Single-Photon Avalanche Diodes for Cancer Diagnosis

Veerappan, Chockalingam

DOI
10.4233/uuid:7db13e84-9ced-4c9c-94fa-8c2a14ad6679

Publication date
2016

Document Version
Final published version

Citation (APA)
Veerappan, C. (2016). *Single-Photon Avalanche Diodes for Cancer Diagnosis*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:7db13e84-9ced-4c9c-94fa-8c2a14ad6679

Single-Photon Avalanche Diodes for Cancer Diagnosis

Ph.D. Thesis

Chockalingam Veerappan

Single-Photon Avalanche Diodes for Cancer Diagnosis

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus Prof. ir. K.Ch.A.M. Luyben,
voorzitter van het College van Promoties,
in het openbaar te verdedigen op donderdag 24 maart 2016 om 12.30 uur

door

Chockalingam VEERAPPAN
elektrotechnisch ingenieur (ir)

geboren te Coimbatore, India.

Dit proefschrift is goedgekeurd door de promotor:
Prof. dr. ir. E. Charbon

Samenstelling promotiecommissie:

Rector Magnificus	voorzitter
Prof. dr. ir. E. Charbon	Technische Universiteit Delft, promotor
Onafhankelijke leden:	
Prof. dr. ir. A. J. P. Theuwissen	Technische Universiteit Delft
Prof. dr. D. R. S. Cumming	University of Glassgow, United Kingdom
Prof. dr. A. Nathan	University of Cambridge, United Kingdom
Prof. dr. C. V. Hoof	IMEC, Belgium
Dr. Erik Jan Lous	ams Netherlands BV, Eindhoven
Dr. I. Rech	Politecnico di Milano, Italia
Prof. dr. ir. P. J. French	Technische Universiteit Delft (reserve lid)

ISBN # 978-94-6186-620-2

Thesis Cover was designed by Reshu Gupta.

... to my parents

Summary

In cancer cell research and in cancer diagnosis equipment, such as positron emission tomography(PET) and single-photon emission computed tomography(SPECT), photonic sensors enabling single-photon sensitivity are required. Until recently, photo multiplier tubes (PMTs) were the sensor of choice. Although a PMT provides the required sensitivity, its size and the need for a high voltage (typically around 1kV) has limited it from being used in high density arrays. In recent past this limitation was overcome when single-photon avalanche diodes (SPADs) have been integrated in CMOS technology. CMOS SPADs have led to the realization of massively parallel arrays of high density single-photon detectors. Today SPAD arrays have surpassed PMTs in terms of spatial resolution and in the capability to time stamp individual photons. This, in addition with magnetic resonance imaging (MRI) compatibility, low voltage requirements (around 25V), and small form factor, has opened new avenues in cancer research. Although, SPAD arrays have a potential to be a viable technology in future cancer diagnostic equipment and in cell research, low sensitivity, high dark noise and high data generation rate have limited its use. This thesis addresses some of these challenges.

In this thesis to ease circuit integration and to reduce electrical crosstalk between SPADs, we focused on improving the single photon sensitivity of the substrate isolated SPAD. In Chapter 3, two design techniques were presented, to enable wide spectral sensitivity. The first technique, is based on the wide depletion junction and, the second, is based on the narrower depletion junction but with a wider photon collection region. For the first technique, three designs were proposed using the p$^+$/deep nwell, pwell/deep nwell and pwell/p-epitaxy/buried-n (p-i-n) junctions. The designs achieved photon detection probability (PDP) greater than 40% from 440nm to 620nm, at the excess bias voltage higher than 8V. The achieved spec-

tral sensitivity is higher and wider than the other known substrate isolated CMOS SPADs. Also when using wide depletion junction, dark noise originating from the tunneling contribution was lowered, resulting in low dark count rate. For instance the p-i-n diode based SPAD measured a DCR of 1.5 cps/μm^2 at 11V excess bias. For the second technique, a device was designed using pplus/nwell junction and with deep nwell and buried-n acting as its photon collection region. At 4V excess bias, the designed device achieved almost the same PDP as the wide depletion junction based SPAD. The pros and cons of these two design techniques are also discussed in this Chapter.

Further, in this thesis, a detailed study was carried out to comprehend the influence of the guard region and the SPAD periphery on the dark noise. The study results are presented in Chapter 4. The results have suggested that the dark noise can also originate from the guard region either due to the guard junction breakdown and/or due to its ineffectiveness at the operating bias voltage. In case of the device periphery the breakdown of the parasitic junction formed between the periphery and one of the main junction terminal was shown to be one of the reasons for the dark noise. In addition, Chapter 4 also discusses in detail the design techniques to thwart/suppress the dark noise that originate from guard region and/or device periphery.

One of the limitation of the SPAD is low fill factor resulting from the inactive area occupied by guard region, periphery, and the circuitry. In this thesis, two different SPAD designs were proposed to reduce the insensitive area. The designs are presented in Chapter 5. The first design, emulates optical microlenses, where the SPAD is designed to operate at or near its guard junction breakdown. Under such operating conditions some of the photocarriers created in the guard junction is sensed by the main junction through the lateral avalanche propagation. Also since the presented design can host transistors on the top of the guard region, without requiring the need for the device periphery, we believe the presented design can lead to the next generation pixel arrays with high density. The second design extends the first, where both the main junction and the guard region were operated above the breakdown. The presented design resulted in 22.2% improvement in fill factor when compared to the conventional design presented in Chapter 3.

When building an imaging system using SPAD sensors, a high data rate resulting from pixel granularity poses a challenge for the data acquisition. In applications such as PET, 100's of SPAD imagers are required to operate in parallel. In Chapter 6, a scalable and a flexible data acquisition system was proposed, based on the sensor network architecture for the PET system. In the proposed approach, every photon module also comprising around 25 SPAD based sensors along with the data processing and communication unit acts as a sensor node. In this configuration, a data network established between the nodes is utilized to perform the distributed data processing to reduce the data in-situ in the system. The proposed approach was shown to be effective for preclinical, brain and clinical PET systems.

Contents

List of Figures

List of Tables

Nomenclature

APCR	Afterpulsing compensated count rate
APP	Afterpulsing probability
APD	Avalanche photodiode
CDR	Coincidence detection packet
CEU	Coincidence engine unit
CP	Coincidence packet
CPG	Coincidence packet generator
CU	Communication unit
DCR	Dark count rate
DAQ	Data acquisition
DPCU	Data processing and communication unit
DPCU	Data processing unit
DTCR	Dead time compensated count rate
DSM	Deep sub-micron
EMCCD	Electron multiplying charge coupled device
EH	Event history

FPGA	Field programmable gate array
FF	Fill factor
FDG	Fludeoxyglucose
FCS	Fluorescence correlation spectroscopy
FLIM	Fluorescence-lifetime imaging microscopy
FRET	Förster resonance energy transfer
FWHM	Full-width at half-maximum
IPC	Input packet controller
LET	Light emission test
LOR	Line of response
MRI	Magnetic resonance imaging
MCR	Measured count rate
OPC	Output packet controller
PMT	Photo multiplier tube
PDP	Photon detection probability
PSPMT	Position sensitive photo multiplier tube
PET	Positron emission tomography
PCB	Printed circuit board
QE	Quantum efficiency
RAM	Random access memory
RAPD	Reach through avalanche photodiode
STI	Shallow trench isolation
SRH	Shockley Read Hall
SiPM	Silicon photo multiplier
SPAD	Single-photon avalanche diodes

SPECT	Single-photon emission computed tomography
TSV	Through silicon via
TDC	Time to digital converter

Chapter 1

Introduction

The world health organization (WHO) has estimated that in 2012 around 14 million people were diagnosed with cancer, of which 8.2 million people have died [5]. It further predicts an increase in cancer incidence by 70% in the next two decades [5]. Today, the success of the cancer treatment depends on its early diagnosis. The state-of-the-art diagnosing equipment is too costly to be available to everyone. In many countries, even today, affording the state-of-the-art equipment is often not possible.

Scientists, in collaboration with medical doctors and engineers, are working on various research projects aiming at building the next generation of low-cost cancer diagnosing equipments. This dissertation is the outcome of one such research activity carried out in targeting the development of a low cost photonic sensor, which has the potential to be used in the next generation cancer diagnosing equipment and in cancer cell studies.

1.1 Photonic sensor

Photonic sensors enabling single-photon detection are currently used in cancer diagnosis e.g. in positron emission tomography (PET) [6] and in single-photon emission computed tomography (SPECT) [7]. In cancer cell studies that use either fluorescence-lifetime imaging microscopy (FLIM) [8], förster resonance energy transfer (FRET) [9] or fluorescence correlation spectroscopy (FCS) [10] single-photon sensors are often required. Until recently, photo multiplier tubes (PMTs) [11, 12] were the sensor of choice in the above mentioned applications. Although, PMTs provide the required sensitivity and timing accuracy, high voltage requirements and its integration complexity when realizing large high density arrays pose a limitation on the application itself.

Researchers have looked at alternatives. The electron multiplying charge coupled device (EMCCD) [13] enables single-photon detection; but the need for gating has limited its use in PET and SPECT. Linear avalanche photodiodes (APDs) [14] are another option; however, the poor timing performance and its associated multiplication noise has limited its use as well. Another alternative researchers have looked at is the single-photon avalanche diode (SPAD) [15], also known as the Geiger mode APDs. SPADs enable single-photon detection with picosecond timing accuracy [16]. The free running property of the SPAD without the need for gating has made it an ideal solid-state replacement for PMTs.

1.2 SPAD

A SPAD is a photodiode, designed to operate above its breakdown voltage. In this mode of operation, also referred to as the Geiger mode, the SPAD is sensitive to single photons. Any electrical engineer when introduced to the SPAD for the first time will be baffled - how can a diode be biased above its breakdown? A general understanding on the diode breakdown is from its I-V characteristics. It needs to be noted that the I-V characteristics represents only the diode static behavior. Under transient conditions, a diode designed with an avalanche breakdown can momentarily be biased at or above its breakdown voltage, provided the diode is devoid of any free carrier to trigger an avalanche.

In a SPAD operated in Geiger mode, a photocarrier generated in the depletion region or in its vicinity can initiate an avalanche breakdown through impact ionization. A high current resulting from the diode breakdown is used to detect single photons. Also, since the statistics involved in the avalanche build-up is in the order of few tens of picoseconds, the photon arrival time can also be measured precisely.

1.2.1 Evolution

1.2.1.1 SPAD in custom process

One of the first SPAD in silicon was designed in a dedicated process. The design is referred to as the reach-through APD (RAPD); it is based on p^+-Π-p-n (Figure 1.1a) [1,17]. Although RAPD achieved high sensitivity, its poor timing performance and the need for a high voltage have limited its use.

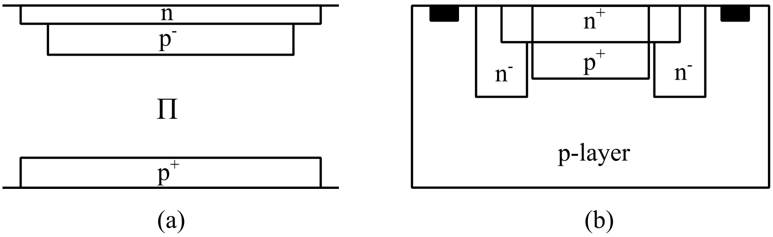


Figure 1.1: Cross section of (a) reach-through APD [1], (b) planar SPAD [2].

Later on, researchers have gradually switched their attention to planar structures, investigated since the 1970s by Cova and others [18]. A planar SPAD cross section is shown in Figure 1.1b. Core of the SPAD is the main junction formed between the n$^+$ and p$^+$. In this design, to avoid the premature edge breakdown at the n$^+$ edges, guard region is designed using n$^-$. The contact to the anode, is provided at the device periphery and using the p-diffusion layer. The area covered by the main junction is also referred to as the active area, as it is the only region that is sensitive to single photon. The main advantage of this design is the low voltage requirements and a better timing performance than RAPD. Planar SPADs designed in silicon have evolved over years aided by novel design and fabrication process enhancements [2, 19–28]. A major drawback in using dedicated processes is its limitation in realizing large arrays.

1.2.1.2 SPAD in CMOS technology

With the introduction of planar SPADs in CMOS [29] (Figure 1.2) and further in deep-submicron (DSM) CMOS [30], the realization of massively parallel arrays of photon detector with time-of-arrival circuitry [31–34] and on-chip data processing was made feasible [35–41]. The state-of-the-art SPAD arrays in CMOS technology have even surpassed PMTs in terms of their spatial resolution and in the capability to time stamp individual photons. These features opened new avenues in cancer research.

For instance, in a scintillator study, SPAD arrays have for the first time enabled the simultaneous mapping of scintillation photons both temporally and spatially [42]. The generated data set of the scintillation is used to improve the timing estimation in PET systems [43, 44]. SPAD's magnetic resonance imaging (MRI) compatibility has enabled the possibility of realizing dual PET-MRI systems [45, 46]. SPAD sensors are used in endoscope probes to help monitor cancer cells during surgery [47,48]. In cancer cell studies, for example, when using FLIM,

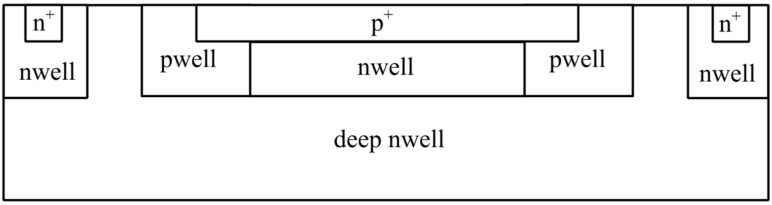


Figure 1.2: Conventional CMOS SPAD cross section. In this design the main junction is engineered between p$^+$ and nwell. A guard ring is designed using pwell. Contact to the cathode, is provided using the deep nwell and the nwell placed at the device periphery.

SPAD arrays have enabled the simultaneous imaging of multiple cells, without the need for scanning [49, 50]. In applications like super resolution microscopy, which has the potential to break new ground in cancer studies, SPAD based sensors are used paving the way to a effective, compact microscopes with nanometric resolution [51, 52].

1.3 Motivation

Though CMOS SPAD arrays have shown their potential as a viable technology for future cancer diagnosis, it needs to be accepted that the CMOS SPADs are still in their budding phase. Some of the challenges faced when using SPAD in cancer research and in diagnosis are presented in this section.

1.3.1 Sensitivity

In cancer cell studies that use either the FLIM, FCS or FRET, a high sensitivity in green to red wavelengths is required. In conventional CMOS SPAD the main junction is engineered near the silicon surface [30, 53–55]. Such designs have resulted in a narrower sensitivity spectrum centered at blue (Figure 1.3).In [56], [57] and in [58] wide spectral sensitive SPADs were reported either using the substrate as one of the junction nodes or the substrate acting as a photon collection region. In either of these designs, the depletion region formed with the substrate requires isolation from transistors and also from the adjacent SPADs to reduce crosstalk. Though a SPAD array designed with the substrate acting as its anode was reported in [59], at the time of the writing of this thesis the impact of the crosstalk was not clear.

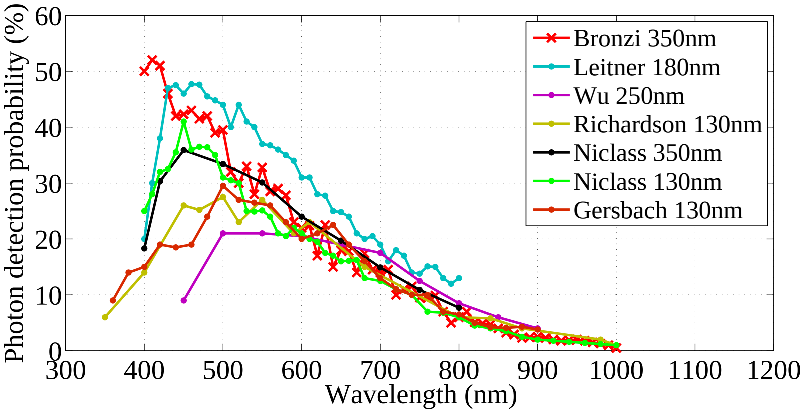


Figure 1.3: State-of-the-art substrate isolated SPAD single photon sensitivity. Note: single photon sensitivity for SPAD is measured as the photon detection probability (PDP), refer Chapter 2 for more details.

1.3.2 Fill factor

The device fill factor defined as the ratio between the sensitive (active) area and the total area, is lower in SPADs, mainly due to the insensitive area occupied by the guard region and the periphery. In CMOS SPAD arrays, the fill factor is reduced even further due to the presence of the on-chip circuitry. For example in one of the biggest SPAD TDC array built till date, the pixel fill factor is around 1% [31].

To improve pixel fill factor researchers have looked at various options. In [38–40, 60–65] novel architectures were proposed for CMOS SPAD arrays to obtain a high fill-factor. Architecturally high fill factor was achieved by reducing pixel granularity and/or by sharing intelligence/processing across pixels. Another option that researchers have looked at is the use of a microlens or prism [66–68], where high fill factor can be achieved without the need to lower either the pixel granularity or the on-chip processing. However, until now microlens misalignments arising due to manufacturing tolerances have resulted in reliability issues, and also in concentration non-uniformity across the arrays.

1.3.3 Dark noise

The dark noise of a CMOS SPAD is generally higher than in other technologies. In DSM CMOS process, the conventional p^+/nwell or n^+/pwell based SPADs have higher noise mainly due to the tunneling [30,53,54]. Though novel techniques were proposed to reduce tunneling noise [55], the observed noise level is still higher. Recently in [69], it was shown that the noise in DSM CMOS SPADs can be brought down to acceptable values using either the imaging process or by the use of additional implants. It needs to be noted that the use of special processing steps will increase the cost of the process.

1.3.4 Data generation rate

Though SPAD arrays provide higher spatial resolution, it has a concomitant drawback of being data intensive [70]. This is a major issue, when using it in applications like PET and SPECT, where hundreds of SPAD arrays are required to operate simultaneously. For example, a clinical PET system designed using SPAD sensors will generate around 525 Gbps (Chapter 6). To handle such high data rates, novel data acquisition (DAQ) techniques are required.

1.4 Contribution

This dissertation deals with two challenges associated with using CMOS SPADs in cancer research and in diagnosis. The first lies in improving CMOS SPAD performance and the other in addressing data acquisition complexity when using SPADs in a PET system. The contribution of this thesis is presented in this section.

1.4.1 Wide spectral response SPAD

This thesis is mainly focused on substrate isolated SPAD design, as it can ease circuit integration complexity and can reduce crosstalk. For the substrate isolated SPAD, two different techniques for the active area design were studied, to increase sensitivity in the green to red regions where applications like FLIM, FCS and FRET operate.

The designs proposed in this thesis achieved wide spectral response attaining >40% PDP from 440nm to 620nm. When compared to the state-of-the-art devices, the SPADs presented in this work achieved 19% and 45% improvement in PDP in

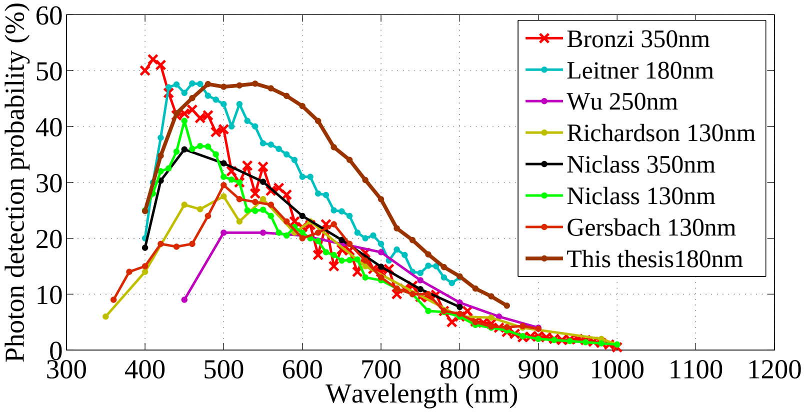


Figure 1.4: This thesis vs the state-of-the-art substrate isolated SPAD PDP.

green (540nm) and in red (640nm) respectively. Figure 1.4 compares the state-of-the-art with a design from this thesis.

1.4.2 Low noise SPAD

Generally it is believed that the dark noise originates in the SPAD active area. In the study carried out in this thesis, it is shown that also the guard region and the device periphery can contribute the SPAD dark noise. Different design techniques for the guard region and periphery were also proposed to reduce their impact on the dark noise.

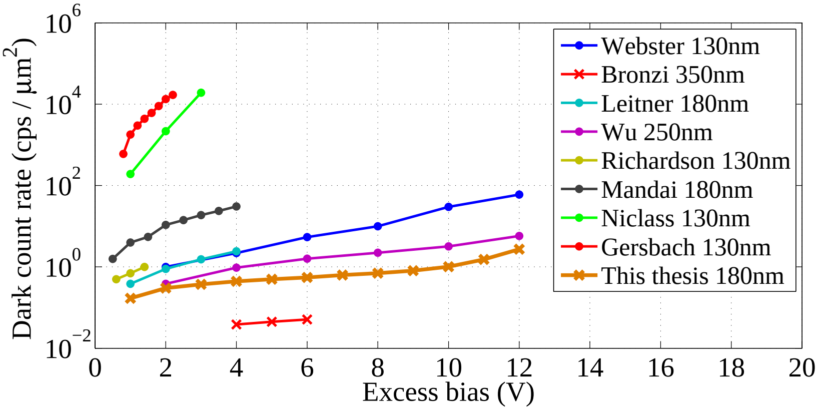


Figure 1.5: This thesis vs state-of-the-art CMOS SPAD dark noise. Note: dark noise for SPAD is reported as the dark count rate, for more details refer Chapter 2.

Further in this thesis, a SPAD designed using p-i-n diode based configuration achieved lower noise than the state-of-the-art CMOS SPAD operated at or above 10V excess bias. Figure 1.5 compares the state-of-the-art devices with the p-i-n device presented in this work.

1.4.3 Fill factor enhancement

In this thesis two different SPAD designs were proposed to reduce the insensitive area. The first design, emulates the optical microlens, where the SPAD is designed to operate at or near its guard junction breakdown. Under such operating conditions some of the photocarriers created in the guard junction is sensed by the main junction through the lateral avalanche propagation. Also since the presented design can host transistors on the top of the guard region, without requiring the need for the device periphery, we believe the presented design can lead to the next generation pixel arrays with high density. The second design extends the first, where both the main junction and the guard region were operated above the breakdown. The presented design resulted in 22.2% improvement in fill-factor when compared to the conventional designs.

1.4.4 Networking technique

For data acquisition, when using SPADs in PET system, a sensor network based approach is proposed in this thesis. In the proposed approach, the data reduction is achieved by performing data processing and noise reduction using the network. The proposed approach was shown to be capable of performing real time data acquisition, when using SPADs in all three PET modalities - preclinical, brain and clinical systems.

1.5 Organization

This thesis is organized as follows:

- *Chapter 2* presents the SPAD characterization techniques and a metrological approach used in this thesis.

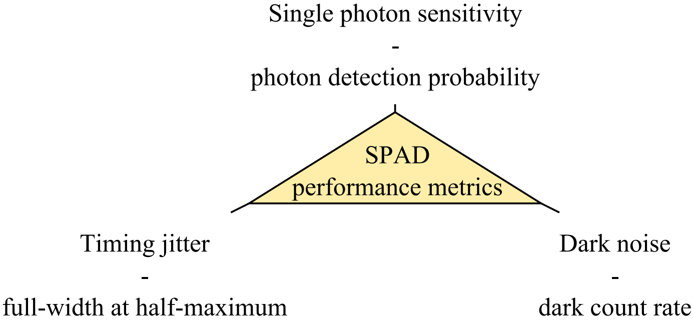
- *Chapter 3* presents active area design techniques for substrate isolated CMOS SPADs.

- *Chapter 4* presents guard region engineering and periphery design techniques for CMOS SPADs.
- *Chapter 5* presents techniques to improve CMOS SPAD fill factor.
- *Chapter 6* presents the sensor network based data acquisition system when using SPADs in PET application.
- *Chapter 7* summarizes the thesis by listing the achievements and by proposing the future work.

As a general note, the SPAD designs reported in this thesis were all fabricated in 180nm CMOS technology - in three different tapeouts. Since we observed some minor process variations from one run to the other, readers are requested not to compare the performance of one device with the others, unless otherwise stated explicitly.

Chapter 2

Metrology


Single photon sensitivity
-
photon detection probability
SPAD
performance metrics
Timing jitter
-
full-width at half-maximum
Dark noise
-
dark count rate

This chapter presents SPAD characterization techniques used throughout this thesis. The techniques presented in this chapter are based on the literature and from the experience garnered during the course of this thesis work. For readers, to understand the characterization procedures and to comprehend measurement results, the physics behind the SPAD operation is presented. In addition, the influence of the experimental setup on the measurement results are also discussed.

SPAD's are characterized in terms of dark noise, single-photon sensitivity and timing jitter. Generally, dark noise is measured as the dark count rate (DCR), single-photon sensitivity as the photon detection probability (PDP), and timing jitter as the full-width at half-maximum (FWHM) of the statistical distribution in single-photon detection time. In addition to the above stated performance parameters, when characterizing a SPAD, its breakdown voltage is also required to be measured.

Chapter organization: Section 2.1 presents the passive and active techniques used to perform the quench and recharge operations. Section 2.2, present diode breakdown voltage measurement techniques. Section 2.3, 2.4 and 2.5 presents the techniques to characterize dark noise, single-photon sensitivity and timing jitter. Section 2.6, presents the experimental setup used in this thesis for the SPAD characterization. Section 2.7, summarizes the chapter.

2.1 Quench and recharge techniques

Although SPADs enable single-photon sensitivity with picosecond timing accuracy, it is not a self-contained device. For every instance when the diode enters into breakdown, the avalanche must be quenched to avoid device from overheating. Generally, external circuitry is used to perform the quench operation. The quench circuitry on detection of a high current, lowers the diode voltage to near breakdown - the condition at which the avalanche is quenched. Further, to aid detect subsequent photons, the diode is recharged back to its original operating voltage by a recharge circuitry. During the quench and recharge time the SPAD is considered, to a first approximation, insensitive to photons. Hence, this time period is referred to as the dead time.

As will be explained in Section 2.3, 2.4 and 2.5, the SPAD measurement results also depend on the quench and the recharge circuitry used. This section provides an insight into the working principles of different quench and recharge techniques. The techniques reported in literature [71] are broadly classified into two categories namely: passive and active.

2.1.1 Passive technique

In passive technique, a resistor greater than 100kΩ is placed in series with the diode (Figure 2.1a). In this configuration, the avalanche current induces a voltage drop across the resistor [72]. As a result, the diode bias voltage is lowered to near the breakdown the condition at which the avalanche is quenched.

To understand the influence of this technique on the measurement results, and to select appropriate resistance values, a detailed understanding on its operation is required.

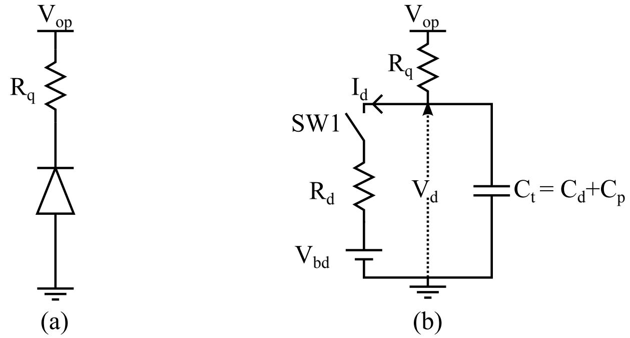


Figure 2.1: (a) SPAD with a quenching resistor, (b) equivalent SPAD model.

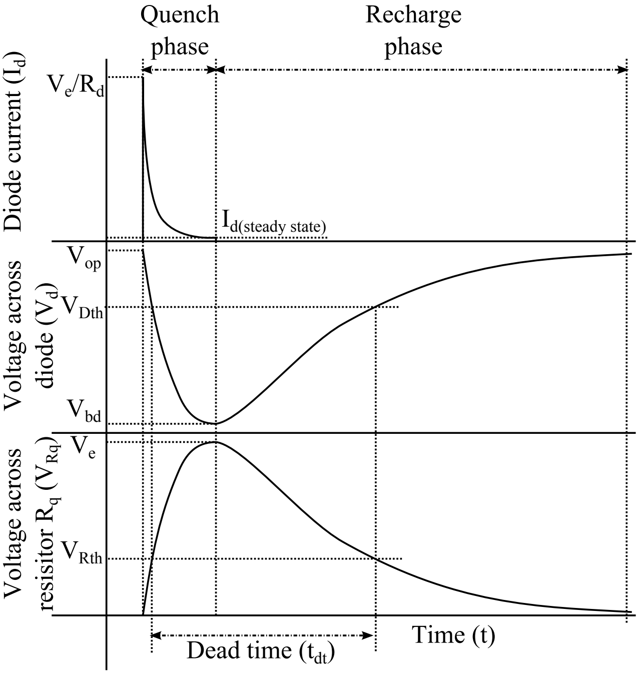


Figure 2.2: Timing diagram when using passive quench and recharge operation.

2.1.1.1 Quench operation

Figure 2.1b presents a simplified circuit model of a SPAD, along with its quench resistor (R_q). In this model, the resistor (R_d) represents the diode space charge resistance and the capacitor (C_t) represents the total capacitance seen across the diode (C_t includes diode junction capacitance - C_d and the parasitic capacitance- C_p, which is introduced by the experimental setup and the readout). Note: depending on the diode geometry R_d can vary from hundreds of ohms to few kilo ohms. Under idle conditions, the switch SW1 is open, the diode current (I_d) is zero and the voltage across the diode (V_d) is V_{op}. On the onset of the avalanche, the switch (SW1) is closed. Assuming the avalanche build-up statistics is negligible, the diode current increases instantaneously to V_e/R_d ($V_e = V_{op} - V_{bd}$), while the diode voltage still remains at V_{op} - as the capacitance C_t resists any instantaneous changes to the voltage (Figure 2.2). In this configuration, the generated avalanche current discharges the capacitor C_t over time. As a result the diode voltage and the current reduces exponentially following the time constant (t_q) as defined in Equation 2.1.

$$t_q = C_t \times \frac{R_d.R_q}{R_d + R_q} \approx C_t \times R_d(sinceR_d \ll R_q) \tag{2.1}$$

In steady state, when the capacitance is fully discharged the diode current and voltage reaches a constant value as described by the following equations.

$$I_{d(steadystate)} = \frac{V_{op} - V_{bd}}{R_q + R_d} \approx \frac{V_e}{R_q} \tag{2.2}$$

$$V_{d(steadystate)} = V_{bd} + I_{d(steadystate)}R_d \tag{2.3}$$

The reduction in V_d and I_d lowers the carrier ionization coefficients and the number of available carriers to sustain an avalanche. As a consequence the chance for an avalanche to get quenched increases. In [71], it was stated that when the diode steady state current is below 100μA the avalanche is in self-quenching phase. The statistics involved in carrier multiplication during the quench phase introduces a jitter in quench time. In practice, the quench time jitter is lowered by increasing the resistance of R_q [71]. At higher values of R_q, the diode's steady state current is well below 100μA, when the chance for an avalanche to be quenched is higher.

2.1.1.2 Recharge operation

Once the avalanche is successfully quenched, the SPAD is recharged back to its original operating voltage (V_{op}) via the quench resistor (R_q). In this technique, the diode voltage is recharged exponentially following the time constant defined as $R_q \times C_t$ (Figure 2.2). For recharge, the resistance of R_q has to be low to reduce the recharge time. As this requirement is in contrast to the quench phase, one has to judiciously choose the value for R_q such that the diode is quenched and also the recharge time is optimized. In practice, when using the passive technique, the recharge time is generally more than 10 times higher than the quench time. Further, it needs to be noted that, during the recharge phase, although the SPAD is not fully charged to its final voltage it is still biased above breakdown. In principle, SPADs can trigger an avalanche during the recharge time, but with lower probability than when fully recharged. This specific property does influence some of the SPAD characterization results, which will be discussed later in this chapter.

2.1.1.3 Avalanche detection

For avalanche detection, generally the voltage pulse formed across the resistor/SPAD is used (Figure 2.2). In practice, the avalanche is assumed to be detected when the voltage across the resistor/SPAD crosses a certain set threshold during the quench phase (Figure 2.2).

2.1.1.4 Dead time measurement

The dead time is measured as the time difference, at which the voltage across the resistor/SPAD, crosses the set threshold during the quench and the recharge phase (Figure 2.2).

2.1.2 Active technique

In the active approach the quench and recharge operations are performed by connecting the SPAD directly (or through a low resistive path) to V_{bd} (for quenching) and to V_{op} (for rechargeing) in real time. The main advantage of this approach is the controlled quench and recharge time. However, when compared to the passive technique, this approach requires more complex circuitry.

2.1.2.1 Quench operation

Figure 2.3 presents a simplified schematic. In this scheme, on successful detection of an avalanche, the diode is quenched by closing the switch SW2. During the quench phase, capacitance C_t is discharged rapidly following a time constant $R_{SW2} \times C_t$, where R_{SW2} represents the on-state resistance of the switch SW2.

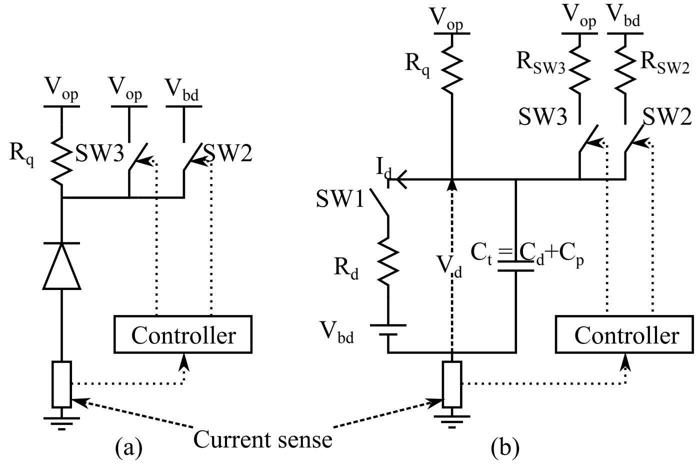


Figure 2.3: (a) SPAD with active quench and recharge circuitry, (b) equivalent SPAD model.

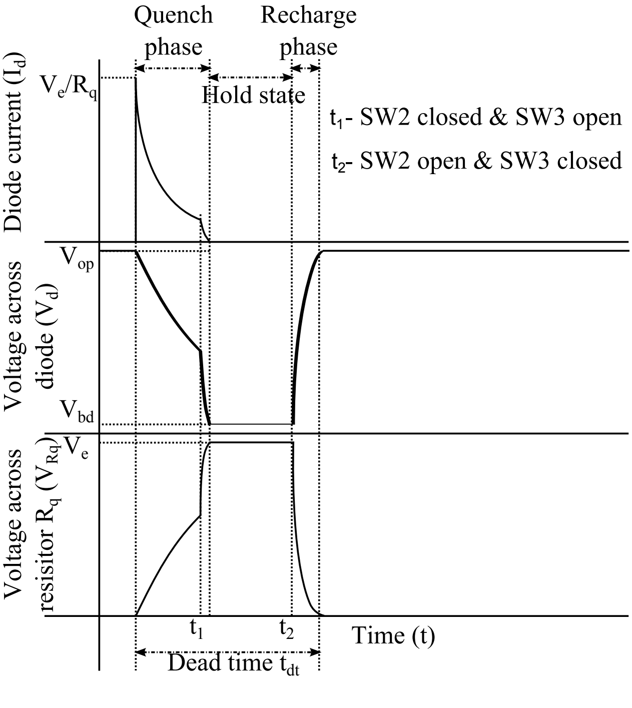


Figure 2.4: Timing diagram when using active quench and recharge operation.

In steady state when the capacitance is fully discharged, the diode voltage is very close to V_{bd} as the resistance R_{SW2} is very small and the diode current I_d is negligible. The low I_d and V_d results in negligible quench time jitter. Further, when using this technique it is feasible to hold-on the SPAD, in the quench state for a definite period of time, by configuring the controller accordingly. This feature is useful in certain measurements, where it is required to extend the dead time (Section 2.3, 2.4 and 2.5).

2.1.2.2 Recharge operation

For recharge, the switch SW2 is opened and the switch SW3 is closed in succession. In this configuration the diode is recharged following a time constant as defined by $R_{SW3} \times C_t$, where R_{SW3} represents the on state resistance of switch SW3. After recharge, the switch SW3 is opened, to let the SPAD detect the next photon. Also in this technique, during the recharge phase the SPAD is biased above its breakdown. Hence, there is a certain chance that the SPAD can trigger an avalanche during the recharge time. In case, if the SPAD triggers, the avalanche will not be quenched until the recharge phase is completed. This could lead to two situations, one - the SPAD gets physically damaged, and two - the SPAD fires as soon as it comes out of the recharge phase. Chances of a SPAD getting damaged is very low as the recharge time is generally very small. Hence, when using active technique for measurements, one can expect the SPAD to fire as soon as it comes out of the recharge phase. This could add some level of uncertainty in some measurements, which will be discussed later in Section 2.3.

2.1.2.3 Avalanche detection

For avalanche detection, generally the output of the current sense circuit is used. When it not accessible, as with passive quenching, the voltage across the diode or the resistor R_q can also be used with a set threshold.

2.1.2.4 Dead time measurement

In the active technique, during the quench and the recharge operation for all practical purposes one can assume that the SPAD is insensitive to photons. Hence, when using this technique, the time period starting from the initiation of the avalanche to the end of the recharge phase is considered dead time (Figure 2.4).

Note: in this section we discussed the quench and the recharge operation applied to the cathode, in principle it can also be applied to the anode.

2.2 Breakdown voltage

One of the first measurements that needs to be performed on a given SPAD is the breakdown voltage measurement. This section discusses four different techniques to measure/estimate the SPAD breakdown voltage.

2.2.1 I-V measurement

I-V Measurement is a straight-forward technique, where the voltage across the diode is swept in reverse bias mode while measuring the current through the diode. The voltage at which the current increases by more than an order of magnitude or higher is generally referred to as the breakdown voltage. However, since in SPADs the breakdown occurs due to the avalanche, there could be some glitches in the I-V measurements. Glitches occur either due to the statistics involved in the creation of a primary carrier or due to a statistical fluctuation in the avalanche build-up. The glitches, in principle, can be thwarted either by increasing the integration time for every measurement point and/or by illuminating the SPAD with a very small amount of light.

2.2.2 Light emission test (LET)

In this measurement, the diode is reverse biased through a current limiting circuit. The bias voltage at which the diode starts emitting light corresponds to the breakdown voltage, and the spatial position from where the light is emitted represents the region of the junction under breakdown. The measurement error introduced by this technique depends heavily on the sensitivity of the microscope and the size of region under breakdown. Further, when using this technique one needs to take care of the series resistance introduced internally in the SPAD and also from the external setup. As the current sourced by the diode at the breakdown could be as high as 1mA, possible IR drop could influence the diode bias voltage. Though LET is a crude technique for breakdown measurements, though it is the only technique that can provide the spatial information on the region under breakdown. This information is quite useful when testing the diodes for the first time.

2.2.3 Sweep and subtract method

Though the I-V measurement is a reliable technique for breakdown measurements, it requires direct access to the anode and the cathode. For instance, when required to measure the breakdown voltage across an array of SPADs, or with an integrated quench and recharge circuitry, we need to resort to other techniques. One such technique is the sweep and subtract method [73]. In this method, when using the quench and recharge circuitry, the SPAD bias voltage (V_{op}) is swept slowly until we start observing avalanche pulses at the output. The voltage at which the avalanche pulse starts to appear at the SPAD output is measured as the breakdown voltage. In situations when the SPAD is coupled to a comparator or to an inverter, the diode breakdown voltage is obtained by subtracting the comparator/inverters threshold voltage from the diode bias voltage. However, the sweep and subtract method is not an accurate measurement technique where we heavily depend on the observation time window, configuration of the circuitry and in some instances on the oscilloscope specifications.

2.2.4 Fit-to-SPAD count rate method

Another technique to determine the diode breakdown voltage with integrated electronics is the fit-to-SPAD count rate method [73]. In this technique, the SPAD is operated in Geiger mode and the avalanche count rate is measured at different bias voltages. A plot made with the diode bias voltage - in x-axis and the count rate - in y-axis, is fit with a straight line. The x-intercept of the fit represents the diode breakdown voltage.

The accuracy of this method relies on an assumption that the count rate is a linear function of the diode bias voltage. For this assumption to be true, it is advisable to perform the measurements with the SPAD being illuminated with light, so that the exponential dependence of the dark noise with the voltage can be avoided [73]. Also, a possible saturation in count rate should also be take into consideration when using this technique. Figure 2.5 presents the impact on breakdown voltage measurements when performing the experiments in linear, dark and in saturated conditions.

2.3 Dark noise

In Geiger mode, one of the first characterizations that needs to be performed is the dark noise. Dark noise represents spurious avalanche pulses that are triggered in

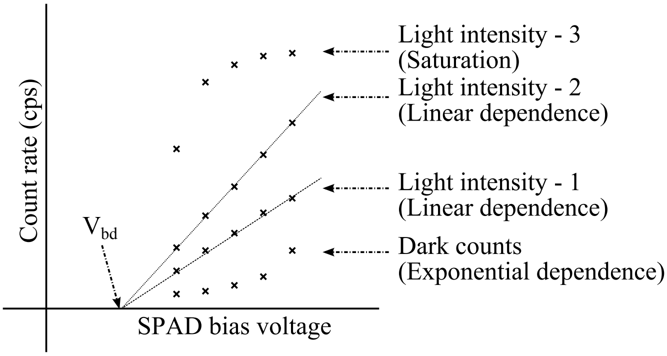


Figure 2.5: Breakdown voltage measurement using the fit-to-SPAD count rate method.

dark conditions, where there are no photons. Generally, dark noise is measured as the dark count rate (DCR), where the DCR is defined as the average count of SPAD output pulses observed under dark condition in one second.

2.3.1 Dark count rate

DCR is composed of two components, namely the primary and secondary pulses [71]. Primary pulses (or Primary DCR) are the random avalanche triggers attributed to the natural process of carrier generation, which could be either due to the thermal generation or band-to-band tunneling, or a trap assisted processes or a combination of these processes. Secondary pulses (also known as afterpulses) on the contrary, are correlated to primary pulses in time, and are due to trapping and de-trapping of a carrier created during previous avalanches.

