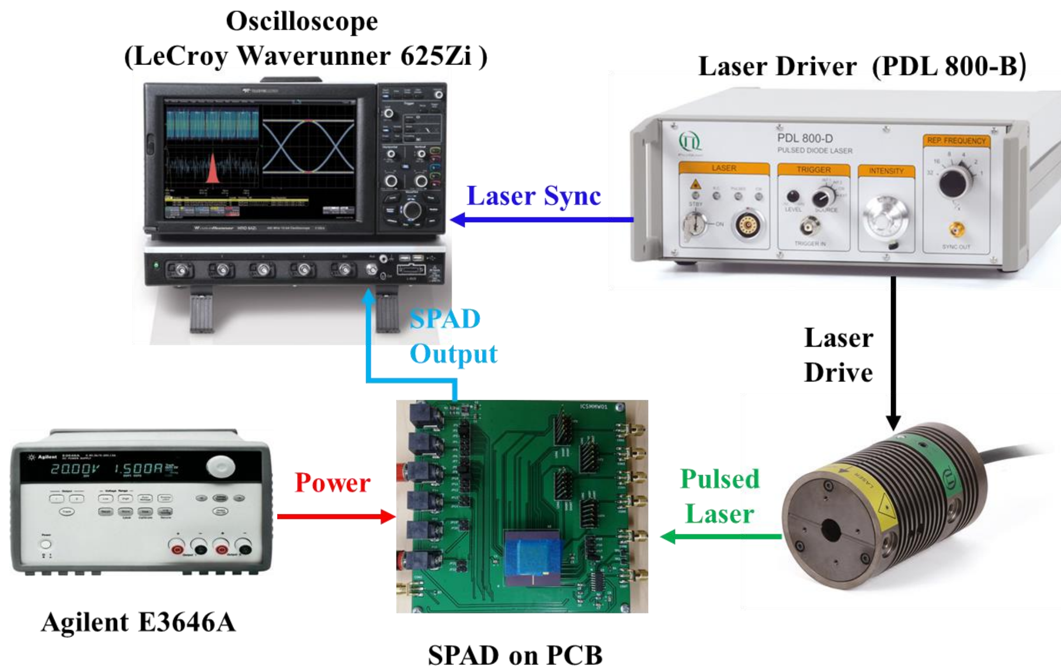


Figure 5-23: Timing jitter measurement setup.

5.4.4. Timing Jitter

The timing jitter was measured using pulsed lasers with a high repetition rate. Figure 5-23 shows the measurement setup of the timing jitter. The time difference between the observed SPAD pixel output pulse and the synchronized output of a PDL 800-B laser driver was recorded by the LeCroy oscilloscope. The full-width at half-maximum (FWHM) of the time difference histogram is reported as the timing jitter. Two laser heads LDH-P-C-690 (685 nm, 70 ps FWHM) and LDH-P-C-405 (405 nm, 50 ps FWHM) were used for the

timing jitter measurement. The system timing jitters are 148 ps for the 405 nm laser and 172 ps for the 685 nm laser when the V_{ex} is 0.5 V, as shown in Figure 5-24. Considering the pulse width of the lasers, the timing jitters for the SPAD pixel are estimated to be 139 ps ($\sqrt{148^2 - 50^2}$) for the 405 nm laser and 157 ps ($\sqrt{172^2 - 70^2}$) for the 685 nm laser. As shown in Fig. 5-24, the diffusion tail for 685 nm laser is more apparent than the 405 nm laser. This is because the p^+/n-well SPAD is a shallow junction that has a peak PDP at 420 nm as shown in Fig. 5-22. This means that photons with a shorter wavelength (405 nm) are more likely to be absorbed inside the depletion region when compared to the photons with a longer wavelength (685 nm), thus showing less diffusion tail in the histogram of the timing jitter.

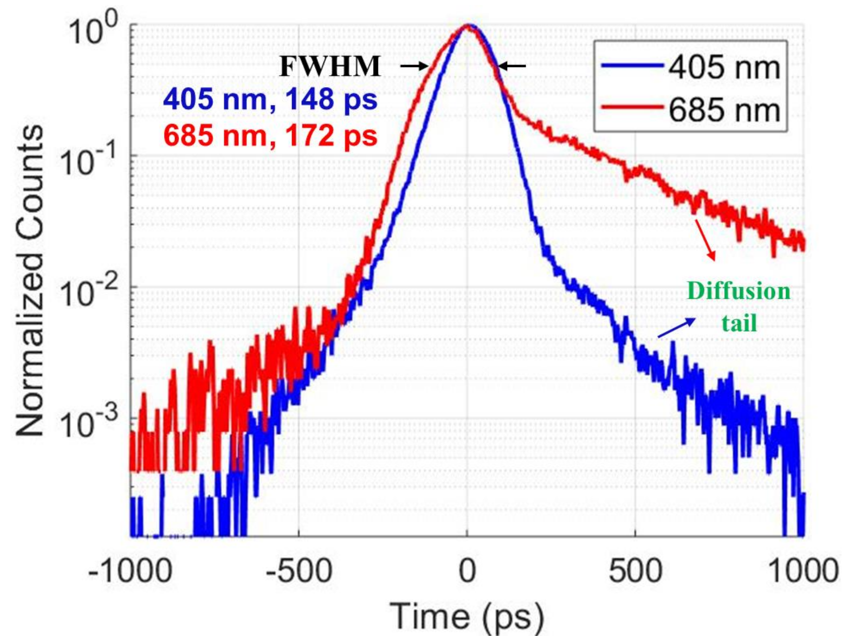


Figure 5-24: Timing jitters at 0.5 V excess voltages using 405 nm and 685 nm pulsed lasers.

5.4.5. Summary and Comparison

Table 5-1 shows the summary of our work and a comparison to other SPAD pixels with AQR circuits. Note that the technique of using the positive feedback loop to reduce the quenching time was reported in other AQR circuit designs [128], [168], [169]. However, our pixel was designed in a low-cost standard CMOS process, and it achieved

the shortest quenching time and highest count rate. Additionally, the advanced technology node allows the AQC circuit to occupy the smallest area of 21.8 μm^2.

5.5. Conclusions

In this work, a SPAD pixel consisting of a p^+/n-well SPAD and a compact and high-speed active quench and reset (AQR) circuit is designed and fabricated using a standard 65 nm CMOS process. The positive feedback loop in the AQR circuit resulted in the quenching time being reduced to 0.1 ns. Due to the fast response of the transistors in this advanced 65 nm technology node, the minimum dead time (DT) of this pixel is 3.35 ns. This DT corresponds a count rate of ~300 Mcps. The SPAD pixel is measured to have a dark count rate of 21 kHz, a peak photon detection probability of 23.8% at 420 nm wavelength and a timing jitter of 139 ps using a 405 nm pulsed laser when it is biased at a 0.5 V excess voltage. Due to the short quenching time, negligible afterpulsing is observed even at a low temperature of -35 °C.

Table 5-1: Comparison of our work to other SPAD pixels with active quench and reset circuits

Year, Ref.	Tech. (nm), Junction	SPAD Dia. (μm)	V_{BD} V_{ex} (V)	Quench Time (ns)	Dead Time (ns)	Max. CR (Mcps)	Peak PDP (%) @ λ (nm)	DCR (cps/ μm^2)	Jitter (ps) @ λ (nm)	AP (%) @ hold-off (ns)	Circuit area (μm^2)	Simple FOM $
2013 [128]	350, p^+/nw	20	25 6	1	20	50	-	25	-	1.3 @ 20	672	0.074
2018, [168]	350, n^+/pw	80	21.7 6.6	0.48 #	9.5-17 #	105 #	35.1 @ 635 22 @ 850	6.1	-	4.8 @ 9.5 0.9 @ 17	17420	0.013
2019 [171]	180 HV, Customized	50	36.1 5	4	>6.2	160	-	-	-	5 @ 6.2 1.4 @ 8	32400	0.0012
2019 [172]	180 Stand., p^+/nw	8	9.98 2	-	20-30	50	30.4 @ 530 3.8 @ 530	10.3k	92 @ 905	1 @ 30	400 †	-
2020 [169]	180 Stand., pw/DNW	10	15.5 3.5	<0.7 #	>5 #	200 #	34 @ 450 2.5 @ 850	20.5	-	0.75 @ 4	306	0.93
2017 [82]	65 BSI CIS/ 40nm ## pw/DNW	15.8 *	12 4.4	-	>8	125	29.5 @ 660	4 ††	75 @ 700	0.08 @ 8	25.5	-
2018 [167]	65 Stand., p^+/nw	20	9.9 1.5	-	4-110	250	8 @ 470 **	2.8k	7.8	10 @ 100	540	-
This work	65 Stand., p^+/nw	10	9.9 0.5	0.1 #	>3.35 #	300 #	23.8 @ 420	233	139 @ 405 157 @ 685	~ 0	21.8	137.6

$ A simple figure-of-merit for a quench and reset circuit is defined as：Count rate [Mcps] / (Quenching time [ns] $\times$ Circuit area [μm^2]).
†Estimated from the microphoto.
#Simulation results.
##Back side illuminated 3D IC: SPAD in a 65 nm image sensing process; circuits in a 40 nm CMOS.
*Estimated from the total SPAD area.
†† Read from the curve (maximum value).
**PDE value was reported but FF was not provided.

Chapter 6
Differential Quench and Reset Circuit for SPADs

6.1. Introduction

Unlike the traditional avalanche photodiodes (APD), SPADs are biased above their breakdown voltage, which is called Geiger mode operation. The voltage difference between the biasing voltage and breakdown voltage is defined as the excess voltage (V_{ex}). In Geiger mode, the electric field in the depletion region of a SPAD is so high that it enables the SPAD to maintain a self-sustaining avalanche process if the SPAD is triggered by even a single carrier. Therefore, a quench and reset (QR) circuit is required to enable the normal operation of a SPAD. The time between the initiation and the end of the avalanche of a SPAD is defined as the "quenching time" while the time needed to reset a SPAD to its initial Geiger mode after quench is called the "reset time". The sum of the "quenching time" and "reset time" is the whole time needed for the detection of a photon, which is also called the "dead time" (DT) of a SPAD. In general, the QR circuit can be categorized into two groups: passive quench-reset (PQR) circuit and active quench-reset (AQR) circuit [131].

PQR circuits generally consist of a large resistor (usually > 50 kΩ) connected in series to the anode or cathode of the SPAD [110]. Once the avalanche is initiated, the rapid discharging current causes the internal resistance of SPAD decreases to a small value (several hundred Ohms), thus lowering the voltage across the SPAD. When the voltage across the SPAD drops below the breakdown voltage and the current flowing through the SPAD is less than a threshold (~ 100 microamperes), the avalanching process cannot be

----

Adapted from W. Jiang, R. Scott, and M. J. Deen, "Differential Quench and Reset Circuit for Single-Photon Avalanche Diodes," *Journal of Lightwave Technology*, 2021, doi: 10.1109/JLT.2021.3111119 (Appendix A)

self-sustaining, and is quenched [109]. After quench, the voltages on the terminals of the SPAD are charged back to their initial biasing voltages by a small current which flows through the large quench resistor. This leads to a large RC time constant during the reset, with typical "reset time" in the range of microseconds. Even though the PQR circuits are easily implemented and occupy only a small area, the long "reset time" means a long DT, resulting in a low count rate (CR) for the SPAD, which limits their performance for high-speed applications. To overcome these disadvantages, AQR circuits were proposed and developed in many SPAD designs.

In an AQR circuit, a comparator is usually used to detect the avalanche of the SPAD to generate the output pulse. In order to perform the active quench and reset, the output pulse of the SPAD will be delayed by preset times before producing controlled quench and reset pulses. The activation of the quench control signal will reduce the voltage across the SPAD below the breakdown voltage while activation of the reset control signal will bring the biasing voltage of the SPAD back to its initial biasing condition to detect the next photon [31]. In general, the voltage across the SPAD will remain below the breakdown voltage for a preset time (hold-off time) before the reset pulse is enabled after quench. During the hold-off time, the carriers (photon-generated or from dark noise) in the space charge region of the SPAD will not generate secondary pulses because the biasing voltage is lower than the breakdown voltage. Therefore, a longer hold-off time in AQR circuits results in a lower afterpulsing (AP), but a longer DT leads to a reduced CR [128]. When comparing to the DT of PQR circuit, the DT in AQR circuit is much smaller since a typical hold-off time in AQR circuit can be in nanoseconds.

At present, both PQR circuits and AQR circuits are widely used in many SPAD applications. However, the traditional PQR circuits and AQR circuits only use one QR circuit connected to either the cathode or anode of the SPAD as shown in Figures 6-1(a) and (b). Therefore, the output of SPADs with existing PQR and AQR circuits is a single-ended signal that is easily influenced by common-mode (CM) noise or interferences. In this work, we propose a differential quench and reset (DQR) circuit for the SPAD, which consists of a pair of QR circuits connected to both the anode and the cathode of the SPAD,

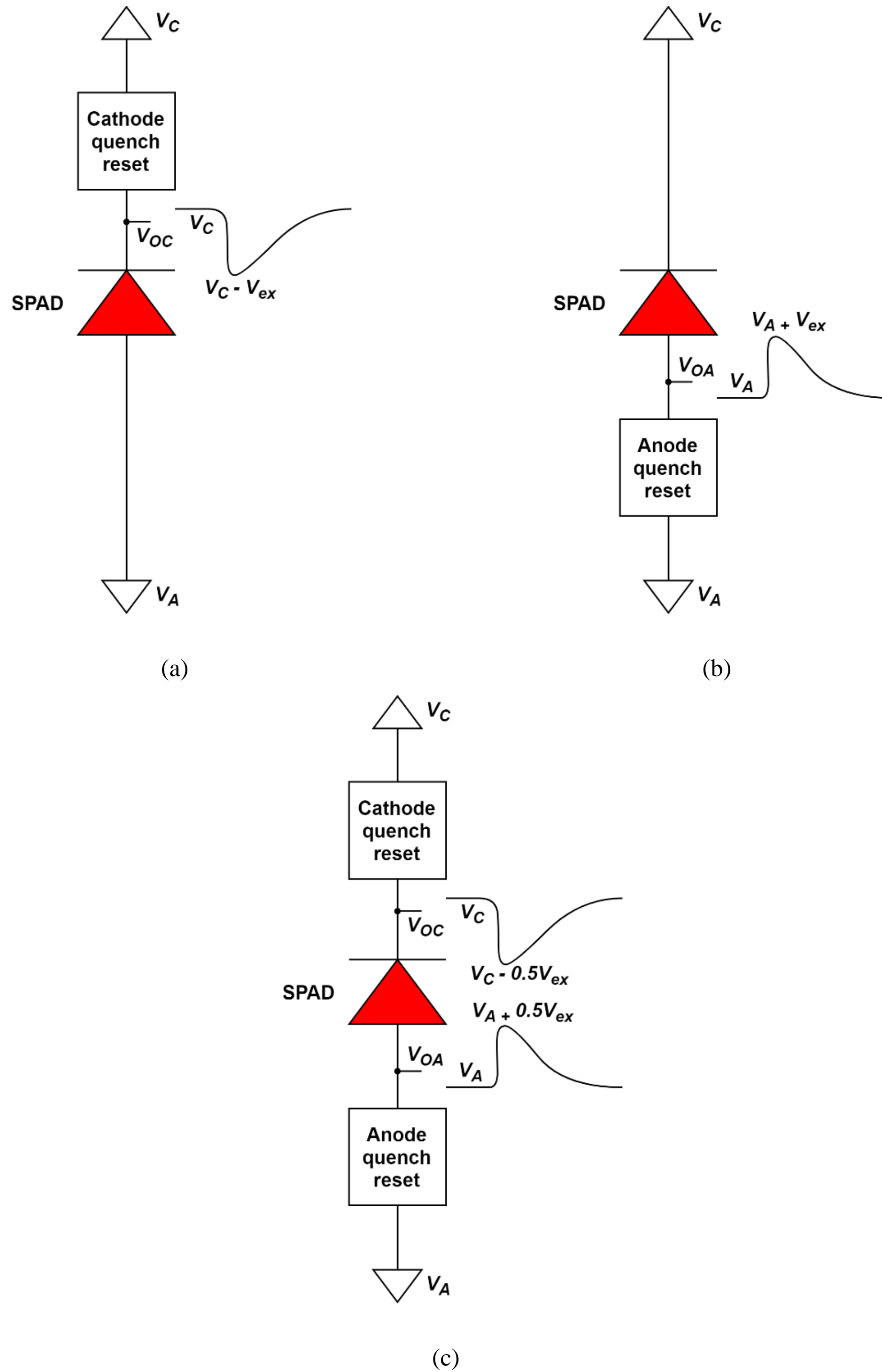


Figure 6-1: SPAD with a quench and reset circuit at the (a) cathode and (b) anode. (c) SPAD with a differential quench and reset circuit that contains two quench and reset circuits on both the cathode and anode.

as shown in Figure 6-1(c). When an avalanche occurs in the DQR configuration, the voltage of the SPAD's cathode (V_{OC}) will drop from the biasing voltage of the cathode (V_C) while the voltage of the SPAD's anode (V_{OA}) will increase from the biasing voltage of the anode (V_A). That is, the SPAD can be quenched through both terminals. The avalanching process is fully quenched when the voltage across the SPAD reaches its breakdown voltage (V_{OC} - $V_{OA} = V_{BR}$). Like the traditional single QR circuit, the DQR can be implemented in either a passive or an active configuration.

