DESIGN OF CMOS SPADS TOWARDS HIGH-PERFORMANCE IMAGERS

DESIGN OF TIME-GATED CMOS SPADS TOWARDS HIGH-PERFORMANCE IMAGERS

By
YAMN CHALICH, H.B.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree of Master of Applied Science

McMaster University
Hamilton, Ontario

Master of Applied Science (2020)
(Electrical and Computer Engineering)

TITLE: Design of Time-Gated CMOS SPADs Towards High-Performance Imagers

AUTHOR: Yamn Chalich,
H.B.Sc. University of Toronto, Scarborough, Canada

SUPERVISOR: Dr. M. Jamal Deen

NUMBER OF PAGES: xxi, 124

Lay Abstract

The CMOS (Complementary Metal-Oxide-Semiconductor) technology process has made accessible the development of highspeed digital circuitry and sophisticated image sensor arrays. SPADs (Single-Photon Avalanche Diodes) capable of detecting single photons for ultimate sensitivity can be fabricated in a standard CMOS process and paves the way for low-cost, next-generation, high-performance, and miniature imaging solutions. In this thesis, a time-gated SPAD implemented in a smaller CMOS technology node not typically investigated in the literature was tested and showed strong overall performance despite suffering from increased noise. A multi-junction SPAD capable of increased photon detecting capabilities was also explored and a time-gated, dual-junction solution was designed and simulated to take advantage of both structures while overcoming the difficulties of integrating them. Finally, a low-cost, highspeed camera is built with standard commodity hardware and a CMOS image sensor to demonstrate the capabilities and ease of developing high-performance imagers comparable to or better than commercial solutions.

Abstract

SPAD (Single-Photon Avalanche Diode) sensors, capable of detecting down to the single photon level for ultimate sensitivity, have shown great promise as the photodetector of choice for next-generation devices used in positron emission tomography, fluorescence lifetime imaging, light detection and ranging, and more. SPAD fabrication has shifted recently towards custom technologies, 3D stacked designs, and post-processing steps (micro-lenses) to improve performance at the expense of increased cost and complexity. This thesis explores time-gating and multi-junction techniques to improve SPAD performance in standard planar CMOS (Complementary Metal-Oxide-Semiconductor) processes to take advantage of their potential for monolithic integration with other mixed-signal circuitry for simple, low-cost, high-performance imaging solutions.

An unbuffered triple-junction SPAD was fabricated to investigate the potential for wavelength distinction, however the top two junctions (n+/p-well and p-well/deep n-well) showed excessive noise, and the deepest junction exhibited a similar spectral response to the top junction, potentially due to a large influence from the process layers over the active area. A time-gated SPAD pixel based on the top junction was also designed and fabricated in the TSMC standard 65 nm CMOS process with a fill-factor of 28.6%. At an excess voltage of 300 mV, it achieved a peak photon detection efficiency of ~3.5% at 440 nm, <1% afterpulsing probability for hold-off times >22ns, and <200 ps timing jitter.

Lastly, the potential for high-performance CMOS imaging systems was demonstrated through the development of a prototype open-source, low-cost, highspeed camera built with standard commodity hardware. It achieved 211 frames per second (fps) at its maximum resolution of 1280x1024, and up to 2329 fps at a 256x256 resolution, with a cost well under $1000 USD. It was found to be very competitive to current low-cost, commercial highspeed cameras using a new figure-of-merit comparison and was tested for biological microscopy applications involving *C. elegans* worms.

Acknowledgements

I would like to express my gratitude to Dr. M. Jamal Deen for giving me the opportunity to research and study under his guidance. He has taught me how to ask meaningful questions, engage and communicate with those in other professions, and set goals to do great work. I have grown not only as a researcher but as an individual thanks to him.

It was also a memorable experience working with the other members of our research group: Wei Jiang, Sumit Majumder, Abu Ilius Faisal, Arif Ul Alam, Ahmed Elsharabasy, Hythm Afifi, Mahdi Naghshvarianjahromi, Ryan Scott, and Si Pan. They have my appreciation for their valuable suggestions, discussions, and feedback on my research. Wei Jiang deserves special mention for acting as a mentor throughout my studies and providing me much needed support to see my work through.

Furthermore, I am grateful to Dr. Nicola Nicolici and his student Alex Lao for their fruitful discussions in digital design, Dr. Mrwan Alayed for his expertise and suggestions on my research, and Tyler Ackland and Joe Peric for their technical assistance. My other committee members Dr. Shahram Shirani and Dr. Qiyin Fang also have my gratitude for their valuable advice on my research and for taking the time to review my thesis.

Last but not the least, I am grateful for the support that I have received from my friends and family. I sincerely thank my parents for their unconditional love and encouragement during this academic journey. This thesis is dedicated to them.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ADC	Analog-to-Digital Converter
AP	Afterpulsing
APD	Avalanche Photodiode
AQR	Active Quench and Reset
AR	Anti-reflective
ASIC	Application-Specific Integrated Circuit
AXI	Advanced eXtensible Interface
BCD	Bipolar-CMOS-DMOS
BSI	Backside Illumination
BTBT	Band-to-Band Tunnelling
CCD	Charge-Coupled Device
CIS	CMOS Image Sensor
CMOS	Complementary Metal-Oxide-Semiconductor
CPU	Central Processing Unit
CW	Continuous Wave
DCR	Dark Count Rate
DDR	Double Date Rate
DFF	Data Flip-Flop
DJTG	Dual-Junction Time-Gated
DMOS	Double-Diffused Metal-Oxide-Semiconductor
DNW	Deep n-well
DOI	Diffuse Optical Imaging
DOS	Diffuse Optical Spectroscopy
DOT	Diffuse Optical Tomography
DRAM	Dynamic Random-Access Memory
dSiPM	Digital Silicon Photomultiplier
DSM	Deep Sub-Micron
DToF	Distribution Time-of-Flight
EHP	Electron-Hole Pair
FB	Frame Buffer
FD	Frequency Domain
FF	Fill Factor or Flip-Flop

FIFO	First-In First-Out
FLIM	Fluorescence Lifetime Imaging Microscopy
FMC	FPGA Mezzanine Card
fNIRS	Functional Near Infrared Spectroscopy
FoM	Figure-of-Merit
FPGA	Field-Programmable Gate Array
FR	Free-Running
FSM	Finite-State Machine
G-APD	Geiger-mode Avalanche Photodiode
GaAs	Gallium Arsenide
GFP	Green Fluorescent Protein
GR	Guard Ring
HDL	Hardware Descriptive Language
HV	High Voltage
IAT	Inter-Arrival Time
InP	Indium Phosphide
IoT	Internet of Things
IRF	Instrument Response Function
LiDAR	Light Detection and Ranging
LUT	Look-Up Table
LVDS	Low-Voltage Differential Signalling
MCP	Microchannel Plate
MP	Megapixels
MRI	Magnetic Resonance Imaging
NA	Numerical Aperture
ND	Neutral Density
NIR	Near Infrared
OLED	Organic Light-Emitting Diode
PB	Pushbutton
PCB	Printed Circuit Board
PDE	Photon Detection Efficiency
PEB	Premature Edge Breakdown
PGM	Portable Grey Map
PL	Programmable Logic
PMOD	Peripheral Module
PQR	Passive Quench and Reset
PS	Processing System

QE	Quantum Efficiency
QKD	Quantum Key Distribution
PET	Positron Emission Tomography
PLL	Phase-Locked Loop
PMT	Photomultiplier Tube
PS-MCP	Position-Sensitive Microchannel Plate
SDD	Source-Detector Distance
SiPD	Silicon Photodetector
SiPM	Silicon Photomultiplier
SNR	Signal-to-Noise Ratio
SoC	System-on-Chip
SPAD	Single-Photon Avalanche Diode
SPI	Serial Peripheral Interface
SRH	Shockley-Read-Hall
STI	Shallow Trench Isolation
SW	Switches
TAT	Trap-Assisted Tunnelling
TBI	Traumatic Brain Injury
TDC	Time-to-Digital Converter
TG	Time-Gated
TJ	Triple-Junction
ToF	Time-of-Flight
TR	Time-Resolved
TSV	Through-Silicon Vias
TTR	Transit Time Response
TTS	Transit Time Spread
VCSEL	Vertical-Cavity Surface-Emitting Laser

List of Symbols

$\alpha_{n(p)}$	Impact ionization coefficient for electrons (holes) [1/m]
$I_{n(p)}$	Electron (Hole) current [A]
M_n	Multiplication factor
Φ_{SPAD}	Number of photons per second counted by a SPAD [1/s]
Φ_{IN}	Number of photons per second incident on a SPAD [1/s]
α	Absorption coefficient [1/m]
T	Absolute temperature [K]
k	Boltzmann's constant [eV/ K]
q	Electron charge [C]
c	Speed of light [m/s]
h	Planck's constant [eV·s]
$\hbar$	Reduced Planck's constant [eV·s]
n_i	Intrinsic carrier concentration [1/cm^3]
n	Non-equilibrium electron concentration [1/cm^3]
p	Non-equilibrium hole concentration [1/cm^3]
$\tau_{n(p)}$	Electron (Hole) lifetimes [µs]
N_t	Trap concentration [1/cm^3]
E_i	Intrinsic Fermi level [eV]
E_t	Recombination center energy level [eV]
$m^*_{n(p)}$	Effective electron (hole) mass [kg]
m^*_t	Effective mass of tunneling electrons for silicon [kg]
$\sigma_{n(p)}$	Capture cross area of electron (hole) [cm^2]
Γ	Field-effect enhancement factor
P_{pair}	Total triggering probability from both electrons and holes
S	Depletion layer area [µm^2]
$G_{SRH,TAT}$	Trap-assisted thermal generation and tunneling generation rate [1/cm^3·s]
$E(x)$	Local electric field strength at depth position *x* [V/cm]
P_{AP}	Afterpulsing probability (%)
P_{AV}	Avalanche probability (%)
W_e	Effective depletion width [µm]
C	Capacitance [F] or Cost [$ USD]
DCR_0	Dark count rate not including afterpulsing [Hz or cps]
DCR_{PR}	Primary dark count rate extracted from IAT exponential fit [Hz or cps]
V_{SPAD}	Voltage applied to the SPAD [V]

V_{BD}	Breakdown voltage of the SPAD [V]
V_{EX}	Excess voltage above breakdown [mV]
V_{out}	Output voltage of the SPAD [V]
V_{reset}	Voltage that re-arms the SPAD at the cathode [V]
V_c	Cathode Voltage of the SPAD [V]
E_g	Bandgap energy [eV]
E_a	Activation energy [eV]
T_{on}	Gate window [ns]
T_g	Gating Period [ns]
r_{opt}	Optical resolution [µm]
M	Magnification
RES_{max}	Maximum resolution at set framerate [fps]
FPS_{max}	Maximum framerate at set resolution [fps]
N_{pixels}	Number of pixels
RT	Recording time [s]
P	Power [W]

Declaration of Academic Achievement

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

- **Chapters 1 and 2:** I conducted the literature review and summarized the research results.
- **Chapter 3:** I designed the SPAD structures and performed the majority of the simulations and layout of the SPAD pixel structures with assistance from Wei Jiang on performing some simulations and layout to meet the fabrication deadline.
- **Chapter 4:** I designed the printed circuit board used to test the designed SPADs and Tyler Ackland assembled the board. I carried out all the measurements reported in characterizing the SPADs.
- **Chapter 5:** I wrote the source code used to interface with and program the image sensor with guidance from Dr. Nicolici and his student Alex Lao. I also designed the printed circuit board for the image sensor and Tyler Ackland assembled the board. The application of the camera involving a nematode system was carried out in collaboration with Dr. Bhagwati Gupta's research group in the Department of Biology, McMaster University.

Chapter 1
Introduction

1.1. Single Photon Detection

A photon is a quantum, or smallest discrete quantity, of electromagnetic radiation. First introduced by Planck in 1900 in an effort to theoretically explain blackbody radiation measurements, it was later used by Einstein in 1905 to explain the photoelectric effect [1], a phenomenon that was key to the emergence of photodetectors. When a photon impinges on a semiconductor or material with a bandgap with energy greater than its bandgap energy, an electron-hole pair (EHP) is generated. This can create current when under an externally applied electric field, and with a proper gain or amplification mechanism, single photon detectors are able to produce an electrical signal from the absorption of just one photon.

Single photons became detectable thanks to the advent of commercial photomultiplier tubes (PMTs) in 1936 [2], with high performance sophisticated PMTs becoming commercially available from the 1960s. It was the detector of choice for a long time due to its great sensitivity, picosecond temporal response, and low noise per unit area. Research has since shifted towards solid-state solutions fabricated in Silicon CMOS (complementary metal-oxide-semiconductor) technology to offer lower power consumption, lower operating voltages, much smaller size, less fragility and susceptibility to magnetic fields, and considerably lower cost [3]. One such photodetector is the avalanche photodiode (APD), which became available to experimenters in the 1990s [4], [5]. It is only capable of detecting very weak optical signals when operated in the linear mode. It is not until they are biased in the Geiger mode (above breakdown) that they become capable of single-photon detection and are thus known as single-photon avalanche diodes (SPADs). SPAD arrays known as silicon photomultipliers (SiPMs) have since been a topic of great research interest as they have evolved to have some of the best-in-class performance and time-

resolved sensing performance, as well as the benefits of being more easily integrable with analog and digital circuitry in planar silicon processes [6].

SPADs for single-photon detection has shown promise in a wide field of applications [1], [3] including, but not limited to, Fluorescence Lifetime Imaging Microscopy (FLIM) [7], Positron Emission Tomography (PET) [8], Diffuse Optical Imaging (DOI) [9], Raman spectroscopy [10], Light Detection and Ranging (LiDAR) applications [11], and Quantum Key Distribution (QKD) [12]. In fact, when compared to emerging detector technologies, a shallow junction Si SPAD held a higher figure-of-merit (FoM) than one based on superconducting nanowires and held the next highest FoM compared to a quantum dot based detector, both requiring expensive and sophisticated cooling techniques [12]. Despite this, SPAD structures, in-pixel circuitry and array integration continue to be researched as there is still much room for improvement, especially in less optimized standard CMOS processes. Many groups worldwide have tackled design and performance issues such as thermal noise, tunnelling effects, detection efficiency and timing performance. Furthermore, these performance parameters can be dependent on each other, meaning certain design choices and trade-offs are required based on the proposed application.

In the following section, the operating principles and performance parameters of PMTs, APDs, and SPADs are described in more detail, followed by a comparison of the technologies to justify SPADs as the technology of choice and focus for this thesis.

1.2. Evolution of Single Photon Detectors

1.2.1. Photomultiplier Tubes

A photomultiplier tube (PMT), shown in Figure 1-1, utilizes a photocathode coated with a photosensitive material that generates electrons through the photoelectric effect. When photons impact the photocathode, electrons are generated from the photosensitive material which travel through a vacuum glass tube that contains several dynodes set at ever increasing voltages. Due to the strong electric fields, the electrons at each stage obtain enough energy to generate more electrons through secondary emission at the next dynode.

Capacitors help maintain the voltage on the dynodes during the final few multiplication events. After several stages, a final multiplication gain on the order of 10^6 can be achieved upon reaching the anode and a large detectable current pulse is generated which is distinguishable from noise [8].

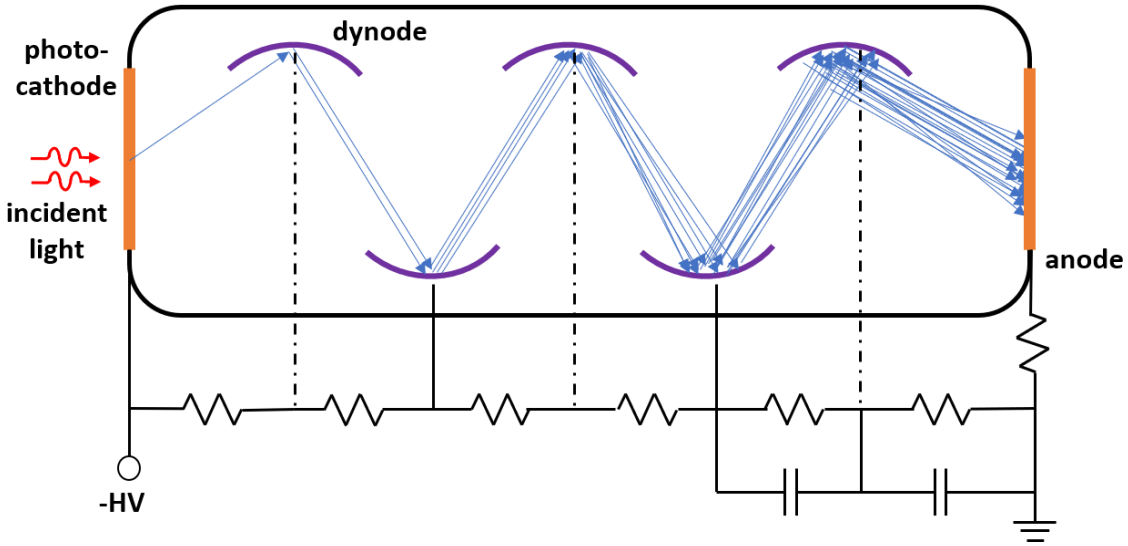


Figure 1-1: Basic structure of a photomultiplier tube (PMT) with the cathode at some negative high voltage (-HV) and the anode at ground.

Another utilization of secondary electron emission to detect single photons is found through microchannel plate (MCP) PMTs. MCPs contain many micron-sized channels (usually 3-10 μm in diameter) that function similarly to the glass tubes of a conventional PMT containing dynodes. The channels are coated with a conductive emissive dynode material which generate electrons through secondary emission as the photogenerated (and subsequently generated) electrons bounce between the inner walls of the channel. These electrons ultimately hit the anode plate as with a typical PMT to generate a current pulse and two or three MCPs can be placed in series to achieve higher gains. If the anode plate is replaced by an array of separate anodes, the device is then called a position-sensitive PMT (PS-PMT) since positional information can be retrieved from which anode generated the pulse [8]. A PS-PMT containing MCPs is illustrated in Figure 1-2.

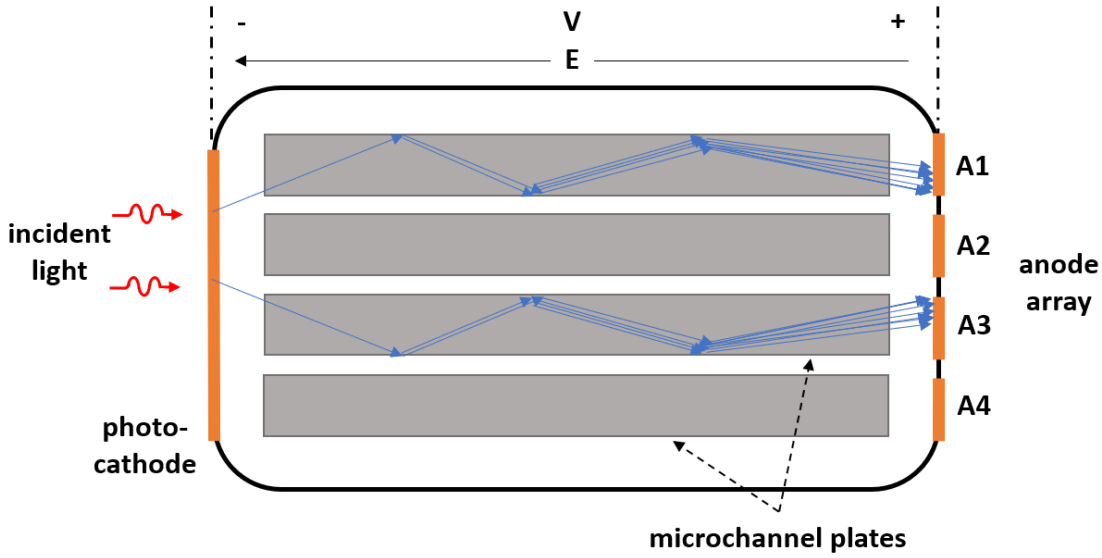


Figure 1-2: Illustration of a position sensitive microchannel plate (PS-MCP) where the anode is segmented instead of singular.

In the ideal case, the PMT only produces output pulses due to photon detection, but it is possible that even when kept in the dark, electrons emitted from the photocathode or dynodes by thermionic or field emission are multiplied and generate a false output. This noise is characterized by a dark count rate (DCR) which is the number of pulses (counts) generated per second when not illuminated by light. While these dark counts are undesirable, PMTs had the advantage of low noise and lower sensitivity to temperature variations owing to the vacuum tube construction. PMTs can have a DCR of tens of counts per second depending on the cathode material and dynode chain design [13].

There exists some variation in the time it takes between a photon creating a photoelectron at the photocathode and an output current pulse being detected at the anode. The average of the distribution created by this variation is known as the electron transit time response (TTR) and the standard deviation of it is the transit time spread (TTS). The TTS is also known as time resolution or timing jitter [13], an important term used also for the timing performance of SPADs since it represents the uncertainty in the arrival time of the photon. In MCP PMTs, a high timing jitter performance as short as 25 ps can be achieved by placing the first microchannel plate very close to the cathode [14].

Another important performance parameter that is common among photodetectors is their photon detection efficiency (PDE). For a PMT, this is primarily determined by the internal quantum efficiency (QE) of the photocathode material and the collection efficiency of the first dynode stage. The QE is the ratio of the number of generated EHPs to the number of incident photons and is wavelength dependent, with a typical value of ~25% for a PMT [15]. The probability that these photoelectrons land on the first dynode stage is coined the collection efficiency. Due to the multiplication of electrons by secondary emission at the first stage, an output signal is still possible even if electrons are lost at later dynode stages due to their trajectories. Thus, the collection efficiency at latter dynode stages are not considered as important and only the first dynode stage is considered [16].

In spite of the relatively good PDE, low noise and low time jitter, PMTs requires very high voltages in excess of 1000 V, are sensitive to magnetic fields, and are bulky with a complicated mechanical structure. The high voltage requirements and costly assembly of PMTs limit their usability in low-cost, compact, or portable applications. The magnetic field sensitivity also severely hinders PMTs from being integrated into multi-modal PET/MRI (Magnetic Resonance Imaging) systems for medical imaging. The readout electronics required to analyze and interpret the PMT signals are also usually implemented separately onto a PCB (printed circuit board) increasing the complexity of the system [2], [8]. As a result, research shifted to solid-state solutions including APDs and later SPADs to overcome the limitations of magnetic field sensitivity, size, cost, and integration capabilities with potential full system-on-chip (SoC) solutions.

1.2.2. Avalanche Photodiodes

An avalanche photodiode (APD) is a p-n or p-i-n photodiode whose gain mechanism is given by avalanche multiplication. These diodes operate at a relatively large reverse bias but below the breakdown voltage of the material. The high electric field produced in the depletion region as a result of the large reverse bias allows charge carriers (preferably photogenerated) that enter the depletion region to accelerate and gain enough energy to generate an EHP upon colliding with the lattice. This process is known as impact ionization

and can be done with both positive and negative charge carriers. However, in APDs, as illustrated in Figure 1-3, the primary charge carrier involved in this avalanche process is the electron. Each collision by the initial charge carrier and subsequently generated charge carriers results in a multiplication event that generates an avalanching current. The impact ionization coefficient of a charge carrier is denoted α (α_n for electrons and α_p for holes) and represents the number of EHP produced per unit length by the carrier. Due to the avalanche process being stochastic in nature, not every injected or photogenerated carrier leads to the same multiplication. This results in noise that is denoted the excess noise factor F. It is dependent on the ratio of the hole to electron impact ionization coefficients α_p/α_n (or α_n/α_p for hole injection/multiplication). A small ratio is preferred, indicating that one charge carrier is more dominant in the final result of the avalanche multiplication, resulting in less variation and thus noise. While this ratio (α_p/α_n) can depend on the electric field strength, it is ~0.5 for germanium and III_V compound semiconductors but only ~0.02 for silicon. This resulted in the majority of commercial APD detectors being made of silicon [17].

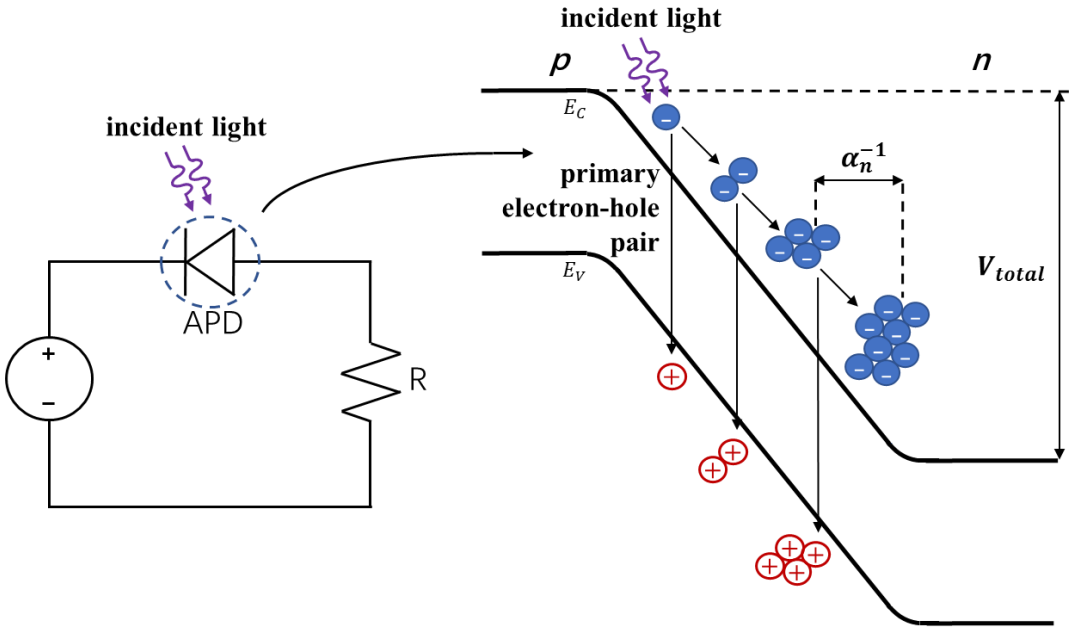


Figure 1-3: The electron dominated impact ionization process in avalanche photodiodes (APDs) with electrons injected from the p-side of the depletion layer. The average distance between each electron multiplication event is α_n^{-1} and V_{total} is the applied reverse bias including the built-in potential [8].

Since the APD is an analog detector, the value of the amplified current is measured and used to indicate the detection of light. Furthermore, a large and stable gain is preferred to improve performance in low-level light conditions. However, the avalanche multiplication events do not distinguish between photogenerated or thermally generated EHPs. Thus, the noise is also amplified, and because of the analog signal, this noise is not measured in counts as with the PMT, but is reported as a dark current. Thus, a high signal-to-noise ratio (SNR) is important in distinguishing between the amplified dark current and amplified signal. Due to the variation in dark current and gain with temperature and reverse bias, there usually exists an optimal operating condition to maximize the SNR.

Another key performance parameter of the APD is its spectral response, or responsivity, indicating the amount of current (A) generated per unit power (W) of light of varying wavelengths that is incident on the detector. Similarly, the QE is also wavelength dependent, indicating the percentage of photons that reach the depletion region and trigger an avalanche. This is dependent on the light transmitted through the semiconductor, the proportion of photons that are absorbed inside and within one diffusion length outside of the depletion region to trigger an avalanche, and the current collected from non-recombined carriers.

While a very good photodetector, the APD is not suitable for single photon detection due to its limited gain [2]. Compared to a PMT that is capable of a gain around 10^6, an APD can only reach gains up to 1000 [15], with most Si APDs having a gain around 50 at the typically reported wavelength of 420 nm for PET applications [18]. They are, however, important in the understanding of the evolution of Si photodetectors and the operating principles of SPADs.

1.2.3. Single-Photon Avalanche Diodes (SPADs)

Where APDs are said to operate in the linear mode, SPADs operate above the breakdown voltage V_{BD}, in what is known as the Geiger-mode (in analogy to a Geiger-Muller detector), attaining the name Geiger-mode APD (G-APD) [2]. The amount of voltage applied past this breakdown point is known as the excess voltage V_{ex}. Such a device can

detect single photons since the detector is put in a state whereby a single photon can trigger a self-sustaining avalanche process and the gain is large enough to give a detectable current. Thus, these devices take the more commonly used name of SPADs. Since a SPAD can detect single photons similar to a PMT, an array of SPADs is also known as a Silicon PhotoMultiplier (SiPM). The distinction between the modes of operation of a diode are presented in Figure 1-4.

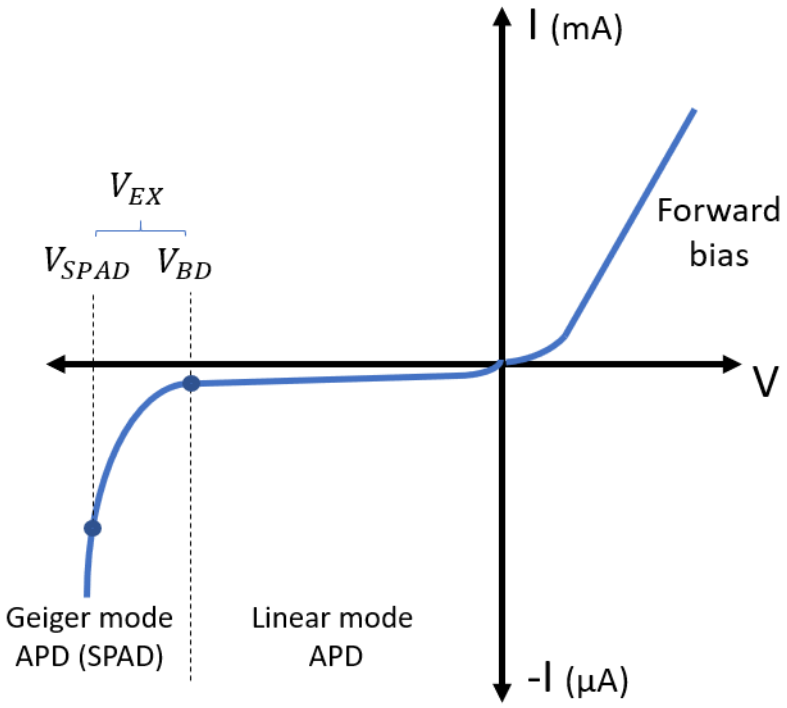


Figure 1-4: IV characteristics of a diode specifying modes of operation based on the applied bias. A single-photon avalanche diode is biased (V_{SPAD}) above the breakdown voltage (V_{BD}) by an excess voltage (V_{EX}).

While a depletion region by definition is depleted of free charges due to minority carrier diffusion, there exists a reverse saturation current independent of voltage (but dependent on temperature) that can range from microamperes (in germanium diodes) to nanoamperes (in silicon diodes) [4]. In silicon, this reverse current typically rises from the nanoamp range to the milliamp range during Geiger-mode avalanche breakdown in as little as 1 ns [15], reflective of a 10^6 multiplication gain. It is the fast rise time associated with the avalanche response that allows SPADs to be suitable for Time-of-Flight (ToF)

applications, more-so than PMTs [8]. When reverse biased past breakdown, the slope of the band bending of the depletion region becomes extreme and an electric field as high as $10^5 – 10^6$ V/cm is developed [3], [19]. Free carriers that enter the depletion region either by photogeneration, thermal generation, or diffusion are accelerated to such high speeds as a result of the large electric field, that the avalanching process becomes self-sustaining. In this state, the ionization coefficient of holes is not neglected as in the case of the APD, but instead plays a significant role in the multiplication gain, allowing holes to also generate enough EHPs to continue the avalanching process.

Following the derivation found in [19], the multiplication gain can be theoretically examined. Assume the initial reverse-biased electron current I_{n0} enters the depletion region from the p-side ($x = 0$), the electron current $I_n(x)$ will move through the depletion region (of width W) and increase due to the avalanche process, at which point the current at $x = W$ can be written as

$$I_n(W) = M_n I_{n0} \tag{1-1}$$

where M_n is the multiplication factor. The same process occurs from the n-side due to holes, and the total current is constant through the junction in steady state. One can write an expression for the incremental electron current at some point x as

$$dI_n(x) = I_n(x)\alpha_n dx + I_p(x)\alpha_p dx \tag{1-2}$$

where α_n and α_p are the electron and hole ionization coefficients, representing the number of electron-hole pairs generated per unit length by an electron or hole, respectively. Making the approximation that the electron and hole ionization rates are equal, it is possible to write

$$1 - \frac{1}{M_n} = \int_0^W \alpha \, dx \tag{1-3}$$

The avalanche breakdown voltage V_{BD} is defined as the voltage at which M_n approaches infinity. Since the ionization rates are strongly dependent on electric field and the electric field is not constant through the depletion region, the above equation is not easy to evaluate. V_{BD} is best determined experimentally and is ill-defined in literature as it is difficult to accurately predict [20]. Although SPADs theoretically have infinite gain due to

their self-sustaining avalanche process, the current must be immediately quenched, either passively or actively, to prevent destruction of the device.

It is evident that SPADs (or SiPMs) carry the advantage moving forward towards future imaging devices. The comparisons against the other described photodetector technologies are summarized and combined from [8], [15], [21] into Table 1-1. Here, QE refers to quantum efficiency and is related to the photon detection efficiency (PDE) of the photodetector. The reported value is at a wavelength of 420 nm since a large market and focus for photodetectors in biomedical imaging is on PET systems, where a scintillator is used to convert high energy gamma rays into visible light at 420 nm before hitting the photodetector. Performance parameters are described in more detail in Chapter 2 along with a closer look into the design of SPAD arrays to form SiPMs. It is shown that SPADs have unrivalled levels of miniaturization and portability, low fabrication costs, low voltage requirements, and high levels of integration while offering performance on par with or better than PMTs and APDs.

Table 1-1: Comparison of single-photon detector technologies.

	PMT	**APD**	**SiPM**
Gain	10^6	50-1000	~10^6
Bias (V)	>1000 V	300-1000	~15-80
QE @ 420 nm (%)	~25	~70	>50% (PDE)
Magnetic Field Compatibility	No	Yes	Yes
ToF Capability	Limited	No	Yes
Signal/Readout	Analog/Complex	Analog/Complex	Digital/Simple
Price/channel ($)	>200	~100	~50

1.3. Research Challenges and Motivation

While SPADs fabricated in a custom process can achieve good performance, they are typically costly and/or lack the ability to be monolithically integrated with high-speed electronic circuits. SPADs fabricated in standard digital CMOS technology can achieve this integration as well as high volume production to allow reduced costs. Unfortunately, the

fixed doping profiles and process layers of standard processes means fixed junction depths and widths as well as unremovable passivation and dielectric layers that can severely limit a SPAD's photon detecting capabilities [3]. Also, moving to smaller technology nodes to take advantage of integration with faster, low-power digital circuitry, leads to higher noise from increased tunneling effects due to higher doping profiles producing thinner depletion regions. Although this can improve a SPAD's timing performance, it also reduces its photon detection efficiency at the same time. Furthermore, it can prevent transistors being directly connected to the SPAD junctions due to the large bias voltages required to properly operate the SPAD, especially if multiple junctions are used. These performance parameters are explored in more detail in Chapter 2, but it is clear that the trade-offs involved must be carefully considered when analyzing a proposed application. Even then, the technology limitations can severely impact expectations. Our goal is to improve the typical limitations experienced when designing SPADs in smaller, standard CMOS technology nodes while still reaping the benefits. Ultimately, it is expected that SPADs will become commonplace and may even become foundry IP blocks, with large industries exploiting SPADs in high volumes in the mobile/consumer areas, such as for automotive and Internet of Things (IoT) applications [7].

One particular application we hope to target with our work is the diffuse optical spectroscopy (DOS) of tissues and organs. In DOS, light in the optical window of 600-1000 nm (red and near infrared red (NIR)) is used to perform structural and functional imaging where the absorption of water is very low, and scattering is dominant [22]. Image reconstruction (2D or 3D slices) can be done through the data collected and systems which can do this are known as DOI (imaging) or DOT (tomography) systems. Images can be obtained through one of two system geometries: transmittance or reflectance [9], [23]. In the transmittance geometry, the light is transmitted through the object, from the source(s) on one side, to the detector(s) on the other. This can be done for thin objects (less than 8 cm thick) such as muscles, breasts, and the heads of newborn babies. For objects that are thicker or have high absorption, the reflectance geometry is employed where the source(s)

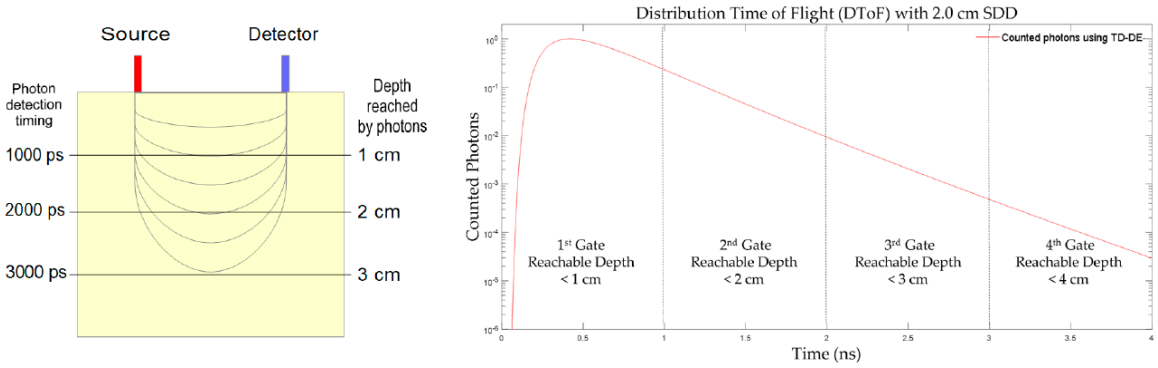
and detector(s) are on the same side and the distance between them is known as the source-detector distance (SDD).

The three methods of DOS are continuous wave (CW), frequency domain (FD), and time-resolved (TR) [9]. CW can only monitor variation of optical properties and not the values themselves, restricting its capability for structural imaging. FD and TR can quantify the absolute values of the optical properties, but depth discrimination (in the reflectance mode) is more challenging in FD than TR. The TR method proves to be the most powerful by utilizing timing information of the detected photons, but has suffered from high cost, complexity, and size limitations. SPADs have allowed the TR approach to become more technically and economically feasible and are shown to produce good results with great potential for future small-scale systems with high imaging performance [7], [9], [23], [24].

Many benefits can be derived for medical applications using DOS systems. For example, it can detect and localize strokes in the brain due to ischemic (lack of blood flow) or hemorrhagic (internal bleeding) conditions, as well as the hemorrhaging that might occur during traumatic brain injury (TBI) [25]. Monitoring can also be done of the brain's oxygen levels, especially post concussion when reports of mild hypoxia (lack of oxygen) can occur [26]. It can also be used in optical mammography for tumor detection, localization, and cancer treatment evaluation. Furthermore, functional DOS, or equivalently, functional NIRS (fNIRS), can measure hemodynamic changes by observing changes in the optical properties due to variation in tissue oxygen saturation (StO2) and brain blood flow during functional activities [9].

To obtain the necessary data in TR-DOI systems, very narrow width laser pulses enter the target and a timing histogram known as the distribution time-of-flight (DToF) is produced with the detected re-emitted photons. This is illustrated in Figure 1-5 where a source and detector are placed in reflectance mode [9]. Information from the DToF such as the logarithmic slope, along with the location of the sources and detectors, is used to extract the optical properties of points in the target using an inverse problem solver for the diffusion equation. Time-gating techniques with SPADs [23] have been employed to limit background noise and increase sensitivity to photons arriving from different depths, thus

improving SNR and depth detection. This is also shown in Figure 1-5b) and will be one technique that is further explored in this work in developing high-performance SPADs.