DCR is characterized in three stages. In the first stage, the total DCR including primary and secondary pulses is measured. In the second stage, secondary pulses are characterized. Finally, in the third stage the primary pulses are characterized in isolation. This section presents the DCR measurement technique. Secondary and primary pulse characterization is discussed in the Section 2.3.2 and in Section 2.3.3 respectively.

2.3.1.1 Measurement technique

Dark noise including primary and secondary pulses is measured by counting SPAD output pulses, under dark conditions, for a time period (t_p). DCR is then calculated by normalizing the measured counts to one second.

2.3.1.2 Discussion

DCR measurement results for a given SPAD at certain excess bias voltage and temperature depends on the following parameters:

1. Integration time (t_p) - Shot noise associated with counting is dependent on the integration time. For DCR measurements or for any measurements that involves counting SPAD output pulses, integration time (t_p) should be made as high as possible to alleviate the shot noise effect.

2. SPAD dead time - As will be explained in the Section 2.3.2, secondary pulse or afterpulsing contribution towards the DCR increases with the reduction in SPAD dead time. Hence, to obtain a worst case DCR result, one needs to choose the lowest possible dead time for the SPAD.

3. Quench and recharge technique - When using active techniques for the quench and the recharge operation, the SPAD cannot trigger an avalanche during dead time (Quench+hold+recharge time). This limits the measured count rate to $1/dead\ time$. The effect of dead time on the count rate measurement can be estimated using the model [74] presented in Equation 2.4.

$$m = \frac{n}{1 + nt_{dt}} \tag{2.4}$$

m - Measured count rate

n - Actual count rate

t_{dt} - SPAD dead time

In passive techniques, the situation is more complicated, because the dead time is not fixed or well defined. In situations when the SPAD fires before it gets recharged to its set threshold voltage for detection, the resulting avalanche could go undetected, and as a consequence the dead time increases (Figure 2.6). The impact on measured count rate with varying dead time is modeled [75] in Equation 2.5. Though the presented model was originally built for Geiger Müller counter, while in [74] and [76] the authors have successfully adapted it for the SPAD and have shown the measurement result to match the estimation.

$$m = \frac{nexp(-nt_{r1})}{1 + nt_q} \tag{2.5}$$

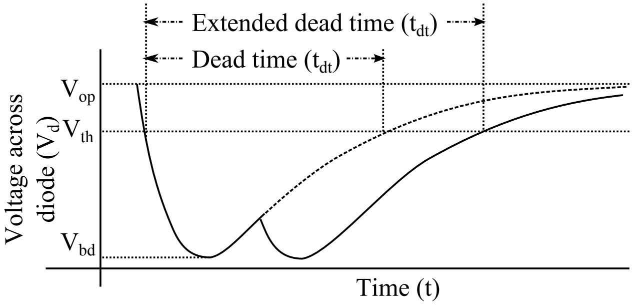


Figure 2.6: Extended dead time effect when using passive techniques for quench and recharge operations.

m - Measured count rate

n - Actual count rate

t_q - Quench time

t_{r1} - Time difference between the start of the recharge phase and the time at which the SPAD bias voltage reaches the set threshold.

Figure 2.7 presents the simulation results when using the active and passive technique for the quench and the recharge operation. The simulations were performed using the model presented in Equation 2.4 and 2.5. The results suggest that the measured count rate when using active techniques will follow the actual count rate to a longer range than with the use of passive techniques. Hence, when given a choice one should choose active techniques for the quench and the recharge operation, as it is relatively less prone to measurement errors.

Further, one needs to be aware that when the actual count rate is higher than $1/dead\ time$, the count rate saturates in case of active techniques, whereas with the passive techniques the count rate drops.

Since the DCR measurement results depend on the SPAD dead time and the choice of the quench and the recharge circuitry, it is important to report the used measurement setup and the dead time configuration when presenting the DCR results.

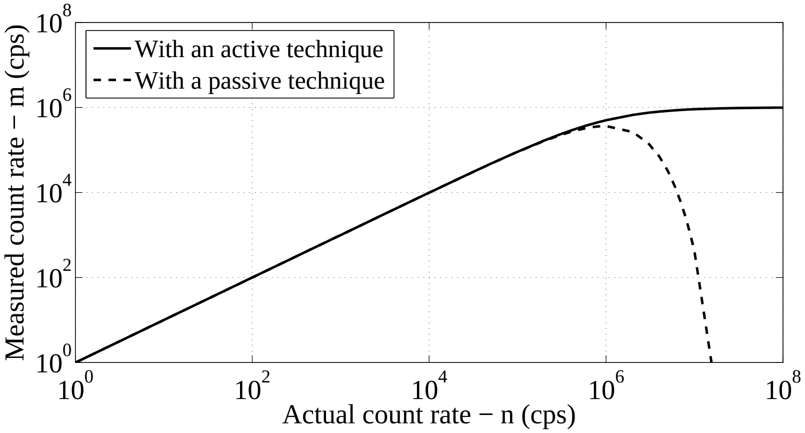


Figure 2.7: Simulation result performed using the model presented in Equation 2.4 and 2.5. For simulations t_{dt} was 1μs, t_q was 10ns and t_{r1} was 1μs.

2.3.2 Secondary pulses

In SPADs during an avalanche a carrier can get captured in a trap. Some of the traps in silicon, hold the carrier for a certain time period before releasing them. In situations when a trapped carrier is released after the SPAD is been brought out of the quench phase, there is a certain probability that the released carrier can trigger a consecutive avalanche. The SPAD output pulses generated by this process are referred to as secondary pulses or afterpulses. Secondary pulses are correlated in time with primary pulses, and the correlation time depends on the life time of the carrier in the trap. In practice, a small fraction of the SPAD output pulses are due to afterpulsing, it is measured as the afterpulsing probability.

Although, afterpulsing is presented under the dark noise section, one needs to be aware that the secondary pulses could also result due to an avalanche initiated by a photon.

2.3.2.1 Measurement technique

Afterpulsing probability is measured using the inter-avalanche time histogram technique [73]. The technique was originally proposed in [77] as time correlated carrier counting technique. In this technique, a histogram is built using the time periods measured between two consecutive avalanches (Figure 2.8). In situations when the DCR is composed of only primary pulses, the poissonian nature of the primary pulse will result in an exponential distribution. However, when secondary pulses are included, the measured histogram will have a multi-exponential behavior [77]. Since in silicon, the carrier life time in a trap, is of the order of a few nano seconds to micro seconds [77], one can safely assume that the exponential distribution observed after 20μs of the inter-avalanche time represents the primary pulses. An exponential curve fitted to the histogram data points measured above 20μs of inter-avalanche time will represent the primary pulse. The area under the fitted exponential curve represents the primary pulse count. The secondary pulse count is then obtained by subtracting the primary pulse count from the total measured avalanche pulses. Afterpulsing probability is calculated using Equation 2.6.

$$APP = \frac{Secondary\ pulse\ count}{Total\ avalanche\ count} \tag{2.6}$$

APP - Afterpulsing probability

Also, from measured data, it is feasible to estimate afterpulsing probability for SPAD dead times higher than these used for the measurements. Estimation is performed also using Equation 2.6, but considering data points only starting from the inter-avalanche time corresponding to the dead time of interest. An example is shown in Figure 2.8 - as one could see the afterpulsing probability reduces with increase in dead time. For the presented example, afterpulsing becomes negligible at around t_{ap}, implying that by configuring the device above t_{ap} dead time, the measured counts will be free of secondary pulses. This specific property is exploited in the next subsection to characterize primary pulses in isolation.

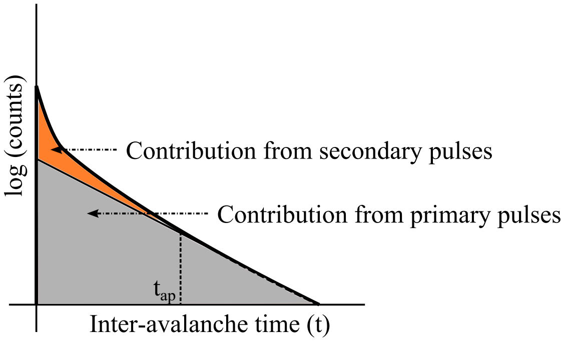

log (counts)
Contribution from secondary pulses
Contribution from primary pulses
t_ap
Inter-avalanche time (t)

Figure 2.8: Inter-avalanche time histogram highlighting contributions from primary and secondary pulses.

2.3.2.2 Discussion

For a given SPAD, the afterpulsing probability result depends on the following:

1. SPAD dead time - A carrier released from a trap during the SPAD dead time cannot trigger an avalanche or cannot result in a detectable secondary pulse. Hence to characterize the worst case probability, it is advisable to choose the minimum possible dead time so that the secondary pulses originating from the carriers released earlier in time can also be included in the measurement results.

2. Primary pulse count rate - The technique presented in this section needs not be performed under dark conditions, it can also be carried out under low light conditions. When using light one needs to assure that the light intensity does not change during the course of the measurements, and also its intensity is low enough to not suppress the afterpulsing effects. Afterpulsing could go undetected when the photon rate or the primary DCR rate is higher than $1/carrier\ life\ time\ in\ a\ trap$. In practice, before measurements it is not feasible to know the properties of the trap. One possible way to assure that afterpulsing is not suppressed is to perform afterpulsing measurements with two different light conditions.

3. Parasitic capacitance - Afterpulsing depends on the probability of a trap to capture a carrier. It depends on the number of carriers that flow through the diode during the quench phase, when the capacitor C_t is discharged. The number of carriers discharged by the capacitor C_t is given by N_q which is proportional to $V_e \times C_t$. Since C_t also contains parasitic capacitance C_p introduced by the experimental setup, it is vital to reduce the parasitic capacitance to reduce the loading effect. However, if the SPAD is characterized to study the defects or for process optimization purposes it is not a bad idea to increase C_p on purpose.

4. Excess bias voltage - Since the number of carriers discharged by the capacitor is proportional to V_e for a given C_t. One can expect an increase in afterpulsing probability with increase of excess bias voltage.

5. Quench and recharge techniques - In the quench operation when using passive techniques, the capacitance C_t discharges only through the diode. Whereas, in active techniques only a small fraction of the capacitance C_t is discharged through the diode and the remaining charge flows through the low resistance

path (through SW2 in Figure 2.3). As the charge flow through the diode is lower in active quenching one can also expect a lower afterpulsing probability.

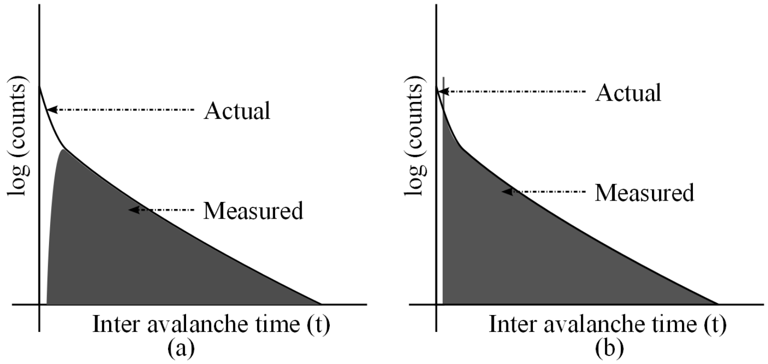


Figure 2.9: (a) afterpulsing measurement when using passive recharge. (b) afterpulsing measurement when using active recharge.

In the recharge operation when using passive techniques, after the quench operation the diode voltage is progressively recharged, when the probability for the SPAD to fire also increases progressively. This property could lead to a deformation in the inter-avalanche time histogram (Figure 2.9a). This deformation in principle can suppress some of the afterpulsing effects. Thus, when using passive recharge, one could overestimate the SPAD afterpulsing performance. Whereas in active technique, the recharge time is very fast and also the SPAD cannot fire during the recharge phase, these properties avoid the deformation in inter-avalanche time histogram (Figure 2.9b). However, if an avalanche is triggered during the recharge time, the SPAD can fire immediately as soon as it comes out of dead time. This in principle can result in an unexpected increase of counts of the first bin in the inter-avalanche time histogram (Figure 2.9b). As it is not possible to distinguish if the counts observed in the first bin is due to afterpulsing or from other influences, one needs to also consider it as a part of afterpulsing. Generally this effect is negligible. Hence, an ideal choice for afterpulsing measurement would be to use active quench and recharge.

Note: the inter-avalanche time histogram technique presented in this section could also be used to characterize the crosstalk between SPADs.

2.3.3 Primary pulses

Primary pulses are due to carriers generated by natural process - band to band tunneling, traps or a combination of them.

2.3.3.1 Measurement technique

To isolate primary pulses from secondary pulses, the SPAD dead time is set such that the afterpulsing becomes negligible. The primary pulses are then counted under dark conditions for a certain time period (t_p). The measured results are then normalized to yield the counts per second. As like in DCR measurement, it is vital to have a larger integration time t_p to reduce the shot noise in counting results. To alleviate afterpulsing generally a dead time bigger than 1μs is required to be set, in such situations to negate any ambiguity on the measurement results due to the dead time configuration, dead time compensation needs to be applied to the results using Equation 2.7 .

$$DTCR = \frac{MCR}{1 - (MCR \times t_{dt})} \tag{2.7}$$

$DTCR$ - Dead time compensated count rate
MCR - Measured count rate
t_{dt} - SPAD dead time

2.3.3.2 Discussion

Primary pulses are characterized to comprehend the major source of dark noise. Although there is no accurate technique available to find the source of dark noise, it is possible to study if the major source of noise is due to a tunneling process or to a Shockley Read Hall (SRH) process (traps). For this study it is required to measure the primary pulse count rate at various temperatures and excess biases.

In [72,78], it was stated that the dark noise generated by tunneling processes, is almost insensitive to temperature variations, whereas the noise that originates from the SRH process increases with temperature. On the other hand the tunneling noise is highly dependent on bias voltage when compared to the noise generated by the SRH process.

For a detailed analysis, in [79], a dark count rate spectroscopy technique was proposed. In this technique, the SPAD activation energy is estimated using the DCR results obtained at various temperatures [79]. The statistical data collected from other similarly designed SPADs in the same die is used to study the source of noise from the spread/distribution of the activation energy and by calculating the capture cross section.

When performing primary pulse measurements one needs to take care of the following:

1. Dead time - To characterize primary pulses in isolation it is vital to choose the SPAD dead time, such that afterpulsing becomes negligible. Hence, before performing this measurement, afterpulsing probability needs to be characterized.

2. Quench and recharge technique - As the primary pulse characterization also requires the counting of SPAD output pulses, it is preferable to use active quenching and recharge (Section 2.3.1). Active techniques when compared to passive ones introduce less error on count rate measurements, and more importantly the dead time is well defined.

2.4 Single-photon sensitivity

For SPADs, single-photon sensitivity is measured as the photon detection probability (PDP). As the name suggests, PDP represents the probability that a photon incident on the active area gets detected.

2.4.1 Measurement technique

PDP is measured by comparing the SPAD's sensitivity with that of a reference diode. For this measurement, the reference photodiode and the device under characterization are illuminated with a same light intensity of a known wavelength. In most measurement setups, a mono-chromator is used to select the specific wavelength of light from a lamp emitting a wide spectrum of wavelengths. An integrating sphere is used to diffuse and scatter the mono-chromator's light output uniformly across the reference photodiode and the SPAD under measurement. For every measurement the reference diode is used to evaluate the photon count rate incident on the SPAD active area, and the photons detected by the SPAD is measured by finding the difference between the SPAD output pulse rate and its DCR. The SPAD PDP is then evaluated using the formula below.

$$PDP = \frac{MCR - DCR}{PCR} \tag{2.8}$$

MCR - Measured count rate.

PCR - Photon count rate incident on the active area. It is measured using the reference photodiode.

2.4.2 Discussion

Theoretically, PDP is defined as the product of the device quantum efficiency (QE) and the avalanche triggering probability. QE is the ratio between the number of photocarriers that reach or are generated in the depletion region, to the total photons incident on the active area. QE depends on the silicon absorption coefficient - which is a function of wavelength and the SPAD design. The avalanche triggering probability is the probability that a free carrier present in the depletion region can trigger an avalanche. It mainly depends on the excess bias voltage.

For a given SPAD, PDP measurement results depend on the following:

1. Excess bias voltage - Theoretically, the higher the excess bias voltage the higher the avalanche triggering probability. However, due to the saturation in carrier ionization coefficient, at field strengths higher than 5×10^5 V/cm, avalanche triggering probability tend to saturate. Thus for a given SPAD, PDP increase with excess bias voltage and then tends to saturate.

2. Afterpulsing - It is vital to assure that the measured photon count rate is not influenced by afterpulsing. In practice, afterpulsing is removed either by increasing the SPAD dead time or by compensating for afterpulsing on the measured count rate using the formula below.

$$APCR = (1 - APP) \times (MCR) \tag{2.9}$$

   $APCR$ - Afterpulsing compensated count rate

   APP - Afterpulsing probability

3. Dead time - To precisely evaluate the PDP, dead time compensation on the count rate also needs to be performed using the Equation 2.7.

4. Light intensity - When performing PDP characterization one needs to ensure that the measured count rate is not affected due to the saturation resulting from the dead time. Hence it vital to perform the measurements in low light conditions.

5. Quench and recharge technique - As with DCR measurements also for PDP it is preferable to use active quench and recharge technique. Since the PDP

measurement involves a count rate measurement, the discussion presented for dark noise (Section 2.3.1) is also relevant here.

6. SPAD size - Generally it is assumed that the PDP across the drawn active area is uniform. However in reality due to the diffusion of guard region dopants, a relatively lower field strength region created at the periphery of the active area. This effect could lead to variations in PDP depending on the SPAD size.

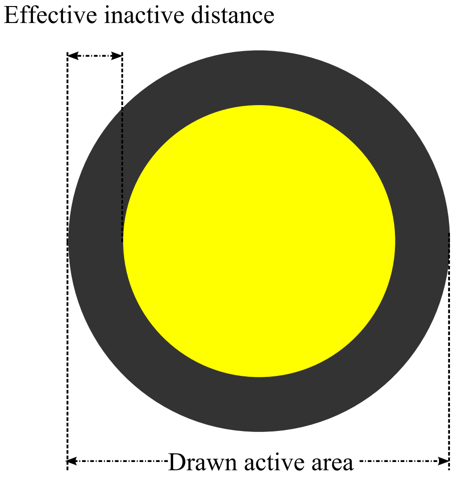


Figure 2.10: A circular SPAD top showing only the drawn active area and effective inactive distance.

In prior work [73] authors have assumed that the region affected by dopant diffusion is insensitive to photons. In this thesis we argue that although the field strength is relatively lower at the periphery, the diffusion affected region can still be sensitive to photons, but with lower PDP. To study the impact of the dopant diffusion on the PDP, we have defined a term *effective inactive distance* in this thesis.

The effective in-active distance is the width of a virtual region placed at the periphery of the drawn active area (Figure 2.10). The virtual region represents the effect of dopant diffusion. In this region the PDP is assumed to be zero. The effective inactive distance is evaluated experimentally by matching the PDP of two circular SPADs designed with different diameters using the Equation 2.11.

$$\frac{pdp_1}{pdp_2} = \frac{\pi(r_1 - d)^2}{\pi(r_2 - d)^2} \times \frac{\pi r_2^2}{\pi r_1^2} \tag{2.10}$$

$$d = r_1 \, \frac{1 - \sqrt{pdp_1/pdp_2}}{1 - (r_1/r_2)\sqrt{pdp_1/pdp_2}} \tag{2.11}$$

d - Effective inactive distance.

pdp_1 - Device-1 PDP calculated considering the drawn active area.

pdp_2 - Device-2 PDP calculated considering the drawn active area.

r_1 - Radius of the drawn active area for device-1.

r_2 - Radius of the drawn active area for device-2.

It needs to be noted that the effective inactive distance is a relative term that could be used to compare the impact of dopant diffusion across different designs, excess bias voltages and wavelengths. Experimental results for a particular SPAD design are presented in Chapter 3.

2.5 Timing jitter

Timing jitter for SPADs represents the fluctuation in photon detection time. Generally, the statistical distribution in photon detection time is composed of two components namely the gaussian and an exponential term. The gaussian component is introduced due to the avalanche build-up dynamics, and the exponential component is from the photocarriers that diffuse to reach the multiplication region.

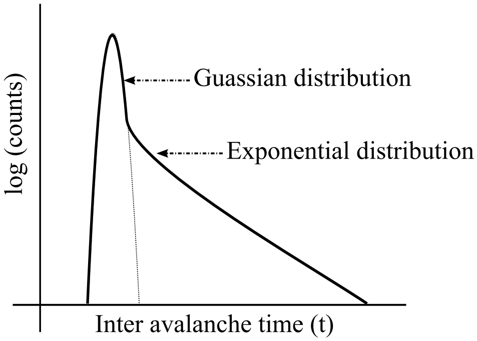


Figure 2.11: Statical distribution of photo-response of the SPAD.

2.5.1 Measurement technique

In practice, timing jitter is measured using a pulsed laser source. For this measurement, the SPAD active area is placed perpendicular to the light path. Under such experimental conditions, the time difference between the photon detection time and the laser firing time is measured. The statistical distribution of the measured time differences represents the SPAD timing jitter. Generally the timing jitter for SPADs is reported in terms of the full-width-at-half-maximum (FWHM) of the statistical distribution of the photo-response of the SPAD.

2.5.2 Discussion

To define the experimental conditions and to understand the implications of various measurement setup parameters on timing jitter, one needs to understand the physics behind the avalanche build-up. In CMOS SPADs, the avalanche builds up in two stages [80]. In the first stage - the avalanche grows locally at the seed point, and in the second stage - the avalanche spreads laterally to the other regions through the multiplication assisted diffusion process. The current resulting from the first stage is influenced the most by the avalanche build up statistics. The resulting current grows during the second stage, following the rate at which the avalanche spreads to the other regions of the active area.

At given bias conditions, the avalanche spread speed depends on the spatial position of the seed point. For instance if an avalanche is initiated at the center of the active area it can then spread in all directions, as the multiplication region is present all around it. Whereas for an avalanche initiated at the periphery, the direction in which the avalanche can spread is limited, as the multiplication region is not all around it. Hence, in addition to the jitter introduced by the avalanche build-up statistics of the first stage, also the statistics involved in creating a seed point spatially across the active area also introduces the jitter.

Until now the discussion was the case in which the seed point for an avalanche is one. In the case in which an avalanche is seeded in multiple points independently, the avalanche current in principle can spread faster when compared to the single seed point case. In such situations the statistical fluctuation in the current raise time also depends on the variance in the number of independent avalanche speed points that are created. In [73], it was shown that in addition to the the Gaussian component of the SPAD timing jitter also the exponential component is impacted by multiple seed points. In practice, multiple seed points are created either when light intensity or when DCR is very high.

For a given SPAD when using a laser emitting a specific wavelength, the timing jitter result depends on the following:

1. Excess bias voltage - The avalanche build-up dynamics is a function of excess bias voltage. The higher the excess bias voltage the lower the statistical fluctuation in avalanche build-up. Hence, when increasing excess bias voltage the timing jitter reduces.

2. Light distribution - As stated earlier, the timing jitter in SPADs could also be introduced from the statistics involved in creating seed point across the active area. Hence to obtain repeatable results it is advisable to illuminate the whole SPAD active area with uniform light intensity.

3. Light intensity - Since SPADs are generally used in photo starved applications, it is vital to perform the timing jitter experiments with the light intensity reduced down to single-photon level. At higher light intensities or when the DCR is very high, measured timing jitter can be worse due to pile-up.

4. Avalanche detection threshold - Avalanche build-up statistics increases with the increase in avalanche current. Hence, to obtain better timing jitter result, it is advisable to set lower threshold in current/voltage for avalanche detection.

5. Parasitic capacitance - Parasitic capacitance introduced by the measurement setup will affect the avalanche current raise time. Hence to reduce the impact of the measurement setup on timing jitter measurement it is vital to reduce the parasitic capacitance.

6. Quench and recharge technique - Also for timing jitter measurements it is preferable to use active quench and recharge. Active quench reduces the impact on measurement result due to the parasitic capacitance and from afterpulsing. Further, since the jitter measurement is sensitive to the bias voltage one should avoid using passive recharge as during passive recharge there is a chance that the SPAD can fire with a lower excess bias voltage. Hence it is preferable to use active recharge for timing jitter measurements.

2.6 Experimental setup

For the discussion presented in Section 2.3, 2.4 and 2.5, it will be evident that for SPAD characterization an ideal choice would be to use active quench and recharge technique. Hence in this thesis, device characterization was performed using active quench and recharge setup, shown in Figure 2.12. In the present configuration, a fast comparator detects an avalanche event; a field programmable gate array (FPGA) directs the tri-state buffer to quench the device for a certain time (programmable) and then to recharge. The programmable quench time feature of this design facilitates the device characterization to be performed with various dead times.

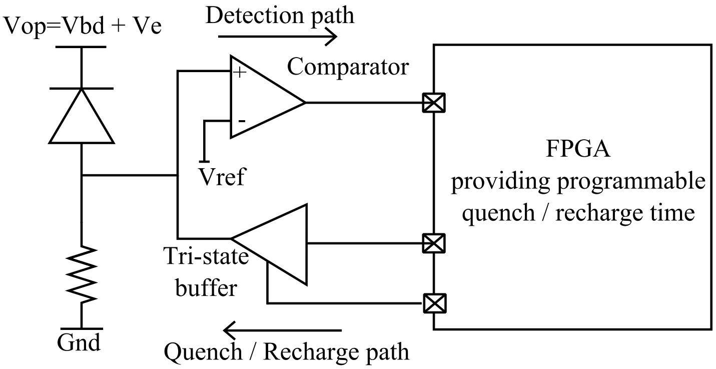


Figure 2.12: Experimental setup used for SPAD characterization.

In this setup, the avalanche is quenched initially using the resistor, which then is taken over by the active quenching circuitry after a loop delay time. The loop delay measured from avalanche ignition to the start of active quenching is around 7ns. Rise and fall time of the level translator along with the parasitic capacitance seen at anode, results in a minimum attainable dead time of 300ns.

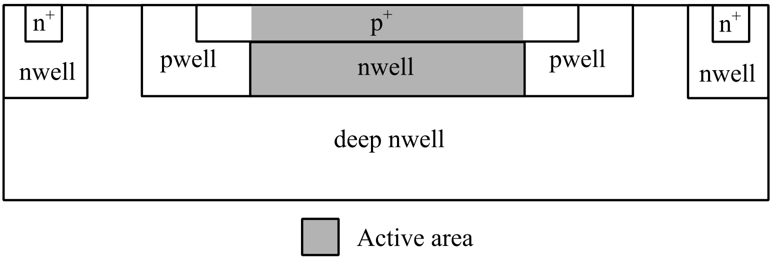
2.7 Summary

In this chapter we discussed the SPAD characterization procedures for dark noise, single-photon sensitivity, and timing jitter. In addition, SPAD breakdown voltage measurement techniques were also presented. A summary of the discussion is presented below.

- SPADs are not self contained devices. It requires the quench and recharge circuitry for operation.
- Measurement results depend on the quench and recharge circuitry used.
- In literature two techniques namely the passive and active techniques were reported for the quench and the recharge operation.
- The passive technique is simple but not very effective for SPAD characterization. Whereas the active technique is complex but very effective for SPAD characterization, due to its tight control on measurement dead time.
- SPAD breakdown voltage measurements can be performed by using any of the four techniques presented in this chapter. I-V measurements are the most reliable techniques of all. LET is a crude technique but provides spatial information on the region under breakdown. The sweep-and-subtract method and the fit-to-SPAD count rate method are two additional techniques that can be used when the SPAD is integrated with electronics.
- Dark noise for SPADs is composed of primary and secondary pulse/afterpulses.
- Dark noise for SPADs is characterized in three stages: first stage is the total DCR characterization, including primary and secondary pulse. Second stage is the secondary pulse characterization. The third stage is the primary pulse characterization.
- Single-photon sensitivity for SPAD is measured as the PDP at various wavelengths, when using a reference photodiode with a known sensitivity.
- For timing jitter, the statistical distribution of the photo-response is measured using a pulsed laser source, when the light intensity falling on the SPAD is adjusted to single-photon level.

Chapter 3

Active area design



This chapter focuses on the active area design for CMOS SPADs. In SPADs, the active area plays a crucial role in determining the device sensitivity, noise and timing performance. Till date, a number of CMOS SPADs have been investigated with the active area designed with the p$^+$/nwell and n$^+$/pwell as main junctions. Though different guard ring structures [30,53,81] were experimented with to avoid edge breakdown, it was found that dark noise was very high, mainly due to band-to-band tunneling and trap-assisted tunneling that results from reduced annealing and drive-in diffusion steps [30]. In certain designs, additional noise has emerged from deep traps in shallow trench isolation [81]. However, a novel design reported in [55], utilizing two lightly doped layers sucha as between pwell and deep nwell, has helped reduce tunneling noise. Further, in [69] the noise was reduced down to 0.05 cps/μm^2 with the use of special enrichment implants. Though the dark noise was reduced, a drawback of these designs is the low sensitivity in green to red wavelengths, where applications like FLIM, FRET, FCS, etc. operate.

In [56], [57] and [58] wide spectral sensitive SPADs were reported, either using substrate as one of the junction nodes or substrate acting as a photon collection region. In either of these designs the depletion region formed with the substrate requires isolation from the transistors and also from the adjacent SPADs to reduce crosstalk. This is an especially serious problem when large arrays of SPADs are required. Hence in this chapter, the design techniques enhancing the spectral response of the conventional designs with substrate isolation was focused upon, as it can reduce crosstalk and can ease circuit integration.

In SPADs, spectral response can be theoretically enhanced by either designing the main junction present in the active area with a wide depletion region and/or by widening the quasi neutral region, from where the photocarriers can be acquired through diffusion process.

In this chapter, two design techniques for the active area are discussed. The first technique, using wide depletion junction for the active area is presented in Section 3.1. The second technique, using a narrower depletion junction but with wider photon-carrier collection region is presented in Section 3.2. The pros and cons of the two techniques are discussed in detail in their corresponding sections. A summary is presented in Section 3.3.

3.1 Wide depletion junction

SPADs designed using wide depletion junction, in principle can result in wider spectral response and can also have lower tunneling contributions towards DCR. However, when using wide depletion junctions, a higher excess bias operation is required to enhance the avalanche triggering probability and to improve timing performance [72]. For SPADs designed in custom technology, where the breakdown is in the order of hundreds of volts, an excess bias voltage greater than 30V is required [82]. This creates the need for attenuation to ensure safe operation the front-end circuits. In CMOS technology, one can reduce the depletion region width, leading to a device breakdown of 40V or less and the excess bias voltage to about 10V. Though a 10V excess bias is still higher than the standard rail-to-rail voltage in deep-sub-micron CMOS circuits, it can be easily attenuated to acceptable values using simple circuits while having a minimal impact to performance, as proposed in [57, 59], thus maintaining full compatibility with large SPAD arrays.

In this section the three SPADs designed with relatively wide depletion junction than the conventional 180nm CMOS SPADs [30, 53, 54, 69] is presented. The devices are designed in CMOS technology using p^+/deep nwell, pwell/deep nwell and pwell/p-epitaxy/buried-n (p-i-n) as its main junction.

3.1.1 p$^+$/deep nwell junction

p$^+$/deep nwell SPAD designed and fabricated in 180nm CMOS technology [3] is presented in this subsection. SPAD achieves wide spectral sensitivity enabling greater than 40% PDP from 440nm to 620nm wavelength at 10V excess bias. For a 12μm active diameter SPAD, the DCR is 17 cps at 2V and 1.45k cps at 10V excess bias, while the afterpulsing probability is less than 0.3% with 300ns dead time at 10V excess bias, and timing jitter is 70ps (FWHM) when using 405nm wavelength laser.

3.1.1.1 Design

The cross section of a circular SPAD designed for a 12μm active area diameter is presented in Figure 3.1.

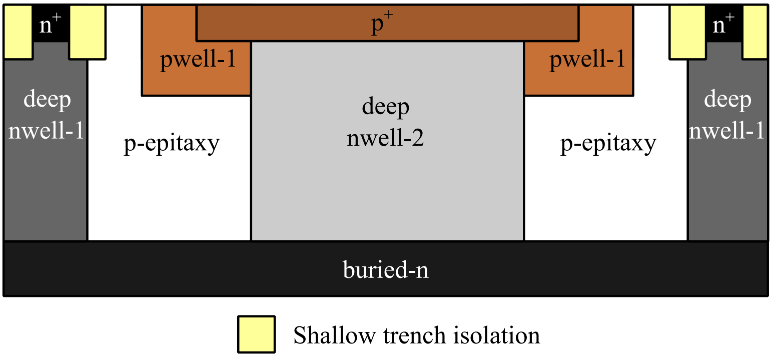


Figure 3.1: p$^+$/deep nwell SPAD: device cross section - fabricated in 180nm CMOS technology.

In this design the main junction was engineered to be between p$^+$ and deep nwell-2. The deep nwell-2, having lower dopant concentration near the junction, helps reduce tunneling noise and also enhances the spectral response with wider depletion than conventional designs. However, for wide depletion devices higher excess bias is generally required to maximize PDP [72]. To enable operation with higher excess bias, in this design, the guard ring was optimized not only to avoid premature edge breakdown but also to be effective for higher bias conditions (Chapter 4). Buried-n enabling substrate isolation provides contact to deep nwell-2, avoiding the need to counter dope the guard ring (pwell-1) with deep nwell as in [30, 69, 83]. The counter-doping was avoided to improve guard region's effectiveness and to provide an p-epitaxy layer around the guard ring, thereby increasing its breakdown voltage.

For the presented design, the breakdown voltage of the guard ring is 35.7V which is 12.2V higher than the main junction. Thus for the presented design, the device can be operated until 12V of excess bias without the guard region entering into breakdown. Note: the guard region breakdown voltage measurements were performed on a test structure designed identically to the reported SPAD but without deep nwell-2.

3.1.1.2 Photon detection probability

The PDP of the device was measured at various bias conditions and for incident light wavelengths from 400nm to 860nm. The results are presented in Figure 3.2a.

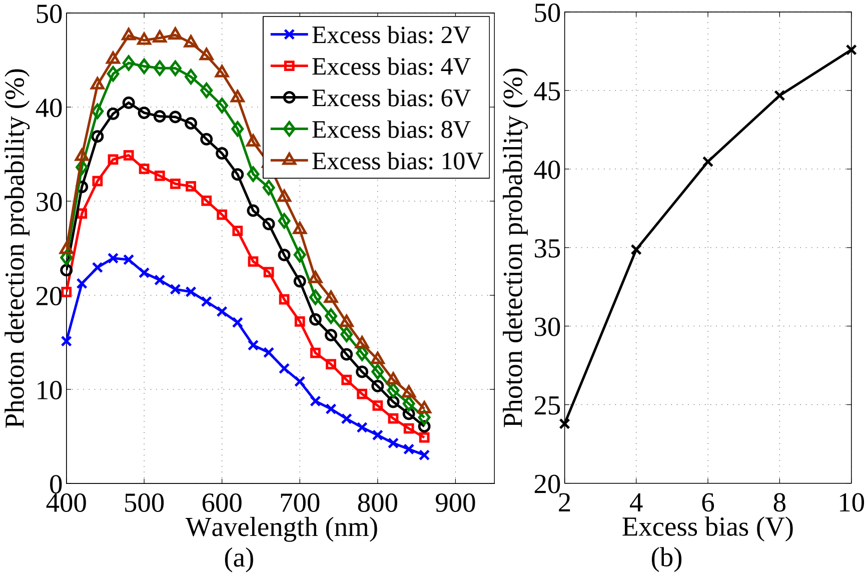


Figure 3.2: p$^+$/deep nwell SPAD: (a) PDP vs wavelength, (b) PDP at 480nm wavelength for various excess bias voltages.

A 12μm active diameter device attains a peak PDP of 47.6% at 480nm when biased at 10V excess bias. Further, the device achieves >40% PDP from 440nm to 620nm, and >30% from 420nm to 680nm at 10V excess bias. The achieved wide PDP profile is attributed to the wide depletion in deep nwell-2 and to the device design, facilitating operation at higher excess bias. Furthermore, for this device, the rate of increase in PDP tends to decrease at higher excess biases (Figure 3.2b), implying that biasing the SPAD over 6V excess bias can result in reducing the impact of breakdown and supply voltage variations on PDP, thus improving PRNU(photo-response non-uniformity) across the array when realized in imagers.

State-of-the-art comparison: among the various substrate isolated devices, The SPADs designed with p$^+$/nwell junction, as reported by Bronzi [69], Leitner [54], Niclass [30] and Gersbach [53], have resulted in narrower PDP profiles due to the formation of a shallower junction. Although, Richardson [55] improved DCR by design using pwell/deep nwell junction, the device spectral response remained identical to conventional designs [30,53,54,84]. A SPAD similar to Richardson's [55], designed in high voltage CMOS process by Wu [85] has resulted in a wider PDP profile due to the use of lightly doped high voltage pwell. However, the peak PDP is $<$25% even at 15V excess bias. The device presented in this sub section attain $>$40% PDP from 440nm to 620nm at 10V excess bias, the achieved performance is superior to the other known substrate isolated CMOS SPADs (Figure 3.3).

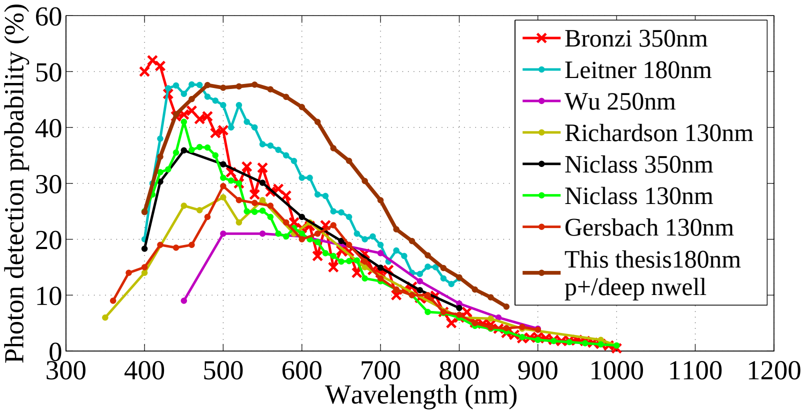


Figure 3.3: p$^+$/deep nwell SPAD: state-of-the-art PDP comparison.

Impact of guard ring's dopant diffusion: for the SPAD design presented in this subsection, the diffusion of the guard ring (pwell-1) dopants into the deep nwell-2 will result in the reduction of the drawn active area. The electric field simulation (Figure 3.4) performed at the device breakdown, supports the expectation with lower field strength in the region affected by the dopant diffusion. However, at higher excess biases the spread in high electric field to the region impacted by the dopant diffusion (Figure 3.4) will aid improve sensitivity.

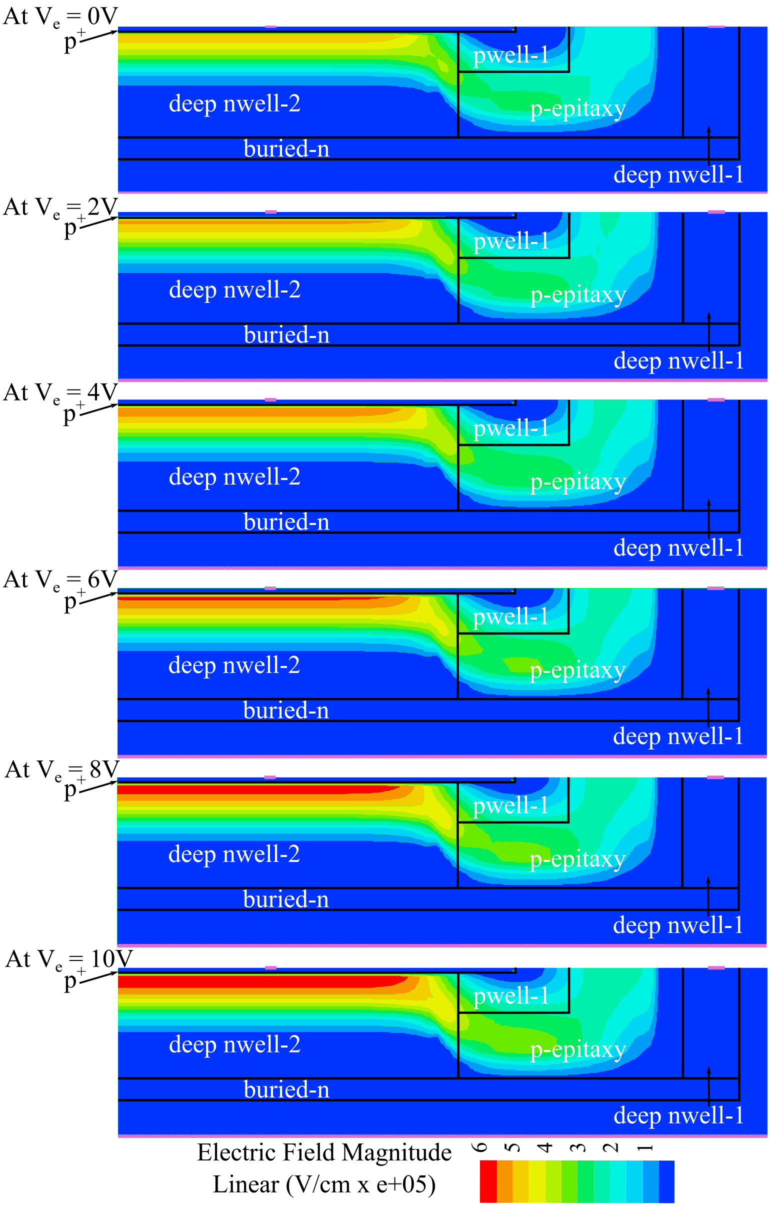


Figure 3.4: p$^+$/deep nwell SPAD: TCAD electric field simulation performed at various excess bias voltages.

To experimentally validate the simulations, the effective inactive distance was estimated using the technique presented in Chapter 2. Results of the effective inactive distance, evaluated using the PDP data from a 12μm (Figure 3.2) and a 6μm active diameter (Figure 3.6) device, are presented in Figure 3.5.