6.2. Operational Principle and Simulations

6.2.1. Differential Passive Quench and Reset

Figure 6-2(a) shows a simplified equivalent electrical model for a SPAD. In this model, C_{SPAD} is the SPAD's junction capacitance, C_{CS} is the parasitic capacitance from the cathode to the substrate, C_{AS} is the parasitic capacitance from the anode to the substrate, R_{SPAD} is a series avalanche resistance, a DC voltage source represents the breakdown voltage V_{BR} and a switch S_W is used to indicate the state of the avalanche. The operational principle of DQR circuit is described using this electrical model.

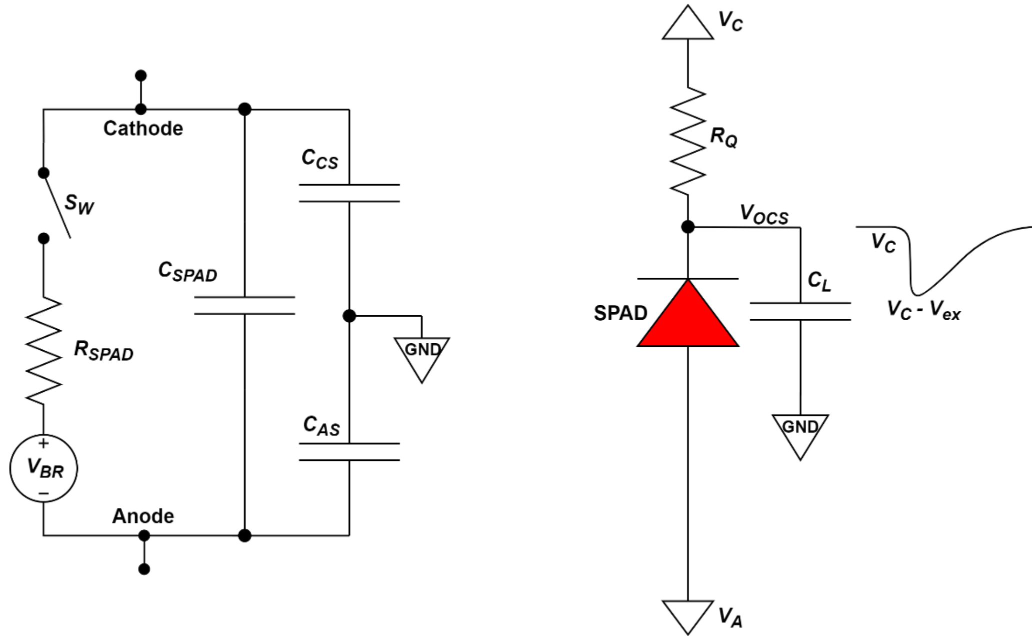
Figure 6-2(b) shows a traditional single passive quench and reset (SPQR) circuit using a quench resistor R_Q connected to the SPAD's cathode. When an avalanche is triggered, the switch S_W closes to indicate the initiation of the avalanche. Then, the SPAD's capacitor C_{SPAD} discharges through the resistor R_{SPAD}. The discharge time constant τ_{QS} of the SPQR circuit is given by:

$$\tau_{QS} = \left(R_Q \parallel R_{SPAD}\right)(C_{SPAD} + C_{CS} + C_L)\ . \tag{6-1}$$

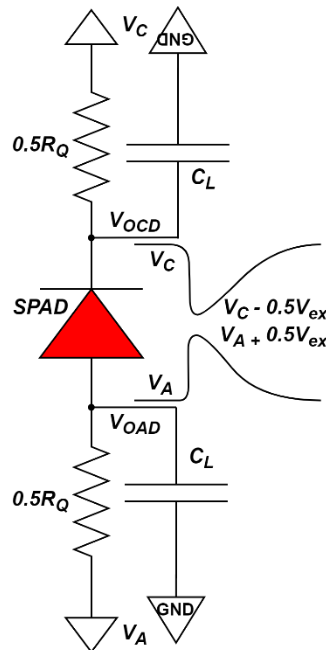
Since $R_Q \gg R_{SPAD}$, the quenching time constant τ_{QS} can be simplified as:

$$\tau_{QS} \approx R_{SPAD}C_{SPAD} + R_{SPAD}(C_{CS} + C_L)\ . \tag{6-2}$$

During the quenching time, the output voltage at the SPAD's cathode drops from V_C to a voltage ($V_C - V_{ex}$) to stop the avalanche. Therefore, the amplitude of the output pulse is V_{ex}. The amount of charge (Q_{QS}) flowing through the SPAD's junction during the avalanching process is expressed as:



(a) (b)



(c)

Figure 6-2: (a) Equivalent circuit model for the SPAD. (b) Traditional single passive quench and reset circuit with a quench resistor connected to the SPAD's cathode. (c) Differential passive quench and reset circuit consisting of two resistors connected to cathode and anode of the SPAD.

$$Q_{QS} = V_{ex}C_{SPAD} + V_{ex}(C_{CS} + C_L)\,. \tag{6-3}$$

After quenching, the cathode will be charged back to V_C by the current flowing through the cathode quench resistor R_Q. The reset time constant is given by:

$$\tau_{RS} = R_Q C_{SPAD} + R_Q(C_{CS} + C_L)\,. \tag{6-4}$$

Because $R_Q \gg R_{SPAD}$, τ_{RS} is much larger than τ_{QS}. Therefore, the dead time of SPQR is dominated by its reset time.

Figure 6-2(c) shows a differential passive quench and reset (DPQR) circuit with two quench resistors R_{QC} and R_{QA} connected to the cathode and anode, respectively. The parasitic capacitance C_{CS} and C_{AS} are assumed to be equal to simplify the analysis. In addition, the cathode and anode are assumed to have the same load capacitance (C_L). Therefore, the total capacitance from the cathode to the ground $(C_{CS} + C_L)$ is equal to the total capacitance from the anode to the ground $(C_{AS} + C_L)$. In this configuration, when an avalanche occurs, the voltage at the cathode decreases from V_C while the voltage at the anode increases from V_A. Like the traditional SPQR circuit, the avalanche is quenched when the voltage across the SPAD is reduced below V_{BR}. Therefore, the sum of the voltage changes on the cathode and anode is equal to the excess voltage (V_{ex}). That is, the sum of the voltage drops on the cathode quench resistor (R_{QC}) and anode quench resistor (R_{QA}) is equal to V_{ex}, which is the same as the voltage across the quench resistor (R_Q) in the SPQR configuration shown in Figure 6-2(b). To obtain the same avalanche quenching effect, the sum of the resistance of R_{QC} and R_{QA} is designed to have the same value of the single quench resistor (R_Q) in the traditional SPQR configuration, as shown in Eq. 6-5.

$$R_Q = R_{QA} + R_{QC}\,. \tag{6-5}$$

When R_{QC} and R_{QA} have the same value, their resistance is equal to $0.5R_Q$. With the same total capacitance at the cathode and anode in the DPQR circuit, the voltage at cathode will reduce to ($V_C - 0.5V_{ex}$) while the voltage at the anode will increase to ($V_A + 0.5V_{ex}$) until the voltage across the SPAD is below V_{BR} to quench the avalanching process. The quenching time constant for the DPQR is given by:

$$\tau_{QD} \approx R_{SPAD}C_{SPAD} + 0.5R_{SPAD}(C_{CS} + C_L)\,. \tag{6-6}$$

which is smaller than the quenching time constant τ_{QS} for the SPQR circuit.

During the avalanching process, the voltage change on C_{SPAD} is V_{ex} while the voltage change on the cathode is only $0.5V_{ex}$. Therefore, the amount of the charge (Q_{QD}) flowing through the SPAD's junction for the DPQR during the avalanching process is expressed as:

$$Q_{QD} = C_{SPAD}V_{ex} + 0.5(C_{CS} + C_L)V_{ex} \, . \tag{6-7}$$

Compared to Q_{QS} in Eq. 6-3, the total charge flowing through the SPAD junction during an avalanche in the DPQR circuit is reduced by $0.5(C_{CS} + C_L)V_{ex}$. This reduced charge results in a lower probability for the trap centers in the depletion region to capture carries, thus reducing the afterpulsing (AP). It is worth noting that the charge difference is highly dependent on the capacitance $(C_{CS} + C_L)$ between the cathode and substrate. If this capacitance is much smaller than C_{SPAD}, the avalanching charge difference between the SPQR and DPQR circuits will also be small. On the contrary, if $(C_{CS} + C_L)$ is much larger than C_{SPAD}, then the avalanching charge difference will be large enough so as that avalanching charges for the DPQR circuit (Q_{QD}) approaches ~50% of the avalanching charges for SPQR circuit (Q_{QS}).

After quench, the cathode will be charged back to V_C while the anode will be discharged back to V_A for detection of the next photon. The reset time constant τ_{RD} for the DPQR is given by:

$$\tau_{RD} \approx R_Q C_{SPAD} + 0.5R_Q(C_{CS} + C_L) \, . \tag{6-8}$$

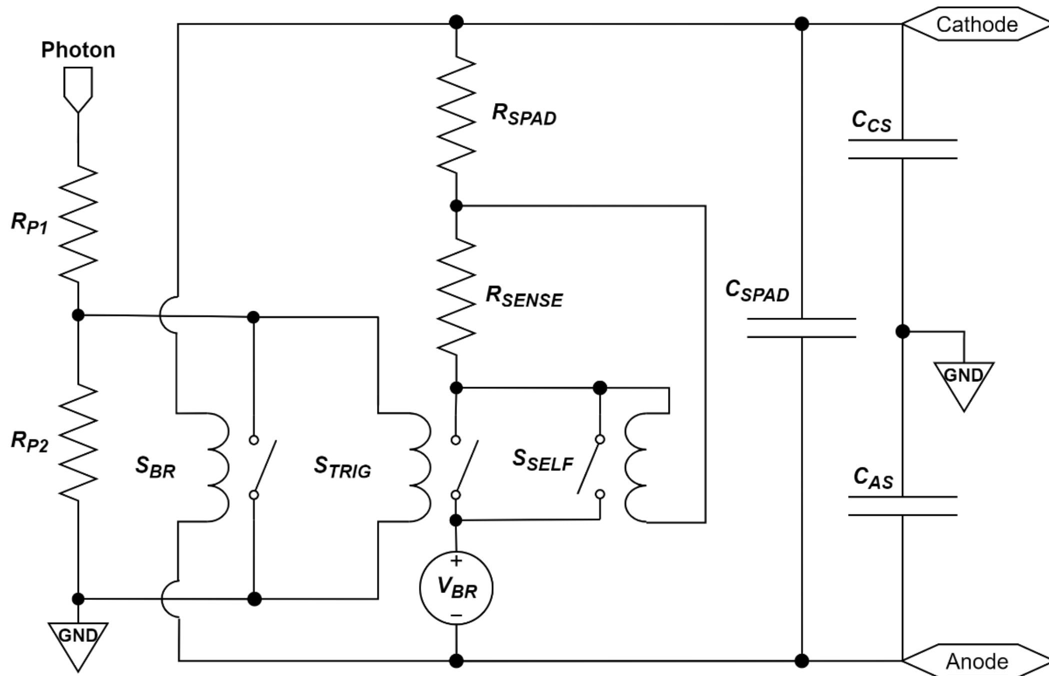
This indicates that the improvement of the reset time also depends on the capacitance $(C_{CS} + C_L)$ between the cathode and substrate. If the sum of C_{CS} and C_L is much larger than C_{SPAD}, the reset time constant for the DPQR will be only ~50% of the reset time constant of the SPQR circuit. Since the DT in the PQR configuration is dominated by the reset time, the DT of the DPQR circuit can also be reduced by ~50%, doubling the CR. Note that the reduced τ_{RD} means that the voltage across the SPAD is recharged back to its initial biasing voltage faster, thus resulting in a higher probability for the released carriers from the trap centers to generate AP. As shown in Eq. (7), the avalanching charges in the DPQR circuit are less than the SPQR circuit, which indicates a lower probability of

AP [109]. Therefore, the reduced reset time has a negative impact on the AP while the reduced avalanching charges has a positive impact on AP in a DPQR circuit. A comparison of the AP performance for the SPQR and DPQR circuits will be described in Section 6.3.

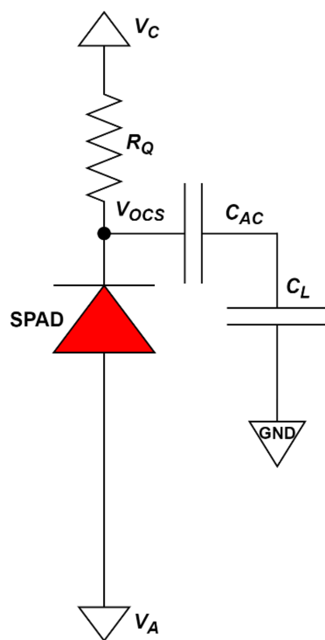
6.2.2. Simulations

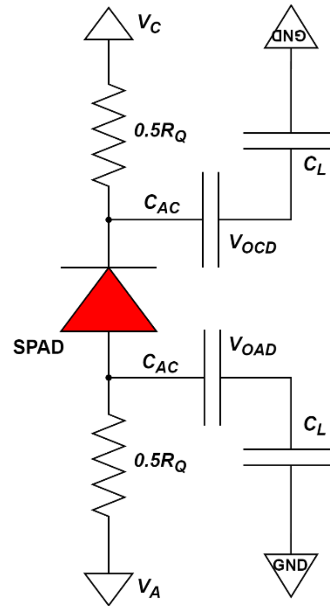
Simulations were performed in Cadence to investigate the operational principle of the DPQR circuit. A simplified version of the SPICE model from [108] was used for the SPAD in the simulations. Figure 6-3(a) shows the SPICE model for the SPAD, in which inductors combined with switches denote relays that switch on or off according to the threshold conditions. Initially, a very short pulse simulates an incident photon and closes the S_{TRIG} switch. Then, the avalanche process is initiated to discharge the SPAD junction capacitor (C_{SPAD}) through the SPAD avalanching resistor (R_{SPAD}), resulting in a fast and large current flowing through the junction. The current flowing through R_{SENSE} causes S_{SELF} to be closed to sustain the avalanching current. Simultaneously, the switch S_{BR} also closes, thus opening S_{TRIG}. Therefore, the width of the simulated photon pulse has no influence on the avalanche process. The quench of the avalanche process is determined by the current threshold of the S_{SELF} switch, where a value of 100 μA was used for simulation models and analytical calculations [109] and thus was also used in this model. To simplify the analysis, the parasitic capacitor of the cathode to the substrate (C_{CS}) and the parasitic capacitor of the anode to the substrate (C_{AS}) were set to have the same value of 100 fF. A R_{SPAD} 600 Ω and a C_{SPAD} 250 fF were used in this model [173].

Figure 6-3(b) shows the simulation setup for the traditional SPQR circuit. A 100 kΩ quench resistor connects to the cathode and a large AC-coupling capacitor C_{AC} is used to block the DC component of the output V_{OCS}. The load capacitance C_L is dominated by the capacitance (9.5 pF) of the passive probe of the oscilloscope. The simulation setup for the DPQR circuit is shown in Figure 6-3(c), in which two 50 kΩ resistors connect to both the cathode and anode. Similarly, AC-coupling capacitors (C_{AC}) are applied on both output terminals (V_{OCD} and V_{OAD}) to block the DC components, thus making the voltages V_{OCD} and V_{OAD} an ideal differential pair with the CM voltage at 0 V.



(a)





(b) (c)

Figure 6-3: (a) Simplified SPICE model for the SPAD. (b) Simulation setup for traditional single passive quench and reset circuit with the AC-coupled output. (c) Simulation setup for differential passive quench and reset circuit with AC-coupled differential outputs.

The simulation waveform is shown in Figure 6-4. In the DPQR configuration, it is seen that the voltage at the cathode (V_{OCD}) is a negative pulse while the voltage at the anode (V_{OAD}) is a positive pulse, and they can be regarded as a differential pair. The voltage across the SPAD can be obtained from the differential signals and is shown as a single-ended signal (V_{OCD} - V_{OAD}) in Figure 6-4. Compared to the 2.48 μs reset time (10 - 90%) for the output of SPQR (V_{OCS}), the reset time in the DPQR circuit is reduced by half (1.24 μs). Since the load capacitance C_L is much larger than the SPAD junction capacitance, the reset time constant for the SPQR circuit (Eq. (4)) can be simplified as $R_Q C_L$ and the reset time constant of the DPQR circuit (Eq. (8)) can be simplified as $0.5 R_Q C_L$. This is the reason why the reset time for the DPQR circuit is reduced by 2 times in the simulation results.