(a) (b)

Figure 1-5: a) A TR-DOS measurement example in the reflectance mode; b) the resulting distribution time-of-flight (DToF) histogram of detected photons [9].

To further improve a SPAD's effectiveness, they need to have higher sensitivity in the NIR range as well as have short timing jitter <500 ps [9]. However, SPADs typically suffer from having low NIR sensitivity due to their shallow junctions as well as low PDE in general if implemented in a standard CMOS process. Furthermore, tackling this issue using older technologies with thicker depletion regions for improved photon detection leads to worse timing jitter. Utilizing deeper junctions, or multiple junctions in parallel, can lead to designs that can extract more information from the detected wavelengths of light. This tackles the reduced PDE across the spectrum found in smaller standard CMOS technologies. The different junction depths and associated depletion regions allows for the detection and distinction of shorter vs. longer wavelengths and also widens the total spectrum of detected light when taken as a whole, thus improving the PDE. This can lend itself to functional imaging applications by distinguishing between the different colours of oxygenated vs. de-oxygenated hemoglobin and other physiological markers.

Should SPADs fabricated in the standard CMOS process reach performance levels comparable to some of the high-performing custom SPADs, it can prove revolutionary in the development of accessible and affordable solutions for medical imaging and monitoring applications. One foreseeable outcome is a portable and low-cost wearable cap that can

perform functional brain imaging of newborns through TR-DOS. This can be modified for instant, on-the-spot diagnosis of traumatic brain injury (e.g. concussion) of adults in accidents and sports related injuries. The ability for improved monitoring of blood oxygen levels can also help enhance the effectiveness of current practices such as pulse oximetry of patients after surgery among other applications. This concept is illustrated in Figure 1-6. Miniaturization is assisted through using pulsed laser diodes with low power (<2 mW) and narrow width pulses (<200 ps) [9] as photon sources, as well as vertical-cavity surface-emitting lasers (VCSELs) [27] which are predicted to be used for next generation TR-DOS systems. The pulsed light and time-gated SPADs can also help significantly reduce the risk of unsafe skin temperature increase [28]. Such an idea was recently explored to some extent [29] with an imager that will be capable of determining oxygenation of preterm-infant brains with high spatial resolution. However, it consisted of many bulky and expensive components and was meant as a bedside imaging device rather than being portable.

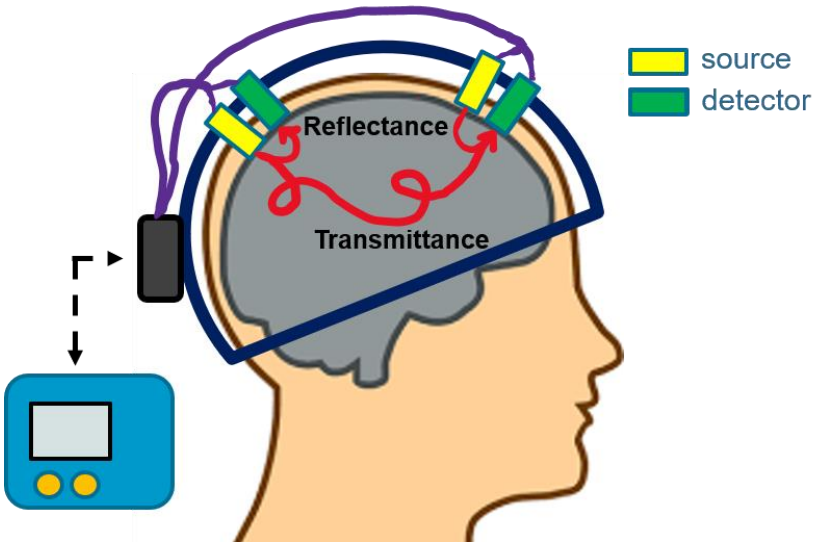


Figure 1-6: Concept of low-cost, portable TR-DOI (time-resolved diffuse optical imaging) system using safe, low-power diodes and high-performance SPADs.

Ultimately, we hope to improve the photon detecting capabilities of SPADs in smaller CMOS technology nodes and utilize TG techniques to pave the way towards low-cost, miniature, and portable TR-DOI devices. The reduced healthcare costs associated with the

early diagnosis of brain disease, the quick and low-cost diagnosis of brain trauma, and improved monitoring of existing patients will provide significant economic benefits in terms of reduced health care costs in both the short and long term.

1.4. Research Contributions

This research was focused on the design of advanced CMOS SPAD pixel structures to enhance the performance of SPADs designed in a standard CMOS processes. The eventual integration of such chips into compact and low-cost commercial products is explored by providing an example using a custom CMOS image sensor PCB, off-the-shelf components and development boards. The major contributions are more specifically summarized as follows:

- **Design and simulation of a triple-junction (TJ) SPAD structure in the TSMC 65 nm standard CMOS process as well as potential time-gated (TG) pixel architectures.** The top shallow junction in the TJ SPAD was incorporated into a TG SPAD pixel design with pulse generation circuitry that was previously designed in an older technology node for performance comparison. The top two junctions were used to create a novel dual-junction time-gated (DJTG) SPAD pixel and modified pulse generation circuitry in order to improve photon detection and overcome limitations in the biasing of multiple junctions with active quench and reset circuitry.
- **Fabrication and test of the TJ SPAD structure and the TG SPAD pixel.** The top junction of the SPAD was characterized through the TG pixel design and achieved a fill-factor of ~28.6%, peak photon detection efficiency of ~3.5% at 440 nm wavelength, <1% afterpulsing probability for hold-off times >22ns, and <200 ps timing jitter all at an excess voltage of 0.3 V. The deepest junction of the TJ SPAD was also characterized through an unbuffered design and the performance of both junctions were compared to those in the existing literature.

- **Design and implementation of a low-cost, highspeed, customizable camera and creation of a highspeed camera figure-of-merit.** The camera was designed using readily available commercial components such as the PYTHON 1300 image sensor, a MicroZed development board with a Zynq 7020 SoC, FMC carrier board, and pushbuttons and switches. The camera prototype had plenty of on-chip resources available (~5% utilization) for further customizability, a cost of ~$650 USD, and could achieve 211 fps at max resolution (1280x1024) and up to 2329 fps at a 256x256 resolution. Using the created figure-of-merit, the camera beats similar machine vision cameras and is comparable to existing low-cost highspeed commercial cameras in value but at a lower price point. The camera was tested in a biological setting to better reveal the physiological behaviour of *C. elegans* worms.

Publications:

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

1.5. Thesis Organization

In Chapter 1, the importance of single photon detection and its various applications is introduced. The operating principles, strengths, and limitations of the most prominent technologies for single photon detectors being PMTs, APDs, and SPADs, are then described and compared with each other. Following that is the motivation for using SPADs as the technology of choice for future detectors that can be very small, highly integrated, scalable, and low-cost. Finally, a brief summary of the main contributions of this research and the structure of this thesis are described.

In Chapter 2, the performance parameters associated with CMOS SPADs are explained and a review of the current literature is performed. More specifically, the various performance parameters of the SPAD such as the PDE, DCR, and timing jitter are explained and the trade-offs between them are examined. This is followed by research into the various structures and design considerations made in the fabrication of SPAD pixels and arrays, as well as a look at trends associated with state-of-the-art structures such as advanced on-chip

processing and 3D stacking. Finally, the design of multi-junction and time-gating structures are investigated to improve a SPAD's PDE and SNR.

In Chapter 3, the design and layout of a triple-junction SPAD structure and its use in time-gated pixel architectures is given using the TSMC standard 65 nm CMOS process. A time-gated pixel utilizing the top shallow junction of the SPAD along with its corresponding pulse generation circuitry is designed and its operation is simulated. A novel time-gated pixel design utilizing the top two junctions is also described as well as the pulse generation used to accomplish this.

In Chapter 4, fabricated chips containing the triple-junction SPAD and time-gated pixel designs are tested and their performance is evaluated. The time-gated pixel is used to characterize the top shallow junction performance and an unbuffered design is used to similarly test the deepest junction. The characterization is done according to the various performance parameters described in Chapter 2 such as PDE, DCR, AP, and timing jitter. The results are summarized and compared against similar SPADs in literature.

In Chapter 5, the implementation of a low-cost, highspeed camera is provided. The selection of standard commodity parts is explained to create a low-cost but capable system. The RTL and C design upon which the camera runs are provided as well as a link to the project code. A new figure-of-merit for highspeed cameras is also described to compare our design against existing machine vision and low-cost highspeed cameras. Finally, the results of applying the camera to better observe the physiological behaviour of *C. elegans* worms is given.

In Chapter 6, this thesis is concluded with a summary of the work done and recommendations are provided for future improvements and research ideas in time-gated and multi-junction SPAD designs.

Chapter 2
SPAD Performance and Literature Review

2.1. Performance Parameters

SPAD performance is dependent on a number of parameters, some of which are mutually exclusive and can not be improved together. Sacrifices are typically made on one parameter in order to improve another, depending on the application. Thus, with many SPADs being developed offering high performance in some areas and low performance in others (such as with photon timing vs. photon counting), it can be difficult to compare SPADs against one another. The following subsections will explore these performance parameters, including photon detection efficiency (PDE), count rate, dead-time, dark count rate (DCR), afterpulsing (AP), and timing jitter.

2.1.1. Photon Detection Efficiency (PDE)

The PDE is the ratio of the number of detected photons (Φ_{SPAD}) to the number of incident photons (Φ_{IN}). It can also be theoretically calculated as the product of the geometric fill-factor (*FF*), the quantum efficiency (*QE*) and avalanche triggering probability (P_{AV}) [3],

$$PDE(V_{ex}, \lambda) = P_{AV}(V_{ex}) \cdot QE(\lambda) \cdot FF = \frac{\Phi_{SPAD}}{\Phi_{IN}}. \quad (2\text{-}1)$$

The QE, as a function of wavelength (λ), indicates the percentage of incident photons that produce EHPs. The P_{AV} dictates whether these EHPs trigger an avalanche or not and is largely dependent on the excess voltage (V_{ex}). This makes increasing the excess voltage a straightforward method for increasing the PDE, but comes at the cost of a higher probability of dark counts and a higher AP due to the increase in avalanche charges flowing through the junction. The FF is the ratio of active SPAD area to the total imaging or pixel area and increases as technology nodes scale down. This is illustrated in Figure 2-1 with regards to

the implementation of CMOS SPADs [30]. Higher FF improves the PDE performance by allowing more area for incident light to be detected, and can be done either by directly increasing the SPAD's active area, moving to a smaller technology node to reduce the size of surrounding circuitry, or having SPADs share n-wells [31]. A more detailed look into SPAD structures in standard CMOS technology is described in Section 2.2.1.

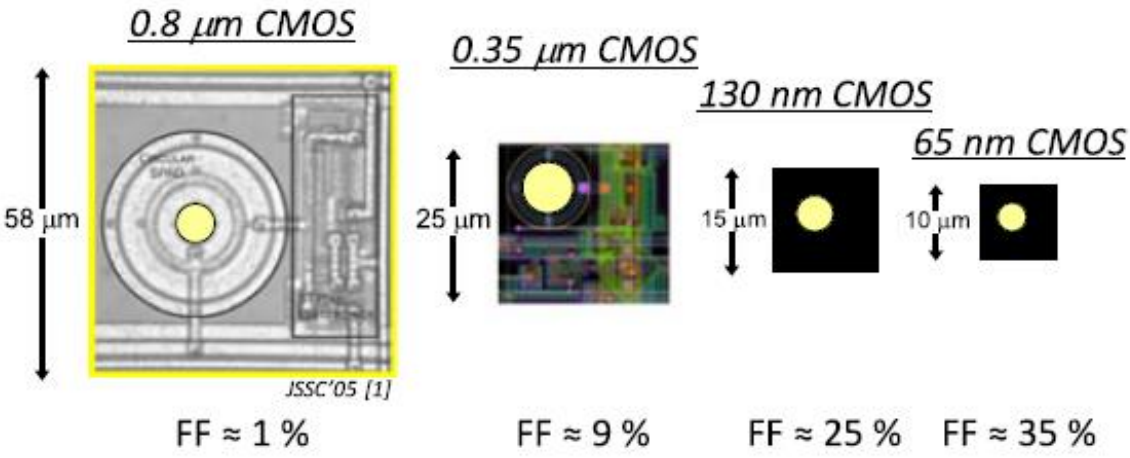


Figure 2-1: Example of SPAD fill factor increasing with the scaling of technology nodes; the yellow circles indicate the SPAD active area [30] (© 2018 IEEE).

Semiconductor material plays a large role in the operation and performance of photon detection. Although silicon is the standard material for standard CMOS processes, it is not ideal for photogeneration, primarily due to its indirect bandgap structure. The bandgap is the energy range where theoretically no electron states can exist and represents the energy which a valence electron must obtain to move into the conduction band and become a mobile charge carrier. Large bandgaps (3.5 – 6 eV) represent insulators, smaller bandgaps (~1 eV) are semiconductors, and those with very small or no bandgaps are conductors [32]. As seen in Figure 2-2, the indirect bandgap of materials like Silicon and Germanium requires an extra phonon interaction to generate a free carrier as opposed to a simple photon-induced vertical transition such as in materials like Gallium Arsenide (GaAs) and Indium Phosphide (InP).

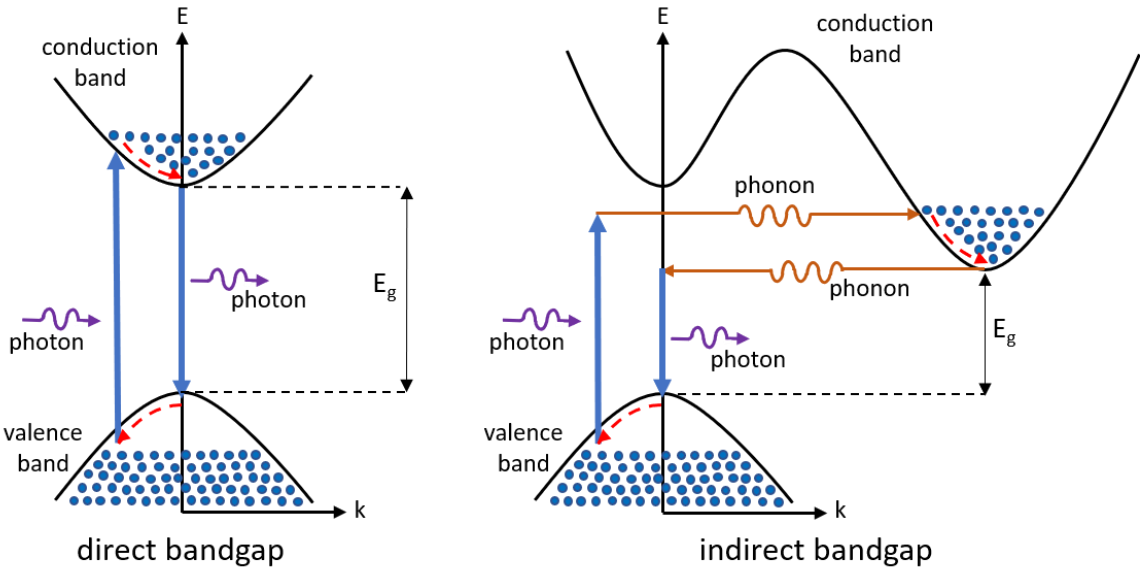


Figure 2-2: Direct vs. indirect bandgap materials in the photogeneration of electron-hole pairs.

Before resulting in an EHP, either by a direct or indirect transition, the photon must first be absorbed. The absorption coefficient determines the penetration depth and range of wavelengths to which a material is opaque. Figure 2-3a) shows the absorption characteristics of common semiconductor materials [33]. Since photons carry energy proportional to its frequency (or inversely proportional to its wavelength), there exists a cut-off wavelength dependent on the material in order to overcome its bandgap and is ~1.1 µm for silicon, for example. The material becomes transparent to light with wavelengths greater than this cut-off. On the other hand, there exists a lower limit to the wavelength due to the high energy photons being reflected, scattered or absorbed at the surface of the material. Figure 2-3b) demonstrates how the intensity (or power P) of light decreases exponentially as it travels through a material following the Lambert-Beer Law, and is given by

$$P(x) = P(x_0)e^{-\alpha(x-x_0)} \tag{2-2}$$

where α is the absorption coefficient and is inverse to the penetration depth. The absorption/penetration depth is the distance a photon travels into a material before its intensity or power decays to 1/e of its surface value.

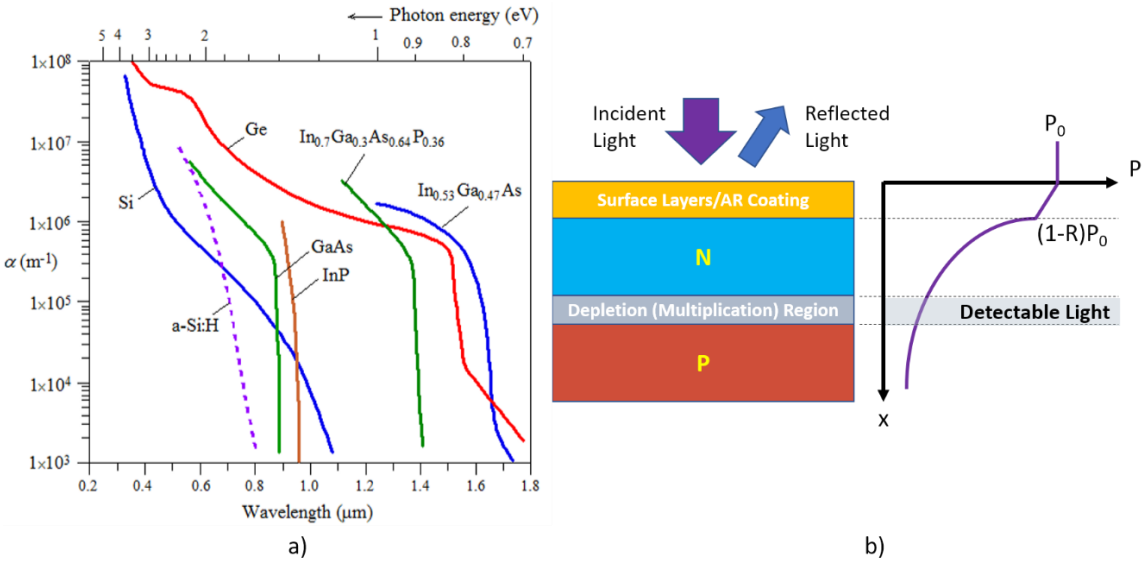


Figure 2-3: a) Comparison of the absorption coefficient of several common semiconductor materials [33]; b) The absorption characteristics of a diode dependent on depth and the properties of the surface layers which may have anti-reflective (AR) properties

CMOS technology with several layers above active devices can adversely affect the absorption probability. A thick top passivation layer used to protect chips from external contaminants post fabrication reduces the number of photons that can pass through. Furthermore, the photons must still pass through several dielectric layers of varying refractive indices before reaching the active, photosensitive area. While nothing can usually be done about this in standard CMOS digital/RF technologies, the dielectric above the active region can be optimized in the more expensive CMOS Image Sensor (CIS) technology to minimize reflections. Other expensive post-processing steps, such as the use of micro-lenses, can help focus light onto the active area of a pixel [3], [11]. In fact, cylindrical micro-lenses offer one of the most enhanced fill factors (~50% increase for collimated light) among SPAD sensors [34]. Removal of the top passivation layer can be done in standard CMOS technology and can lead to improvements in performance [3]. However, other experiments indicate lower performance [35], perhaps due to a damaged or scratched device. Typical PDE performance of SPADs implemented in standard CMOS technologies is approximately a few percent, and is dependent on the excess voltage and wavelength [36]–[39].

2.1.2. Count Rate and Dead Time

When an avalanche is triggered, the circuit must be quenched immediately, either passively with a large resistor on the order of a few hundred kΩ [4], or actively with the proper front-end circuitry to accelerate the process. This brings the voltage of the SPAD below breakdown, stopping the avalanche process and prevents overheating and destruction of the device. The SPAD must then be reset either passively or actively to be ready for the next detection event. This sequence is illustrated in Figure 2-4, with the full process happening over a time period known as the deadtime.

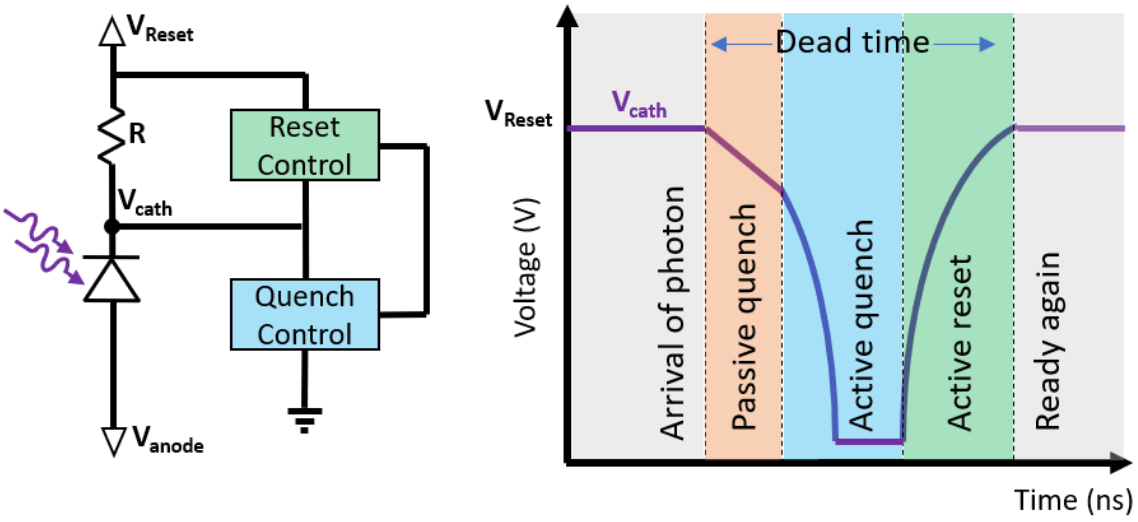


Figure 2-4: Schematic diagram of a SPAD with active quench and reset and the corresponding behaviour of the SPAD cathode voltage upon the onset of an avalanche breakdown. The time for the SPAD to be ready for the next avalanche trigger is called the dead time.

If the quench and reset is done passively, such as with a large resistor, the voltage reset is slow and the total deadtime can be on the order of 100 ns and is not well defined. While the quench process itself may still be acceptably fast, it is the reset that suffers due to the RC time to charge the capacitance of the SPAD. Active circuitry that allows for detection and feedback to pull the SPAD voltage down immediately after avalanche can drastically increase operating speed as well as lead to reduced avalanche charge and afterpulsing. The disadvantage of this additional circuitry is the area occupation and reduced fill factor, sometimes preventing the integration of large or closely packed detector arrays [40]. On

the other hand, prolonging the deadtime to reduce the effects of AP limits highspeed operation and can introduce non-linearity in the response of the SPAD to increased light intensity, affecting its photon counting ability [41]. This fundamentally limits the maximum count rate at which a SPAD can operate. The SPAD can be held below breakdown for periods longer than the deadtime, known as the hold-off time to match circuit operation or reduce noise. Thus, proper quench and reset circuitry around a SPAD is usually implemented to maximize speed performance, and is investigated further in Section 2.2.3.

2.1.3. Dark Count Rate (DCR)

When a SPAD is reverse biased past its breakdown voltage, it will produce a digital pulse when a free charge carrier is generated in the depletion region, triggering an avalanche response. This means the process is not limited to only photo-generated EHPs. There are several carrier generation mechanisms that can interfere and trigger counts, even when the SPAD is kept in the dark, thus producing "dark counts". Minority carriers that are generated outside, but move into, the avalanche multiplication region also play a role in the dark count rate, being associated with a collection probability that depends on the carrier mobility, diffusion constant, and recombination lifetime. Thus, the DCR is said to be the product of the breakdown triggering probabilities within the multiplication region, the total dark carrier generation (thermal or tunneling induced), and the collection probabilities of diffused carriers into the multiplication region [8]. The resulting DCR is an important performance parameter for determining a SPAD's signal-to-noise ratio (SNR) and resolution, which can for instance determine the depth to which TR-DOI systems can image [3], [9]. The DCR is expressed as a counts per second (cps or Hz) and is typically reported per μm^2 due to varying SPAD active areas.

Many forms of noise sources can exist in semiconductor devices. Of importance to the topic of SPADs are Shockley-Read-Hall (SRH) thermal generation-recombination (GR) and trap-assisted [3] or band-to-band tunneling (TAT and BTBT, respectively) [3], [42], [43]. These noise sources are represented in Figure 2-5. According to SRH statistics, crystal defects and/or impurities can create forbidden energy levels near the middle of the bandgap

which act as generation-recombination (GR) centers, thus releasing carriers that could potentially trigger a false dark count. Since this noise source is temperature dependent, it can be reduced at lower temperatures.

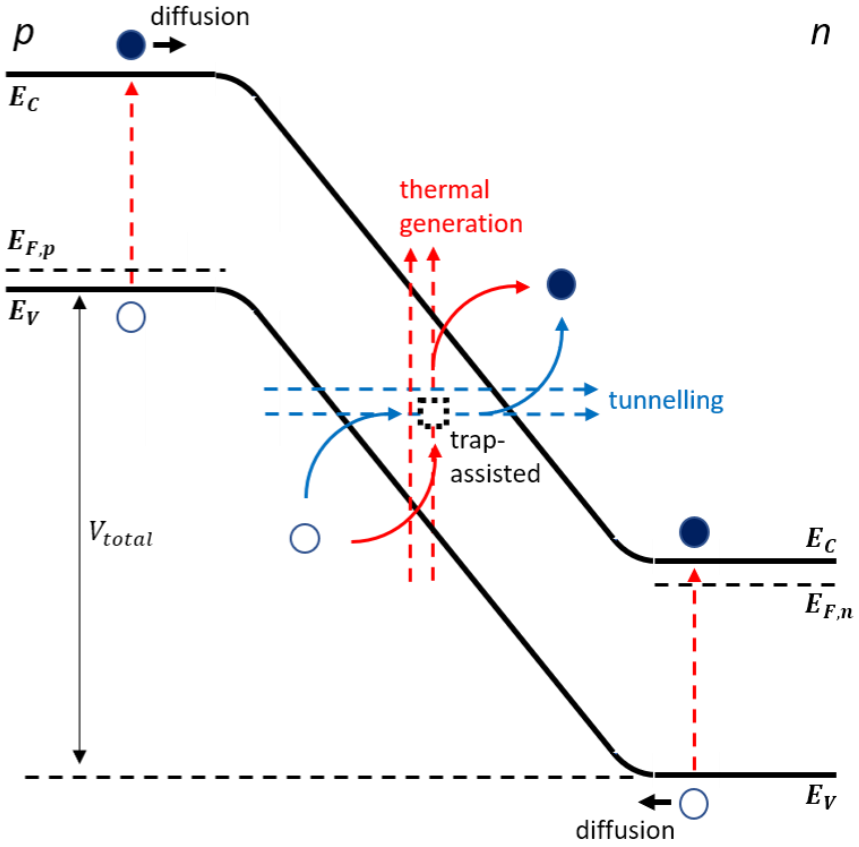


Figure 2-5: Noise sources include thermal generation and tunneling (that can be trap-assisted or band-to-band) which can trigger an avalanche and contribute to the dark count rate (DCR). The V_{total} is the sum of the built-in voltage and applied reverse bias.

DCR declines less rapidly at low temperature since tunneling is weakly temperature dependent. Tunneling also becomes more prominent in smaller Deep-Sub-Micron (DSM) CMOS processes as the junction dimensions scale down and the doping concentrations increase, resulting in thinner depletion regions and more abrupt junctions [3]. Thermally generated dark counts tend to saturate at large excess voltages and thus DCR is typically plotted as a function of excess voltage to investigate tunneling effects, with an expected exponential dependence. In silicon, the non-equilibrium carriers' indirect thermal generation in the depletion layer dominates the direct band-to-band thermal generation,

thus the latter can be ignored. Also considering trap-assisted thermal generation, the generation rate of non-equilibrium carriers via traps can be calculated according to SRH recombination theory [42] as

$$G_{SRH,TAT} = \frac{pn - n_i^2}{\frac{\tau_n}{1+\Gamma_n}\left[p + n_i \exp\left(\frac{E_i - E_t}{kT}\right)\right] + \frac{\tau_p}{1+\Gamma_p}\left[n + n_i \exp\left(\frac{E_t - E_i}{kT}\right)\right]} \tag{2-3}$$

$$\tau_{n(p)} = \sqrt{\frac{m^*_{n(p)}}{3kT}} \cdot \frac{1}{N_t \sigma_{n(p)}}, \qquad n = n_i \exp\left(\frac{E_F^n - E_i}{kT}\right), \qquad p = n_i \exp\left(\frac{E_i - E_F^p}{kT}\right).$$

Here, k is Boltzmann's constant, T the absolute temperature, n_i is the intrinsic carrier concentration, n and p are non-equilibrium electron and hole concentrations, respectively, $\tau_{n(p)}$ are the electron (hole) lifetimes, N_t the recombination center density, E_i the intrinsic Fermi level, E_t the recombination center energy level, $m^*_{n(p)}$ the electron (hole) effective mass, and $\sigma_{n(p)}$ the capture cross area of electron (hole). The strong electric field in the avalanche region leads to TAT and a field-effect enhancement factor Γ was introduced into the SRH model to describe this. Assuming $\Gamma = \Gamma_n = \Gamma_p$ (for both electrons and holes), if the applied electric field is not more than 9×10^5 V/cm, then

$$\Gamma = 2\sqrt{3\pi}\frac{|E(x)|}{F_\Gamma}\exp\left(\left(\frac{E(x)}{F_\Gamma}\right)^2\right), \qquad F_\Gamma = \frac{\sqrt{24 m_t^* (kT)^3}}{q\hbar} \tag{2-4}$$

where $E(x)$ is the local electric field strength at depth position x, m_t^* is the effective mass of the tunneling electrons for silicon, q is electron charge, and $\hbar$ is the reduced Planck's constant. The Poole-Frenkel effect is negligible at a strong electric field when compared to the tunneling effect, so it is not considered here. These generated carriers have a probability to trigger an avalanche given as $P_{pair}(x)$, the total triggering probability from both electrons and holes. With a depletion layer area S and an upper and bottom boundary positions of the multiplication region as W_1 and $W_1 + W_2$, respectively, the DCR produced is

$$DCR_{SRH} + DCR_{TAT} = S \cdot \int_{W_1}^{W_1+W_2} P_{pair}(x) \cdot G_{SRH,TAT}(x)dx \tag{2-5}$$

$$P_{pair}(x) = P_e(x) + P_h(x) - P_e(x)P_h(x).$$

If the electric field strength exceeds 7×10^5 V/cm, BTBT becomes the dominant noise source which can be modeled using statistical factors and user-definable parameters for silicon according to [42]. Finally, the total DCR can be written as

$$DCR_{TOTAL} = DCR_{SRH} + DCR_{TAT} + DCR_{BTBT}. \quad (2\text{-}6)$$

Key model parameters of avalanche trigger probability $P_{pair}(x)$ and electric field profiles $E(x)$ are provided by Geiger mode TCAD simulation. It was concluded that TAT was the main source of DCR, with BTBT becoming the dominant origin of DCR for scaled DSM CMOS technologies [42]. This is evident, for example, with an 8 µm diameter octagonal SPAD in 90 nm standard CMOS showing a 16 kHz DCR at 130 mV excess bias [44], while a same sized SPAD in 65nm standard CMOS yielded a 150 kHz DCR at the same excess bias [45]. As excess bias is increased, SRH noise increases slowly while the tunneling noise increases exponentially. With the scaling down of technologies, BTBT will eventually dominate as the doping levels of active areas, and consequently the electric field, are further enhanced.

2.1.4. Afterpulsing (AP)

A dark noise mechanism unique to SPADs is AP. During a Geiger avalanche, the flowing free charges can fill the deep-level energy traps caused by the previously mentioned semiconductor impurities and defects. These traps are characterized by finite lifetimes and the charges can be released at random time intervals which trigger statistically correlated Geiger pulses if the excess bias is re-established before the traps can be released. Dark counts from thermal generation follow a Poisson distribution whereas afterpulsing does not [3], [46]. Due to AP avalanches having the probability to create more subsequent afterpulses, there exists a positive feedback loop which can be described by a geometric series. With P_{AP} as the AP probability and DCR_0 the DCR without AP, the total DCR is given by [47]

$$\begin{aligned} DCR_{Total} &= DCR_0 \cdot P_{AP} + DCR_0 \cdot {P_{AP}}^2 + \cdots + DCR_0 \cdot {P_{AP}}^n \\ &= DCR_0 \cdot \sum_{n=0}^{\infty} {P_{AP}}^n = \frac{DCR_0}{1 - P_{AP}}. \end{aligned} \quad (2\text{-}7)$$

Since the avalanche current density and duration affect the number of filled traps, the total parasitic capacitance of the SPAD plays a key role in the AP probability. The probability of AP is given approximately as

$$P_{AP} \approx \frac{C}{q} N_t \sigma_n W_e V_{EX} \tag{2-8}$$

where N_t is the electron trap concentration, σ_n is the electron trap cross section, W_e is the effective depletion width and C the SPAD's capacitance [48]. The W_e is dependent on the doping concentrations in the process where as N_t and σ_t are mostly determined by contamination and damage during the fabrication – all factors a designer has no control over when using a standard process. Since W_e is also dependent on V_{EX}, that is one of the most influencing factors and is controllable by the user. However, reducing V_{EX} limits a SPAD's PDE and timing performance. Often, it is best to have properly designed front-end circuitry to reduce the capacitance on the SPAD and its quenching time [49].

A high AP can significantly affect not only the detector sensitivity, but also the maximum count rate. This is because a common way to reduce AP is by extending the hold-off time of the SPAD to allow trapped charges to be released. In fact, DCR_0 can be found experimentally by using a very long hold-off time so that virtually all trapped charges have enough time to be released [50]. When cooled to low temperatures, AP becomes the dominant source of dark noise in SPADs since the lifetime of the traps increase with temperature. Given that impurities and crystal defects affect the DCR and noise performance, a low DCR can also prove as an indicator of a good CMOS process and the quality of the fabrication [3]. It is evident that a trade-off must occur when selecting an appropriate hold-off time that balances an acceptable AP probability with operating frequency. Proper trap depopulation and the minimization of avalanche charge through optimized front-end circuitry is key to achieving both high count rates and low noise.

2.1.5. Timing Jitter

The instrument response function (IRF) of a SPAD to a laser pulse is typically composed of a Gaussian peak followed by an exponential tail, as illustrated in Figure 2-6. Photons

absorbed in the depletion region trigger a fast response with a timing jitter only dependent on the statistical fluctuation of the avalanche build-up time, thus giving the Gaussian peak. However, photons absorbed deeper into the neutral region must diffuse slowly to reach the depletion region and trigger an avalanche to be detected [51]. The time constant of the exponential tail gives the average time it takes for the carrier to diffuse to the depletion region. Longer wavelength photons produce this behaviour since they have a longer absorption depth, thus missing the top shallow junction that most SPADs utilize. An IRF with little to no diffusion tail can indicate that the SPAD junction is at the optimal depth for the wavelength of light absorbed.

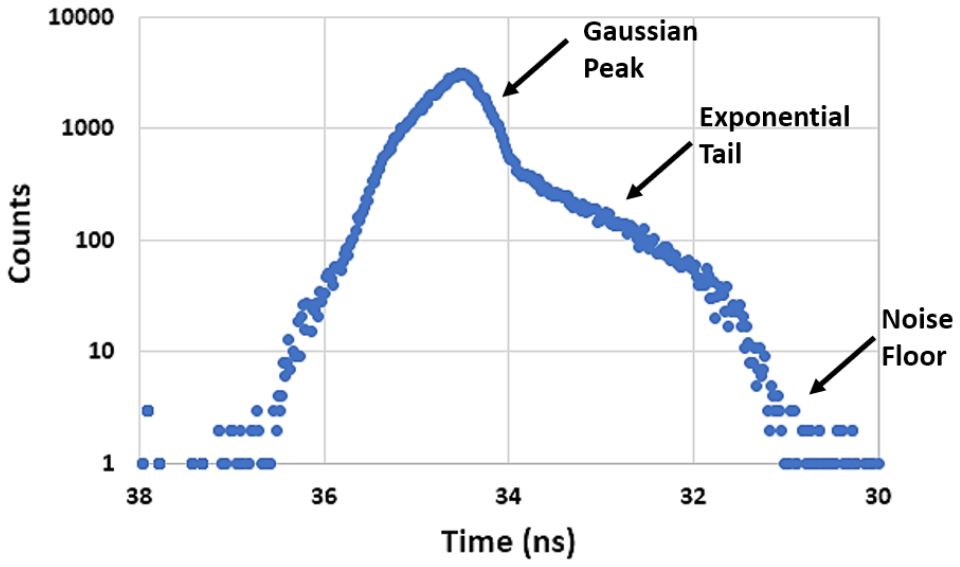


Figure 2-6: Example of the temporal impulse response of a SPAD, noting the Gaussian peak and exponential tail in addition to the noise floor. Same data as in Figure 4-19 a) (V_{ex} = 500 mV).

Narrowing the width of the Gaussian peak provides more precise timing information. Doping concentrations play a significant role as they influence the thickness of the depletion, or active, region. As technology scales down, the thinner junctions with higher electric fields which produces lower breakdown voltages and PDE, but better timing precision [2]. A sharper build-up time, and thus a smaller time jitter, can also be obtained either by designing SPADs with smaller active areas or applying a higher excess bias during operation. A higher excess voltage increases the avalanching probability and speeds up the

turn-on transient of the avalanche current, however this comes at the cost of higher DCR and AP [3].

2.2. Deep Sub-Micron (DSM) CMOS Architectures

2.2.1. Multi-Junction Structures

With a focus on SPADs implemented in the standard Silicon CMOS process, p-n junctions can be implemented in a number of ways, as seen in Figure 2-7. In a standard process with a p-type substrate (p-sub) and a deep n-well (DNW) structure as a result of ion implantation, there can be both lateral and vertical SPADs, although it has been shown that the lateral junctions offer the worst performance [32]. The thickness of the layers also contributes to potential optical cross-talk between adjacent pixels/cells [2]. As displayed in Figure 2-7, it is possible to have up to three vertically stacked p-n junctions in some standard DSM CMOS technologies by utilizing the n+/p-well, p-well/DNW, and DNW/p-sub junctions. Due to the third junction sharing a terminal with the p-sub, one terminal is limited to being connected to ground, potentially complicating bias conditions in multi-junction designs.

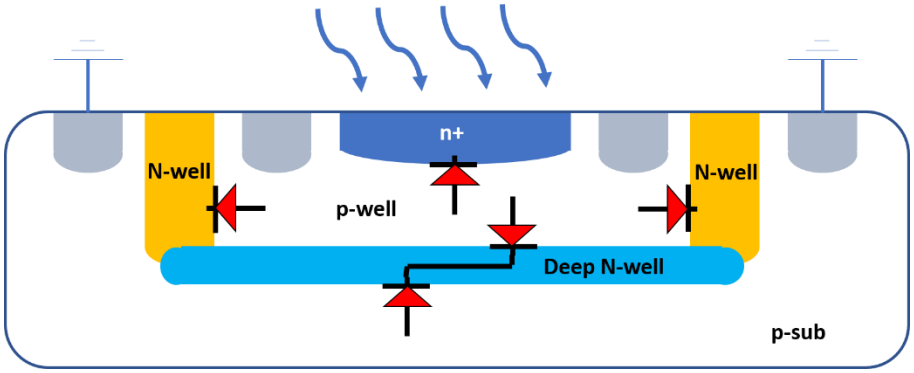


Figure 2-7: some p-n junctions available now in a standard DSM CMOS technology.