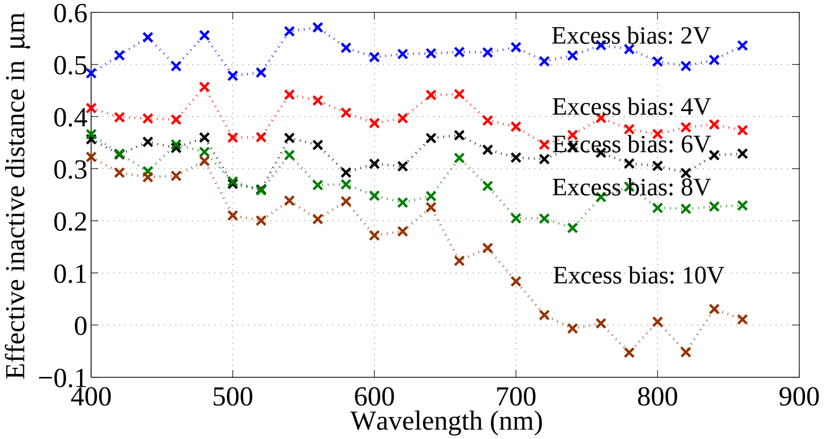


Figure 3.5: p$^+$/deep nwell SPAD: effective inactive disatance estimation.

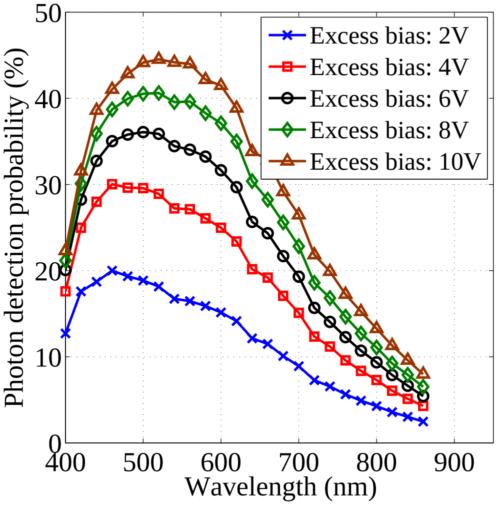


Figure 3.6: p$^+$/deep nwell SPAD: PDP vs wavelength, for a device design with 6μm active diameter.

As expected, the effective inactive distance reduced with the increase in excess bias; and at higher excess biases, the effective inactive distance was found to reduce with the wavelength. At 10V excess bias, the effective inactive distance was estimated to be negative for wavelength above 740nm. This implies that at 10V excess bias low energy photons absorbed deeper into the silicon are acquired from outside of the drawn active area. For this design the migration of the high electric field to beneath the guard ring is the reason for the reduction in the effective inactive distance with the increase in wavelength.

3.1.1.3 Dark count rate

The total DCR including primary and secondary pulses was measured with 300ns dead time. The measurement results for three different dies at 25°C is shown in Figure 3.7. Though a statistical variation in count rate was observed in different devices, no outlier or high DCR SPAD was found. The observed DCR is 0.15 cps/μm^2 at 2V excess bias and 12.84 cps/μm^2 at 10V excess bias.

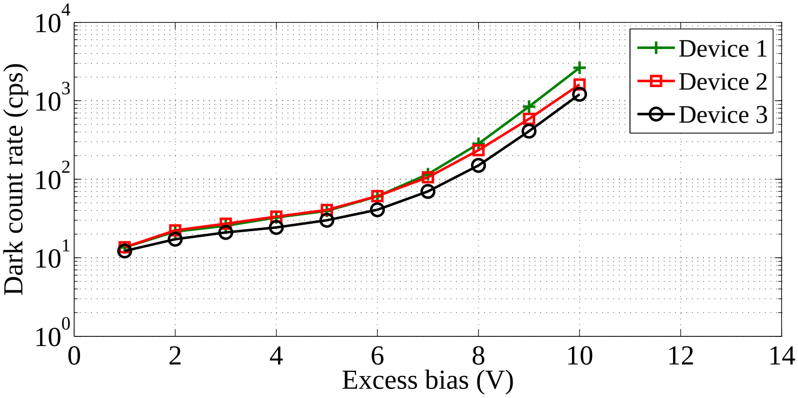


Figure 3.7: p$^+$/deep nwell SPAD: DCR characterization for three devices with the dead time tuned to 300ns.

Secondary pulse / afterpulsing were characterized using the inter-avalanche time histogram method as explained in Chapter 2. In this measurement, avalanche inter-arrival statistics were collected, with the dead time tuned to 300ns when using the experimental setup from Chapter 2. Due to the presence of afterpulses a deviation from the expected exponential distribution is observed (Figure 3.8). Using external quenching and recharge, the afterpulsing probability was measured to be 0.03% at 2V excess bias and 0.3% at 10V. The observed increase in afterpulsing probability with excess bias is due to the increase in charge flow that leads to the increase in carrier trapping probability.

Further, the histogram developed with avalanche inter-arrival time, shown in

Figure 3.8a, reveals that even at 10V excess bias, afterpulsing reduces to a negligible value within 1μs of inter-avalanche time. Hence, configuring the setup with dead time higher than 1μs, it is possible to eliminate afterpulsing from primary pulse characterization.

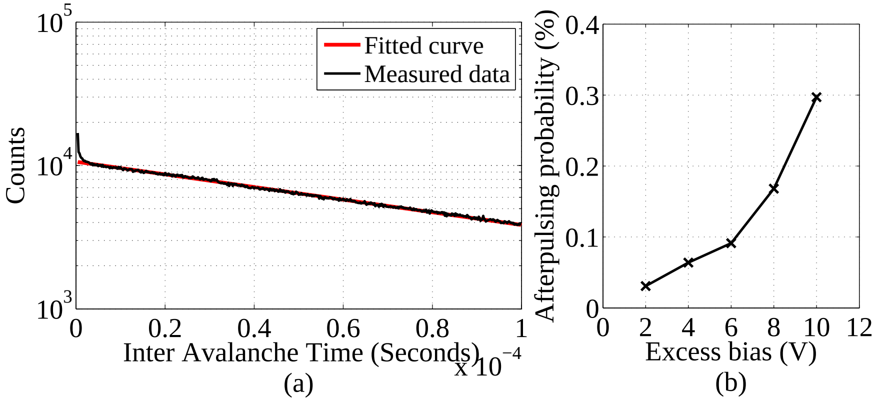


Figure 3.8: p$^+$/deep nwell SPAD: afterpulsing probability measurement results, performed with the device deadtime configured to 300ns, a) afterpulsing experimental results at 10V excess bias, b) afterpulsing probability measured at various excess biases.

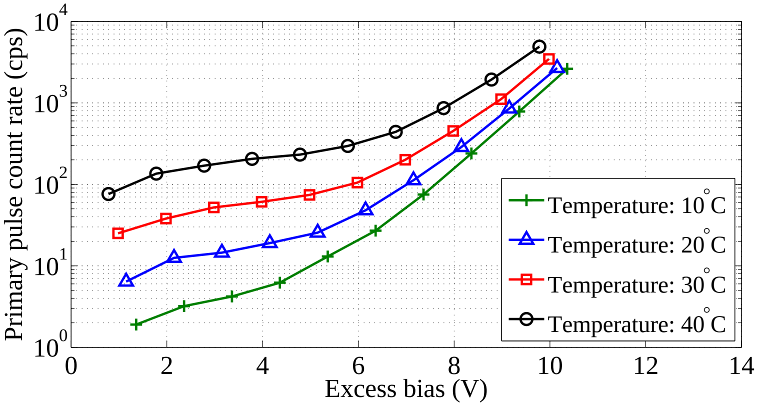


Figure 3.9: p$^+$/deep nwell SPAD: primary DCR characterization performed at various temperatures with the dead time tuned to 10μs.

Primary pulse is characterized in isolation by operating the devices with a dead time of 10μs. The detailed characterization performed for one of the devices at various temperatures (Figure 3.9) suggests an onset and increase in tunneling contribution roughly after 6V excess bias and SRH contribution at low voltages. The reason for the observed tunneling contribution after 6V excess bias is studied in detail in Chapter 4.

State-of-the-art comparison: Figure 3.10 emphasizes the fact that the attained noise performance is better than the most of the CMOS SPADs, owing to the use of the wide depletion junction, that has helped reduce the tunneling noise at low excess biases. However, when compared to [30, 53, 54, 69] the proposed SPAD has higher DCR. It is to be noted that in [69], a special enrichment implant was used for the design, whereas the device in this subsection was realized without any modification to the process.

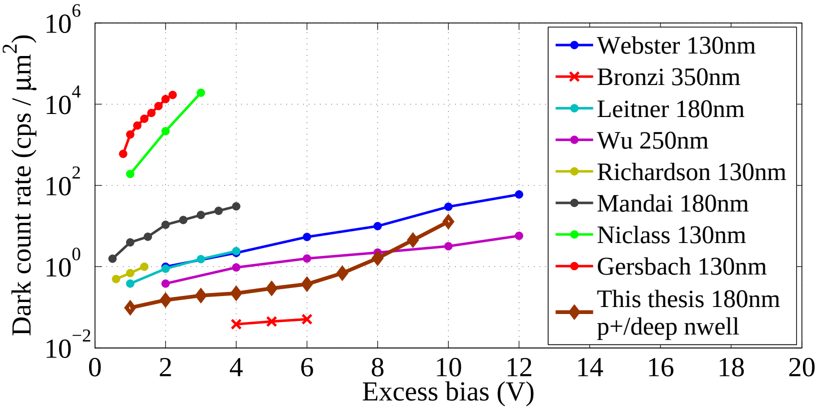


Figure 3.10: p$^+$/deep nwell SPAD: state-of-the-art DCR comparison.

3.1.1.4 Timing jitter

Jitter measurements performed using two different lasers emitting 405nm and 637nm wavelength light are presented in Figure 3.11 for various excess biases. Note: in Figure 3.11, the measured counts were normalized.

In contrast to the timing response obtained with the blue laser, for the red laser a slow exponential tail was observed. The presence of a slow exponential tail is due to the photocarriers created in the quasi-neutral region (towards the lower end of the deep nwell-2 and buried-n) diffusing toward the depletion region. At higher excess bias, jitter reduces due to the increase in field strength. Note that the measurement includes the jitter contributions from the laser (25ps for blue and 37ps for red), external circuitry, and also from the oscilloscope. Using the experimental setup as explained in Chapter 2, the device achieved 70ps FWHM and 86 ps FWHM at 10V excess bias when using a 405nm and a 637nm laser, respectively.

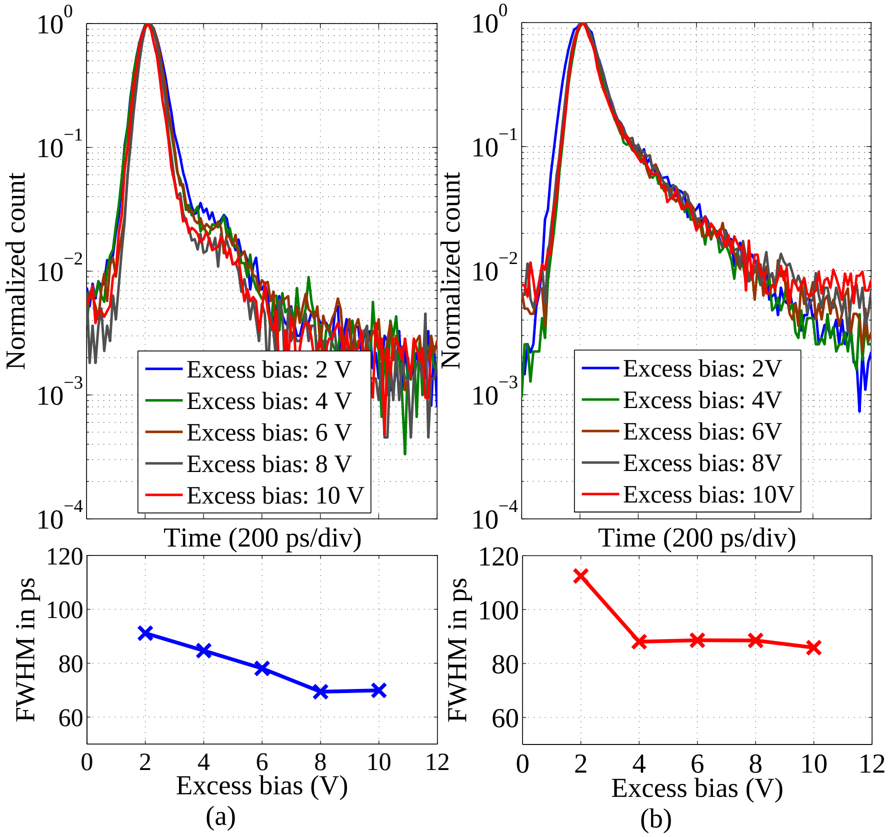


Figure 3.11: p$^+$/deep nwell SPAD: timing jitter measurements results obtained using (a) 405nm (b) 637nm laser.

3.1.2 pwell/deep nwell junction

For the SPAD design presented earlier, the diffusion of the guard ring dopants into the drawn active area has negatively influenced the device sensitivity. The impact on dopant diffusion is an issue when designing devices with smaller dimensions. For example for the SPAD presented in the previous subsection at 480nm PDP reduced from 47.6% to 42.8% when reducing the device diameter from 12μm to 6μm.

A possible approach to circumvent the dopant diffusion issue, is to use enhancement mode designs. In enhancement mode designs no additional implant or a diffusion layer is used in the guard region. A problem with the conventional enhancement mode designs, is the low fill factor resulting from the need to place the periphery far apart from the main junction's p^+ or n^+. For the p^+/deep nwell-2 junction designed without the guard ring will require at least 5μm distance between the anode and the periphery. The inactive space introduced between the p^+ and the periphery will negatively influence device fill factor as the device dimension shrinks.

In this subsection, a novel enhancement device is proposed that combines the benefits of both the enhancement mode approach and the guard ring based design.

3.1.2.1 Design

The cross section of a circular device fabricated in 180nm CMOS technology is presented in Figure 3.12.

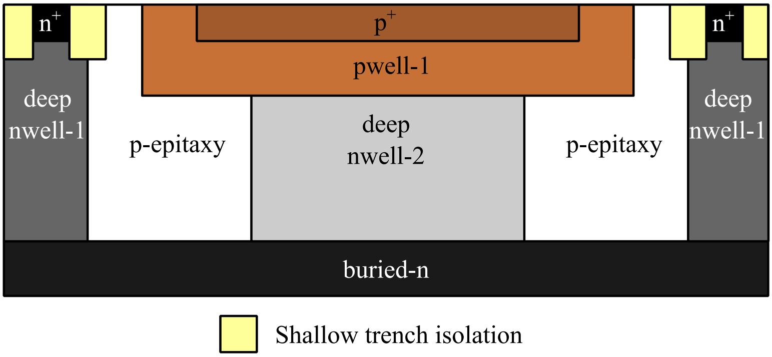


Figure 3.12: pwell/deep nwell SPAD: device cross section - fabricated in 180nm CMOS technology.

This design extends the conventional enhancement mode design with the use of pwell-1 as its anode and deep nwell-2 as its cathode. In this approach, the pwell-

1 extending outside of the deep nwell-2 creates an interface with the periphery, similar to that of a guard ring based design (Section 3.1.1). Hence, for the fabricated device, as in guard ring based design (Section 3.1.1), only 2μm spacing is required between the pwell-1 and the periphery.

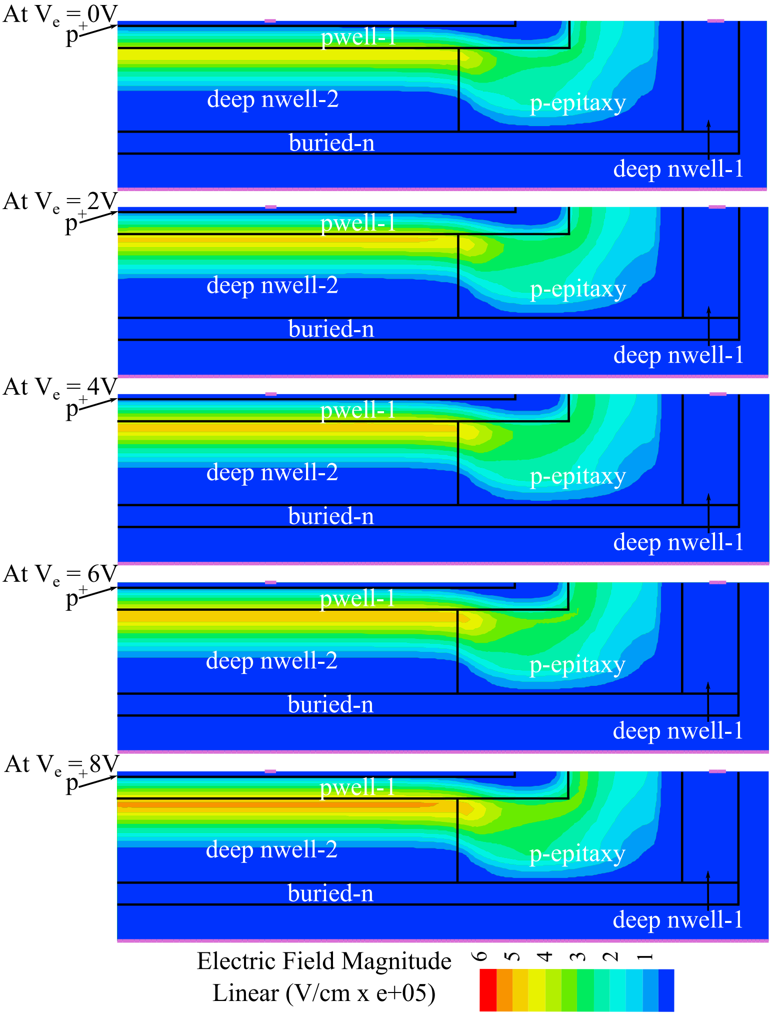


Figure 3.13: pwell/deep nwell SPAD: TCAD electric field simulation performed at various excess bias voltage.

The design, in addition to being robust to pwell-1 diffusion, utilizes deep nwell-2 diffusion to collect photons from outside of the drawn active area. At higher bias voltage the field strength (Figure 3.13) in the deep nwell-2 diffusion region reaches a value where it can aid photon detection through multiplication assisted diffusion process. For the fabricated device the measured breakdown voltage is 26.6V.

3.1.2.2 Photon detection probability

To evaluate the proposed concept, two devices were fabricated with 6μm and 12μm diameter for the deep nwell-2 (active area). The PDP[1] results are presented in Figure 3.14. As can be seen above 4V excess bias, the 6μm active diameter device has slightly higher PDP. This is because in smaller devices a relatively higher percentage of photons are acquired from the deep nwell-2 dopant diffusion region.

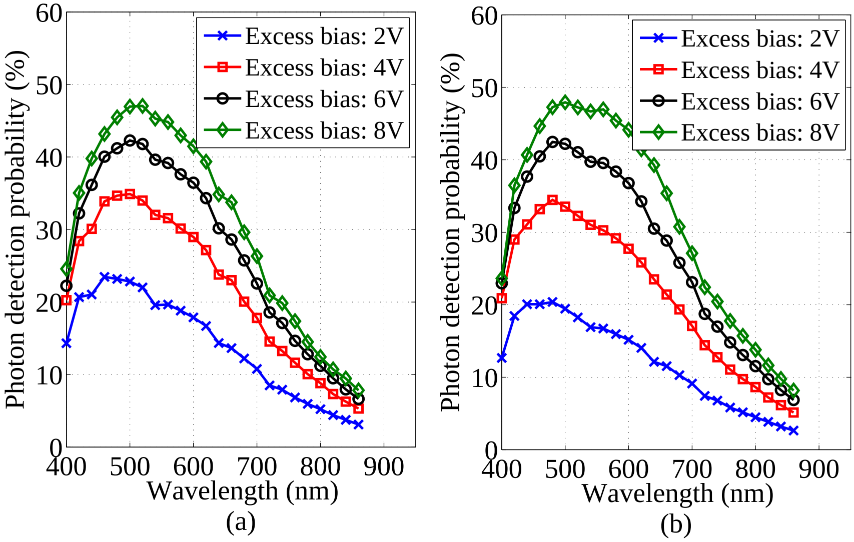


Figure 3.14: pwell/deep nwell SPAD: PDP vs wavelength for device designs with (a) 6μm and (b) 12μm diameter for the active area.

The observed increase in PDP when reducing the device active area is in contrast to the guard ring based designs. This is because, in case of the guard ring based designs, the active area reduces due to the dopant diffusion, whereas in case of the proposed design the active area has increased due to the deep nwell-2 dopant diffusion. The effective inactive distance evaluated using the PDP data is presented in Figure 3.15. As expected, above 4V excess bias, the effective inactive distance

[1] PDP was evaluated considering the drawn active area size.

was found to be negative, implying that photons are acquired from outside of the drawn active area i.e. from the deep nwell-2 diffusion region.

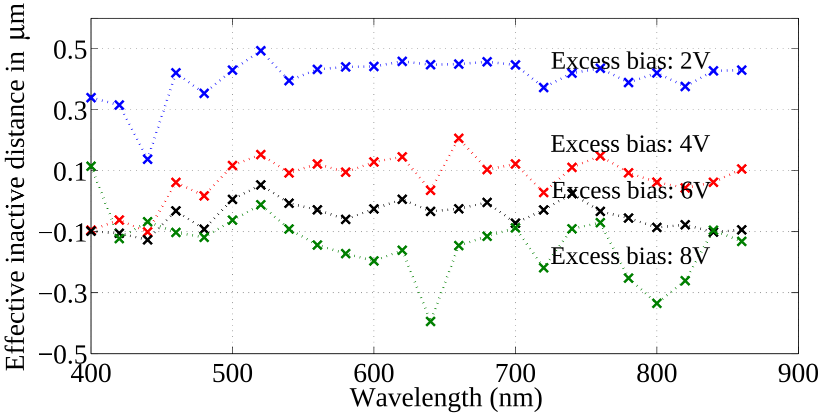


Figure 3.15: pwell/deep nwell SPAD: effective inactive distance estimation.

State-of-the-art comparison: the SPAD presented in this subsection achieves almost the same performance as that of the SPAD presented in Section 3.1.1.2. When compared to the p$^+$/nwell based designs [30, 53, 54, 69], the presented design has wider and higher PDP. The reason for such a high performance is due to the wider depletion junction and to the device design that allows 8V of excess bias operation.

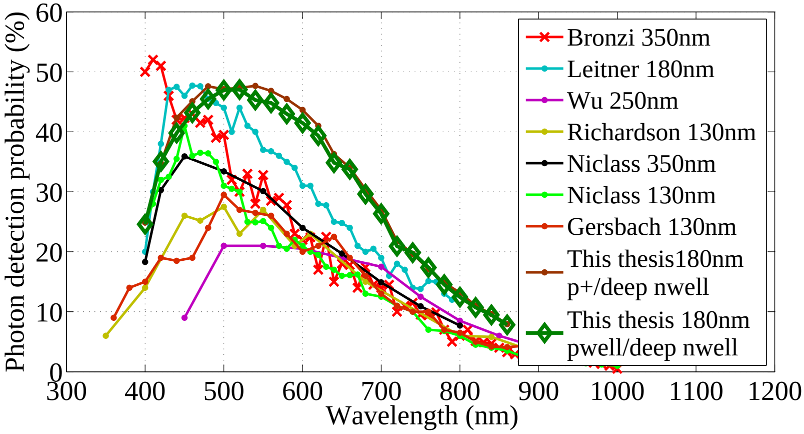


Figure 3.16: pwell/deep nwell SPAD: state-of-the-art PDP comparison.

3.1.2.3 Dark count rate

The DCR measurement results of four 12μm active diameter devices at 25°C are presented in Figure 3.17. For these measurements a device dead time of 300ns was chosen. At 2V excess bias DCR of 35 cps was measured and at 8V DCR becomes 2917 cps was measured. The detailed characterization comprehending the various component of the DCR is presented below.

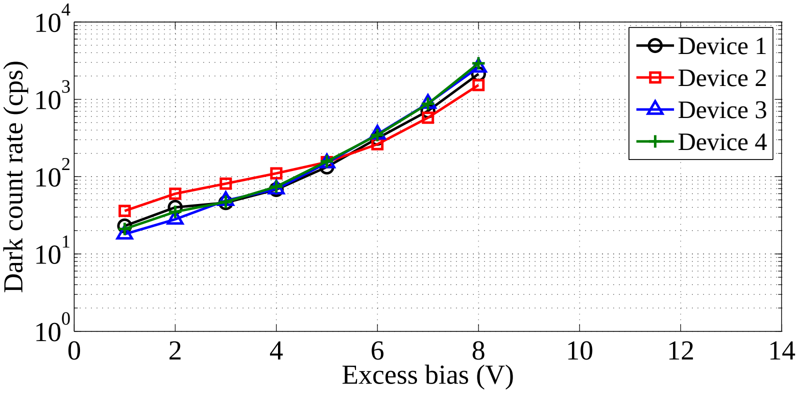


Figure 3.17: pwell/deep nwell SPAD: DCR characterization for four devices with the dead time tuned to 300ns.

Secondary pulses or afterpulsing was characterized using the inter-avalanche time histogram method described in Chapter 2. The histogram built using inter-arrival times of the SPAD is shown in Figure 3.18.

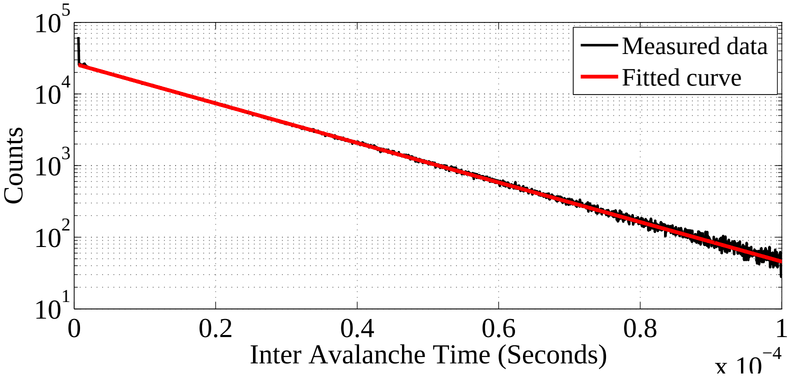


Figure 3.18: pwell/deep nwell SPAD: afterpulsing probability measurement results at 10V excess bias, performed with the device deadtime configured to 300ns.

At 8V excess bias, an afterpulsing probability of 0.86% was measured, when the device dead time is 300ns. Further, from the measurements it can be seen that when the inter-avalanche time is 2.1μs afterpulsing becomes negligible.

Primary pulses are characterized in isolation when the device dead time was tuned to 10μs. To comprehend the source of the dark noise, a device was characterized at various temperatures. The results presented in Figure 3.19, highlight that the primary pulse count rate is relatively less dependent on the temperature than to the SPAD presented in Section 3.1.1, implying that the device could have more dark noise originating due to the tunneling. This is in contrast to the expectation that a SPAD with a relatively wider depletion region should have lower tunneling noise. We believe the ineffectiveness of the pwell-1 in protecting from p$^+$ edges (Chapter 4) has resulted in higher tunneling noise.

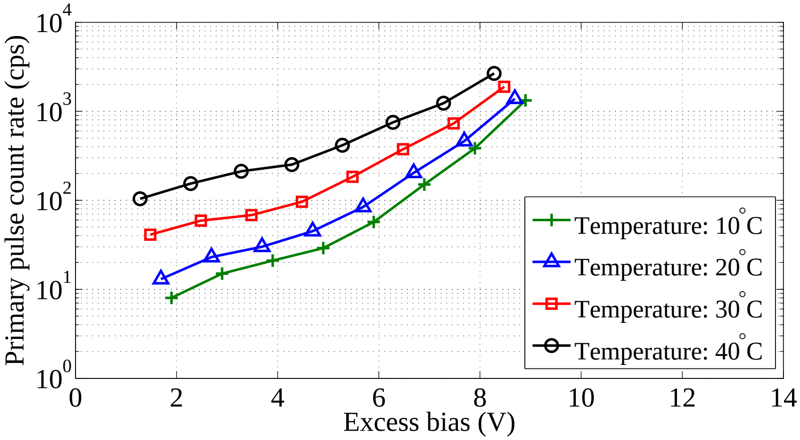


Figure 3.19: pwell/deep nwell SPAD: primary DCR characterization performed at various temperatures with the dead time tuned to 10μs.

State-of-the-art comparison is presented in Figure 3.20. As can be seen for this device, DCR is slightly higher than the p$^+$/deep nwell device presented in Section 3.1.1.The reason for the same is discussed in the primary pulse characterization section. However, when compared to other devices designed with p$^+$/nwell junctions, this device has lower noise, thanks to the use of wide depletion junction that has led to lower tunneling contributions.

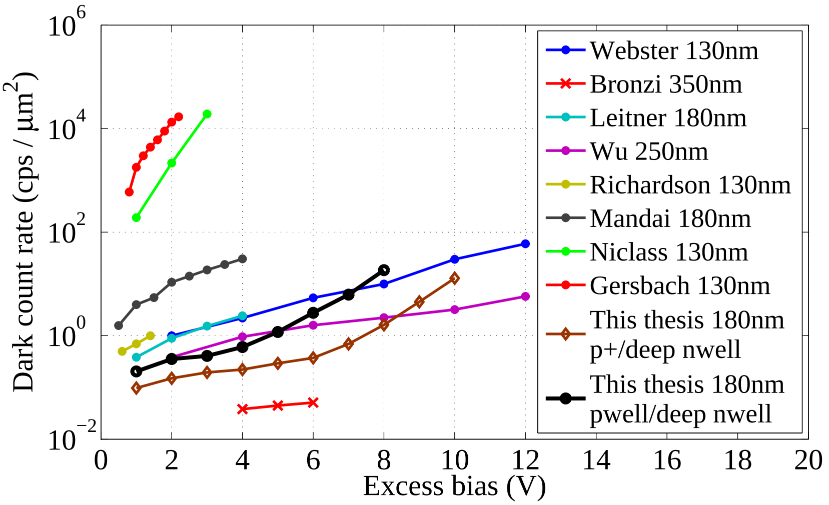


Figure 3.20: pwell/deep nwell SPAD: state-of-the-art DCR comparison.

3.1.2.4 Jitter

Timing jitter measurements performed on a 12μm active diameter device when using 405nm and 637nm laser sources are presented in Figure 3.21. The results highlight the fact that the device has slightly higher jitter when compared to the device presented in Section 3.1.2. This slight increase in timing jitter can be attributed to the low electric field region created around the active area and to the wider depletion region of this device.

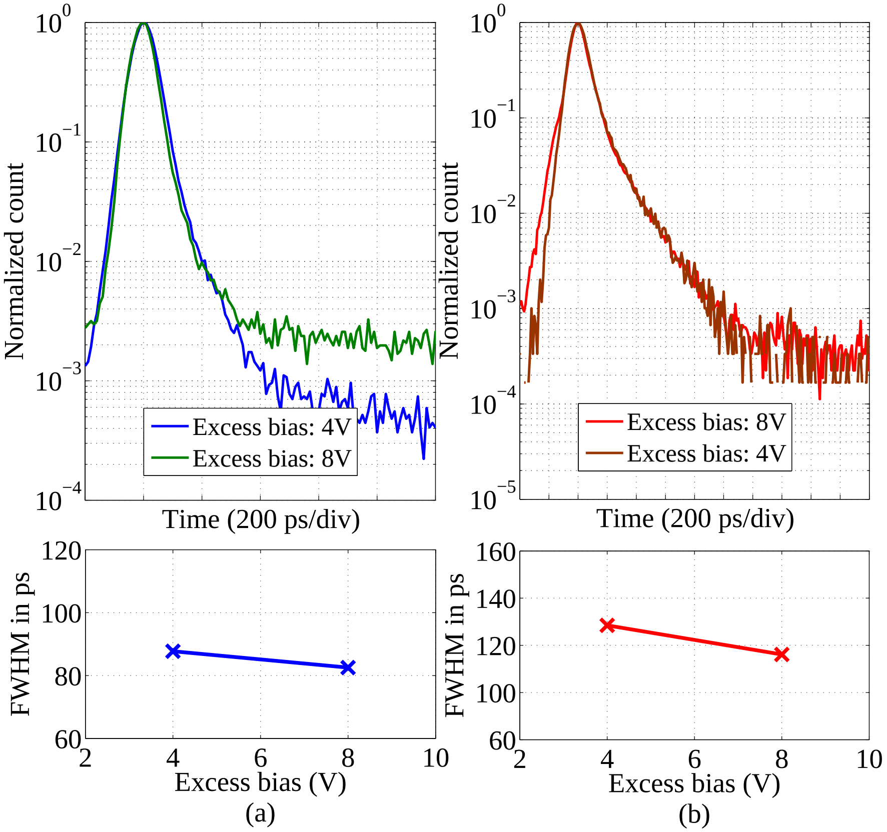


Figure 3.21: pwell/deep nwell SPAD: timing jitter measurements results performed using (a) 405nm (b) 637nm laser.

3.1.3 p-i-n junction

The SPAD designs presented in the previous subsections achieved wider sensitivity. However, the noise performance is higher when compared to the state-of-the-art low noise SPAD [69]. The detailed dark noise characterization performed at various temperatures has shown that at 10V excess bias, a major contributor to DCR is tunneling. In this section, p-i-n diodes operating in Geiger mode to reduce tunneling noise at high operating voltages are presented. The proposed structure is inspired by vertical reach-through configurations [86–89] that were proposed in early SPAD designs before they could be integrated in CMOS technologies. Wide depletion achieved in these designs has resulted in wider sensitivity profile and lower tunneling noise. However, reach-through designs were at the time several hundred microns thick and were incompatible with CMOS technology, while the proposed design using p-i-n configuration is roughly around 1-2μm thick and is fully compatible with CMOS technology.

A p-i-n photodiode is one of the most widely used structures in conventional cameras. High quantum efficiency and good frequency response achieved using the wide intrinsic region in p-i-n diodes has led to its use in optical communications. In CMOS technology, lateral [90, 91] and vertical [92] p-i-n diodes are operated in proportional APD mode. In this thesis, CMOS p-i-n diode is extended to Geiger mode operation not only to reduce tunneling noise but also to achieve a wider sensitivity profile. In addition, an adequate timing performance is also expected from p-i-n diodes mainly due to its intrinsic good frequency response achieved in proportional APD mode.

In this section, two variants of the same design with different depletion widths were studied. One of the variants achieved PDP $>40\%$ from 460nm to 600nm, with DCR <1.5 cps/μm^2 at 11V of excess bias. Another important advantage of the proposed design is the so-called PDP compression occurring at and above 11V. At this voltage, the PDP becomes saturated, i.e. practically insensitive to excess bias voltage variations, thus enabling arrays that are robust to supply and breakdown voltage variations. This property is especially important in multi-megapixel designs.

3.1.3.1 Design

The p-i-n diode is realized using a pwell/p-epitaxy/buried-n junction (Figure 3.22) in 180nm CMOS technology [93]. In this design pwell functions as anode and non-retrograde buried-n acts as cathode. In this construction, buried-n assures substrate isolation while deep nwell-1 provides a contact to buried-n. In contrast to [55] where a virtual guard ring is realized using retrograde deep nwell, in the presented design premature edge breakdown is avoided utilizing pwell lateral diffusion and a lightly doped p-epitaxy around it.

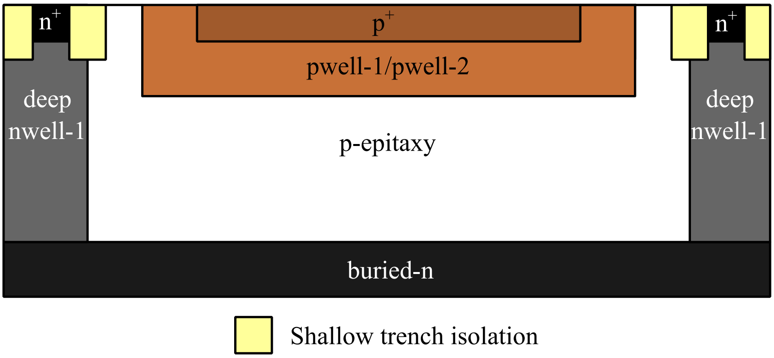


Figure 3.22: pwell/p-epitaxy/buried-n SPAD: device cross section - fabricated in 180nm CMOS technology.

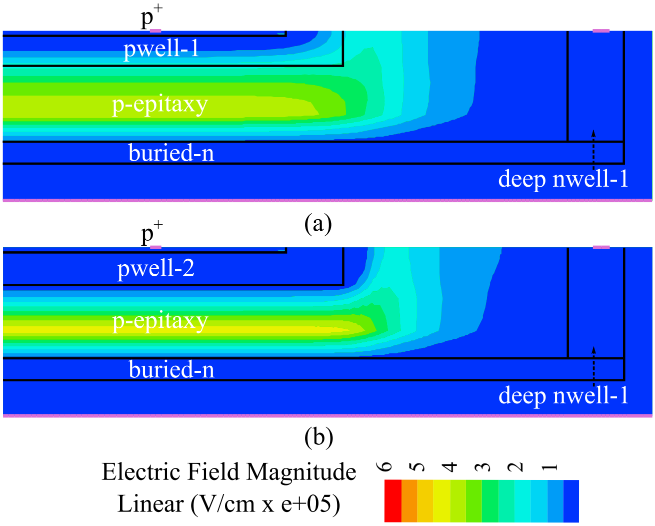


Figure 3.23: pwell/p-epitaxy/buried-n SPAD: (a) electric field simulation results for device-1. (b) electric field simulation result for device-2.

Two design variants realized with different pwell implants (pwell-1 and pwell-2), were simulated using MEDICI. Electric field simulation results presented in Figure 3.23 for device-1 (using pwell-1) and device-2 (using pwell-2) highlight uniform high field beneath pwell and lower field strength at the edges. Further, it needs to be noted that in this design, due to the use of non-retrograde buried-n, the peak electric field is observed at the interface between the buried-n and the p-epitaxy layer.

The following sections discuss detailed characterization of device-1 and 2; characterizations were performed on circular devices with a 12μm diameter for the pwell.

3.1.3.2 Device Variant 1

Device-1 designed with pwell-1 results in a breakdown voltage of 36.5V. pwell-1 is shallower than pwell-2 and thus it enables a wider depletion region than device-2.

3.1.3.2.1 Dark count rate

DCR measurements performed on 4 devices when using 300ns dead time at 25°C is presented in Figure 3.24. For device-1 a DCR of 2195 cps was measured at 4V excess bias. Detailed characterization on primary and secondary pulses contributing to total DCR is discussed bellow.

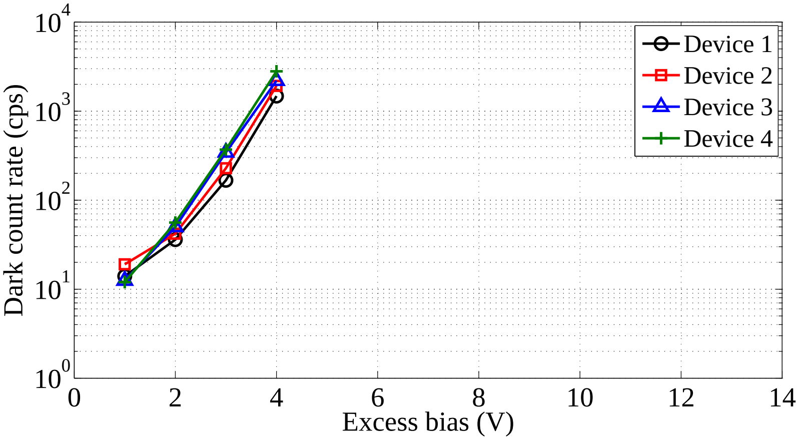


Figure 3.24: pwell-1/p-epitaxy/buried-n SPAD: DCR characterization for four devices with the dead time tuned to 300ns.

Secondary pulses or Afterpulsing measurement results are presented in Figure 3.25. For device-1 with a dead time of 300ns at 4V excess bias afterpulsing probability of 0.34% was measured, when using the inter-avalanche time histogram technique.

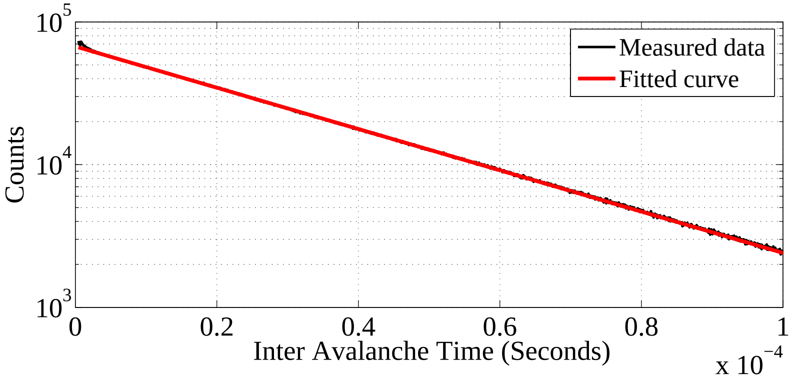


Figure 3.25: pwell-1/p-epitaxy/buried-n SPAD: afterpulsing probability measurement results at 10V excess bias, performed with a deadtime of 300ns.

Primary pulses: generally, in wide depletion devices, lower tunneling noise is expected. However, temperature measurements have shown the contrary (Figure 3.26). Above 3V of excess bias, primary pulse count rate is more dependent on voltage than on temperature, suggesting that a major contributor to noise is tunneling. It is believed that higher tunneling contribution is due to the exposure of p$^+$ edge to high electric field that resulted from widening of the depletion region due to increased excess bias.