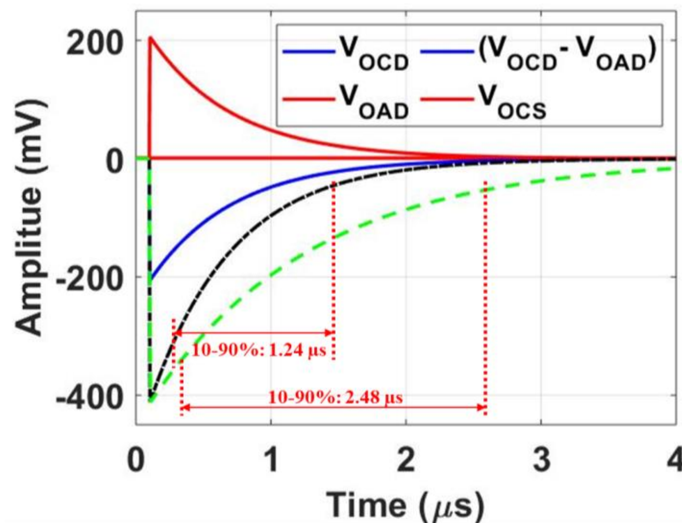


Figure 6-4: Simulated waveforms of traditional single passive quench and reset circuit and differential passive quench and reset circuit.

6.3. Measurements and Discussions

6.3.1. Measurement Setup

Experiments were performed to investigate the properties of the DPQR circuit, including: reset time, dark count rate (DCR) and AP. Figure 6-5(a) shows the setup to measure the DCR and AP of the SPAD. A SPAD is installed on a printed circuit board (PCB) to generate an output when the SPAD is triggered by dark carriers. Two DC power

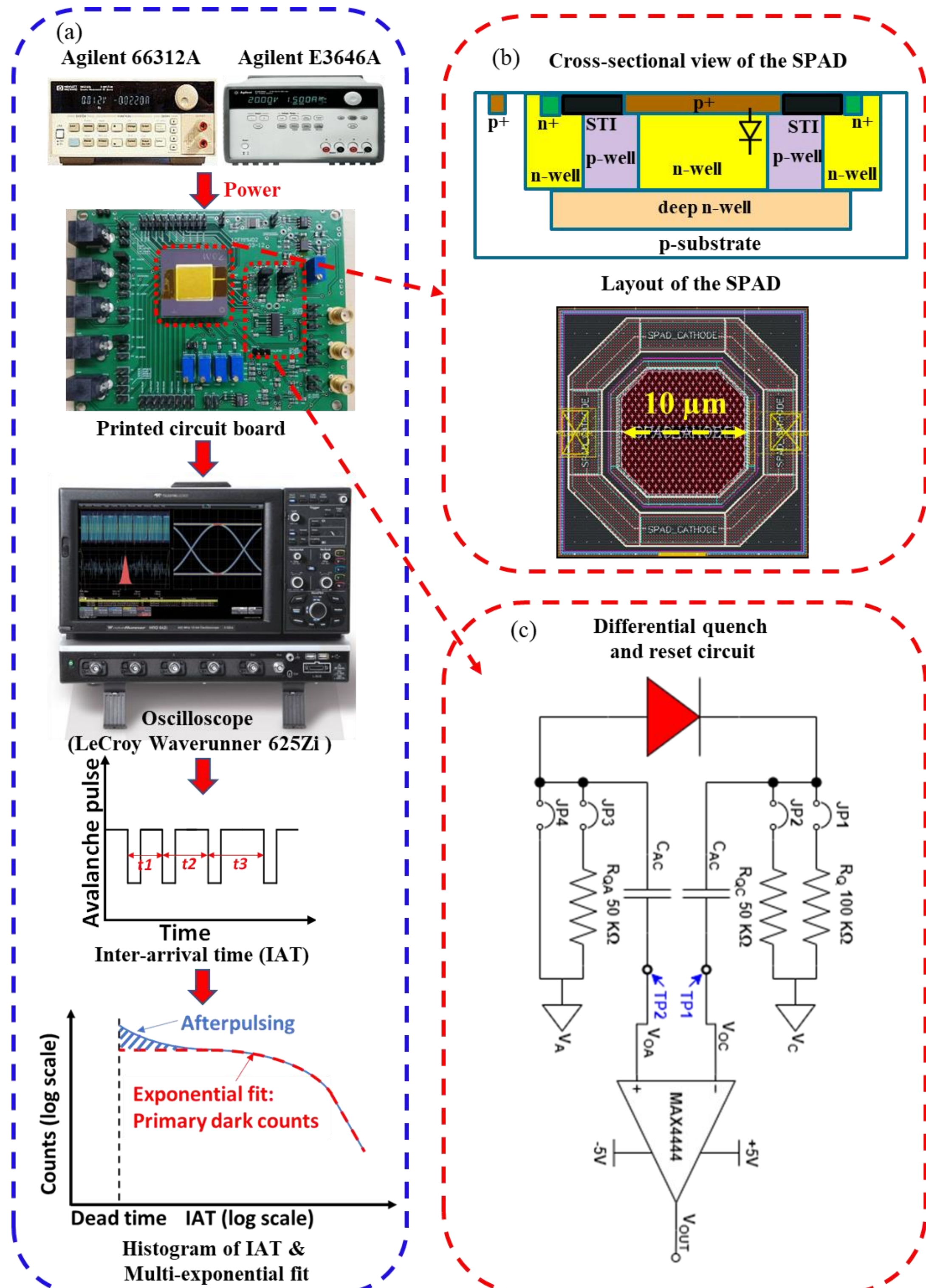


Figure 6-5: Setup to measure the DCR and AP of traditional single passive quench and reset circuit and differential passive quench and reset circuit.

supplies (Agilent 66312A and E3646A) provide voltages needed by the PCB. The SPAD output is then measured by a high-speed digital oscilloscope (LeCroy Waverunner 625Zi) to record the inter-arrival time (IAT) of the output pulses as shown in Figure 6-5(a). Then, the DCR and AP can be obtained by applying multi-exponential fits on the IAT histograms.

The SPAD used in the measurements is a p^+/n-well SPAD with a p-well guard ring that was designed and fabricated using a 65 nm standard CMOS process. The cross-sectional view and the layout of the SPAD are shown in Figure 6-5(b). The p^+ layer was extended into the p-well guard ring to push the shallow trench isolation (STI) away from the active region of the SPAD, thus lowering the dark noise [73]. The n-well cathode was connected by the deep n-well (DNW) layer. The SPAD has an octagonal shape with a diameter of 10 μm. To eliminate the influence from process variations, the same SPAD was used in both the SPQR circuit and DPQR circuit that were implemented at the PCB level. As shown in Figure 6-5(c), when jumpers JP1 and JP4 are short and JP2 and JP3 are open, the SPAD operates with the SPQR circuit, in which the SPAD cathode connects to the cathode biasing voltage (V_C) through a 100 kΩ quench resistor while the SPAD anode directly connects to the anode biasing voltage (V_A). When JP2 and JP3 are short and JP1 and JP4 are open, the SPAD operates with the DPQR circuit, in which both SPAD terminals connect to their biasing voltages through a 50 kΩ quench resistor. To ensure the same capacitive load for both SPQR and DPQR circuits, AC-coupling capacitors are connected to both cathode and anode even though the anode voltage is a constant DC voltage (V_A) in the SPQR configuration. After blocking the DC component, the anode output (V_{OC}) and cathode output (V_{OA}) are then sent to a differential-to-single-ended receiver (MAX4444) to convert the differential-ended signals to a single-ended signal (V_{OUT}).

6.3.2. Waveforms of the SPQR and DPQR Outputs

The outputs of the SPAD with the SPQR and DPQR were measured at the test points TP1 and TP2 using passive oscilloscope probes (9.5 pF capacitance). In the SPQR configuration, the cathode output measured at TP1 is the waveform of V_{OCS} shown in Figure 6-6, while the anode output at TP2 has a constant DC voltage of 0 V because the anode

connects to the DC anode biasing voltage (V_A) directly. The reset time (10 - 90%) for V_{OCS} is 2.78 µs. In the DPQR configuration, the cathode output at TP1 and the anode output at TP2 are denoted as V_{OCD} and V_{OAD} in Figure 6-6, respectively. It is seen that V_{OCD} is a negative pulse while V_{OAD} is a positive pulse. The single-ended signal converted from the differential signals (V_{OCD} and V_{OAD}) is shown as (V_{OCD} - V_{OAD}) in Figure 6-6, which has a reset time (10 – 90%) of 1.64 µs. The reset time for the DPQR circuit is reduced to ~59% of the reset time for SPQR circuit, which is not exactly 50% presented by the simulation results shown in Figure 6-4. This is because of the difference in the parasitic capacitances at the cathode and anode. As shown in the cross-sectional view of the SPAD in Figure 6-5(b), the n-well cathode should have a larger parasitic capacitance than the p^+ anode since the n-well and DNW are surrounded by the p-well guard ring and p-substrate. The effect of the larger parasitic capacitance at the cathode can also be observed on the waveforms of V_{OCD} and V_{OAD}. It is seen that the differential cathode output V_{OCD} shows a longer reset time than the differential anode output V_{OAD}. Therefore, this differential signal is not ideally symmetric.

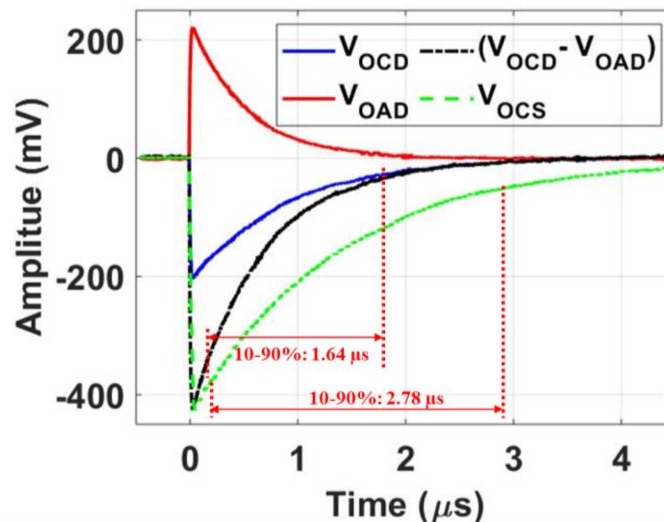


Figure 6-6: Measured waveforms of traditional single passive quench and reset circuit and differential passive quench and reset circuit.

Note that the RC time constant (τ) can be extracted through the rise time and then used to calculate the load capacitance of the circuits. The rise time from 10 – 90% is around 2.2

times of the time constant τ. Taking the rise time of the SPQR circuit as an example, the time constant τ is calculated to be 1.26 (2.78/2.2) μs. The quench resistor is 100 kΩ, therefore the total load capacitance is calculated to be 12.6 pF. The total load capacitance consists of 5 main parts: the input capacitance of the oscilloscope probe, the SPAD's junction capacitance, the capacitance from the pin pad and bonding wire of the SPAD package, the input capacitance of the chip on the PCB and the parasitic capacitance from PCB. Subtracting the 9.5 pF capacitance from the oscilloscope probe, the capacitance from other parts is 3.1 pF, which is a reasonable value.

6.3.3. DCR and AP

It is known that AP increases when the temperature decreases. That is because that the lifetime of trapped carriers increases as temperature decreases, thus leading the carriers released from the trap centers to have a higher probability to trigger secondary pulses [174]. Therefore, the DCR and AP measurements were intentionally performed at a low operating temperature (-35 °C) to show a higher AP for the SPAD under test. The excess voltage was varied from 0.7 to 1.1 V in steps of 0.1 V.

The measured results are shown in Figure 6-7. The SPAD with the SPQR circuit and the DPQR circuit presented almost the same primary DCR at the same V_{ex}, as shown in Figure 6-7(a). The mean of the ratio of DCR for the SPAD with the SPQR and the DPQR (DCR_{SPQR}/DCR_{DPQR}) is 1.01 as shown in Figure 6-8. This result is expected since the DCR of a SPAD is mainly determined by the intrinsic properties of the SPAD such as doping concentrations of the junction, guard ring structures, and defects during the fabrication, and not by the quench and reset circuits. In contrast, the quench and reset circuit greatly impacts the AP performance. To reduce the AP, the SPAD should be quenched as soon as possible to lower the probability of carriers being captured by the trap centers. In addition, the voltage across the SPAD should be kept below the breakdown voltage for a time that is long enough to ensure that the majority of trapped carriers are released before the SPAD is re-biased beyond the breakdown voltage [171]. The AP measurements with different excess voltages at -35 °C are shown in Figure 6-7(b), which indicates that the SPAD with the

DPQR circuit has a smaller AP than the SPAD with the SPQR circuits. The mean of the ratio of AP (AP_{SPQR}/AP_{DPQR}) for the SPAD with the SPQR and the DPQR is 1.94 as shown in Figure 6-8.

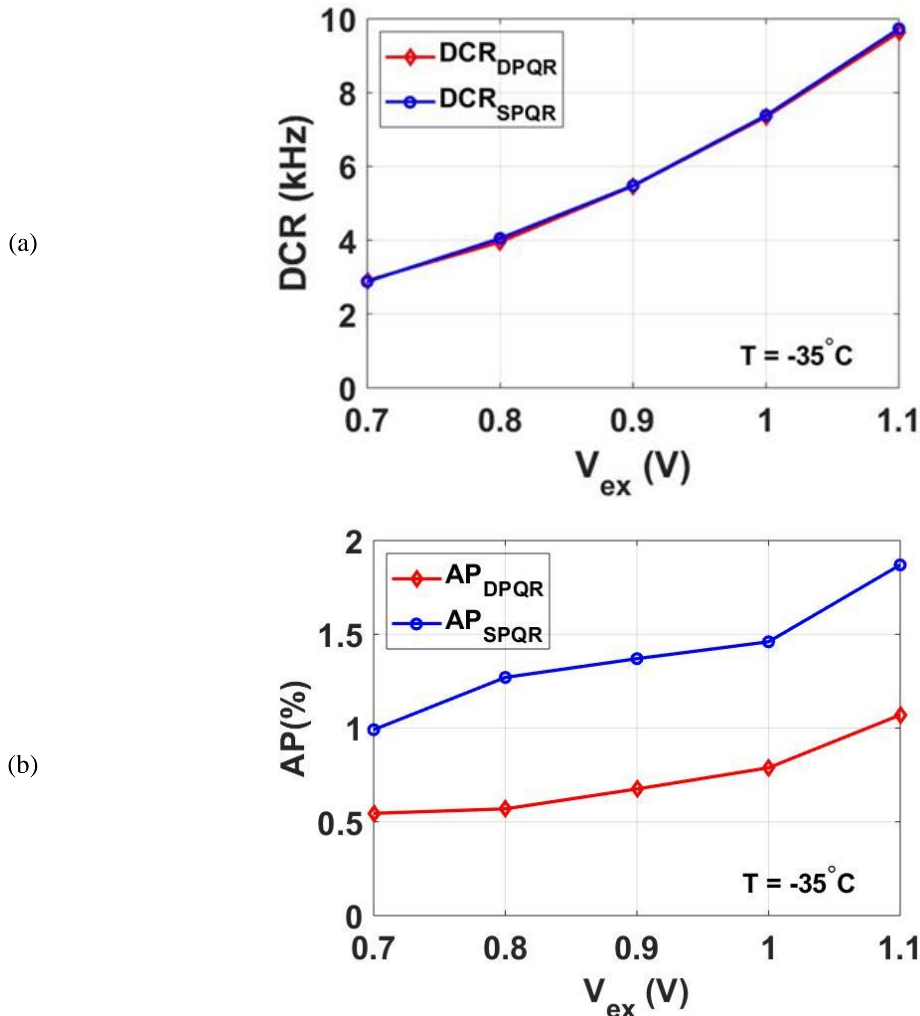


Figure 6-7: (a) Primary dark count rate (DCR) and (b) afterpulsing (AP) of the SPAD with the traditional single passive quench and reset (SPQR) and differential passive quench and reset (DPQR) as a function of excess voltage (V_{ex}) at -35 °C.

As analyzed in Section 6.2.1, there are two effects that impact the performance of the AP in the DPQR circuit. First, the reduced avalanching charges lower the AP. Second, the reduced reset time increases the AP. Note that C_L is much larger than C_{SPAD} due to the large capacitance of the passive oscilloscope probes in this measurement setup. Therefore, the avalanching charges in the DPQR circuit decrease by ~2 times, thus reducing the AP.

However, the reset time for the DPQR circuit is only 59% of that in the SPQR circuit as shown in Figure 6-6, which increases the AP of the SPAD with the DPQR circuit. A lower AP (Figure 6-7(b)) was measured in the SPAD with the DPQR circuit when compared to the SPQR circuit. This result indicates that the impact due to the reduced avalanching charges outweighs the impact from the reduced reset time on the performance of AP in the SPAD with the DPQR circuit in our measurement setup. The DQR circuit can also be implemented with an active reset to form a differential passive quench and active reset (PQAR) circuit in which the reset time is controllable and can be much smaller. Therefore, the influence from the reduced reset time of the differential PQAR circuit can be controlled. If the hold-off time for a single PQAR circuit and a differential PQAR circuit are set to be the same, the AP of the differential PQAR circuit can always be made lower when compared to the single PQAR circuit due to the reduced avalanching charge.