Most SPAD designs rely on just one junction and is typically the top-most shallow one. While this aids in the detection of shorter wavelengths, detection of longer wavelengths diminishes. The idea of a dual-junction was proposed in [52] using a p+/n-well shallow planar junction as well as a deep non-planar junction that coincided with the guard ring in the IBM 180 nm process. They demonstrated reliable operation of the device and expanded

the spectral response of the device. The technique was later applied in a 130 nm low-voltage CMOS image sensor (CIS) process that incorporated a deep n-well structure allowing for both junctions to be planar (p-well/deep n-well and deep n-well/p-sub) [53]. The junctions could be operated concurrently or separately, and the junction at which a photon is detected can be uniquely distinguished by the differing dead time of the detection pulses, easily picked up by simple digital circuitry. Achieving proper biasing of both junctions at the same time runs the risk of reaching operating voltages not suitable for the digital circuitry in the technology. Therefore, the comparator for avalanche detection was done through AC coupling, and the SPADs had to be passively quenched. The structure showed a peak PDE of ~40% when both junctions operated together compared to ~30% when operating separately.

Triple-junction (TJ) [54] and quad junction [55] designs have also been fabricated to distinguish between wavelengths of incident light, however front-end circuitry and integration with other digital circuitry is not discussed. The AC-coupling usually required to interface with these multi-junction designs hinder this capability as capacitors are large components and not ideal for large array structures. If this can be overcome, the increased PDE and wavelength distinction can make for better performance and more versatile designs.

2.2.2. Guard Rings and Noise Reduction in the DSM CMOS Process

Guard ring structures are important for SPAD design to ensure the region around the junction is free from noise sources and to reduce the electric field intensity at sharp junction edges that can result in premature edge breakdown (PEB). Ideally, the electric field intensity that can trigger avalanche multiplication should be spread uniformly across the planar region of the active area. In DSM CMOS technologies, a silicon dioxide (SiO_2) shallow trench isolation (STI) is used to prevent punch-through and latch-up in circuits and was first utilized to act as a guard ring for a SPAD fabricated in a 180 nm CMOS technology [56]. The higher dielectric strength of SiO_2 means a much more compact guard ring can be created using STI to improve FF. However, STI can create defects at the SiO_2-Si boundary

that function as generation-recombination centers and cause an increase in DCR and AP if they directly contact the active area of the SPAD. Diffusion GR structures have been shown to be effective and most commonly used, whereby a p-well diffusion GR (for p+/n-well junctions) or n-well diffusion GR (for n+/p-well junctions) help reduce the doping and thus electric field intensity at the edges [8]. This helps position the STI further away from the active region which allows carriers generated at the SiO_2-Si boundary to more likely recombine than diffuse into the active region.

Novel techniques have been employed as well such as a field gate over the perimeter [57] that proved effective at reducing PEB with no FF loss. The poly layer around the perimeter of the p+/n-well structure meant a negative gate voltage could be applied to lower the electric field at the edge of the junction, reducing PEB and creating a more planar electric field. It was demonstrated to offer high SNR under high-illumination conditions and high sensitivity under low-light conditions, with a DCR as low as 2 Hz/μm^2 at room temperature. To reduce the band-to-band tunnelling in SPADs made in DSM CIS technology, it was demonstrated that a process layer intended for pinned-photodiode formation could be employed to reduce DCR [58]. The available p- implant fulfills the role of a glove-like passivation implant around the STI to reduce dark current in pinned-photodiodes, as well as acting like a guard ring structure. This forms a less abrupt, graded junction, thus reducing tunnelling dark counts. The breakdown voltage due to this change increased to around 12.4 V compared to the usual 9.6 V reported for p+/n-well SPADs. The circular SPAD device tested achieved a DCR of 40 Hz at 25℃ with a diameter of 8 µm. The perimeter field gate, STI GR, and p-well diffusion GR (p+/n-well) structures are illustrated in Figure 2-8.

Another layer of interest is the silicide blocking layer, which blocks the formation of silicide used in the DSM CMOS process to reduce sheet and contact resistance [59]. It is, however, a well-known source of reverse leakage current in reverse-biased pn junctions that was shown to increase DCR, and the removal of silicide on the active area of a SPAD showed a more sudden and rapid rise in the current during breakdown compared to one with silicide which indicates non-silicide SPADs to be of higher quality [41]. Even with

respect to APDs, one study used silicide on only the optical window and not the optical window and electrodes which showed higher responsivity and 63% higher photodetection bandwidth [60].

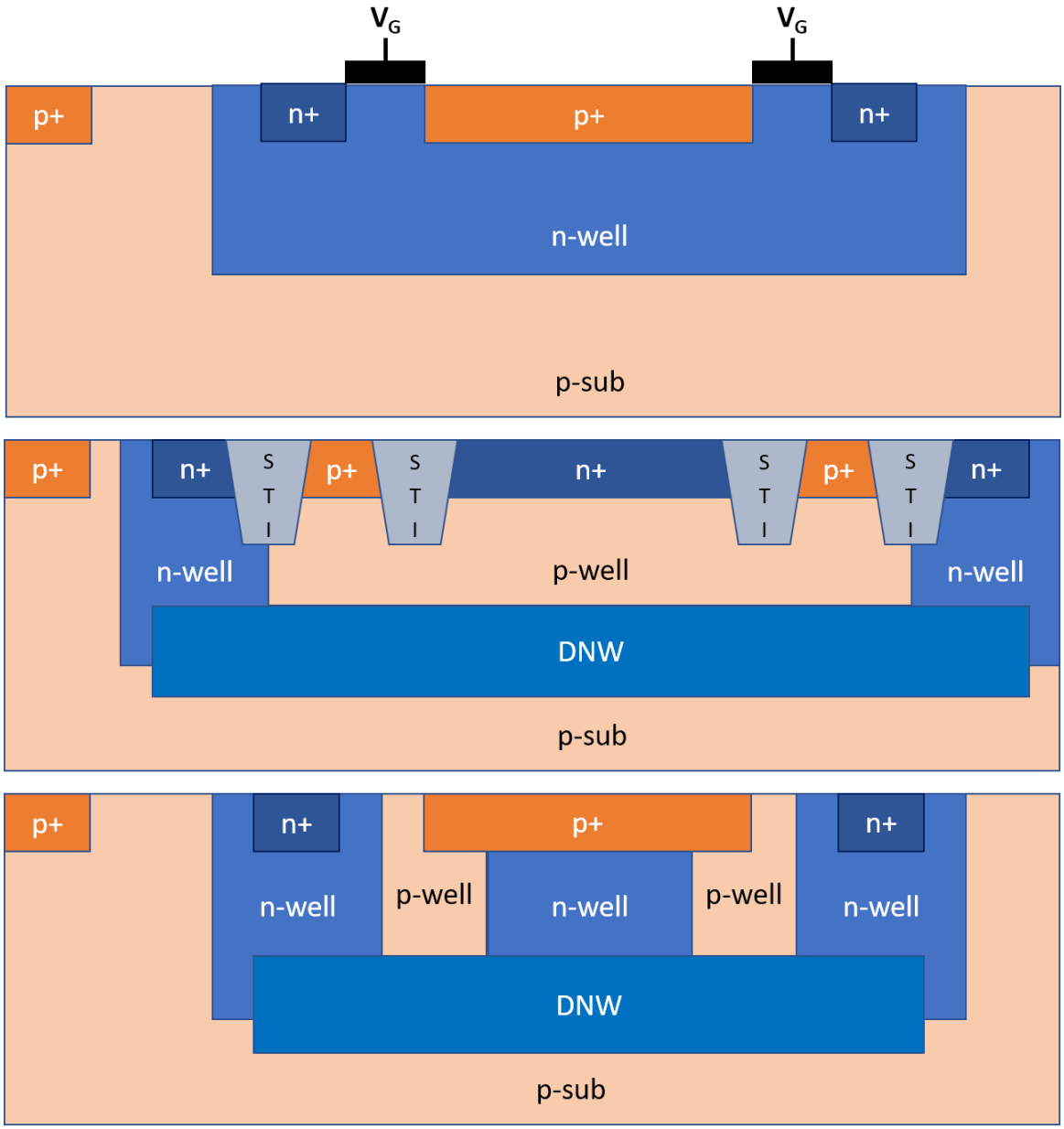


Figure 2-8: Perimeter field gate (top), STI GR (middle), and p-well diffusion GR for a p+/n-well SPAD (bottom).

2.2.3. In-Pixel Circuitry and Time-Gating

In order to design a functioning SPAD, circuitry must be in place to sufficiently quench the avalanching process and then reset the SPAD to be ready for the next avalanche event (either photon or noise triggered). This means bringing the junction voltage below breakdown upon detection of an avalanche event and later re-establishing the excess voltage above breakdown. In a passive quench and reset (PQR) circuit, a large quench resistor with values ranging from 50 to 500 kΩ is typically used. Due to the simplicity of PQR circuits, they are a great solution for reducing pixel size, increasing FF, and reducing parasitics. For this reason, PQR circuits continue to be used even today in commercially available SiPMs [8]. However, the use of a large resistance to properly quench the avalanche process leads to a long recharge RC time constant to re-establish the potential across the SPAD junction capacitance. This can lead to long dead times of several hundred nanoseconds and consequently low count rates.

More sophisticated active quench and reset (AQR) circuits have been employed to reduce and control the dead time of SPADs and consequently any AP effects [50], [61]–[63]. Here, integrated CMOS circuitry is typically used to detect the avalanche and control the quenching process by opening a path for the cathode to a voltage level that reduces the SPAD voltage below breakdown (usually ground). After a pre-determined (hold-off) time, the reset circuitry connects the cathode directly to V_{Reset}, bypassing the quench resistor and allowing the SPAD to reach the excess voltage more quickly and be ready for the next avalanche event. The duration of the hold-off time can be set to reduce the effects of afterpulsing to an acceptable level with typical values on the order of tens of nanoseconds [8]. Using such circuitry ultimately requires area and is usually implemented in-pixel, negatively impacting the FF of some SPAD designs, however this has been alleviated over time as smaller DSM CMOS technologies have been used. AQR circuitry can be utilized to effectively time-gate a SPAD and control the bias of one of the SPAD terminals so as to bring it below or above breakdown on demand through some input trigger or clock. The effect of this is illustrated in Figure 2-9.

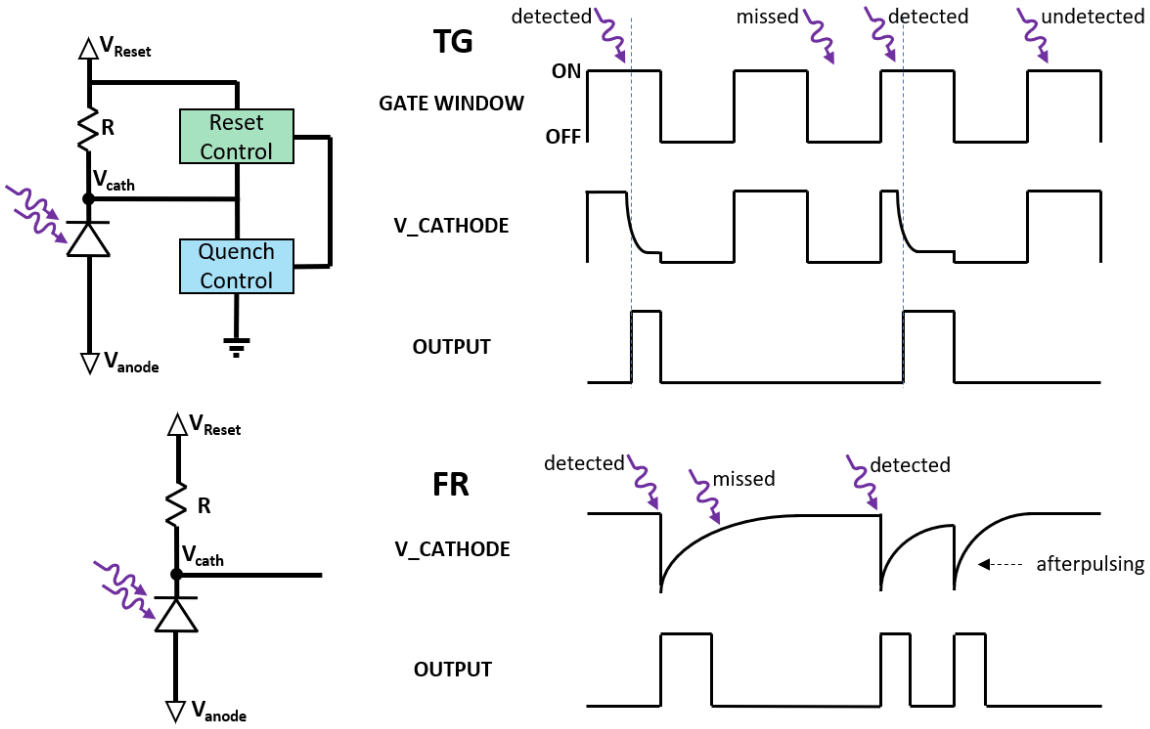


Figure 2-9: Comparison of the operation and output of time-gated (TG) and free-running (FR) SPADs.

A comparison of free-running vs. time-gating was done in [50] and showed that shallow-junction SPADs with larger (~100 μm^2) active areas can achieve substantially reduced AP effects when operated in the time-gated mode. With time-gating, optimal hold-off time and temperature conditions can be found to achieve <1% AP. Time-gated SPADs have demonstrated success for applications like Raman Spectroscopy [10] and show great promise when used in TR-DOI [23]. When compared to a non-gated SPAD, the time-gated one increased depth detection by up to 65% in reflectance DOT which is known to have less depth sensitivity when compared to transmittance geometry. This helps improve performance in situations with thick organs or organs that are too strongly absorbing.

The benefit of SPADs implemented in DSM CMOS technology is that other digital circuitry for processing and storage can be added on-chip and even in-pixel to create sophisticated SiPMs [64]. An example of a high-performance SPAD pixel is found in Figure 2-10 [65]. The 130 nm CMOS PQAR SPAD pixel contained an in-pixel counter and analog-to-digital converter (ADC) for high-speed operation. Time-to-Digital Converters

(TDCs) can also be included to obtain precise timing information of incident photons for biomedical imaging and ranging applications [66]. This, however, also decreases the FF and the added circuitry for active quenching and reset increases the parasitic capacitance on the sense node. Instead of trading PDE for dead-time reduction, using multiple SPADs per pixel by sharing a common well allows for parallel counting with minimal penalty in fill-factor [31], [67]. As SiPMs evolve to process the energy and timing information on-chip instead of requiring separate ASICs (Application-Specific Integrated Circuits) for handling the analog output, they become known as digital SiPMs (dSiPMs) for their fully digital readout.

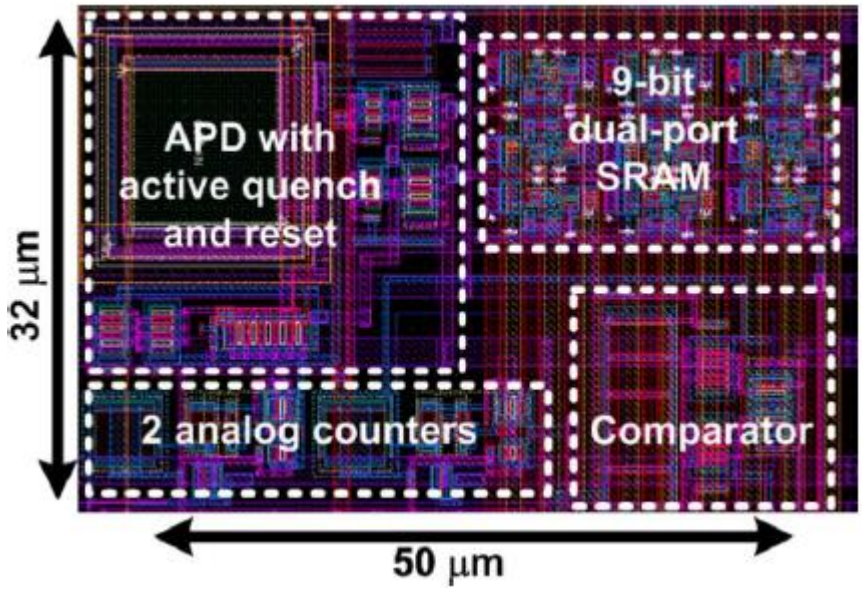


Figure 2-10: SPAD pixel layout with in-pixel analog counting and SRAM for parallel pixel analog-to-digital conversion [65] (© 2011 IEEE).

2.3. State-of-the-Art Performance and Structures

The key performance parameters and typical values reported in literature for SPADs fabricated in standard, image sensor, and high-voltage CMOS technologies are summarized in Table 2-1 [7], [21], [68].

Table 2-1: Typical ranges of key performance parameters in CMOS SPADs.

Parameter	**Value Range**
SPAD pitch (μm)	30-50
Peak PDE (%)	1-50
DCR (Hz/μm^2)	0.3-100
Fill-Factor (%)	1-60
Timing Jitter (ps)	30-100
AP Probability (%)	0.1-10
Dead Time (ns)	10-100

Advanced designs incorporating multiple SPADs per pixel, large array sizes, and on-chip TDCs have been developed, moving towards what is now state-of-the-art. CMOS SPADs are seen to be the best choice for imaging at the single-photon level and that among the different CMOS SPADs, state-of the-art devices are designed in 0.35 μm technologies with built-in TDCs, very low DCR and very large (30-100 μm) diameters [69]. The expense of these chips are their large sizes (5 mm x 5 mm) with just 1k - 2k pixel counts. Many other high performing CMOS SPADs are also fabricated in the 130 nm process where the smaller technology node allows for faster front-end circuitry, smaller pixel pitch, and higher pixel counts despite higher noise. A review on compact SPAD pixel architectures for time-resolved imaging, focusing on time-gating or time-stamping also investigated primarily 0.35 μm HV and 0.13 μm CIS processes [31]. It was stressed that device optimization would inevitably use specialized process options such as with CIS, and that all-NMOS topologies are best for optimized FF to exploit shared n-well SPAD layouts and transistor reuse among other things.

One specialized process option that has attracted interest in SPAD research has become 3D-stacked technology. This dramatically improves FF while enabling increased functionality, better timing, low power, and higher uniformity in all performance parameters. Here, the top-tier chip is the SPAD array with improved FF, while the data processing, compression, and data transfer occurs on the bottom tier, generally fabricated in a more advanced technology node. The two tiers would then be connected by through-

silicon vias (TSVs). This is to take advantage of optimizing both processes individually – the doping levels and profiles of larger technologies offer improved PDE and DCR performance for SPADs, while the smaller technologies enhance the accuracy, speed, power, and compactness of the timing circuitry, pixel-level memory, and processing. Sometimes smaller technology nodes are used for the SPAD layer if it is a custom process such as CIS where PDE improvements can be had. In [30], a 45 nm CIS process was used for the top tier which is back-side illuminated (BSI) while the bottom tier was in 65 nm CMOS. Micro-lenses were also incorporated to improve the FF from 31.3% to 50.6%. Its structure is provided in Figure 2-11.

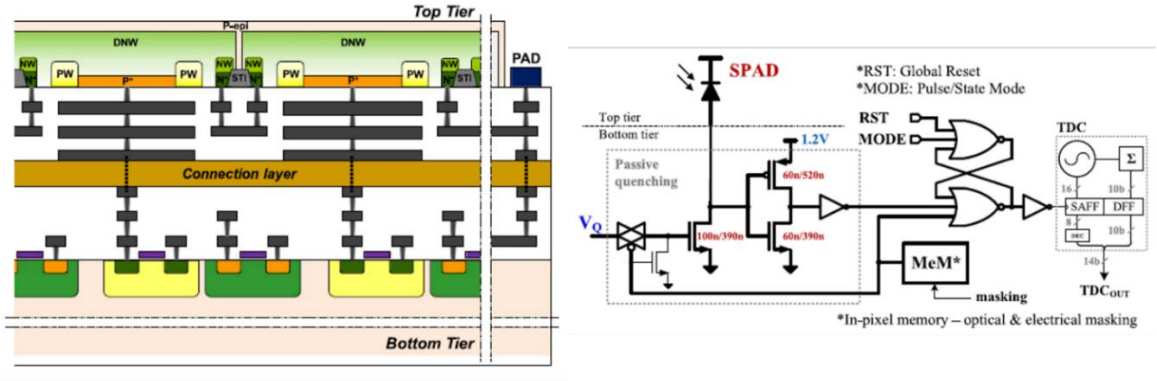


Figure 2-11: a) Cross-section of a back-illuminated 3D integrated SPAD, b) Schematic diagram showing the bottom-tier with passive quench, reset, memory, and TDC circuitry [30] (© 2018 IEEE).

Another special structure in a 0.16 µm BCD (Bipolar-CMOS-DMOS (Double Diffused Metal Oxide Semiconductor)) technology was devised with sharp timing response and red-enhanced sensitivity, particularly useful for the DOI optical window [70]. NIR photons are primarily absorbed by the n-type side of the depletion layer in a one-sided p+/n junction, meaning avalanches are triggered by holes which have a lower avalanche triggering probability when compared to electrons in Si. By using a high energy boron implant instead of low energy phosphorus to create a deep p/n+ junction, the avalanche becomes mostly initiated by minority electrons, resulting in higher PDE.

While great performance can be achieved with such custom processes, their main limiting factor is the increased cost and complexity in the fabrication and manufacturing of such devices. Sticking to more standard processes allows for mass production, lower costs and easier integration with existing circuits. Furthermore, CMOS SPADs are usually the only choice when multi-pixel arrays are required [71]. For instance, STMicroelectronics has fully industrialized the production of SPADs which can be fabricated and shipped in millions per week [72]. Utilizing its 130 nm CIS technology node, the SPAD is formed by a p-well and DNW implant, has managed >5% PDE at 850nm, a median DCR of 100 Hz at room temperature, and a dead time of 25 ns (40 MHz maximum count rate).

2.4. Conclusions

Current research on SPADs has been focused around DSM technologies in order to efficiently reduce the cost and size of the sensor and allow for easy scalability and integration with digital processing circuits. In order to maintain the trend of performance improvements, researchers have moved to expensive and custom processes including 3D structures and CIS, BCD, and BSI technologies [11], [30], [67], [70]. 3D technology allows for much higher fill-factor while enabling increased functionality, better timing, low power, and higher uniformity in all performance parameters, thanks to separating the SPAD and supporting circuitry into different tiers. It can also be said that some of the best performing SPADs have been designed a 350 nm or 130 nm CMOS process [53], [65], [67], [69], [71]. Furthermore, all-NMOS pixel structures seem to be necessary as well for optimized FF to exploit shared n-well SPAD layouts and transistor reuse [31], [67].

There is still, however, progress to be made for improving performance in the standard CMOS process in order to maintain low cost, high scalability and integration. SPAD performance is largely pre-determined when using a standard CMOS process, since the fabrication house has already set various parameters such as the number and type of layers, the layer depths, doping concentrations, operating device voltages, and well structures available. What can be decided by a user during design is primarily the SPAD dimensions, shape, pixel structure, junctions and guard rings used, and associated in- and off-pixel

circuitry. To overcome some performance limitations in standard CMOS SPADs, techniques such as multi-junction and TG SPADs have been employed to offset the inherently low PDE and reduce the effects of AP to ultimately increase SNR [9], [10], [23], [31], [50], [52]–[55]. However, these have not yet been explored in much detail in smaller, more advanced DSM CMOS technologies and no design has yet managed to take advantage of both techniques together due to the difficulty of biasing multiple junctions.

The proposed SPAD to tackle this forms three pn junctions: n+/p-well, p-well/deep n-well (DNW) and DNW/p-sub with minimal layer spacing for improved FF. A test TG pixel was also designed based on previous work [10], [73] that utilizes the top junction of this SPAD to observe the changes and improvements with using a more advanced technology node. Furthermore, the design and simulation of a dual-junction time-gated (DJTG) design is explored in an effort to combine the benefits of both without the need for AC-coupling to the SPAD junctions. Due to the lack of SPAD literature in this technology node, an in-depth characterization of the three pn junctions and the TG design is required before more sophisticated, multi-array SPAD structures can be developed.

TSMC's standard 65 nm CMOS technology was chosen for this research due to fabrication being readily available and in following with recent trends with regards to DSM TDC designs. At the current time, going to nodes smaller than this can see reduced fabrication opportunities, higher costs, and further PDE and noise performance degradation as a result of thinner depletion regions and higher tunnelling probabilities.

Chapter 3
Design of Multi-Junction and Time-Gated CMOS SPADs

3.1. Triple-Junction SPAD

3.1.1. SPAD Circuit Modelling

A model similar to the one found in [74] which considers the temporal SPAD behaviour, avalanche build-up and self-quenching mechanisms was used. A diagram of the model is presented in Figure 3-1, where inductors together with switches represent relays that switch on when a threshold is crossed.

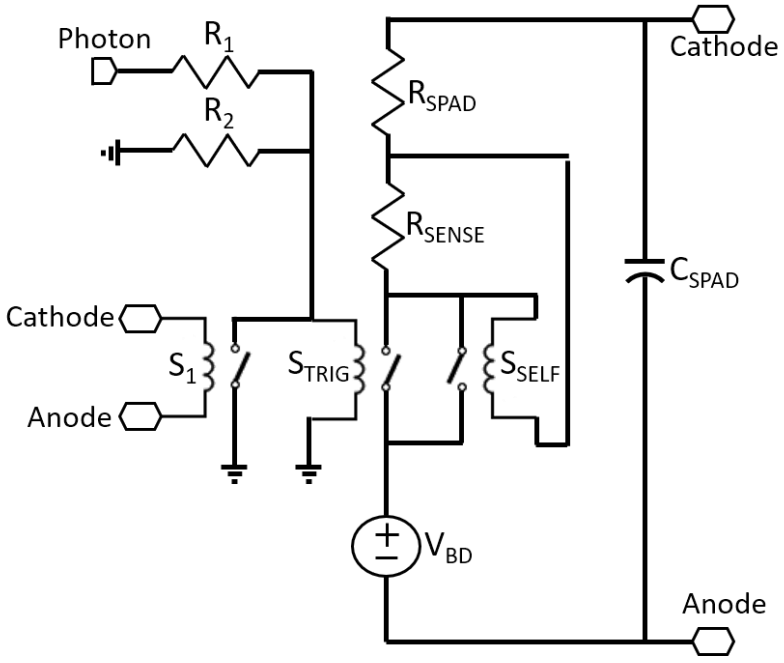


Figure 3-1: SPAD circuit model.

Initially, a pulse source simulates the "Photon" terminal that closes the S_{TRIG} switch. This initiates the avalanche process by discharging the SPAD's capacitance C_{SPAD} through the SPAD resistance R_{SPAD} that creates a fast and large current spike. This immediately closes S_{SELF} to sustain the avalanching current as well as closes S_1 (thus opening S_{TRIG}) which removes the influence of the width of the simulated photon pulse on the SPAD avalanche. The duration of the avalanche is determined by the current threshold of the S_{SELF} switch. Below a certain current threshold, there exists a high probability that no carriers exist within the depletion region after a random time, thus resulting in a quenched avalanche. This threshold value is not well-defined, but a value of 100 µA has been used in the SPAD literature for analytical calculations and simulation models [8] and was thus also used in this model for S_{SELF}. An R_{SPAD} value of 600 Ω was also used as an estimate based off the existing SPAD literature. The model was replicated for each junction and the breakdown and capacitance values for each junction was estimated from the TSMC process documentation.

3.1.2. Schematic and Layout

The SPAD layout is shown in Figure 3-2 with an illustration of the cross-sectional view. An n-well GR for the n+/p-well junction was proposed. However, n-wells designed in this process must touch the DNW and thus would isolate the p-well and prevent access to it, rendering the top shallow junction unusable. In order to fabricate a TJ SPAD in this technology, STI was used as a GR. To obtain the highest FF possible for the SPAD, the design rules of the TSMC 65 nm CMOS process was followed carefully by using the minimum spacing requirements between the process layers such as between the n+ and p+ layers as well as the minimum overlap and extension of the n-well over the DNW. A silicide block layer was used over the active area and surrounding it were metal layers above the ones used for routing which acted as shielding so that photons directly incident on the SPAD would most likely be absorbed in the planar junction regions. The active area of the SPAD was designed to have a side length of 10 µm with corners at 45 degrees to reduce PEB. This makes the total active area to be approximately 100 μm^2. The distance from the

center of the SPAD to the edge of the n-well measured at 7.26 µm resulting in an active area that makes up ~47% of the total SPAD area. A pixel designed with this SPAD will ultimately have FFs <47% due to the area considerations of any extra circuitry required for the pixel.

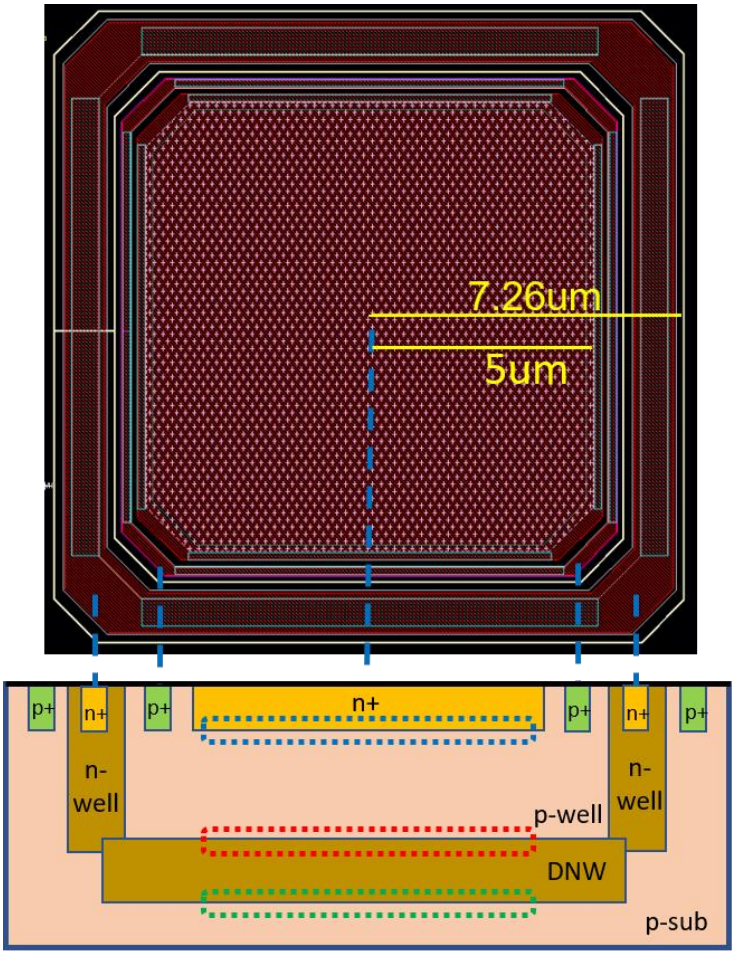


Figure 3-2: Layout of the SPAD with minimal clearance between process layers for highest obtainable FF. Below is a look at a cross-sectional view of the SPAD.

In order to test the breakdown voltage and run tests on the TJ SPAD, the circuit was designed as in the schematic in Figure 3-3. Connections VSPAD1, -HV, and VSPAD23 were used to bias the junctions either simultaneously or independently, and 50 kΩ resistors were made and connected to the junctions to passively quench them upon breakdown. Test points J1_TEST and J23_TEST provide a way to bypass the resistors if needed. With this,

the breakdown voltage of the three junctions can be determined and the cathode voltage waveforms can be observed through an oscilloscope. Future mentions of J1 refers to the top junction (n+/p-well) and J2 and J3 refer to the middle (p-well/DNW) and bottom (DNW/p-sub) junctions, respectively.

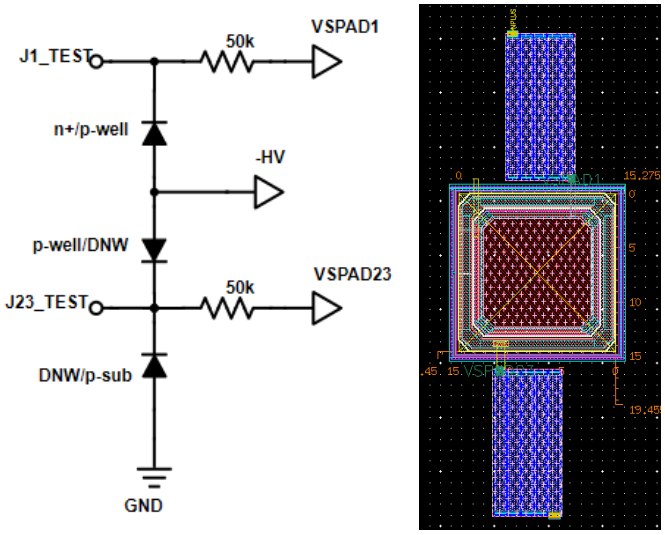


Figure 3-3: Schematic and corresponding layout of the passively quenched triple-junction (TJ) SPAD. The blue rectangles are the n-well resistors.

This PQ TJ SPAD can suffer from the load capacitance on the junction when being tested with an oscilloscope probe and this could affect the SPAD performance. To reduce the load capacitance, an on-chip solution was made where a TJ SPAD was AC-coupled to two output buffers (Figure 3-4) with pull-up transistors; one for J1 and the other for J2/J3 since they share the DNW junction. This means avalanche events by both J2 and J3 cannot be simultaneously detected, however differences in the width of the VOUT2 can potentially be used to differentiate between the two junctions [53]. The AC-coupling was needed since the high voltage levels required to properly bias more than one junction at once is not compatible with the MOSFETs in the technology which are rated for at most 3.3 V.

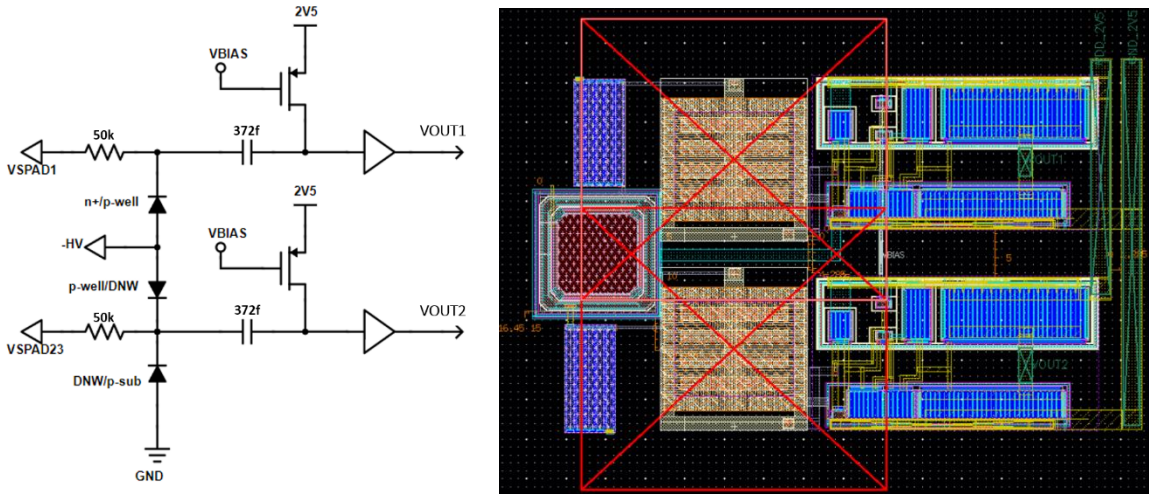


Figure 3-4: Schematic and corresponding layout of the triple-junction (TJ) SPAD with AC-coupled output.

3.1.3. Simulation Results

The simulation of the AC-coupled TJ SPAD in Figure 3-5 includes both pre- and post-layout results. Pre-layout is the ideal schematic simulation behaviour while the post-layout results include the extracted parasitic resistances and capacitances from the design of the layout. It is evident how the added parasitics contributes to increased delays in the recharge of the SPAD which results in wider output pulses. This limits the count rate and overall speed at which the SPAD can operate, as well as increase AP due to the increased number of charges that can be trapped during avalanche as described in Section 2.1.4.

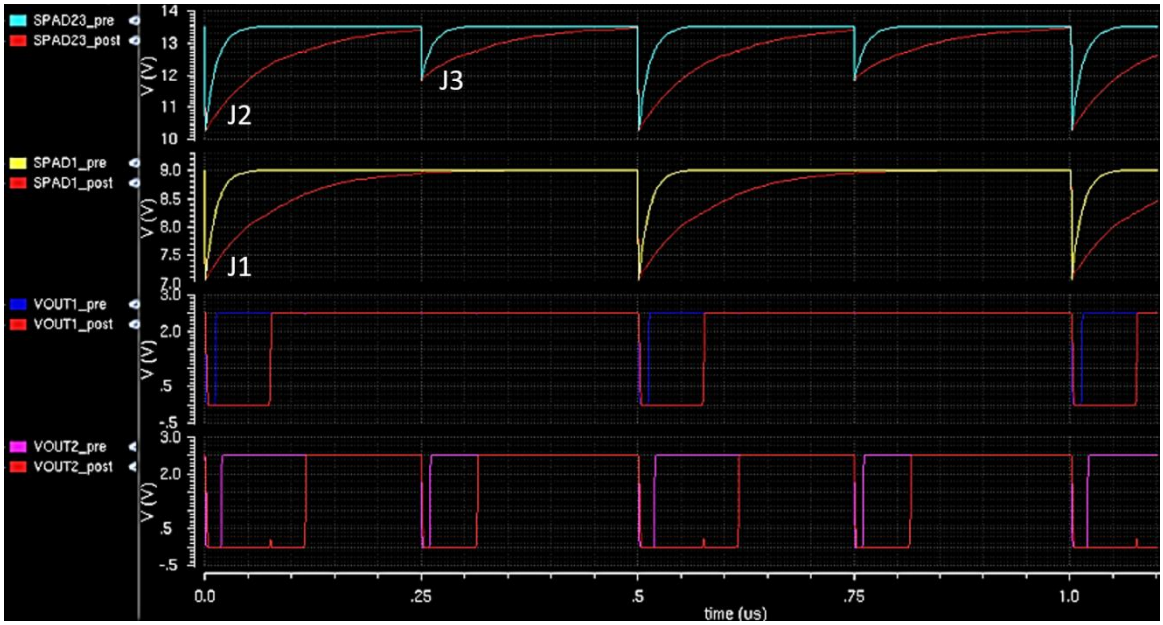


Figure 3-5: Simulation of the three junctions undergoing breakdown and recharge using the SPAD circuit model and the resulting AC-coupled output pulses.

3.2. Time-Gated Single-Junction SPAD

3.2.1. Schematic and Layout

The implemented TG SPAD pixel was based on the design in [10] which was previously fabricated in an IBM 130 nm process and showed good results in its performance and application in Raman Spectroscopy. This allows the test and comparison of our 65 nm CMOS design with the 130 nm CMOS process to observe performance changes and verify results between technologies. To understand the function of the pixel, a schematic is provided in Figure 3-6. Not counting the buffer, the pixel is composed of 5 transistors and requires 3 pulses to allow time-gated (TG) operation. Initially, the 3 pulses are held high to isolate the SPAD cathode (VC) from VDD and hold it and the output (VOUT) at GND. A negative high voltage (-HV) is applied at the anode of the SPAD at all times that keeps the junction voltage just below breakdown. Two simultaneous low input pulses (P1 and P2) closes NMOS M2 to isolate the SPAD cathode from GND and opens PMOS M1 to connects the cathode to VDD for a brief moment. P1 lasts only a brief moment, while P2 stays low

for the entire duration which the SPAD is to be armed, effectively isolating and holding the SPAD cathode at VDD for the length of the gate window. A low P3 pulse occurs as soon as P1 is finished which enables the readout circuit to detect and register an avalanche event throughout the duration of the gate window. Finally, P2 and P3 go high again to close the gate window and bring the SPAD below breakdown again. The process is then repeated, with pulses P1, P2, and P3 acting according to some input trigger or clock.