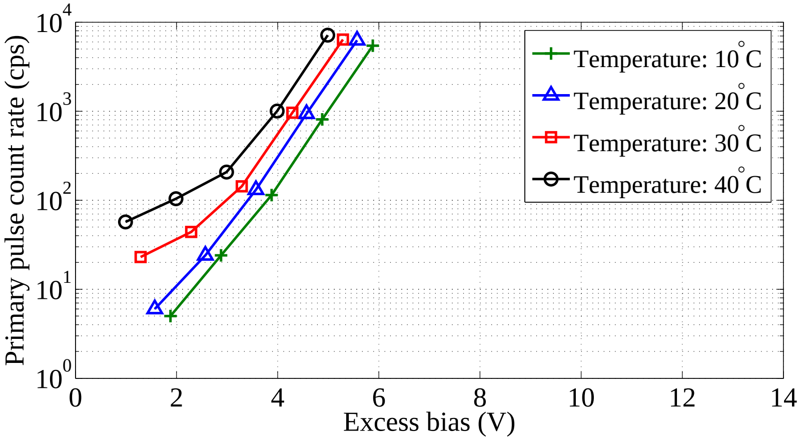


Figure 3.26: pwell-1/p-epitaxy/buried-n SPAD: primary DCR characterization performed at various temperatures with a dead time of 10μs.

3.1.3.2.2 Photon detection probability

In general, devices with wider depletion need to be operated at higher excess bias to enhance PDP. For device-1, high DCR limits device operation to lower excess bias. PDP characterization performed until 4V excess bias is presented in Figure 3.27. Though, wider PDP profile is in-line with depletion thickness, peak PDP is only 27.8% at 4V excess bias.

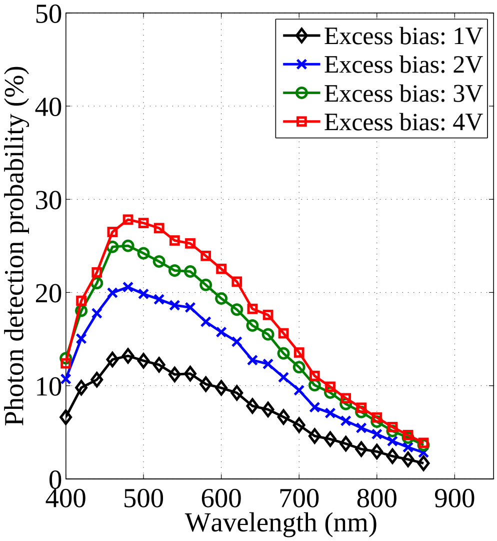


Figure 3.27: pwell-1/p-epitaxy/buried-n SPAD: PDP vs wavelength.

3.1.3.2.3 Timing jitter

Timing jitter measurements performed when using blue(405nm) and red(637nm) laser sources are presented in Figure 3.28 a and b. When using the red source, a jitter (FWHM) of 427ps was measured at 2V and and 223ps at 4V excess bias. The blue source resulted in a jitter (FWHM) of 243ps and 141ps, respectively. By comparing the jitter results with SPADs presented in [30, 55, 69], it can be seen that the jitter obtained with the proposed design is almost two times higher. Thus, higher excess bias operation is required to improve timing jitter performance, even as high DCR limits device operation to 4V excess bias.

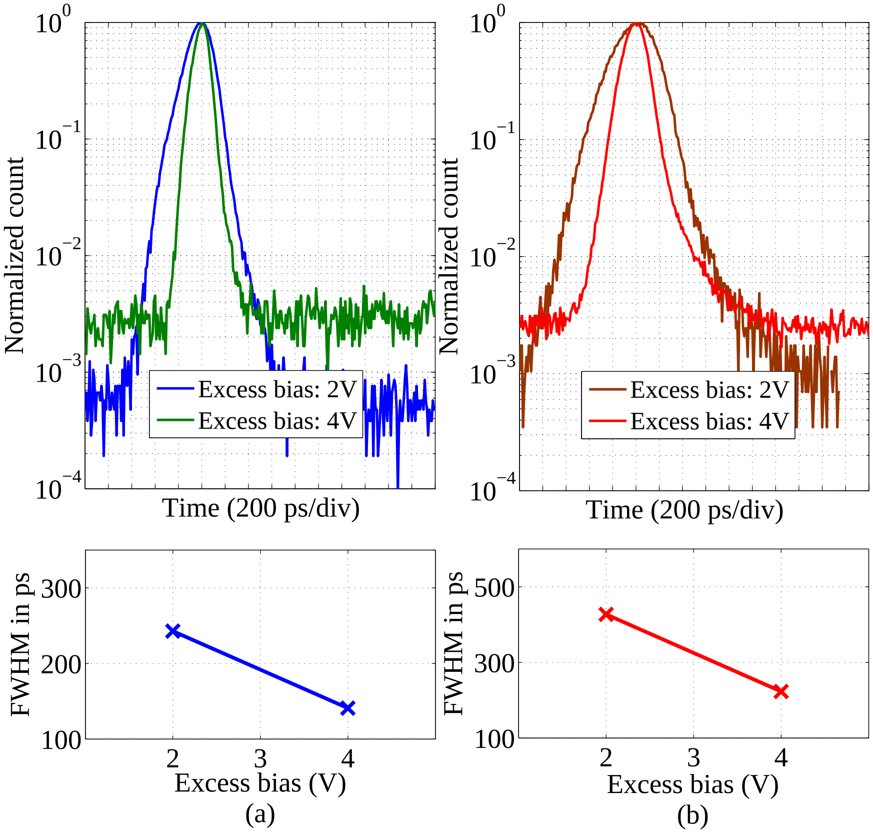


Figure 3.28: pwell-1/p-epitaxy/buried-n SPAD: timing jitter measurements results performed using (a) 405nm (b) 637nm laser.

3.1.3.3 Device Variant 2

For device-2, the expected p^+ edge exposure to depletion region is circumvented by the use of pwell-2 that is deeper and more highly doped than pwell-1. When using pwell-2, a device breakdown voltage of 24.46V was measured. Detailed device-2 characterization is presented in this section.

3.1.3.3.1 Dark count rate

DCR measurements performed on four different devices at 25°C with 300ns dead time is presented in Figure 3.29. Compared to device-1, device-2 has lower noise at 4V excess bias and is operational until 12V of excess bias. At 11V excess bias, the DCR is 1.5 cps/μm^2. The achieved performance is superior to any other deep submicron CMOS SPAD operating above 10V of excess bias [3, 57, 85]. Characterization results of primary and secondary pulses are discussed in the following.

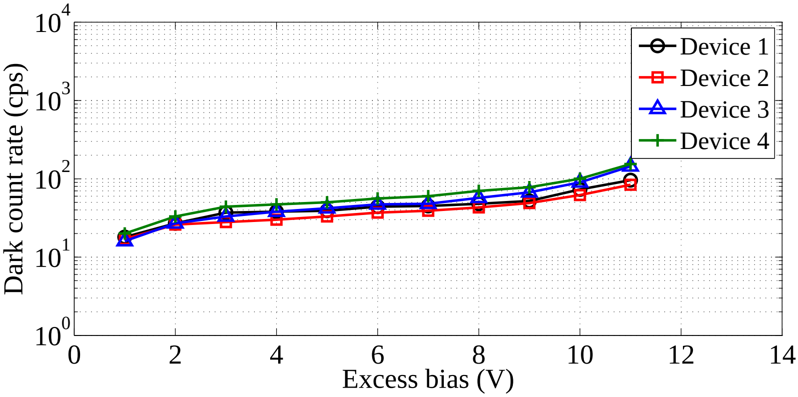


Figure 3.29: pwell-2/p-epitaxy/buried-n SPAD: DCR characterization for four devices with the dead time tuned to 300ns.

Secondary pulses or Afterpulsing measurements performed using inter-avalanche time histogram technique is presented in Figure 3.30. For device-2 when using a dead time of 300ns at 11V excess bias an afterpulsing probability of 7.2% was measured. Further, when comparing the afterpulsing probability of the two devices at their respective operating voltages, i.e. at 11V excess bias for device-2 and at 4V for device-1, it can be seen that device-2 has higher afterpulsing than device-1. The observed difference is due to high excess bias operation of device-2. Whereas at 4V excess bias device-1 has higher afterpulsing probability than device-2 (0.08%). The reason for this behavior is a wider depletion region.

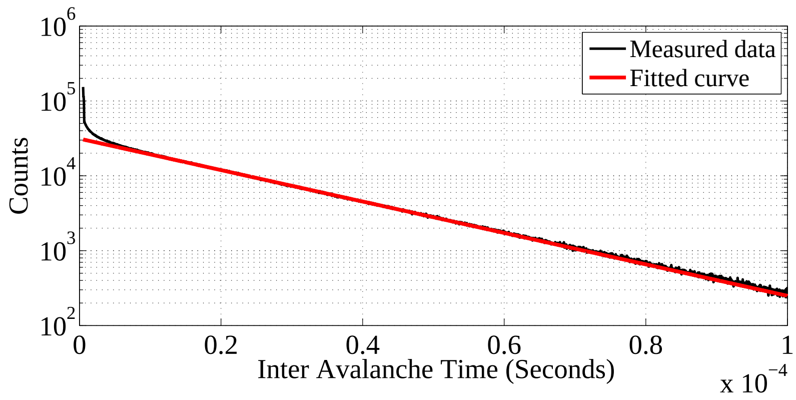


Figure 3.30: pwell-2/p-epitaxy/buried-n SPAD: afterpulsing probability measurement results at 10V excess bias, performed with the device dead time set to 300ns.

Primary pulses characterization performed at various temperatures suggest a lower tunneling contribution towards DCR when compared to device presented in 3.1.1 even at an excess bias voltage as high as 11V (Figure 3.31). The low tunneling contribution is due to the p-i-n configuration and to the use of the pwell-2 as anode.

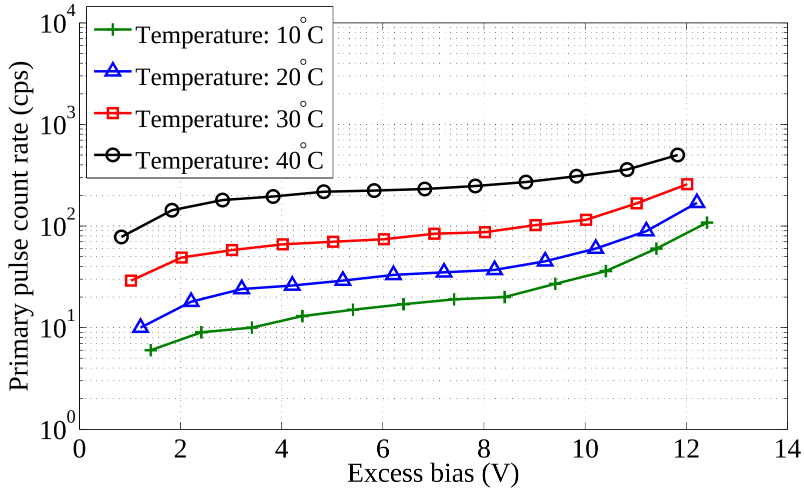


Figure 3.31: pwell-2/p-epitaxy/buried-n SPAD: primary DCR characterization performed at various temperatures with the dead time tuned to 10μs.

State-of-the-art comparison is presented in Figure 3.32. It was shown than the tunneling noise is a major contributor at 10V excess bias [3,57]. For the device presented in this subsection, with the use of p-i-n construction, tunneling noise was lowered resulting in better noise performance. Further, device-2 achieves 1.5 cps/μm^2 at 11V excess bias while state-of-the-art low noise SPADs achieve 0.05 cps/μm^2 at 6V excess bias [69]. Though, device-2 has higher noise than [69], it has to be noted that in device-2 no special implant layer [69], but a standard CMOS technology, was used.

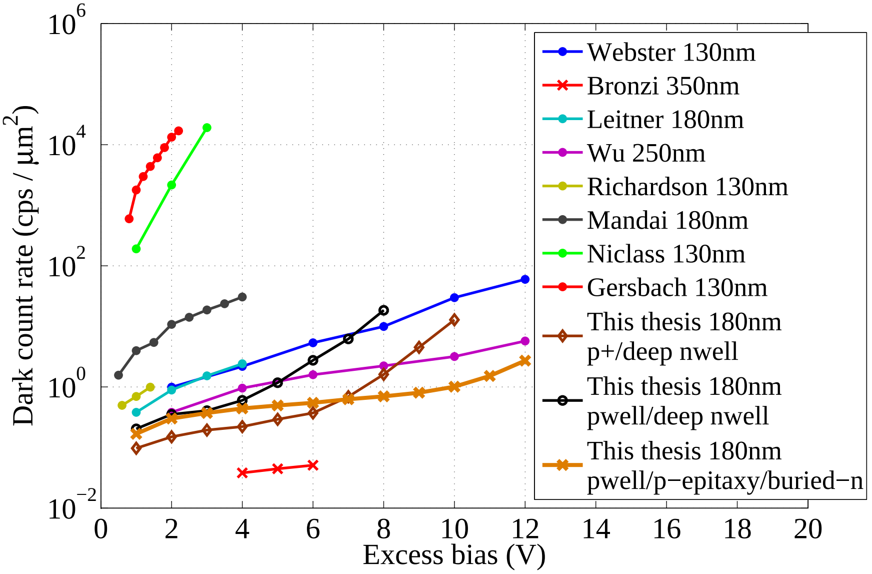


Figure 3.32: pwell-2/p-epitaxy/buried-n SPAD: state-of-the-art DCR comparison.

3.1.3.3.2 Photon detection probability

PDP characterization results are presented in Figure 3.33a. Device-2 achieves PDP greater than 40% from 460nm to 600nm at 11V excess bias. The achieved performance is comparable to the SPADs presented in the previous subsections. By analyzing the PDP at 500nm for various excess bias voltages (Figure 3.33b), it can been seen that the PDP tends to saturate when the excess bias voltage exceeds 11V. This implies that the device is robust to supply voltage and device breakdown variations.

State-of-the-art comparison: Figure 3.34 compares device-2 PDP with state-of-the-art substrate-isolated SPADs. When compared to conventional p$^+$/nwell based designs [30, 53, 69, 81], device-2 achieves a wider PDP profile. Comparing

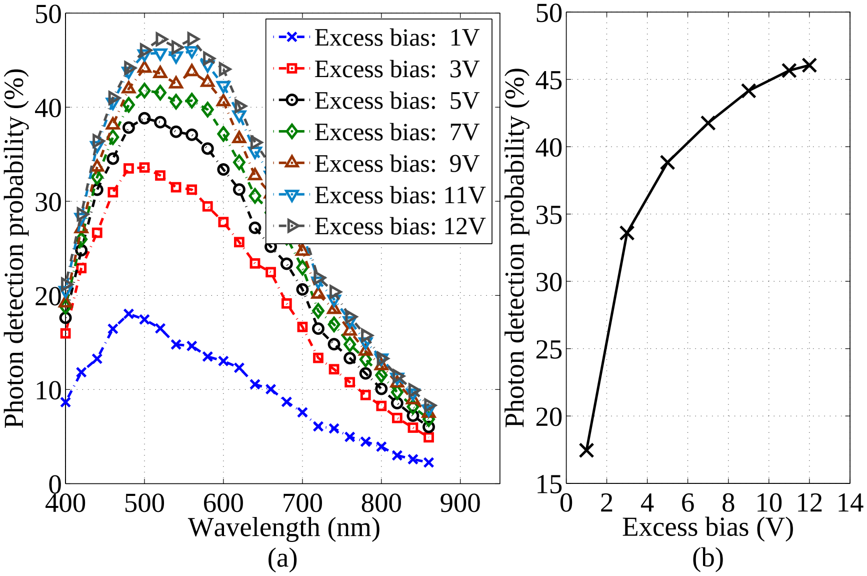


Figure 3.33: pwell-2/p-epitaxy/buried-n SPAD: (a) PDP vs wavelength, (b) PDP at 500nm wavelength for various excess bias voltages.

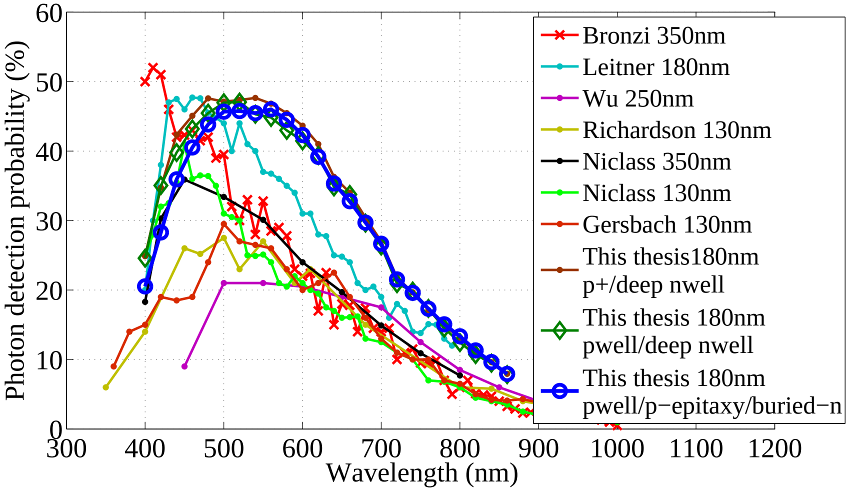


Figure 3.34: pwell-2/p-epitaxy/buried-n SPAD: state-of-the-art PDP comparison.

with the p$^+$/deep nwell-2 based design, device-2 has lower sensitivity in the blue due to the junction formed deeper in silicon. However, above 500nm the sensitivity is almost identical to p$^+$/deep nwell-2 based design. Further, comparing with pwell/deep nwell junction [55, 85], device-2 has a higher and wider PDP profile; thanks to the use of an p-epitaxy layer that has enabled wider depletion.

3.1.3.3.3 Timing jitter

Timing jitter measurements performed when using blue(405nm) and red(637nm) laser sources are presented in Figure 3.35 a and b, respectively. As expected, the jitter of device-2 is improved by increasing excess bias. When using blue laser source, the jitter improved from 133ps at 3V excess bias to 97.2ps at 11V excess bias. For the red laser source, the jitter was 139.5ps at 3V and 100.8ps at 11V of excess bias voltage. Although device-2 needs to be operated at higher excess bias, the main advantage of the proposed design is the reduction in exponential tail that results from carrier diffusion. In narrower depletion junctions formed near the silicon surface, some of the red photons are acquired through photocarrier diffusion process. In the proposed design, the achieved wide depletion region encompasses a major part of the photon collection region resulting in a reduced exponential tail.

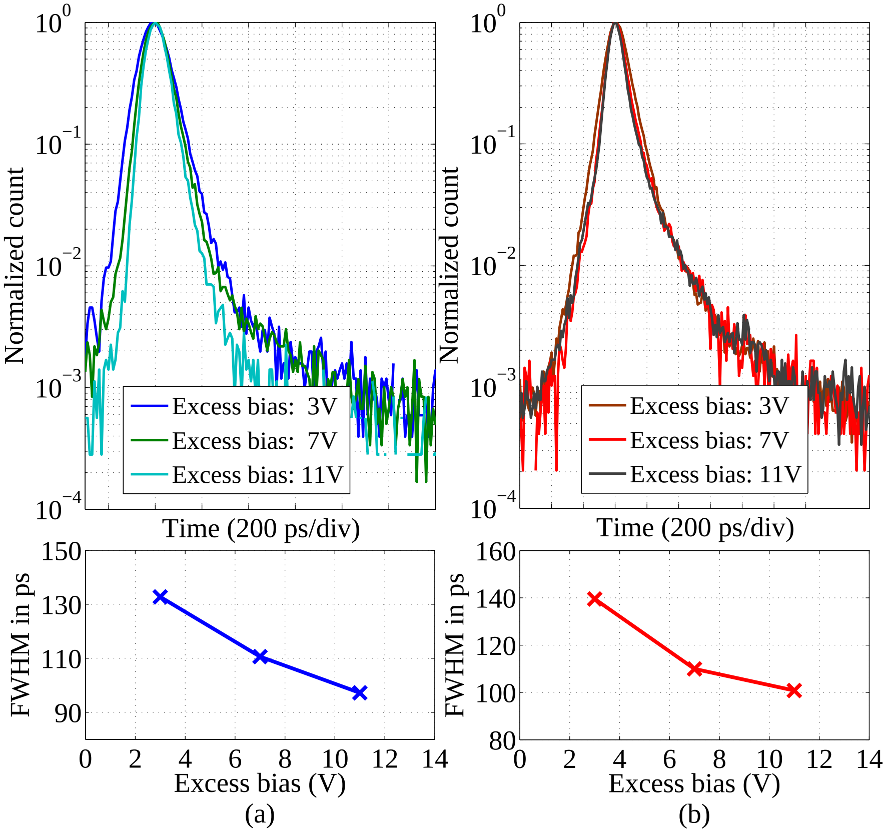


Figure 3.35: pwell-2/p-epitaxy/buried-n SPAD: timing jitter measurements results performed using (a) 405nm (b) 637nm laser.

3.2 Wide photocarrier collection region

Though SPADs designed with a wide depletion junction has resulted in a wider spectral sensitivity, the need to operate the devices at or above 8V excess bias is one of its drawbacks. To circumvent the high excess bias operation, this section presents an active area design technique that uses photocarrier diffusion process. The presented technique, uses a narrower depletion junction with a wider photon carrier collection region for the active area. In such designs, photocarriers acquired through the diffusion process will enhance the sensitivity spectrum. The CMOS SPAD presented in this section achieves almost the same PDP profile as that of the wide depletion junction based designs. The pros and cons of this design technique is discussed in detail in the reminder of this section.

3.2.1 p^+/nwell junction

The conventional p^+/nwell junction PDP profile is enhanced utilizing photocarrier diffusion processes, resulting in a considerable expansion of the sensitivity spectrum of more than 30%. The proposed device achieves PDP greater than 40% from 440nm to 580nm at a 4V excess bias, while the DCR is 16 cps/μm^2. The achieved sensitivity is wider and higher than any other CMOS SPADs when operated at 4V excess bias. When compared to the SPADs designed using similar technique but using substrate as its photon collection region [57], the proposed design has better timing performance in the blue, and in the red it has relatively lower contribution towards the exponential tail.

3.2.1.1 Design

The cross section of the SPAD designed using p^+/nwell junction is presented in Figure 3.36. Deep nwell-2 and highly doped buried-n placed beneath nwell acts as a photon collection region [94]. In this configuration, part of the photo-holes created in the quasi-neutral region of the buried-n, deep nwell-2 and from the lower part of the nwell could reach the multiplication region through the diffusion process. When compared to conventional p^+/nwell devices [30,53,54,69], the proposed device has a wider photon collection region due to the buried-n layer protecting deep nwell-2 from depletion, when the substrate junction is reverse biased.

Electric field simulations (Figure 3.37) performed using MEDICI highlight the effectiveness of buried-n in protecting deep nwell-2 from depletion, when the substrate junction is reverse biased. Further, the minimal depletion observed in buried-

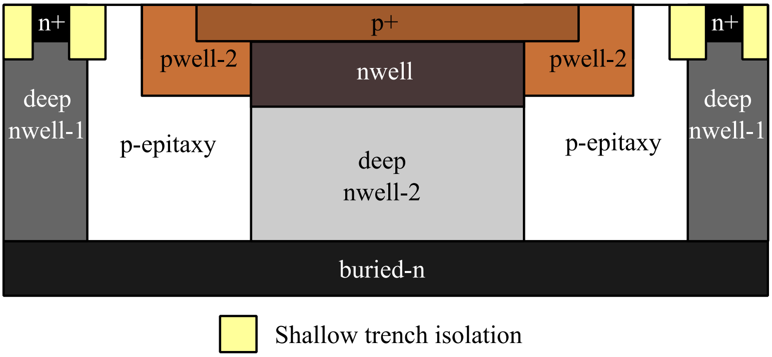


Figure 3.36: p$^+$/nwell SPAD: device cross section - fabricated in 180nm CMOS technology.

n due to its higher doping concentration suggests that almost the entirety of the buried-n volume, along with deep nwell-2, will function as a photon collection region. Note: the simulations were performed when the cathode was biased at the device breakdown, while the anode and the substrate were tied to ground.

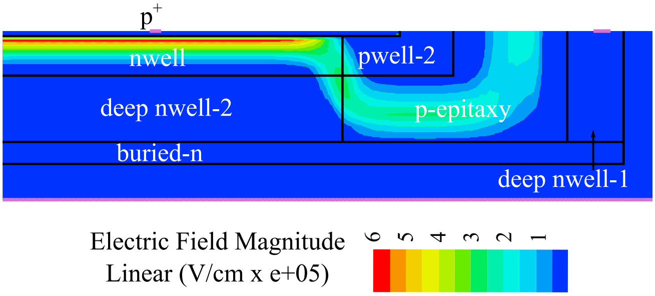


Figure 3.37: p$^+$/nwell SPAD: device electrical field simulation results performed using MEDICI.

In this design the pwell-2, used as the guard ring, protects p$^+$ edges from premature edge breakdown. Further, buried-n enabling substrate isolation, provides contact to deep nwell-2, while avoiding the need to counter dope pwell-2, resulting in improved guard ring effectiveness. For a circular device with a active area diameter of 12μm fabricated in 180nm CMOS technology, a breakdown voltage of 14.64V was measured.

3.2.1.2 Photon detection probability

PDP at various excess bias voltages is presented in Figure 3.38. Measurement results at 4V excess bias are compared with state-of-the-art CMOS SPADs [30,53–55,69,85,95,96] in Figure 3.39.

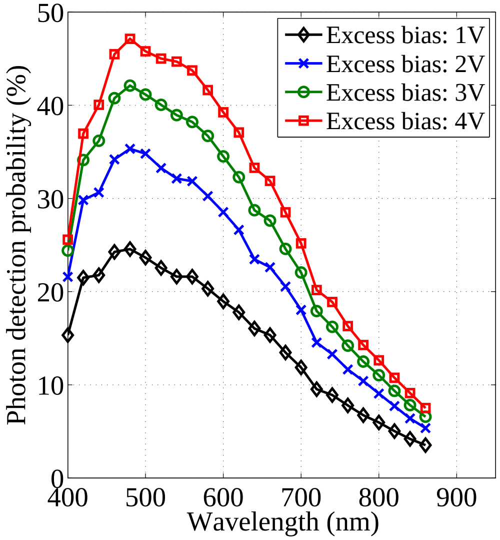


Figure 3.38: p$^+$/nwell SPAD: PDP vs wavelength.

The proposed device surpasses state-of-the-art p$^+$/nwell junction based designs, mainly due to the presence of a wider photon collection region. When operating at 4V excess bias, the design presented here outperforms equivalent wide depletion devices presented in Section 3.1. Further, the reported PDP at 4V excess bias compares favorably with that of the SPAD presented in Section 3.1 at their respective operating excess bias voltages.

When compared to SPADs [57,58] designed using substrate as its photon collection region, it can be seen that the presented design has superior performance at 4V excess bias up until 680nm (Figure 3.40). Higher sensitivity above 680nm in non-substrate isolated SPAD is due to the design allowing photocarrier collection from deep within the substrate through the diffusion process, where most of the lower energy photons are absorbed in silicon. In the presented design the electric field resulting from reverse biasing the buried-n/substrate junction prevents photocarriers generated in the substrate from reaching the multiplication region. Though the use of buried-n reduces the device sensitivity beyond 680nm, it helps in en-

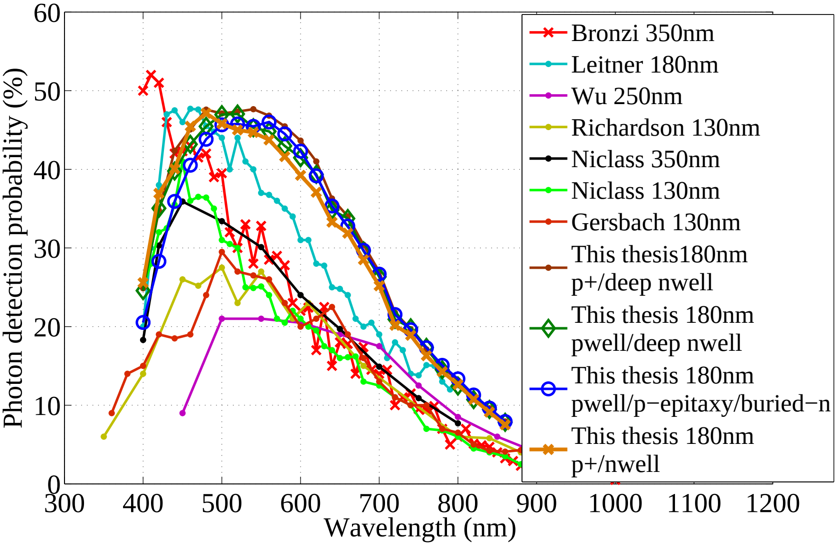


Figure 3.39: p$^+$/nwell SPAD: state-of-the-art PDP comparison.

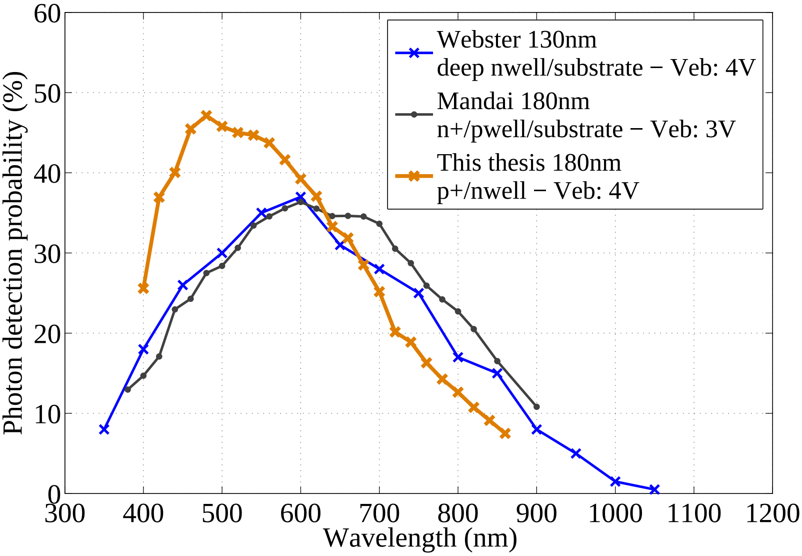


Figure 3.40: p$^+$/nwell SPAD: state-of-the-art PDP comparison with non substrate isolated SPADs.

abling substrate isolation and reducing the exponential tail of the time response of the device, when compared to non-substrate isolated SPADs.

3.2.1.3 Timing jitter

Timing jitter measurements performed using red (637nm) and blue (405nm) lasers are presented in Figure 3.41 a and b, respectively. It can be seen that when using a red laser, a relatively higher number of photons contribute to the exponential tail; suggesting that more red photons are collected through photocarrier diffusion. This observation is in line with our design of the photon collection region, which is deeper into silicon, where most of the red photons are absorbed.

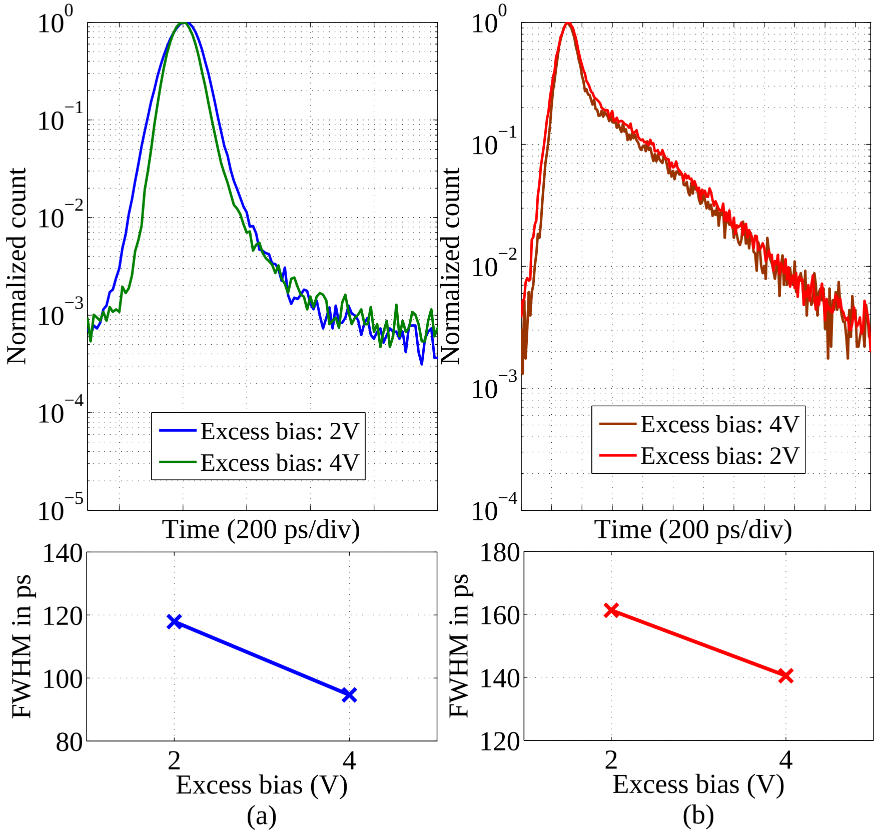


Figure 3.41: p^+/nwell SPAD: timing jitter measurement results performed using (a) 405nm (b) 637nm laser.

Jitter measurements showed a jitter of 141ps FWHM for red (637nm), and 95ps

Table 3.1: State-of-the-art jitter comparison when using blue laser.

	Blue full width at half maximum	Blue full width at 10% maximum
Webster 130nm	1.55ns(443nm)	>5ns
Mandai 180nm	182ps(405nm)	600ps(405nm)
Bronzi 350nm	85.8ps(390nm)	400ps(390nm)
This work 180nm	95ps(405nm)	180ps(405nm)

Table 3.2: State-of-the-art Jitter comparison when using red laser.

	Red full width at half maximum	Red full width at 10% maximum
Webster 130nm	77ps(654nm)	3ns(654nm)
Mandai 180nm	165ps(790nm)	550ps(790nm)
Bronzi 350nm	119ps(780nm)	-
This work 180nm	141ps(637nm)	690ps(637nm)

FWHM for blue (405nm) sources at 4V excess bias. When compared to [57], where the substrate was used as a photon collection region, the presented design has better timing performance in the blue and has lower contribution towards exponential tail in the red. Table 3.1 and 3.2 present the state-of-the-art CMOS SPAD jitter comparison when using blue and red laser sources. Along with generally reported full width at half maximum, full width at 10% maximum is also reported to include the exponential tail in jitter comparison. It can be seen that the presented device has comparable timing performance when using blue sources, and for red sources, as expected, it has slightly worse full width at 10% maximum mainly due to presence of a diffusion tail, as discussed earlier.

3.2.1.4 Dark count rate

DCR measurement results of the four devices at 25°C involving both primary and secondary pulses is presented in Figure 3.42. A DCR of 31 cps at 1V excess bias was measured, and 1.8k cps at 4V.

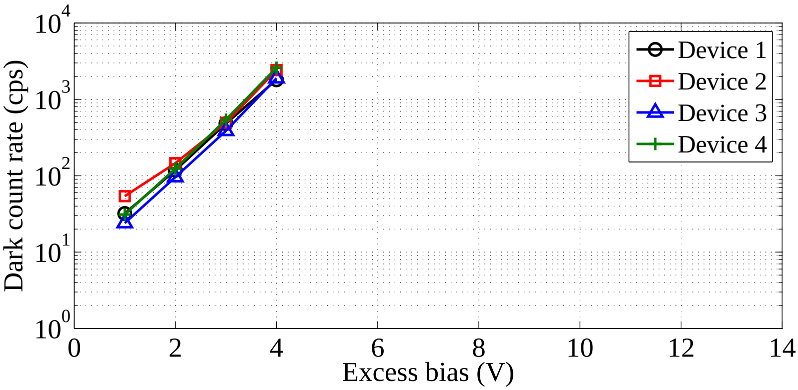


Figure 3.42: p$^+$/nwell SPAD: DCR characterization for four devices with the dead time tuned to 300ns.

Secondary pulses or afterpulsing probability of 0.2% was measured using inter-avalanche time histogram technique at 4V excess bias. Measurement results presented in Figure 3.43 suggests that the afterpulsing is negligible when the inter-avalanche time is more than 1.4μs. For this measurement a device dead time of 300ns was chosen.

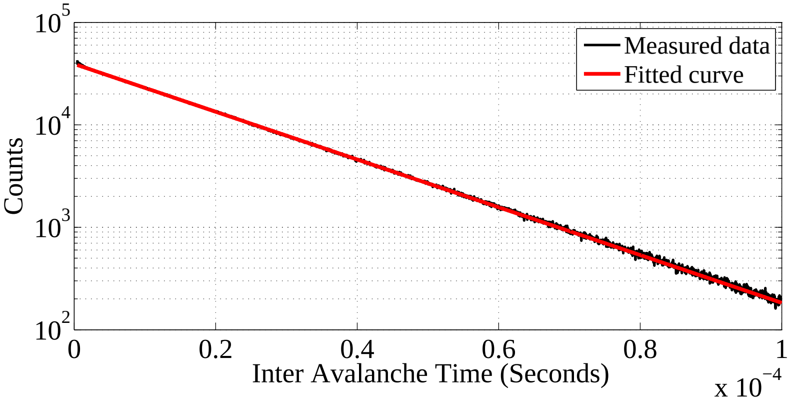


Figure 3.43: p$^+$/nwell SPAD: afterpulsing probability measurement results at 10V excess bias, performed with the device deadtime configured to 300ns.

Primary pulses characterization performed at various temperatures is presented in Figure 3.44. For this measurement, a device dead time of 10μs was chosen, to ensure negligible afterpulsing. The measurement highlights that the DCR dependence on voltage is higher than that on temperature, suggesting that a major contributor to noise is tunnelling.

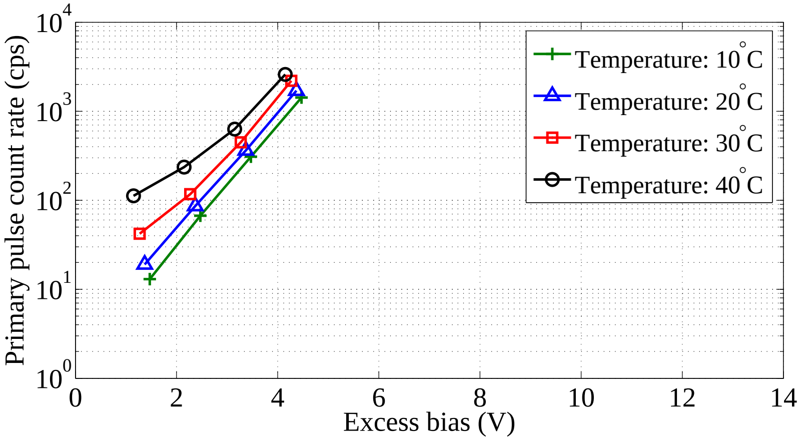


Figure 3.44: p$^+$/nwell SPAD: primary DCR characterization performed at various temperatures with the dead time tuned to 10μs.

State of the art comparison: as can be seen in Figure 3.45, the device presented in this subsection has slightly higher DCR than the state-of-the-art substrate isolated devices [3, 55, 85] mainly due to tunnelling. Low DCR SPADs presented in [69] use special enrichment implant, whereas the SPAD presented in this subsection was designed without any process modifications.

Comparing device sensitivity (Figure 3.39) to DCR (Figure 3.45) at their operating excess bias, it can be seen that the proposed device is comparable to the device presented in Section 3.1. However when compared to the p-i-n diode based the device has higher noise performance.

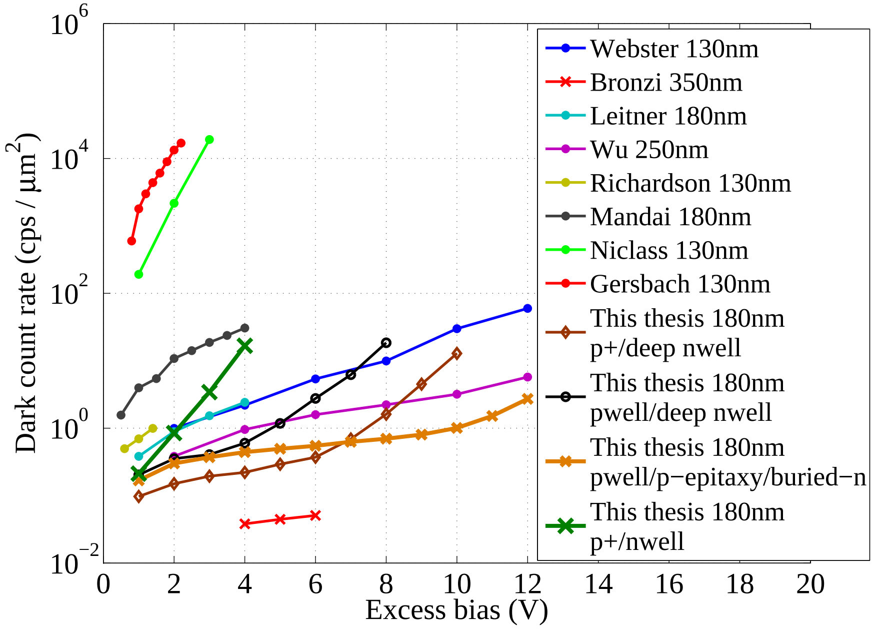


Figure 3.45: p$^+$/nwell SPAD: state-of-the-art DCR comparison.

3.3 Summary

We have implemented 5 different active areas to explore the SPAD performance optimization and to study the design trade-offs in PDP, DCR, timing jitter and device excess bias operation.

We found the following facts to be true.

- Substrate isolated SPAD's spectral response can be improved by using either a wide depletion junction or by using a narrower depletion junction with a wider photon collection region.
- Wide depletion design has lower DCR and better timing performance when compared to the other approach.
- Wide depletion junction based approaches require >8V of excess bias operation, whereas the narrower depletion junction with wide photon collection regions require only 4V excess bias.
- The exponential component of the timing response increases by widening the photon collection region.

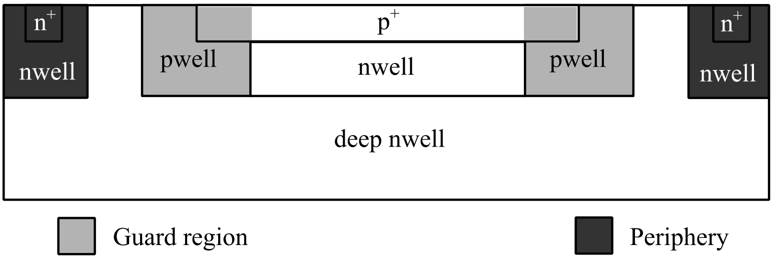
The performance characteristics of the designs discussed in this Chapter is summarized in Table 3.3.