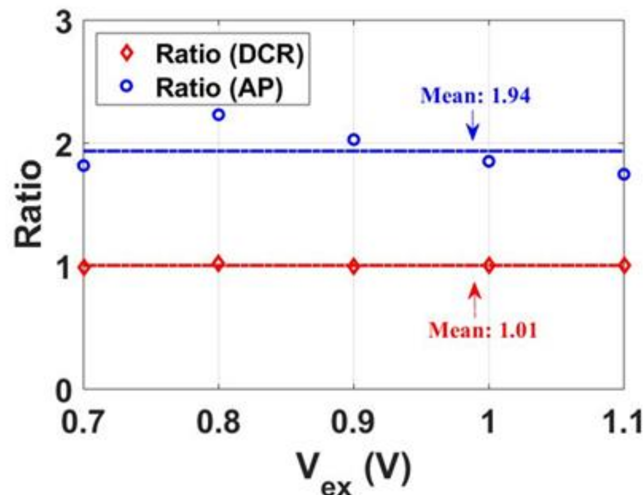


Figure 6-8: The ratio of DCR (DCR_{SPQR}/DCR_{DPQR}) and AP (AP_{SPQR}/AP_{DPQR}) for the SPQR and DPQR configurations.

6.3.4. Other Discussions

Fill Factor: As analyzed in Section 6.2, the values for the two quench resistors in the DPQR circuit are each half of the quench resistor in the SPQR circuit. If the quench resistors are implemented by the commonly used poly-resistor in the circuit design and layout, the configuration of the new DPQR circuit and conventional SPQR circuit will

occupy approximately the same area, thus resulting in an almost same fill factor (FF). Examples of the layouts for the SPQR and DPQR circuits are shown in Figures 6-9(a) and (b), respectively. These two passively quenched SPAD pixels have the same FF. However, the capacitors needed in the AC- coupled interfaces for the DPQR circuit should be considered since they will occupy extra area, thus lowering the effective FF for the whole pixel. It is worth noting that it is possible to share wells between adjacent pixels in the array design using the SPAD with the SPQR circuit. For example, if the array is designed using the p^+/n-well SPAD pixel as shown in Fig. 5(b), the n-well can be shared between adjacent pixels. However, wells cannot be shared between the SPAD pixels with the DPQR circuit since both cathode and anode of the SPAD with the DPQR circuit need to connect to a quench resistor, respectively. This would reduce the effective FF in the array design using the SPAD pixel with DPQR circuits.

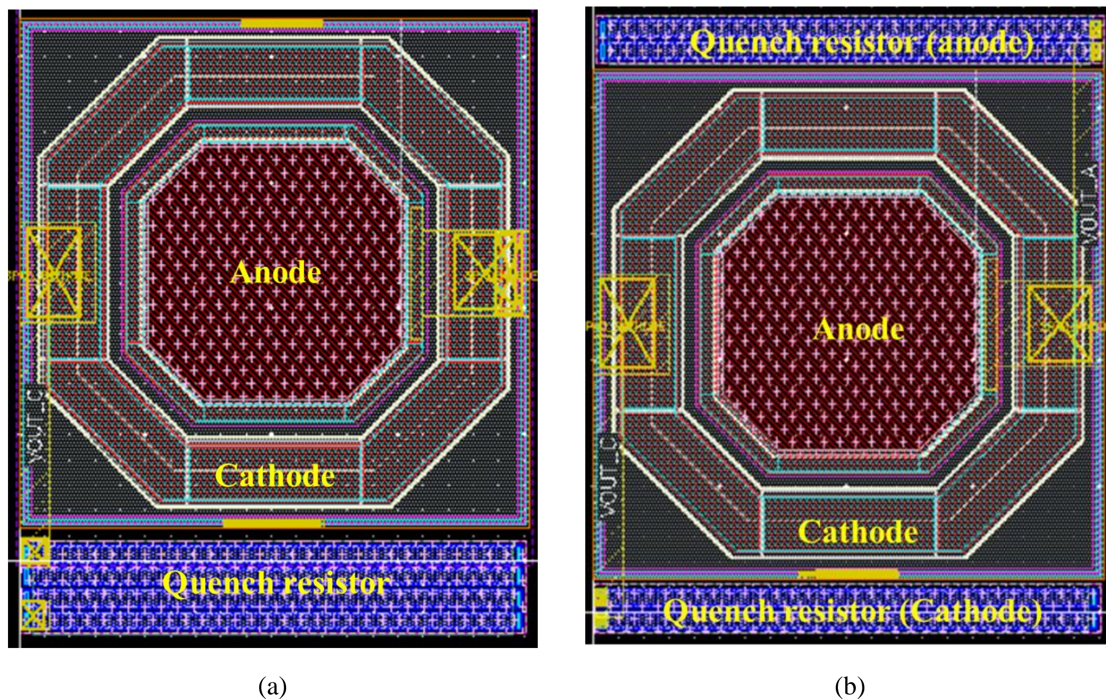


Figure 6-9: (a) Layout of a SPAD with a single passive quench and reset circuit consisting of a 100 kΩ poly resistor. (b) Layout of a SPAD with a differential passive quench and reset circuit consisting of two 50 kΩ poly resistors.

In the SPQR circuit, the SPAD's output can be adjusted to the input voltage range of the following signal conditioning and processing circuits by controlling the biasing

voltages. As an example, for a SPAD fabricated in a 1 V 65 nm CMOS process, the cathode biasing voltage (V_C) can be set to 1 V while the anode biasing voltage is set to a negative high voltage. Under this condition, the SPAD output will be in the range from (1-V_{ex}) to 1 V, so it can be directly sent to the following CMOS circuits for further processing without requiring the AC-coupling capacitor. However, in the DPQR circuit, when the cathode bias voltage is set to ensure that the cathode output voltage is in the acceptable range for the following circuits, the anode output has to be in the range of a negative high voltage. Similarly, when the anode bias voltage is set to ensure the anode output voltage is in the acceptable range for the following circuits, the cathode output will be in the range of a positive high voltage. Therefore, at least one AC-coupling capacitor is needed for the differential readout circuits in the DPQR circuit. This would eventually lower the effective FF for the whole pixel. However, this issue can be mitigated using a high-k capacitor for the AC-coupling interface or the 3-D SPAD structure in which the SPAD and the circuits are implemented in different layers. For example, a MIM capacitor with a size of 4×4 μm^2 in the standard 65 CMOS process has a capacitance of 35.3 fF and it can be used for the AC-coupling interface [92]. The size of this capacitor is ~5% of the size (15×18 μm^2) of an AQR circuit that was also implemented in the same standard 65 nm CMOS process [167]. That is, the area required by the AC-coupling capacitor will not reduce the FF too much if the DQR circuit is implemented in the active configuration to reduce its DT and increase its CR using the standard 65 nm CMOS process.

Noise Rejection: Compared to single signal path circuits, circuits with differential signal paths show improved signal-to-noise ratio (SNR) and signal integrity since the differential configuration can reject CM noise and interferences. To show the benefit of the DPQR circuit, we used the simulation setup shown in Figure 6-3(c). A 3-ns pulse with 150 mV amplitude was added on the biasing voltage to mimic the CM noise. The simulation results are shown in Figure 6-10, where the output V_{OCS} in the SPQR circuit shows a spike when the noise is applied. This spike can potentially cause a false trigger if it exceeds the threshold of the following comparator. Also, even though the differential pair V_{OCD} and

V_{OAD} have spikes in the DPQR circuit, the resulting single-ended signal (V_{OCD} - V_{OAD}) is not impacted by the spikes. That is, the CM noise is rejected in the DPQR circuit.

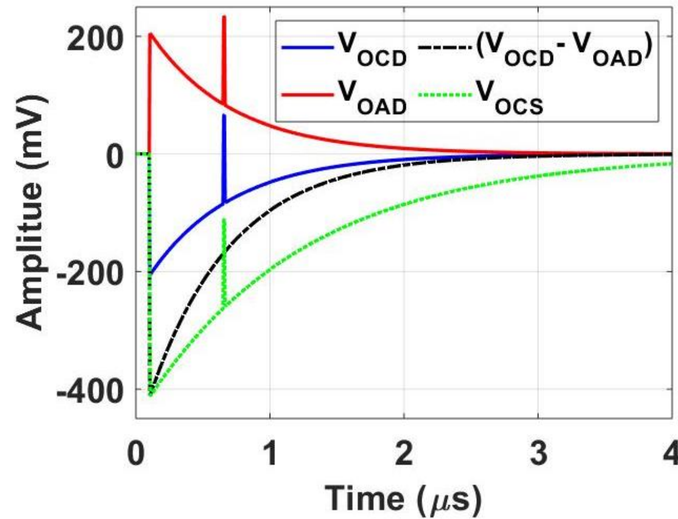
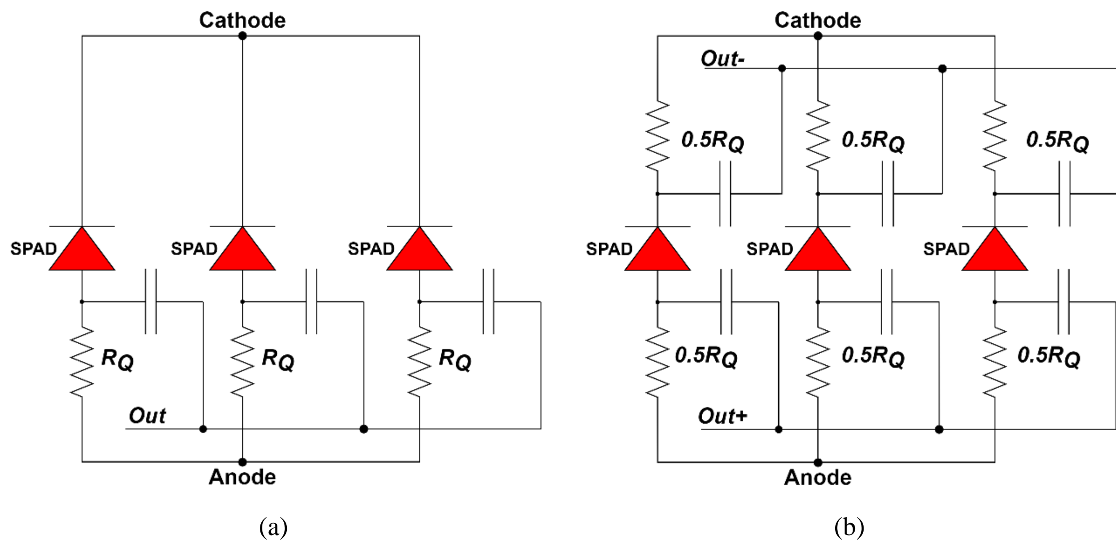


Figure 6-10: Simulated waveforms of traditional single passive quench and reset circuit and differential passive quench and reset circuit with a common-mode noise spike.



(a) (b)

Figure 6-11: (a) Traditional silicon photomultiplier (SiPM) with fast output terminal. (b) Proposed SiPM with differential output signals.

SPAD Array: Passively quenched SPAD pixels can be connected in parallel to form a silicon photomultiplier (SiPM). Figure 6-11(a) shows a traditional SiPM with a fast output terminal (*Out*). The fast output is the sum of individual AC-coupled outputs of all SPAD pixels [175]. Similarly, the SPAD with DPQR circuit can also be implemented in an array to form a differential SiPM with advantages such as: lower DT, higher CR, and improved

SNR due to the CM noise rejection. Figure 6-11(b) shows an example of the implementation of a differential SiPM in which fast differential output signals (*Out+* and *Out-*) are generated from both the cathode and anode of each SPAD.

6.4. Conclusions

In this chapter, we proposed a new differential quench and reset (QR) configuration for single-photon avalanche diodes (SPADs), which consists of two QR circuits on both the cathode and anode to quench and reset the SPAD through both terminals. The differential QR configuration can produce a negative pulse on the cathode as well as a positive pulse on the anode of the SPAD, which can be treated as a differential pair after blocking the DC voltage of the baselines of these two outputs. The analysis and simulations on a single passive quench and reset (SPQR) circuit and a differential passive quench and reset (DPQR) circuit indicated that the DPQR circuit has the potential to reduce the reset time, decrease the total avalanche charge, increase the count rate (CR), and to generate differential outputs when compared to the SPQR circuit.

Measurements were performed on a SPAD fabricated in a standard 65 nm CMOS process. To eliminate the influence of process variations, the same SPAD was measured with a SPQR circuit and a DPQR circuit that were implemented at the printed circuit board (PCB) level. The measurements showed that the DPQR can produce a differential pair of outputs and its reset time is only 59% of the SPQR. The ability to reject the common-mode (CM) noise in the differential configuration was demonstrated in simulations by adding a noise pulse on the biasing voltage of the SPAD. The dark count rate (DCR) and afterpulsing (AP) measurement were performed at a low temperature of -35°C to make differences between structures in the AP performance more obvious. The measurements showed that the SPAD with the SPQR circuit and the DPQR circuit presented almost the same primary DCR, but the AP of the SPAD with DPQR circuit was reduced by 1.9 times. This could be explained by the reduced charge through the SPAD during the avalanche process in the DPQR circuit. However, it is worth noting that the reduction of AP of the SPAD with DPQR circuit is highly dependent on the load capacitance of the SPAD outputs.

We also proposed a differential silicon photomultiplier (SiPM) using an array of DPQR pixels. Compared to the traditional SiPM, it can use the advantages of the differential configuration such as reduced dead time, increased count rate, and reduced AP, especially when there is a large load capacitance. In addition, the differential output pair can help to improve the signal-to-noise ratio since the differential SiPM is able to reject the CM noise and interference.

Chapter 7

Conclusions and Future Work

7.1. Conclusions

Due to its high sensitivity to differences in the metabolic and biological activities at the molecular level, the positron emission tomography (PET) imaging technique has become a powerful tool for acquiring functional images that can be used in a wide variety of clinical and preclinical areas. One of the most important components in the PET imaging system is the PET detector which converts the high-energy gamma rays emitted from an individual to electrical signals. Based on the electrical signals, three types of information: the position where the gamma photon impacts the scintillator; the time when the output pulse from the photosensor arrives; and the energy of the output pulse, can be extracted. Using this position, timing and energy information, the PET image can be obtained through image reconstruction algorithms.

Generally, there are four types of sensors used in PET applications: photomultiplier tubes (PMTs), avalanche photodiodes (APDs), single-photon avalanche diode (SPAD)-based detectors and cadmium zinc telluride (CZT) detectors. Among these four types of sensors, SPAD-based detectors are the most promising for PET applications due to their unique advantages such as high gain to detect even a single photon, excellent timing resolution to employ ToF techniques, compatibility with magnetic field to be integrated with MRI to form a multimodal PET/MRI imaging system, and compact size to reduce the overall volume of the imaging system and to reduce the complexity for the whole system integration. Therefore, the research work in this thesis focuses on the design and optimization of SPADs and their relevant quench and reset circuits for ToF PET imaging applications. To achieve a low fabrication cost, this thesis work specifically focusses on the design of SPADs using standard CMOS technologies.

Prior to the design, an intensive review was performed on sensors for PET applications with a focus on SPAD-based sensors. We reviewed and summarized the design and performance of SPADs in the recent literature. We also discussed three key points that need to be considered in the CMOS SPAD design. First is the junction of the CMOS SPADs. Different junctions such as p^+/n-well, n^+/p-well, p-well/DNW, n-well/p-substrate and DNW/p-substrate are being used for CMOS SPADs. Second is the design of guard ring structures since they not only help to prevent premature edge breakdown to ensure the Geiger mode operation for SPADs, but also have great influences on the noise performance. The commonly used guard ring structures such as shallow trench isolation, diffusion guard rings, virtual guard rings and poly gate guard rings were described and compared. Third is the design of quench and reset circuits. Three types of quench and reset circuits were discussed, including: passive quench and reset circuits, active quench and reset circuits and time-gated circuits.

Following the introduction of the PET application background and literature review, the random telegraph signal (RTS) noise was investigated based on n^+/p-well SPADs that were fabricated using a standard 130 nm technology (Chapter 3). Three parameters were used to quantify the RTS noise: T_{ON}, the time spent in the "on" state of the avalanche; T_{OFF}, which is the time spent in the "off" state of the avalanche; and I, the amplitude of the RTS noise current. The experimental results demonstrated that the mean value of T_{ON} increased with biasing voltage (V_{BIAS}) while the mean value T_{OFF} decreased with V_{BIAS}. The amplitude of RTS noise increased linearly with V_{BIAS}. The measurement results also showed that the temperature had no impact on the RTS noise amplitude that was almost constant as long as the absolute excess voltage across the SPAD was kept unchanged. Additionally, an analytical model was developed to describe the defects which cause RTS noise during the transitional period of the avalanche. Using this model, the effective dimension of the total defects was estimated based on the curve of RTS noise amplitude with biasing voltage. The defects in the depletion region can function as trap centers which play a significant role on the performance of the dark count rate (DCR) and afterpulsing (AP). As a result, the RTS noise of a SPAD which resulted from the defects in the depletion region was expected to

show correlations with its DCR and AP. Three SPAD samples with different effective dimensions of defects were used to measure and compare their values of DCR and AP at a low temperature of -30 °C. The experimental results revealed a trend that the SPAD with a larger effective dimension of defects presents a higher DCR and AP.