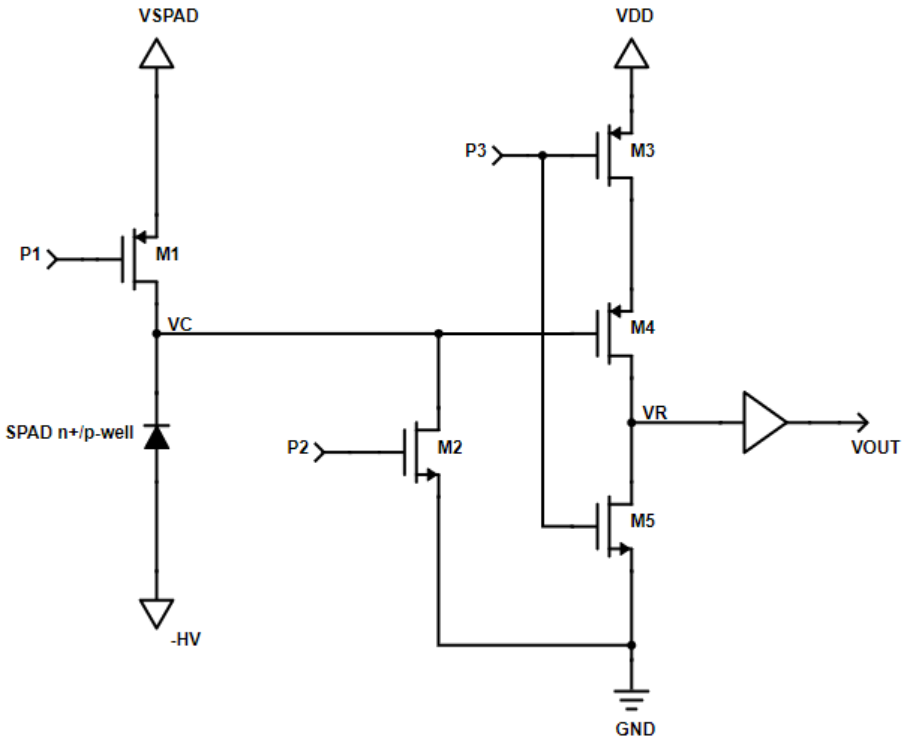


Figure 3-6: Time-gated (TG) SPAD pixel schematic.

Generating the required pulses on-chip is essential to creating short and fast gating with simple synchronization while removing the need for creating a complicated external input. To create a low pulse, the output of a NAND gate can be used where one input to the gate is inverted and slightly delayed from the other. A delay can be created using an inverter and capacitor, where multiple of these can be stacked in a chain to increase the delay. Thus, the creation of the three pulses can be done using the circuit illustrated in Figure 3-7. Given that P3 acts as soon as P1 has finished and lasts the remaining duration of P2, it was much simpler to use this relationship and logically generate P3 as a result of P1 and P2. This cuts

the required area of the on-chip pulse generation considerably since capacitors are costly elements in a DSM CMOS process due to their size. A separately generated P3 as in [10] would require roughly the same amount of space as that of generating P2. Furthermore, given that P3 must start after P1 and end no later than P2 to prevent false outputs, it is important to implement it based on the other two pulses to better ensure tighter operation.

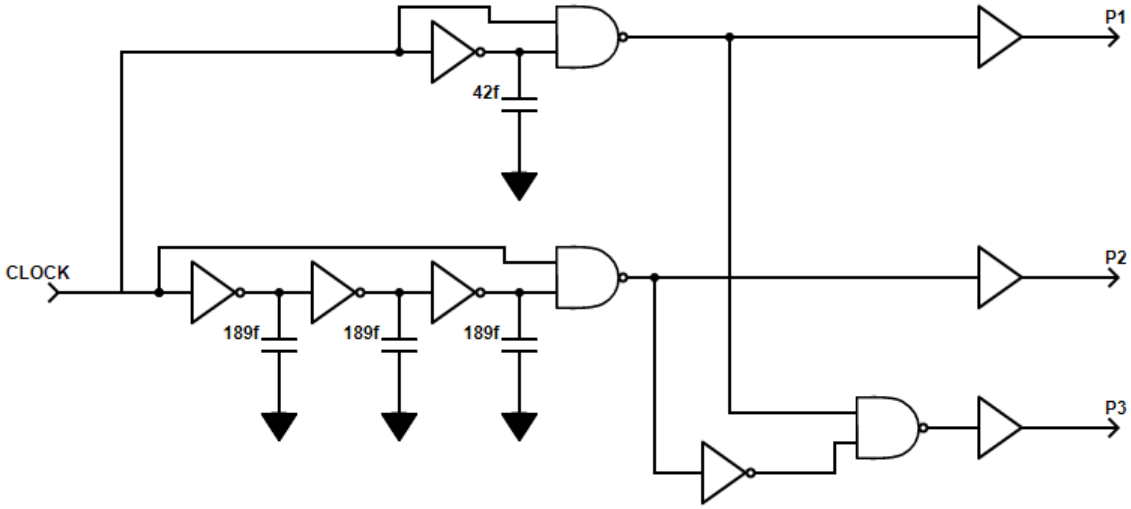


Figure 3-7: On-chip pulse generation circuit schematic.

The duration of P1 was set around 100-200 ps with M1 large enough to allow a rapid reset of the SPAD that limits avalanche events before the SPAD has fully reached the applied excess voltage. This keeps the SPAD response consistent and reduces the probability of avalanches occurring before the gate window (P3). Furthermore, in the context of TR-DOI applications, the fast transition time allows the detection of late photons without being saturated by early photons, thus enhancing the DR and SNR [9]. P2 was set to last ~3 ns to be close to the 3.5 ns window of [10] for comparison, but for an improved theoretical count rate. As a result, P3 was designed as the difference between P2 and P1.

Combining the TG SPAD pixel structure with the pulse generation circuit and an output buffer resulted in the layout presented in Figure 3-8. Here it is clear how costly the metal-insulator-metal capacitors (mimcaps) are with regards to their size within the CMOS process. The mimcaps utilize the top metal layers and are shielded from the bottom to allow the layout of structures and additional circuitry directly below, although it is not

recommended. Fortunately, this pulse generation structure is only needed once if it is used to gate an array of TG SPAD pixels since only the resultant pulses need to be properly buffered and routed to the SPAD array. In this current test configuration, the TG pixel had an area of ~350 μm^2 which sets its FF at ~28.6%, which is better than the 9.8% FF attained in [10].

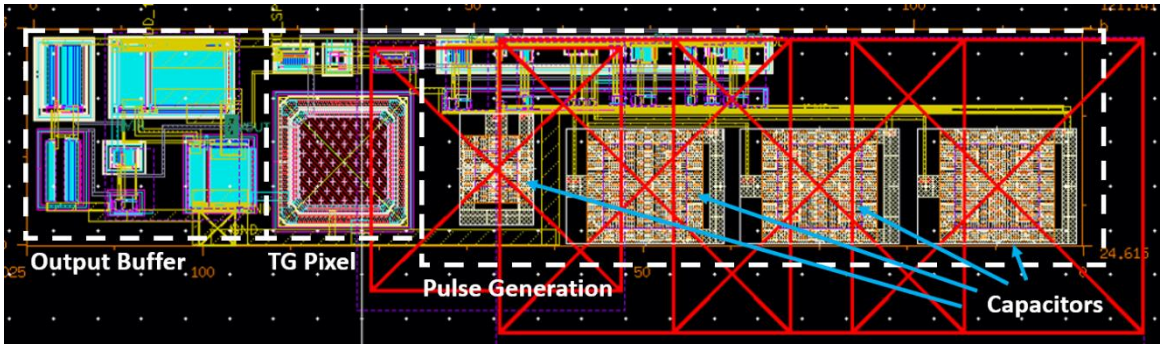


Figure 3-8: Layout of the TG SPAD design including the pixel, pulse generation, and output buffer.

3.2.2. Simulation Results

A simulation demonstrating the operation of the TG SPAD pixel is shown in Figure 3-9 where the pulse generation circuit arms the SPAD for ~3 ns every 10 ns using a 100 MHz clock input. Simulated photon pulses arriving at different times within the gate windows show the change in the width of the SPAD output, and no output occurs when the photon pulse arrives outside a gate window. Since the cathode is fixed to have voltage range of 0 – 1 V (GND to VDD), the excess voltage is similarly limited to only a certain range. The excess voltage is determined by the negative voltage applied to the SPAD anode but should not exceed 1 V else the cathode voltage will drop below GND level upon quenching. Due to the cathode being held at GND during hold-off in-between gates, the SPAD will thus always be biased above breakdown and may avalanche when not intended. A lower limit also exists since the voltage drop has to exceed the threshold of the PMOS M4 at the readout circuit (Figure 3-6) to trigger an output. This gives a usable excess voltage range of roughly 0.3 – 1 V. A leaking effect in the post-layout results is seen after a photon is detected where the SPAD cathode voltage rises slowly after an avalanche occurs. This may arise from the

large M1 transistor sized to quickly recharge the SPAD. Furthermore, slight delays in the pulses are seen due to the increased parasitic capacitances, however the logical generation of P3 from P1 and P2 allow P3 to rise in sync with P2 to prevent process variation mismatches.

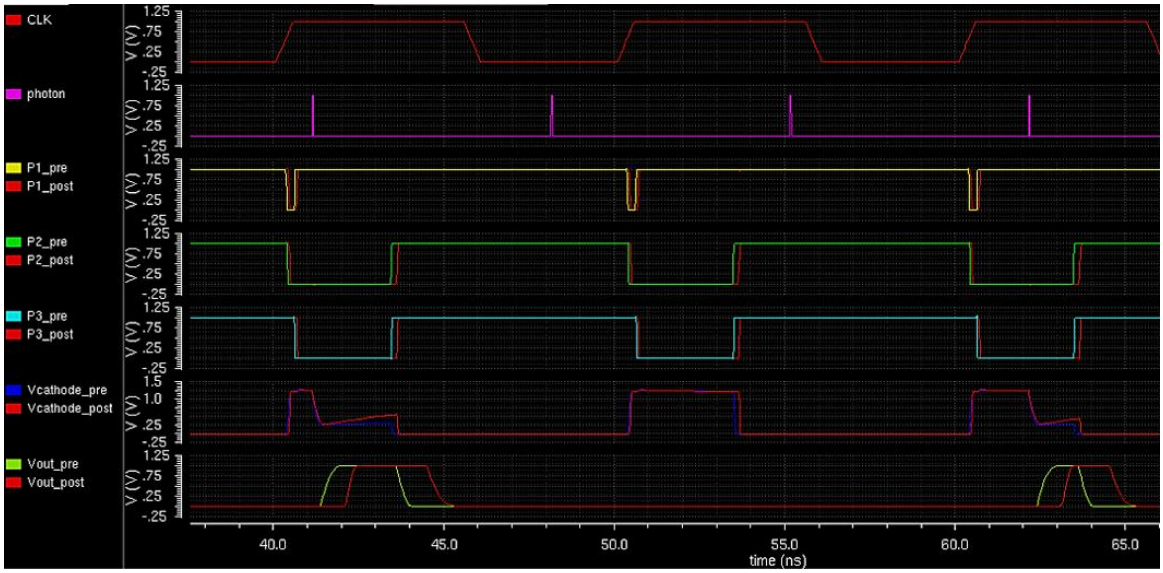


Figure 3-9: Simulation results showing the TG SPAD operation.

3.3. Dual-Junction Time-Gated SPAD

3.3.1. Schematic and Layout

Utilizing two junctions simultaneously in a TG design opens the door to improved count rates, higher PDE, and possible wavelength distinction based on which junction avalanched. To do this, the previous TG design was extended to include the p-well/DNW junction. The deeper DNW/p-sub junction could not be considered due to having the p-sub connected to GND which limited its biasing potential. Using the top two junctions still complicated the biasing since they both shared the p-well. To overcome the strict biasing conditions in multi-junction designs, AC-coupling the cathode or anode of the SPAD to readout circuit is common as was done in the TJ SPAD design above and in previous works such as in [53]. However, a TG design requires active quench/reset circuitry which cannot

be done through AC coupling. The goal was to develop a possible method of gating two junctions of a SPAD while meeting the biasing conditions of the junctions.

Replicating the TG design from section 3.2 for the second junction is not possible. Based on available information from the 65 nm CMOS process, the breakdown voltages of J1, J2, and J3 were estimated to be roughly 9 V, 12.2 V and 11.8 V, respectively. To ensure J1 stays below breakdown when not gated (and the cathode goes down to GND), the -HV node should be capped -9.0 V at the lowest. This requires the J2 cathode to be able to reach ~3.2 V just to reach breakdown and should have more headroom to achieve reasonable excess voltages. The thickest oxide MOSFETs in the technology are rated for 3.3 V which made this impossible.

One method to extend the range of V_{EX} possible is to replicate a design similar to [75] in which biasing is done from the cathode and detection is done instead from the anode using cascoded thick-oxide NMOS rated for 2.75 V. A V_{EX} of 4.4 V can be reached despite this rating because the voltage is distributed between the two cascoded NMOS and limits the midpoint to 2.2 V to ensure the gate oxide integrity of the devices. Creating a DJTG SPAD with J1 and J2 using the anode for readout in a similar manner could mean missed counts and increased difficulty in distinguishing which junction avalanched. Furthermore, due to J2 having a higher breakdown voltage in this technology than the deeper J3, if anything higher than 11.8 V is applied at the J2 cathode in an attempt to bring it above breakdown, then J3 would start to avalanche.

One method to properly bias both junctions would be to safely lower the anode voltage (-HV) below the breakdown of J1 to give J2 more flexibility at its cathode. To do this, the DNW process is exploited by using NMOS transistors with bodies isolated by a DNW. Where typically the body of an NMOS is connected to its source (at GND), one isolated by a DNW can have its body be driven at a custom voltage. If connected to a negative voltage (but not too negative so as to breakdown the p-well body/DNW junction), the terminals of the NMOS can be driven to negative voltages as well while maintaining the integrity of the gate oxide. Such a method has not been properly tested in literature, but a design based on this approach is given in Figure 3-10 with testable voltage parameters.

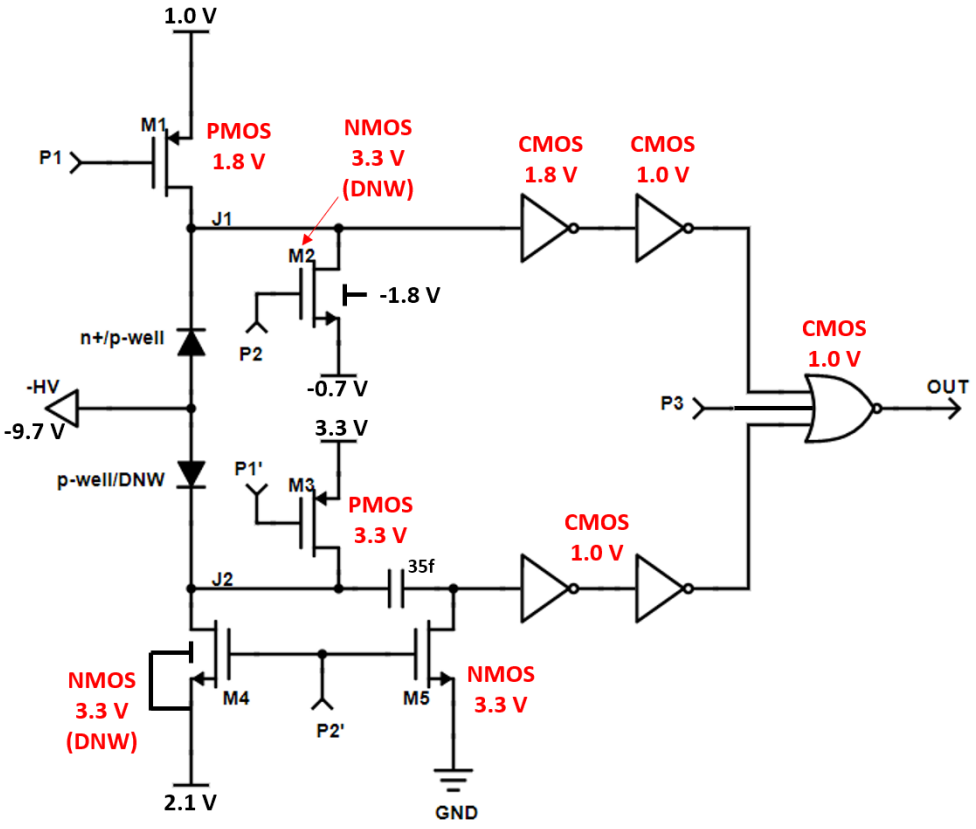


Figure 3-10: Schematic of a dual-junction time-gated (DJTG) SPAD utilizing MOSFETs with different gate oxide thicknesses. Voltages at the various nodes given produce working simulation results.

In this design, the quench and reset of the two junctions are driven by separate pulses in an alternating fashion with a shared readout which prevents missed detection events if both fire simultaneously. Here, an avalanche is detected when one junction drops low during its gate while the other is held-off (also low). Along with pulse P3 being low, a high output pulse is best given by utilizing a NOR gate with those three as input. Any other configuration does not produce an output which is ideal. Furthermore, the -HV node could be pushed down to -9.7 V, with M2 allowing the J1 cathode to reach -0.7 V during hold-off so that J1 can be reduced to just below breakdown when not gated. Through simulation, a body voltage of -1.8 V must be applied to M2 to achieve this. M1 and the inverter attached to the J1 node must use gate oxides rated for 1.8 V to account for the swing between 1 V to -0.7 V. Similarly, M2 had to use the thickest oxide rated for 3.3 V. The pair of inverters buffer the cathode voltage drop upon avalanche. They are also used to step down the voltage

in a simple manner by driving the 1.8 V inverter with a VDD of 1.0 V at the cost of slight voltage spikes during transitions (visible in Figure 3-15).

On the J2 side, breakdown now occurs when its cathode is above 2.5 V which gives more room for excess voltage to be applied and have detectable avalanches. The unconventional voltage change between the reset voltage 3.3 V to 2.5 V required the use of a capacitor to transfer the roughly 1 V change along to the buffer to produce cleaner pulses to the NOR gate. During the gating process, the drain of M5 rises and falls (upon avalanche) together with the J2 node. During hold-off, M5 acts as a pull-down transistor to reset the voltage after the capacitor to GND. Thus, while AC-coupling was used to successfully simulate this design, the goal of attaching active circuitry directly to two junctions of the SPAD and time-gating them was still successful.

In order to properly drive this pixel design, a more robust pulse generation circuit is required than in the original TG design. With the junctions being driven in an alternating fashion, such a design can either double the count rate, or keep the combined count rate similar to a single-junction TG design while reducing AP (due to increased/doubled hold-off time per junction). This was accomplished using the circuit in Figure 3-11.

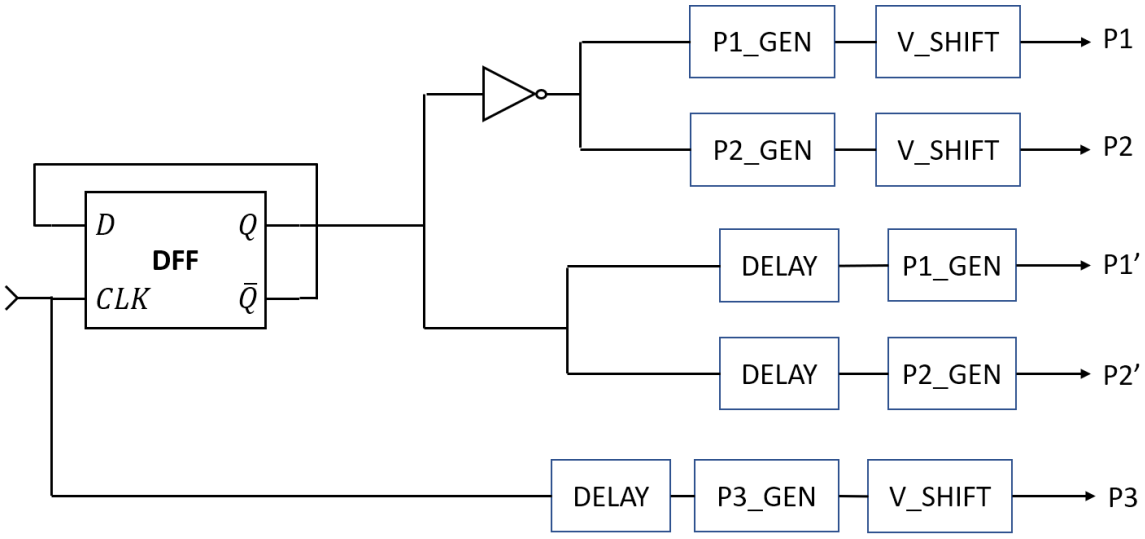


Figure 3-11: Overview of the pulse generation circuit that provides separate and alternating qunch and reset pulses for both SPAD junctions (P1/P2 for J1 and P1'/P2' for J2) using a divide-by-2 D flip-flop (DFF).

A single clock or trigger input that can generate the pulses for both junctions was still desirable for simplicity, so a D flip-flop (DFF) was used to divide the input clock by 2. The output of the DFF was split in two; one line to generate P1' and P2' (for J2) when the clock goes high, and the other is inverted so that P1 and P2 (for J1) is generated when the clock goes low. Since J2 required pulses at 3.3 V, the circuit was designed with an input clock with a high of 3.3 V, and P1_GEN and P2_GEN blocks were designed with 3.3 V input accordingly. However, J1 required 1 V pulses, so voltage shifting was required, similar to the buffers in the DJTG pixel. Due to the TSMC 65 nm CMOS technology having different gate oxide thickness to support MOSFETs rated for 1, 1.8, 2.5, and 3.3 V, a simple voltage shifter circuit (V_SHIFT) was made consisting of a chain of inverters of MOSFETs of decreasing gate oxides being driven at a VDD of the next smallest voltage. Although small voltage and current spikes occur between inverters, simulations showed they were small and short-lived.

Due to the combined output of the DJTG pixel from both SPAD junctions, P3 had to function on every clock cycle unlike P1/P2 or P1'/P2' that occur every other cycle. This meant P3 could not easily be logically created from the other pulses in this design and had to be generated separately. Extra delay (DELAY) blocks using inverters and capacitors were added into the P1', P2', and P3 lines to properly synchronize all the pulses.

The pixel and pulse generation circuit layouts were implemented and are shown in Figure 3-12. Again, it is evident how costly capacitors are in a CMOS process in terms of space. The pixel itself achieved a FF of 13.2% and some of the larger transistors had to be resized due to charge injection that occurred during post-layout simulations discussed below.

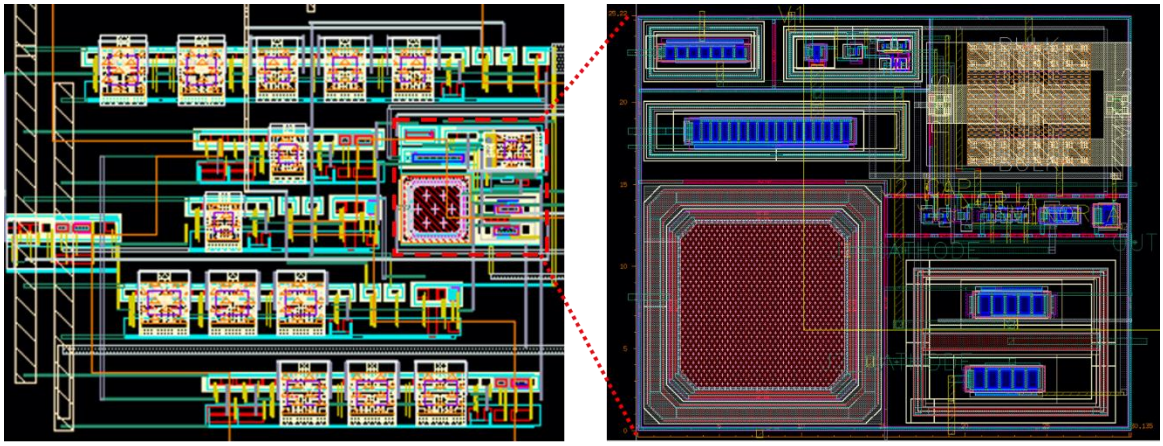


Figure 3-12: The layout of the DJTG pixel together with its pulse generation and a zoom-in of the pixel.

3.3.2. Simulation Results

Figure 3-13 shows simulation results of the DJTG before considering parasitic effects from layout using a 100 MHz clock as the input trigger to the new pulse generation. The figure shows the junction cathodes, their individual buffer outputs that feed into the final NOR gate (norA for J1 and norB for J2), and finally the output of the NOR gate. The pulses were omitted for clarity, but are similar in effect and gate length to the original pulse generation. The only difference is P1, P2, and P3 have a high of 1 V, while P1' and P2' have a high of 3.3 V. Furthermore, P3 occurs every clock cycle while P1/P2 and P1'/P2' alternate to arm the two junctions in an alternating fashion. Once again, different output pulse widths are observed depending on when the photon arrives within the gate window, and of course shows no output when no photon trigger is present within a gate window. In comparison to the results of the TG pixel in Figure 3-9, the junctions here have more than double the hold-off time to effectively reduce AP effects while maintaining an overall equivalent count rate.

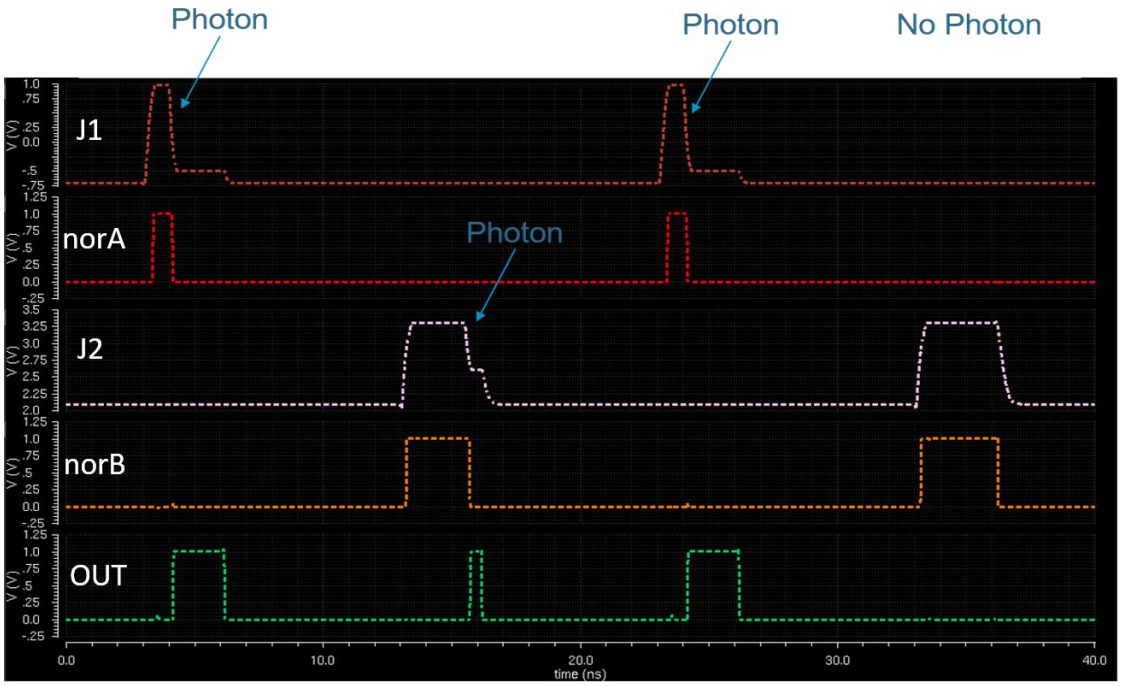
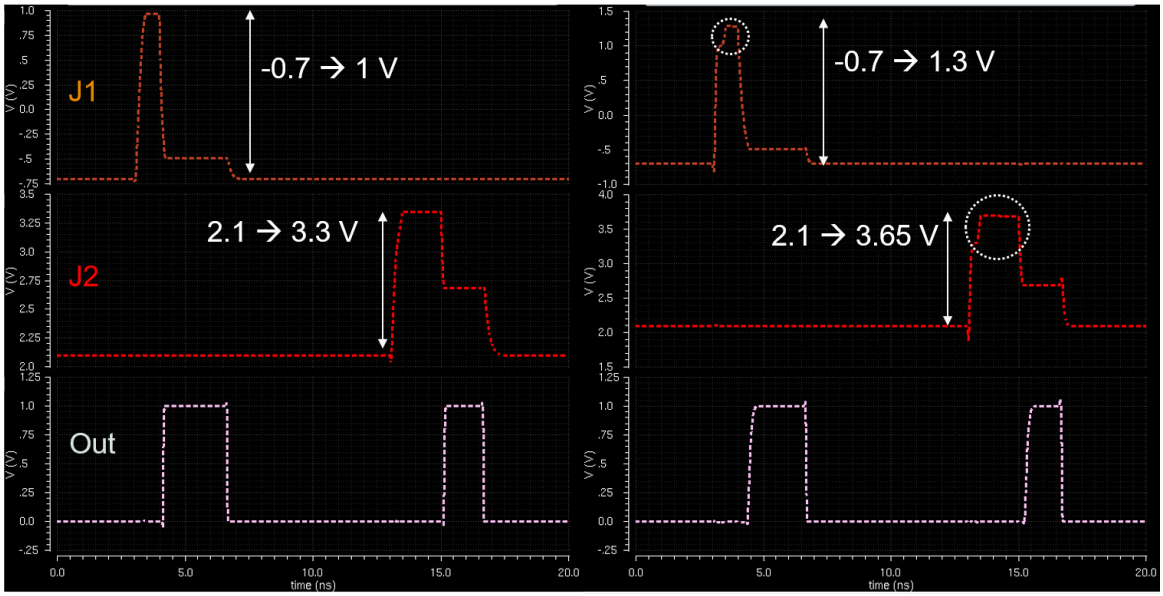


Figure 3-13: Schematic simulation results illustrating the function of the DJTG with both junctions being gated in an alternating fashion.

When comparing pre- vs. post-layout simulations, modifications had to be made to tackle certain issues. The many capacitors required to create the pulse generation circuit and lack of logic to generate P3 as was done with the pulse generation for the TG pixel resulted in mismatched pulses and incorrect gating, thus a comparison figure was not produced. However, the capacitors were re-sized and corrected to be used in conjunction with the post-layout extractions of the DJTG pixel. This produced gate windows where the cathode showed signs of charge injection due to having large MOSFETs for fast gating. The effect is shown in Figure 3-14 that shows the simulation comparison between pre- and post-layout. The overshoot is evident the moment P1 (P1') pulses go back high, M1 and M3 are turned off, and the cathodes are isolated.



(a) (b)

Figure 3-14: Pre-layout parasitic extraction (a) vs. post-layout extraction (b) simulation results. Charge injection effects become apparent due to large transistors used.

Finally, the side-effects of the simple voltage level shifter implemented is seen in Figure 3-15. The mid-buffer waveform shows the voltage output of the first inverter in the buffer with gate oxides rated for 1.8 V. The source of the PMOS in the inverter is connected to 1.0 V to match the gate oxide requirements of the next inverter. The post-layout simulation reveals the voltage spikes that can occur as result of this method which may impact the reliability and longevity of such a design.

Overall, while the DJTG design was simply a proof of concept, the need for careful post-layout considerations for a more robust design is evident through these simulations. The added information due to parasitic capacitances and resistance reveal the inconsistencies of delays induced by capacitors and the effects of charge injection and voltage spikes.

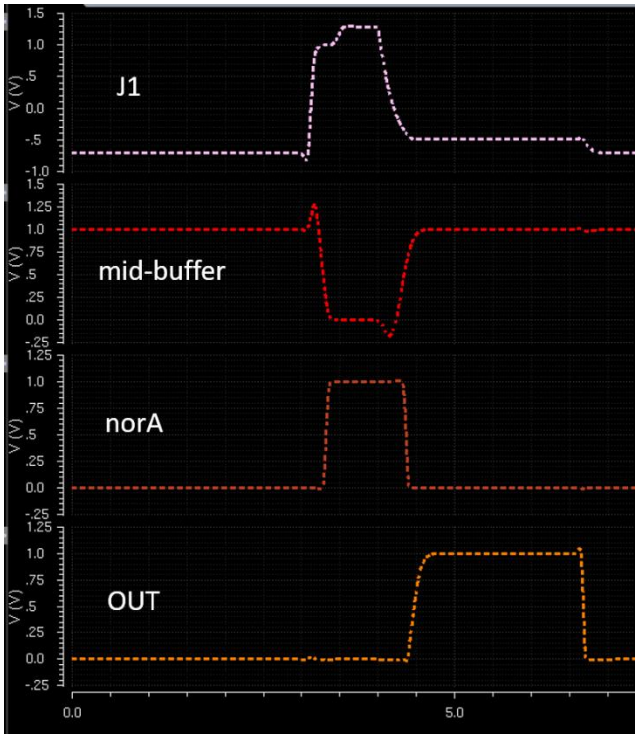


Figure 3-15: Post-layout simulation of the voltage spikes that can occur during transitions between inverters in the simple voltage level shifter chain that was implemented.

3.4. Conclusions

A TJ SPAD (J1: n+/p-well, J2: p-well/DNW, J3: DNW/p-sub) was designed on the basis of improved PDE and possible wavelength distinction due to which junction avalanched. A circuit model was used to simulate the SPAD avalanche response for the three junctions. The SPAD was designed in a square shape with corners cut at 45 degrees to reduce PEB. A high FF of 47% was achieved for the SPAD active area relative to its structure. This came at the cost of using STI for the GR structure which was expected to produce higher than expected noise when fabricated and tested. Simulations showed the effect of parasitics influencing the recharge rate of the SPADs which affects its dead time and ultimately count rate. The TJ SPAD was implemented using an unbuffered design as well as an AC-coupled design to reduce the load capacitance on the SPAD junctions.

A TG pixel design based on [10] previously implemented in a 130 nm process was also implemented and simulated. The smaller 65 nm process allowed a much better FF of

~28.6% which is better than the 9.8% of the previous design, however the use of 1.0 V transistors meant the excess voltage was limited to range of roughly 0.2 - 1.0 V. The pixel structure used only 5 transistors and required 3 pulses (P1, P2, and P3) to operate: P1 reset the SPAD cathode voltage above breakdown, P2 isolates the cathode from GND, and P3 dictates the gate window in which an output can be read. The pulse generation circuit was improved from the previous by logically generating P3 from P1 and P2, saving valuable space and improving the reliability. Its operation was also successfully simulated with post-layout parasitic extractions.

Finally, a DJTG pixel was designed and simulated as a proof of principle for creating TG pixels where two junctions that share an anode are both quenched and reset using active circuitry from their cathodes. A new pulse generation circuit was also developed to gate the junctions in an alternating fashion. With both junctions in use, the combined count rate could either double or be consistent with the previous TG design with the added benefit of longer hold-off times for reduced AP effects. To accomplish this, NMOS transistors meant to quench the junctions were implemented in a DNW. This allowed the body of the NMOS to be biased at a voltage other than GND so that negative voltages could be reached. The design was implemented and simulated to show proper operation despite challenges with regards to the capacitor delays, charge injection, and voltage spikes. Such meticulous crafting of the MOSFETs has its design flaws with regards to reliability and limited/restricted operation at specific excess voltages. Also, such a procedure is heavily reliant on the breakdown voltages of the junctions used when sharing a node with each other. Ultimately, this design was not tested in the following chapter, but demonstrates the plausibility of manipulating the MOSFETs implemented within a DNW to isolate its body and drive unique voltages for future, more robust implementations.

Chapter 4
Measurement and Characterization of CMOS SPADs

4.1. Design of Printed Circuit Board for Testing

The SPAD designs were implemented on a 1.5x0.8 mm^2 area chip that was fabricated and packaged using a 68-pin PGA package. A printed circuit board (PCB) was designed to interface with a SPAD chip and is shown in Figure 4-1.

Figure 4-1: Printed circuit board (PCB) designed for testing the SPAD chips.

It supplies the necessary supply voltages to all logic circuitry through one input 3.3 V connection (top left) which is also split and regulated down to 2.5 V, 1.8 V, and 1.0 V supplies. Three other inputs (left) provided separate and variable biases to the possible three

junctions of the various designs. On the other side, SMA connectors were used to connect the SPAD output signals to an oscilloscope and other testing equipment. Due to the many designs on chip and the fact that most are tested independently from each other, jumpers were a simple solution in reducing the number of required inputs and outputs. One SMA input on the left is used to provide a CLK input for the TG design and is terminated with a 50 Ω resistor. The schematic design for the PCB is provided in Appendix A-1.

4.2. Results and Discussions

4.2.1. Breakdown Voltage

Determining the SPAD's breakdown voltage is the first step in the evaluation of its performance. A consistent breakdown voltage among a number of devices helps give initial indication as to the reliability of the process and the design of the SPAD. The breakdown voltages of five TJ SPADs (Figure 4-2) were determined at room temperature using an Agilent B1500A Semiconductor Device Analyzer by varying the SPAD junction voltage until a sharp increase in current was observed. The data was transferred to Matlab where the breakdown voltage was determined as the point which the current increased by a factor of around 3 to 4 times the previous data point, represented in Figure 4-2a). The result is presented in Figure 4-2b) which shows the average breakdown voltages of junctions J1 (n+/p-well), J2 (p-well/DNW), and J3 (DNW/p-sub) to be 9.52 V, 12.59 V, and 12.35 V, respectively. The standard deviation of breakdown voltages for J1, J2, and J3 are rounded as 0.05, 0.07, and 0.03, respectively.

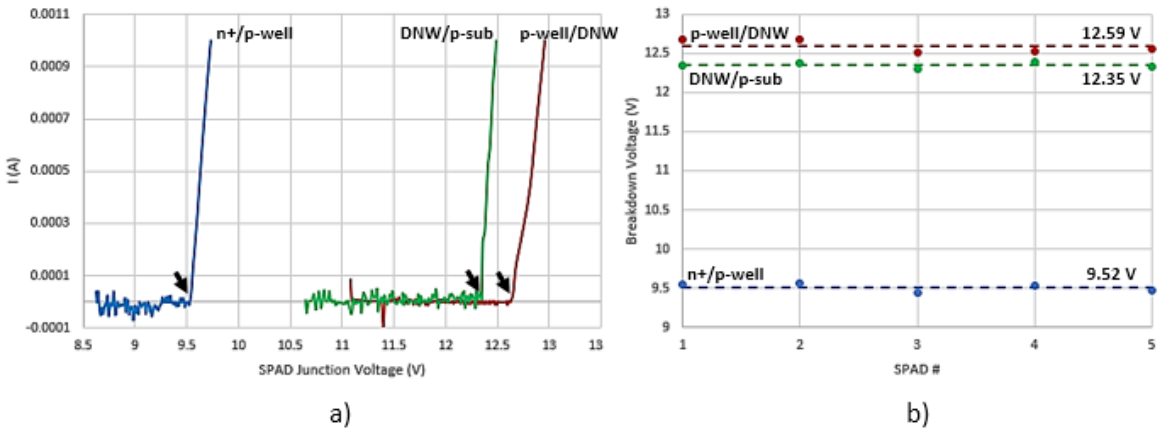


Figure 4-2: Measurement of the junction breakdown voltages: a) The IV curve from which the breakdown voltage was extracted; b) The average of five SPADs

These results are similar to a published result showing a breakdown voltage of 9.1 V for J1 using a standard 65 nm CMOS process [45]. It is also consistent with documentation provided by TSMC for this CMOS process. As previously mentioned, SPADs designed in DSM CMOS technologies have shallower junctions and relatively low junction breakdown voltages on the order of tens of Volts due to their higher doping concentrations and consequently narrower depletion widths. The results obtained shows how the more highly doped shallow junction has a lower breakdown voltage than the more lightly doped wells and substrate forming the deeper junctions. This, however, also means that junctions like J1 have a higher chance for PEB and higher DCR due to tunnelling effects. This was illustrated in a comparison which found that devices with breakdown voltages in the range of 9.4 to 11.4 V had a large DCR compared to devices with breakdown voltages in the range of 23.1 to 27.5 V which had the best DCR performance and typically fabricated in HV CMOS [41].