Table 3.3: Summary on SPAD characteristics.

	p^+ / deep nwell	pwell / deep nwell	p-i-n variant-1	p-i-n variant-2	p^+ / nwell
Active area diameter (μm)	12	12	12	12	12
Excess bias (V)	10	8	4	11	4
DCR (cps)	1453	2917	2195	154	1887
PDP >40% (nm)	440-620	440-600	-	460-600	44-580
Timing jitter 405nm/637nm (FWHM in ps)	70/86	83/116	141/223	97/101	95/141

Chapter 4

Guard region and periphery design



This chapter presents the design techniques for guard region and periphery design in CMOS SPADs. In SPADs, the guard region is used to avoid a premature edge breakdown at the sharp edges of either the anode or the cathode of the main junction. The device periphery is used to provide a contact to the main junction terminal placed deeper in substrate. Although, the guard region and the periphery are required for device operation, the area occupied by them is insensitive to photons. A general approach is to reduce their size to improve fill factor. Alternatively, in literature [55] novel guard region designs have also been reported to reduce insensitive areas. In all reported designs, it is generally assumed that the SPAD performance is not dependent on the guard region design or on the periphery design. The results presented in this chapter, will show that in addition to the main junction, also the guard region and the device periphery can contribute to SPAD DCR. Further, in this chapter, the design techniques to suppress DCR that originates from guard region and periphery are also presented.

In SPADs, a high voltage applied across the main junction, is also seen across some of the parasitic junctions formed either with the guard region and/or with the periphery. The impact on DCR is studied in this chapter, considering the influence of these parasitic junctions. Of the two design approaches presented in Chapter 3, the designs with the wide depletion junction are most likely to be affected, mainly due to its need to be biased at higher voltages. Hence, in this chapter, wide depletion junction based designs are considered for the study. However, it needs to be noted that the conclusions drawn on the influence of the parasitic junctions can be generalized to any SPAD design.

This chapter is organized as follows: Section 4.1 and Section 4.2 presents the guard region design and the periphery design from CMOS SPADs. Section 4.3 presents a summary of this chapter. Note: in this chapter the guard region is also referred to as the guard ring owing to its shape.

4.1 Guard region

This section presents a detailed study carried out to comprehend the influence of the guard region design on the SPAD's DCR. The results of this study are organized in three subsections where the influence on DCR due to the guard junction breakdown, guard ring sizing and the effectiveness of the guard ring with respect to the device curvature are presented.

4.1.1 Breakdown

4.1.1.1 Design-1

Conventionally, in a CMOS SPAD designed with a p^+/deep nwell or with a p^+/nwell junction, premature edge breakdown at the sharp edges of the p^+ is avoided by the use of a pwell guard ring. When implementing such designs on a p-substrate, the deep nwell is placed beneath the pwell to provide contact to the cathode.

To comprehend the implication of the parasitic pwell/deep nwell junction, a device with p^+/deep nwell-2 as its main junction was fabricated in 180nm CMOS technology. The device cross section is presented in Figure 4.1a. For the fabricated device, the difference in breakdown voltage between the main junction and the parasitic junction is 4.3V (Figure 4.1b). Note: pwell-2/deep nwell-2 breakdown was measured using the test structure designed similarly to the SPAD (Figure 4.1a) but with pwell-2 covering p^+.

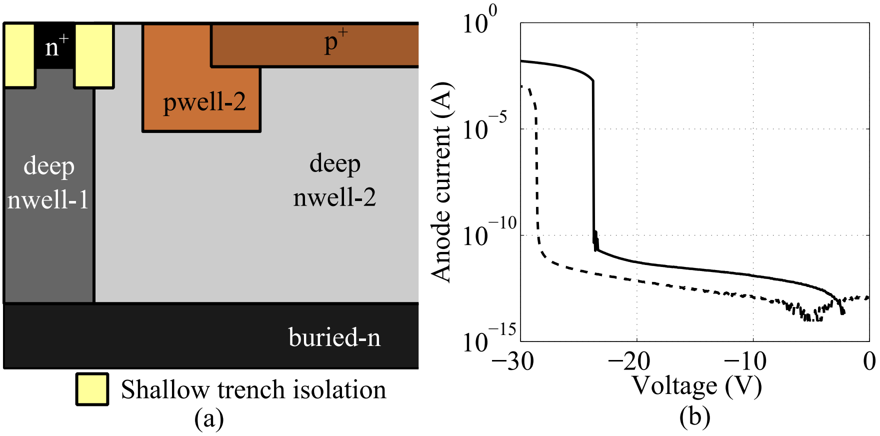


Figure 4.1: Guard region design-1: (a) one half of the device cross section, (b) I-V measurement results demonstrating the main junction and the guard junction breakdown.

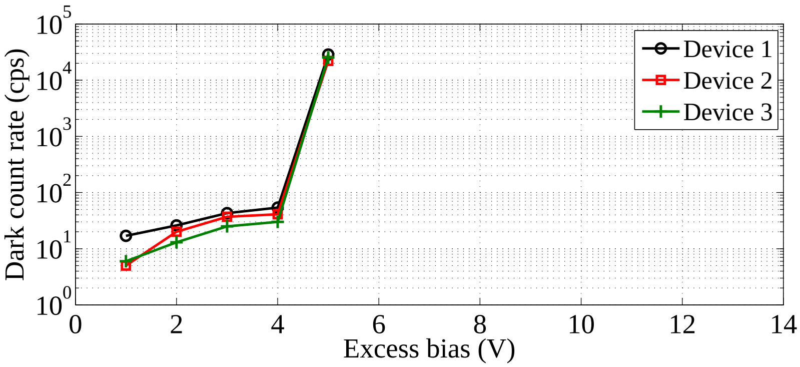


Figure 4.2: Guard region design-1: DCR characterization for three devices with the dead time tuned to be 400ns.

DCR measurements performed on three identically designed devices with 12μm diameter for the active area is presented in Figure 4.2. The results have shown a drastic increase in DCR after 4V excess bias. To comprehend, if the guard junction breakdown is indeed the cause of high DCR measured above 4V excess bias, design-2 with a higher guard junction breakdown was fabricated.

4.1.1.2 Design-2

In this design (Figure 4.3a) the deep nwell-2 is removed from beneath the pwell, and the electrical contact to the main junction's cathode (deep nwell-2) is provided using buried-n. With the use of pwell-1 relatively more lightly doped than pwell-2, and with a presence of a p-epitaxy layer between the pwell-1 and the buried-n, the guard junction breakdown has increased to 35.7V, which is 12.2V higher than the main junction (Figure 4.3b). Thus with the modified design, the SPAD can be operated until 12V excess bias without the guard junction entering into breakdown. Note: the guard junction breakdown measurement was performed on a test structure designed identically to the SPAD structure but without the deep nwell-2.

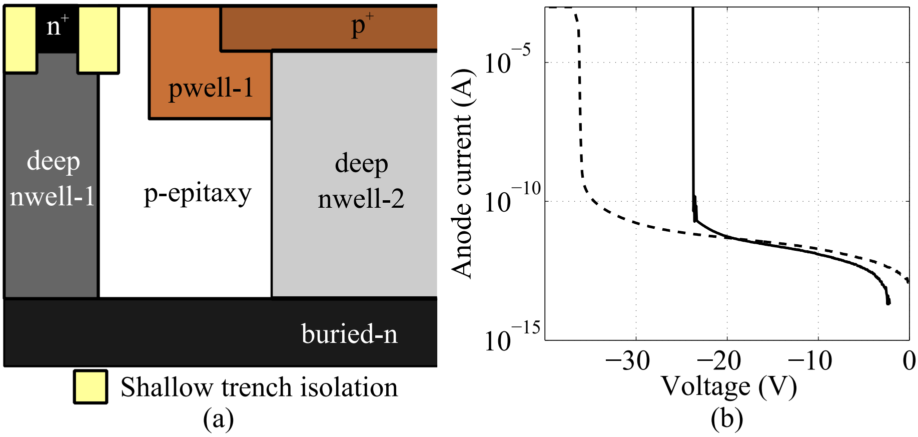


Figure 4.3: Guard region design-2: (a) one half of the device cross section, (b) I-V measurement results demonstrating the main junction and the guard junction breakdown.

The DCR characterization results for four devices designed with a 12μm diameter for active area is presented in Figure 4.4. When compared to design-1, design-2 has lower noise when operated above 5V excess bias. This observation suggest that the reason for high DCR observed in case of design-1 is due to the guard junction breakdown.

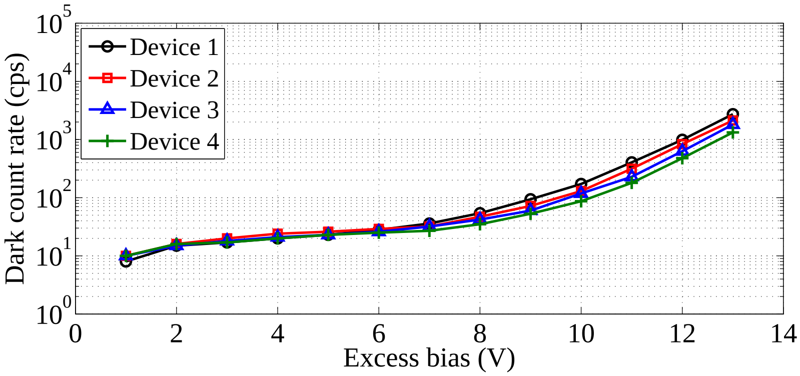


Figure 4.4: Guard region design-2: DCR characterization for four devices with the dead time tuned to 400ns.

Further, it needs to be noted that, design-2 when operated above its guard junction breakdown (at 13V excess bias) has 10 times lower noise than design-1 (at 5V excess bias). Design-1 has a DCR of 25.4k cps when operated at 0.7V above its guard junction breakdown (at 5V excess bias), and design-2 has a DCR of 1.8k cps at 0.8V above it guard junction breakdown (at 13V excess bias). The reason for the observed difference is due to the design. Electric field simulation performed on the guard region of device-1 and -2, at their respective breakdown voltages is shown in Figure 4.5.

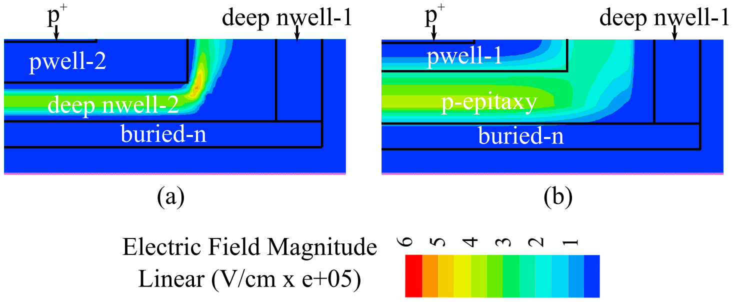


Figure 4.5: Electric field simulation result at guard junction breakdown: (a) design-1, (b) design-2.

In case of the design-1, a high electric field seen at the pwell-2 edges has resulted in tunneling. DCR measurements performed at various temperatures corroborate the simulations on tunneling, as it shows a higher DCR dependence on voltage than on temperature after 4V excess bias. Whereas, in the case of design-2, a high electric field seen at the pwell-2 edges in design-1 is avoided by the placement of a lightly doped p-epitaxy layer around pwell-1 and by utilizing the lateral diffusion of pwell-1 dopant. This design modification in the guard region has helped reduce tunneling noise. DCR measurements performed at the various temperatures have shown a relatively lower dependence on voltage than on temperature, when compared to design-1.

The results presented in this subsection imply that the SPAD DCR in addition to the main junction properties is also dependent on the guard junction design. Also from the results, it can be inferred that by designing the guard region with care, it is feasible to operate both the guard and the main junction simultaneously in Geiger mode. In the next chapter a dual junction SPAD designed to enable Geiger mode operation to both the guard and the main junction is presented.

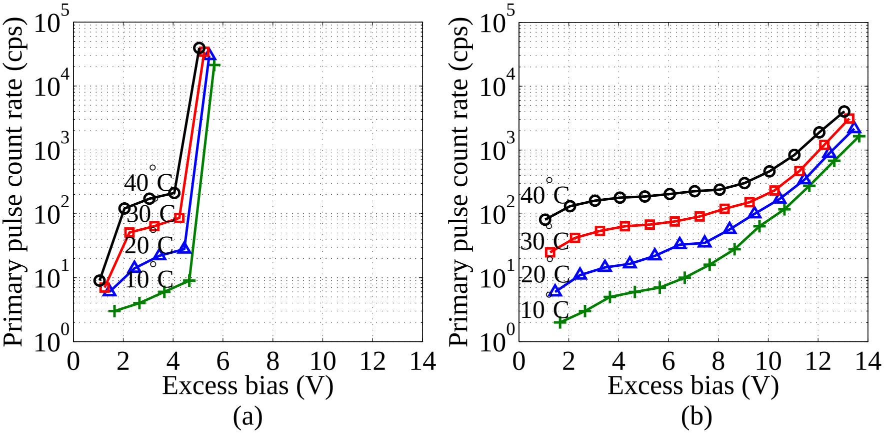


Figure 4.6: DCR characterization performed at various temperatures with the dead time tuned to 10μs: (a) design-1, (b) design-2.

4.1.2 Size

The design-2 presented in the previous subsection was shown to be effective until 12V excess bias. However, it needs to be noted that the SPAD reported in the previous subsection, was designed with a 4μm wide guard ring. A SPAD with a 12μm diameter for active area and a 4μm wide guard ring, will attain a fill factor of only 18.36%. To improve device fill factor, it is vital to reduce the insensitive area occupied by the guard ring.

When reducing the dimensions of the guard ring, it is important not only to assure the prevention of premature edge breakdown but also to be effective at all operating excess bias conditions. DCR measurements from four devices, designed with a 12μm diameter for the active area, and different guard ring widths (4, 3, 2 and 1μm) are presented in Figure 4.7. Measurements have shown that when reducing the guard ring size, the DCR has increased. For instance, at 6V excess bias, the DCR increased by 8.4k cps when reducing the guard ring width from 4μm to 1μm.

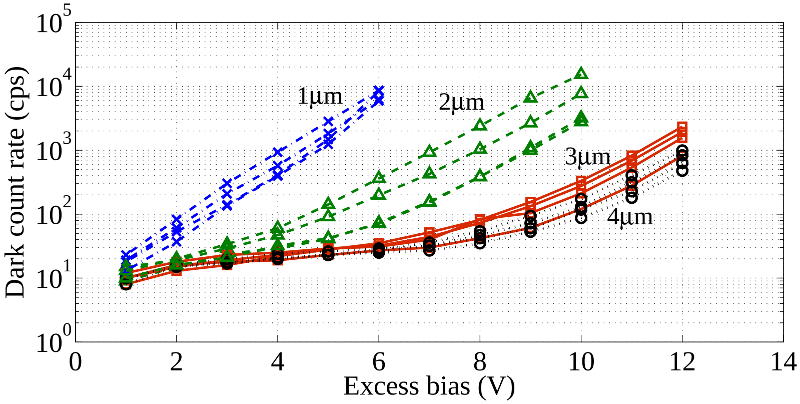


Figure 4.7: DCR characterization results for device designed with different guard ring width. Measurements for every design variant was performed on 4 devices with the dead time tuned to 400ns.

Electric field simulations (Figure 4.8) performed at 10V excess bias have shown that the p$^+$ edges are well protected when using a 4μm wide guard ring. However, when using a 1μm wide guard ring the depletion in pwell-1 has exposed p$^+$ edges to high electric field, which in principle can contribute to tunneling noise. DCR measurements performed at various temperatures are shown in Figure 4.9. When compared to a 4μm wide guard ring based device, the DCR of a 1μm wide guard ring based device is more dependent on voltage than on temperature, implying that the device with a 1μm wide guard ring is more heavily affected by tunneling contributions.

Thus for a guard ring to be effective at higher excess bias conditions a wider guard ring is required. A drawback in using a wider guard ring is the low fill factor. A technique to circumvent the use of a wider guard ring is to reduce the depletion in pwell-1 by increasing the pwell dopant concentration. A device similar to that presented in Figure 4.3, but with pwell-2 instead of pwell-1 was thus designed.

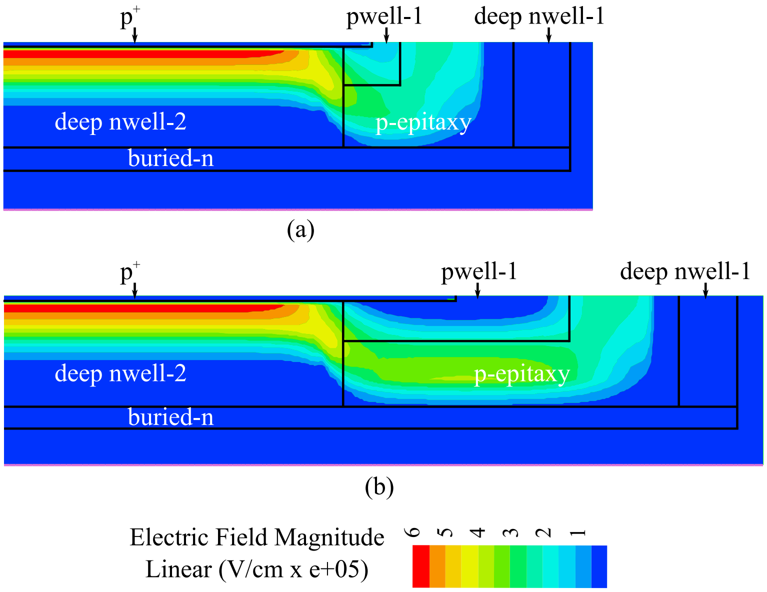


Figure 4.8: Electric field simulation result at 10V excess bias: (a) device with 1μm wide guard ring, (b) device with 4μm wide guard ring.

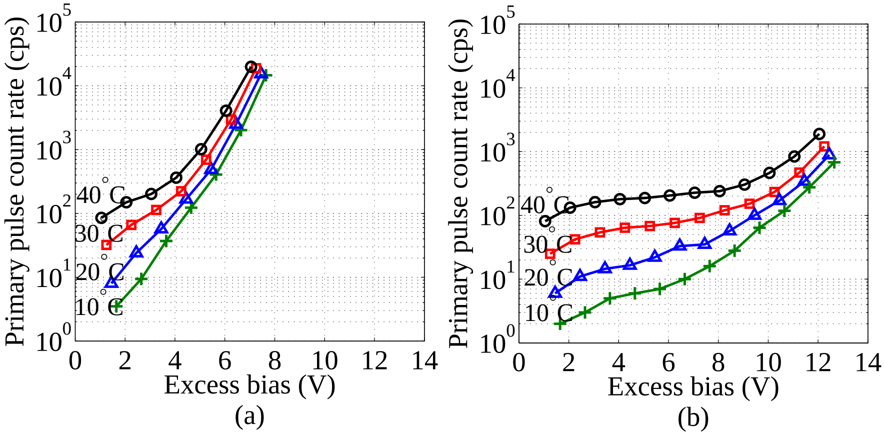


Figure 4.9: DCR characterization results performed at various temperatures with the dead time tuned to 10μs: (a) device with 1μm wide guard ring, (b) device with 4μm wide guard ring.

Note: pwell-2 is more highly doped than pwell-1.

The DCR measurement results for the device designed with 2μm wide guard ring and 12μm diameter for the active area are presented in Figure 4.10. As can be seen when compared to the devices designed with pwell-1, pwell-2 based designs have lower DCR at higher excess bias. However, it needs to be noted that when increasing pwell dopant concentration, the guard junction breakdown is reduced. For the presented design with pwell-2, the guard junction breakdown is just 0.2V higher than the main junction; implying that for all practical situations both the main junction and the guard junction are operated in Geiger mode simultaneously. This is the also reason why the pwell-2 based designs have slightly higher DCR than the designs with pwell-1 at a lower excess bias conditions.

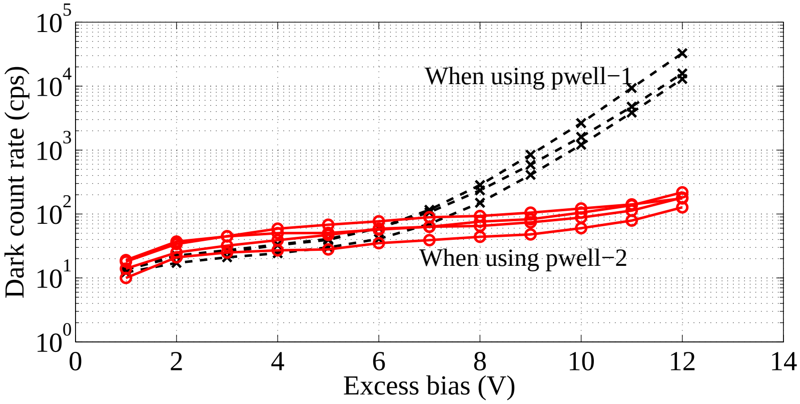


Figure 4.10: DCR characterization results for devices designed with different pwell implants. Measurements for every design variant was performed on 4 devices with the dead time tuned to 400ns.

4.1.3 Device curvature

The effectiveness of a guard region, in addition to the bias voltage, also depends on the properties of the p$^+$. In CMOS technologies, p$^+$ traits such as its doping, depth, dopant lateral and vertical diffusions are all fixed by process. However, at the design stage p$^+$ curvature can be varied either when realizing circular SPADs or when designing rectangular SPADs with rounded corners. This subsection discusses the guard ring effectiveness with respect to p$^+$ curvature.

To study the impact on DCR by p$^+$ curvature, different devices were fabricated with 4μm wide guard ring. Design specifications are listed in Table 4.1. DCR measurements normalized with respect to the active area are presented in Figure 4.11. The observed increase in DCR with active area size is in-line with the literature. However, above 2V excess bias, for devices designed with p$^+$ diameter less than 12μm, DCR increased when reducing the active area size.

Table 4.1: Design specifications for devices designed with different curvatures.

p$^+$/deep nwell-2 junction diameter (μm)	p$^+$ diameter (μm)	Guard ring width (μm)
4	8	4
8	12	4
12	16	4
16	20	4
20	24	4
24	28	4

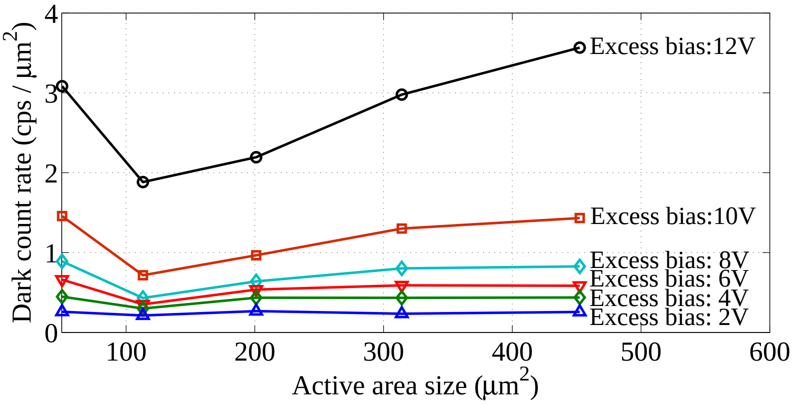


Figure 4.11: DCR normalized with respect to the active area for devices designed with different curvatures. Each measurement point is the mean DCR obtained from 4 devices when the dead time was tuned to 400ns.

We believe that the reason for the observed behavior is due to the guard region's ineffectiveness at higher p+ curvatures, resulting in more tunneling contribution. DCR measurements performed at various temperatures on two devices designed with 8μm diameter for p+ and 28μm diameter for p+ corroborate our expectation on tunneling, with a smaller diameter device being more dependent on voltage than on the temperature.

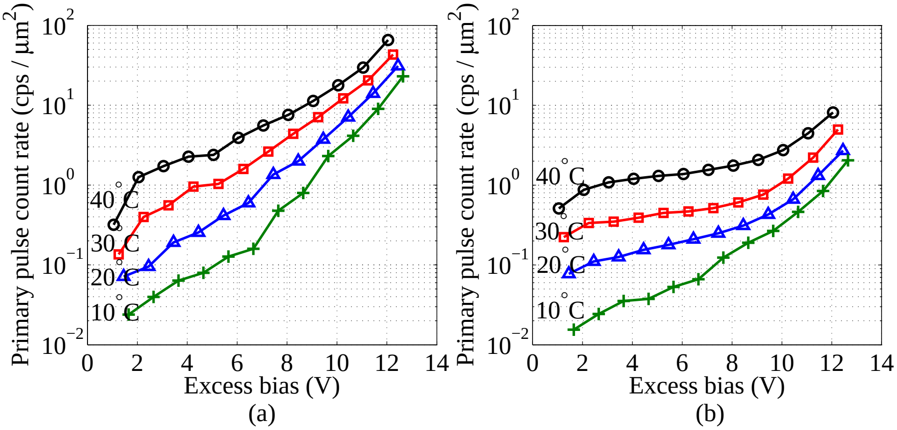


Figure 4.12: DCR characterization results performed at various temperatures with the dead time tuned to be 10μs: (a) device with 8μm diameter for p+, (b) device with 28μm diameter for p+.

4.2 Periphery

In this section, the impact on DCR due to the periphery design is presented. For the SPADs presented in Chapter 3, deep nwell-1 is used in periphery to provide contact to the cathode. In those designs, deep nwell-1 forms a parasitic junction with the pwell-1/pwell-2. The impact on DCR due to the parasitic deep nwell-1/p-epitaxy/pwell junction is studied for the p-i-n configuration. The p-i-n diode based SPAD was chosen, as it emulates the worst case scenario with high electric field beneath the pwell and close to the parasitic junction under study.

4.2.1 Design-1

Design-1 is the same as that of the p-i-n diode with pwell-2 as presented in Chapter 3. When optimizing the SPAD for fill factor it is vital to reduce the inactive space present between deep nwell-1 and pwell-2. In doing so, the field strength in the parasitic deep nwell-1/p-epitaxy/pwell-2 junction increases. High electric field could in principle influence the device DCR by injecting carriers either by tunneling or by activating some of the traps. To study the impact on reducing the distance between the pwell-2 and deep nwell-1, four different devices were fabricated with different distances between pwell-2 and deep nwell-1.

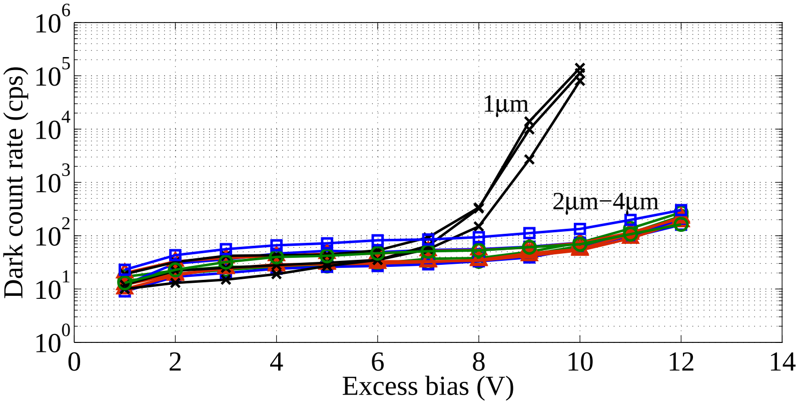


Figure 4.13: Design-1: DCR characterization results for devices designed with different spacing between pwell-2 and deep nwell-1. Measurements for every design variant was performed on 3 devices with the dead time tuned to 400ns.

Measurement results highlight that up until 12V excess bias, DCR remains unaffected when reducing the distance between pwell-2 and deep nwell-1 from 4μm to 2μm. However, at 1μm distance between pwell-2 and deep nwell-1, the DCR after 7V excess bias was measured to increase faster, when compared to the other design variants. Device simulations (Figure 4.14) performed with 1μm distance between pwell-2 and deep nwell-1, suggest the parasitic junction breakdown to be around 31V, implying that the high DCR observed when reducing the distance between pwell-2 and deep nwell-1 could be due to the parasitic junction breakdown.

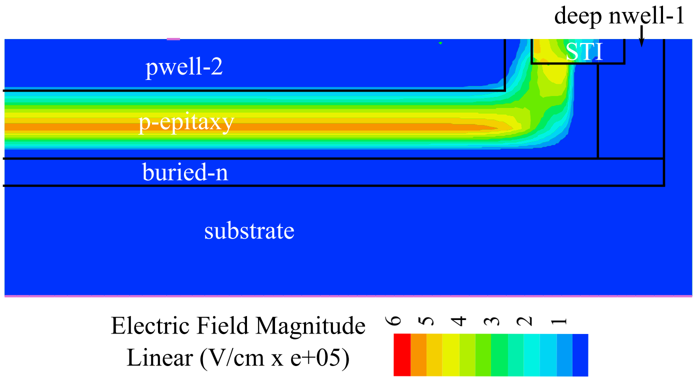


Figure 4.14: Electric field simulation result performed at the periphery junction breakdown when the spacing between pwell-2 and deep nwell-1 is 1μm.

4.2.2 Design-2

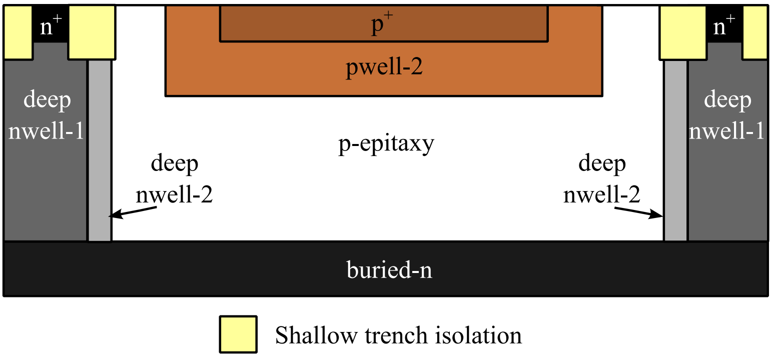


Figure 4.15: Periphery design-2, device cross section.

The modified design presented in Figure 4.15 utilizes a more lightly doped deep nwell-2 in the periphery. The lower doping when compared to deep nwell-1 has reduced the electric-field strength in the parasitic junction (Figure 4.16). The DCR results for the device designed with 1μm distance between the pwell-2 and deep nwell-2 is shown in Figure 4.17. As can be seen, when using the deep nwell-2 in the periphery, the DCR has reduced when compared to design-1. The results imply that in addition to the active area and the guard region also the device periphery can contribute to the SPAD DCR.

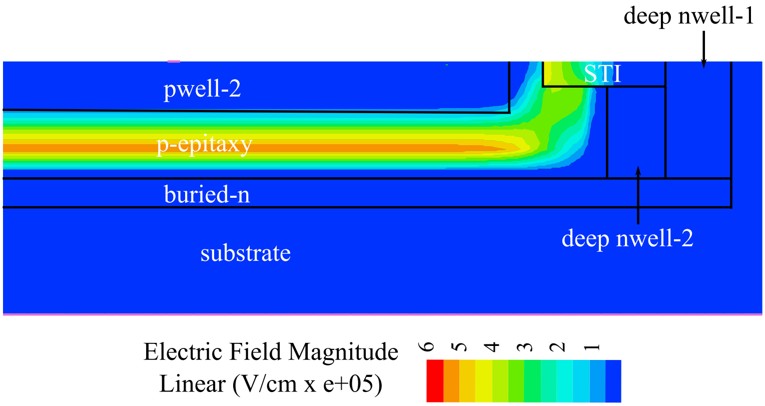


Figure 4.16: Electric field simulation result performed at the periphery junction breakdown when the spacing between pwell-2 and deep nwell-2 is 1μm.

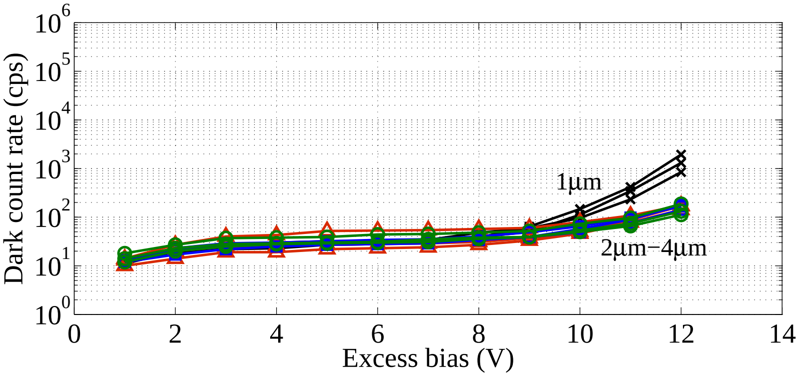


Figure 4.17: DCR characterization results comparing design-1 and design-2 when the spacing between pwell-2 and deep nwell-1 is 1μm. Measurements for every design variant were performed on 4 devices with the dead time tuned to be 400ns.

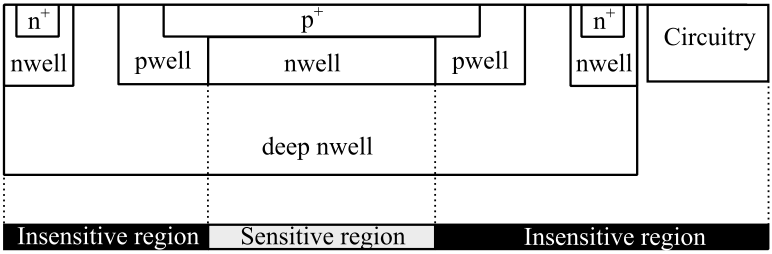
4.3 Summary

We studied the impact of the guard region and the periphery structures on the performance of the SPADs by making assumptions and subsequently verifying them by means of specific designs. From the experiments we formulated several principles listed hereafter can serve as guidelines for optimization of SPADs.

- In addition to the active area SPAD DCR is also influenced by the guard region and the device periphery.
- DCR originates from the guard region either due to the guard junction breakdown or due to the complete depletion of the guard ring or due to the device curvature.
- Influence on DCR due to the guard junction breakdown depends on the guard region properties.
- Depletion in guard region can be reduced either by increasing the guard region dimensions or by increasing the dopant concentration of the guard layer implant.
- DCR that originates from the periphery can be reduced either by increasing the distance between the active area and the periphery or by reducing the periphery's dopant concentration.

Chapter 5

Fill factor optimization



SPAD fill factor is affected by the presence of inactive guard region, periphery and circuitry. This chapter presents two design technique, focused on improving CMOS SPAD fill factor.

The first technique presented in Section 5.1 emulates the functionality of a optical microlens. The integrated microlens exploits the avalanche propagation phenomenon. In contrast to an optical microlens, the proposed technique is more robust to manufacturing tolerances and can in principle scale to large arrays, providing a better concentration uniformity. Measurement results show a PDP as high as 64.85% at 600 nm when the SPAD is biased at 2V of excess bias. The microlens configuration is compatible with backside illumination and the use of integrated devices on the guard rings, thereby enabling the high density pixels and extremely high fill factors.

The second technique discussed in Section 5.2 is based on a novel dual junction SPAD. In this design, the second junction realized in the guard region of the single junction SPAD has resulted in a fill factor improvement of 22.2% and device sensitivity improvement of around 50%. To reduce circuit complexity, the design ensures substrate isolation, while the overall operation is similar to that of a conventional single-junction SPAD. Also the proposed dual junction SPAD is suitable for PDP engineering. In this design, two junctions having the same breakdown voltage but with different PDP profiles are designed to operate simultaneously, so as to maximize the sensitivity spectrum. Using the proposed PDP engineering technique, the device sensitivity profile was modified during the design phase by adjusting the ratio between two junction active areas. An empirical model was formulated to estimate the PDP *a priori* and it was shown to closely match the measurements.

5.1 Electrical microlens

In a conventional SPAD design, guard region, periphery, and the readout circuitry limit the pixel fill factor. To recover the loss of incident photons in the guard region, microlenses or prisms have been used, so as to concentrate light from insensitive to sensitive areas of the pixel. However, due to manufacturing tolerances, concentration non-uniformity and reproducibility issues, robust solutions have not been achieved for arrays larger than 1k pixel [67]. Alternatively, backside-illuminated (BSI) designs [97] have been proposed for APDs and SPADs. However, it is not clear at the time of the writing of this thesis, whether BSI will be a viable option for SPADs in the near future.

This section, addresses the challenges associated with optical microlens, with a concentration technique that is not optical but electrical in nature. The technique is based on lateral avalanche propagation (LAP) [80, 98, 99]; its purpose is to enable photocarriers generated in the guard region, usually subject to recombination, to trigger Geiger pulses. In contrast to drift, that moves photocarriers along the field lines without multiplication, LAP acts along the junction perpendicular to the electric field (parallel to the surface), thus enabling radial movement of carriers towards the center of the structure. In the past, the use of LAP has been limited to the active area only, for position-sensitive SPADs [100] and fix-position noise reduction [101]. In this section, we expand the use of LAP far beyond the active area for acquiring photocarriers from below the guard ring regions and from other deep-well regions hosting CMOS circuits.

5.1.1 Design

To enable the creation of an avalanche under the guard ring region, the SPAD is designed to operate near the guard region breakdown voltage i.e. the breakdown voltage of the guard ring is 2V higher than that of the multiplication region [102]. Further, to achieve identical spectral response everywhere, the guard ring region and multiplication regions were designed to have virtually the same depth. The proposed design is shown in Figure 5.1. The figure depicts the cross section of the device.

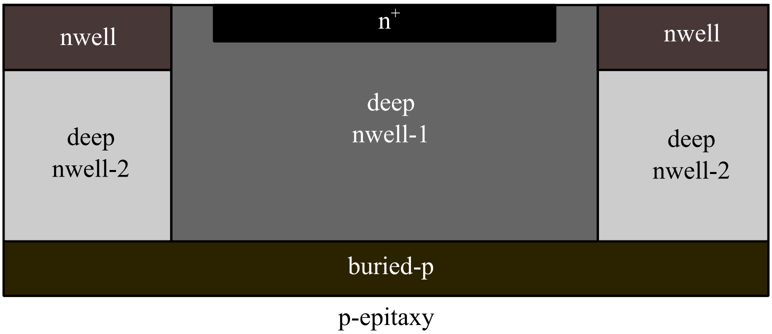


Figure 5.1: Electrical microlens: device cross section fabricated in 180nm CMOS technology.

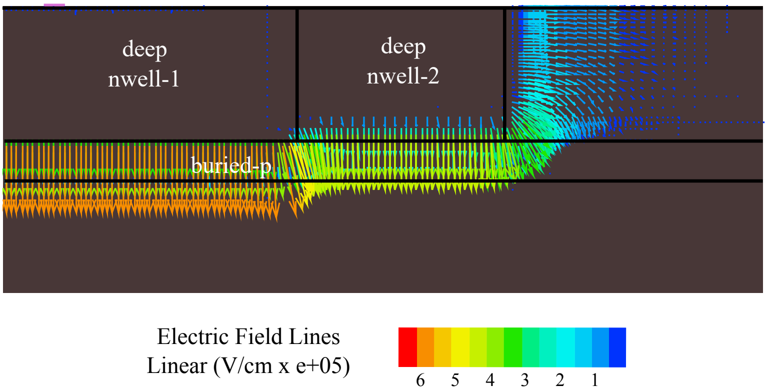


Figure 5.2: Electrical microlens: TCAD electric field simulation performed at device breakdown.

The device was carefully designed using the appropriate semiconductor layers and junctions to achieve adequate electric field profiles. Electric field simulations were performed using MEDICI. Figure 5.2 shows the amplitude and direction of the electric field in the proposed device, as a result of a simulation.

To validate the proposed technique, a device with a guard ring (deep nwell-2) width of 4μm was fabricated with an active area (deep nwell-1) diameter of 12μm. The device breakdown voltage is 26.4V.

5.1.2 Characterization

5.1.2.1 Experimental setup

The devices were operated in Geiger mode using external quenching and recharge circuitry. The designed circuitry provided fixed quench and recharge time thereby eliminating any impact on dead time. The circuitry is shown in Figure 5.3. A fast comparator detects an avalanche event and a feedback loop controls the recharge of the SPAD providing enough current to recharge the parasitic capacitance present at the cathode of the SPAD. The anode is placed at a negative voltage corresponding to the breakdown voltage, while the excess bias voltage is applied on the other pin of the quenching resistor.

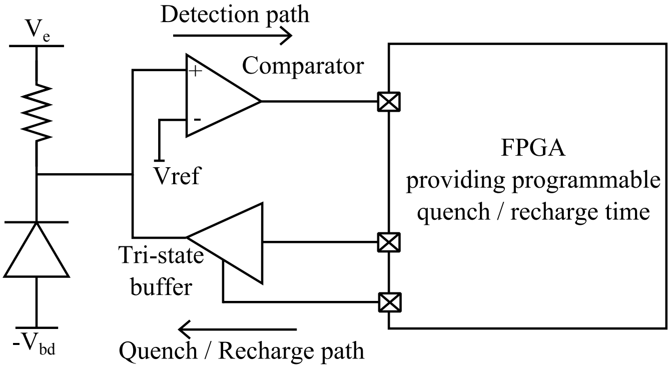
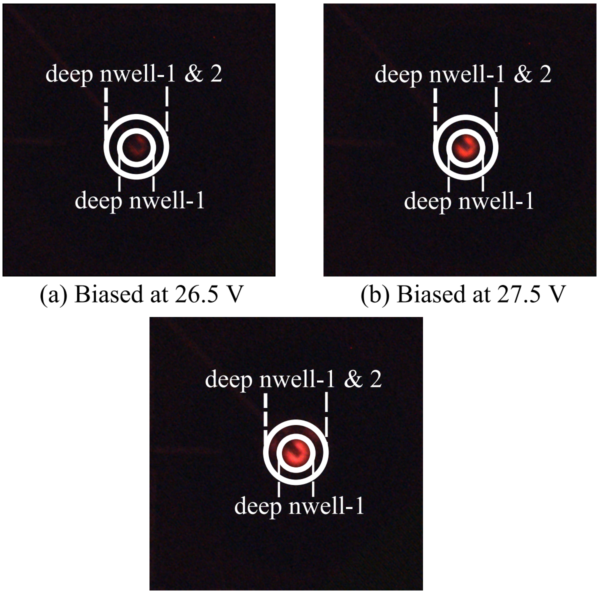


Figure 5.3: Electrical microlens: experimental setup.