The function of the poly gate as guard rings for n^+/p-well SPADs was investigated in Chapter 4. Three types of n^+/p-well SPADs with different poly gate configurations: no poly gate (SPAD_NG), a dummy floating poly gate (SPAD_DG), and a field poly gate connected to the n^+ cathode (SPAD_FG) were designed and fabricated in a low-cost standard 180 nm CMOS process. TCAD simulations were performed to show the effectiveness of guard rings to prevent premature edge breakdown (PEB). The noise performance: dark count rate (DCR) and afterpulsing (AP) of these three SPADs were measured at different excess voltages and different temperatures. A comparison of measured results between SPAD_NG and SPAD_DG or SPAD_FG revealed that the poly gate greatly helped to reduce the DCR and AP. This was because the poly gate over the n-well guard ring pushed the noisy Si-SiO_2 interface of the STI away from the active region of the SPAD. The best results were obtained from SPAD_FG because the voltage on the poly gate reduced the magnitude of the electric field distribution in the guard ring region. Therefore, the likelihood that carriers in the guard ring can enter the active region of the SPAD and initiate the dark counts was reduced, resulting in improved noise performance. In addition, the SPAD_FG in this work was implemented in a standard CMOS process, which does not require extra mask layers in the fabrication, thus reducing the effort for integration of SPADs with other CMOS signal conditioning and processing circuits. Overall, the n^+/p-well field gate SPAD was verified to be an effective structure for improving the noise performance of CMOS SPADs.

Following this, a high-speed SPAD pixel was presented in Chapter 5. This pixel consisted of a p^+/n-well SPAD integrated with a compact and high-speed active quench and reset (AQR) circuit. This pixel was designed and fabricated using a standard 65 nm CMOS process. Thanks to the positive feedback loop in the AQR circuit, the quenching time for this pixel was only 0.1 ns according to the post-layout simulation results. Due to the fast

response of the transistors in this advanced 65 nm technology node, the minimum dead time of this pixel was ~3.35 ns, which corresponds a maximum count rate of ~300 Mcps. The measurements showed that the SPAD pixel had a dark count rate of 21 kHz, a peak photon detection probability of 23.8% at 420 nm wavelength and a timing jitter of 139 ps using a 405 nm pulsed laser and an excess voltage of 0.5 V. Due to the short quenching time, negligible afterpulsing was observed even at a low temperature of -35 °C.

Finally, a new differential quench and reset (QR) configuration for single-photon avalanche diodes (SPADs) was proposed in Chapter 6. This differential QR circuit consisted of two QR circuits on both the cathode and anode to quench and reset the SPAD through both terminals. In the differential QR configuration, a negative pulse was generated on the cathode of the SPAD while a positive pulse was generated on its anode. These two signals could be regarded as a differential pair after blocking their DC voltages. Through the analysis and simulations on a single passive quench and reset (SPQR) circuit and a differential passive quench and reset (DPQR) circuit, it was found that the DPQR circuit had the potential to reduce the reset time, to decrease the total avalanche charge, to increase the count rate (CR), and to generate differential outputs when compared to the SPQR circuit. Measurements were performed on a SPAD fabricated in a standard 65 nm CMOS process. To eliminate the influence of process variations, the same SPAD was measured with a SPQR circuit and a DPQR circuit that were implemented at the printed circuit board level. The measurements showed that the DPQR circuit can produce a differential pair of outputs with a reset time that was only 59% of the SPQR circuit. A noise pulse was added on the biasing voltage of the SPAD during the simulation. The simulated waveforms showed that the differential configuration had the ability to reject the common-mode noise. The measurements on the dark count rate (DCR) and afterpulsing (AP) showed that the SPAD with the SPQR circuit and the DPQR circuit presented almost the same primary DCR, but the AP of the SPAD with DPQR circuit was reduced by 1.9 times. The similar value of DCR in both SPQR and the DPQR circuits was expected because the DCR of a SPAD is mainly determined by the intrinsic properties of the SPAD such as doping concentrations of the junction, guard ring structures, and defects during the fabrication, and

not by the quench and reset circuits. However, the reduced AP could be explained by the reduced charge through the SPAD during the avalanche process in the DPQR circuit. In addition, the issue of fill factor of DPQR was discussed and a differential silicon photomultiplier (SiPM) using an array of DPQR pixels was proposed. Compared to the traditional SiPM, it could exploit advantages of the differential configuration such as reduced dead time, increased count rate, and reduced AP, especially when there is a large load capacitance. Also, the differential output pair from the differential SiPM is able to reject the common mode noise and interference, thus improving the signal-to-noise ratio in the detection system.

7.2. Future Work

In order to design optimized SPADs using standard CMOS technologies, many techniques and proposals to address the research challenges in terms of lowering the dark noise, increasing the photon detection efficiency and improving the timing jitter were studied and investigated in recent years. Advancements in the design and fabrication of CMOS SPADs allow them to achieve competitive performance at the system level when compared to SPADs fabricated in customized technologies. However, there are still some future challenges that need to be overcome to push the limits on the performance of standard CMOS SPADs. In this section, we will first describe some recommendations for the improvement or extended work based on the works that presented in this thesis, then discuss other potential future works that can improve the performance of the CMOS SPADs for PET applications.

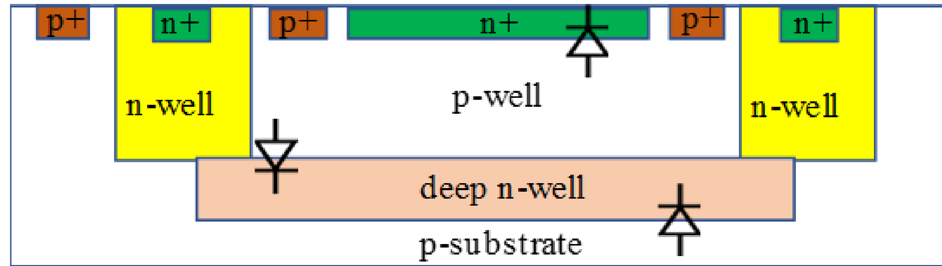
1) **RTS Noise:** In the work of random telegraph signal (RTS) noise of the SPAD presented in Chapter 3, the three SPAD samples used in the measurement had the same size due to the limitation on the availability of SPADs with different sizes. The RTS noise measurement on SPADs with different sizes may help us to have a better understanding of the behaviour of RTS noise in SPADs. In addition, some research showed RTS might be the origin of $1/f$ noise in devices [152]. In our work, we showed that RTS noise correlates with the DCR and AP of the SPAD. Therefore,

we can expect that 1/*f* noise of a SPAD should also present some correlations with its DCR and AP.

2) **Poly Gate SPAD:** In the work of the poly gate SPAD design presented in Chapter 4, we investigated the function of the poly gate on the noise performance of the n^+/p-well SPADs. Through measurements, we found that the SPAD with a dummy poly gate showed better noise performance than the SPAD without a gate. The noise performance could be further improved by connecting the poly gate to the n^+ cathode of the SPAD. Through the TCAD simulations, we found that connecting the poly gate to the cathode of the SPAD helped to reduce the electric field in the n-well guard ring region near the edge of the depletion region of the SPAD, thus reducing the dark counts. The analysis and explanations of the results in this work were based on TCAD simulations. However, it would be useful for more analysis to be performed based on semiconductor physics theory in future work. One potential topic could be the physical mechanism of the influence from the field gate on the sources of the dark noise such as carrier diffusion, and carrier generation and recombination in the field gate SPADs.

3) **AQR Circuit:** In the active quench and reset (AQR) circuit design presented in Chapter 5, a conservative design strategy was used in the design of the pulse generation circuit that generates a 1.5 ns delayed pulse with a width of 1.5 ns to reset the SPAD. This means that the SPAD is held off for a minimum 1.5 ns after quench, and then reset by a 1.5 ns wide pulse to its initial condition for the detection of the next photon. That is, 3 ns is needed for this pulse generation circuit. This time is a large fraction of the minimum dead time (3.35 ns). Therefore, optimization on the operating time of the pulse generation can potentially lower the dead time, thus increasing the count rate of this SPAD pixel. Additionally, thanks to the short quenching time of this AQR circuit, the AP is negligible. Therefore, the hold-off time controlling circuit is not necessary for this pixel. We can save some layout area to make the AQR circuit more compact and also reduce the power consumption by removing the hold-off time controlling circuit from the pixel design.

4) **DQR Circuit:** In the work of the differential quench and reset (DQR) circuit for SPADs, we described the operation principle based on the differential passive quench and reset (DPQR) circuit. Moreover, the measurements were also based on the DPQR circuit that was implemented at the printed circuit board level. Since the content in this work was all based on the DPQR circuit, one extended work for DQR circuit could be the design and implementation of differential active quench and reset (DAQR) circuit. The DAQR circuit can be designed and fabricated using standard CMOS technologies. Another future work can be the design of the differential silicon photomultiplier (SiPM) as shown in Figure 6-11. This differential SiPM is an array of SPAD pixels with DPQR circuits, which can also be implemented in standard CMOS technologies.
5) **Triple-junction SPAD:** SPADs fabricated in standard CMOS technologies usually suffer from a low PDE due to two main reasons: the photon reflection and absorption of a thick passivation layer and the internal dielectric oxide insulation layers, and a thin depletion region. One potential solution to increase the PDE is to use multi-junction SPADs. A dual-junction SPAD implemented in a 130 nm CIS process that incorporated the deep n-well (DNW) layer was proposed in [92]. This SPAD had two planar junctions: p-well/DNW and DNW/p-sub. The measurements showed that when the junctions were operated concurrently, the PDE could be improved by ~30% when compared to the PDE when only one junction of the SPAD was operated. In addition, one could distinguish which junction was triggered by a photon by observing the output pulse width from different junctions. With proper guard ring structures, it is possible to design to a triple-junction SPAD that consists of three stacked planar junctions: n^+/p-well, p-well/DNW and DNW/p-sub as shown in Figure 7-1. It is expected that a higher PDE and a broader band can be achieved in the triple-junction SPAD than the single-junction SPAD. Similarly, the triple-junction SPAD also has the ability to distinguish the wavelength of the photon due to the different depths of these three junctions.



(a)

Figure 7-1: Three stacked junctions: n^+/p-well, p-well/DNW, and DNW/p-substrate junctions.

6) **PDE Model:** Many models have been developed for the dark noise of SPADs to investigate the different sources of the noise. Models were also proposed and validated for afterpulsing and timing jitter. The simulation results from these models are in good agreement with the relevant measurement results, which shows their potential to predict the performance of the SPAD and to guide the optimization in the SPAD design. However, the model for the photon detection probability, one of the important performance parameters, was investigated in few publications. In order to be detected by a SPAD, photons go through the following steps.

   1) First, photons should pass through a thick passivation layer and multiple inter-metal dielectric oxide insulation layers to arrive the surface of the SPAD.
   2) Then, the arrived photons penetrate the neutral layer to reach the depletion region to be absorbed.
   3) The absorbed photons then generate electron-hole pairs.
   4) The electron-hole pairs must undergo impact ionization to trigger an avalanching process due to high biasing voltage.
   5) The avalanching process is detected by the readout circuit to register a photon-detection event.

As described, many steps are involved in the process, thus increasing the complexity of the development of the PDE model. Among those steps, the thickness of the passivation layer and the inter-metal dielectric (IMD) layers can vary up to 20% due to process variations. This makes it more difficult for the model to

accurately predict the PDE of the SPAD. Therefore, the method to obtain the information about the thickness of the passivation layer and IMD layers, and optical properties of these layers needs to be investigated in order to build a reliable PDE model for the SPAD. One possible solution is to measure the PDEs using a vertically stacked multi-junction SPADs. Here, we take triple-junction SPAD as an example. Although three junctions should present different responses of the PDE due to their different junction structures and depths, they all have the same passivation layer and IMD layers above the silicon surface. To make the description easier, three junctions of the SPAD are given special names: JT for the top junction, JM for the middle junction and JB for the bottom junction. For the PDE measurement, each junction can be biased individually at one time to obtain three PDE responses. Also, two of three junctions can be biased together to get three PDE responses. The combination can be JT and JM, JT and JB, or JM and JB. Additionally, all three junctions can be biased together to get the PDE response. In total, we should get seven PDE responses. Each response can be viewed as a function that contains a same condition: same passivation layer and IMD layers. Therefore, it is possible to derive the properties of the passivation layer and IMD layers by solving these seven functions.

7) **dSiPM:** Besides the low manufacturing cost, SPADs fabricated in standard CMOS technologies also present the potential to be integrated with other CMOS signal conditioning and processing circuits to form a complete detection system. One of the most important signal conditioning and processing circuits are time-to-digital converters (TDCs) that digitize the timing information of the input signal. The integration of the SPAD arrays with TDCs are called digital silicon photomultipliers (dSiPMs), which provides the digital timing information of the SPAD output. Thanks to the integration, the dSiPM has improved timing performance since it reduces the jitter from the interconnection between SPADs and TDCs if they are implemented and packaged in different chips. This will eventually improve the signal-to-noise ratio for the PET imaging system if dSiPM-based PET detectors are

used. Additionally, the use of dSiPMs also make the PET system more compact and consume less power.

8) **3-D dSiPM:** For the integration of the dSiPM, one important structure is the 3-D dSiPM, in which SPADs and TDC can be fabricated on different wafers using different processes, then vertically connected by through-silicon vias as shown in Figure 7-2 [44]. In planar integration, SPADs and TDCs are on the same wafer and use the same technology, which requires a compromise in the performance between SPADs and TDCs. However, in the 3-D structure, SPADs can be fabricated using less-scaled technologies that have lower doping concentration resulting in thicker depletion regions, thus resulting in a higher PDE and a lower DCR caused by tunneling. Similarly, TDCs can be implemented in a more advanced CMOS technology so as to achieve shorter delay times, higher operational frequency and lower power consumption. That is, the SPAD and TDC tiers can be optimized and fabricated in separate processes by using the 3-D structure. Therefore, 3-D dSiPMs may achieve excellent performance given that the SPAD and TDC are both optimized using their preferred technologies. However, it is worth noting that the cost for the 3-D integration would be much higher than the planar process.

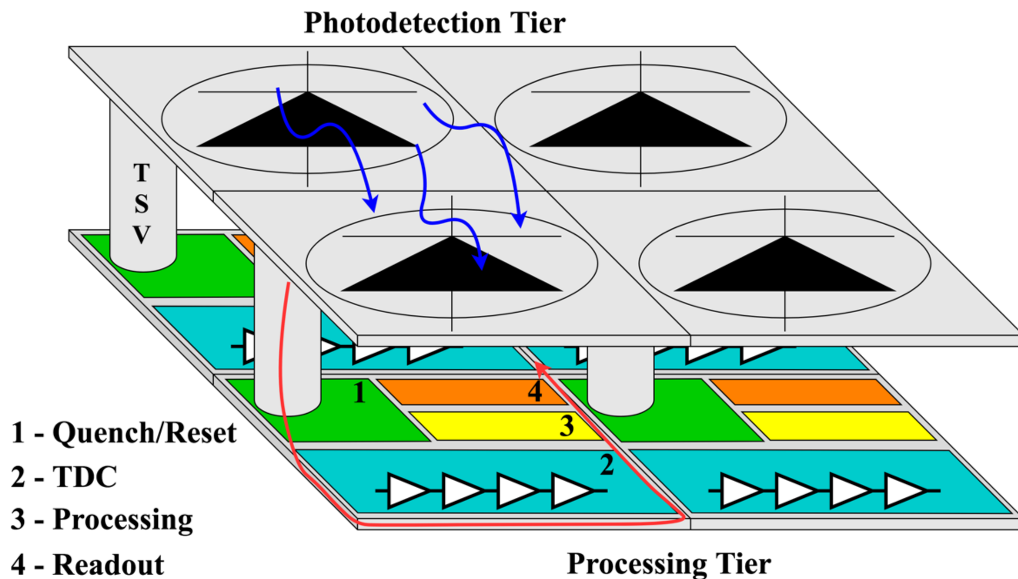


Figure 7-2: Illustration of a 3D pixel structure for a SPAD-based sensor [44]. (© 2021 IEEE)

References

[1] M. L. Giger, “Machine Learning in Medical Imaging,” *J. Am. Coll. Radiol.*, vol. 15, no. 3, pp. 512–520, Mar. 2018, doi: 10.1016/j.jacr.2017.12.028.

[2] E. M. Rohren, T. G. Turkington, and R. E. Coleman, “Clinical Applications of PET in Oncology,” *Radiology*, vol. 231, no. 2, pp. 305–332, May 2004, doi: 10.1148/radiol.2312021185.

[3] A. Kadir *et al.*, “Dynamic Changes in PET Amyloid and FDG Imaging at Different Stages of Alzheimer’s Disease,” *Neurobiol. Aging*, vol. 33, no. 1, pp. 198.e1-198.e14, Jan. 2012, doi: 10.1016/j.neurobiolaging.2010.06.015.