To further test the breakdown characteristics of the junctions and its relation to noise, its dependence on temperature was also investigated. It is expected that the breakdown voltage of SPADs increases with temperature due to an increase in the rate of phonon scattering. This makes it more difficult for electrons and holes to reach the required threshold energy for impact ionization to act as an initial trigger for avalanche breakdown [76]. Furthermore, a higher temperature coefficient is an indication that the DCR is

composed mainly of thermal generation or whereas a lower temperature coefficient indicates a higher tunnelling contribution. The temperature coefficient of the breakdown voltage was determined by once again measuring the breakdown voltage of five TJ SPAD devices as above but with the use of a temperature chamber to vary the ambient temperature between -30°C and 30°C as illustrated in Figure 4-3. Figure 4-4 contains the results of the temperature dependence by extracting the slopes of the graphs. The shallow n+/p-well junction showed an average temperature coefficient of 5.8 mV/°C, similar to the 5 mV/°C found in [45]. The deepest DNW/p-sub junction showed the most consistent behaviour between SPADs with an average of 8.5 mV/°C, while the middle p-well/DNW junction showed the most variation in measurement with values between 8.1 and 9.4 mV/°C, averaging at 8.9 mV/°C. The greater variation in the p-well/DNW junction means that it may be more likely influenced by fabrication defects. All the subsequent measurements that depend on temperature variations take into account the variation in breakdown voltage. The excess voltage is adjusted accordingly to maintain consistency and accuracy in the reported results.

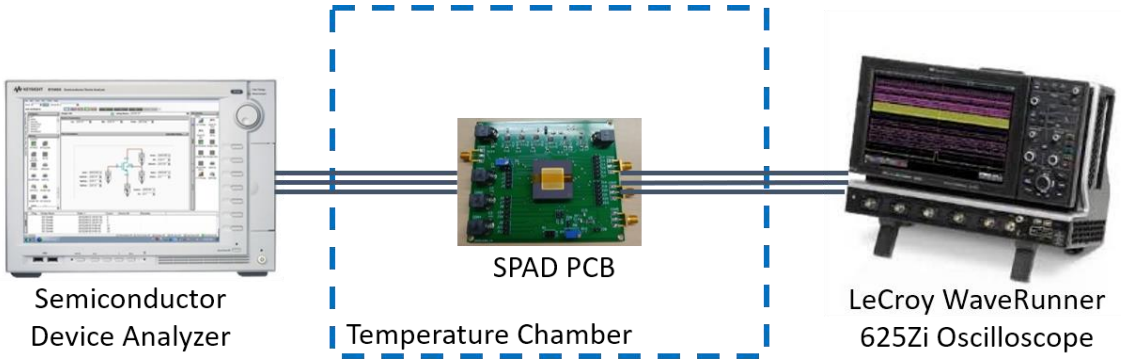


Figure 4-3: The setup for testing breakdown temperature dependence with the device analyzer and for noise measurements with the oscilloscope.

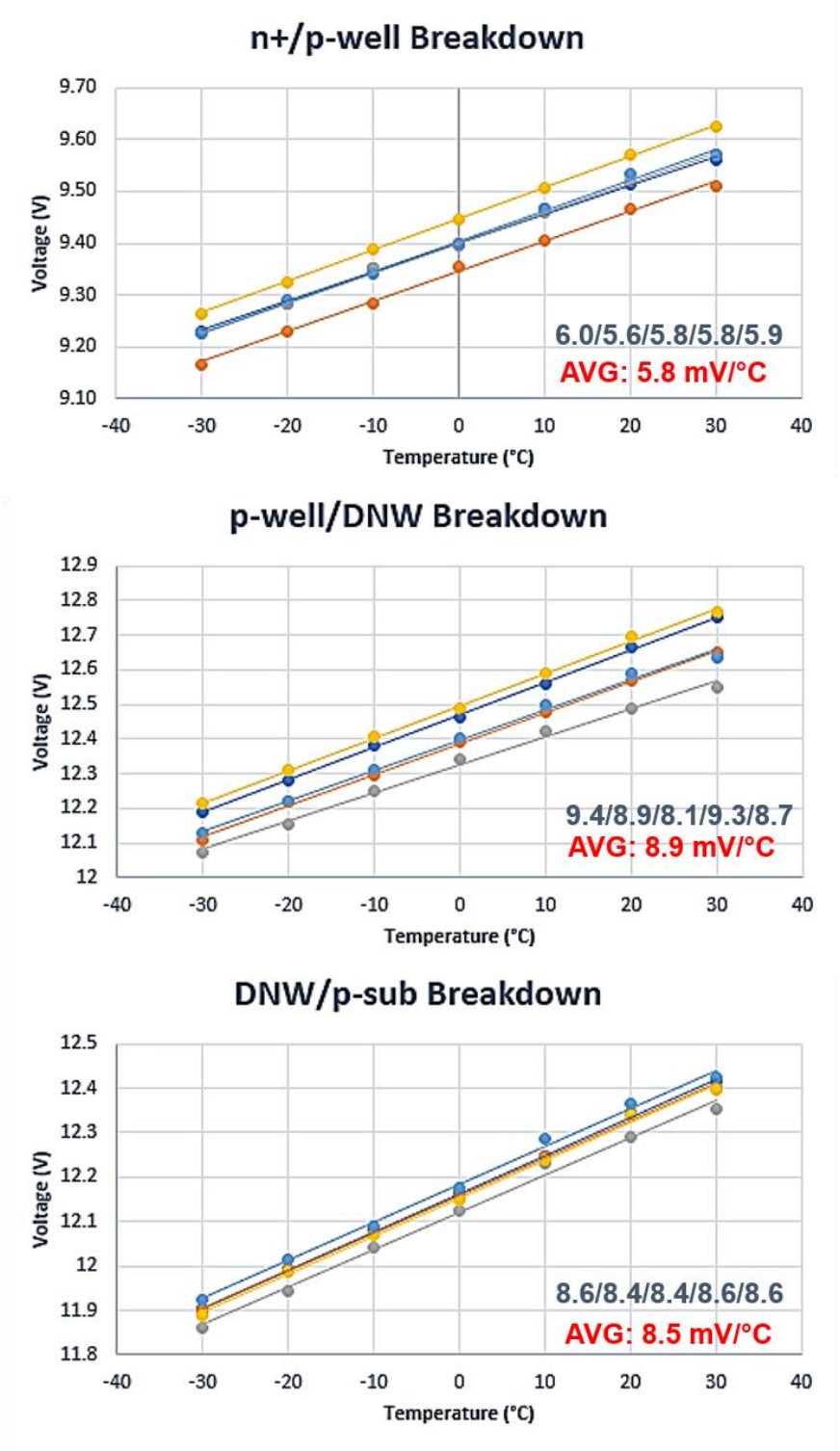


Figure 4-4: Test results of the breakdown voltage dependence on temperature for each junction.

A LeCroy WaveRunner 625Zi highspeed sampling oscilloscope was used to probe the test points of the TJ SPADs and record the waveforms of each junction during breakdown in a dark environment. The result obtained is shown in Figure 4-5 and is representative of what was observed among all the several SPADs tested. Here, the top two junctions appeared to continuously break down due to noise and were unable to properly reset to the set excess voltage before avalanching. This prevents these two SPAD junctions from properly responding to light and renders them untestable. It was predicted that without buffers, these junctions suffered from the high capacitive load of the oscilloscope probe (9.5 pF). Unfortunately, the AC-coupled implementation (Figure 3-4) also had poor performance since low excess voltages could not be used for a sufficient voltage drop to trigger output pulses, and high excess voltages resulted in too much noise.

Fortunately, the third junction (DNW/p-sub) exhibited much less noise and displayed a cleaner waveform that responded to sources of light. Furthermore, the TG J1 SPAD was found to function and respond to light albeit a high DCR thanks to the low load capacitance of its front-end circuitry. As a result, the following tests to analyze the DCR, PDE, and timing jitter of the SPADs were performed on the unbuffered third junction J3 of the TJ SPAD and the TG SPAD based on the top shallow junction J1. This prevented the proper analysis of the junctions working simultaneously so the benefits of a multi-junction SPAD could not be thoroughly tested.

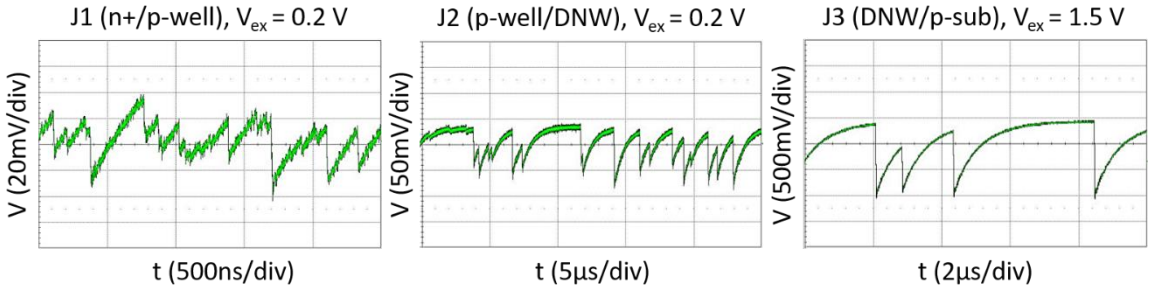


Figure 4-5: The top junctions (J1 & J2) showed high levels of noise and afterpulsing with J1 constantly under breakdown and unable to reach the set excess voltage. J3 showed the expected behaviour.

4.2.2. DCR and AP

Analysis of the DCR of the TG J1 SPAD requires that the time of the gate window be considered. Therefore, a dark count probability per gate window (DCP_{GW}) is typically evaluated [73]. The effective DCR is calculated from the DCP_{GW} based on the duration of the gate window (3 ns in this case) that the SPAD is armed to avalanche either due to photons or noise. This equation is given as

$$DCR = \frac{DCP_{GW}}{T_{ON}} = \frac{DCP_{GW}}{3\ ns}. \tag{4-1}$$

Due to the high susceptibility to noise of these SPADs, the relationship between DCR and excess voltage was tested to determine an excess voltage range to maintain an acceptable SNR and proper operation. The DCR of J3 and the TG J1 SPAD junctions is given in Figure 4-6 as a function of V_{ex} at room temperature. An exponential fit to the data matches the expectation that DCR is exponentially dependent on V_{ex} for a limited range until saturation effects take over. The plots indicate that this saturation occurs above an excess voltage of 0.5 V for both J1 and J3 and is a result of more avalanches occurring during recharge or moments of hold-off/dead time which are not counted. This is further influenced by the selection of the threshold used in the case of the unbuffered J3 SPAD which can neglect counts that occur before proper reset of the excess voltage. For the TG J1 SPAD, deviation from the exponential fit is also seen below 0.3 V. These missing counts can be explained by avalanches that occur late in the time window which may not properly drive the threshold requirement of the PMOS responsible for the readout of counts. Thus, the remaining measurements on the SPADs were done at excess voltages around or below 0.5 V, but above 0.3 V for the TG SPADs.

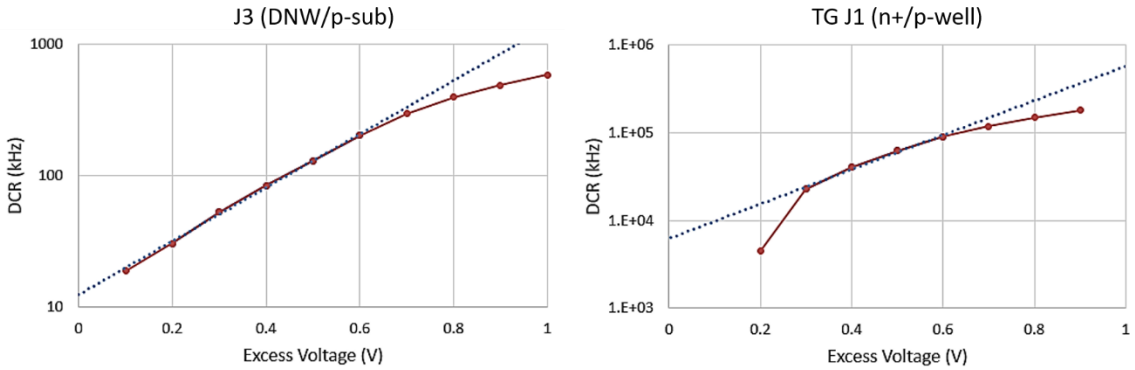


Figure 4-6: Dark count rate (DCR) of J3 (left) and dark count probability per gate window (DCP_{GW}) of the time-gated (TG) J1 (right) both as a function of excess voltage and at room temperature.

The primary dark counts due to thermal noise and tunneling are known to follow Poisson statistics [41] where individual counts are mutually independent of each other, the probability of a count is proportional to the duration of the time interval, and more than one count occurring within an interval is negligible. As a result, first-order inter-arrival times (IATs) are independently identically distributed random variables. This should provide the temporal information to distinguish primary dark counts due to thermal generation from that of AP. This is done by determining the deviation of the IAT distribution from the ideal Poisson distribution, which follows an exponential decay probability density function in the case of zero AP given by

$$f(\tau) = \lambda \exp(-\lambda\tau) \tag{4-2}$$

where λ is a constant rate representing the mean primary dark count rate (DCR_{PR}) [41]. The unbuffered J3 junction of the TJ SPAD was tested for its noise characteristics by measuring the IAT between avalanche pulses in the observed waveform of dark counts, as shown in Figure 4-7. The LeCroy WaveRunner 625Zi highspeed sampling oscilloscope was used to acquire approximately 50k to 500k IATs at a range of temperature and excess bias voltage points using the setup from Figure 4-3. The large number of counts is required to collect enough points for the tail of the distribution since the exponential nature of the IAT distribution means short delays are much more likely than long delays between pulses. A 50% threshold was used for the majority of the data collected and a few sets of data were taken at the 100 mV (from VDD) threshold for AP comparison. The IAT information from

an example waveform of J3, the choice of thresholds, and the resulting histogram is given in Figure 4-7. With enough counts collected, the first non-zero value in the IAT histogram gives the dead time T_{DT} of the SPAD signifying the smallest amount of time where it is completely insensitive.

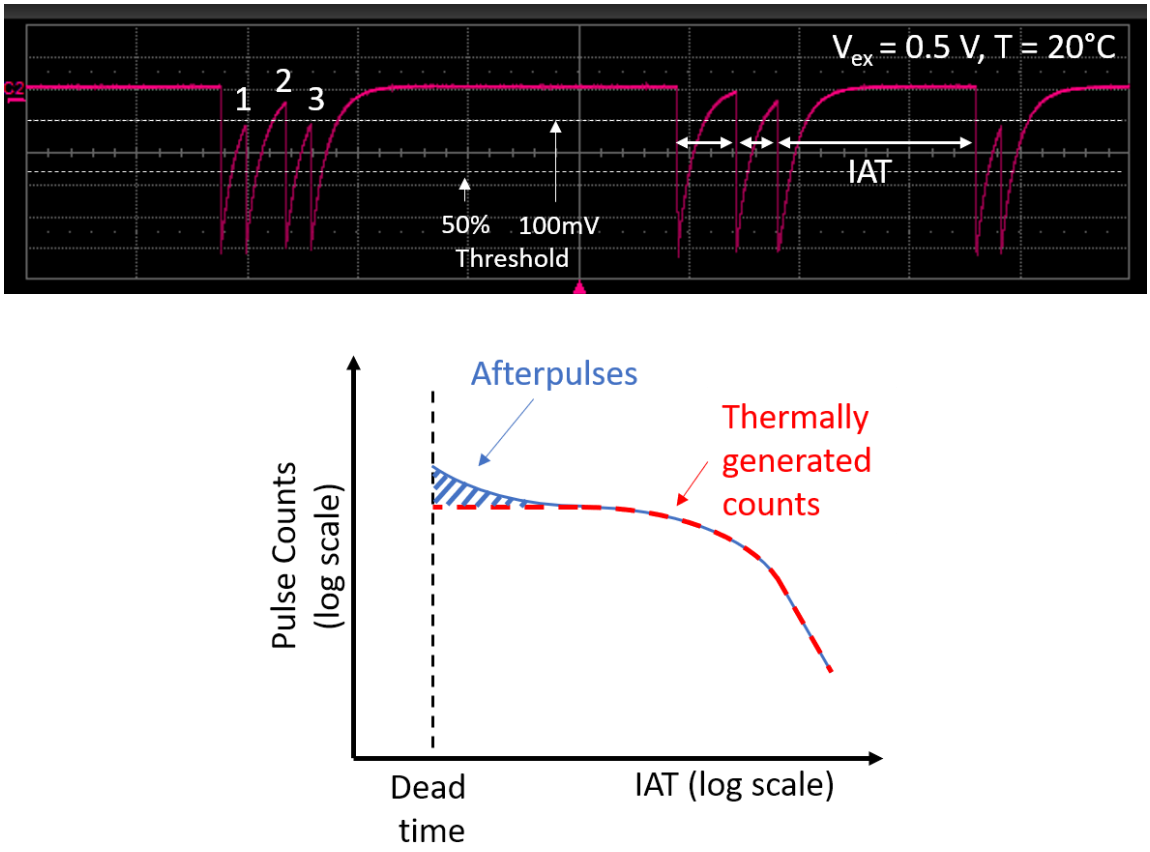


Figure 4-7: Waveform (above) from J3 illustrating the inter-arrival time (IAT) between pulses and the resulting histogram (below) on a log-log scale. The effect that the selected threshold has on the pulses counted and their IATs is illustrated in the waveform where a 100 mV threshold from VDD of the SPAD will skip pulses 1 and 3, extending the perceived dead-time and IAT, while pulse 2 would be counted.

Figure 4-8 shows IAT distributions of a J3 SPAD with exponential fits at two different temperatures (-20 and 20°C) and two excess voltages (0.3 and 0.5 V) using a 50% threshold. At the higher excess voltage and temperature, a saturation of early counts starts to become noticeable which corresponds with the saturation discussion of Figure 4-6. A high DCR due to thermal generation at higher temperatures can mask AP effects, but as the

temperature decreases, the AP effects become noticeable since thermal dark counts decrease and trap lifetimes become longer. An increased excess voltage at lower temperatures also contribute to greater avalanching probability of these trapped charges.

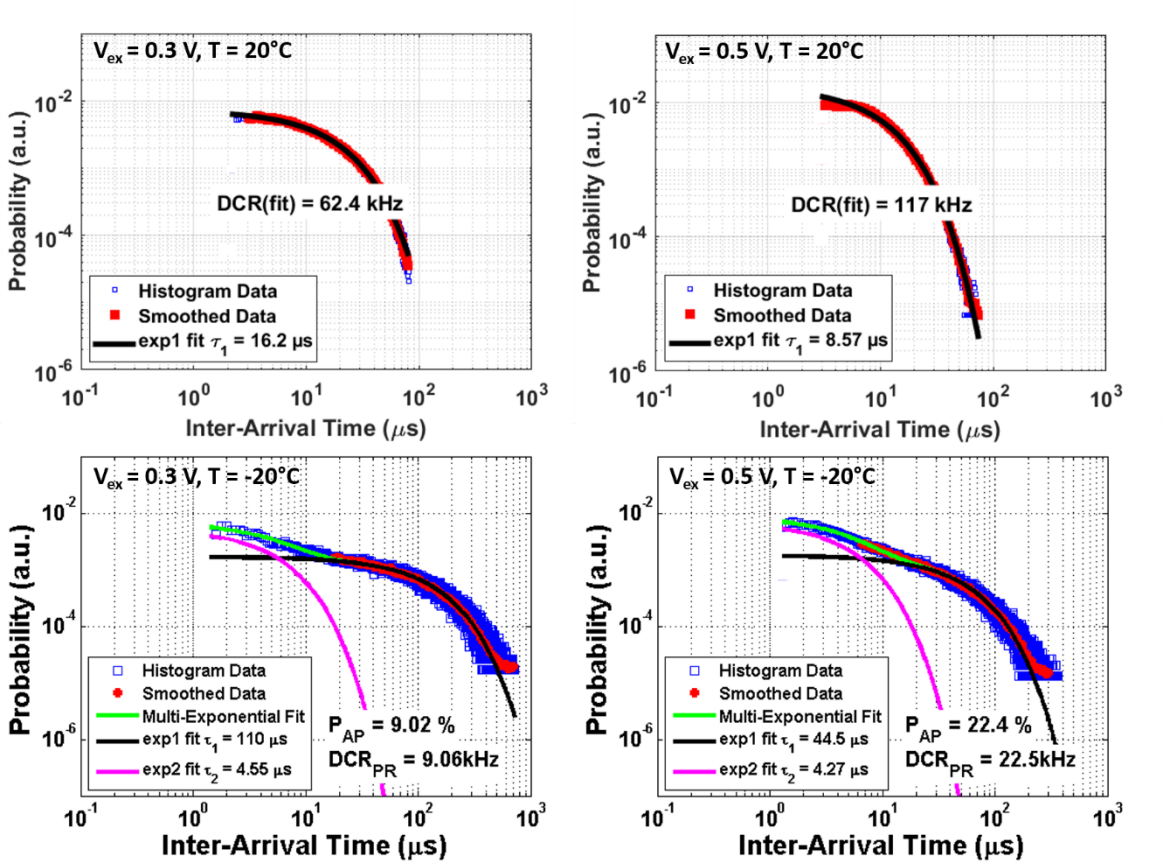


Figure 4-8: IAT distributions of dark counts at V_{ex} = 0.3 and 0.5 V, and T = 20 and -20°C. When operating the unbuffered J3 SPAD at higher temperatures, the effects of AP is masked as opposed to lower temperatures below 0°C where it becomes more apparent.

In Figure 4-9, the difference in AP and dead time that occurs with different threshold selection can be seen. The primary DCR due to thermal generation and tunnelling sources remains fairly consistent, however, the perceived dead-time of the SPAD and AP probability is affected. As observed in Figure 4-9, there is a shift in the IAT distribution between the two thresholds used. The plot with the tighter 100 mV threshold is shifted more to the right compared to the 150 mV threshold due to more points being counted as retriggering events that extend the perceived IATs, as was illustrated with points 1 and 3 in Figure 4-7. This also lowers the calculated AP probability since pulses that would be

considered AP are missed and reclassified as retriggering events [77]. More research can be done to determine proper thresholds to be used in certain situations and a more detailed theoretical description of the process can lead to simulations of IAT statistics for more accurate and faster time-to-market designs SPADs. Overall, the J3 SPAD can be said to have a dead time around 1-3 µs and negligible AP around room temperature.

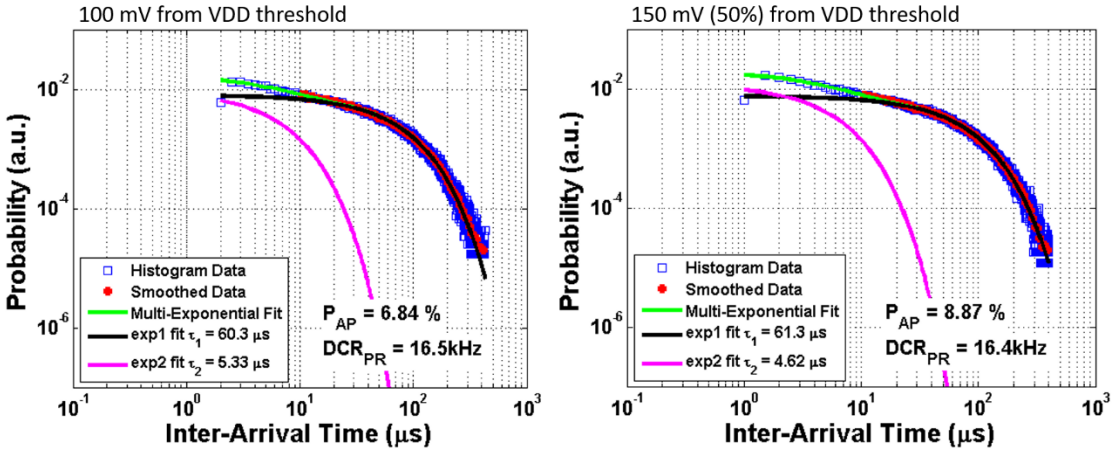


Figure 4-9: The IAT distributions and exponential fits of the unbuffered J3 SPAD at V_{ex} = 0.3 V, T = -10°C plotted on a log-log scale with different thresholds. The second exponential fit is a result of after-pulses which can be influenced by the threshold selection. The afterpulsing probabiltiy is lower with the selection of a threshold that is closer to VDD (left, 100 mV below VDD) compared to lower (right, 150 mV below VDD).

These dark count measurements are important in evaluating the quality of the fabrication process, the result of the layout on SPAD quality, and the corresponding sources of dark counts. The DCR can be expressed as a function of temperature according to the Arrhenius relationship [76]

$$DCR \propto T^2 \exp\left(-E_a/kT\right) \tag{4-3}$$

where k is Boltzmann's constant, T is the absolute temperature, and E_A is the activation energy. The magnitude of E_A gives insight into the primary defect type leading to the measured dark counts and how it changes according to temperature. When considering only thermal generation, E_A should roughly correspond to half the bandgap energy ($E_A=E_g/2$) since a mid-gap activation energy indicates mid-gap defects which are the most efficient GR centers in accordance with SRH GR theory. If E_A is equal to the band gap energy (1.1 eV for silicon), this can indicate band-to-band GR and diffusion of minority carriers such

as from contact regions [78]. An activation energy below the mid-gap means an increased density of non-mid-gap GR centers and higher contributions from tunnelling or field-assisted effects. This is especially true and is expected with the SPADs fabricated in the 65 nm CMOS process.

The DCR of the J3 junction was recorded by the LeCroy oscilloscope at a temperature range of -30 to 30°C and a V_{ex} from 0.2 to 0.5 V. The corresponding Arrhenius plots are presented in Figure 4-10 where a change in activation energy is evident when sweeping temperatures by extracting the slopes of the curves. The E_A above 20°C sees a slight decrease as the excess voltage increased from 0.4 to 0.5 V as seen in the figure. As the GR mechanism is expected to show little dependence on excess voltage, this meant tunneling effects are still quite dominant at higher temperatures given a high enough excess voltage. Below -10°C, the E_A saw an overall decrease which reflected the growing dominance from tunnelling effects. This is further demonstrated with the sharper decrease in E_A from an excess voltage of 0.4 to 0.5 V as tunnelling effects are known to vary with excess voltage [41].

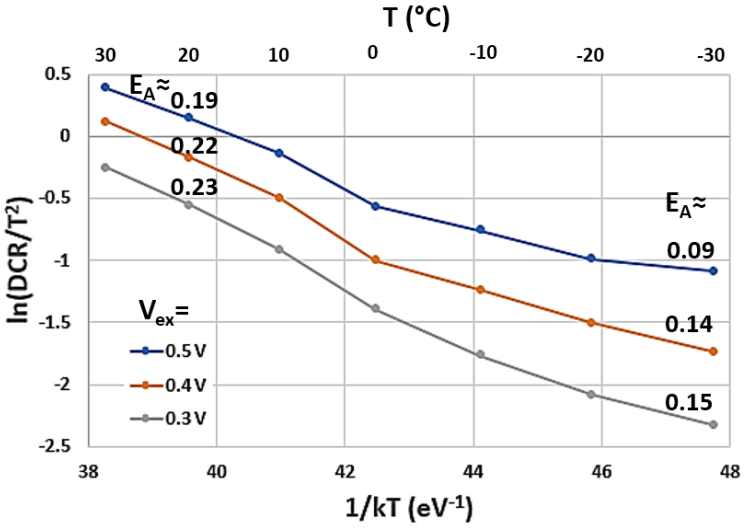


Figure 4-10: Arrhenius plot of the dark count rate (DCR) of J3 (DNW/p-sub) showing the activation energy (E_A) dependence on temperature and various excess voltages (V_{ex}).

The tunneling dominance is further displayed when running the same measurements on the TG J1 SPAD where the more highly doped and thinner depletion region lends itself

to higher tunnelling probabilities. The results were plotted in Figure 4-11, where a change in slope is not evident with a change in temperature. The y-axis has been shifted due to using the DCP_{GW} given by

$$\ln(DCP_{GW}/T^2) \propto -E_a/kT + \ln(3\,ns). \quad (4\text{-}4)$$

A straight line in an Arrhenius plot indicates dominance of one DCR component over the other [31], [38] and with very low activation energies hovering above 0.10, this means trap-assisted tunneling dominates at all tested temperatures. This again shows variation with excess voltage as expected of tunnelling mechanisms.

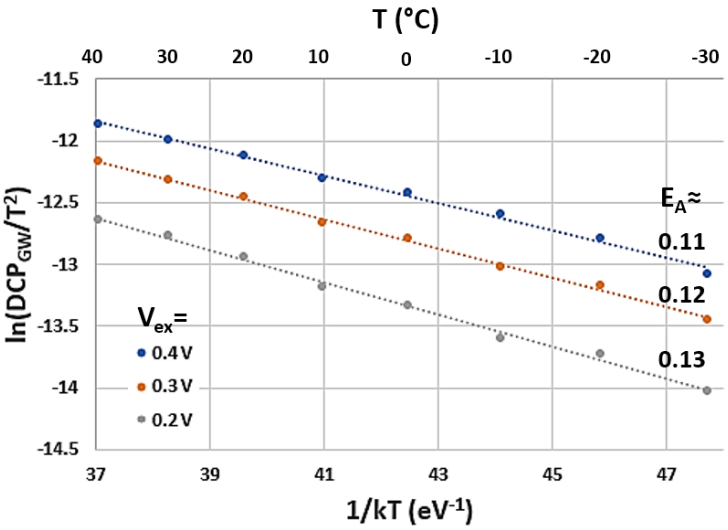


Figure 4-11: Arrhenius plot of the dark count probability per gate window (DCP_{GW}) of time-gated (TG) J1 (n+/p-well) showing the activation energy (E_A) dependence on temperature and various excess voltages (V_{ex}).

Since the J1 shallow junction is time-gated, it is important to determine optimal gating frequencies to find a suitable hold-off time and limit the effects of AP. The gating frequency f_g of the TG SPAD was varied to observe the effects of the hold-off time on AP, which is expected to only occur as the gating frequency passes a certain threshold. Otherwise, the measured DCR as a result of time-invariant thermal generation or tunneling is expected to remain constant regardless of the set hold-off time. The results shown in Figure 4-12 show a relatively constant DCR above a gating period of ~50 ns (below a 20 MHz gating frequency). For an excess voltage of 0.3 V, the AP probability remained below 1% up to a

gating period of 25 ns (40 MHz gating frequency). With a gate window of ~3 ns, this means the hold-off time where an AP probability <1% can be achieved is ~22 ns. The dependence of AP on excess voltage is also demonstrated and AP is seen to increase significantly with even smaller gating periods.

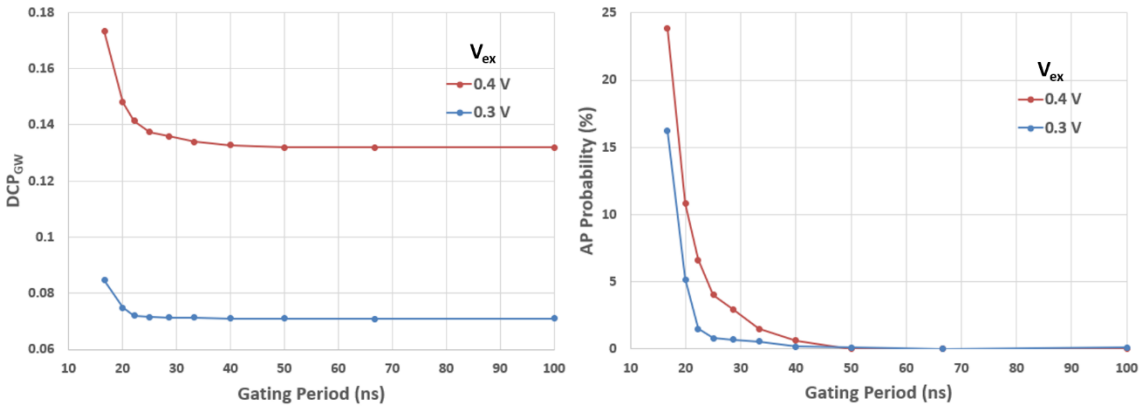


Figure 4-12: The dark count probability per gate window (DCP_{GW}) of the time-gated (TG) J1 (n+/p-well) SPAD (left) and the corresponding afterpulsing (AP) probability (right) as a function of gating period.

4.2.3. PDE

The PDE is calculated as the ratio of the number of photons per second counted by the SPAD (Φ_{SPAD}) to the incident number of photons per second (Φ_{IN}). The full measurement setup used to obtain the total incident photon counts and compare it to the counts from the SPAD is illustrated in Figure 4-13. The total incident photons required obtaining the power of the light at the location of the SPAD using a silicon photodetector (SiPD) and optical power meter calibrated to the wavelength selected through bandpass filters. The light is supplied by a xenon lamp and neutral density (ND) filters help control the intensity of light so as not to saturate the SPAD. With this setup, the incident number of photons on the SPAD can be determined as

$$\Phi_{IN} = P_{SiPD}\left(\frac{\lambda}{hc}\right)\left(\frac{A_{SPAD}}{A_{SiPD}}\right) \tag{4-5}$$

where P_{SiPD} is the power measured by the calibrated SiPD, λ is the wavelength of incident light, h is Planck's constant, c the speed of light, A_{SPAD} the active area of the SPAD (100 μm^2), and A_{SiPD} the active area of the SiPD (1 cm^2). The power reading of the SiPD is

converted into photon flux through the conversion factor of λ/hc depending on the wavelength and is corrected for the difference in active area between the SiPD and the SPAD. All values were converted to SI units for the calculation.

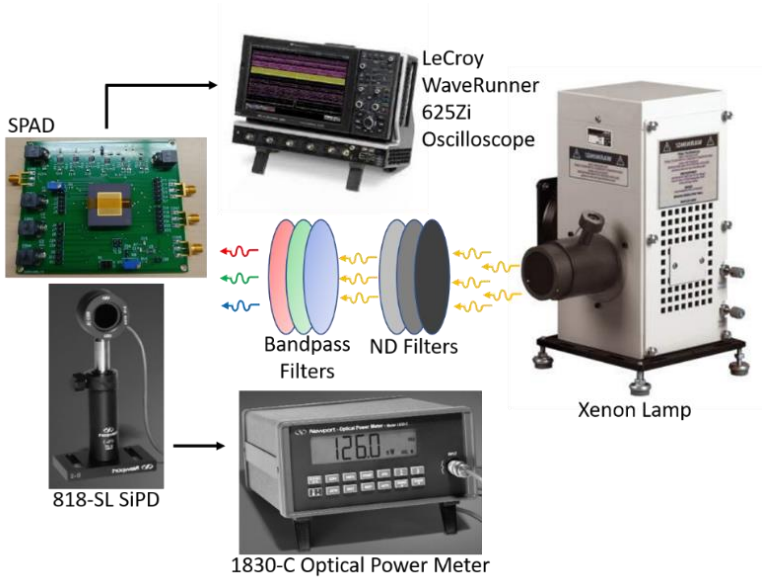


Figure 4-13: Experimental setup for photon detection efficiency (PDE) measurement.

To determine Φ_{SPAD}, the DCR is subtracted from the number of counts reported by the oscilloscope for the SPAD for the respective excess voltages tested. In order to determine a suitable intensity of light, it was important to make sure the counts are sufficiently higher than DCR, but below saturation so as to not miss counts. If the counts are similar to the DCR noise floor, inaccuracies can inflate the observed PDE. Figure 4-14 reflects this behaviour and the PDE is seen to stabilize when the number of SPAD counts approaches ~10x the DCR noise floor. AP effects did not need to be corrected for since the previous DCR results of J3 showed negligible AP at room temperature and excess voltage at or below 0.3 V, while the TG J1 was gated at a frequency of 10 MHz to avoid AP as well.

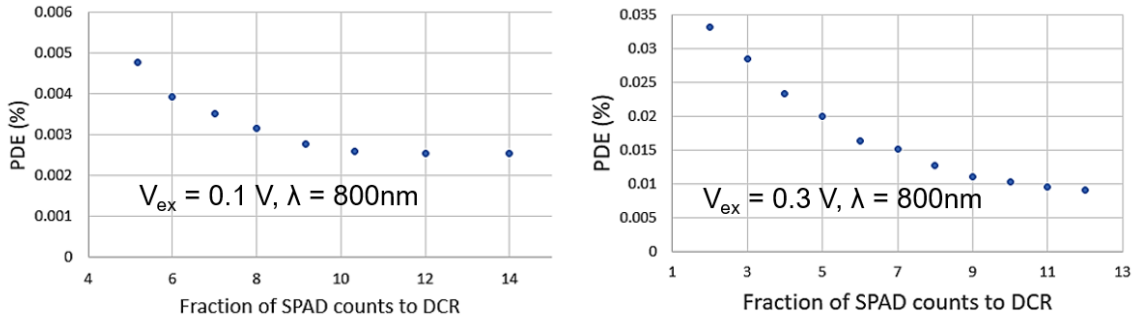


Figure 4-14: The photon detection efficiency (PDE) as a function of the signal-to-noise (SNR) ratio which is the fraction of measured counts to the dark count rate (DCR) noise floor. Measurements taken at a wavelength of 800 nm and excess voltages of 0.1 and 0.3 V.

The PDE of the J3 and TG J1 SPAD junctions were measured at a range of wavelengths (λ) and the results are reported in Figure 4-15. Since TG J1 only responds during the specified gate window, the measured photon counts will effectively be reduced by a factor of the duty cycle (T_{ON}/T_G where T_G is the gating period and T_{ON} is the gate window) when using a continuous wave (CW) incident light. This correction is applied in the final result displayed in Figure 4-15. The J3 and TG J1 PDE peaks were ~0.15% and ~3.5%, respectively, both at 440 nm.

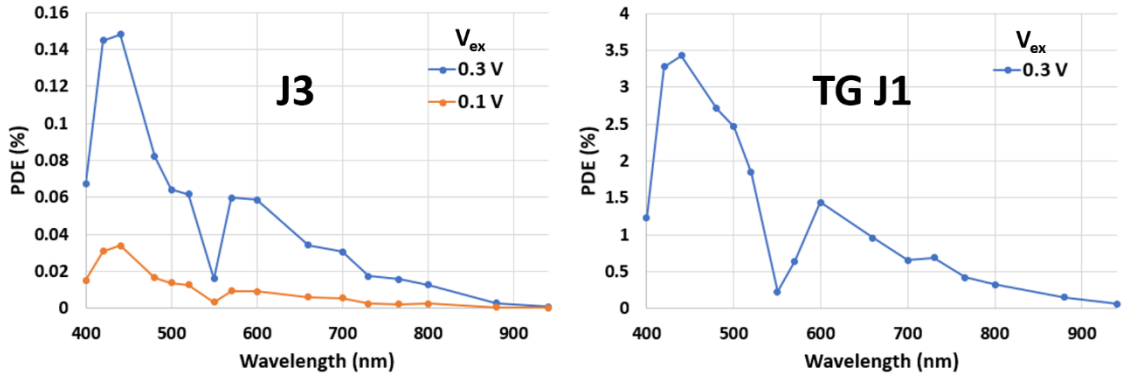


Figure 4-15: PDE measurement results for J3 (DNW/p-sub) on the left, and the time-gated (TG) J1 (n+/p-well) on the right.