5.1.2.2 Light emission test

Further, to study the spatial distribution of high electric field region, light emission tests were performed at various bias conditions. Light emission test results (Figure 5.4) have confirmed that at the device's breakdown voltage, a high electric field region resides only in the designed active area (beneath deep nwell-1). At little above 2V of excess bias, as expected, the high electric field region spreads to the guard ring area (beneath deep nwell-2).



(a) Biased at 26.5 V

(b) Biased at 27.5 V

(c) Biased at 28.5 V

Figure 5.4: Electrical microlens: light emission test

5.1.2.3 Dark count rate

The DCR measurements performed at various temperatures and at various excess bias conditions show a stronger dependence of DCR on excess bias, than temperature. This observation suggests that the major contribution to DCR originates due to tunneling.

5.1.2.4 Afterpulsing probability

The afterpulsing probability measurements were carried out as described in Chapter 2. In these measurements the device was used along with external circuitry tuned to provide 10μs of quench time. The experimental results at various excess biases are presented in Figure 5.6.

Although the measured afterpulsing is 7% or higher, it is to be noted that in the current setup the quench and the recharge circuitries are external, with a high parasitic capacitance due to bond pads and wires. We expect that the integration of external circuitry on to silicon can reduce the afterpulsing probability.

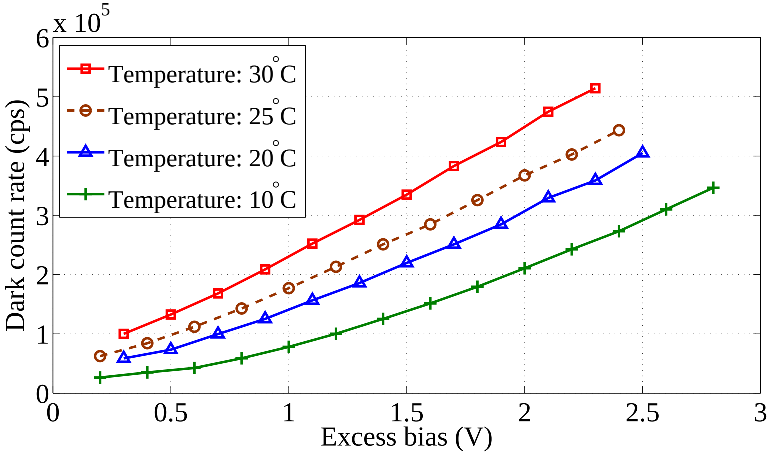


Figure 5.5: Electrical microlens: DCR characterization performed at various temperatures with the dead time tuned to 10μs.

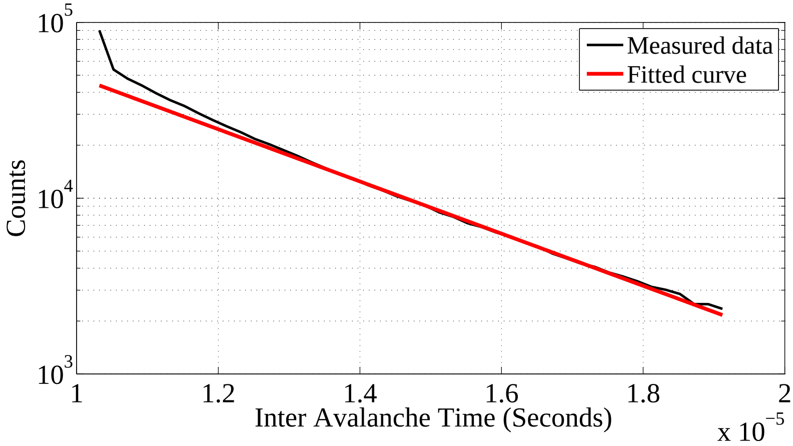


Figure 5.6: Electrical microlens: afterpulsing probability measurement results at 2V excess bias, performed with the device deadtime configured to 10μs.

5.1.2.5 Photon detection probability

The PDP measurement results are reported in Figure 5.7. The presented results were obtained considering 12μm (deep nwell-1 size) for active diameter and with afterpulsing compensation. As expected, the PDP increased drastically from 35.24% to 64.85% at a wavelength of 600nm, when biased at 1.5V (below guard region breakdown) and at 2.0V (at guard region breakdown) respectively. The measured result confirms the fact that, when biased at 2.0V of excess bias, the region beneath the guard ring is activated leading to photon sensitivity. As expected, the PDP increase observed here is much greater than the normal increase due to excess bias.

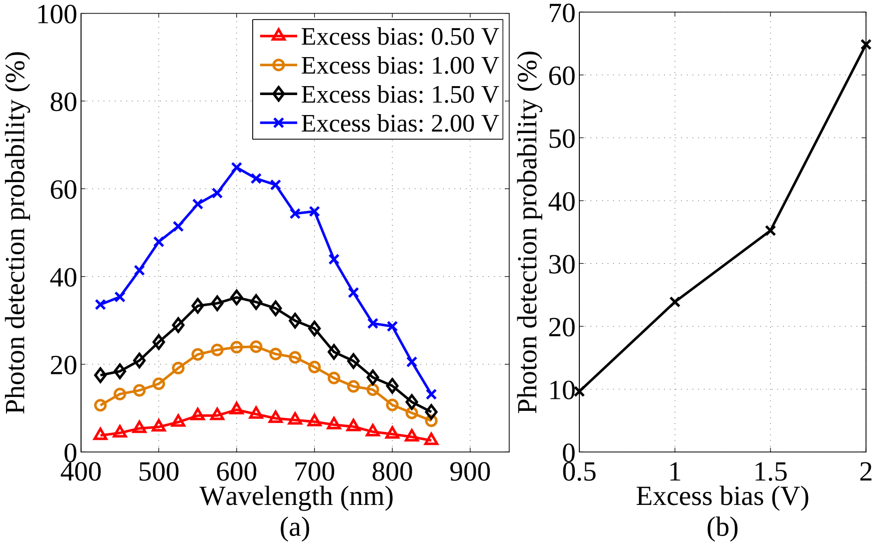


Figure 5.7: Electrical microlens: (a) PDP vs wavelength, (b) PDP at 600nm wavelength for various excess bias voltages.

5.1.3 Evolution

In this section, it was demonstrated that the space occupied by the guard ring can be recovered when operating the SPAD near the guard junction breakdown. As the junctions were formed deeper into the substrate, it can in principle permit the use of transistor on its top. A device fabricated with the nmos transistor on the top of the guard junction is presented in Figure 5.8. Due to process modifications, the designed SPAD was not operational due to a very high dark noise. The change in process is cited as the reason because of the measured change in breakdown voltage for this device when compared to one presented in Figure 5.1. Note: these two devices were fabricated in different tapeout. However, we believe that, on successful operation of both the transistor and the SPAD at the same time, will enable the realization of the next generation high density SPAD arrays.

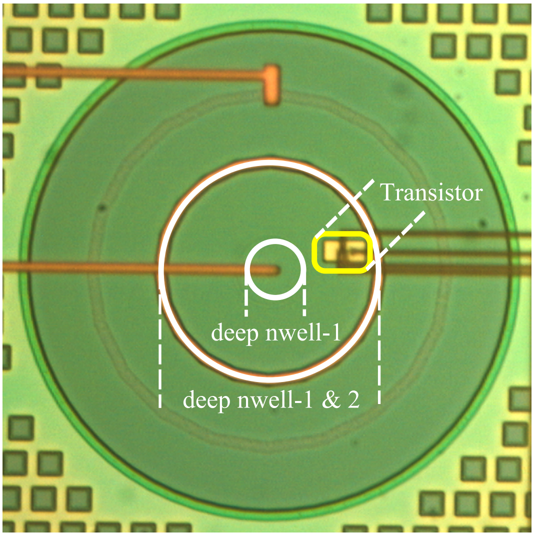


Figure 5.8: Electrical microlens: integrated nmos transistor on the guard ring.

5.2 Dual junction SPAD

The design reported in this section extends the concept presented in the previous section by also enabling Geiger mode operation in the guard ring junction. The dual junction SPAD realized utilizing the insensitive area occupied by the guard region to realize the second active junction has improved the fill factor by 22.2% when compared to the single junction SPADs [3]. In this configuration, a lower DCR of 0.44 cps/μm^2 was achieved at 10V excess bias. The proposed dual junction device, in contrast to conventional design of stacking two junction [103, 104], enables full substrate isolation and an operation similar to that of a single junction SPAD, while simultaneously minimizing electrical crosstalk.

In contrast to SPADs designed in custom technologies [2,19–28], CMOS SPADs are inherently limited to available implant/diffusion layers [30,53,55–57,81]. When the availability of the implant layers are limited or when doping concentrations cannot be modified, the possibility to tune the device spectral response is strongly curtailed.

The device design reported in this section enables spectral response modification without having to change any of the implant/diffusion doping profiles. The presented technique uses two junctions that have distinct and partially complementary spectral responses. In this configuration, the overall PDP is modified by properly engineering the ratio between the junction's active areas. To understand this effect, a model that estimates the PDP as a function of the junction characteristics was shown to match measurements closely. Accurate modeling enabled us to design SPADs with predictable, custom made PDP profiles, without the use of customized layers and dedicated semiconductor technologies. This technique is referred to as the *PDP engineering*.

5.2.1 Design

Figure 5.9 shows the cross section of a circular device designed in 180nm CMOS technology, where two junctions namely p$^+$ / deep nwell-2 and pwell-2 / p-epitaxy / buried-n are fabricated adjacent to each other. In this design, junction-2 (pwell-2 / p-epitaxy / buried-n) is designed around junction-1 (p$^+$ / deep nwell-2); it utilizes p-epitaxial layers as its guard ring. In this configuration, a lateral diffusion of pwell-2 aided by the presence of a lightly doped p-epitaxial layer avoids edge breakdown. Buried-n, designed as a cathode for junction-2, is extended throughout the device, thus enabling substrate isolation and the contact to deep nwell-2.

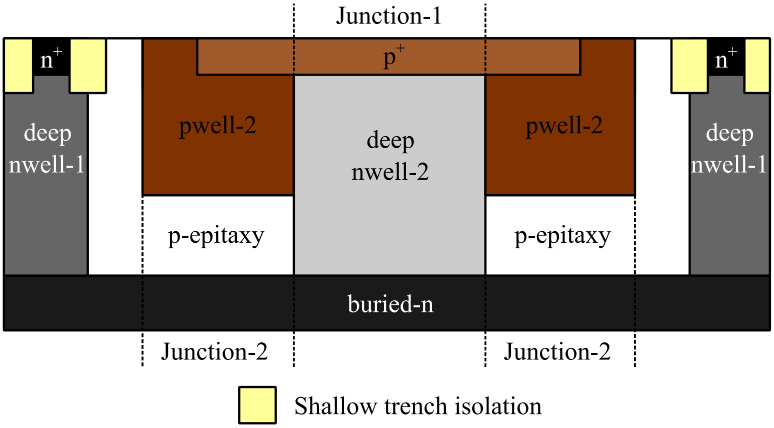


Figure 5.9: Dual junction SPAD: device cross section - fabricated in 180nm CMOS technology.

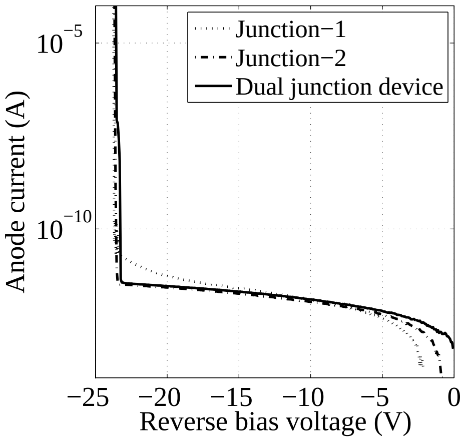


Figure 5.10: Dual junction SPAD: breakdown voltage measurement.

To facilitate two-terminal device operation, anodes and cathodes of two junctions are connected using p^{+} and buried-n respectively. Further, to ease device operation, the junctions are designed with almost identical breakdown voltage. For the designed device, a breakdown voltage of 23.5V and 23.44V was measured for junction-1 and 2 respectively (Figure 5.10). Measurements were performed on two test structures. Test structure-1, designed to characterize p^{+}/deep nwell-2 junction is identical to that of a SPAD presented earlier in [3]. Test structure-2, used for characterizing pwell-2/p-epitaxy/buried-n junction is similar to the presented design (Figure 5.9) but without deep nwell-2.

5.2.2 Characterization

5.2.2.1 Experimental setup

Two-terminal device operation, along with almost identical junction breakdown, has resulted in a device that can be operated as any other conventional SPAD. The device was operated using external active quench and recharge circuitry as discussed in Chapter 2. The experimental setup is shown in Figure 5.11. FPGA enables programmable quenching and recharge time with a minimum attainable dead time of 300ns.

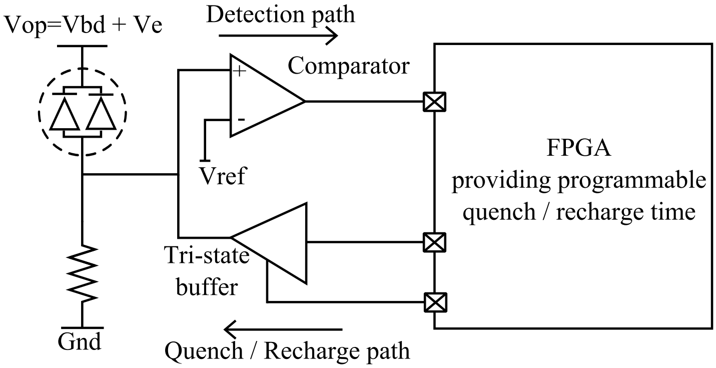


Figure 5.11: Dual junction SPAD: experimental setup.

5.2.2.2 Photon detection probability

The PDP measured for a circular device designed with a 12μm diameter for junction-1 and 2μm width for junction-2 is presented in Figure 5.12a.

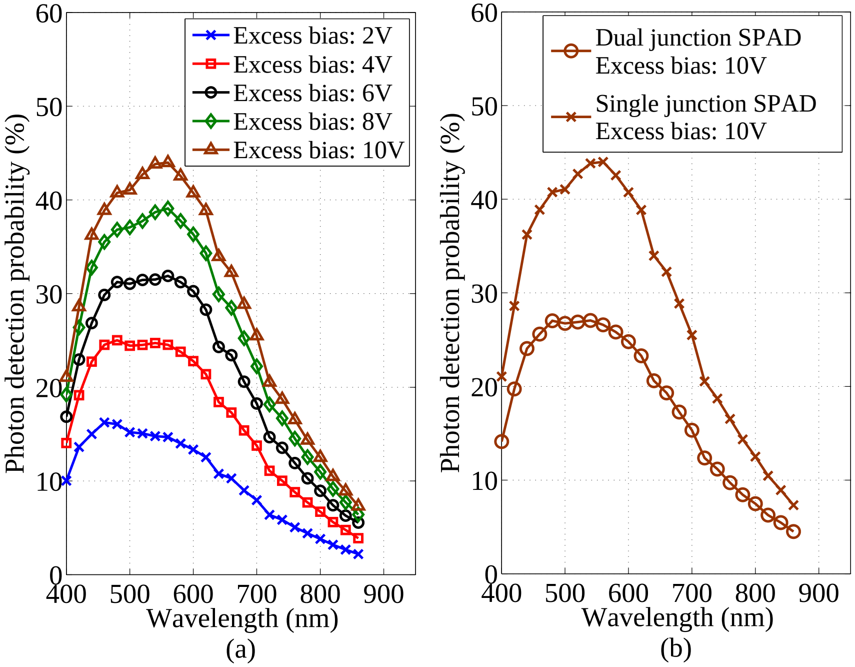


Figure 5.12: Dual junction SPAD: (a) PDP for a device designed with 12μm diameter for junction-1 and 2μm width for junction-2. (b) PDP for a single junction SPAD [3]. Note: In (a) and (b) PDP was calculated considering the device active area to be 16μm diameter (12μm from junction-1 + 4μm for pwell-2).

For this device a PDP greater than 40% was measured from 480nm to 600nm at 10V excess bias. In Figure 5.12b the PDP results at 10V excess bias are compared with the single junction SPAD [3], which is similar to that in Figure 5.12b. For comparison, the PDP of a single junction SPAD was calculated considering the device active area to be 16μm in diameter (12μm from junction-1 + 4μm for pwell-2). The observed increase in device sensitivity by around 50% is due to the realization of the second active junction in the area originally used by the guard ring [3]. This design facilitates control on the PDP profile. In Section 5.2.3, PDP engineering techniques are discussed, along with the PDP estimation model and its validation.

5.2.2.3 Timing jitter

Timing jitter measurements performed using a 637nm pulsed laser source are presented in Figure 5.13. At 10V excess bias, the device achieved a FWHM response of 103ps. In contrast to conventional single junction devices, the dual junction SPAD presented in this section has two distinct peaks originating from the two junctions.

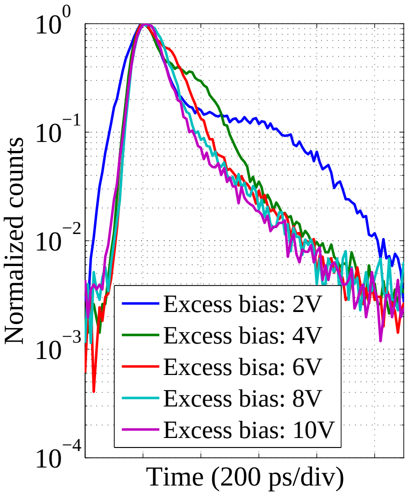


Figure 5.13: Dual junction SPAD: timing jitter measurement result when using 637nm laser.

The origin of the two peaks was studied in detail using two circular devices designed with a diameter of 12μm for junction-1 and junction-2 sized to be 2μm (device-1) and 1.5μm (device-2). The results obtained when using a 637nm laser source at 2V excess bias are presented in Figure 5.14 a. Comparing the timing measurement results for two devices, it can be seen that the second peak shifted in time has decreased in magnitude as the width of junction-2 is reduced, suggesting that the origin of second peak is indeed junction-2.

To further substantiate this inference, measurements performed using red and blue lasers were compared in Figure 5.14 b. For these measurements, the device was designed as a circle with a diameter of 12μm for junction-1 and a width of 2μm for junction-2. Junction-2, formed deeper in silicon, has lower sensitivity to blue photons than to red photons. As expected, the second peak was found to reduce in size when using a blue rather than a red laser. The observed results confirm that

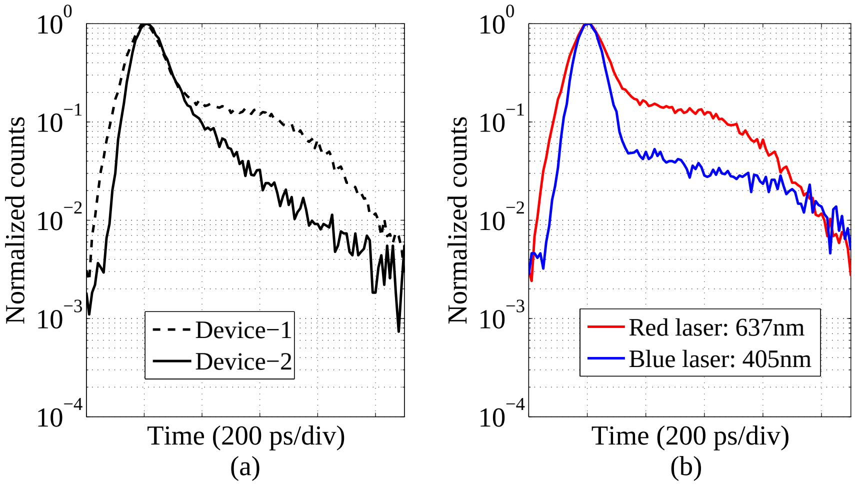


Figure 5.14: Dual junction SPAD: (a) comparison of timing measurement results of device-1 and device-2 at 2V excess bias when using 637nm laser, (b) comparison of timing measurement results of device-1 when using 637nm and 405nm laser.

the two peaks in Figure 5.13 are caused by the presence of two junctions placed laterally.

Although the observed peaks could deteriorate the device timing performance at low excess bias, it could in principle be used to distinguish the photons collected from junction-1 and -2, thereby enabling colour differentiation or in realization of position sensitive SPADs.

5.2.2.4 Dark count rate

The DCR characterization was performed for a device with a diameter of 12μm for deep nwell-2 and 2μm width for pwell-2.

DCR characterization involving primary and secondary pulses was performed on four devices at 25°C and presented in Figure 5.15. For this measurement, a device dead time of 300ns was chosen to include both primary and secondary pulses in DCR measurements. Under such experimental conditions, a DCR of 18 cps and 88 cps was observed at 2V and 10V excess bias, respectively. When compared to state-of-the-art dual junction SPADs, the observed noise performance is superior to the vertically stacked junctions based designs [103] and [104].

The **Afterpulse characterization** performed using the inter-avalanche time histogram technique, is presented in Figure 5.16.

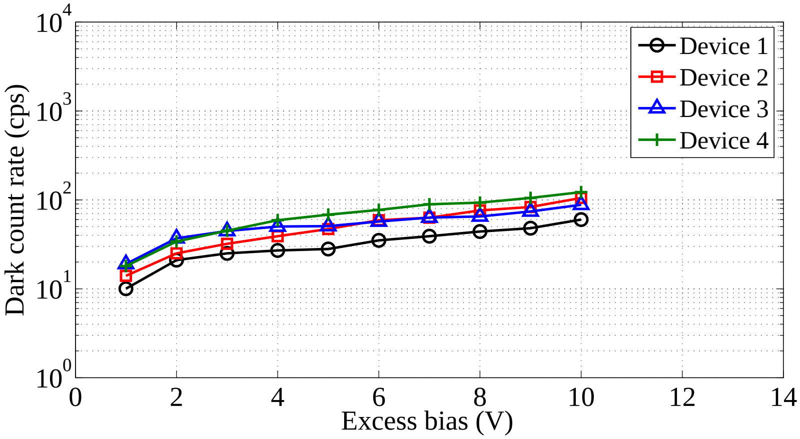


Figure 5.15: Dual junction SPAD: device-3 DCR characterization for four devices with the dead time tuned to be 300ns.

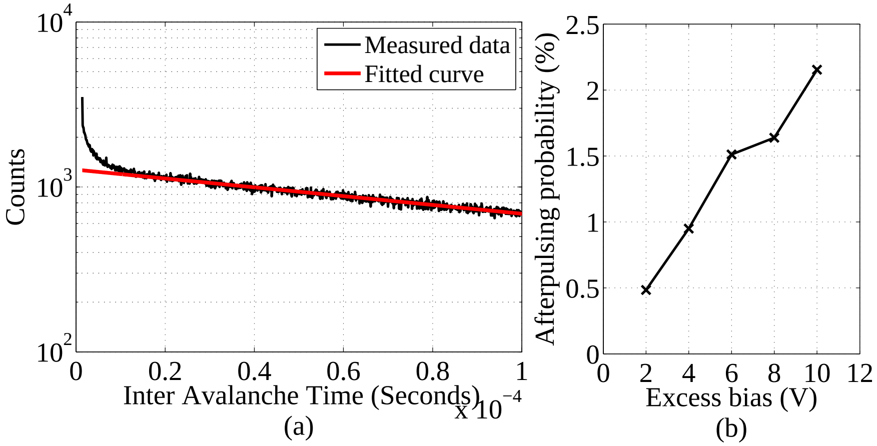


Figure 5.16: Dual junction SPAD: afterpulsing probability measurement results, performed with the device deadtime configured to 300ns, a) afterpulsing experimental results at 10V excess bias, b) afterpulsing probability measured at various excess biases.

The measurement result shows an increase in afterpulsing probability from 0.48% at 2V excess bias to 2.14% at 10V excess bias. The observed increase in afterpulsing is attributed to an increase in carrier trapping probability resulting from an increase in charge flow at high excess biases. It has to be noted that the measurements were performed using the experimental setup described earlier, when the device dead time was tuned to be 300ns.

Primary pulse characterization performed at various temperatures (Figure 5.17a) suggests that a major contributor to noise is SRH and not the tunneling. For this measurement, a device dead time of 10μs was chosen to ensure negligible afterpulsing.

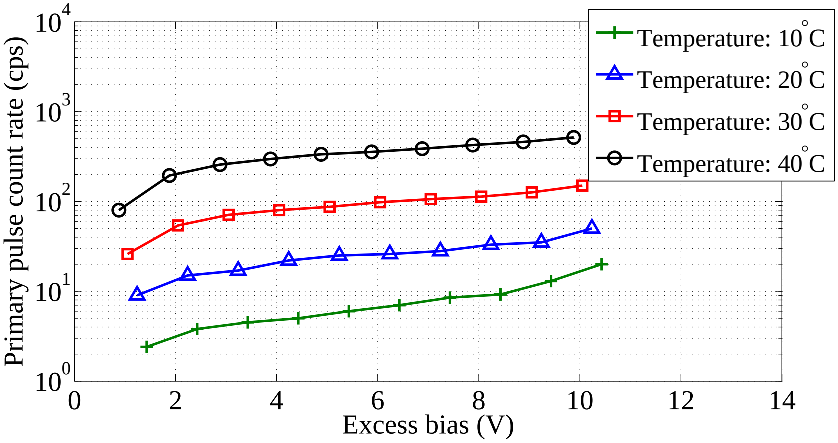


Figure 5.17: Dual junction SPAD: Device-3 Primary DCR characterization performed at various temperatures with the dead time tuned to 10μs.

5.2.3 PDP engineering

For a SPAD designed with two active junctions as shown in Figure 5.9, the overall device PDP can be modified by adjusting the ratio between two junction's active area. Figure 5.18 shows the PDP for a device designed with 6μm diameter for junction-1 and 2μm width for junction-2. For this device, the ratio between junction areas (junction-1: junction-2) is 0.56. Comparing PDP with the device presented in Section 5.2.2.2 where the ratio between two junctions was 2.77, it can be seen that the device with higher junction-2 area has a PDP profile shifted towards green-red region.

This empirical observation was studied to understand the underlying mechanism through a PDP model that was validated with the experimental results. The remainder of this section discusses the model and its validation.

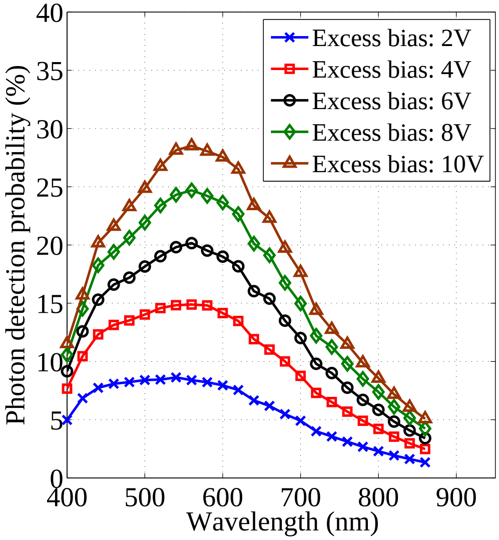


Figure 5.18: Dual junction SPAD: PDP for a device designed with 6μm diameter for junction-1 and 2 μm width for junction-2.

5.2.3.1 PDP estimation and modeling

For a dual junction SPAD, as proposed in this section, the overall device PDP can be modelled as an area-weighted sum of contributions from two junctions (Equation (5.1)).

$$PDP = \frac{A_{J1}}{A_T}.PDP_{J1} + \frac{A_{J2}}{A_T}.PDP_{J2}, \tag{5.1}$$

A_{J1} - Junction-1 active area
A_{J2} - Junction-2 active area
A_T - Junction-1 active area + Junction-2 active area
PDP_{J1} - Junction-1 PDP
PDP_{J2} - Junction-2 PDP

Equation (5.1) assumes PDP to be uniform across a junction's active area. In practice, the lateral diffusion of pwell-2 and deep nwell-2 leads to non-uniformities in PDP at the junction periphery. The electric field simulation performed using MEDICI confirmed the impact of dopant diffusion, as the cause of a reduction in field strength at the periphery of deep nwell-2 and along the inner edges of pwell-2. The model incorporating dopant diffusion is presented in Equation (5.2).

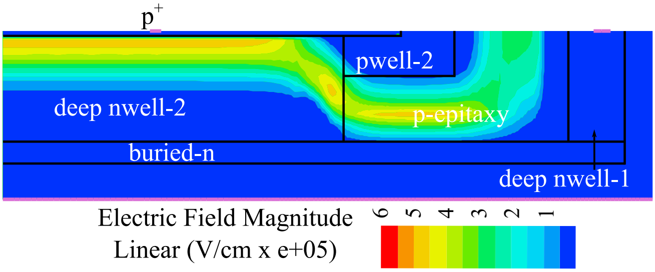


Figure 5.19: Dual junction SPAD: TCAD electric field simulation performed at breakdown.

$$PDP = \frac{A_{JE1}}{A_T}.PDP_{J1} + \frac{A_{JE2}}{A_T}.PDP_{J2} + \frac{A_{diff}}{A_T}.PDP_{diff} \tag{5.2}$$

A_{JE1} - Junction-1 effective area after deducting diffusion affected area

A_{JE2} - Junction-1 effective area after deducting diffusion affected area

A_{diff} - Diffusion affected region area

PDP_{diff} - Diffusion affected region PDP

Though Equation (5.2) is comprehensive, in practice it is not feasible to determine precisely the PDP of the diffusion affected region (PDP_{diff}) or its area (A_{diff}). To incorporate dopant diffusion in the model, the terms representing diffusion affected regions are combined, p$^+$/deep nwell-2 junction, and a segment of the pwell-2/p-epitaxy/buried-n junction. These terms are combined into a single term $PDP_{primarydevice}$, resulting in Equation (5.4). $PDP_{primarydevice}$ represents a primary device PDP designed to be similar to the one for which the PDP is being estimated. Thus, using Equation (5.4), the device PDP can now be estimated by extrapolating the primary device's PDP.

$$PDP = \left(\frac{A_{J2} - A_i}{A_T}.PDP_{J2}\right) + \left(\frac{A_i}{A_T}.PDP_{J2} + \frac{A_{JE1}}{A_T}.PDP_{J1} + \frac{A_{diff}}{A_T}.PDP_{diff}\right) \tag{5.3}$$

$$PDP = \left(\frac{A_{J2} - A_i}{A_T}.PDP_{J2}\right) + PDP_{primarydevice} \tag{5.4}$$

A_i - Junction-2 area of primary device

5.2.3.2 Model validation

Three circular devices fabricated in 180nm CMOS technology were used for validating the proposed model. The devices were designed with junction-1 sized 12μm in diameter, and junction-2 width as 2μm (device-1), 1.5μm (device-2) and 1μm (device-3). PDP measurement results for device-1 and 3, along with the estimated PDP for pwell-2/p-epitaxy/buried-n junction, are presented in Figure 5.20. pwell-2/p-epitaxy/buried-n junction PDP estimation was performed by solving Equation (5.4), when using PDP data from device-1 and 3.

The model of Equation (5.4) was validated for device-2 considering device-3 as primary device. The estimated PDP is compared with measurements in Figure 5.21 for various wavelengths and excess biases. Though the estimated value matched the measured data closely, a gradual shift in estimated PDP from measurements was observed when increasing the excess bias. The reason for the observed difference is the impact on PDP from the outer edges of pwell-2, which the model does not take completely into account. Device-to-device variation could also be a reason for the estimation error.

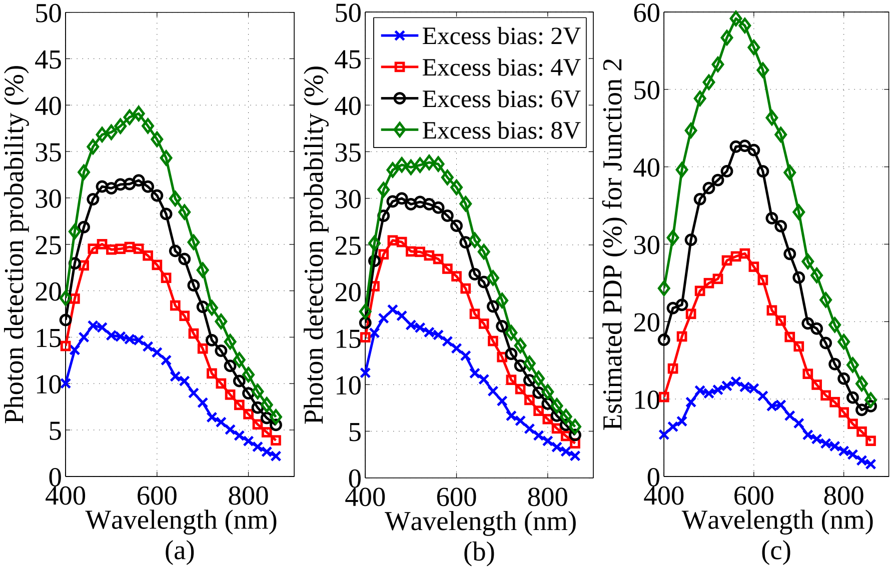


Figure 5.20: Dual junction SPAD: (a) device-1 PDP (b) device-3 PDP and (c) estimated junction-2 PDP.

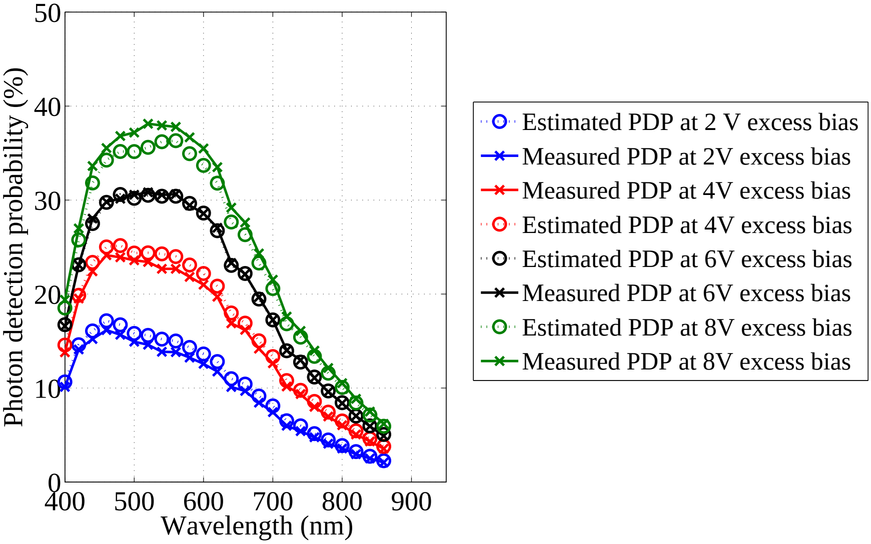


Figure 5.21: Dual junction SPAD: estimated and measured PDP for device-2.

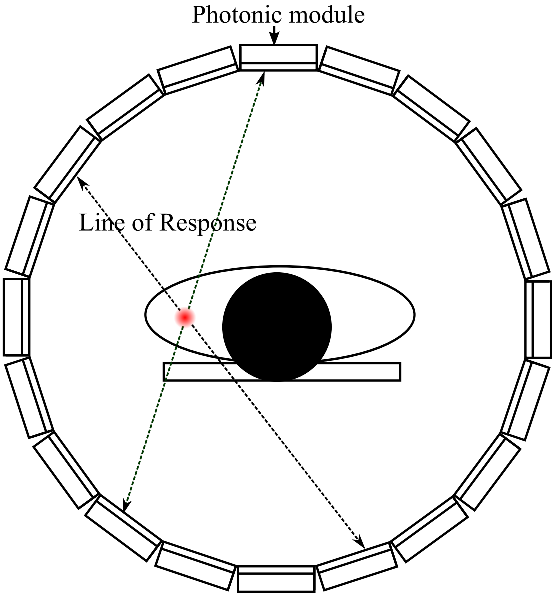
5.3 Summary

We studied ways of increasing fill factor in SPADs using alternatives to optical microlens arrays. The study was completed by means of several designs that exposed the causes of various non-idealities in designs. From the study we deduced a number of rules that can be used to guide the design of SPADs with improved fill factor.

- Electrical microlens design avoids the need for periphery by enabling contact through substrate.
- Electrical microlens uses lateral avalanche propagation to collect photocarriers from the guard region, by operating the device near the guard junction breakdown.
- Dual junction SPAD was designed by operating the guard junction and the main junction above breakdown simultaneously.
- Dual junction SPAD design improved device fill factor by 22.2%, when compared to the conventional single junction SPAD.
- Dual-junction SPAD design enables substrate isolation and operation similar to that of a conventional SPAD.
- Dual junction SPAD has two peaks in the timing jitter measurements, originating from two junctions.
- Dual junction SPAD enables device PDP engineering, without requiring any modification to the process.

Chapter 6

SPADs in PET: A Case Study



The work presented in this Chapter was carried out within the framework of SPADnet project[1]. The goal of the SPADnet project [46] was to build a PET system using SPAD sensors.

[1] European Union funded FP7 project (http://www.spadnet.eu/)

PET is a medical imaging technique used mainly in cancer diagnostics. For PET, prior to imaging the patient is injected with fludeoxyglucose (FDG). FDG accumulates in the region of higher metabolic activities. The radio active isotope ^{18}F attached to the FDG undergoes a positron emission decay inside the body. The emitted positron, annihilates with an electron to result in two gamma photons of 511keV, that are almost 180^0 apart to each other. The photonic modules placed in multiple rings around the patient detect the two gamma events, also referred to as the true coincidence or true events. The line of response (LoR) formed between two gamma events is collected for multitude of gamma photon pairs. The image reconstruction algorithm such as the filtered back projection used on the collected data to obtain the spatial map of higher metabolic activity regions such as cancer cells.

In PET the photonic module is made of scintillator and photonic sensor. Scintillator blocks the incoming gamma photon to result in visible light photons which is then detected by the photonic sensors. Photonic sensors used in PET have evolved over the years from PMT [11] to PSPMT [105] (position sensitive photo multiplier tube) and to analog SiPM [106] (silicon photo multipliers), addressing the need for higher pixel granularity and timing response, while ensuring MRI compatibility [107] and reduced size/cost. More recently, there has been growing interest in using CMOS integrated SPAD based sensors or digital SiPMs [4, 38, 108].

The inherent digital property of SPADs and their migration to deep sub-micron CMOS processes have enabled the realization of digital photonic sensors with built-in intelligence. Deep-submicron CMOS SPADs, along with 3-D integration, will lead to highly granular pixel arrays capable of time stamping individual photons, the ideal sensing solution for PET systems. Digital SiPM based sensors [4,38,108], when used in PET systems, will produce a large amount of data for each potential gamma event, prompting the need to handle the generated data efficiently and accurately. Further, when using digital SiPMs, multiple rings of photonic modules are required to scale the dimensions of preclinical, clinical and brain PET systems. This chapter focuses mainly on techniques to handle the challenges associated with data acquisition when hundreds of digital SiPMs are used in the PET system.

6.1 DAQ challenges and the proposed approach

A preclinical PET system, when designed using 5cmx5cm SPADnet sensor tile, will generate around 420 Gbps of data every second (Table 6.1). In brain and in clinical PET modalities generated data rate will be around 262.5 Gbps and 525

Gbps (Table 6.1). Comprehending the data acquisition challenges, two techniques were considered for data reduction.

1. *Estimating gamma event properties close to the sensor tile:* PET image reconstruction algorithms require energy, timing and scintillation coordinates for every gamma event. Estimating these parameters, close to the sensor tile, will reduce data transfer requirements significantly (Table 6.2).

2. *Implementing a noise filtering technique in the PET system:* of the detected gamma events only the true events are required for image reconstruction. True events contributing to less than 10% of the total events [6], can be filtered using well established techniques such as pile-up reduction, energy windowing and coincidence detection [6]. Implementing noise filtering techniques in the PET system will further reduce the data transfer rate to a manageable level (Table 6.2).

Table 6.1: Raw data rate analysis when using SPADnet sensor [4].

PET Modalities	Configuration	Raw data /gamma event (kb)	Gamma event /photonic module /second	Expected data rate (Gbps)
Preclinical	2 rings of 10 modules	4.2	5×10^6	420
Brain	5 rings of 25 modules	4.2	0.5×10^6	262.5
Clinical	5 rings of 50 modules	4.2	0.5×10^6	525

Table 6.2: Implication of data rate reduction.

PET Modalities	Raw data rate (Gbps)	Data rate after estimation (Gbps)	True event data rate (Gbps)
Preclinical	420	6.4	0.64
Brain	262.5	4.0	0.4
Clinical	525	8.0	0.8

6.2 Photonic module construction

The entire SPADnet photonic module [46] is depicted in Figure 6.1.

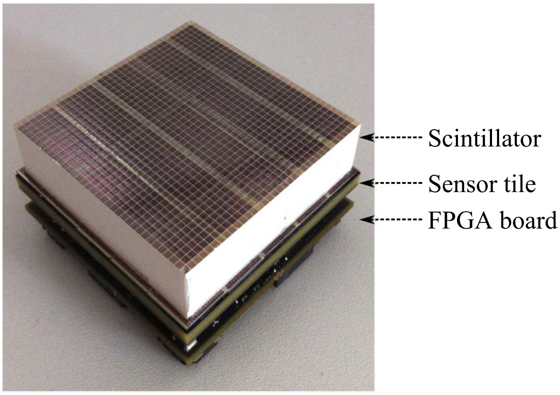


Figure 6.1: SPADnet photonic module.

In SPADnet 25 arrays of 16x8 mini SiPMs are tightly abutted on a single PCB to form a sensor tile [46]. High sensor concentration is enabled by TSV (through silicon via) connections from the front-side of each sensor device to their backside, replacing conventional wire bonding [46]. The sensor tile is then interfaced to an FPGA based PCB on its back, where the data processing and communication Unit (DPCU) is designed to reside [46]. The SPADnet photonic module comprising the sensor tile and DPCU will function as an autonomous sensing, computing and communication unit.