[4] F. Y. Keng, “Clinical Applications of Positron Emission Tomography in Cardiology: A Review.,” *Ann. Acad. Med. Singapore*, vol. 33, no. 2, pp. 175–82, Mar. 2004.

[5] M. A. Nader and P. W. Czoty, “Brain Imaging in Nonhuman Primates: Insights into Drug Addiction,” *ILAR J.*, vol. 49, no. 1, pp. 89–102, Jan. 2008, doi: 10.1093/ilar.49.1.89.

[6] J. S. Kim *et al.*, “Performance Measurement of the microPET Focus 120 Scanner,” *J. Nucl. Med.*, vol. 48, no. 9, pp. 1527–1535, Sep. 2007, doi: 10.2967/jnumed.107.040550.

[7] K. Sato *et al.*, “Performance Evaluation of the Small-animal PET Scanner ClairvivoPET Using NEMA NU 4-2008 Standards,” *Phys. Med. Biol.*, vol. 61, no. 2, pp. 696–711, Jan. 2016, doi: 10.1088/0031-9155/61/2/696.

[8] R. Prasad, O. Ratib, and H. Zaidi, “Performance Evaluation of the FLEX Triumph X-PET Scanner Using the National Electrical Manufacturers Association NU-4 Standards,” *J. Nucl. Med.*, vol. 51, no. 10, pp. 1608–1615, Oct. 2010, doi: 10.2967/jnumed.110.076125.

[9] H. Zhang *et al.*, “Performance Evaluation of PETbox: A Low Cost Bench Top Preclinical PET Scanner,” *Mol. Imaging Biol.*, vol. 13, no. 5, pp. 949–961, Oct. 2011, doi: 10.1007/s11307-010-0413-y.

[10] K. Herrmann *et al.*, “Evaluation of the Genisys4, a Bench-Top Preclinical PET Scanner,” *J. Nucl. Med.*, vol. 54, no. 7, pp. 1162–1167, Jul. 2013, doi: 10.2967/jnumed.112.114926.

[11] “Ingenuity TF PET/MR system.” https://www.philips.ca/healthcare/product/HC882380/ingenuity-tf-petmr-system/documentation (accessed Mar. 20, 2019).

[12] H. Zaidi *et al.*, “Design and Performance Evaluation of A Whole-Body Ingenuity TF PET-MRI System.,” *Phys. Med. Biol.*, vol. 56, no. 10, pp. 3091–106, May 2011, doi: 10.1088/0031-9155/56/10/013.

[13] “SIGNA^{TM} PET/MR.” https://www.gehealthcare.com/en/products/magnetic-resonance-imaging/3-0t/signa-pet-mr (accessed Mar. 20, 2019).

[14] “Biograph mMR.” https://www.siemens-healthineers.com/molecular-imaging/pet-mr/mi-biograph-mmr (accessed Mar. 20, 2019).

[15] A. Boss, M. Weiger, and F. Wiesinger, “Future Image Acquisition Trends for PET/MRI,” *Semin. Nucl. Med.*, vol. 45, no. 3, pp. 201–211, May 2015, doi: 10.1053/j.semnuclmed.2014.12.002.

[16] H. Jadvar and P. M. Colletti, “Competitive Advantage of PET/MRI,” *Eur. J. Radiol.*, vol. 83, no. 1, pp. 84–94, Jan. 2014, doi: 10.1016/j.ejrad.2013.05.028.

[17] D. A. Torigian *et al.*, “PET/MR Imaging: Technical Aspects and Potential Clinical Applications,” *Radiology*, vol. 267, no. 1, pp. 26–44, Apr. 2013, doi: 10.1148/radiol.13121038.

[18] Saha G. B., *Basics of PET Imaging: Physics, Chemistry, and Regulations*, 3rd ed. Cleveland, OH, USA: Springer International Publishing, 2015.

[19] L. H. C. Braga *et al.*, “A Fully Digital 8 x 16 SiPM Array for PET Applications With Per-Pixel TDCs and Real-Time Energy Output,” *IEEE J. Solid-State Circuits*, vol. 49, no. 1, pp. 301–314, 2014, doi:

10.1109/jssc.2013.2284351.

[20] K. Iniewski, *Electronics for Radiation Detection*. CRC Press, 2011.

[21] S. R. Cherry, J. A. Sorenson, and M. E. Phelps, *Physics in nuclear medicine*. Elsevier/Saunders, 2012.

[22] Z. Cheng, M. J. Deen, and H. Peng, "A Low-Power Gateable Vernier Ring Oscillator Time-to-Digital Converter for Biomedical Imaging Applications," *IEEE Trans. Biomed. Circuits Syst.*, vol. 10, no. 2, pp. 445–454, 2016, doi: 10.1109/TBCAS.2015.2434957.

[23] H. Peng and C. S. Levin, "Recent Development in PET Instrumentation.," *Curr. Pharm. Biotechnol.*, vol. 11, no. 6, pp. 555–71, Sep. 2010, Accessed: Nov. 29, 2018. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/20497121.

[24] T. F. Budinger, "Time-of-flight Positron Emission Tomography: Status Relative to Conventional PET.," *J. Nucl. Med.*, vol. 24, no. 1, pp. 73–8, Jan. 1983, Accessed: Apr. 30, 2019. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/6336778.

[25] J. S. Karp, "Time-of-Flight PET.," *advancing molecular imaging & therapy*, 2006. http://snmmi.files.cms-plus.com/docs/PETCENews_Fall06.pdf (accessed Mar. 04, 2019).

[26] W. Becker, *Advanced Time-Correlated Single Photon Counting Techniques*, vol. 81. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005.

[27] V. C. Spanoudaki and C. S. Levin, "Photo-Detectors for Time of Flight Positron Emission Tomography (ToF-PET)," *Sensors*, vol. 10, no. 11, pp. 10484–10505, Nov. 2010, doi: 10.3390/s101110484.

[28] Hamamatsu Photonics K.K., "Photomultiplier Tubes – Basics and Applications," 2007. https://www.hamamatsu.com/resources/pdf/etd/PMT_handbook_v3aE.pdf (accessed Nov. 14, 2019).

[29] K. Ng, *Complete Guide to Semiconductor Devices*. New York: McGraw-Hill, 1999.

[30] Hamamatsu, "Si APD (avalanche photodiode) Selection Guide.," 2019. https://www.hamamatsu.com/resources/pdf/ssd/si_apd_kapd0001e.pdf (accessed Jun. 03, 2019).

[31] D. Palubiak, M. M. El-Desouki, O. Marinov, M. J. Deen, and Q. Fang, "High-Speed, Single-Photon Avalanche-Photodiode Imager for Biomedical Applications," *IEEE Sens. J.*, vol. 11, no. 10, pp. 2401–2412, 2011, doi: 10.1109/JSEN.2011.2123090.

[32] A. Gallivanoni, I. Rech, and M. Ghioni, "Progress in Quenching Circuits for Single Photon Avalanche Diodes," *IEEE Trans. Nucl. Sci.*, vol. 57, no. 6, pp. 3815–3826, Dec. 2010, doi: 10.1109/TNS.2010.2074213.

[33] R. Mita and G. Palumbo, "High-Speed and Compact Quenching Circuit for Single-Photon Avalanche Diodes," *IEEE Trans. Instrum. Meas.*, vol. 57, no. 3, pp. 543–547, Mar. 2008, doi: 10.1109/TIM.2007.911691.

[34] D. P. Palubiak and M. J. Deen, "CMOS SPADs: Design Issues and Research Challenges for Detectors, Circuits, and Arrays," *IEEE J. Sel. Top. Quantum Electron.*, vol. 20, no. 6, pp. 409–426, 2014, doi: 10.1109/JSTQE.2014.2344034.

[35] M. Mazzillo *et al.*, "Silicon photomultipliers development at STMicroelectronics," May 2013, vol. 8773, p. 877302, doi: 10.1117/12.2014035.

[36] T. K. Lewellen, "Recent Developments in PET Detector Technology," *Phys. Med. Biol.*, vol. 53, no. 17, pp. R287–R317, Sep. 2008, doi: 10.1088/0031-9155/53/17/R01.

[37] M. F. Bieniosek, P. D. Olcott, and C. S. Levin, "Time Resolution Performance of an Electro-optical-coupled PET Detector for Time-of-flight PET/MRI," in *2011 IEEE Nuclear Science Symposium Conference Record*, Oct. 2011, pp. 2531–2533, doi: 10.1109/NSSMIC.2011.6152683.

[38] S. Il Kwon *et al.*, "Development of Small-Animal PET Prototype Using Silicon Photomultiplier (SiPM): Initial Results of Phantom and Animal Imaging Studies," *J. Nucl. Med.*, vol. 52, no. 4, pp. 572–579, Apr. 2011, doi: 10.2967/JNUMED.110.079707.

[39] H. S. Yoon *et al.*, "Initial results of simultaneous PET/MRI experiments with an MRI-compatible silicon photomultiplier PET scanner.," *J. Nucl. Med.*, vol. 53, no. 4, pp. 608–14, Apr. 2012, doi:

10.2967/jnumed.111.097501.

[40] J. W. Cates, M. F. Bieniosek, and C. S. Levin, "Highly multiplexed signal readout for a time-of-flight positron emission tomography detector based on silicon photomultipliers," *J. Med. Imaging*, vol. 4, no. 1, p. 011012, Mar. 2017, doi: 10.1117/1.jmi.4.1.011012.

[41] H. J. Choe, Y. Choi, W. Hu, J. Yan, and J. Ho Jung, "Development of capacitive multiplexing circuit for SiPM-based time-of-flight (TOF) PET detector," *Phys. Med. Biol.*, vol. 62, no. 7, pp. N120–N133, Mar. 2017, doi: 10.1088/1361-6560/aa5f9b.

[42] M. El-Desouki, M. J. Deen, Q. Fang, L. Liu, F. Tse, and D. Armstrong, "CMOS image sensors for high speed applications," *Sensors*, vol. 9, no. 1, pp. 430–444, Jan. 2009, doi: 10.3390/s90100430.

[43] Z. Cheng, X. Zheng, M. J. Deen, and H. Peng, "Recent Developments and Design Challenges of High-Performance Ring Oscillator CMOS Time-to-Digital Converters," *IEEE Trans. Electron Devices*, vol. 63, no. 1, pp. 235–251, Jan. 2016, doi: 10.1109/TED.2015.2503718.

[44] R. Scott, W. Jiang, and M. J. Deen, "CMOS Time-to-Digital Converters for Biomedical Imaging Applications," *IEEE Rev. Biomed. Eng.*, pp. 1–1, 2021, doi: 10.1109/RBME.2021.3092197.

[45] T. Frach, G. Prescher, C. Degenhardt, R. Gruyter, A. Schmitz, and R. Ballizany, "The Digital Silicon Photomultiplier — Principle of Operation and Intrinsic Detector Performance," in *2009 IEEE Nuclear Science Symposium Conference Record (NSS/MIC)*, 2009, pp. 1959–1965, doi: 10.1109/NSSMIC.2009.5402143.

[46] L. Verger *et al.*, "Characterization of CdTe and CdZnTe Detectors for Gamma-Ray Imaging Applications," *Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip.*, vol. 458, no. 1–2, pp. 297–309, Feb. 2001, doi: 10.1016/S0168-9002(00)00874-3.

[47] C. S. Levin, "New imaging technologies to enhance the molecular sensitivity of positron emission tomography," *Proc. IEEE*, vol. 96, no. 3, pp. 439–467, Mar. 2008, doi: 10.1109/JPROC.2007.913504.

[48] H. B. Barber, "Applications of Semiconductor Detectors to Nuclear Medicine," *Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip.*, vol. 436, no. 1–2, pp. 102–110, Oct. 1999, doi: 10.1016/S0168-9002(99)00605-1.

[49] J. F. Butler, C. L. Lingren, and F. P. Doty, "Cd/sub 1-x/Zn/sub x/Te Gamma Ray Detectors," *IEEE Trans. Nucl. Sci.*, vol. 39, no. 4, pp. 605–609, 1992, doi: 10.1109/23.159673.

[50] J. E. Toney *et al.*, "Uniformity of Cd1 − xZnxTe Grown by High-pressure Bridgman," *Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip.*, vol. 380, no. 1–2, pp. 132–135, Oct. 1996, doi: 10.1016/S0168-9002(96)00373-7.

[51] X. Zheng, Z. Cheng, M. J. Deen, and H. Peng, "Improving the Spatial Resolution in CZT Detectors Using Charge Sharing Effect and Transient Signal Analysis: Simulation Study," *Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip.*, vol. 808, pp. 60–70, Feb. 2016, doi: 10.1016/J.NIMA.2015.11.006.

[52] "Vereos Digital PET/CT." https://www.usa.philips.com/healthcare/product/HC882446/vereos-digital-petct-proven-accuracy-inspires-confidence (accessed Mar. 20, 2019).

[53] Philips Healthcare, "Truly Digital PET Imaging: Philips Proprietary Digital Photon Counting Technology.," 2016. https://philipsproductcontent.blob.core.windows.net/assets/20170523/ed2389b512114e209f2fa77c0145f267.pdf (accessed Mar. 29, 2019).

[54] "Discovery MI." https://www.gehealthcare.com/en/products/molecular-imaging/discovery-mi (accessed Mar. 20, 2019).

[55] GE Healthcare, "Lightburst Digital Detector: A White Paper.," 2017. https://static1.squarespace.com/static/5a25ac3c914e6b61f6b983d8/t/5a31f8a8ec212deac346e90d/1513224366471/LDD_whitepaper_Final+JB50594XX.pdf (accessed Mar. 20, 2019).

[56] "Biograph Vision." https://www.siemens-healthineers.com/molecular-imaging/pet-ct/biograph-vision (accessed Mar. 20, 2019).

[57] J. J. van Sluis *et al.*, "Performance Characteristics of the Digital Biograph Vision PET/CT System,"

J. Nucl. Med., p. jnumed.118.215418, Jan. 2019, doi: 10.2967/jnumed.118.215418.

[58] "Celesteion^{TM} PUREViSION Edition PET/CT." https://us.medical.canon/products/computed-tomography/celesteion/technology/ (accessed Mar. 20, 2019).

[59] J. S. Reddin *et al.*, "Performance Evaluation of the SiPM-based Siemens Biograph Vision PET/CT System," in *2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)*, Nov. 2018, pp. 1–5, doi: 10.1109/NSSMIC.2018.8824710.

[60] J. et al. Griesmer, "Philips Ingenuity TF PET/MR Calibration Overview.," 2012. https://philipsproductcontent.blob.core.windows.net/assets/20170523/bdc837ad1a64431b9834a77c01545929.pdf (accessed Mar. 20, 2019).

[61] N. Ojha *et al.*, "Philips Ingenuity TF PET/MR Performance," 2011. https://philipsproductcontent.blob.core.windows.net/assets/20170523/6e7323e7323e4410b61fa77c01545b76.pdf (accessed Mar. 20, 2019).

[62] Y. Qin, M. M. R. Howlader, M. J. Deen, Y. M. Haddara, and P. R. Selvaganapathy, "Polymer Integration for Packaging of Implantable Sensors," *Sensors Actuators B Chem.*, vol. 202, pp. 758–778, Oct. 2014, doi: 10.1016/J.SNB.2014.05.063.

[63] I. M. Antolović *et al.*, "Optical-stack Optimization for Improved SPAD Photon Detection Efficiency," in *Quantum Sensing and Nano Electronics and Photonics XVI*, Feb. 2019, vol. 10926, p. 99, doi: 10.1117/12.2511301.

[64] C. Jackson, L. Wall, K. O'Neill, B. McGarvey, and D. Herbert, "Through silicon via developments for silicon photomultiplier sensors," in *Optical Components and Materials XII*, Mar. 2015, vol. 9359, p. 93591A, doi: 10.1117/12.2076893.

[65] A. Arbat *et al.*, "High voltage vs. High Integration: a Comparison Between CMOS Technologies for SPAD Cameras," Aug. 2010, p. 77801G, doi: 10.1117/12.860482.

[66] S. Cova, A. Lacaita, and G. Ripamonti, "Trapping Phenomena in Avalanche Photodiodes on Nanosecond Scale," *IEEE Electron Device Lett.*, vol. 12, no. 12, pp. 685–687, Dec. 1991, doi: 10.1109/55.116955.

[67] W. J. Kindt and H. W. Van Zeijl, "Modelling and Fabrication of Geiger Mode Avalanche Photodiodes," *IEEE Trans. Nucl. Sci.*, vol. 45, no. 3, pp. 715–719, Jun. 1998, doi: 10.1109/23.682621.

[68] "The 10 ps challenge." https://the10ps-challenge.org/ (accessed Oct. 10, 2019).

[69] I. Rech, G. Luo, M. Ghioni, H. Yang, X. S. Xie, and S. Cova, "Photon-Timing Detector Module for Single-Molecule Spectroscopy With 60-ps Resolution," *IEEE J. Sel. Top. Quantum Electron.*, vol. 10, no. 4, pp. 788–795, Jul. 2004, doi: 10.1109/JSTQE.2004.833975.