The PDE of J3 was expected to be higher than that of J1 given the fact that it is a deeper junction with lower doping concentrations which leads to a wider depletion region. Furthermore, since it is located deeper, it should respond more favourably to longer wavelengths when compared to the shallower J1. Seeing as how both junctions responded similarly as a function of wavelength (peaking around 440 nm and a dip around 550 nm)

suggests that the maxima and minima in the PDE observed may be dominated by the stack of dielectric layers/silicon forming a Fabry-Perot resonator in a standard CMOS process [79]. This is usually corrected for in the more optimized CIS process where the effect of the junction depth on PDE is better observed [53]. The maxima as well as the dip around 550 nm in the PDE- λ characteristics is similar to that of another standard 65 nm CMOS SPAD reported in literature [45].

4.2.4. Timing Jitter

Measurement of the timing jitter of the TG J1 SPAD was done using the experimental setup in Figure 4-16. The gating frequency from the function generator is fed to both the SPAD PCB and the laser driver through a passive delay unit. This is used to synchronize the laser pulse (of 685 nm wavelength) with the gate window of the SPAD.

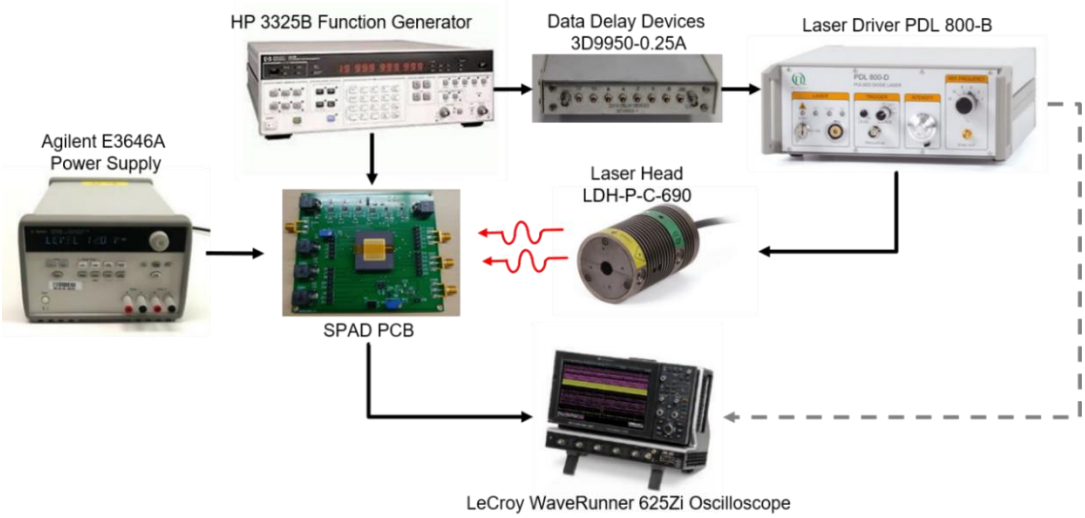
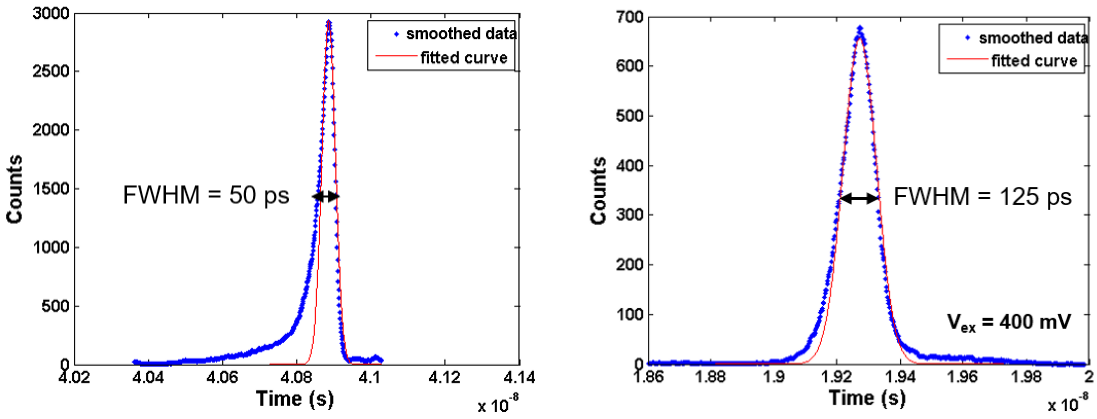


Figure 4-16: Timing Jitter experimental setup for the TG SPAD.

The instrument response function (IRF) of the setup is the result of the histogram created by the oscilloscope when it compares the arrival time of the observed SPAD pulse to the reference pulse from the laser driver. The uncertainty in this time difference typically manifests in a Gaussian distribution where the FWHM is reported as the IRF and is equivalent to the timing jitter. Since this IRF is of the total system, it is a convolution of all component IRFs and is estimated using the quadratic sum

$$IRF_{system} \approx \sqrt{\sum_{component,i} IRF^2_{component,i}} \quad (4\text{-}6)$$

which can be used to extract and estimate component IRFs, one of which is the SPAD. In taking data, the laser intensity was adjusted to give a count rate at least 10x above the DCR floor and at most just below saturation so as to avoid the influence of arrival of random dark counts. An example of the resulting histogram in the relative arrival time of the detected counts with respect to the laser driver sync out is given in Figure 4-17 for both a commercial (PD-050-CTD) SPAD and the TG J1 SPAD. The data was smoothed and fitted with a Gaussian curve to extract FWHM values.



(a) (b)

Figure 4-17: The IRF of the system using: (a) a commercial (PD-050-CTD) SPAD; (b) using the TG J1 SPAD when operating at 400 mV and a gating frequency of 20 MHz.

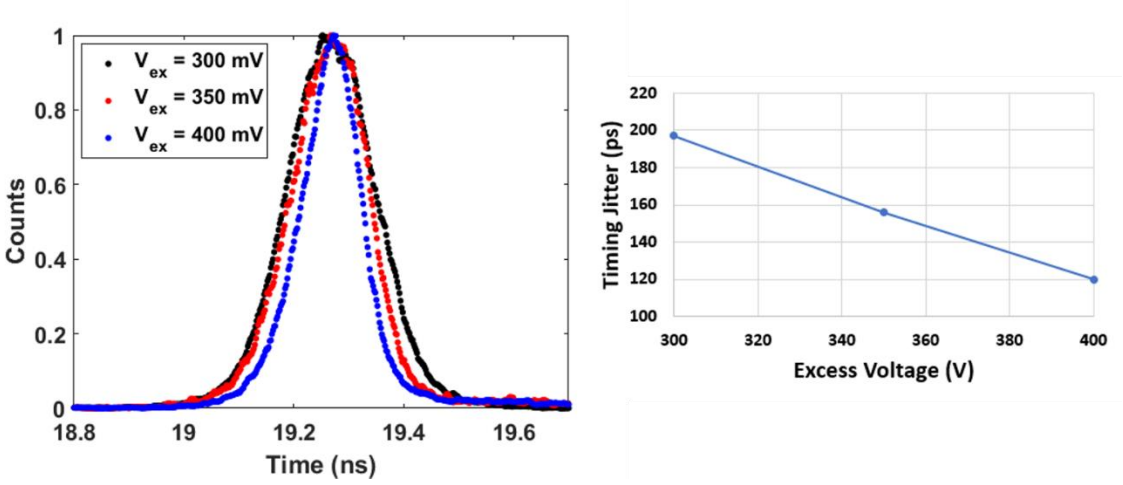
Using the known jitter of 30 ps for the laser head (IRF_{laser}) and 27 ps FWHM for the commercial SPAD (IRF_{PD}) through a test report, the jitter contributions of all other sources in the measurement setup (IRF_{other}) can be estimated as:

$$IRF_{other} = \sqrt{IRF^2{}_{system} - IRF^2{}_{laser} - IRF^2{}_{PD}} = \sqrt{50^2 - 30^2 - 27^2} \approx 30 \text{ ps.}$$

This jitter includes sources such as the sync out and external trigger in of the laser driver. From this, the jitter of the TG J1 SPAD can be calculated in a similar using the information from Figure 4-17 (b) to give:

$$IRF_{SPAD} = \sqrt{IRF^2{}_{system} - IRF^2{}_{laser} - IRF^2{}_{other}} = \sqrt{125^2 - 30^2 - 30^2} \approx 118 \text{ ps.}$$

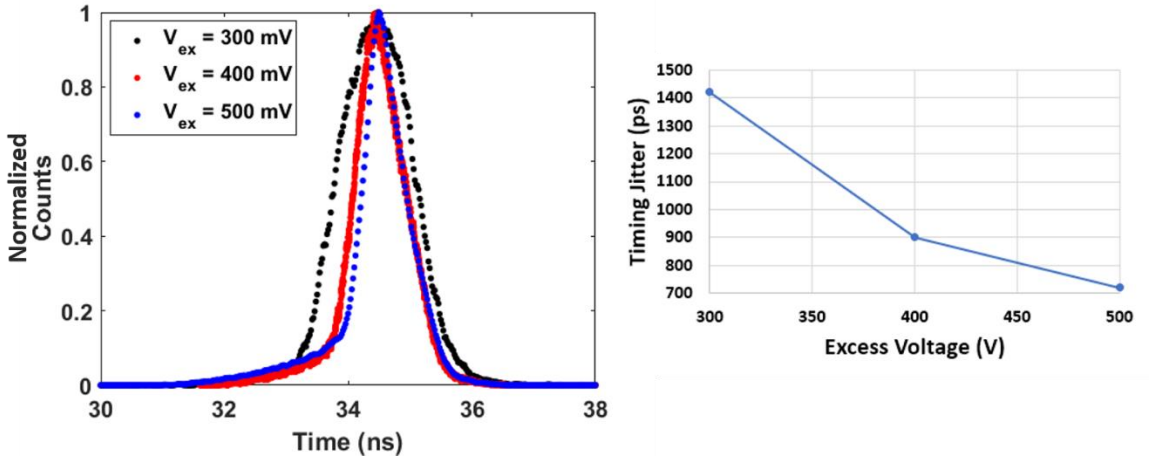
The measurements were also taken at an excess voltage of 300 and 350 mV and displayed relative to each other in Figure 4-18 (a). The timing jitter dependence on excess voltage is given in Figure 4-18 (b) with values calculated similarly as above, resulting in timing jitters of 156 ps and 197 ps for excess voltages of 350 mV and 300 mV, respectively.



(a) (b)

Figure 4-18: Multiple timing jitter measurements of the TG J1 SPAD: a) the histograms at 300, 350, and 400 mV excess voltages after being normalized and then smoothed; b) the relation of timing jitter decreasing as excess voltage increases.

As demonstrated in Figure 4-18, an increase in the SPAD's excess voltage resulted in sharper Gaussian peaks with smaller FWHM, or timing jitter, values. The electric field enhancement due to the increased excess voltage means there is less statistical fluctuations in the avalanche build-up time and thus a sharper response. Higher excess voltages are expected to decrease the jitter even more, but the higher DCR contributions would begin to negatively impact the SNR. A similar test was run on the unbuffered J3 SPAD, with the prediction that it will have a higher timing jitter compared to the thinner top junction. Deeper junctions with higher breakdown voltages indicates lower doping levels, wider depletion regions, and thus reduced timing performance. Furthermore, the unbuffered design introduced further sources of uncertainty with using the oscilloscope to set a threshold and detect avalanches through a probe with relatively high capacitive load. Figure 4-19 shows the result of this, with higher excess voltages helping to reduce the jitter but ultimately staying around a value of 1 ns.



(b) (b)

Figure 4-19: Multiple timing jitter measurements of the J3 SPAD showing reduced performance due to being a deep junction and unbuffered: a) the histograms at 300, 400, and 500 mV excess voltages after being normalized and then smoothed; b) the relation of timing jitter decreasing as excess voltage increases.

4.3. Conclusions

This has been a preliminary evaluation of a SPAD fabricated in a standard 65 nm CMOS process as well as its performance using an established gating technique with proven success in the larger 130 nm CMOS technology node. The breakdown characteristics of a TJ configuration (J1: n+/p-well, J2: p-well/DNW, J3: DNW/p-sub) and the noise and performance of J1 and J3 in this technology is the first in literature which provides valuable information for SPAD designers in this technology node.

In detail, the breakdown voltages and temperature dependence of all three junctions of the TJ SPAD were measured. The high noise associated with junctions J1 and J2, an unbuffered design, and relatively high load capacitance of the oscilloscope probes used, prevented proper testing of the expected PDE enhancements when all junctions work simultaneously. Tunneling mechanisms dominates as a noise source in this small DSM technology, especially with the thinner and more highly doped J1. Nonetheless, the bottom junction J3 (DNW/p-sub) of the TJ SPAD as well as the TG SPAD pixel based on J1 functioned appropriately and were both characterized according to the performance parameters introduced such as the break-down voltage, DCR, AP probability, PDE, and

timing jitter. The respective performance of each is summarized in Table 4-2 and compared against other SPADs fabricated in standard CMOS processes, with an emphasis on those also fabricated in a 65 nm process.

Table 4-1: Comparison of our work to other SPADs fabricated in a standard CMOS process.

Year, Ref.	Standard Technology, ***Junction***	V_{BD}/V_{ex} (V)	Peak PDE (%)	DCR (cps/μm^2)	Timing Jitter (ps FWHM)	AP (%)	FF (%)
2014, [80], [81]	130 nm, p+/n-well	12.13/1.5	0.3 @ λ = 425 nm	28.0	<198 @ λ = 408 nm	-	<1
2017, [38]	150 nm, p+/n-well	18/3	10 @ λ = 450 nm	0.4	42 @ λ = 831 nm	<1 @ 0.15 μs hold-off	39.9
2018, [82]	180 nm HV, p+/n-well	16.8/4	0.55 @ λ = 480 nm	2	260 @ λ = 640 nm	-	<1
2019, [36]	180 nm, p-well/p-epi/DNW	25.46/6.5	5.25 @ λ = 520 nm	0.26	110 @ λ = 637 nm	-	10.5
2013, [45]	65 nm, n+/p-well	9/0.4	5.5† @ λ = 420 nm	15.6 k	235 @ λ = 637 nm	<1 @ 5 μs dead time	-
2018, [37]	65 nm (TSMC), p+/n-well	9.9/1.5	8 @ λ = 470 nm	2.8 k	7.8 @ λ = 410 nm	<10 @ 0.1 μs hold-off	-
Our work	65 nm (TSMC), n+/p-well	9.5/0.3	3.5 @ λ = 440 nm	230 k	<200 @ λ = 685 nm	<1 @ 0.022 μs hold-off	28.6
	65 nm (TSMC), DNW/p-sub	12.35/0.3	0.15 @ λ = 440 nm	550	<1420 @ λ = 685 nm	<1 @ 2 μs dead time	-

† PDP value reported (PDE = PDP*FF)

While most SPADs were designed with the p+/n-well shallow junction, we required the use of the n+/p-well shallow junction to create a TJ SPAD. The measured J3 SPAD performance suffered due to being unbuffered, increasing the capacitance on the junction when tested with the oscilloscope probe. While this led to an increased dead-time and AP probability, the reduced doping and cleaner fabrication of the deeper junction allowed for a low DCR. Nonetheless, its characterization is valuable as a reference for future designs.

The TG J1 suffered from a high DCR which accounted for ~7% of possible counts when biased at an excess voltage of 0.3 V. A potentially smaller SPAD active area with an improved GR structure is expected to see substantial improvements on DCR that would offset a potential negative impact on PDE. The time-gated design limits the excess voltage to 1 V, but provides reduced AP probability and highspeed operation. The excess voltage could be pushed further if a 3.3 V reset PMOS was used instead of one rated for 1 V that was used to give a faster and sharper turn on response for the SPAD. In comparison with other 65 nm SPADs, our TG J1 SPAD showed comparable or better performance aside from the high DCR. It had a timing jitter better than that of [45] at a lower excess voltage and is estimated to see comparable jitter performance with [37] if the excess voltage could reach as high. In providing a better comparison against time-gated designs, Table 4-2 is provided which includes some relevant and recent time-gated designs and the previous design which this work was based off.

Table 4-2: Comparison of our work to other time-gated SPAD devices

Year, Ref.	2014, [83]	2019, [36]	2019, [39]	2014, [10], [73]	Our Work
Tech.	350 nm HV	180 nm	180 nm CIS	130 nm	65 nm
Activation Energy (<-10°C/> 20°C)	-	-	0/1.1	0.17/0.38	0.12
Pixel Format	1024x8	512x512	1024x500	1	1
Active Area/FF (μm^2)/(%)	255/44.3	28/10.5	6/7	100/9.8	100/28.6
V_{BD}/V_{ex} (V)	19.6/3	25.46/6.5	23/3.3	11.3/1.4	9.5/0.3
Peak PDE (%)	9.6 @ 465 nm	5.25 @ 510 nm	0.74 @ 520 nm	3 @ 510 nm	3.5 @ 440 nm
DCR (cps/μm^2)	22.4	0.26	0.065	30	230 k
Timing Jitter (ps)	-	110 @ 637 nm	-	60 @ 532 nm	118 @ 685 nm
AP Probability (%)	-	-	-	<1 @ 16 ns hold-off	<1 @ 22 ns hold-off
Gating Frequency or Image Framerate	<950 Hz	<98k fps	24k fps	<100 MHz	<100 MHz
Gating Window (ns)	0.7	≥5.75	3.8	3.5	3

It can be seen that most designs are shifting to full array implementations with imaging capabilities and high framerates similar to a standard CMOS image sensor with built-in pixel memory and on-chip gating controls [36], [39]. This limits their speed of operation in their readout architecture and are described with a framerate instead of a gating frequency. In comparison with the previous TG design in a 130 nm process, we managed to maintain comparable AP, highspeed operation, and PDE all with a higher FF and lower excess voltage. If similar excess voltages could be A higher timing performance is also expected using similar excess voltages. Again, while noise appears to be the main limiting factor in our design, the remaining results show good and comparable performance with other SPAD pixels using just a relatively low excess voltage in comparison and thus shows good promise for future improvements.

To overcome the high noise contributions due to the STI GR, a novel GR such as in [57] could be used, or switching to a p+/n-well junction with a p-well GR, but this would prevent a TJ SPAD design. If the characterization of the junctions is known, then a more accurate DJTG design can be pursued as well to allow fabrication and testing. Along with optimization in the sizing and layout of the time-gating circuitry and avoidance of capacitors to perform the pulse generation, its performance can be greatly improved through tighter timing operation and higher FF.

Chapter 5
Development of a Low-Cost, User-Customizable, Highspeed Camera

5.1. Introduction

As CMOS SPADs continue to improve and dSiPMs become more accessible, this paves the way for the development of low-cost, miniature, and portable high-performance imaging solutions. As a proof of principle, this can be demonstrated using available CMOS image sensors today which incorporate integrated circuits for low-power, on-chip processing, and highspeed readout that allow for the design of cost-effective, highspeed cameras. As a result, high-speed imaging is now seeing a widespread adoption in the mainstream consumer market, with start-ups offering more affordable alternatives [84], [85] and flagship commercial smartphones such as the Samsung Galaxy S lineup recently boasting high-speed 960fps imaging as a key selling point [86]. Nevertheless, such devices can still be considered expensive with prices that hover in the thousands of dollars and consumer phones are also not particularly designed for industrial and research applications due to their form factor, battery life and lack of configurability. Upon investigating available commercial cameras, there appears to be a missing market of relatively very low-cost (low to mid hundreds of dollars) high-speed cameras that can offer framerates in the hundreds to low thousands at no less than a 256x256 resolution with enough storage for at least a few seconds of footage at no compression.

We provide a working prototype built upon a Zynq SoC (System-on-Chip) containing a CPU (Central Processing Unit) and FPGA (Field-Programmable Gate Array) with a CMOS image sensor that can lead the way towards the development of these low-cost high-speed cameras. Although this technology is well known in the industry, our use and

organization of off-the-shelf components with a custom FPGA implementation provide a framework for a build-your-own style, standalone (no host computer required) camera that is competitive in the market in terms of cost and functionality. Our approach is similar to the approach of some start-ups such as Kron Technologies Inc. [84] and The Slow Motion Camera Company Limited [85] that create high-speed cameras in the low to mid thousands of dollars using custom architectures and efficient optimizations. For instance, Kron Technologies Inc. uses a ten-year-old ARM Cortex A8 CPU for the user interface, and a $35 Lattice FPGA to grab data from its sensor at high speeds where an analogous FPGA in a Phantom camera can cost thousands of dollars [87].

In fact, the performance benefits of FPGAs play a large role in the implementation and optimization of the high-speed designs required to grab and process data from an image sensor. FPGAs run at clock speeds orders of magnitude slower than their embedded processor equivalents, but their high degree of parallelization can create spatially-oriented circuits that allow application speedups in both latency and throughput compared to a software implementation [88]. Since FPGA designs also implement the data and control paths, they can offer fully deterministic performance, avoiding the fetch and decode pipeline that can severely limit software execution on a processor [89], [90]. Furthermore, they have become important in fields such as deep learning due to their efficiency as hardware accelerators, outperforming GPUs in performance per Watt for some important subroutines [91]. Despite their excellent potential performance and versatility, FPGAs have been referred to as a “specialist architecture” [91] and coding them in hardware descriptive languages (HDLs) is “strenuous” [89]. Compared to how software programming has evolved, hardware programming on FPGAs is still very low level in terms of abstraction, more difficult to debug, and uses more complex EDA tools among other things [90], leading to a lack of hardware programming literacy among scientists and engineers.

We show that nowadays, with some FPGA programming experience and standard commodity parts, it is possible to build a simple high-speed camera prototype with a cost similar to those found in the machine vision market – under $1000 and preferably in the low-to-mid hundreds. The machine vision market of industrial image-based inspection

provides a suitable area for comparisons since they also utilize CMOS imagers with high-speed potential, but many such commercial cameras, such as those from FLIR Systems Inc. [92], are not standalone and depend on device link throughput to a host computer, thus limiting framerates. While such design choices are understandable in that industry to allow for very compact designs and cost reductions, this leaves a gap in the market that we hope to fill and bring attention to.

We present a low-cost high-speed camera designed in the price range of machine vision CMOS cameras, but with the functionality and benefits of more expensive standalone cameras. First, the prototype design implemented using a Zynq SoC Microzed development board, an FMC carrier board, and a custom camera sensor printed circuit board (PCB) is described, serving as a framework by which others can develop custom low-cost high-speed cameras. Then, a figure-of-merit (FoM) is also created to analyze and compare the performance of some available standalone and machine vision high-speed cameras with our own to show the value proposition of our design. Next, we explore the camera's customizability in terms of both hardware and software to suit various research conditions. Lastly, we demonstrate the camera's capabilities by recording low frequency spatial vibration and show that it can outperform commercial entry-grade cameras used routinely in biological [93] and biomedical imaging [94] research, as well as for teaching.

5.2. Hardware Design

5.2.1. Component Selection

Being one of the costliest components in a camera, proper selection of an image sensor for this market area is crucial. Although CCD (Charge-Coupled Device) image sensors can offer a better signal-to-noise ratio and dynamic range suitable for still images, CMOS image sensors have managed to significantly catch up, and offer a variety of benefits over CDD imagers as a result of sharing a standard fabrication procedure with integrated circuits. CMOS image sensors offer lower power, in-pixel charge-to-voltage conversion for high-speed readout, on-chip peripheral components such as PLLs (Phase-Locked Loops)

and ADCs (Analog-to-Digital Converters) for complete system integration, and all at a lower cost [64]. For our design, the PYTHON 1300 CMOS image sensor was chosen since it had a great price-to-performance, allowing us to comfortably reach near or above 1000 fps at respectable resolutions while remaining cost effective. It has an array size of 1280x1024 pixels (1.3 MP), and can reportedly reach a very fast 210 fps at full resolution, and up to 2235 fps at 256x256 [95]. The sensor is configurable through a serial peripheral interface (SPI) and outputs 6 low-voltage differential signaling (LVDS) pairs (a 360 MHz clock, 1 sync channel, and 4 data channels). The monochrome version was chosen over the Bayer colour version because this allows the Bayer pattern demosaic algorithm to be skipped and the lack of RGB filters enhances light sensitivity by allowing up to 3x more light to hit each pixel. The improved sensitivity allows the monochrome sensor to be better suited for scientific research, even though the sensor itself is only marketed for machine vision, motion monitoring, security, and 2D barcode scanning [95].

Next, we looked to use off-the-shelf components to build a camera in a standalone form factor with the memory and hardware necessary to maximize the capabilities of this sensor. The MicroZed development board with the Zynq-7000 SoC family [96] offered a great option in this regard for being a cost-optimized solution containing both programmable logic (PL) (based on Artix-7 FPGA) and a processing system (PS) (dual-core ARM Cortex-A9 processor). Of the low-end models, the more expensive Zynq 7020 SoC was selected due to the increased on-chip resources to speed up the debugging process, but due to the very low (<5%) resource usage of our design, even the cheapest model chip could work in an optimized implementation. Although the design is primarily PL-based, the inclusion of the PS allowed for a simple method to transfer stored frames to an SD Card for permanent storage and also helps provide a wider range of future design considerations. Of the Zynq-based development boards, the MicroZed was selected due to its versatile MicroHeader expansion, proven capabilities to handle the high-speed LVDS signals required to interface with the image sensor, and above all, its large 1 GB DDR3 memory for acting as a buffer for frames captured at high framerates.

In order to access the necessary I/O banks in the PL through the MicroHeaders, a carrier board was used. There already exists a MicroZed embedded vision development kit that includes a PYTHON 1300 module, but the overall cost of the evaluation kit is $1049 USD, with the camera module alone being $500 USD [97]. The kit also requires licensed IP and is limited in its customizability and standalone capabilities. We decided to choose the cheaper and more versatile FMC (FPGA Mezzanine Card) carrier board and designed our own camera PCB at a fraction of the cost. The FMC connector can support up to 36 LVDS pairs making this configuration capable of handling much more than the 6 pairs we require if a more capable (and more expensive) sensor is needed. Another benefit of the FMC carrier board was its multi-purpose Peripheral MODule (PMOD) headers which was used to expand the functionality of the camera through pushbuttons, switches, and a VGA output (by Digilent Inc.) to give the camera its standalone quality. A custom-cut acrylic cover protected the system and a C-Mount-threaded lens mount was incorporated to allow mounting onto microscope systems or attach lenses, but it is recommended that future designs use a 3D-printed housing solution to further lower costs. A spacer was also included to lift the C-mount to provide better focus with attached lenses. The flat, open design of our prototype also meant it required no cooling. Figure 5-1 shows a picture of the assembled system with the described components and PMOD attachments.

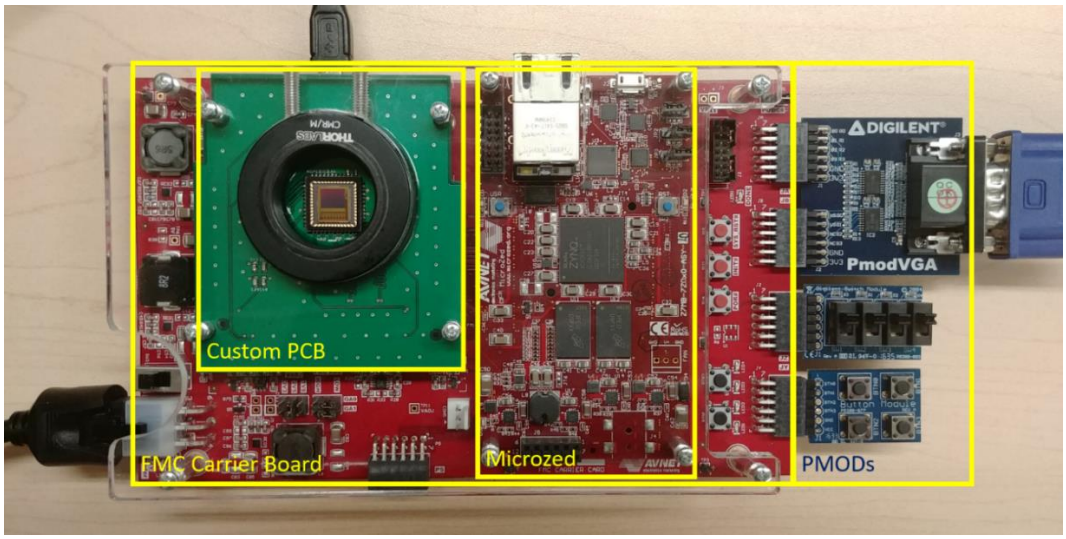


Figure 5-1: Full camera prototype containing the Microzed, FMC carrier board, custom camera sensor module and PMOD extensions. An acrylic protective cover and c-mount lens provide a temporary makeshift solution for safe mounting.

5.2.2. Image Sensor PCB Design

The design required a custom PCB for the image sensor that could supply power to the sensor and route its I/O to the FMC connector pins appropriately. Two supply voltages of 3.3 V and 1.8 V were required, as well as another pixel supply voltage of 3.3 V. The tolerances for each differ and range from 1.5% to 5%, thus requiring separate high-quality voltage regulators to meet the specs when used in conjunction with a general-purpose 5V micro-USB power supply. Although the FMC connector can supply a 12 V power line, the PCB was initially designed to be powered independently for testing and possibly act as a custom carrier board for the MicroZed in the future, so the external power was kept in the design.

The LVDS output of the sensor has a common mode voltage of 1.25 V for transferring pixel, timing and sync data out, while the single-ended signals operate at 3.3 V for sensor configuration, monitoring, and triggering. The LVDS standard required setting the voltage level of the carrier board, and thus the I/O bank used, to 2.5 V to act as a proper receiver. Furthermore, the differential traces on the PCB were designed to be length matched through meandering and set to 100 Ω differential impedance. With the Zynq chip's I/O bank limited to operating at 2.5 V, a voltage level shifter was required to translate the single ended signals between 2.5 and 3.3 V. To better isolate the LVDS lines and improve RF performance, a 4-layer board was used where the 2 internal layers were grounded and via stitching connected the 4 ground planes together. The majority of the components were placed on the back of the PCB so as to minimize the obstruction to any housing solution, such as with the C-mount used. A close-up of the front and back of the PCB is presented in Figure 5-2 and its schematic is provided in Appendix A-2.

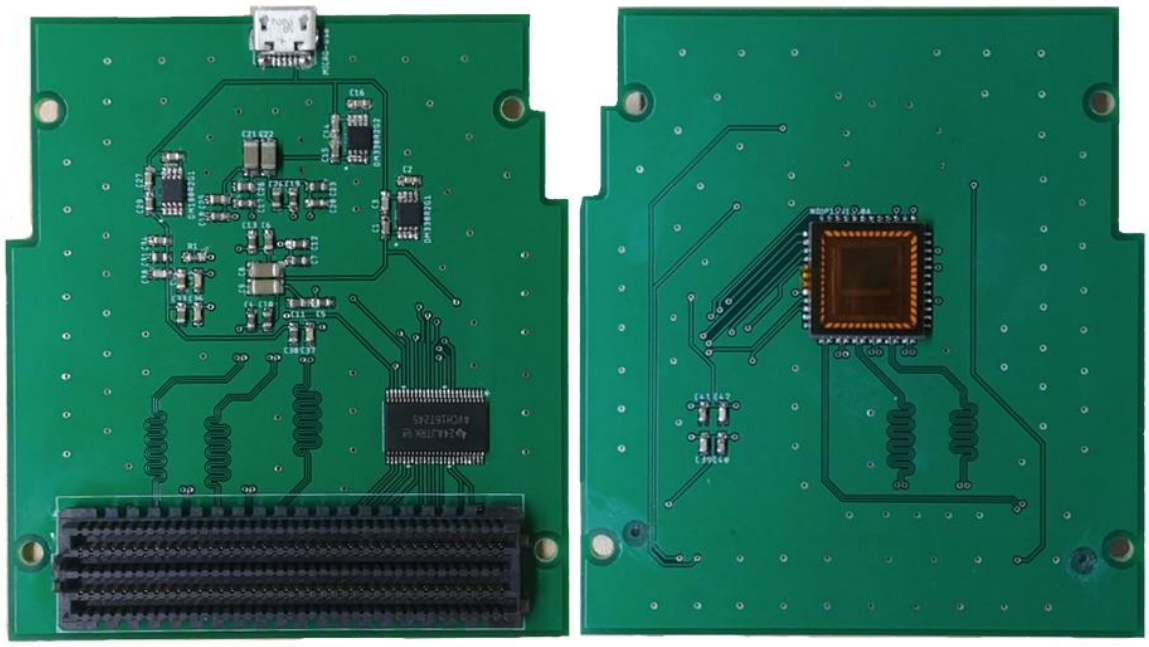


Figure 5-2: PYTHON 1300 Camera sensor PCB design with an FMC connector.

5.2.3. HDL Design

From here, our codebase (written primarily in SystemVerilog) must power the camera sensor correctly, configure it through the SPI protocol, align and process the LVDS input, and finally store or display the captured frames, all operated using pushbuttons (PB) and switches (SW). The flow of this design is displayed in Figure 5-3 and is developed using the free Vivado WebPACK version 2017.4.1 (in which Zynq-7000 SoCs are supported devices. The code can be found through the link found in Appendix A-3.

The LVDS data channels of the image sensor transmitted serially at a double data rate (DDR) of 720 Mbps over the nominal 360 MHz clock sent through the clock channel. A Xilinx reference design (XAPP1017) was used and slightly modified to receive and deserialize the data, synced to the differential clock channel from the sensor. The pixel data was then organized in order due to the unique readout order of the sensor and then sent simultaneously to two different FIFOs (first-in first-out) before being processed and stored into RAM. FIFOs allow their input and output ports to operate at different clock speeds making them useful for designs that cross clock domains. One FIFO facilitates the storage

of frames in frame buffers (FB) for real-time display on a monitor while the other captures all incoming frames for picture or video storage.

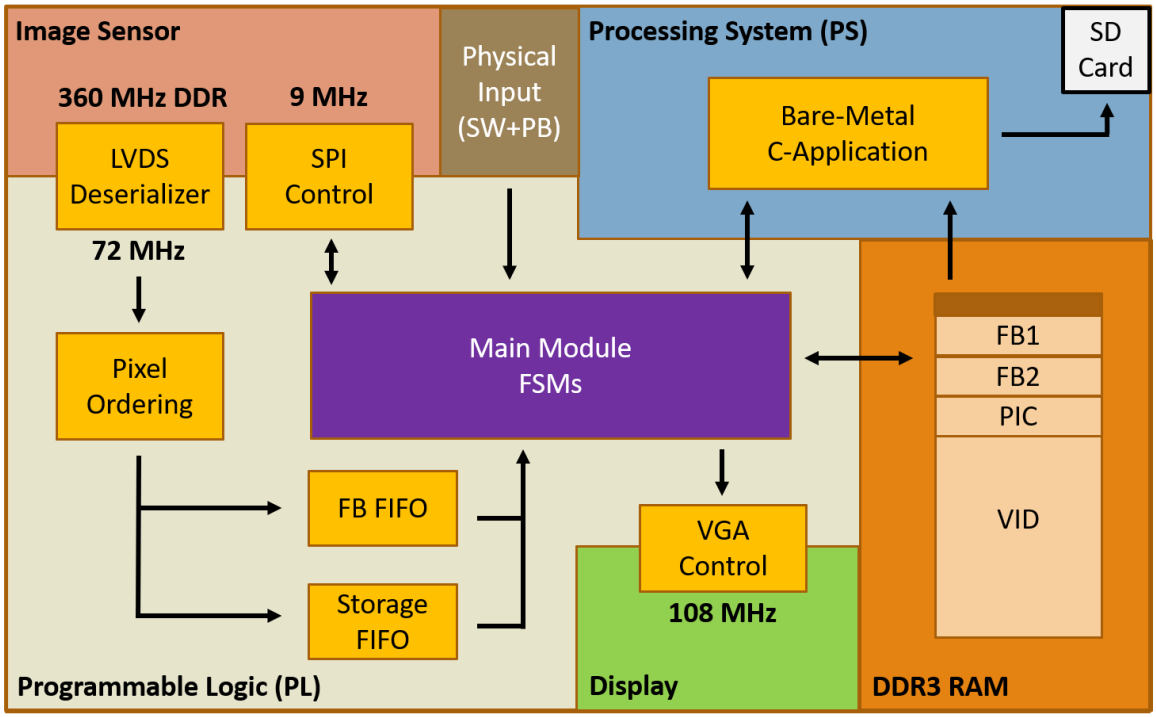


Figure 5-3: Design flow including camera configuration, pixel storage, and display output.

Switching of the FBs is automated based on the completion of the display of one frame and the buffering of the next. When one FB is being filled with pixel data, the other is being read for display. If the display of FB1 is finished and FB2 is not yet completely filled, then FB1 is simply displayed again. Once FB2 is filled with the next frame and the display of FB1 is done, a switch occurs; FB2 is now sent off for display while FB1 fills with the next frame. This design allows for display on a monitor with any framerate that is at or below the framerate of the image sensor as long as the pixel clock speeds required for the display are supported. The VGA control module also contains a FIFO that allows for pre-fetching pixels to ensure that pixels are always available and ready for display in case the RAM is busy with other operations. While the VGA PMOD can output 4-bits of R, G, and B, which can produce $2^{(4+4+4)} = 4096$ different colours, the camera outputs in greyscale, so the 4 MSBs of each pixel is instead fed into the three RGB channels resulting in only $2^4 = 16$

shades of greyscale on the monitor. Future improvements can include an HDMI output on the custom PCB to avoid the lack of pins required for a VGA display, but VGA was used to ensure compatibility with as many displays as possible.

The main module contains the various combinational logic and finite state machines (FSMs) required to power and configure the image sensor, manage control signals and data flow to facilitate all modules working together, and also handle user input through switches and pushbuttons. The required SPI uploads were stored in order using on-chip block RAM for efficient handling. A reset switch was also used as a global reset and to trigger the power down sequence of the image sensor. Switches were used to boot the camera in pre-configured resolutions and framerates by altering the SPI commands sent, while pushbuttons were used to take pictures, start and stop video capture, and initiate SD card storage, with the LEDs providing feedback on the actions taken. The main module interfaces with the DDR3 (double data rate) DRAM (dynamic random-access memory) through a custom Advanced eXtensible Interface (AXI) design that manually handles the data transfer and handshaking protocols. Although the pixels are captured in 10-bit mode, the pixels are reduced to 8-bits before being stored in DRAM. This is not a limitation in the hardware but instead an intentional trade-off since our specific camera uses benefitted more from the 25% extra recording time gained as opposed to an increased dynamic range.

The design does not rely on the MicroBlaze soft processor or any IPs requiring purchase, and circumvents the need for the USB3, GigE, and CoaXPress cores that Xilinx advertises in their machine vision designs [98]. The USB3 or GigE links are what some low-cost machine vision cameras from FLIR utilize, but their connection throughput ultimately hinder the capabilities of the image sensor. For instance, the BFS-U3-13Y3M-C camera uses the USB3 Vision interface with a 380 MB/s maximum throughput limit and the same PYTHON 1300 image sensor as our design, but can only reach 170 fps at maximum resolution (1280x1024) compared to the reported 210 fps from the image sensor datasheet, and only up to 663 fps at a 320x240 resolution [99].

A few IP blocks included with the Vivado webPACK were required to initialize and reset the Zynq PS, as well as generate the required clocks, FIFOs, and memory-mapped

registers for communication between the PL and PS. These registers were used to start the transfer of frames to the SD Card for permanent storage. A bare-metal C-application (written in Xilinx SDK 2017.4) that used only one of the ARM cores and did not require Linux stored the captured frames to the SD card using the FatFS library. A small section at the top of the DRAM was reserved to store the application. The pictures were stored as portable grey map (PGM) files, while videos were stored as binary files with the SD card formatted with a large cluster size to maximize transfer speeds. The binary files are then processed on a computer and encoded into videos using the free, open-source FFmpeg tool.