6.3 Networked approach

To facilitate scalability to different PET modalities namely preclinical, clinical and brain, a sensor network based approach is used. In the SPADnet system, every photonic module acts as a sensor node [109] (Figure 6.2). DPCU, placed beneath the sensor tile, acquires sensor generated data and processes it into data packets comprising the estimated values of energy, timing and spatial coordinates of the scintillation. Data packets after passing through pile-up reduction and energy windowing filters, are then communicated to the network. The network performs coincidence detection and true event transfer in real time to an external computer for reconstruction. The remainder of this chapter focuses on the design and realization of the network.

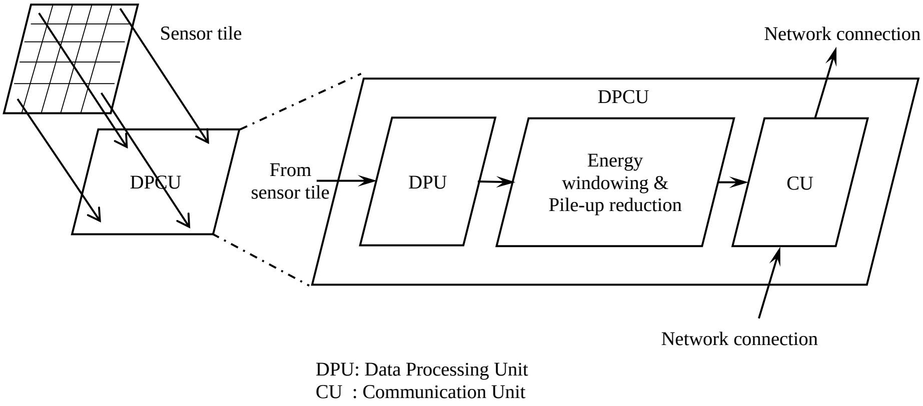


Figure 6.2: Data flow from photonic module to network.

The network architecture and the techniques discussed from here on can be used with any digital SiPM sensors, provided the data processing, pile-up reduction and energy windowing are designed specifically for that sensor type.

6.4 Network topology

In a multi-ring based PET system, the photonic module acting as a network node is connected to its neighbours, forming a cylindrical mesh topology (Figure 6.3). In the proposed network, in addition to the photonic modules, a functional node called snooper is included in every ring. In this configuration, one of the snooper node designated as a master, acts a bridge between the network and the PC, where the other snooper nodes aid master node functionality. In this network, a high-speed bidirectional serial communication link is used for inter node communication, due to its compactness and high data rate. The snooper-to-PC communication is established using giga bit Ethernet connectivity.

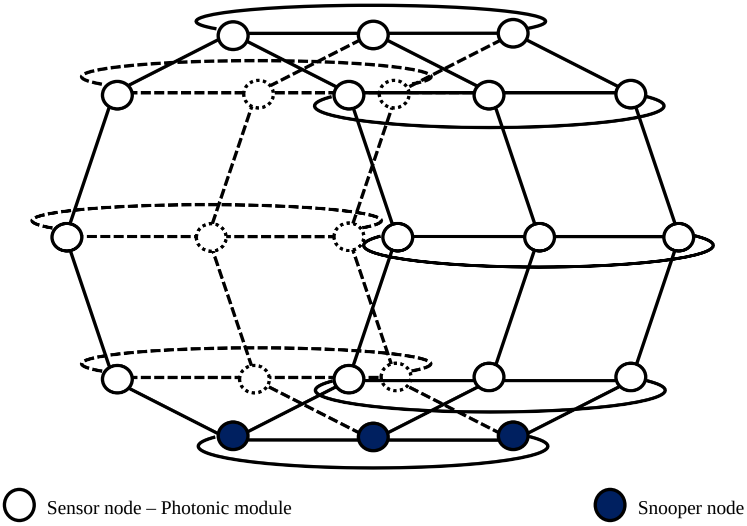


Figure 6.3: Proposed cylindrical mesh topology.

6.5 System level description

The network is designed to perform the following tasks

1. True event acquisition
2. Singles acquisition
3. Raw data acquisition
4. Sensor/node configuration

The singles and raw data acquisition are included to perform sensor tile test and characterization.

6.5.1 True event acquisition

For true event acquisition, the coincidence detection needs to be performed on the network. Coincidence detection [6] is performed by identifying two gamma events, that are detected within a short time window. To tag the detected coincidence events as a true event pair, the identified gamma events have to be within the field-of-view of the photonic module that has detected its pair [6].

Considering the complexity involved in performing search operation in identifying true events, the coincidence detection is generally performed using a dedicated processor [110–112]. To perform such complex tasks in a network, two possible approaches are feasible, namely, single point coincidence detection [112] and distributed coincidence detection [113–115].

1. In *single point coincidence detection*, data packets generated in various nodes across the network are communicated to the snooper. The snooper then performs the coincidence detection. The disadvantage of this approach is the snooper design complexity, and its scalability to different PET modalities.

2. In *distributed coincidence detection*, all network nodes are designed to perform coincidence detection simultaneously by comparing the locally generated events with the events that traverse the network nodes. When compared to single point coincidence detection, distributed coincidence detection is scalable as it distributes the work load across an arbitrarily high number of nodes.

Since scalability is the main motivation to use the network for data acquisition, the distributed coincidence detection technique has been chosen.

6.5.1.1 Distributed coincidence detection

For distributed coincidence detection, a search-algorithm needs to be implemented in every node, to identify whether the arriving packet is in coincidence with any of the events detected in that node. The implementation of an extensive search algorithm across the history of all the events detected in that node is impractical. Hence, a search space is reduced by designing the network to provide a lower, well-defined packet latency (or lower variance in packet latency).

Lower variance in packet latency will ensure that the packets arrive at a specific node within a certain time period after being detected. Thus, using this approach it is possible to update the event search space continuously, by erasing the events that are detected before the expected packet latency. Implying, that when using this technique the event search space will depend on the packet latency, and the efficiency of this technique will depend on the latency variance. Thus to perform the distributed coincidence detection effectively, the network is required to provide lower packet latency and its variance.

6.5.1.2 Two stage coincidence detection

To reduce packet latency and its variance, the true event pair is formed in two stages in the network. In the first stage, coincidence detection is performed. In the second stage, the detected coincidence events are paired to form a true event pair. To aid two-stage data transfer bi-directional network is partitioned into two unidirectional channels, namely the coincidence detection channel and the data transfer channel. In this approach, coincidence detection is performed using the dedicated communication link, unhindered from the other packet types. The reminder of this subsection elaborates the two-stage data transfer.

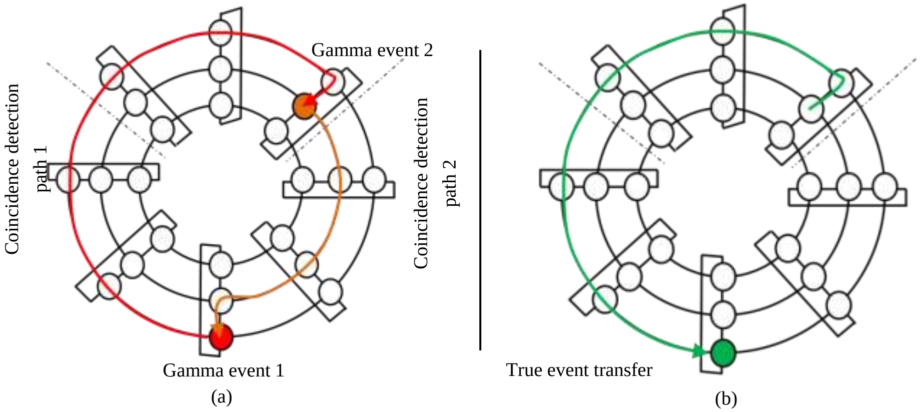


Figure 6.4: Two Stage Coincidence Detection: (a) stage 1 - Coincidence detection, (b) stage 2 - Coincidence pair formation.

Stage 1 - Coincidence detection: in case of coincidence detection to reduce packet latency smaller data packet of 32 bits carrying the Gamma event timing information and node ID is circulated in the coincidence channel. As the packet traverses the field-of-view requirements a copy of the same is made to circulate the network in the axial direction. Given the field-of-view requirements the distributed coincidence detection unit present in every DPCU performs coincidence detection by comparing the received information on the gamma event's timing with its own event history. Upon successful detection of the coincidence pairs, the event present in history is tagged accordingly, and after a certain time the tagged true event is communicated to the data transfer channel.

Stage 2 - Coincidence pair formation: the data transfer channel, upon receiving the tagged event, transfers the packet either to its pair's or it holds it until its pair arrive. The implemented network logic arbitrates on which packet to be transferred and which to wait for depending on the packet timing information. The arbitration logic presented in Algorithm 1 assures equal utilization of network resources across all nodes, by randomizing the selection of the node that transfers its packet. Once the pair has been formed using the data transfer channel, it is then transferred to the PC via the snooper. It should be noted that the data transfer channel is designed to act in the opposite direction as opposed to the coincidence channel, to help enable the nodes monitor the status of their neighbors using node status packets.

Algorithm 1 Arbitration logic when forming true packet pair.

```
procedure SELECTION OF COINDENCE DETECTED PACKET
    if Gamma event timing is a odd number then
        Transfer packet from node with bigger node ID
    else
        Transfer packet from node with smaller node ID
    end if
end procedure
```

6.5.2 Raw data acquisition

To test and characterize sensor tile, the network was designed to acquire raw data from every photonic module individually. Raw data generated in a photonic module is packetized into a set of smaller data packets, and is then transferred to the snooper using the data transfer channel. The snooper collects the network packet, packetizes it into a bigger raw data packet and then transfers it to the computer for further processing.

6.5.3 Singles acquisition

Further, to test the PET system, the network is also designed to perform singles acquisitions. A singles packet comprising the estimated values of energy, timing and scintillation coordinates of the gamma event is transferred to the PC via the snooper. Also for this mode of operation, a data transfer channel is used for communication.

As the singles data rate can be higher than the Ethernet capacity, there could arise a situation when the snooper is not able to transfer data packets to the PC. In such situation, some gamma events need to be dropped. When dropping packets it is vital to ensure no gamma event is lost while its pair is being maintained in the network. Hence a flow control strategy as detailed in the next section is used.

6.5.4 Sensor/node configuration

To facilitate sensor/node configuration, a set of registers are included in every node. The registers provided with read and write access are controlled via the network using the data transfer channel from the PC.

For the *register write operation*, up on receiving the command and the data from the PC, snooper transfers it to the node.

For the *register read operation*, up on receiving the command from the PC, the snooper sends a read request packet to the node. The node then responds with a set of data packets comprising configuration register information to the snooper. The snooper then communicates it to the PC.

6.6 Packet handling techniques

The network partitioned into two data channels handles seven different packet types. The coincidence channel handles the node status and the coincidence packet. The data transfer channel handles the node status, configuration, coincidence detected, true event, singles and raw data packet.

This section discusses techniques such as packet routing, flow-control and scheduling schemes that permit various packet types to coexist in a single data channel effectively. Since the two channels do not interfere with each other, they are treated as two standalone networks from the perspective of the packet routing, flow-control and scheduling schemes.

6.6.1 Routing strategy

In a network, the packet routing algorithm defines a path for a packet to travel from a origin node to its destination node. For the network presented in this chapter, static routing technique is used for all data packet types. In static routing, a packet entering the network will always follow a fixed path, irrespective of the network status. The static routing scheme was chosen as it is simple to realize. Further, when compared to the adaptive techniques, the chosen scheme requires unidirectional communication links and it is not prone to live-locks. Live-locks in principle can deteriorate packet latency performance.

For the cylindrical mesh topology, the packets are routed first along the radial axis until they reaches the axial ring of their destination node, and then they travel in the axial direction to their final destination. The used routing strategy is similar to the conventional X-Y routing scheme [116].

6.6.2 Scheduling algorithm

In a network, the scheduling algorithm is used to decide on a packet that needs to be transferred to the next node. For the proposed network, the scheduling algorithm is designed to function in two stages. In the first stage, the packet type that needs to be transferred is selected. Then, in the second stage, a specific packet is chosen from a selected packet type.

For the coincidence channel handling coincidence and status packets, the status packet is given the highest priority. Further, for the status packets, first-come-first-serve policy is used for scheduling. Whereas, for the coincidence packet the oldest-packet-first scheduling scheme is used to improve latency performance. The status packets were given the highest priority to facilitate the network decide on the packet flow using the network status.

For the data transfer channel handling 7 different packet types, the packet priority is assigned in the following order:

1. Status packet
2. Configuration packet (CP)
3. Coincidence detected packet (CDP)
4. True event, raw data and singles packet

The status packet is given the highest priority as it dictates the flow control of the other packet types. The configuration packet is given the second highest priority to facilitate PC control nodes even when network is congested by other packet types. The coincidence detection packet is given higher priority when compared to the DAQ packets (true event, raw data and singles), to improve latency performance when forming a true event pair. Last the true event, raw data and singles packets are given the same priority, because at any point in time the network will be configured for only one mode of operation.

For status, configuration, true event, raw and singles packet, a first-come-first serve policy is used in selecting a packet for transfer. For coincidence detected packet, oldest-packet-first scheduling scheme is used. To aid the design realization of the above discussed scheduling schemes, a virtual channel (VC) [117,118] based approach is used. A VC is a first-in first-out buffer, placed along the packet flow. In the proposed design, every packet type is assigned to at least one VC. Using this design strategy a packet can traverse the network unhindered from the network congestion that could arise from the lower priority packet types.

Further, in the design multiple VCs are assigned for the coincidence packet and for the coincidence detected packet types, to reduce the design complexity in finding the oldest packet. In the proposed scheme, the incoming packets are assigned to various VCs in round robin fashion, with a special VC assigned to the packets originating in that specific node. In this configuration, the oldest packet is determined by comparing the timing of the packet present at the top of every VCs.

6.6.3 Flow control

The network flow control logic decides when to transfer a packet to the next node. For the proposed network an adaptive flow control technique is used. For a network functioning as a data acquisition system, the packet dropping probability is a critical performance metric. In case of the PET application, it is critical to ensure that all nodes maintain equal packet dropping probability during the entire data acquisition. This condition will ascertain that at any point in time, no gamma event is lost while its pair is being processed in the network. To ensure equal packet dropping probability, an adaptive flow control strategy was devised.

In the proposed network, DAQ packets (true-event, singles and raw data) are transferred to the next node only when the receiving node's network resource utilization is less than the current node. To aid this process status packets are transferred to the neighbouring nodes periodically. The flow control technique used for DAQ packets will aid the network achieve almost equal utilization of buffer occupancy in every node. Note: buffers are the VCs included in every node and for every packet type.

Whereas coincidence packets and coincidence detected packets are transferred only if the receiving nodes buffer VC occupancy is less than 80%. In situations when a node's buffer VC occupancy for CPs or CDPs reaches 80% or above, a status packet is broadcast to all nodes to stop acquiring any new gamma event. Under such situation the node that raised the stop flag needs to send a restart command to all nodes to start acquiring a gamma event. A re-start command is sent only when the buffer VC occupancy of the involved node reach below 10%.

Further, for intra-node communication a protocol was devised to allow transfer of data packets only when the receiving module within a node is free to receive it. This protocol aids the inter-node data flow in shutting down the gamma event acquisition when network is not able to accept any more events.

6.7 Network start-up

At the network start-up, all nodes are required to establish the communication link with their neighbours. To assure the network is operational, the communication links are established progressively.

In the designed start-up strategy, first at the power up, DPCU establishes axial communication links. From then onwards starting from the snooper node, the link between the axial rings is setup one after the other in clockwise direction until all links in the radial axis are operational. Finally, the snooper nodes establish their axial communication links, starting from the master snooper node. In this strategy if any of the nodes fail to setup a link, then one or more of the communication links in the master snooper node are not operational. Thus by monitoring the status of the master snooper node, it is feasible to know the network start-up status. The initialization algorithm used for the DPCU and for the snooper node is presented in Algorithm 2 and 3.

Algorithm 2 DPCU start-up algorithm.

```
procedure NETWORK START-UP
    Activate → EAST, WEST and SOUTH channel
    while EAST, WEST and SOUTH Channels are SETUP do
        Activate → NORTH Channel
    end while
end procedure
```

Algorithm 3 Snooper start-up algorithm.

```
procedure NETWORK START-UP
    if Snooper node is master then
        Activate → NORTH and SOUTH channel
        while NORTH and SOUTH channel are SETUP do
            Activate → EAST and WEST channel
        end while
        if All 4 channels are activated then
            return set intialization successful flag
        end if
    else
        Activate → NORTH, SOUTH and WEST channel
        while NORTH, SOUTH and WEST channel are SETUP do
            Activate → EAST Channel
        end while
    end if
end procedure
```

6.8 Network node failure detection

For the network presented in this chapter, there could arise a situation in which one of the nodes might stop operating. Under such a situation it is required to stop acquiring data. Because, the image reconstruction algorithms are designed only when all photonic modules are operational. To aid the network stop acquiring data, a packet-free technique is proposed.

In the proposed approach, the nodes are designed to switch-off all communication links when one of them fails to function. When using this technique, if any of the nodes or a communication link fail to function, the whole network will be switched off progressively. From then on, network nodes switch to the initialization mode to restart the network.

For fault diagnosis: in a situation in which the network is not able to initialize, fault diagonsis can be performed by communicating to individual nodes from the PC.

6.9 Data processing and communication unit

The DPCU [119] is designed to comprise the following modules.

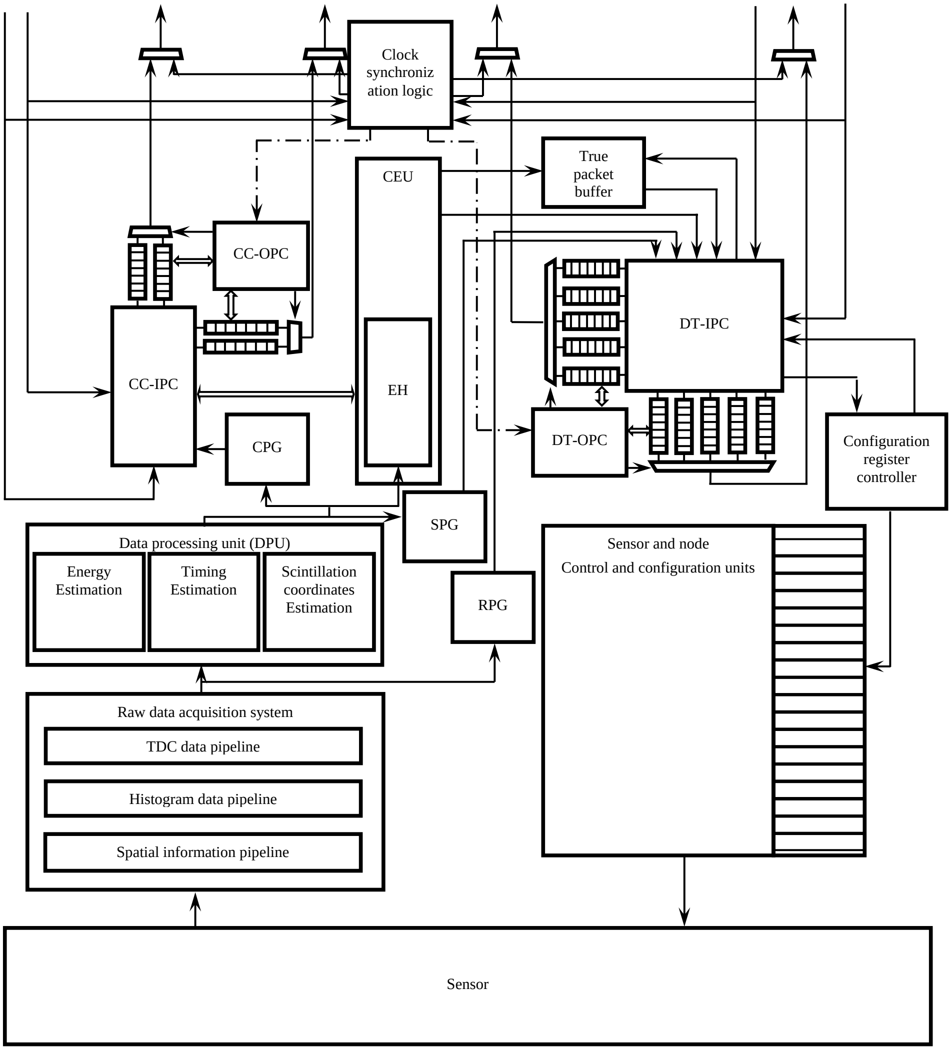
1. Sensor control unit
2. Data processing unit (DPU)
   (a) Energy estimation
   (b) Timing estimation
   (c) Scintillation coordinates estimation
3. Communication Unit (CU)
   (a) Coincidence channel unit
   (b) Coincidence engine unit (CEU)
   (c) Data transfer channel unit

The design architecture is presented in Figure 6.5. The sensor control and the data processing units are sensor dependent designs, hence it is not discussed in this chapter.

6.9.1 Coincidence channel unit

In the presented design, the DPU generated data packet is communicated to the coincidence channel unit. In the coincidence channel unit, the data packet is stored in the event history, while a copy of the same is used to generate a coincidence packet. The coincidence packet received from the network and from the coincidence packet generator is arbitrated into various virtual channels adhering to the packet routing scheme (Section 6.6) in the input packet controller. The output packet controller, which houses the scheduling and flow control algorithms, decides on a packet from the VCs output and transfers it to the communication link.

In this design the input packet controller selects the coincidence packets based on the photonic module's field-of-view configuration, and transfers the selected CPs to the coincidence engine unit.


Clock synchronization logic
True packet buffer
CEU
CC-OPC
DT-IPC
CC-IPC
EH
CPG
DT-OPC
Configuration register controller
SPG
Data processing unit (DPU)
Energy Estimation
Timing Estimation
Scintillation coordinates Estimation
RPG
Sensor and node Control and configuration units
Raw data acquisition system
TDC data pipeline
Histogram data pipeline
Spatial information pipeline
Sensor

CC-OPC: Coincidence channel output packet controller

CC-IPC: Coincidence channel input packet controller

DTC-OPC: Data transfer channel output packet controller

DTC-IPC: Data transfer channel input packet controller

EH: Event History

RPG: Raw packet generator

CPG: Coincidence packet generator

SPG: Singles packet generator

CEU: Coincidence engine unit

Figure 6.5: DPCU architecture.

6.9.2 Coincidence engine unit

In a multi-ring system the expected rate of CPs arriving at the CEU is much higher than that of a single-ring system, due to the fact that the coincidence detection needs to be performed across the rings. In case of a clinical system, the expected packet rate could be as high as 125 million packet per second. To handle such high packet rate, a high throughput design is required. Hence a novel CEU design performing coincidence detection in one clock cycle was developed. The architecture is shown in Figure 6.6.

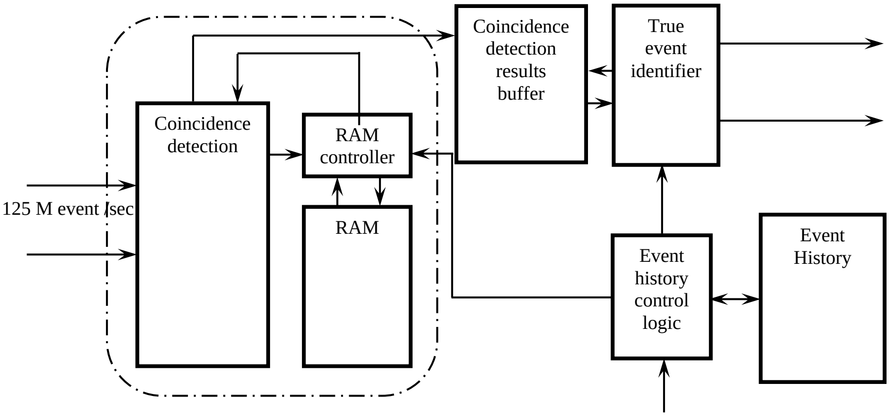


Figure 6.6: CEU architecture.

In this architecture, a RAM size of 18kb (available in Xilinx FPGAs [120]) is used, where every bit in the RAM represents a time unit. Up on detection of a gamma event, the RAM bit addressed by its timing information is flagged using the RAM controller, and after a certain time, dictated by the network's expected packet latency, the flagged bit is nullified. Thus the timing information of the entire gamma event history is stored in a user friendly format, thereby allowing the search to be performed in one clock cycle. The coincidence detection block performs the coincidence detection on the arriving packets and transfers it to the results buffers. The true event identifier tags the gamma event packets as a single or true or multiple event, depending on the results of the coincidence detection.

6.9.3 Data transfer channel unit

The functionality of the input packet controller and the output packet controller are the same as described for the coincidence channel unit. However, the data transfer network is designed to operate on six different data packet types such as the status, coincidence detected, sensor configuration, true event pair, singles and raw data packet. For every packet type at least one specifically dedicated VC is included in the design for both axial and for radial communication. Hence for these reasons the number of VCs present in the true packet network is higher than the coincidence packet network.

6.10 Snooper

This section presents the snooper node's design. The overall snooper architecture is presented in Figure 6.7. The design comprises the following modules.

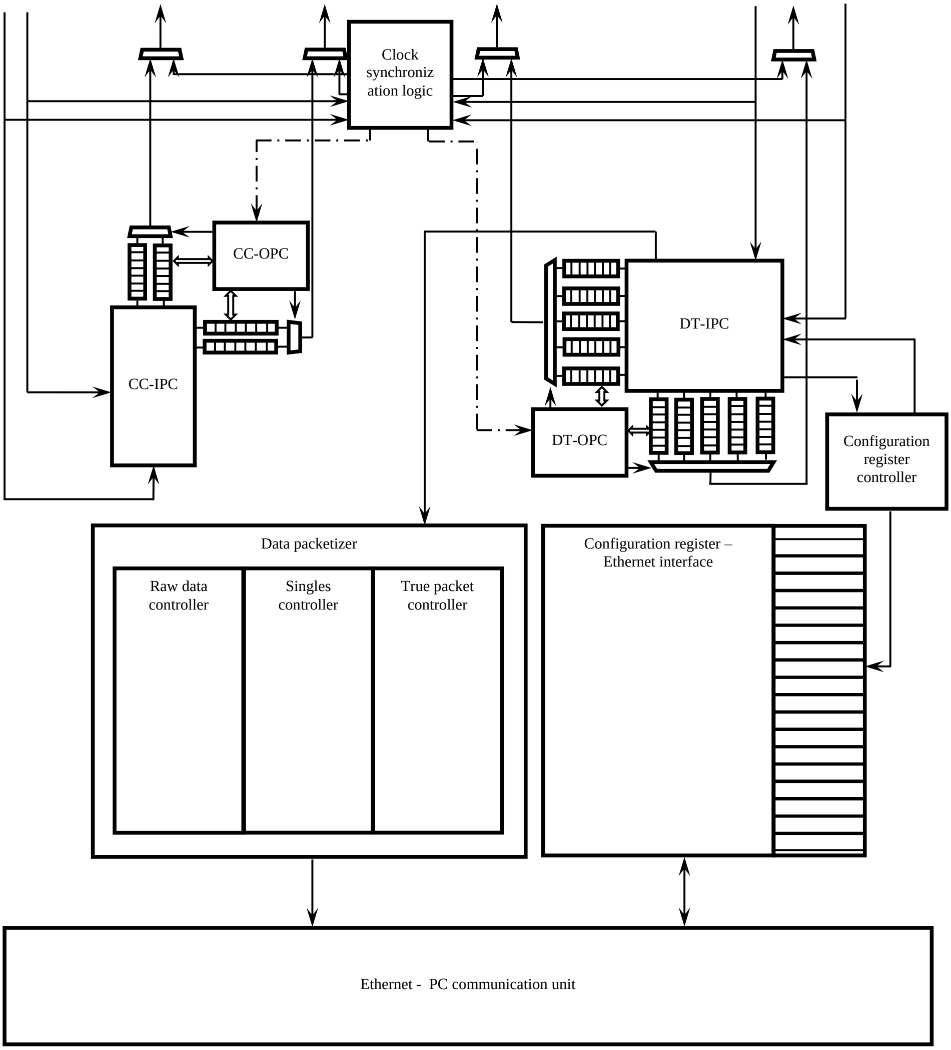
1. Coincidence channel unit
2. Data transfer channel unit
3. Ethernet-PC communication unit

6.10.1 Coincidence channel unit

The coincidence network unit design is identical to the DPCU's coincidence network unit, but for the snooper, the designs related to the coincidence engine unit and the logic used in handling the sensor generated coincidence packets are removed.

6.10.2 Data transfer channel unit

This unit is also designed in a similar fashion as the DCPU design. However, in case of the master snooper node, the data transfer unit is facilitated with an additional logic to divert the DAQ packets circulating in the network to an accumulator connected to the Ethernet controller. In this design the diversion of the DAQ packet happens only when the accumulator is free to receive it. In case the accumulator is not free to receive, the true-event packets are recirculated into the network following the data-flow strategy discussed in Section 6.6. Further, the master snooper node is also included with a additional logic to handle the configuration packets communicated to/from the PC through an interface module.


Clock synchronization logic
CC-OPC
CC-IPC
DT-IPC
DT-OPC
Configuration register controller
Data packetizer
Raw data controller
Singles controller
True packet controller
Configuration register – Ethernet interface
Ethernet - PC communication unit

CC-OPC: Coincidence channel output packet controller

CC-IPC: Coincidence channel input packet controller

DTC-OPC: Data transfer channel output packet controller

DTC-IPC: Data transfer channel input packet controller

Figure 6.7: Snooper architecture.

6.10.3 Ethernet - PC communication unit

The ethernet-PC communication unit is a module that is included only in the master snooper node. This module is designed to transfer the data collected in the accumulator to the PC, and to handle the configuration data to/from the PC. The implemented design achieves a data communication rate, as high as 105 MB/sec, which is very close to the theoretical maximum for a Giga-bit Ethernet connectivity.

6.11 Prototype model

To validate the proposed networking concepts, two scaled-down prototype models were built using the FPGAs produced by Xilinx Inc [120].

6.11.1 Single ring system

The network built to validate the single ring system, comprises 10 DPCU nodes and 1 master snooper node are shown in Figure 6.8. For this network, the DPCU was ported on to a custom designed Spartan-6 based FPGA board, and the snooper in Virtex-6 based ML605 board from Xilinx Inc [120]. In this setup a 2Gbps serial communication link was established for internode communication using the GTP (for Spartan-6) and GTX (for Virtex-6) transceivers. The Aurora link-layer protocol providing 32 bit interface was used, along with the 8/10 bit encoding scheme for forward error corrections. To facilitate packets of varying sizes to use aurora interface, a packetizer and a de-packetizer was designed as an interface between the aurora and the network core. In the implemented design both the DPCU and the snooper where operated at 62.5 MHz to match the data communication rate and the aurora data interface.

In this system, the network nodes where synchronized using a scalable approach build on a hard-wired clock distribution scheme and using the exchange of the data packets between the nodes. When using this hybrid approach timing synchronization in the order of few picoseconds was achieved. The details of the same can be found in [121, 122].

Using this setup all modes of operation were tested successfully using artificially injected data packets generated within every node. The data packets were generated periodically in all/selected nodes every few clock cycles once, as the goal of this setup is to test the network operation and to obtain the real time parameters for simulations. The simulator and the simulation results are discussed in Section 6.12 and Section 6.13 respectively.

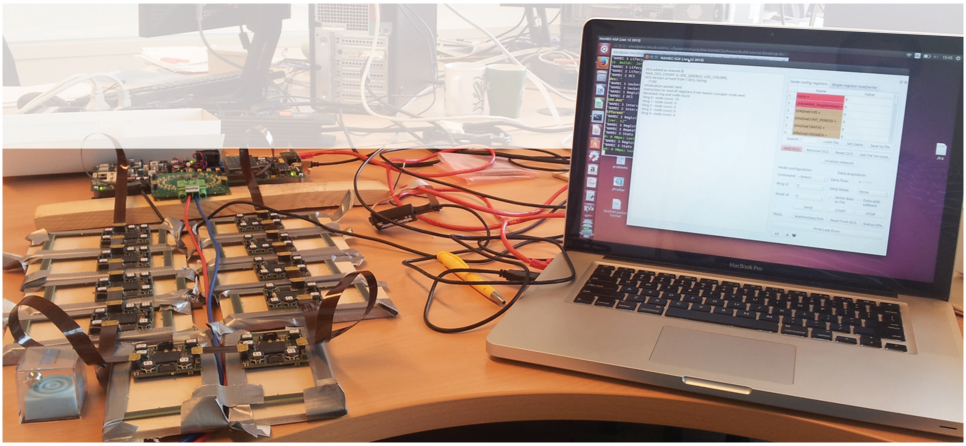


Figure 6.8: Prototype model comprising 10 DPCU nodes and 1 master snooper node. DPCU was ported into Spartan-6 based FPGA boards that are custom designed for SPADnet photonic module. Snooper was ported on to ML605 board from Xilinx Inc.

6.11.2 Multi ring system

The multi ring system was tested using an another prototype model built also using a FPGA produced by Xilinx Inc [120]. The designs were implemented in a development kit denominated ML605 [120] (Virtex-6).

In this configuration, each ML605 board houses two nodes from adjacent rings, realizing a network of 2 rings of 5 nodes each (Figure 6.9). Further, to facilitate synchronization of timestamp generation for gamma events across various nodes, a clock was distributed to all the nodes from a centralized source based on the LMK00301 board from Texas Instruments. Using the current setup, the network was operated at an inter-node communication speed of 3.2Gbps.

6.12 Network simulator

To enable our network scalability study, network simulators were built in Matlab. This section focuses on the simulator design.

Simulating a network starting from a gamma event generation to the transfer of true event is not practical, as it requires performing coincidence detection and true event transfer simultaneously. Considering the simulation time and the design complexity, two simulators, one emulating the coincidence detection channel

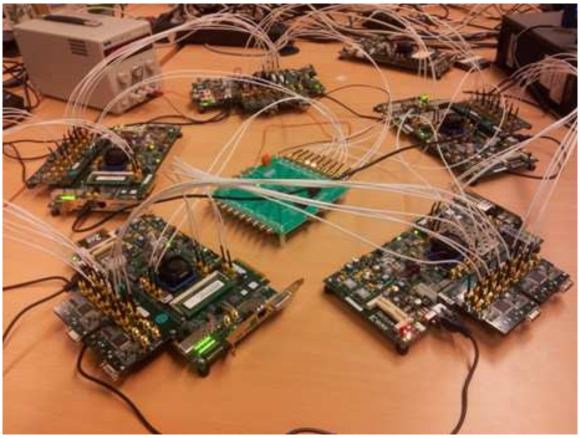


Figure 6.9: Prototype model comprising 2 ring of 5 nodes each. Two DPCU/snooper nodes from adjusent rings are ported into one Virtex-6 based ML605 board from Xilinx Inc.

and the other emulating the data transfer channel were built [123]. The following subsection discusses the techniques used in designing the simulators.

6.12.1 Data transfer simulator

In the data transfer channel, node buffer VC occupancy is a critical parameter that needs to be monitored to track a loss of data packets. Bandwidth requirements (node-node and network-PC) also need to be studied, so as to identify data transfer bottlenecks. To comprehend the requirements, simulation time and memory requirements, in contrast to discrete-time simulation, an event-driven approach was chosen. To optimize the simulator performance, the node buffer status was modelled using the rate-of-change concept, and the simulation was run only when either the status of the buffer changes to/from full Figure 6.10. Using the presented technique, simulations were run faster due to lower memory requirements and fewer iterations, while the speed-up was attained without any loss in accuracy. The complexity involved in realizing the event-driven simulation is one of the simulator drawbacks.

6.12.2 Coincidence detection simulator

For a simulator to perform coincidence detection, the transfer of the packets from a node to the next has to be monitored. Hence, for this simulator the event-

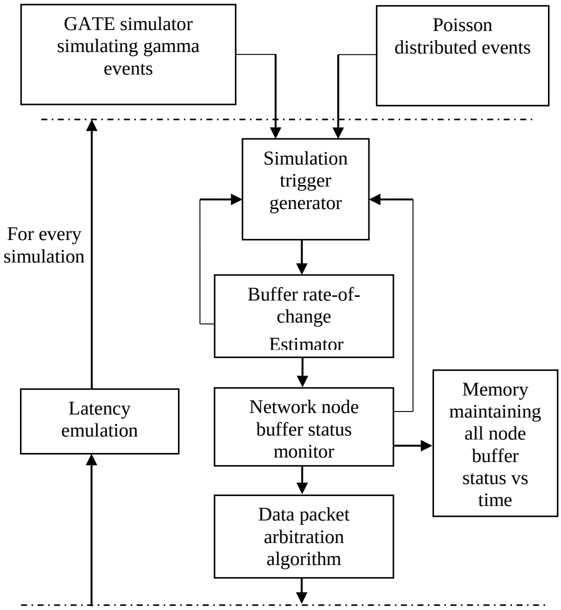

GATE simulator simulating gamma events
Poisson distributed events
Simulation trigger generator
For every simulation
Buffer rate-of-change Estimator
Latency emulation
Network node buffer status monitor
Memory maintaining all node buffer status vs time
Data packet arbitration algorithm

Figure 6.10: Data transfer channel simulator architecture.

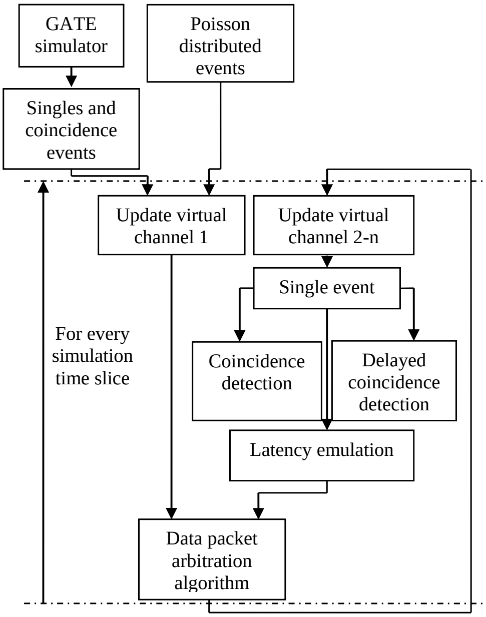

GATE simulator
Poisson distributed events
Singles and coincidence events
Update virtual channel 1
Update virtual channel 2-n
Single event
For every simulation time slice
Coincidence detection
Delayed coincidence detection
Latency emulation
Data packet arbitration algorithm

Figure 6.11: Coincidence channel simulator architecture.

driven simulation technique was implemented by triggering the simulation iteration, whenever a packet is ready to be transferred to the other node. When compared to the cycle based simulation technique, the event-driven technique will require fewer simulations and less memory. The simulator built to simulate the coincidence detection, is presented in Figure 6.11.

6.13 Scalability study

The network scalability to various PET modalities was studied using the simulators presented in Section 6.12. To ensure, that simulations match the real system, the simulator was coupled with the GATE simulator to emulate gamma event generation. To accurately model packet latency and to incorporate data bandwidth variations introduced due to protocol overhead, simulations where performed based on the results obtained from the hardware emulators built using the field-programmable gate arrays (Section 6.11).

6.13.1 Coincidence detection channel

A set of simulations were carried out to understand the network's capability and to perform coincidence detection. In the presented simulations, equal distribution of gamma event following Poisson statistics were injected into every node.

Simulation results in case of the preclinical dimensions (5 rings of 10 photonic modules each) are shown in Figure 6.12. In these simulations, latency involved for a coincidence packet to find its pair was monitored. To perform coincidence detection using a network, convergence in packet latency is required. Simulation results have shown that in case of a preclinical system with a node-to-node communication speed of 1 Gbps and 2 Gbps, packet latency was found to diverge when the incident gamma event per photonic module is around 1.6 million events per second and 3.2 million events per second respectively. The reason for the observed behaviour is evident from the bandwidth saturation observed in communication link interlinking the nodes.

For clinical (5 rings of 50 photonic modules each) and for brain PET (5 rings of 25 photonic modules each) dimensions, simulation results are presented in Figure 6.14 and Figure 6.13 respectively. As observed in preclinical system simulations, in clinical and in brain PET, the packet latency was found to diverge when either the axial or the radial link bandwidth saturated. In case of brain PET, the wider field-of-view requirements has led to bandwidth saturation earlier than in clinical

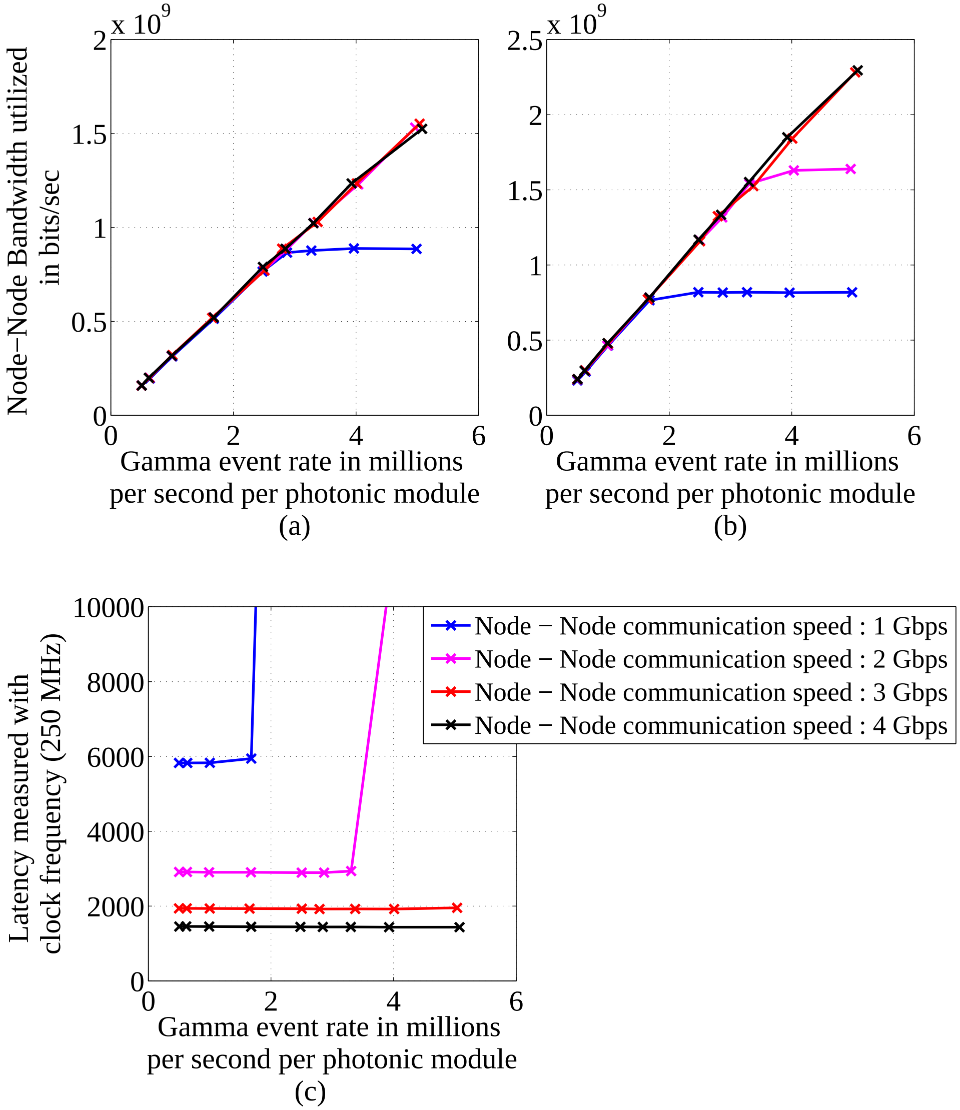


Figure 6.12: Preclinical PET simulation results: (a) radial ring node-node bandwidth utilization. (b) axial ring node-node bandwidth utilization. (c) maximum coincidence packet latency.