[70] F. Acerbi *et al.*, "Characterization of Single-Photon Time Resolution: From Single SPAD to Silicon Photomultiplier," *IEEE Trans. Nucl. Sci.*, vol. 61, no. 5, pp. 2678–2686, Oct. 2014, doi: 10.1109/TNS.2014.2347131.

[71] C. L. F. Ma, M. J. Deen, L. E. Tarof, and J. C. H. Yu, "Temperature dependence of breakdown voltages in separate absorption, grading, charge, and multiplication InP/InGaAs avalanche photodiodes," *IEEE Trans. Electron Devices*, vol. 42, no. 5, pp. 810–818, May 1995, doi: 10.1109/16.381974.

[72] C. L. Forrest Ma, M. J. Deen, L. E. Tarof, and J. C. H. Yu, "Temperature Dependence of Breakdown Voltages in Separate Absorption, Grading, Charge, and Multiplication InP/InGaAs Avalanche Photodiodes," *IEEE Trans. Electron Devices*, vol. 42, no. 5, pp. 810–818, 1995, doi: 10.1109/16.381974.

[73] N. Faramarzpour, M. J. Deen, S. Shirani, and Q. Fang, "Fully Integrated Single Photon Avalanche Diode Detector in Standard CMOS 0.18-μm Technology," *IEEE Trans. Electron Devices*, vol. 55, no. 3, pp. 760–767, Mar. 2008, doi: 10.1109/TED.2007.914839.

[74] D. P. Palubiak, Z. Li, and M. J. Deen, "Afterpulsing Characteristics of Free-Running and Time-Gated Single-Photon Avalanche Diodes in 130-nm CMOS," *IEEE Trans. Electron Devices*, vol. 62, no. 11, pp. 3727–3733, Nov. 2015, doi: 10.1109/TED.2015.2475126.

[75] Z. Cheng, D. Palubiak, X. Zheng, M. J. Deen, and H. Peng, "Impact of Silicide Layer on Single

Photon Avalanche Diodes in a 130 nm CMOS Process," *J. Phys. D. Appl. Phys.*, vol. 49, no. 34, p. 345105, Sep. 2016, doi: 10.1088/0022-3727/49/34/345105.

[76] T. Leitner *et al.*, "Measurements and Simulations of Low Dark Count Rate Single Photon Avalanche Diode Device in a Low Voltage 180-nm CMOS Image Sensor Technology," *IEEE Trans. Electron Devices*, vol. 60, no. 6, pp. 1982–1988, Jun. 2013, doi: 10.1109/TED.2013.2259172.

[77] J. A. Richardson, E. A. G. Webster, L. A. Grant, and R. K. Henderson, "Scaleable Single-Photon Avalanche Diode Structures in Nanometer CMOS Technology," *IEEE Trans. Electron Devices*, vol. 58, no. 7, pp. 2028–2035, Jul. 2011, doi: 10.1109/TED.2011.2141138.

[78] H. Finkelstein, M. J. Hsu, S. Zlatanovic, and S. Esener, "Performance Trade-offs in Single-photon Avalanche Diode Miniaturization," *Rev. Sci. Instrum.*, vol. 78, p. 103103, 2007.

[79] W. J. Kindt, *Geiger Mode Avalanche Photodiode Arrays*. Delft, The Netherlands: Delft Univ. Press, 1999.

[80] J. A. Richardson, L. A. Grant, and R. K. Henderson, "Low Dark Count Single-photon Avalanche Diode Structure Compatible with Standard Nanometer Scale CMOS Technology," *IEEE PHOTONICS Technol. Lett.*, vol. 21, no. 14, pp. 1020–1022, 2009.

[81] C. Accarino *et al.*, "Low Noise and High Photodetection Probability SPAD in 180 nm Standard CMOS Technology," in *Proceedings - IEEE International Symposium on Circuits and Systems*, May 2018, vol. 2018-May, pp. 1–4, doi: 10.1109/ISCAS.2018.8351173.

[82] S. Lindner, S. Pellegrini, Y. Henrion, B. Rae, M. Wolf, and E. Charbon, "A High-PDE, Backside-Illuminated SPAD in 65/40-nm 3D IC CMOS Pixel With Cascoded Passive Quenching and Active Recharge," *IEEE Electron Device Lett.*, vol. 38, no. 11, pp. 1547–1550, 2017, doi: 10.1109/LED.2017.2755989.

[83] H. Xu, L. Pancheri, G.-F. D. Betta, and D. Stoppa, "Design and Characterization of a p+/n-well SPAD Array in 150nm CMOS Process," *Opt. Express*, vol. 25, no. 11, p. 12765, May 2017, doi: 10.1364/OE.25.012765.

[84] G. Humer, M. Peev, C. Schaeff, S. Ramelow, M. Stipčević, and R. Ursin, "A Simple and Robust Method for Estimating Afterpulsing in Single Photon Detectors," *J. Light. Technol.*, vol. 33, no. 14, pp. 3098–3107, Jul. 2015, doi: 10.1109/JLT.2015.2428053.

[85] G. Bonanno *et al.*, "Precision Measurements of Photon Detection Efficiency for SiPM Detectors," *Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip.*, vol. 610, no. 1, pp. 93–97, Oct. 2009, doi: 10.1016/J.NIMA.2009.05.117.

[86] B. Nakhkoob, M. M. Hella, and S. Ray, "High speed photodiodes in standard nanometer scale CMOS technology: a comparative study," *Opt. Express, Vol. 20, Issue 10, pp. 11256-11270*, vol. 20, no. 10, pp. 11256–11270, May 2012, doi: 10.1364/OE.20.011256.

[87] A. Bandyopadhyay, M. J. Deen, L. E. Tarof, and W. Clark, "A simplified approach to time-domain modeling of avalanche photodiodes," *IEEE J. Quantum Electron.*, vol. 34, no. 4, pp. 691–699, Apr. 1998, doi: 10.1109/3.663452.

[88] 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, doi: 10.1109/16.641363.

[89] A. Tosi *et al.*, "Fast-gated single-photon avalanche diode for extremely wide dynamic-range applications," in *Design and Quality for Biomedical Technologies II*, 2009, vol. 7170, pp. 124–134, doi: 10.1117/12.809205.

[90] Y. Ardeshirpour, M. J. Deen, and S. Shirani, "2-D CMOS Based Image Sensor System for Fluorescent Detection," in *Canadian Conference on Electrical and Computer Engineering 2004 (IEEE Cat. No.04CH37513)*, pp. 1441–1444, doi: 10.1109/CCECE.2004.1349674.

[91] H. Finkelstein, M. J. Hsu, and S. C. Esener, "Dual-junction Single-photon Avalanche Diode," *Electron. Lett.*, vol. 43, no. 22, p. 1228, 2007, doi: 10.1049/el:20072355.

[92] R. K. Henderson, E. A. G. Webster, and L. A. Grant, "A Dual-Junction Single-Photon Avalanche Diode in 130-nm CMOS Technology," *IEEE Electron Device Lett.*, vol. 34, no. 3, pp. 429–431, Mar.

2013, doi: 10.1109/LED.2012.2236816.

[93] Y. Ardeshirpour, M. J. Deen, and S. Shirani, “Evaluation of complementary metal-oxide semiconductor based photodetectors for low-level light detection,” *J. Vac. Sci. Technol. A Vacuum, Surfaces, Film.*, vol. 24, no. 3, p. 860, May 2006, doi: 10.1116/1.2190652.

[94] H. Finkelstein, M. J. Hsu, and S. C. Esener, “STI-Bounded Single-Photon Avalanche Diode in a Deep-Submicrometer CMOS Technology,” *IEEE Electron Device Lett.*, vol. 27, no. 11, pp. 887–889, Nov. 2006, doi: 10.1109/LED.2006.883560.

[95] M. Gersbach *et al.*, “A Low-noise Single-photon Detector Implemented in a 130 nm CMOS Imaging Process,” *Solid. State. Electron.*, vol. 53, no. 7, pp. 803–808, Jul. 2009, doi: 10.1016/J.SSE.2009.02.014.

[96] R. M. Field, J. Lary, J. Cohn, L. Paninski, and K. L. Shepard, “A low-noise, single-photon avalanche diode in standard 0.13 μm complementary metal-oxide-semiconductor process,” *Appl. Phys. Lett.*, vol. 97, no. 21, p. 211111, Nov. 2010, doi: 10.1063/1.3518473.

[97] C. Niclass, M. Gersbach, R. Henderson, L. Grant, and E. Charbon, “A 130-nm CMOS single-photon avalanche diode,” in *Optoelectronic Devices: Physics, Fabrication, and Application IV*, Sep. 2007, vol. 6766, p. 676606, doi: 10.1117/12.728878.

[98] J. M. Pavia, M. Scandini, S. Lindner, M. Wolf, and E. Charbon, “A 1x400 Backside-Illuminated SPAD Sensor With 49.7 ps Resolution, 30 pJ/Sample TDCs Fabricated in 3D CMOS Technology for Near-Infrared Optical Tomography,” *IEEE J. Solid-State Circuits*, vol. 50, no. 10, pp. 2406–2418, 2015, doi: 10.1109/JSSC.2015.2467170.

[99] E. Charbon, H.-J. Yoon, and Y. Maruyama, “A Geiger Mode APD Fabricated in Standard 65nm CMOS Technology,” in *2013 IEEE International Electron Devices Meeting*, Dec. 2013, pp. 27.5.1-27.5.4, doi: 10.1109/IEDM.2013.6724705.

[100] L. H. C. Braga *et al.*, “A CMOS mini-SiPM Detector with in-pixel Data Compression for PET Applications,” in *2011 IEEE Nuclear Science Symposium Conference Record*, 2011, pp. 548–552, doi: 10.1109/NSSMIC.2011.6154110.

[101] F. Zappa, S. Tisa, A. Gulinatti, A. Gallivanoni, and S. Cova, “Complete Single-photon Counting and Timing Module in a Microchip,” *Opt. Lett.*, vol. 30, no. 11, p. 1327, Jun. 2005, doi: 10.1364/OL.30.001327.

[102] D. Mosconi, D. Stoppa, L. Pancheri, L. Gonzo, and A. Simoni, “CMOS Single-Photon Avalanche Diode Array for Time-Resolved Fluorescence Detection,” in *2006 Proceedings of the 32nd European Solid-State Circuits Conference*, Sep. 2006, pp. 564–567, doi: 10.1109/ESSCIR.2006.307487.

[103] D. Shin, B. Park, Y. Chae, and I. Yun, “The Effect of a Deep Virtual Guard Ring on the Device Characteristics of Silicon Single Photon Avalanche Diodes,” *IEEE Trans. Electron Devices*, vol. 66, no. 7, pp. 2986–2991, Jul. 2019, doi: 10.1109/TED.2019.2913714.

[104] M. Dandin and P. Abshire, “High Signal-to-Noise Ratio Avalanche Photodiodes With Perimeter Field Gate and Active Readout,” *IEEE Electron Device Lett.*, vol. 33, no. 4, pp. 570–572, Apr. 2012, doi: 10.1109/LED.2012.2186112.

[105] Y. Wang, X. Jin, S. Cao, Y. Peng, and J. Luo, “Design and Measurement of Ring-Gate Single Photon Avalanche Diode with Low Dark Count Rate,” *IEEE Photonics J.*, vol. 12, no. 3, Jun. 2020, doi: 10.1109/JPHOT.2020.2993654.

[106] Z. Cheng, X. Zheng, D. Palubiak, M. J. Deen, and H. Peng, “A Comprehensive and Accurate Analytical SPAD Model for Circuit Simulation,” *IEEE Trans. Electron Devices*, vol. 63, no. 5, pp. 1940–1948, May 2016, doi: 10.1109/TED.2016.2537879.

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

[108] F. Zappa, A. Tosi, A. D. Mora, and S. Tisa, “SPICE Modeling of Single Photon Avalanche Diodes,” *Sensors Actuators A Phys.*, vol. 153, no. 2, pp. 197–204, Aug. 2009, doi: 10.1016/J.SNA.2009.05.007.

[109] 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, Apr. 1996, doi: 10.1364/AO.35.001956.

[110] V. Savuskan, M. Javitt, G. Visokolov, I. Brouk, and Y. Nemirovsky, "Selecting Single Photon Avalanche Diode (SPAD) Passive-Quenching Resistance: An Approach," *IEEE Sens. J.*, vol. 13, no. 6, pp. 2322–2328, Jun. 2013, doi: 10.1109/JSEN.2013.2253603.

[111] C. Veerappan *et al.*, "A 160×128 Single-photon Image Sensor with on-pixel 55ps 10b Time-to-digital Converter," in *2011 IEEE International Solid-State Circuits Conference*, Feb. 2011, pp. 312–314, doi: 10.1109/ISSCC.2011.5746333.

[112] A. J. González *et al.*, "The MINDView brain PET detector, feasibility study based on SiPM arrays," *Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip.*, vol. 818, pp. 82–90, May 2016, doi: 10.1016/j.nima.2016.02.046.

[113] A. J. Gonzalez *et al.*, "Initial Results of the MINDView PET Insert Inside the 3T mMR," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 3, no. 3, pp. 343–351, Aug. 2018, doi: 10.1109/trpms.2018.2866899.

[114] C. Niclass, M. Gersbach, R. Henderson, L. Grant, and E. Charbon, "A single photon avalanche diode implemented in 130-nm CMOS technology," *IEEE J. Sel. Top. Quantum Electron.*, vol. 13, no. 4, pp. 863–869, Jul. 2007, doi: 10.1109/JSTQE.2007.903854.

[115] L. Neri *et al.*, "Note: Dead Time Causes and Correction Method for Single Photon Avalanche Diode Devices," *Rev. Sci. Instrum.*, vol. 81, no. 8, p. 086102, Aug. 2010, doi: 10.1063/1.3476317.

[116] A. Rochas, P.-A. Besse, and R. S. Popovic, "Actively Recharged Single Photon Counting Avalanche Photodiode Integrated in an Industrial CMOS Process," *Sensors Actuators A Phys.*, vol. 110, no. 1–3, pp. 124–129, Feb. 2004, doi: 10.1016/J.SNA.2003.08.003.

[117] S. Tisa, A. Tosi, and F. Zappa, "Fully-integrated CMOS Single Photon Counter," *Opt. Express*, vol. 15, no. 6, pp. 2873–2887, Mar. 2007, doi: 10.1364/OE.15.002873.

[118] Z. Li and M. J. Deen, "Towards a Portable Raman Spectrometer Using a Concave Grating and a Time-gated CMOS SPAD," *Opt. Express*, vol. 22, no. 15, p. 18736, Jul. 2014, doi: 10.1364/OE.22.018736.

[119] G. Acconcia, I. Labanca, I. Rech, A. Gulinatti, and M. Ghioni, "Note: Fully integrated active quenching circuit achieving 100 MHz count rate with custom technology single photon avalanche diodes," *Rev. Sci. Instrum.*, vol. 88, no. 2, Feb. 2017, doi: 10.1063/1.4975598.

[120] M.-J. Lee and W.-Y. Choi, "Performance Optimization and Improvement of Silicon Avalanche Photodetectors in Standard CMOS Technology," *IEEE J. Sel. Top. Quantum Electron.*, vol. 24, no. 2, pp. 1–13, Mar. 2018, doi: 10.1109/JSTQE.2017.2754359.

[121] W. Jiang, Y. Chalich, R. Scott, and M. J. Deen, "Time-Gated and Multi-Junction SPADs in Standard 65 nm CMOS Technology," *IEEE Sens. J.*, 2021, doi: 10.1109/JSEN.2021.3063319.

[122] E. A. G. Webster, J. A. Richardson, L. A. Grant, D. Renshaw, and R. K. Henderson, "A Single-Photon Avalanche Diode in 90-nm CMOS Imaging Technology With 44% Photon Detection Efficiency at 690 nm," *IEEE Electron Device Lett.*, vol. 33, no. 5, pp. 694–696, May 2012, doi: 10.1109/LED.2012.2187420.

[123] L. Pancheri and D. Stoppa, "Low-noise single Photon Avalanche Diodes in 0.15 μm CMOS technology," in *2011 Proceedings of the European Solid-State Device Research Conference (ESSDERC)*, 2011, pp. 179–182, doi: 10.1109/ESSDERC.2011.6044205.

[124] C. Niclass and M. Soga, "A miniature Actively Recharged Single-photon Detector Free of Afterpulsing Effects with 6 ns Dead Time in a 0.18 μm CMOS Technology," *Electron Devices Meet. (IEDM), 2010 IEEE Int.*, pp. 14.3. 1--14.3. 4, 2010.

[125] C. Veerappan and E. Charbon, "A Substrate Isolated CMOS SPAD Enabling Wide Spectral Response and Low Electrical Crosstalk," *IEEE J. Sel. Top. Quantum Electron.*, vol. 20, no. 6, pp. 299–305, 2014.

[126] C. Veerappan and E. Charbon, "CMOS SPAD Based on Photo-Carrier Diffusion Achieving PDP >40% from 440 to 580 nm at 4 v Excess Bias," *IEEE Photonics Technol. Lett.*, vol. 27, no. 23, pp.