5.3. Results and Discussions

5.3.1. Camera Specifications

The configurations implemented are shown in Table 5-1, utilizing three switches for simple switching between the eight-state combination of resolutions and framerates. Maximum framerates were calculated by measuring the number of clock cycles between Frame Start signals from the sensor to obtain the time per frame and then inverting the value. Although the PYTHON 1300 datasheet reported a maximum of 2235 fps at 256x256, we were able to surpass this. The commonly used framerates of 60 and 30 fps were made available at max resolution, while all resolutions allowed for a roughly halved framerate in exchange for increased exposure in low light conditions. Storage of a full video of roughly 1 GB to the SD card takes approximately 1.5 minutes.

Table 5-1: List of 8 camera configurations supported by 3 PMOD switches.

Resolution	**Framerates**	**Record Time (s) (8-bit pixels, ~1 GB storage)**
1280x1024 (SXGA)	211 (MAX)	3.86
	100	8.14
	60	13.57
	30	27.13
640x480 (VGA)	817 (MAX)	4.25
	400	8.69
256x256	2329 (MAX)	6.99
	1000	16.29

This design comes with benefits that may be absent from some commercially available cameras. For instance, the VGA display outputs a continuous live feed from the camera even during recording due to the large DRAM bandwidth and separate VGA frame buffer and storage pipelines. Some cameras, especially those that require a host computer, have to disable the live display during recording. With our addition of PBs and SWs, no host computer is required and the camera can be operated in standalone. This removes the need for complicated software that may require training, troubleshooting, or maintenance. On top of the standard function to record until the DRAM is full, the option to make the DRAM a cyclic buffer was also implemented. This allowed for continuous recording that captures the past several seconds (depending on resolution and framerate) until stopped – a useful feature for events that are difficult to time. Resource utilization for this design came to under 5% of the available PL resources on the Zynq 7020 SoC, as seen in Table 2.

This helped the camera achieve a low power consumption, with a Tenma 72-7745 multimeter measuring a peak current draw of 0.283 A (3.396 W) at the 6-pin 12 V power connector of the carrier board during recording. Combined with the power dissipation of the custom PCB coming primarily from the image sensor at 0.620 W, this gave the whole camera a power consumption of roughly 4 W, comparable to machine vision cameras in its price range. Additionally, this gives an indication of the relative simplicity of our custom camera design, allowing for more functionality to be implemented or a lower end chip to be used for future implementations that are cheaper and more optimized. A cost breakdown of the camera is detailed in Table 3, with the overall cost of the camera totaling $656 USD. It is estimated that a full custom board with only the necessary components of this design can be achieved for one third of this price when considering bulk pricing.

Table 5-2: Resource utilization of the Zynq 7020 PL for this design.

	Look-Up Tables (LUTs)	**Flip-Flops (FFs)**	**Block RAM (# 36 Kb Blocks)**
Resources Used	2748	3978	9
Zynq 7020 PL Total [96]	53200	106400	140
Percent Usage	5.2 %	3.7 %	6.4 %

Table 5-3: Cost breakdown of the camera.

Component	Cost (USD)
MicroZed 7020	$213
FMC Carrier Board	$175
Custom Image Sensor PCB	$150
Acrylic cover + C-mount + Spacer	$90
PMODs (VGA + PBs + SWs)	$22 ($9 + $8 + $5)
5 V micro USB power supply	$6
Total:	$656

5.3.2. Figure-of-Merit Comparison

To compare our camera against other high-speed cameras on the market, we developed a figure-of-merit (FoM) that takes into consideration important performance parameters of a camera and outputs a single overall performance number, dependent on the framerate and resolution the user is targeting. The equation developed and used was

$$FoM = \frac{\sqrt{RES_{max}} \cdot RT_{RES_{max}} \cdot FPS_{RES_{max}}}{C \cdot P} \cdot \frac{FPS_{max}}{FPS_{RES_{max}}},$$

$$= \frac{\sqrt{RES_{max}} \cdot RT_{RES_{max}} \cdot FPS_{max}}{C \cdot P}, \quad (5\text{-}1)$$

where RES_{max} is the number of pixels at the maximum resolution for a fixed target framerate, $RT_{RES_{max}}$ and $FPS_{RES_{max}}$ are the recording time (in seconds) and framerate at RES_{max}, respectively, FPS_{max} is the maximum frames per second for a fixed target resolution, C is cost in USD, and P the power consumption in Watts. A square root of RES_{max} was taken to obtain a square dimension and scale the FoM down. The targets set for this comparison was the specifications achieved with our camera: 2329 fps and 256x256 pixels. Thus, the cameras investigated are reported at two configurations: the first to determine their maximum resolutions up to and around 2329 fps (if the camera cannot reach the target framerate, the highest reported framerate configuration is used), and again at a configuration as close to 256x256 pixels as possible to determine a maximum achievable framerate. Separating these two targets allows a proper comparison of the cameras' recording times at a given framerate, as well as give an indication of their high-speed

capabilities to offset their cost if they can surpass the target framerate at the target resolution. The separation is also important since the product of the number of pixels with framerate is typically not constant across different configurations of the same camera due to pixel readout architectures and overheads.

The units of the FoM are $N_{pixels}/\$ \cdot W$, that is, the number of pixels per unit cost and power, where the higher the number, the better. The time dependence associated with framerates is cancelled out due to including recording times at a fixed framerate. This is almost inverse to some metrics used such as in [100] that give a cost per HD frame per second, where the lower the number, the better. In comparison, our FoM provides the benefits of the target separation discussed, the inclusion of recording times for frame buffering capabilities, as well as power consideration, an important metric when faced with factors like heat dissipation and portability. We have nonetheless included an equivalent Cost/FPS metric for comparison, with frames normalized to a 256x256 resolution using the FPS_{max} values. Table 4 contains the FoM and Cost/FPS comparisons between our camera, some machine vision cameras and some of the best affordable standalone cameras with publicly available specifications and prices.

Although resolutions close to 256x256 was chosen for FPS_{max}, the fps4000 Lite comes with a fixed configuration of 720p, potentially skewing its FoM. For the battery powered and portable Chronos 1.4 and fps4000 Lite cameras, the power was calculated by researching the typical W·h rating of the battery and using the maximum reported usage/recording time for the camera. The machine vision cameras can record for much longer periods thanks to their reduced framerates and dependence on the memory of a host computer, but for proper inclusion to the FoM comparison against standalone cameras, the reported internal buffers of the machine vision cameras were used to calculate a theoretical recording time. While the above comparison is based on pure standalone hardware performance and high-speed functionality for the price, it is important to note that the commercial cameras reported may be more user friendly, have better portability, or have more advanced software features and options compared to our camera. The benefits of our

design are better realized when considering the potential for added functionality and processing capabilities due to its modularity and low resource usage.

Table 5-4: FoM comparison of several low-cost, high-speed cameras against our prototype.

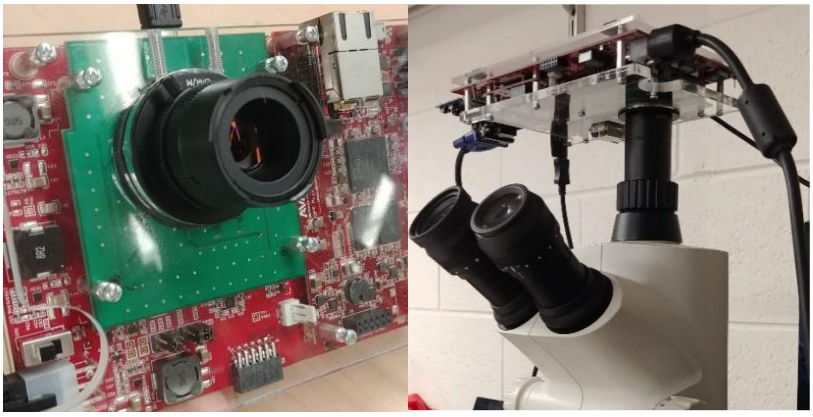
Category	Camera	RES_{max} @ FPS (up to ~2329 fps)	Record Time (s) (w/ 8-bit pixels)	FPS_{max} @ RES (down to ~256x256)	Power (W)	Cost (USD)	FoM (N_{pixels}/ $·W)	Cost/FPS (256x256 frames) ($/frame)
Stand-alone	Chronos 1.4 [84]	1024x576 @ 2359	6.17	15200 @ 336x252	19.2[a]	2999	1251	0.15
	fps4000 Lite [85]	1280x720 @ 950	60	950 @ 1280x720	11.1[b]	2421[c]	2036	0.18
	Edgertronic SC1 [101]	640x480 @ 2324	22.41[d]	7181 @ 320x240	30	5495	541	0.65
	Sony RX100 V [102]	912x308 @ 1000	7	1000 @ 912x308	2.6	1000	1427	0.23
Machine Vision	GS3-U3-23S6M-C [103]	416x240 @ 697	1.84[d]	697 @ 416x240	4.5	1045	86	0.98
	BFS-U3-04S2M-CS [104]	320x240 @ 997	3.13[d]	997 @ 320x240	3	365	791	0.31
	BFS-U3-13Y3M-C [105]	320x240 @ 663	4.71[d]	663 @ 320x240	3	415	696	0.53
	FL3-GE-13S2C-CS [106]	320x240 @ 136	3.06[d]	136 @ 320x240	2.5	695	66	4.36
	Our Prototype Camera	256x256 @ 2329	6.99	2329 @ 256x256	4	656	1588	0.28

[a] Assuming EN-EL4a (28.9 W·h) batteries for the reported 1.5 hours battery life when recording
[b] Assuming 2 x 18650 Li ion (3.7 V, 3000 mA·h) batteries for the maximum reported 2 hours battery life
[c] Converted from Pound sterling
[d] Calculated using internal buffer at RES_{max} config. (top to bottom: 16 GB, 128 MB, 240MB, 240 MB, 32 MB)

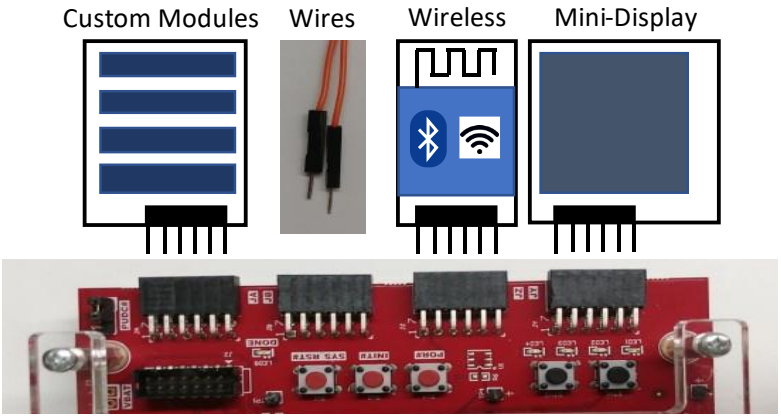
5.3.3. Camera Customizability

Various aspects of the design and build of our camera system allow for modifications that can range from simple to complex. This includes adding software or hardware code to replacing or creating custom PMOD attachments and boards. Currently, the camera

contains a C-mount for attachable lenses and mounting on microscope systems, along with PMOD VGA output, pushbuttons, and switches. These can be substituted for different modular attachments that are commercially available, such as the Bluetooth or 96x64 RGB OLED (organic light-emitting diode) display PMODs from Digilent Inc. [107]. Example code and resources are provided for them online and they give the added functionality to control the camera remotely or offer a miniature real-time display, omitting the need of a monitor. Custom attachments are also viable and remaining/exposed pins can be manually connected to. These types of hardware modifications are illustrated in Figure 5-4.



(**a**)



(**b**)

Figure 5-4: Examples of operation and some hardware customizability examples: (**a**) Camera with attachable lenses or mounted onto a microscope; (**b**) Different modular attachments or custom solutions to fit the user's needs.

More complex modifications can include the addition of an HDMI output on the custom PCB along with a custom HDMI Verilog module to bypass the bit-depth limitation of the VGA output. Additionally, should higher framerates or higher resolutions be needed for an application, more demanding image sensors can be substituted given the 36 LVDS pairs that the FMC connection allows for. The PCB and Verilog code would have to be modified to handle the extra data lines, but the overall cost of the camera would primarily rise only with respect to the image sensor. With the MicroZed Zynq SoCs having up to 24 LVDS pairs per I/O bank capable of 950 Mb/s of DDR data and a DRAM bandwidth above 4 GB/s (1066 MT/s on an 32-bit wide interface) [108], custom carrier boards or fully custom board designs can maximize I/O usage and minimize size, power, and cost.

Regarding the codebase, extra software can be written on the processor for a more sophisticated user experience, and the USB and Ethernet ports allows the camera to communicate with a host computer if preferred. However, care is to be taken on how much memory is required for the application should it have to be stored in DRAM, affecting the recording length. Moreover, the current very low utilization of the programmable logic means a wide range of processing capabilities can also be implemented. With FPGA circuits being highly parallel and spatially oriented on the fabric, extra Verilog modules can be added to the code base with little to no impact on existing performance as long as the signals and data are properly buffered to meet timing requirements.

Simple features that may be difficult to find in commercial cameras are readily implementable due to the user having access to all the internal signals in the Verilog code. For instance, individual frames of a video can be tagged during recording using input pulses to mark specific events, proving useful in synchronized testing environments involving other sensors. Accessing the frame synchronization information from the image sensor in Verilog and reading an input pulse from a wire on a spare PMOD pin (at the appropriate voltage level of the carrier board) makes this relatively easy to implement. The large remaining resources can be utilized to create and test custom image processing algorithms such as real-time image classification [109] and object detection [110]. Some of this research requires separate high-speed and potentially expensive cameras as input, such as

in [111] where two high-speed camera heads are used with a platform consisting of two FPGAs to create a high-speed vision system. The same team also described a fast multi-object feature extraction algorithm [112] implemented on one of the FPGAs with low resource usage capable of fitting within the remaining PL resources from our design. A system based around our camera design provides a cost-effective platform with customizability suitable for building and testing such systems and algorithms.

5.4. Application

5.4.1. General Applications

A convenient sampling rate or framerate for an event is typically several times the event frequency, but can be lower as long as it is greater than twice the event frequency required by the Nyquist Sampling Theorem [113]. Thus, cameras that can achieve framerates into the low thousands are useful in a wide range of applications that require up to millisecond or sub-millisecond time resolutions. A recent proposal for a low-cost test rig for impact experiments on dummy heads suggested framerates equal to or above 200 fps [114]. Tests have also been done on the free-fall drop impact of portable products at 1000 fps [115] and automotive crash testing cameras typically run between 1000-4000 fps to meet automotive safety compliance standards [116]. Furthermore, such cameras can be used for biomechanics research such as was done with the analysis of wing flap aerodynamics of the hummingbird [117] as well as their comparison to that of bats [118], with recordings from 500 – 2000 fps. Another interesting example of biomechanical research investigated the relationship between trans-glottal airflow and vocal fold vibration (recorded at 1900 fps at 256x64 resolution) [119]. Moreover, popular technology reviewers use high-speed imaging to perform comparisons or verify commercial product claims. One such test compared the input lag associated with Nvidia’s G-Sync and AMD’s FreeSync technologies using a Sony FS700 camera at 960 fps [120] since the propagation of a signal from keyboard to monitor is on the order of milliseconds.

Low-cost high-speed camera systems are suitable for many of these applications and more since they do not require such high-end equipment or features and can be run at reduced resolutions. In addition, the cost benefit of such cameras becomes important when considering applications that may currently be done using other sensor technologies. An example of this is the measurement of human reaction time, which can be difficult to accurately and inexpensively measure. While low-cost techniques have been developed to tackle this such as with wireless motion sensors [121], the use of a low-cost high-speed camera remains a simple and effective solution capable of providing much higher accuracy to either replace, test, or calibrate existing systems.

A demonstration of the camera's high-speed capabilities is in recording sound vibrations, akin to research from [122] that used high-speed cameras to "image sound". Tests were run at 2200 fps at 700x400 resolution using a Phantom V10 camera to observe the low frequency spatial vibrations of objects and obtain speech and sound from the resulting high-speed video. One clear limitation of this technique is the use of expensive cameras to record video at high enough framerates to cover a suitable range of frequencies. We demonstrate capturing the vibrations of an inexpensive 440 Hz tuning fork at 2329 fps with 256x256 resolution. Its frequency, as well as the 120 Hz flicker from the ambient fluorescent lighting, was determined accurately from an FFT of the changes in pixel intensity of a cropped section of the frames focused on the vibration of one fork tine, as seen in Figure 5-5. The experiment was run with no extra light sources resulting in dark frames due to the low exposure associated with such a high framerate. This was done to test the success of such an experiment without the need of investing in extra lighting equipment.

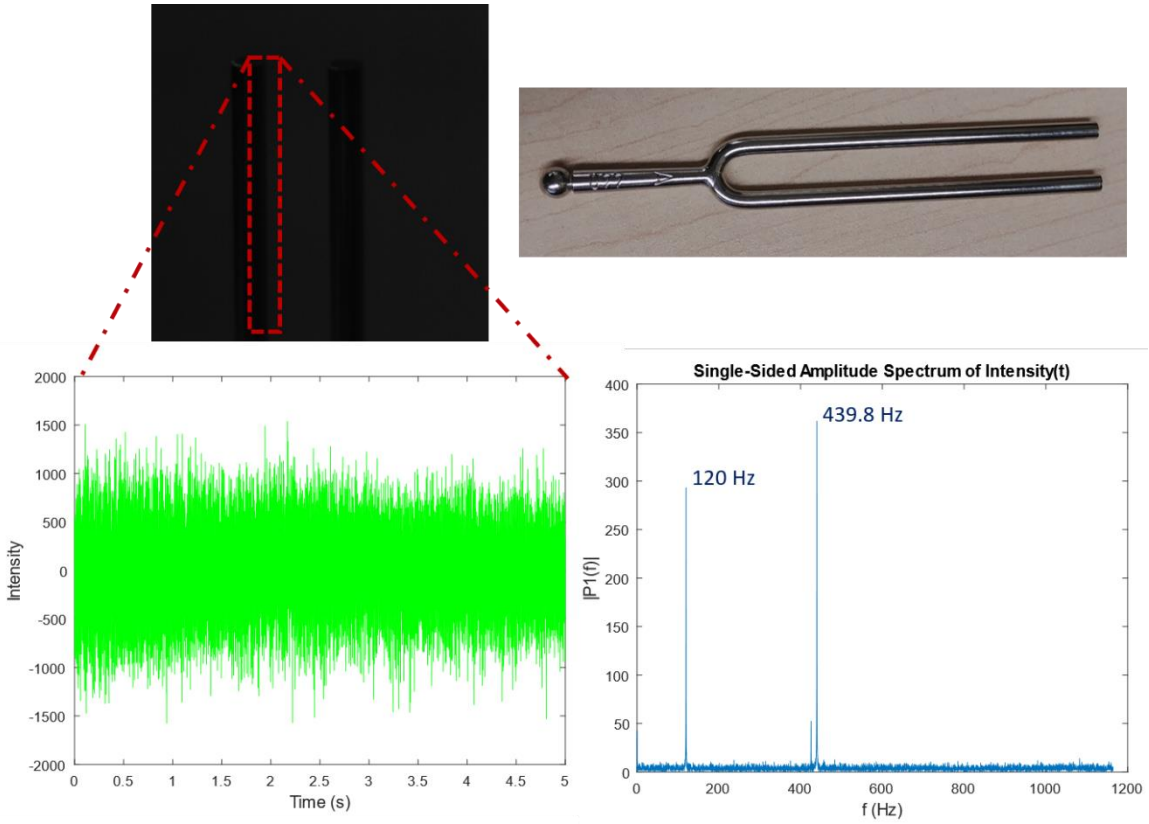


Figure 5-5: A 440 Hz tuning fork captured at 2329 fps using only ambient lighting. Fourier analysis identified the fork's frequency as well as the 120 Hz flicker of the ambient fluorescent lighting.

A key research field where such cameras can be most useful is in microscopy [123]. Here, strong sources of light usually accompany microscope systems, creating an environment that is well-suited for high-speed capture. Depending on the wavelength (λ) of light and the numerical aperture (NA) of the objective and condenser lenses used, the optical resolution (r_{opt}) achievable by a microscope is given by the Rayleigh criterion [93] as

$$r_{opt} = \frac{1.22\lambda}{NA_{obj} + NA_{cond}} \tag{5-2}$$

and represents the minimum resolvable distance. Assuming green light (λ = 525 nm) and a microscope system with NA_{obj} of 0.65 and NA_{cond} of 0.95, the optical resolution would be ~0.4 µm. With a magnification (M) of the objective lens and a target sampling rate within the optical resolution obtained from Equation 2, the pixel size is calculated as

$$pixel\ size = \frac{M \cdot r_{opt}}{sampling\ rate} \tag{5-3}$$

Three to six pixels within the optical resolution was shown to be ideal for cell microscopy [124], satisfying the Nyquist Sampling Theorem without excessive oversampling that can create false details due to secondary diffraction. Thus, using the optical resolution of 0.4 µm previously calculated, along with a magnification of 10x and a target sampling rate of 3, a required pixel size of ~5.3 µm can be calculated. The small 4.4 µm pixel size of the PYTHON 1300 image sensor selected for our camera makes it suitable for this example and other common microscope configurations.

5.4.2. Application Involving a Nematode System

Our camera was compared against an existing FL3-GE 13S2C-C machine vision colour camera (included in the FoM table) currently in use as a microscope camera to test single image quality of *C. elegans*. *C. elegans* (commonly referred as nematode or worm) is the leading animal model to study the conserved biological pathways and disease phenomena in a laboratory setting [125], [126]. The animals are transparent with adults of about 1 mm length that feed on microbes, primarily bacteria. Since their first introduction by Brenner [127], *C. elegans* have been used extensively in biomedical research in wide areas such as neurobiology, cell biology, and aging. Among other advantages, these animals offer a short life cycle, small size, compact genome and the ease of propagation. Figure 5-6 shows a few frames taken with both cameras at their native resolutions and at 30 fps for comparable results. Our camera shows similar or better image quality for a lower cost, while being standalone and customizable. This gives it the added benefit of unhindered high-speed capture, real-time display even during recording, continuous recording with a cyclic frame buffer, and no need for a host computer.

C. elegans has been used extensively for research on aging-related processes. The hermaphrodites have a short lifespan of around 3 weeks, making it possible to study age-associated changes in tissues and processes [128]. Genetic experiments in *C. elegans* have uncovered the roles of many aging-associated genes and pathways that are conserved across

eukaryotes [129], [130]. The findings have revealed that aging is a progressive increase in fragility that results in increased mortality rate as a function of time [131]. Two of the age associated markers in *C. elegans* are body bending and pharyngeal pumping (i.e., feeding). The rates of these physiological markers decline as animals get older. However, since the existing procedures typically involve manual scoring of these changes [132], the quantification is labor intensive, subjective, and prone to error. Furthermore, alterations in phenotypes cannot always be reliably detected. Our camera allows for the high-speed observation and recording of subtle changes during aging of these animals. The camera is also capable of capturing various other behavioral processes such as defects in mating, thrashing rate, defecation length cycle and real time rolling. All these behaviours were captured, and snapshots are given in Figure 5-7.

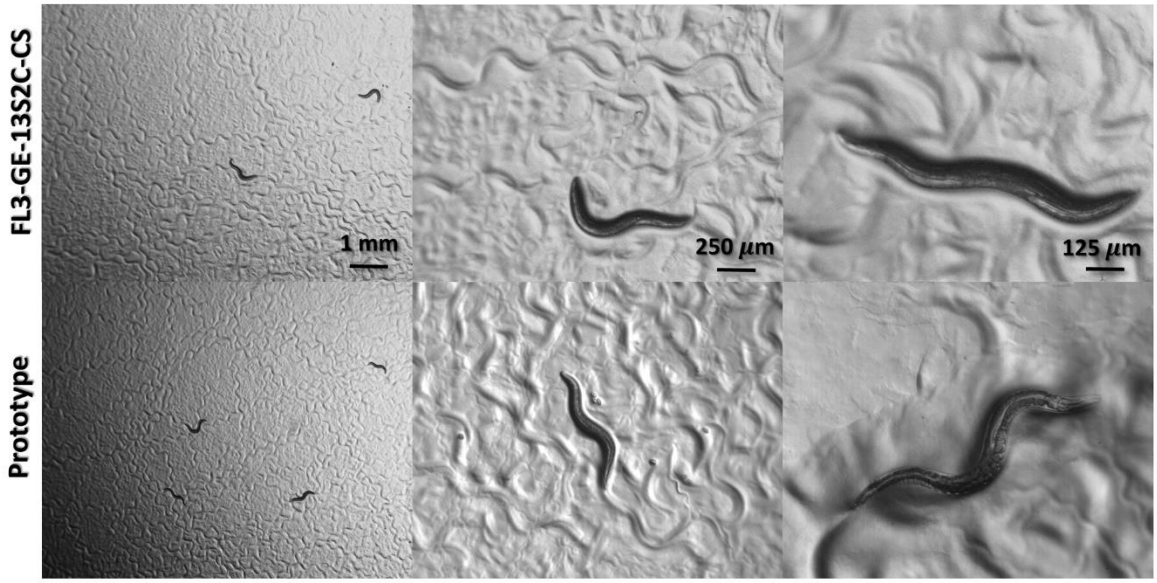


Figure 5-6: A comparsion of images at 3 different magnifications using the FL3-GE 13S2C-CS (top row) and our prototype (bottom row) at their maximum resolutions (1288x964 and 1280x1024, respectively) and 30 fps for both. All the images show day-1 wild type adult animals.

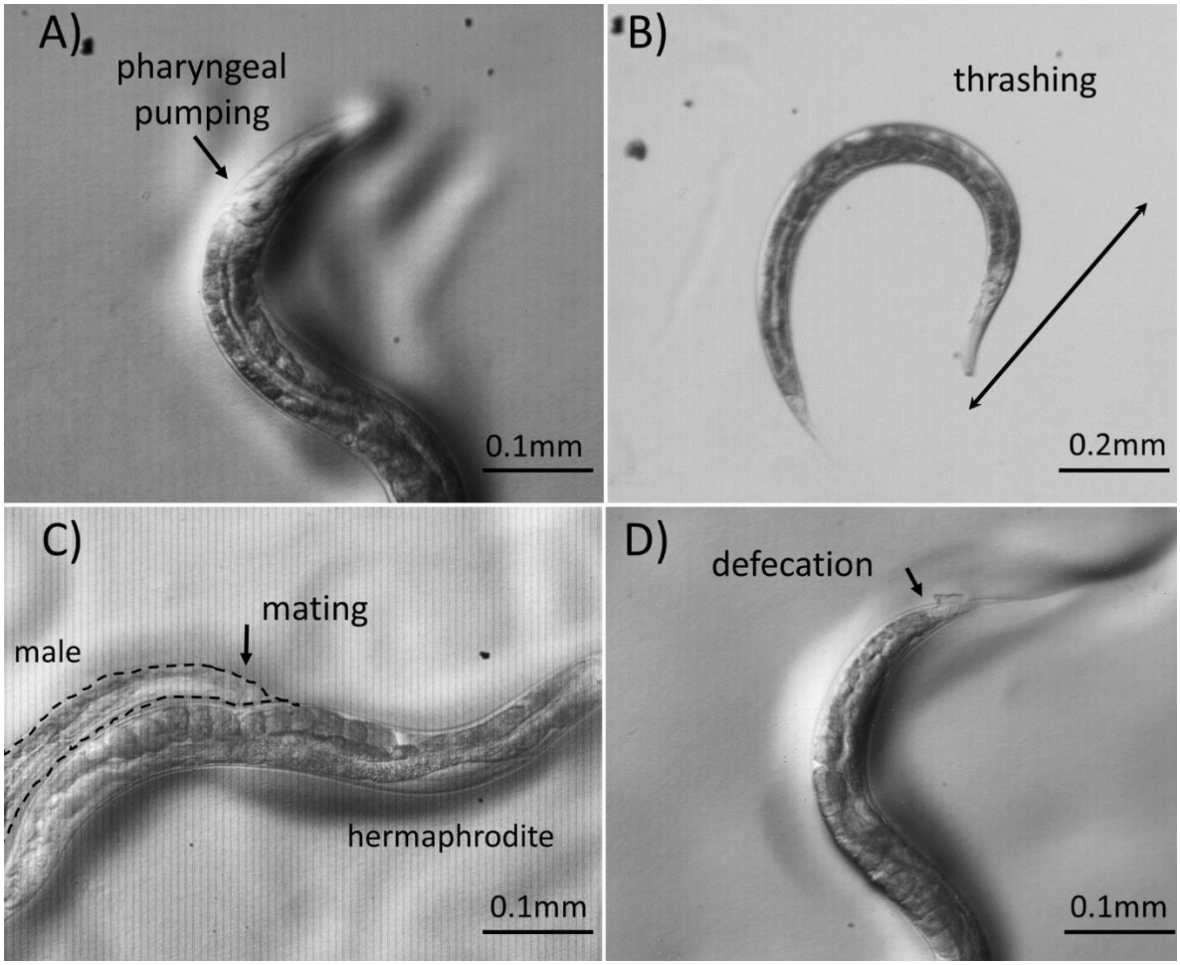


Figure 5-7: Aging-associated phenotypes and other nematode specific behaviour are made easier to analyze by capturing videos at high speed. Still images of various processes taken from the videos of respective speed include: (**a**) Pharyngeal pumping at 100 fps; (**b**) Thrashing at 210fps; (**c**) Mating behaviour at 210fps and 8x analog gain; and (**d**) Defecation at 210fps.

While this high-speed imaging and recording is possible due to strong microscope light sources, the imaging of fluorescence, for instance, is difficult at high framerates due to the potentially very low fluorescent light levels. These situations favour large pixel sizes for better light collection, especially if trying to record at high framerates, putting the small (4.4 µm) pixel size of the image sensor used in our camera at a disadvantage. However, by recording at higher exposures (reduced framerates down to 30 fps) or modifying the digital or analog gain (at the cost of potential noise increase), it becomes possible. This is demonstrated in Figure 5-8, where a green fluorescent protein *(GFP)*-tagged *C. elegans* worm was recorded and imaged with increased exposure at 30 fps.

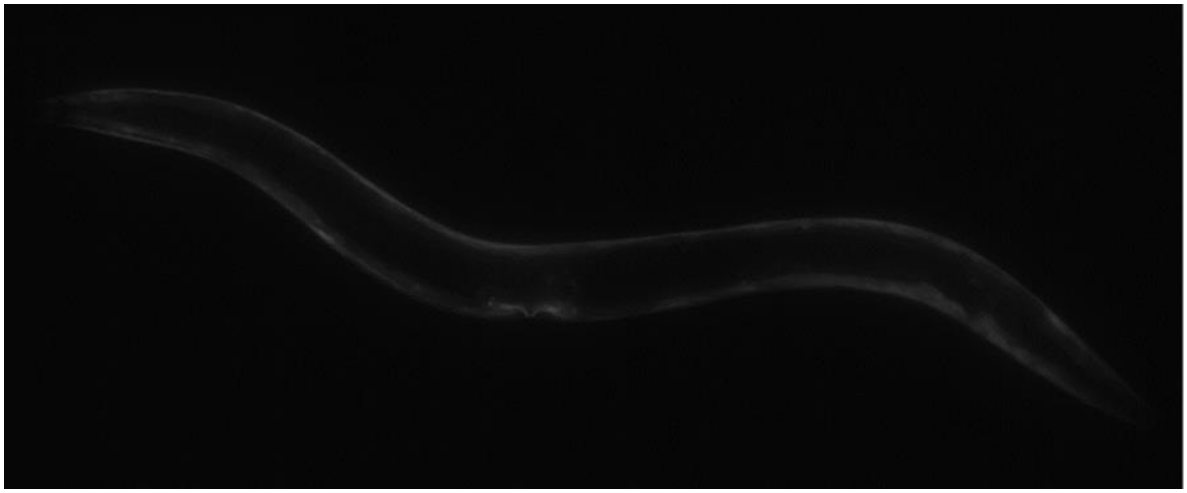


Figure 5-8: GFP fluorescence captured at 30 fps. Image composed of two pictures stitched together to obtain a wider view. Transgenic animal expressing myo-3p::GFP has been used.

5.5. Conclusions

We developed a low-cost, standalone, high-speed camera prototype to showcase the capability and cost-effectiveness of CMOS image sensors for high-performance imagers. It is capable of >2000 fps at reduced resolutions using a Zynq system-on-chip development environment and a custom, open-source codebase. The camera was shown to be competitive to some commercial standalone and machine vision high-speed cameras using a developed figure-of-merit showing their price-to-performance values. It avoided the hardware restrictions that hindered other low-cost machine vision cameras while offering similar functionality to more expensive standalone cameras. The customizability of the camera as a result of the low programmable logic resource usage and potential for custom or commercial modular attachments is explored, providing a cost-effective platform for the development and testing of high-speed imaging systems and image processing algorithms. The camera proved successful in the capture of low frequency spatial vibrations in low light conditions and was tested against an existing microscope camera, revealing greater details of processes used to investigate age-associated physiological changes in *C. elegans.*

Chapter 6
Conclusions and Future Work

6.1. Conclusions

In summary, this thesis focused on the design and test of novel SPAD structures and pixel architectures in standard CMOS processes for reduced cost and complexity. Such performance gains can benefit applications such as time-resolved diffuse optical imaging (TR-DOI), where a miniature, low-cost, and portable TR-DOI prototype device was proposed that could determine brain oxygenation levels in pre-term or new-born infants, as well as assist in on-the-spot diagnosis of traumatic brain injury (TBI) such as concussion in accidents and sports related injuries. The first stage of this research involves the development of an optimal photodetector that can be integrated with low-cost digital circuitry, have high timing precision, and enhanced photodetecting capabilities with potential wavelength distinction. Thus, a triple-junction (TJ) SPAD and a time-gated (TG) pixel design was fabricated in the TSMC standard 65 nm CMOS process and their performance was tested in order to determine the feasibility of the technology for future array designs. Using an available CMOS image sensor on the market today, a low-cost and customizable highspeed camera was also developed as a proof of principle that demonstrates the ease of interfacing with existing CMOS photodetectors and building an imaging device with available hardware on the market.

In developing an optimal photodetector, the performance improvements and reduced costs associated with the evolution of single photon detection from photomultiplier tubes to solid-state single-photon avalanche diodes (SPADs) was summarized. A review of SPAD performance parameters, design considerations, and state-of-the-art structures was also given. It was revealed that although very good overall performance can be had through custom technologies and 3D stacking, fabrication in a standard CMOS process offers the

most cost-effective approach towards future monolithic integration with other standard digital circuitry for robust system-on-chips. The smaller the technology node, the faster and more power and space efficient the surrounding digital circuitry. However, the main limiting factors in using smaller, standard CMOS processes are the high doping profiles which can lead to junctions with increased noise due to tunnelling, and the inability to modify or remove some process layers that may inhibit performance, such as the dielectric stack over the active area that significantly reduces PDE. Improved PDE, potential wavelength distinction, and greater SNR performance can be beneficial for a wide range of applications, but TR-DOI was focused on as one potential application where significant improvements can be made.

A TJ SPAD in the TSMC standard 65 nm CMOS process was designed to improve PDE performance through widening the spectral response and allowing wavelength distinction due to the varying expected wavelength response of the junctions at different depths. The three junctions from top to bottom were the n+/p-well, p-well/deep n-well (DNW), and the DNW/p-substrate. Using an unbuffered design, the breakdown voltage and temperature coefficients of the three junctions were determined, but high noise in the top two junctions prevented the proper testing of their photodetecting capabilities. It was believed the use of a shallow trench isolation (STI) guard ring (GR) and the increased tunnelling due to the higher doping of the upper junctions all contributed to this increased noise, especially when a high load capacitance oscilloscope probe was required to test the unbuffered design. The bottom junction showed expected functionality and was tested further for performance in photon detection efficiency (PDE), dark count rate (DCR), and timing jitter. Although a low DCR (550 cps/μm^2) was obtained due to the lower doping and reduced sources of defects, a low PDE (0.15% @ 440 nm) was also obtained that did not show an expected peak shift towards longer wavelengths.

The performance of the top shallow junction was testable through a time-gated (TG) pixel design that consists of only 5 transistors and achieving a fill factor (FF) of 28.6%. An associated pulse generation circuit takes an input trigger to produce a fixed gate window of ~3 ns in which the SPAD is armed so that it can be synchronized with only events of interest

and avoid background light, thus increasing SNR. The small gate window allows a fast count rate and a fast transition time to be armed lends itself to improved distinction between early and late arriving photons in TR-DOI applications. However, a high DCR severely impacted performance limiting excess voltages to around 0.3 V and count rates to <100 MHz to obtain good SNR between detected photons and the noise floor. The associated hold-off time when the SPAD is gated allows trapped carriers to be released without triggering an AP event and an AP probability <1% was achieved with a gating frequency of 40 MHz (hold-off time of ~22 ns). Timing jitters of <200 ps and <125 ps were obtained at excess voltages of 0.3 V and 0.4 V respectively, making it much under the 500 ps requirement to be suitable for TR-DOI applications as stated in literature. An impressive 3.5% peak PDE (@ 440 nm) was also obtained when biased at an excess voltage of 0.3 V. If higher excess voltages could be attained despite the excessive noise observed, the overall performance results could be comparable to or better than the results of other SPADs in literature that were obtained at relatively much higher excess voltages.

A dual-junction time-gated (DJTG) design was simulated to take advantage of both multi-junction and TG structures while overcoming the biasing and fill factor difficulties. The junctions are biased in an alternating fashion thanks to a modified pulse generation circuit so that the count rate could be doubled at no cost to hold-off time for each junction to reduce AP. However, this needs improvement in regards to reliability and increased range of excess voltage operation since it is heavily reliant on junctions with similar breakdown voltages which can be difficult in standard processes.

Separately, the performance of the TG design was found to be beneficial to TR-DOI applications, but the TJ SPAD failed to show noticeable difference in the spectral response of the top and bottom junctions, with both peaking around 440 nm. This was believed to be the result of using a standard process where the dielectric stack above the active area largely influenced the observed maxima and minima. This reveals that pursuing something like the DJTG or multi-junction designs for an optimal photodetector for the proposed NIROT application may not be fully realizable in the 65 nm process used.

Lastly, the low-cost, customizable, highspeed camera that was developed showed performance on-par with existing commercial low-cost, highspeed cameras, but with a cost comparable to low-end machine vision cameras. It utilized a PYTHON 1300 monochrome image sensor and a development board with a Zynq 7020 SoC while only utilizing 5% of its available programmable logic. Along with an open-source design that does not require licensed IP blocks or software, this leads the way for others to develop sophisticated image processing algorithms and improve the design in many ways. A unique figure-of-merit was also created that can compare highspeed cameras and the camera was tested in a biological environment to demonstrate its effectiveness. This ultimately showed the cost-effectiveness and simplicity of developing with CMOS photodetectors, off-the-shelf components and hardware as an example for future envisioned low-cost and portable imaging prototypes.

6.2. Future Work

In developing an optimized SPAD photodetector in smaller, standard CMOS technologies, the challenges encountered reveal potential research areas for future studies in optimizing performance. Furthermore, new applications can be investigated with regards to the design techniques applied in this research. These are listed below:

- As technology continues to scale down, it is clear that performance is becoming more dependent on the understanding of noise sources. Thus, progress can be made by characterizing defects to more deeply understand the physics related to charge transfer, after-pulsing, dark current, and associated time constants. Furthermore, while dark counts and photon counts are believed to be identical when detected, differentiating between them may be possible using statistical methods involving threshold selection, IAT variations, and machine learning algorithms to identify variations in directly observed SPAD junction waveforms. Threshold selection was already shown to affect the amount of detected AP events, so an optimal threshold may be obtainable and IAT statistics may eventually become possible to simulate with a proper model. This can lead to a faster and more cost-efficient development time.