2445–2448, 2015, doi: 10.1109/LPT.2015.2468067.

[127] C. Veerappan and E. Charbon, “A Low Dark Count p-i-n Diode Based SPAD in CMOS Technology,” *IEEE Trans. Electron Devices*, vol. 63, no. 1, pp. 65–71, 2016, doi: 10.1109/TED.2015.2475355.

[128] D. Bronzi, S. Tisa, F. Villa, S. Bellisai, A. Tosi, and F. Zappa, “Fast sensing and quenching of CMOS SPADs for minimal afterpulsing effects,” *IEEE Photonics Technol. Lett.*, vol. 25, no. 8, pp. 776–779, 2013, doi: 10.1109/LPT.2013.2251621.

[129] D. Stoppa, D. Mosconi, L. Pancheri, and G. Lorenzo, “Single-Photon Avalanche Diode CMOS Sensor forTime-Resolved Fluorescence Measurements,” *IEEE Sens. J.*, vol. 9, no. 9, pp. 1084–1091, 2009.

[130] C. Bruschini, H. Homulle, and E. Charbon, “Ten years of biophotonics single-photon SPAD imager applications: retrospective and outlook,” in *Multiphoton Microscopy in the Biomedical Sciences XVII*, Feb. 2017, vol. 10069, p. 100691S, doi: 10.1117/12.2256247.

[131] W. Jiang, Y. Chalich, and M. J. Deen, “Sensors for Positron Emission Tomography Applications,” *Sensors*, vol. 19, no. 22. 2019, doi: 10.3390/s19225019.

[132] M. Alayed and M. J. Deen, “Time-Resolved Diffuse Optical Spectroscopy and Imaging Using Solid-State Detectors: Characteristics, Present Status, and Research Challenges,” *Sensors*, vol. 17, no. 9, p. 2115, Sep. 2017, doi: 10.3390/s17092115.

[133] A. R. Ximenes, P. Padmanabhan, M.-J. Lee, Y. Yamashita, D. N. Yaung, and E. Charbon, “A 256×256 45/65nm 3D-stacked SPAD-based direct TOF image sensor for LiDAR applications with optical polar modulation for up to 18.6dB interference suppression,” in *2018 IEEE International Solid - State Circuits Conference - (ISSCC)*, Feb. 2018, pp. 96–98, doi: 10.1109/ISSCC.2018.8310201.

[134] D. Chitnis and S. Collins, “A SPAD-Based Photon Detecting System for Optical Communications,” *J. Light. Technol.*, vol. 32, no. 10, pp. 2028–2034, May 2014, doi: 10.1109/JLT.2014.2316972.

[135] W. Shockley and W. T. Read, “Statistics of the recombinations of holes and electrons,” *Phys. Rev.*, vol. 87, no. 5, pp. 835–842, Sep. 1952, doi: 10.1103/PhysRev.87.835.

[136] E. O. Kane, “Theory of tunneling,” *J. Appl. Phys.*, vol. 32, no. 1, pp. 83–91, Jan. 1961, doi: 10.1063/1.1735965.

[137] C. L. F. Ma, M. J. Deen, and R. H. S. Hardy, “Excess currents as a result of trap-assisted tunneling in double barrier resonant tunneling diodes,” *https://doi.org/10.1139/p92-161*, vol. 70, no. 10–11, pp. 1005–1012, Oct. 2011, doi: 10.1139/P92-161.

[138] E. Simoen, B. Dierickx, C. L. Claeys, and G. J. Declerck, “Explaining the Amplitude of RTS Noise in Submicrometer MOSFET’s,” *IEEE Trans. Electron Devices*, vol. 39, no. 2, pp. 422–429, 1992, doi: 10.1109/16.121702.

[139] M. J. Deen, S. L. Rumyantsev, and M. Schroter, “On the origin of 1/f noise in polysilicon emitter bipolar transistors,” *J. Appl. Phys.*, vol. 85, no. 2, pp. 1192–1195, Jan. 1999, doi: 10.1063/1.369256.

[140] M. Sandén, O. Marinov, M. J. Deen, and M. Östling, “A new model for the low-frequency noise and the noise level variation in polysilicon emitter BJTs,” *IEEE Trans. Electron Devices*, vol. 49, no. 3, pp. 514–520, Mar. 2002, doi: 10.1109/16.987124.

[141] S. R. Li, W. McMahon, Y. L. R. Lu, and Y. H. Lee, “RTS noise characterization in flash cells,” *IEEE Electron Device Lett.*, vol. 29, no. 1, pp. 106–108, Jan. 2008, doi: 10.1109/LED.2007.910776.

[142] I. H. Hopkins and G. R. Hopkinson, “Random telegraph signals from proton-irradiated CCDs,” *IEEE Trans. Nucl. Sci.*, vol. 40, no. 6, pp. 1567–1574, 1993, doi: 10.1109/23.273552.

[143] M. J. Deen, S. Majumder, O. Marinov, and M. M. El-Desouki, “Random telegraph signal noise in CMOS active pixel sensors,” in *Proceedings of the IEEE 21st International Conference on Noise and Fluctuations, ICNF 2011*, 2011, pp. 208–211, doi: 10.1109/ICNF.2011.5994302.

[144] M. A. Karami, C. Niclass, and E. Charbon, “Random Telegraph Signal in Single-Photon Avalanche Diodes,” in *Int. Image Sensor Workshop*, 2009, pp. 1–4, [Online]. Available: https://www.imagesensors.org/Past Workshops/2009 Workshop/2009 Papers/062_paper_karami_tudelft_spad.pdf.

[145] M. A. Karami, L. Carrara, C. Niclass, M. Fishburn, and E. Charbon, “RTS Noise Characterization in

Single-Photon Avalanche Diodes," *IEEE Electron Device Lett.*, vol. 31, no. 7, pp. 692–694, Jul. 2010, doi: 10.1109/LED.2010.2047234.

[146] V. Agarwal, A. J. Annema, S. Dutta, R. J. E. Hueting, L. K. Nanver, and B. Nauta, "Random Telegraph Signal phenomena in avalanche mode diodes: Application to SPADs," in *2016 46th European Solid-State Device Research Conference (ESSDERC)*, Sep. 2016, pp. 264–267, doi: 10.1109/ESSDERC.2016.7599636.

[147] V. Agarwal, A.-J. Annema, S. Dutta, R. J. E. Hueting, L. K. Nanver, and B. Nauta, "Random Telegraph Signal Phenomena in Ultra Shallow p^+n Silicon Avalanche Diodes," *IEEE J. Electron Devices Soc.*, vol. 6, pp. 642–652, 2018, doi: 10.1109/JEDS.2018.2835153.

[148] S. T. Hsu, "Bistable noise in p-n junctions," *Solid State Electron.*, vol. 14, no. 6, pp. 487–497, Jun. 1971, doi: 10.1016/0038-1101(71)90059-1.

[149] O. Marinov, M. J. Deen, and J. A. J. Tejada, "Theory of microplasma fluctuations and noise in silicon diode in avalanche breakdown," *J. Appl. Phys.*, vol. 101, no. 6, p. 064515, Mar. 2007, doi: 10.1063/1.2654973.

[150] R. J. McIntyre, "Theory of microplasma instability in silicon," *J. Appl. Phys.*, vol. 32, no. 6, pp. 983–995, Jun. 1961, doi: 10.1063/1.1736199.

[151] J. R. Srour and J. W. Palko, "Displacement damage effects in irradiated semiconductor devices," *IEEE Trans. Nucl. Sci.*, vol. 60, no. 3, pp. 1740–1766, 2013, doi: 10.1109/TNS.2013.2261316.

[152] F. Di Capua *et al.*, "Random Telegraph Signal in Proton Irradiated Single-Photon Avalanche Diodes," *IEEE Trans. Nucl. Sci.*, vol. 65, no. 8, pp. 1654–1660, Aug. 2018, doi: 10.1109/TNS.2018.2814823.

[153] M. A. Karami, A. Pil-Ali, and M. R. Safaee, "Multistable defect characterization in proton irradiated single-photon avalanche diodes," *Opt. Quantum Electron.*, vol. 47, no. 7, pp. 2155–2160, Jul. 2015, doi: 10.1007/s11082-014-0089-7.

[154] D. P. Palubiak, Z. Li, and M. J. Deen, "Afterpulsing Characteristics of Free-Running and Time-Gated Single-Photon Avalanche Diodes in 130-nm CMOS," *IEEE Trans. Electron Devices*, vol. 62, no. 11, pp. 3727–3733, Nov. 2015, doi: 10.1109/TED.2015.2475126.

[155] M. J. Kirton, M. J. Uren, S. Collins, M. Schulz, A. Karmann, and K. Scheffer, "Individual defects at the Si:SiO2 interface," *Semicond. Sci. Technol.*, vol. 4, no. 12, p. 1116, Dec. 1989, doi: 10.1088/0268-1242/4/12/013.

[156] M. Campajola, F. Di Capua, L. Gasparini, and E. Sarnelli, "Radiation induced degradation in a 150-nm CMOS SPADs device," *J. Instrum.*, vol. 15, no. 08, pp. C08017–C08017, Aug. 2020, doi: 10.1088/1748-0221/15/08/c08017.

[157] A. C. Giudice, M. Ghioni, S. Cova, and F. Zappa, "A process and deep level evaluation tool: Afterpulsing in avalanche junctions," in *European Solid-State Device Research Conference*, 2003, pp. 347–350, doi: 10.1109/ESSDERC.2003.1256885.

[158] B. Korzh, T. Lunghi, K. Kuzmenko, G. Boso, and H. Zbinden, "Afterpulsing studies of low-noise InGaAs/InP single-photon negative-feedback avalanche diodes," *J. Mod. Opt.*, vol. 62, no. 14, pp. 1151–1157, Aug. 2015, doi: 10.1080/09500340.2015.1024294.

[159] D. Palubiak, "CMOS Single-photon Avalanche Diodes and Time-to-digital Converters for Time-resolved Fluorescence Analysis," Ph.D. Thesis, Department of Electrical and Computer Engineering, McMaster University, 2015.

[160] S. M. Sze and K. K. Ng, *Physics of Semiconductor Devices*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2006.

[161] O. A. Amusan, "Analysis of Single Event Vulnerabilities in a 130 nm CMOS technology," Master Thesis, School of Engineering, Vanderbilt University, 2006.

[162] W. Jiang and M. J. Deen, "Random Telegraph Signal in n^+/p-Well CMOS Single-Photon Avalanche Diodes," *IEEE Trans. Electron Devices*, pp. 1–6, 2021, doi: 10.1109/TED.2021.3070557.

[163] E. Kamrani, F. Lesage, and M. Sawan, "Premature edge breakdown prevention techniques in CMOS APD fabrication," *2012 IEEE 10th Int. New Circuits Syst. Conf. NEWCAS 2012*, pp. 345–348, 2012, doi: 10.1109/NEWCAS.2012.6329027.

[164] A. Katz, T. Blank, A. Fenigstein, T. Leitner, and Y. Nemirovsky, "Active-Reset for the N+P Single-Ended SPAD Used in the NIR LiDAR Receivers," *IEEE Trans. Electron Devices*, vol. 66, no. 12, pp. 5191–5195, Dec. 2019, doi: 10.1109/TED.2019.2946015.

[165] E. Kamrani, F. Lesage, and M. Sawan, "Efficient premature edge breakdown prevention in SiAPD fabrication using the standard CMOS process," *Semicond. Sci. Technol.*, vol. 28, no. 4, p. 045008, Apr. 2013, doi: 10.1088/0268-1242/28/4/045008.

[166] A. C. Ulku, C. Bruschini, I. M. Antolovic, S. Weiss, X. Michalet, and E. Charbon, "Phasor-based widefield FLIM using a gated 512×512 single-photon SPAD imager," in *Multiphoton Microscopy in the Biomedical Sciences XIX*, Feb. 2019, vol. 10882, p. 21, doi: 10.1117/12.2511148.

[167] F. Nolet *et al.*, "Quenching Circuit and SPAD Integrated in CMOS 65 nm with 7.8 ps FWHM Single Photon Timing Resolution," *Instruments*, vol. 2, no. 4, 2018, doi: 10.3390/instruments2040019.

[168] R. Enne, B. Steindl, M. Hofbauer, and H. Zimmermann, "Fast Cascoded Quenching Circuit for Decreasing Afterpulsing Effects in 0.35-m CMOS," *IEEE Solid-State Circuits Lett.*, vol. 1, no. 3, pp. 62–65, Mar. 2018, doi: 10.1109/LSSC.2018.2827881.

[169] Y. Xu, J. Lu, and Z. Wu, "A Compact High-Speed Active Quenching and Recharging Circuit for SPAD Detectors," *IEEE Photonics J.*, vol. 12, no. 5, pp. 1–8, Aug. 2020, doi: 10.1109/jphot.2020.3015872.

[170] K. Bach, A. Voerckel, and M. Franke, "Integrated photodetectors in CMOS chips and their spectral sensitivity," *ECS Trans.*, vol. 39, pp. 265–274, Jan. 2011, doi: 10.1149/1.3615202.

[171] F. Ceccarelli, G. Acconcia, A. Gulinatti, M. Ghioni, and I. Rech, "Fully Integrated Active Quenching Circuit Driving Custom-Technology SPADs With 6.2-ns Dead Time," *IEEE Photonics Technol. Lett.*, vol. 31, no. 1, pp. 102–105, Jan. 2019, doi: 10.1109/LPT.2018.2884740.

[172] J. Rhim *et al.*, "Monolithically-Integrated Single-Photon Avalanche Diode in a Zero-Change Standard CMOS Process for Low-Cost and Low-Voltage LiDAR Application," *Instruments*, vol. 3, no. 2, p. 33, Jun. 2019, doi: 10.3390/instruments3020033.

[173] Y. Chalich, "Design of Time-gated CMOS SPADs towards High-performance Imagers," Master Thesis, Department of Electrical and Computer Engineering, McMaster University, 2020.

[174] M. Moreno-Garcia, L. Pancheri, M. Perenzoni, R. del Rio, O. Guerra Vinuesa, and A. Rodriguez-Vazquez, "Characterization-Based Modeling of Retriggering and Afterpulsing for Passively Quenched CMOS SPADs," *IEEE Sens. J.*, vol. 19, no. 14, pp. 5700–5709, Jul. 2019, doi: 10.1109/JSEN.2019.2903937.

[175] A. Muntean *et al.*, "Blumino: the first fully integrated analog SiPM with on-chip time conversion," *IEEE Trans. Radiat. Plasma Med. Sci.*, pp. 1–1, Dec. 2020, doi: 10.1109/trpms.2020.3045081.

Appendix A
Copyright Permission

Sensors for Positron Emission Tomography Applications, Sensors

From: Sensors <sensors@mdpi.com>

Date: Mon, Jul 12, 2021 at 4:57 AM

Subject: Re: Request for permission to reuse the paper in my PhD thesis

To: WEI JIANG <jiangw35@mcmaster.ca>

Cc: sensors@mdpi.com

Dear Wei,

Thank you very much for your email. Please note that Sensors is an open access journal and we follow the conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/), which provides that published material can be re - used without obtaining permission as long as a correct citation to the original publication is given.

If you would like to cite several parts of this publication, please directly contact the authors. The copyrights belong to you.

Hope this does help. Feel free to let me know if you have any other questions.

Kind regards,
Jayleen Chen
Managing Editor

On 2021/7/9 22:49, WEI JIANG wrote:
> Dear Sensors editorial team,
>
> I am writing to request the permission to reuse the contents of the following paper in my
> Ph.D. dissertation. Thank you for your consideration and support.
>
> W. Jiang, Y. Chalich, J.M. Deen, Sensors for Positron Emission
> Tomography Applications, Sensors. 19 (2019). doi:10.3390/s19225019.
> Link: https://www.mdpi.com/1424-8220/19/22/5019#
> <https://www.mdpi.com/1424-8220/19/22/5019#>
>
> Regards,
> Wei Jiang
> Ph.D. Candidate
> School of Biomedical Engineering, McMaster University.

Random Telegraph Signal in n^+ /p-Well CMOS Single-Photon Avalanche Diodes

Author: Wei Jiang
Publication: IEEE Transactions on Electron Devices
Publisher: IEEE
Date: June 2021
Copyright © 2021, IEEE
Thesis / Dissertation Reuse
The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:
Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.
Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:
1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.
If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

Time-Gated and Multi-Junction SPADs in Standard 65 nm CMOS Technology

Author: Wei Jiang

Publication: IEEE Sensors Journal

Publisher: IEEE

Date: 15 May15, 2021

Copyright © 2021, IEEE

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.

2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.

3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]

2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.

3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new

collective works for resale or redistribution, please go to
http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.
If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

Differential Quench and Reset Circuit for Single-Photon Avalanche Diodes

Author: Wei Jiang
Publication: Journal of Lightwave Technology
Publisher: IEEE
Date: September 2021
Copyright © 2021, IEEE
Thesis / Dissertation Reuse
The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:
Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:
1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.
Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.
If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.