- The benefits of a multi-junction SPAD could not be fully explored in this work utilizing a standard 65 nm CMOS process, but still holds great value for future designs. Utilizing junctions at different depths can not only provide colour information and facilitate parallel counting, but can also be used to test the properties of the CMOS process itself. The unique spectral responses, noise, and timing performances of the vertically stacked junctions can reveal information about the doping process and location of defects to better evaluate and refine the process.
- In this work, the pulse generation technique employed for the TG design was refined over a previous work by using digital logic to create one of the pulses based on the other two. However, still relying on capacitors in a CMOS process to generate delays lends itself to uncertainty with process variations and takes up valuable space on silicon. Developing a smaller pulse generation circuit with fine control and reliability over the gating window would prove valuable, especially when wanting to drive an array of TG-SPAD pixels. The array can be modified to have rows of pixels be gated with different time offsets to automate the process of collecting data at various gate delays that is typically done manually. This streamlines and significantly speeds up the data collection process.
- The low-cost, highspeed camera demonstrated the ease of building with CMOS image sensors using available hardware from major retailers, but can still be improved upon. As an open-source prototype, it opens the door to affordable high-performance imaging solutions for both industry and educational purposes. Sophisticated FPGA image processing algorithms can be implemented on the FPGA fabric, custom boards can be fabricated for smaller and cheaper builds, or a more capable image sensors can be used for even higher performance.

References

[1] M. D. Eisaman, J. Fan, A. Migdall, and S. V Polyakov, "Invited Review Article: Single-photon sources and detectors," *Rev. Sci. Instrum.*, vol. 82, p. 71101, 2011.

[2] D. Renker, "Geiger-mode avalanche photodiodes, history, properties and problems," *Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip.*, vol. 567, no. 1, pp. 48–56, Nov. 2006.

[3] 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.

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

[5] S. Cova, F. Zappa, A. Tosi, and M. Ghioni, "Avalanche Diodes and Circuits for Infrared Photon Counting and Timing: Retrospect and Prospect.," in *2006 Digest of the LEOS Summer Topical Meetings*, pp. 7–8.

[6] E. Charbon, "Single-photon Imaging in CMOS," in *Proc.SPIE*, 2010, vol. 7780.

[7] C. Bruschini, H. Homulle, and E. Charbon, "Ten years of biophotonics single-photon SPAD imager applications: retrospective and outlook," 2017, vol. 10069, p. 100691S.

[8] W. Jiang, Y. Chalich, and J. M. Deen, "Sensors for Positron Emission Tomography Applications," *Sensors* , vol. 19, no. 22. 2019.

[9] 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, 2017.

[10] 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.

[11] 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)*, 2018, pp. 96–98.

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

[13] *Photomultiplier Tubes – Basics and Applications*, 3rd Ed. Hamamatsu Photonics K. K., 2006.

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

[15] 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.

[16] Hamamatsu Photonics K. K. Editorial Committee, *Photomultiplier Tubes - Basics and Applications*, 3rd ed. Hamamatsu Photonics K. K. Electron Tube Divison, 2007.

[17] K. Ng, "Complete Guide to Semiconductor Devices," New York: McGraw-Hill, 1999, pp. 425–426.

[18] Hamamatsu, "Si APD (avalanche photodiode) Selection Guide," 2019. .

[19] D. A. Neamen, *Semiconductor physics and devices : basic principles*. McGraw-Hill, 2012.

[20] 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)*, 2016, pp. 264–267.

[21] C. Piemonte and A. Gola, "Overview on the main parameters and technology of modern Silicon Photomultipliers," *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 926. Elsevier B.V., pp. 2–15, 11-May-2019.

[22] S. Lloyd-Fox, A. Blasi, and C. E. Elwell, "Illuminating the developing brain: The past, present and future of functional near infrared spectroscopy," *Neurosci. Biobehav. Rev.*, vol. 34, no. 3, pp. 269–284, 2010.

[23] A. Puszka *et al.*, "Time-resolved diffuse optical tomography using fast-gated single-photon avalanche diodes," *Biomed. Opt. Express*, vol. 4, no. 8, p. 1351, Aug. 2013.

[24] A. Farina *et al.*, "In-vivo multilaboratory investigation of the optical properties of the human head.," *Biomed. Opt. Express*, vol. 6, no. 7, pp. 2609–23, Jul. 2015.

[25] D. J. Davies *et al.*, "Near-Infrared Spectroscopy in the Monitoring of Adult Traumatic Brain Injury: A Review," *J. Neurotrauma*, vol. 32, no. 13, pp. 933–941, Jan. 2015.

[26] D. Milzman *et al.*, "IMPACT OF AIR FLIGHT ON GAMES MISSED POST CONCUSSION IN NHL PLAYERS," *Br. J. Sports Med.*, vol. 48, no. 7, pp. 639.1-639, Apr. 2014.

[27] A. D. Mora *et al.*, "Towards next-generation time-domain diffuse optics for extreme depth penetration and sensitivity," *Biomed. Opt. Express*, vol. 6, no. 5, pp. 1749–1760, 2015.

[28] A. Bozkurt and B. Onaral, "Safety assessment of near infrared light emitting diodes for diffuse optical measurements.," *Biomed. Eng. Online*, vol. 3, no. 1, Mar. 2004.

[29] A. Di Costanzo-Mata *et al.*, "Time-resolved NIROT 'pioneer' system for imaging oxygenation of the preterm brain: Preliminary results," in *Advances in Experimental Medicine and Biology*, vol. 1232, Springer, 2020, pp. 347–354.

[30] M.-J. Lee *et al.*, "High-Performance Back-Illuminated Three-Dimensional Stacked Single-Photon Avalanche Diode Implemented in 45-nm CMOS Technology," *IEEE J. Sel. Top. Quantum Electron.*, vol. 24, no. 6, pp. 1–9, Nov. 2018.

[31] M. Perenzoni, L. Pancheri, and D. Stoppa, "Compact SPAD-Based Pixel Architectures for Time-Resolved Image Sensors," *Sensors*, vol. 16, no. 5, May 2016.

[32] M. J. Deen and P. K. Basu, *Silicon Photonics*. Chichester, UK: John Wiley & Sons, Ltd, 2012.

[33] E. Nocerino, "The Semiconductor Multiplication System for Photoelectrons in a Vacuum Silicon Photomultiplier Tube and Related Front End Electronics," MSc thesis, University of Naples Federico II, 2016.

[34] I. Gyongy *et al.*, "Cylindrical microlensing for enhanced collection efficiency of small pixel SPAD arrays in single-molecule localisation microscopy," *Opt. Express*, vol. 26, no. 3, pp. 2280–2291, Feb. 2018.

[35] B.-L. Berube *et al.*, "Implementation Study of Single Photon Avalanche Diodes (SPAD) in 0.8 µm HV CMOS Technology," *IEEE Trans. Nucl. Sci.*, vol. 62, no. 3, pp. 710–718, Jun. 2015.

[36] 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*, 2019, vol. 10882, pp. 70–77.

[37] 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.

[38] 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.

[39] K. Morimoto *et al.*, "A megapixel time-gated SPAD image sensor for 2D and 3D imaging applications," *Optica*, Mar. 2020.

[40] A. Gallivanoni, I. Rech, and M. Ghioni, "Progress in Quenching Circuits for Single Photon Avalanche Diodes," *IEEE Trans. Nucl. Sci.*, 2010.

[41] D. Palubiak, "CMOS single-photon avalanche diodes and time-to-digital converters for time-resolved fluorescence analysis," PhD thesis, McMaster University, 2015.

[42] Y. Xu, P. Xiang, and X. Xie, "Comprehensive understanding of dark count mechanisms of single-photon avalanche diodes fabricated in deep sub-micron CMOS technologies," *Solid. State. Electron.*, vol. 129, pp. 168–174, Mar. 2017.

[43] J. Rhim *et al.*, "Guard-ring dependence of noise characteristics for single-photon avalanche diodes in a standard CMOS technology," in *2017 IEEE 14th International Conference on Group IV Photonics (GFP)*, 2017, pp. 155–156.

[44] M. A. Karami, M. Gersbach, H.-J. Yoon, and E. Charbon, "A new single-photon avalanche diode in 90nm standard CMOS technology," *Opt. Express*, vol. 18, no. 21, pp. 22158–22166, 2010.

[45] E. Charbon, H.-J. Yoon, and Y. Maruyama, *A Geiger mode APD fabricated in standard 65nm CMOS technology*. 2013.

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

[47] E. Sciacca *et al.*, "Silicon planar technology for single-photon optical detectors,"

IEEE Trans. Electron Devices, vol. 50, no. 4, pp. 918–925, 2003.

[48] 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.

[49] 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.

[50] 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.

[51] 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.

[52] H. Finkelstein, M. J. Hsu, and S. C. Esener, "Dual-junction single-photon avalanche diode," *Electron. Lett.*, vol. 43, no. 22, 2007.

[53] 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.

[54] Y.-C. Tsai, Y.-M. Hsin, Y.-N. Zhong, C.-A. Huang, and F.-P. Chou, "Silicon photodetectors with triple p–n junctions in CMOS technology at 650- and 850-nm wavelengths," *Electron. Lett.*, vol. 52, no. 20, pp. 1707–1708, Sep. 2016.

[55] C. Richard *et al.*, "CMOS buried Quad p-n junction photodetector for multi-wavelength analysis," *Opt. Express*, vol. 20, no. 3, pp. 2053–2061, Jan. 2012.

[56] 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.

[57] 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.

[58] R. Henderson, J. Richardson, and L. Grant, "Reduction of band-to-band tunneling in deep-submicron CMOS single photon avalanche photodiodes," 2009.

[59] R. J. Baker, *CMOS Circuit Design, Layout, and Simulation*, 3rd ed. Wiley-IEEE Press, 2010.

[60] 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.

[61] 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, p. 26103, Feb. 2017.

[62] 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, pp. 124–129, 2004.

[63] 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. 86102, 2010.

[64] M. El-Desouki, M. Jamal 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.

[65] 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.

[66] 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, 2016.

[67] I. Gyongy *et al.*, "A 256x256 , 100-kfps, 61% Fill-Factor SPAD Image Sensor for Time-Resolved Microscopy Applications," *IEEE Trans. Electron Devices*, vol. 65, no. 2, pp. 547–554, Feb. 2018.

[68] C. Bruschini, H. Homulle, I. M. Antolovic, S. Burri, and E. Charbon, "Single-photon avalanche diode imagers in biophotonics: review and outlook," *Light Sci. Appl.*, vol. 8, no. 1, 2019.

[69] F. Villa, R. Lussana, D. Portaluppi, A. Tosi, and F. Zappa, "Time-resolved CMOS SPAD arrays: architectures, applications and perspectives," in *Advanced Photon Counting Techniques XI*, 2017, vol. 10212, pp. 42–49.

[70] M. Sanzaro, P. Gattari, F. Villa, A. Tosi, G. Croce, and F. Zappa, "Single-Photon Avalanche Diodes in a 0.16 μm BCD Technology With Sharp Timing Response and Red-Enhanced Sensitivity," *IEEE J. Sel. Top. Quantum Electron.*, vol. 24, no. 2, pp. 1–9, 2018.

[71] D. Bronzi, F. Villa, S. Tisa, A. Tosi, and F. Zappa, "SPAD Figures of Merit for Photon-Counting, Photon-Timing, and Imaging Applications: A Review," *IEEE Sens. J.*, vol. 16, no. 1, pp. 3–12, 2016.

[72] S. Pellegrini and B. Rae, "Fully industrialised single photon avalanche diodes," in *Advanced Photon Counting Techniques XI*, 2017, vol. 10212, pp. 11–21.

[73] Z. Li, "Miniaturization of Time-Gated Raman Spectrometer with a Concave Grating and a CMOS Single Photon Avalanche Diode," PhD thesis, McMaster University, 2015.

[74] 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, 2009.

[75] 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.

[76] S. M. Sze, *Physics of Semiconductor Devices, 2nd Ed.* NY, USA,: Wiley-Interscience New Y, 1981.

[77] M. Moreno-García, L. Pancheri, M. Perenzoni, R. del Río, Ó. G. Vinuesa, and Á.

Rodríguez-Vázquez, "Characterization-Based Modeling of Retriggering and Afterpulsing for Passively Quenched CMOS SPADs," *IEEE Sens. J.*, vol. 19, no. 14, pp. 5700–5709, 2019.

[78] R. Pagano *et al.*, "Dark Current in Silicon Photomultiplier Pixels: Data and Model," *IEEE Trans. Electron Devices*, vol. 59, no. 9, pp. 2410–2416, 2012.

[79] 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.

[80] R. M. Field, S. Realov, and K. L. Shepard, "A 100 fps, Time-Correlated Single-Photon-Counting- Based Fluorescence-Lifetime Imager in 130 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 49, no. 4, pp. 867–880, 2014.

[81] 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, 2010.

[82] C. Accarino *et al.*, "Low Noise and High Photodetection Probability SPAD in 180 nm Standard CMOS Technology," in *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2018, pp. 1–4.

[83] Y. Maruyama, J. Blacksberg, and E. Charbon, "A 1024x8, 700-ps Time-Gated SPAD Line Sensor for Planetary Surface Exploration With Laser Raman Spectroscopy and LIBS," *Solid-State Circuits, IEEE J.*, vol. 49, pp. 179–189, Jan. 2014.

[84] "Chronos 1.4 high-speed camera." [Online]. Available: https://www.krontech.ca/store/Chronos-1-4-high-speed-camera-p92268927. [Accessed: 24-Jun-2019].

[85] "The fps4000 High Speed Camera." [Online]. Available: https://www.slomocamco.com/cameras/. [Accessed: 24-Jun-2019].

[86] "Galaxy S9 64GB (Unlocked)." [Online]. Available: https://www.samsung.com/us/mobile/phones/galaxy-s/galaxy-s9-64gb--unlocked--sm-g960uzkaxaa/. [Accessed: 26-Jun-2019].

[87] "This Cheap High Speed Camera is Made in Canada!!.," 2018. [Online]. Available: https://www.youtube.com/watch?v=2FJXOXSLc3k. [Accessed: 24-Jun-2019].

[88] S. Sirowy and A. Forin, "Where's the Beef? Why FPGAs Are So Fast," Sep. 2008.

[89] A. R. Brodtkorb, C. Dyken, T. R. Hagen, J. M. Hjelmervik, and O. O. Storaasli, "State-of-the-art in Heterogeneous Computing," *Sci. Program.*, vol. 18, no. 1, pp. 1–33, 2010.

[90] D. Bacon, R. Rabbah, and S. Shukla, "FPGA Programming for the Masses," *Queue*, vol. 11, no. 2, pp. 40:40--40:52, Feb. 2013.

[91] G. Lacey, G. W. Taylor, and S. Areibi, "Deep Learning on FPGAs: Past, Present, and Future." 2016.

[92] "Machine Vision Cameras." [Online]. Available: https://www.flir.com/browse/industrial/machine-vision-cameras. [Accessed: 26-Jul-2019].

[93] G. Cox, *Optical imaging techniques in cell biology*. CRC Press, 2012.

[94] M. Kfouri *et al.*, "Toward a Miniaturized Wireless Fluorescence-Based Diagnostic Imaging System," *IEEE J. Sel. Top. Quantum Electron.*, vol. 14, no. 1, pp. 226–234, 2008.

[95] "NOIP1SN1300A PYTHON 1.3/0.5/0.3 MegaPixels Global Shutter CMOS Image Sensors." [Online]. Available: https://www.onsemi.com/pub/Collateral/NOIP1SN1300A-D.PDF. [Accessed: 24-Jun-2019].

[96] "Zynq-7000 All Programmable SoCs Product Tables and Product Selection Guide.," 2019. [Online]. Available: https://www.xilinx.com/support/documentation/selection-guides/zynq-7000-product-selection-guide.pdf.

[97] "ON Semiconductor PYTHON-1300-C Camera Module.," 2017. [Online]. Available: http://microzed.org/sites/default/files/product_briefs/PB-AES-CAM-ON-P1300C-G-V1.pdf. [Accessed: 05-Jul-2019].

[98] "Machine and Computer Vision." [Online]. Available: https://www.xilinx.com/applications/industrial/machine-vision-systems.html. [Accessed: 24-Jun-2019].

[99] "BFS-U3-13Y3 Technical Reference." [Online]. Available: https://flir.app.boxcn.net/s/9kuaj9ly0wzjhp4t0t18pyaym9n1qtwl/file/418605703793.

[100] "HSC Camera Guide." [Online]. Available: http://www.hispeedcams.com/hsc-camera-guide/. [Accessed: 03-Jul-2019].

[101] Edgertronic, "SC1 - 701 fps @ 720p." [Online]. Available: https://www.edgertronic.com/our-cameras/sc1. [Accessed: 01-Jul-2019].

[102] "Full Specifications and Features DSC-RX100M5A." [Online]. Available: https://www.sony.com/electronics/cyber-shot-compact-cameras/dsc-rx100m5a/specifications. [Accessed: 03-Jul-2019].

[103] "Grasshopper3 USB3 Model: GS3-U3-23S6M-C." [Online]. Available: https://www.flir.com/products/grasshopper3-usb3/?model=GS3-U3-23S6M-C. [Accessed: 04-Jul-2019].

[104] "Blackfly S GigE Model: BFS-PGE-04S2M-CS." [Online]. Available: https://www.flir.com/products/blackfly-s-gige/?model=BFS-PGE-04S2M-CS. [Accessed: 04-Jul-2019].

[105] "Blackfly S USB3 Model: BFS-U3-13Y3M-C." [Online]. Available: https://www.flir.com/products/blackfly-s-usb3/?model=BFS-U3-13Y3M-C. [Accessed: 04-Jul-2019].

[106] "Flea3 GigE Model: FL3-GE-13S2C-CS." [Online]. Available: https://www.flir.com/products/flea3-gige/?model=FL3-GE-13S2C-CS. [Accessed: 04-Jul-2019].

[107] "Pmod Modules & Connectors." [Online]. Available: https://store.digilentinc.com/pmod-modules-connectors. [Accessed: 05-Sep-2019].

[108] "MicroZed™ Zynq™ Evaluation and Development and System on Module

Hardware User Guide.," 2017. [Online]. Available: http://zedboard.org/sites/default/files/documentations/5276-MicroZed-HW-UG-v1-7-V1.pdf. [Accessed: 24-Jun-2019].

[109] M. Qasaimeh, A. Sagahyroon, and T. Shanableh, "FPGA-Based Parallel Hardware Architecture for Real-Time Image Classification," *IEEE Trans. Comput. Imaging*, vol. 1, no. 1, pp. 56–70, 2015.

[110] X. Long, S. Hu, Y. Hu, Q. Gu, and I. Ishii, "An FPGA-Based Ultra-High-Speed Object Detection Algorithm with Multi-Frame Information Fusion," *Sensors*, vol. 19, no. 17, 2019.

[111] I. Ishii, T. Tatebe, Q. Gu, Y. Moriue, T. Takaki, and K. Tajima, "2000 fps real-time vision system with high-frame-rate video recording," in *2010 IEEE International Conference on Robotics and Automation*, 2010, pp. 1536–1541.

[112] Q. Gu, T. Takaki, and I. Ishii, "Fast FPGA-Based Multiobject Feature Extraction," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 1, pp. 30–45, 2013.

[113] M. Versluis, "High-speed imaging in fluids," *Experiments in Fluids*, vol. 54, no. 2. Springer-Verlag, p. 1458, 14-Feb-2013.

[114] A. Y. Alhaddad, J.-J. Cabibihan, A. Hayek, and A. Bonarini, "A low-cost test rig for impact experiments on a dummy head," *HardwareX*, vol. 6, 2019.

[115] E. Tempelman, M. M. S. Dwaikat, and C. Spitás, "Experimental and Analytical Study of Free-Fall Drop Impact Testing of Portable Products," *Exp. Mech.*, vol. 52, no. 9, pp. 1385–1395, Nov. 2012.

[116] "How to choose the right camera for your Automotive Application.," 2018. [Online]. Available: https://photron.com/choose-right-camera-automotive-application. [Accessed: 05-Jul-2019].

[117] D. R. Warrick, B. W. Tobalske, and D. R. Powers, "Aerodynamics of the hovering hummingbird," *Nature*, vol. 435, no. 7045, pp. 1094–1097, Jun. 2005.

[118] R. Ingersoll, L. Haizmann, and D. Lentink, "Biomechanics of hover performance in Neotropical hummingbirds versus bats," *Sci. Adv.*, vol. 4, no. 9, 2018.

[119] S. Granqvist, S. Hertegård, H. Larsson, and J. Sundberg, "Simultaneous analysis of vocal fold vibration and transglottal airflow: Exploring a new experimental setup," *J. Voice*, vol. 17, no. 3, pp. 319–330, Sep. 2003.

[120] "FreeSync vs G-Sync Input Lag Comparison.," 2015. [Online]. Available: https://www.youtube.com/watch?v=MzHxhjcE0eQ&t=437s. [Accessed: 05-Jul-2019].

[121] R. Abbasi-Kesbi, H. Memarzadeh-Tehran, and M. J. Deen, "Technique to estimate human reaction time based on visual perception," *Healthc. Technol. Lett.*, vol. 4, no. 2, pp. 73–77, Apr. 2017.

[122] A. Davis, M. Rubinstein, N. Wadhwa, G. J. Mysore, F. Durand, and W. T. Freeman, "The visual microphone: passive recovery of sound from video," *ACM Trans. Graph.*, vol. 33, no. 4, pp. 1–10, Jul. 2014.

[123] Jaggi, B., Deen, M.J., and Palcic, B. Quantitative light microscope using a solid state detector in the primary image plane. US Patent 4845552, 1989.

[124] S. Majumder, “Random Telegraph Signal Noise in CMOS Image Sensor (CIS) and Use of a CIS in a Low-Cost Digital Microscope,” MASc thesis, McMaster University, 2011.

[125] J. Apfeld and S. Alper, “What Can We Learn About Human Disease from the Nematode C. elegans?,” *Methods Mol. Biol.*, vol. 1706, pp. 53–75, 2018.

[126] M. Markaki and N. Tavernarakis, “Modeling human diseases in Caenorhabditis elegans,” *Biotechnol. J.*, vol. 5, no. 12, pp. 1261–1276, Dec. 2010.

[127] S. Brenner, “The genetics of Caenorhabditis elegans.,” *Genetics*, vol. 77, no. 1, pp. 71–94, May 1974.

[128] M. R. Klass, “A method for the isolation of longevity mutants in the nematode Caenorhabditis elegans and initial results,” *Mech. Ageing Dev.*, vol. 22, no. 3, pp. 279–286, 1983.

[129] C. Kenyon, “The first long-lived mutants: discovery of the insulin/IGF-1 pathway for ageing,” *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 366, no. 1561, pp. 9–16, Jan. 2011.

[130] M. Uno and E. Nishida, “Lifespan-regulating genes in C. elegans,” *npj Aging Mech. Dis.*, vol. 2, no. 1, 2016.

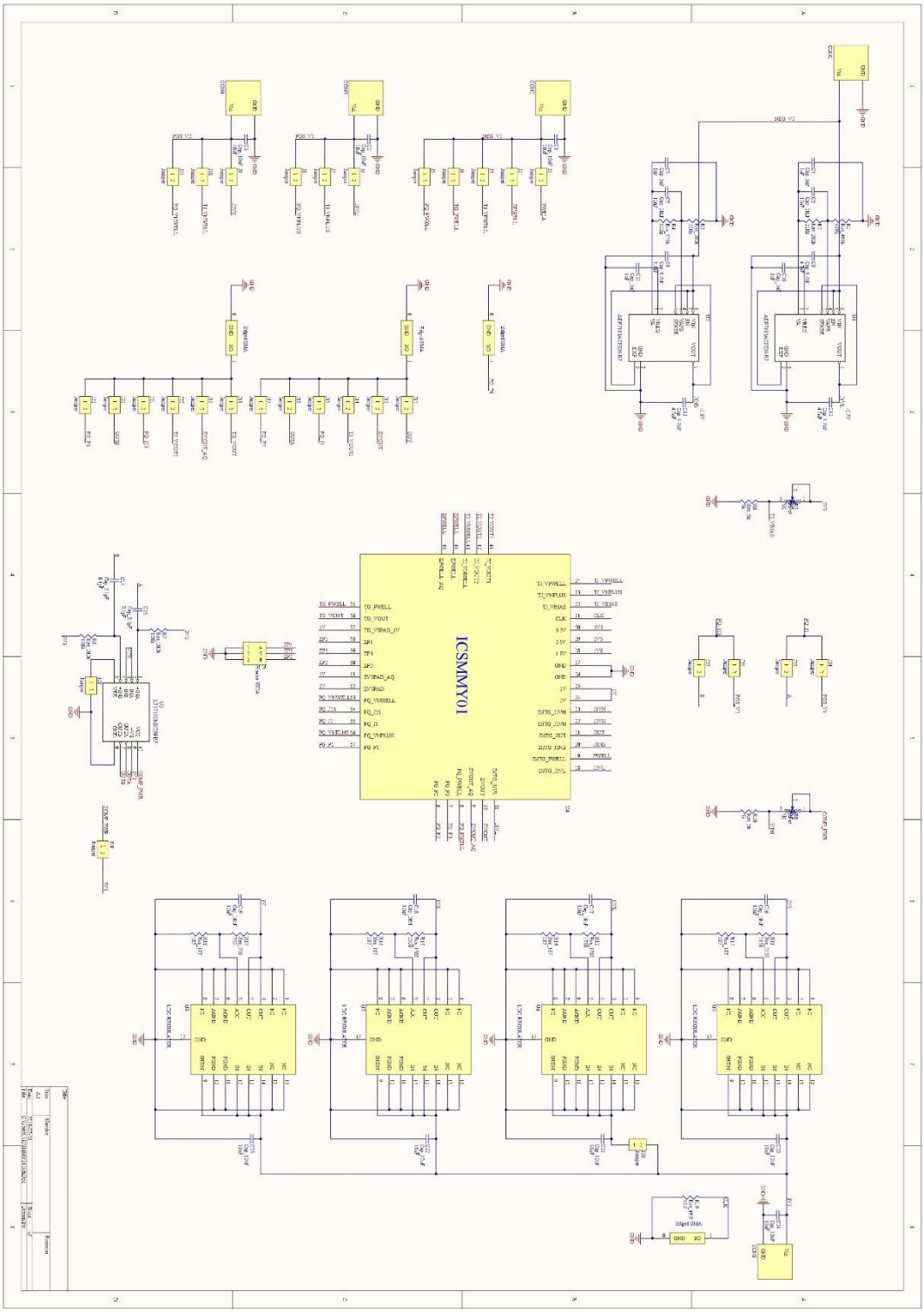
[131] C. E. Finch, *Longevity, senescence, and the genome*. University of Chicago Press, 1994.

[132] C. Huang, C. Xiong, and K. Kornfeld, “Measurements of age-related changes of physiological processes that predict lifespan of Caenorhabditis elegans,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, no. 21, pp. 8084–8089, May 2004.

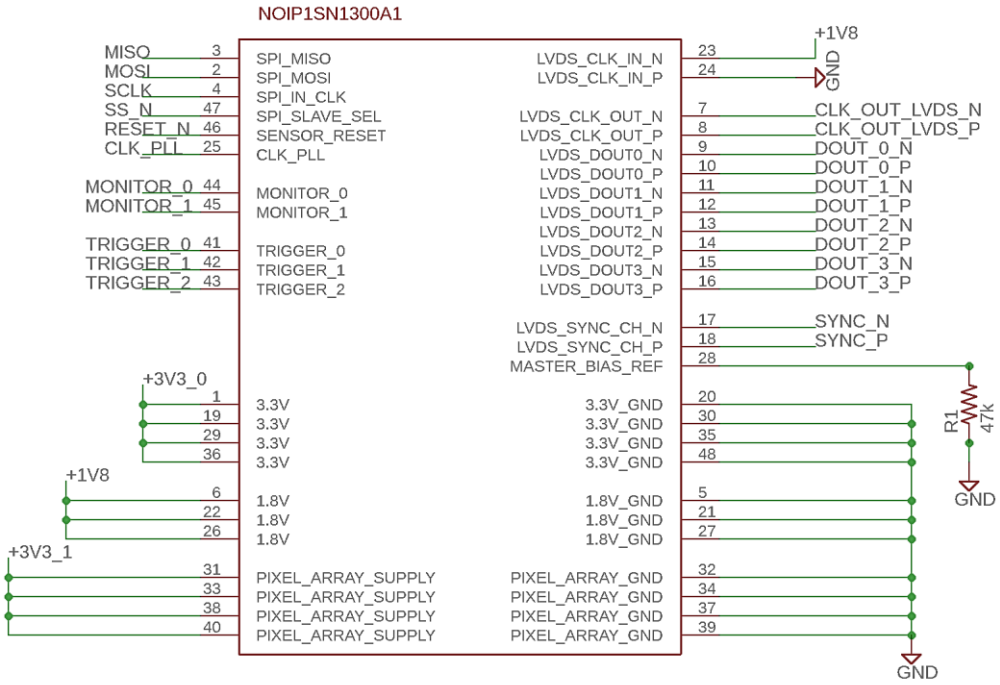
Appendix A

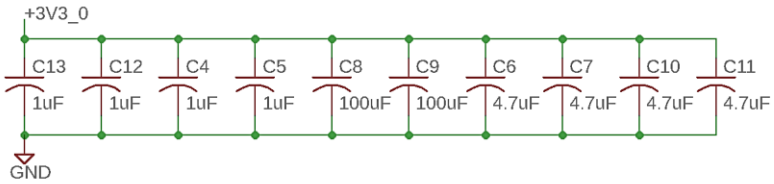
Schematic Designs and Source Code

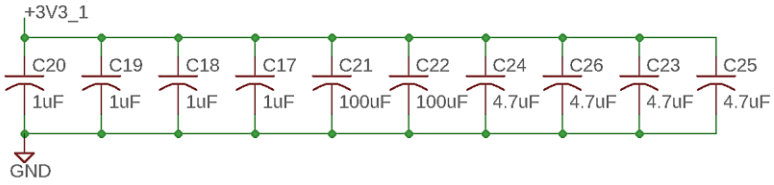
A-1. PCB Schematic for SPAD measurements

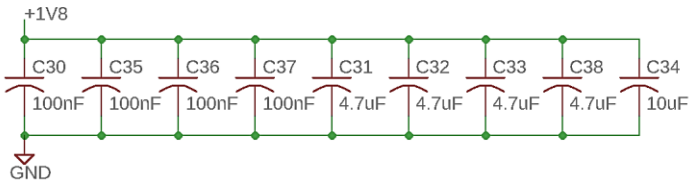


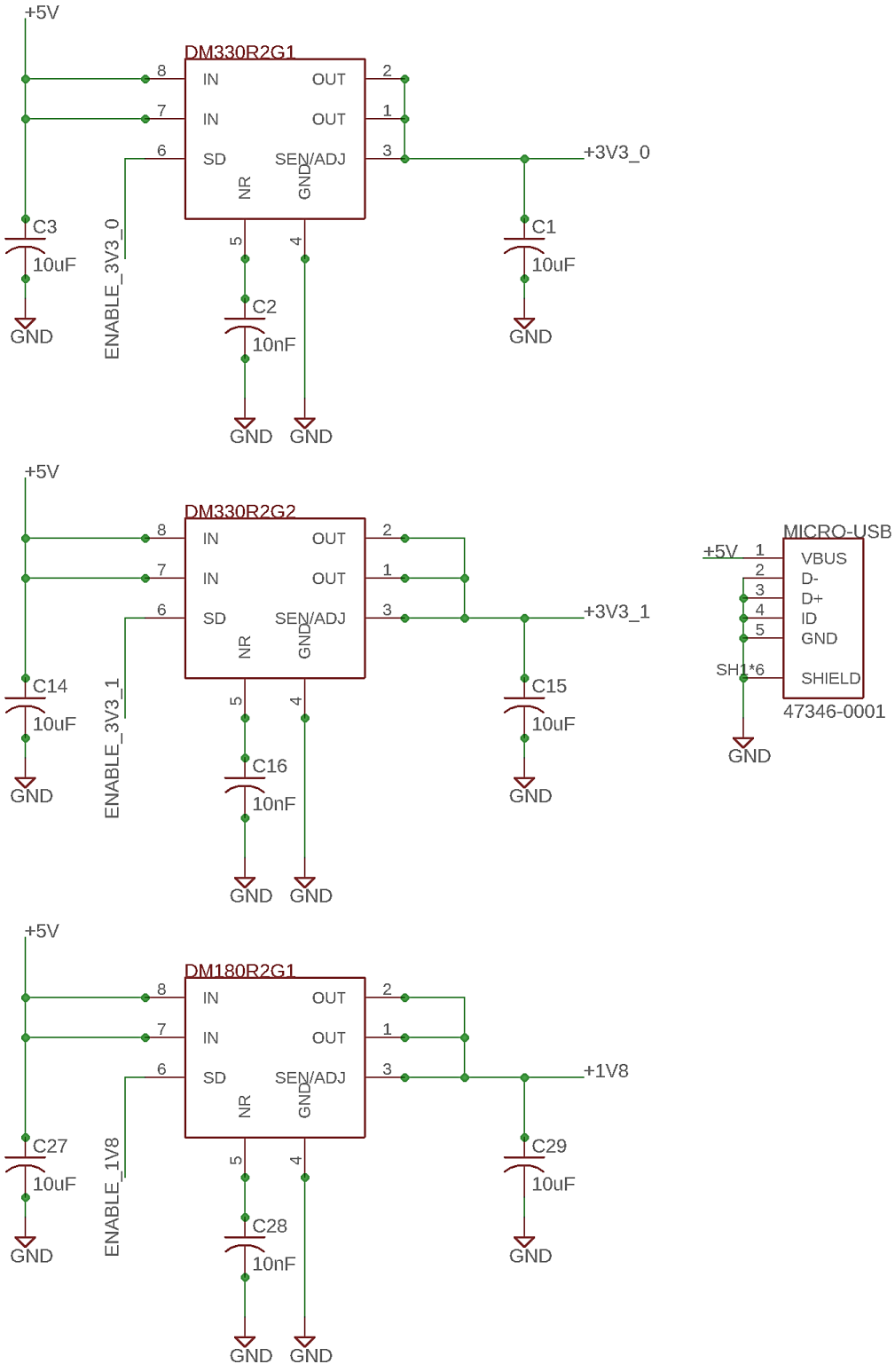
A-2. PCB Schematic for PYTHON 1300 Image Sensor

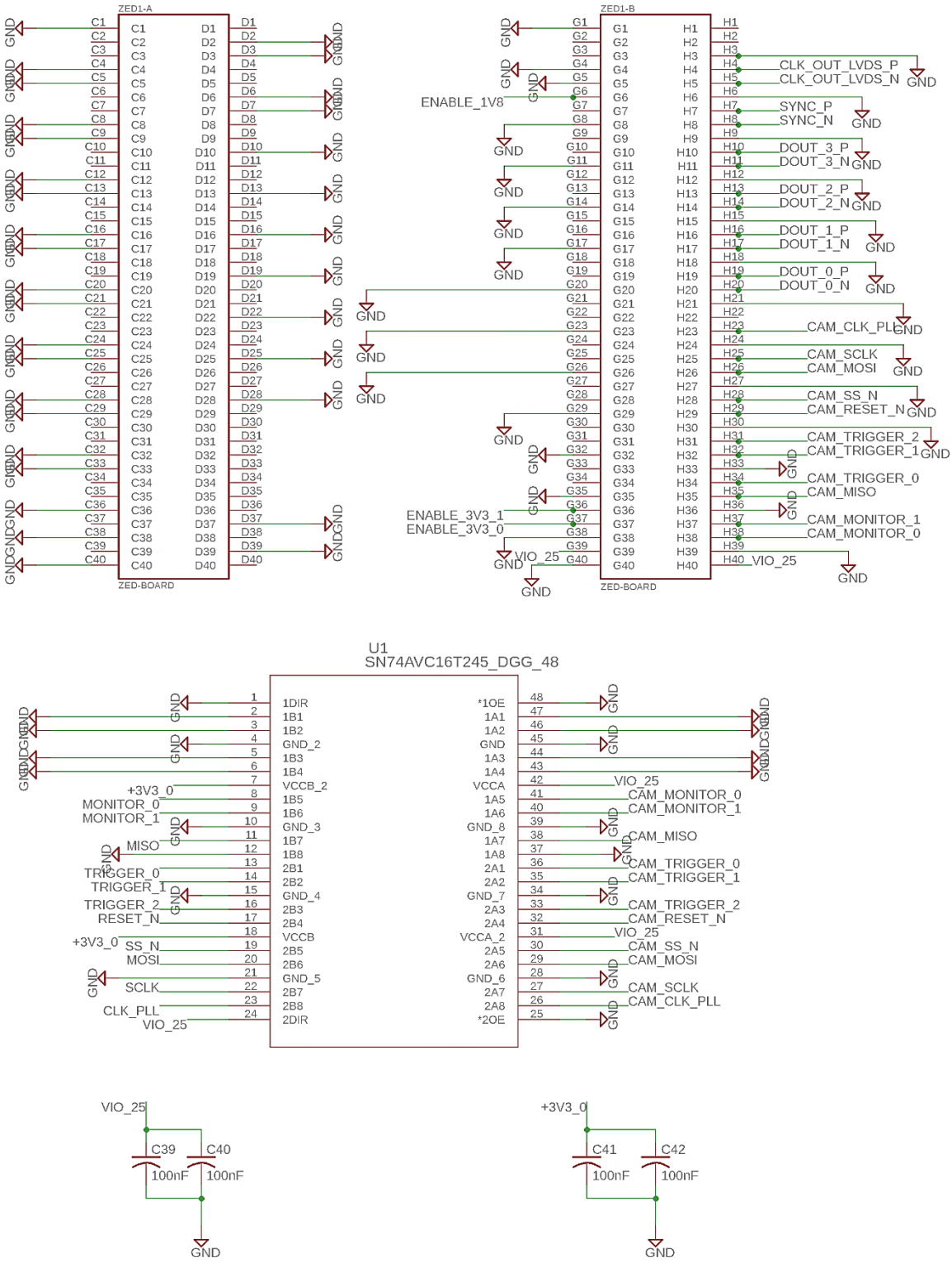









+5V
DM330R2G1
8
IN
7
IN
6
SD
OUT
2
OUT
1
SEN/ADJ
3
NR
5
GND
4
+3V3_0
C3
10uF
GND
ENABLE_3V3_0
C2
10nF
GND
GND
C1
10uF
GND
+5V
DM330R2G2
+3V3_1
C14
10uF
GND
ENABLE_3V3_1
C16
10nF
GND
GND
C15
10uF
GND
MICRO-USB
+5V
1
VBUS
2
D-
3
D+
4
ID
5
GND
SH1*6
SHIELD
47346-0001
GND
+5V
DM180R2G1
+1V8
C27
10uF
GND
ENABLE_1V8
C28
10nF
GND
GND
C29
10uF
GND


ZED1-A
ZED-BOARD
ZED1-B
ZED-BOARD
ENABLE_1V8
ENABLE_3V3_1
ENABLE_3V3_0
VIO_25
CLK_OUT_LVDS_P
CLK_OUT_LVDS_N
SYNC_P
SYNC_N
DOUT_3_P
DOUT_3_N
DOUT_2_P
DOUT_2_N
DOUT_1_P
DOUT_1_N
DOUT_0_P
DOUT_0_N
CAM_CLK_PLL
CAM_SCLK
CAM_MOSI
CAM_SS_N
CAM_RESET_N
CAM_TRIGGER_2
CAM_TRIGGER_1
CAM_TRIGGER_0
CAM_MISO
CAM_MONITOR_1
CAM_MONITOR_0
U1
SN74AVC16T245_DGG_48
+3V3_0
MONITOR_0
MONITOR_1
MISO
TRIGGER_0
TRIGGER_1
TRIGGER_2
RESET_N
SS_N
MOSI
SCLK
CLK_PLL
C39
C40
100nF
C41
C42

A-3. Source Code for the Highspeed Camera

Available online at **https://github.com/yamnchalich/HFRC.git**