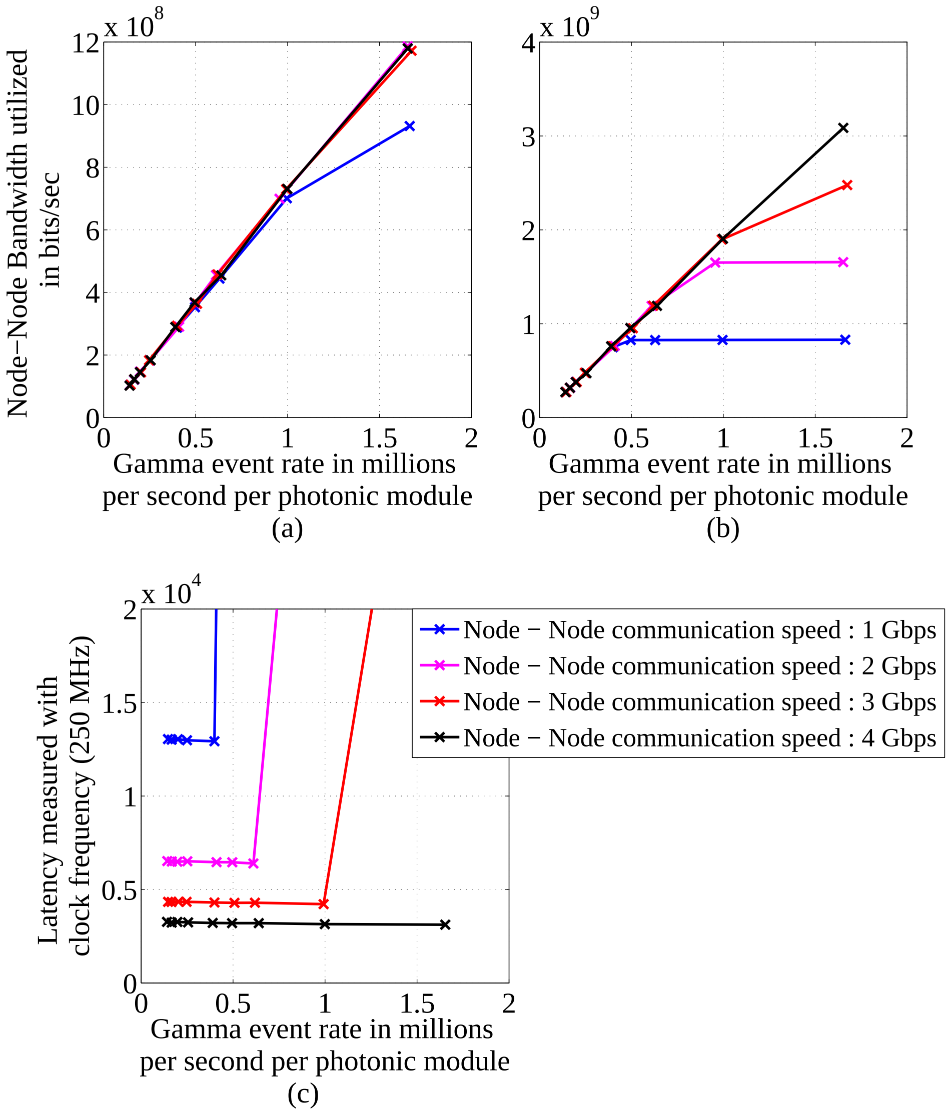
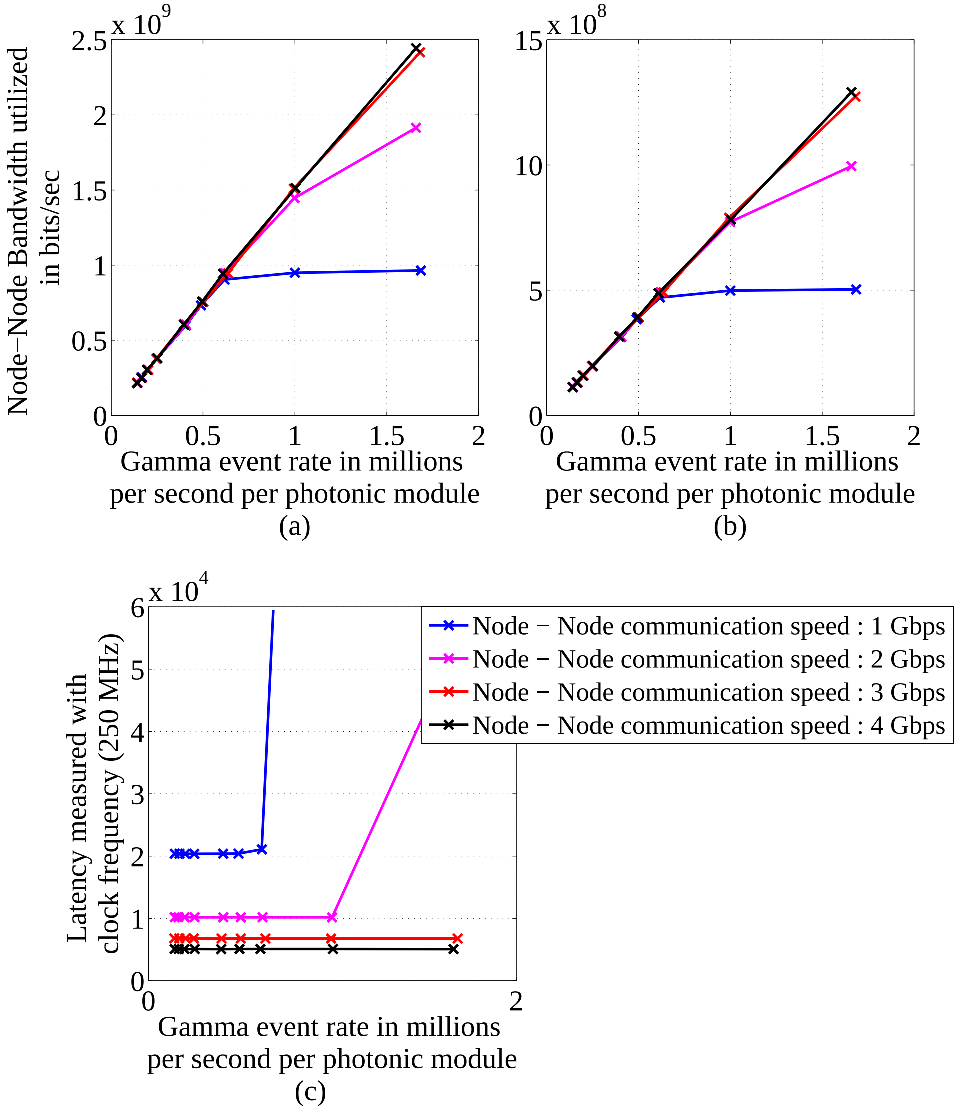


Figure 6.13: Brain PET simulation results: (a) radial ring node-node bandwidth utilization. (b) axial ring node-node bandwidth utilization. (c) maximum coincidence packet latency.



(a)

(b)

(c)

Figure 6.14: Clinical PET simulation results: (a) radial ring node-node bandwidth utilization. (b) axial ring node-node bandwidth utilization. (c) maximum coincidence packet latency.

system, although the brain PET is half the size of the clinical system.

For the digital photonic modules envisioned in the SPADnet project [46], the maximum expected singles rates for preclinical, clinical and brain PET dimensions are of the order of 5 Mcps, 500 kcps, and 500 kcps respectively. These input data rates are lower than the system's dynamic range, hence it can be ascertained that the presented network architecture will be capable of performing coincidence detection, provided the data communication rate reaches at least 3 Gbps, which has been demonstrated feasible [120].

6.13.2 Data transfer channel

Further, the data flow techniques proposed to maintain equal packet dropping probability across the system were tested by performing transient simulations. In this simulation, a huge number of packets were artificially injected into a given node (e.g. node-4) when the network was in a steady state, and then the network resource utilization was monitored. The simulation results (Figure 6.15) have shown that the network is capable to readjust itself when an imbalance occurs. However, the rate at which it readjusts depends on the data communication bandwidth.

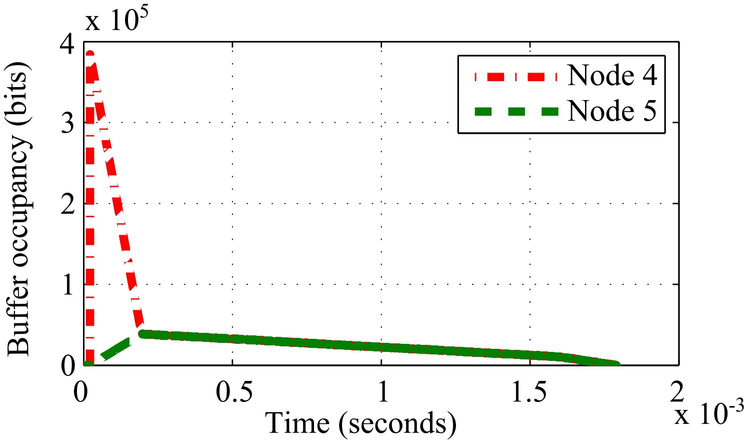


Figure 6.15: Transient simulation results.

6.14 Summary

- The amount of data generated when using SPAD sensors in PET application is very high.
- In SPADnet project [46], data rate was reduced by estimating gamma event parameters close to sensor tile and by performing noise filtering in the PET system.
- Proposed sensor network based approach enables data acquisition scalability to different PET modalities.
- In the presented system photonic modules acts as a sensor node, and are connected in cylindrical mesh topology.
- Two stage data transfer approach was used to perform distributed coincidence detection effectively in the network.
- In addition to true event acquisition, network also enables raw data and singles acquisition.
- Packet routing, flow-control and scheduling algorithms were designed specifically for the proposed network.
- Networking concept was tested successfully in hardware using FPGAs for single ring and for multi ring configurations.
- Custom designed network simulators were used to study the network scalability.
- Simulation study has shown that the network is scalable to different PET configurations.

Chapter 7

Conclusions

This thesis addressed two challenges, faced when using CMOS SPADs in cancer research and diagnosis. First, it presented techniques to improve substrate isolated CMOS SPAD performance. Second, it addressed the data acquisition challenge faced when using SPADs in a PET system. Achievements of this work are summarized chapter wise in Section 7.1. Recommendations for future work are presented in Section 7.2.

7.1 Achievements

7.1.1 Active area design

In Chapter 3, two different techniques for the active area design were proposed. The design techniques were aimed at achieving a wide spectral response for the substrate isolated SPADs.

1. The first technique is based on the wide depletion junction. Three different designs were proposed, designed and implemented in 180nm CMOS technology.

   - Design-1 based on the p$^+$/deep nwell junction, achieved PDP > 40 % from 440nm to 620nm at 10V excess bias, thanks to the guard ring design that facilitated device operation up until 10V excess bias.
   - Design-2 is a novel enhancement mode design. It overcomes the dopant diffusion issue associated with guard ring based designs and the low

fill factor of the enhancement mode design. The device designed using pwell/deep nwell junction achieved PDP $>$ 40% from 440nm to 620nm at 8V excess bias. Unlike the guard ring based designs, design-2 was shown to scale to smaller dimensions without any loss in sensitivity.

- Design-3 is a novel SPAD design based on the p-i-n configuration. Two variations were implemented and characterized. One of the variants achieved PDP >40% from 460nm to 600nm at 11V excess bias. The device achieved a DCR of 1.5 cps/μm^2 at 11V excess bias. The achieved noise performance at 11V excess bias is better than the state-of-the-art CMOS SPADs operated at or above 10V excess bias.

Design wide depletion junction based designs achieved wide spectral response and helped reduce tunneling noise, a drawback of this approach is the need to operate at or above 8V excess bias.

2. The second technique proposed in this chapter was focused on reducing the required excess bias operation. In this approach, a narrower depletion junction is used in combination with a wider photon collection region.

   - The device designed in 180nm CMOS technology using p$^+$/nwell junction achieved PDP greater than 40% from 440nm to 580nm at 4V excess bias. A drawback of this approach is that the design has relatively higher tunneling noise and a slightly worse timing jitter.

Thus, the substrate isolated SPAD PDP can be improved in green to red spectrum region either by designing the active area with a wider depletion region or by using a narrower depletion region with a wider photon collection region. Though the two approaches can achieve similar PDP results, their application will depend heavily on their drawbacks, related to the timing jitter, DCR and the required excess bias voltage.

Figure 7.1 and 7.2 present the achievement of this work by comparing it with the state-of-the-art devices.

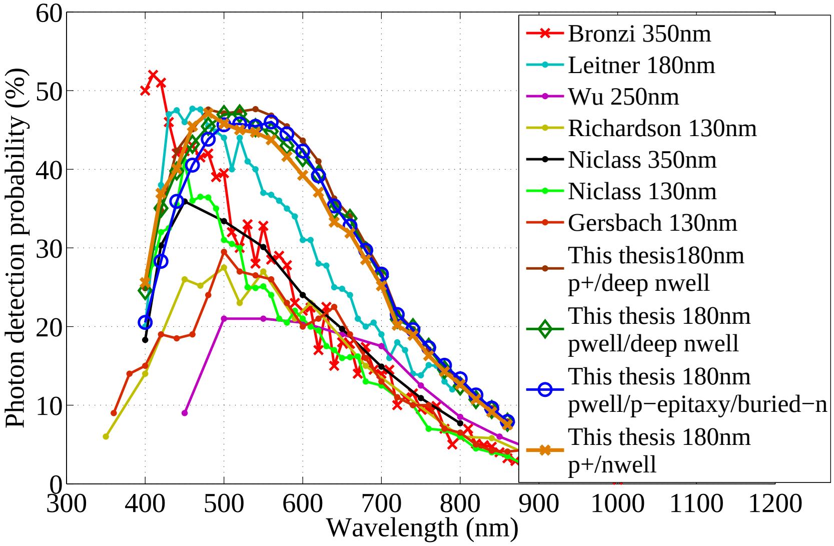


Figure 7.1: State-of-the-art PDP comparison.

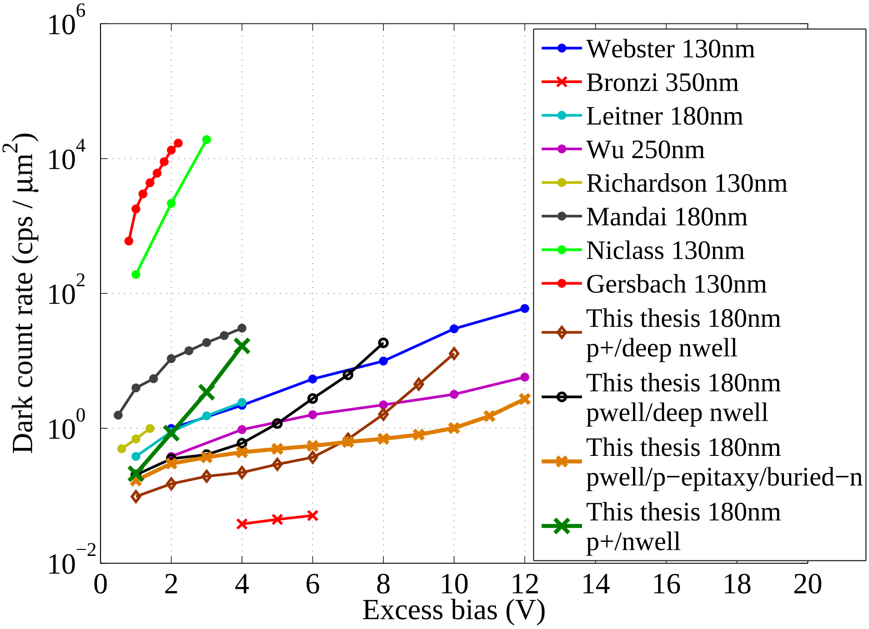


Figure 7.2: State-of-the-art DCR comparison.

7.1.2 Guard ring and periphery design

The main contributions of Chapter 4 lay in understanding the impact on DCR due to the guard region and the periphery design.

1. A study of the guard region design was performed considering the guard junction breakdown, size, dopant concentration and the device curvature.

    - The impact due to the guard region breakdown was studied using two designs. The study results have shown that when operating the device above its guard junction breakdown, the overall device DCR is dependent also on the guard junction properties.

      In one of the designs, guard junction breakdown has increased the device DCR, to limit its excess bias operation to around 4V. In the other design, the guard region designed to breakdown at 12.2V higher than the main junction, has enabled device operation up until 12V excess bias. Further, in case of the second design, even when biased above the guard junction breakdown, the device DCR was not very high, implying that it is also possible to operate both the main junction and the guard junction simultaneously in Geiger mode. The results of this study are used in Chapter 5 to design the dual junction SPAD.

    - In another study, carried out to comprehend the impact of the guard ring size on the device DCR, it was found that the guard ring effectiveness reduced with its size. The results have shown that the device DCR has increased by reducing guard ring dimensions.

    - In a follow-up study performed to understand the guard ring effectiveness for various device curvatures, it was found that a wider guard ring is required to increase its effectiveness for smaller diameter devices. Further, it was also shown that the guard ring effectiveness can also be improved by increasing its doping concentration, but at the cost of reducing guard junction breakdown voltage.

2. The study on the periphery design has shown that the breakdown of the parasitic junction formed between the periphery and the active area can impact the device DCR. Further, it was also shown by measurements that, by reducing the dopant concentration of the periphery, it is feasible to reduce the overall DCR.

Thus, in addition to the active area, the guard region and also the device periphery can contribute to DCR. The increase in DCR due to the guard region and/or the device periphery could limit the operating excess bias voltage, which in turn can effect the PDP and the timing jitter performance. Hence when designing SPADs one needs to take care of the guard region and the periphery design, not only to satisfy their required functionality but also to be able operate at the required excess bias conditions.

7.1.3 Fill factor optimization

In Chapter 5 two SPAD designs were proposed and implemented in 180nm CMOS technology.

1. The first design emulates optical microlens by acquiring photons from the guard region through the lateral avalanche propagation phenomenon. The device is designed deeper in silicon, and by using the substrate as one of the junction terminals. These design choices have helped achieve fill factor closer to 100%, by avoiding the need for a periphery and by providing the capability to house circuitry on the top of the guard ring.

2. In the second design, a novel dual junction SPAD was realized by designing the guard region in addition to the main junction to operate also in Geiger mode. The presented design achieved a 22.2% increase in fill factor when compared to the single junction based design. For the largest fabricated device, the measured DCR is about 88 cps at 10V excess bias. The measured DCR surpasses that of the state-of-the-art dual-junction SPADs proposed in [103] and [104]. Further, the proposed design enables PDP engineering. The designers can modify the device spectral response by adjusting the ratio between two junction areas. This behavior was modeled as an area weighted average of two junctions PDP. The model was shown to match the measurement results closely.

Thus, the generally insensitive guard region can be made sensitive either by designing the SPAD to operated near or above its guard region's breakdown. The resulting improvement in fill factor comes at the cost of the timing performance and increase in DCR. The magnitude of increase in DCR depends on the design, hence by careful design its feasible to make this effect negligible.

7.1.4 SPADs in PET

Chapter 6 presented a novel sensor network based approach for data acquisition, when using SPAD sensors in PET system. The proposed scheme relies on distributed data processing and noise filtering techniques for data reduction. In this approach, the photonic modules comprising the scintillator, sensor tile and an FPGA board form a sensor node. For a multi-ring based PET system a cylindrical mesh topology was proposed. In this scheme, the coincidence detection was performed in a distributed fashion. The coincidence packet latency was optimized using smaller packets in combination with static packet routing, flow control and scheduling algorithms. In addition to the true event acquisition, the network was also designed to perform raw and singles acquisition, to facilitate sensor tile test and characterization. The proposed network architecture was tested in hardware using the FPGAs, and the network scalability was validated using the network simulator built in Matlab.

Thus, the high data rate generated when using SPAD based sensors in PET systems can be handled effectively by performing localized processing and by using sensor network based approach for data filtering. The proposed approach was shown to be flexible to adapt to any sensor design and to be scalable for different PET modalities.

7.2 Future work

- In Chapter 3, two different techniques were proposed for active area design. Though the wide depletion junction based approach showed better performance, the need to operate the SPADs at or above 10V excess bias requires complicated pixel design. In [59], a poly resistor and a decoupling capacitor was used in the pixel design to enable operation above 10V excess bias. The inactive area occupied by poly resistor and the capacitor reduces the pixel fill factor. One of the future research directions could be to optimize the pixel design for wide depletion junction based designs. A research approach could be to investigate the use of the poly resistor and the decoupling capacitor on the top of the guard ring or on the periphery.

- In Chapter 4 it was found that the device to device variation in DCR is higher when a using 2μm wide guard ring. The measured increase in DCR when reducing the guard ring width from 2μm to 1μm was also very high. These two observations suggest that the devices designed with a 2μm wide guard ring are more susceptible to process variations in the pwell implant. Further investigation is required in this direction to understand if the guard region design is one of the causes for the high DCR pixels in an array.

- In Chapter 5 it was shown that with the use of deeper junctions, it is possible to integrate transistors on the top of active area. Detailed study is required to understand the influence on the transistor and on the SPAD performance due to the each other.
- Further in Chapter 5, the timing jitter measurements for the dual junction SPAD have shown to have two peaks corresponding to two junctions. In principle, the two peaks can facilitate color and position sensitivity in SPADs. Future research is required to study the potential use of the dual junction SPAD for colour and position estimation.
- The networking technique, presented in Chapter 6, was evaluated in simulations. Future work would be to test them in real time, when used in PET application. Further, in the current implementation FPGA is used at the back of the sensor tile. Heat generated by the FPGA could in principle can affect the SPAD performance. A potential approach would be to migrate the networking technique to an ASIC, which can reduce the heating.

This dissertation was prepared with the aim to improve the success rate of cancer treatment. But one needs to understand that cancer can be fought more effectively only with prevention.

”prevention is better than cure...”

Bibliography

[1] P. Webb, R. McIntyre, and J. Conradi, “Properties of avalanche photodiodes,” *RCA review*, vol. 35, pp. 234–278, 1974.

[2] M. Ghioni, S. Cova, A. Lacaita, and G. Ripamonti, “New silicon epitaxial avalanche diode for single-photon timing at room temperature,” *Electronics Letters*, vol. 24, no. 24, pp. 1476–1477, 1988.

[3] C. Veerappan and E. Charbon, “A substrate isolated CMOS SPAD enabling wide spectral response and low electrical crosstalk,” *Selected Topics in Quantum Electronics, IEEE Journal of*, vol. 20, no. 6, pp. 1–7, Nov 2014.

[4] L. Braga, L. Gasparini, L. Grant, R. Henderson, N. Massari, M. Perenzoni, D. Stoppa, and R. Walker, “A fully digital 8 × 16 SiPM array for PET applications with per-pixel TDCs and real-time energy output,” *Solid-State Circuits, IEEE Journal of*, vol. 49, no. 1, pp. 301–314, Jan 2014.

[5] “World cancer report 2014.” International Agency for Research on Cancer and others, WHO Geneva, 2014.

[6] P. E. Valk, *Positron emission tomography: basic sciences*. Springer Science & Business Media, 2003.

[7] D. W. Jones, P. Hogg, and E. Seeram, *Practical SPECT/CT in nuclear medicine*. Springer, 2013.

[8] J. McGinty *et al.*, “Wide-field fluorescence lifetime imaging of cancer,” *Biomedical optics express*, vol. 1, no. 2, pp. 627–640, 2010.

[9] S. Lu, Y. Wang, H. Huang, Y. Pan, E. J. Chaney, S. A. Boppart, H. Ozer, A. Y. Strongin, and Y. Wang, "Quantitative FRET imaging to visualize the invasiveness of live breast cancer cells," *PloS one*, vol. 8, no. 3, p. e58569, 2013.

[10] L. Tang, C. Dong, and J. Ren, "Highly sensitive homogenous immunoassay of cancer biomarker using silver nanoparticles enhanced fluorescence correlation spectroscopy," *Talanta*, vol. 81, no. 4, pp. 1560–1567, 2010.

[11] H. Iams and B. Salzberg, "The secondary emission phototube," *Radio Engineers, Proceedings of the Institute of*, vol. 23, no. 1, pp. 55–64, 1935.

[12] H. Kume, K. Koyama, K. Nakatsugawa, S. Suzuki, and D. Fatlowitz, "Ultrafast microchannel plate photomultipliers," *Appl. Opt.*, vol. 27, no. 6, pp. 1170–1178, Mar 1988. [Online]. Available: http://ao.osa.org/abstract.cfm?URI=ao-27-6-1170

[13] J. Hynecek, "Impactron-a new solid state image intensifier," *Electron Devices, IEEE Transactions on*, vol. 48, no. 10, pp. 2238–2241, Oct 2001.

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

[15] H. Dautet, P. Deschamps, B. Dion, A. D. MacGregor, D. MacSween, R. J. McIntyre, C. Trottier, and P. P. Webb, "Photon counting techniques with silicon avalanche photodiodes," *Applied Optics*, vol. 32, no. 21, pp. 3894–3900, 1993.

[16] S. Cova, A. Lacaita, M. Ghioni, G. Ripamonti, and T. Louis, "20-ps timing resolution with single-photon avalanche diodes," *Review of scientific instruments*, vol. 60, no. 6, pp. 1104–1110, 1989.

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

[18] R. H. Haitz, "Mechanisms contributing to the noise pulse rate of avalanche diodes," *Journal of Applied Physics*, vol. 36, no. 10, pp. 3123–3131, 1965.

[19] A. Lacaita, M. Ghioni, and S. Cova, "Double epitaxy improves single-photon avalanche diode performance," *Electronics Letters*, vol. 25, no. 13, pp. 841–843, 1989.

[20] S. Cova, A. Lacaita, M. Ghioni, G. Ripamonti, and T. A. Louis, "20 ps timing resolution with single photon avalanche diodes," *Review of Scientific Instruments*, vol. 60, no. 6, pp. 1104–1110, 1989.

[21] M. Ghioni, A. Gulinatti, P. Maccagnani, I. Rech, and S. Cova, "Planar silicon SPADs with 200-m diameter and 35-ps photon timing resolution," pp. 63 720R–63 720R–9, 2006. [Online]. Available: http://dx.doi.org/10.1117/12.685834

[22] A. Lacaita, S. Cova, M. Ghioni, and F. Zappa, "Single-photon avalanche diode with ultrafast pulse response free from slow tails," *Electron Device Letters, IEEE*, vol. 14, no. 7, pp. 360–362, 1993.

[23] B. Aull, A. Loomis, J. A. Gregory, and D. Young, "Geiger-mode avalanche photodiode arrays integrated with CMOS timing circuits," in *Device Research Conference Digest, 1998. 56th Annual*, 1998, pp. 58–59.

[24] J. C. Jackson, A. Morrison, D. Phelan, and A. Mathewson, "A novel silicon geiger-mode avalanche photodiode," in *Electron Devices Meeting, 2002. IEDM '02. International*, 2002, pp. 797–800.

[25] E. Sciacca *et al.*, "Silicon planar technology for single-photon optical detectors," *Electron Devices, IEEE Transactions on*, vol. 50, no. 4, pp. 918–925, 2003.

[26] M. Ghioni, G. Armellini, P. Maccagnani, I. Rech, M. Emsley, and M. Unlu, "Resonant-cavity-enhanced single-photon avalanche diodes on reflecting silicon substrates," *Photonics Technology Letters, IEEE*, vol. 20, no. 6, pp. 413–415, 2008.

[27] B. Aull, H. Andrew, J. Douglas, B. J. Richard, M.H. and, J. Peter, and J. L. Deborah, "Geiger-mode avalanche photodiodes for three-dimensional imaging," vol. 13, pp. 335–350, 2002.

[28] B. Aull *et al.*, "Laser radar imager based on 3d integration of geiger-mode avalanche photodiodes with two SOI timing circuit layers," in *Solid-State Circuits Conference, 2006. ISSCC 2006. Digest of Technical Papers. IEEE International*, 2006, pp. 1179–1188.

[29] A. Rochas, M. Gani, B. Furrer, P.-A. Besse, R. Popovic, G. Ribordy, and N. Gisin, "Single photon detector fabricated in a complementary metal oxide

semiconductor high-voltage technology," *Review of Scientific Instruments*, vol. 74, no. 7, pp. 3263–3270, 2003.

[30] C. Niclass, M. Gersbach, R. Henderson, L. Grant, and E. Charbon, "A single photon avalanche diode implemented in 130-nm CMOS technology," *Selected Topics in Quantum Electronics, IEEE Journal of*, vol. 13, no. 4, pp. 863–869, 2007.

[31] C. Veerappan *et al.*, "A 160x128 single-photon image sensor with on-pixel 55ps 10b time-to-digital converter," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2011 IEEE International*, 2011, pp. 312–314.

[32] D. Stoppa, F. Borghetti, J. Richardson, R. Walker, L. Grant, R. Henderson, M. Gersbach, and E. Charbon, "A 32x32-pixel array with in-pixel photon counting and arrival time measurement in the analog domain," in *ESSCIRC, 2009. ESSCIRC '09. Proceedings of*, 2009, pp. 204–207.

[33] M. Gersbach, Y. Maruyama, E. Labonne, J. Richardson, R. Walker, L. Grant, R. Henderson, F. Borghetti, D. Stoppa, and E. Charbon, "A parallel 32x32 time-to-digital converter array fabricated in a 130 nm imaging CMOS technology," in *ESSCIRC, 2009. ESSCIRC '09. Proceedings of*, 2009, pp. 196–199.

[34] J. Richardson, R. Walker, L. Grant, D. Stoppa, F. Borghetti, E. Charbon, M. Gersbach, and R. Henderson, "A 32x32 50ps resolution 10 bit time to digital converter array in 130nm CMOS for time correlated imaging," in *Custom Integrated Circuits Conference, 2009. CICC '09. IEEE*, 2009, pp. 77–80.

[35] S. Tisa, A. Tosi, and F. Zappa, "Fully-integrated CMOS single photon counter," *Opt. Express*, vol. 15, no. 6, pp. 2873–2887, Mar 2007. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-15-6-2873

[36] A. Rochas, M. Gosch, A. Serov, P. A. Besse, R. Popovic, T. Lasser, and R. Rigler, "First fully integrated 2-D array of single-photon detectors in standard CMOS technology," *Photonics Technology Letters, IEEE*, vol. 15, no. 7, pp. 963–965, 2003.

[37] F. Zappa, A. Gulinatti, P. Maccagnani, S. Tisa, and S. Cova, "SPADA: single-photon avalanche diode arrays," *Photonics Technology Letters, IEEE*, vol. 17, no. 3, pp. 657–659, 2005.

[38] S. Mandai and E. Charbon, "Timing optimization of a H-tree based digital silicon photomultiplier," *Journal of Instrumentation*, vol. 8, no. 09, p. P09016, 2013. [Online]. Available: http://stacks.iop.org/1748-0221/8/i=09/a=P09016

[39] D. Tyndall, B. Rae, D. Li, J. Richardson, J. Arlt, and R. Henderson, "A 100Mphoton/s time-resolved mini-silicon photomultiplier with on-chip fluorescence lifetime estimation in 0.13 μm CMOS imaging technology," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2012 IEEE International*, 2012, pp. 122–124.

[40] R. Walker, J. Richardson, and R. Henderson, "A 128x96 pixel event-driven phase-domain σ/δ - based fully digital 3D camera in 0.13 μm CMOS imaging technology," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2011 IEEE International*, 2011, pp. 410–412.

[41] F. Villa *et al.*, "SPAD smart pixel for time-of-flight and time-correlated single-photon counting measurements," *Photonics Journal, IEEE*, vol. 4, no. 3, pp. 795–804, June 2012.

[42] J. Meijlink, C. Veerappan, S. Seifert, D. Stoppa, R. Henderson, E. Charbon, and D. Schaart, "First measurement of scintillation photon arrival statistics using a high-granularity solid-state photosensor enabling time-stamping of up to 20,480 single photons," in *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2011 IEEE*, Oct 2011, pp. 2254–2257.

[43] E. Venialgo, S. Mandai, T. Gong, D. R. Schaart, E. Charbon, A. Carimatto, G. Borghi, and D. Schaart, "Time estimation with multichannel digital silicon photomultipliers," *Physics in medicine and biology*, vol. 60, no. 6, pp. 2435–2452, 2015.

[44] M. Fishburn and E. Charbon, "System tradeoffs in gamma-ray detection utilizing SPAD arrays and scintillators," *Nuclear Science, IEEE Transactions on*, vol. 57, no. 5, pp. 2549–2557, Oct 2010.

[45] J. Wehner, B. Weissler, P. M. Dueppenbecker, P. Gebhardt, B. Goldschmidt, D. Schug, F. Kiessling, and V. Schulz, "MR-compatibility assessment

of the first preclinical PET-MRI insert equipped with digital silicon photomultipliers," *Physics in Medicine and Biology*, vol. 60, no. 6, p. 2231, 2015. [Online]. Available: http://stacks.iop.org/0031-9155/60/i=6/a=2231

[46] C. Bruschini *et al.*, "SPADnet: Embedded coincidence in a smart sensor network for PET applications," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 734, pp. 122–126, 2014.

[47] D. Cumming and M. AL-RAWHANI, "Endoscopy capsule with SPAD array for detecting fluorescence emitted by biological tissue," Apr. 16 2015, wO Patent App. PCT/GB2014/053,041. [Online]. Available: http://www.google.com/patents/WO2015052523A1?cl=en

[48] E. Garutti, "EndoToFPET-US a novel multimodal tool for endoscopy and positron emission tomography," in *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2012 IEEE*, Oct 2012, pp. 2096–2101.

[49] X. Michalet, A. Ingargiola, R. Colyer, G. Scalia, S. Weiss, P. Maccagnani, A. Gulinatti, I. Rech, and M. Ghioni, "Silicon photon-counting avalanche diodes for single-molecule fluorescence spectroscopy," *Selected Topics in Quantum Electronics, IEEE Journal of*, vol. 20, no. 6, pp. 248–267, Nov 2014.

[50] D. D.-U. Li, J. Arlt, D. Tyndall, R. Walker, J. Richardson, D. Stoppa, E. Charbon, and R. K. Henderson, "Video-rate fluorescence lifetime imaging camera with CMOS single-photon avalanche diode arrays and high-speed imaging algorithm," *Journal of biomedical optics*, vol. 16, no. 9, pp. 096 012–096 012, 2011.

[51] V. Krishnaswami, S. Burri, F. Regazzoni, C. Bruschini, C. J. van Noorden, E. Charbon, and R. Hoebe, "SPAD array camera for localization based super resolution microscop," in *Focus On Microscopy Conference*, no. EPFL-POSTER-191249, 2013.

[52] V. Krishnaswami, C. J. Van Noorden, E. M. Manders, and R. A. Hoebe, "Towards digital photon counting cameras for single-molecule optical nanoscopy," *Optical Nanoscopy*, vol. 3, no. 1, p. 1, 2014.

[53] M. Gersbach, C. Niclass, E. Charbon, J. Richardson, R. Henderson, and L. Grant, "A single photon detector implemented in a 130nm CMOS imaging

process," in *Solid-State Device Research Conference, 2008. ESSDERC 2008. 38th European*, 2008, pp. 270–273.

[54] T. Leitner *et al.*, "Measurements and simulations of low dark count rate single photon avalanche diode device in a low voltage 180-nm CMOS image sensor technology," *Electron Devices, IEEE Transactions on*, vol. 60, no. 6, pp. 1982–1988, 2013.

[55] J. Richardson, L. Grant, and R. Henderson, "Low dark count single-photon avalanche diode structure compatible with standard nanometer scale CMOS technology," *Photonics Technology Letters, IEEE*, vol. 21, no. 14, pp. 1020–1022, 2009.

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

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

[58] S. Mandai, M. W. Fishburn, Y. Maruyama, and E. Charbon, "A wide spectral range single-photon avalanche diode fabricated in an advanced 180 nm CMOS technology," *Opt. Express*, vol. 20, no. 6, pp. 5849–5857, Mar 2012. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-20-6-5849

[59] E. Webster, R. Walker, R. Henderson, and L. Grant, "A silicon photomultiplier with 30% detection efficiency from 450-750nm and 11.6 μm pitch NMOS-only pixel with 21.6% fill factor in 130nm CMOS," in *Solid-State Device Research Conference (ESSDERC), 2012 Proceedings of the European*, 2012, pp. 238–241.

[60] D. Mosconi, D. Stoppa, L. Pancheri, L. Gonzo, and A. Simoni, "CMOS single-photon avalanche diode array for time-resolved fluorescence detection," in *Solid-State Circuits Conference, 2006. ESSCIRC 2006. Proceedings of the 32nd European*, Sept 2006, pp. 564–567.

[61] Y. Maruyama, J. Blacksberg, and E. Charbon, "A 1024x8 700ps time-gated SPAD line sensor for laser raman spectroscopy and LIBS in space and rover-based planetary exploration," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2013 IEEE International*, 2013, pp. 110–111.

[62] C. Niclass and E. Charbon, "A single photon detector array with 64×64 resolution and millimetric depth accuracy for 3D imaging," in *Solid-State Circuits Conference, 2005. Digest of Technical Papers. ISSCC. 2005 IEEE International*, Feb 2005, pp. 364–604 Vol. 1.

[63] D. Stoppa, L. Pancheri, M. Scandiuzzo, L. Gonzo, G.-F. Dalla Betta, and A. Simoni, "A CMOS 3-D imager based on single photon avalanche diode," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 54, no. 1, pp. 4–12, Jan 2007.

[64] T. Frach, G. Prescher, C. Degenhardt, R. de Gruyter, A. Schmitz, and R. Ballizany, "The digital silicon photomultiplier - principle of operation and intrinsic detector performance," in *Nuclear Science Symposium Conference Record (NSS/MIC), 2009 IEEE*, Oct 2009, pp. 1959–1965.

[65] L. Braga, L. Gasparini, L. Grant, R. Henderson, N. Massari, M. Perenzoni, D. Stoppa, and R. Walker, "An 8x16-pixel 92k SPAD time-resolved sensor with on-pixel 64ps 12b TDC and 100MS/s real-time energy histogramming in 0.13 μm CIS technology for PET/MRI applications," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2013 IEEE International*, 2013, pp. 486–487.

[66] G. Intermite *et al.*, "Enhancing the fill-factor of CMOS SPAD arrays using microlens integration," pp. 95 040J–95 040J–12, 2015. [Online]. Available: http://dx.doi.org/10.1117/12.2178950

[67] S. Donati, E. Randone, M. Fathi, J.-H. Lee, E. Charbon, and G. Martini, "Uniformity of concentration factor and back focal length in molded polymer microlens arrays," in *Conference on Lasers and Electro-Optics 2010.* Optical Society of America, 2010, p. JThE36. [Online]. Available: http://www.opticsinfobase.org/abstract.cfm?URI=CLEO-2010-JThE36

[68] J. M. Pavia, M. Wolf, and E. Charbon, "Measurement and modeling of microlenses fabricated on single-photon avalanche diode arrays for fill factor

recovery," *Opt. Express*, vol. 22, no. 4, pp. 4202–4213, Feb 2014. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-22-4-4202

[69] D. Bronzi *et al.*, "Low-noise and large-area CMOS SPADs with timing response free from slow tails," in *Solid-State Device Research Conference (ESSDERC), 2012 Proceedings of the European*, 2012, pp. 230–233.

[70] C. Veerappan, "Data acquisition system design for a 160x128 single-photon image sensor with on-pixel 55 ps time-to-digital converter," Ph.D. dissertation, TU Delft, Delft University of Technology, 2010.

[71] S. Cova, M. Ghioni, A. Lacaita, C. Samori, and F. Zappa, "Avalanche photodiodes and quenching circuits for single-photon detection," *Appl. Opt.*, vol. 35, no. 12, pp. 1956–1976, Apr 1996. [Online]. Available: http://ao.osa.org/abstract.cfm?URI=ao-35-12-1956

[72] A. Rochas, "Single photon avalanche diodes in CMOS technology," Ph.D. dissertation, Ecole Polytechnique Federale de Lausanne, 2003.

[73] M. Fishburn, "Fundamentals of CMOS single-photon avalanche diodes," Ph.D. dissertation, Delft University of Technology, 2012.

[74] A. Eisele, R. Henderson, B. Schmidtke, T. Funk, L. Grant, J. Richardson, and W. Freude, "185 MHz count rate, 139 dB dynamic range single-photon avalanche diode with active quenching circuit in 130nm cmos technology," 2011, pp. 278–281.

[75] S. H. Lee and R. P. Gardner, "A new g–m counter dead time model," *Applied Radiation and Isotopes*, vol. 53, no. 4, pp. 731–737, 2000.

[76] L. Neri, S. Tudisco, F. Musumeci, A. Scordino, G. Fallica, M. Mazzillo, and M. Zimbone, "Note: Dead time causes and correction method for single photon avalanche diode devices," *Review of Scientific Instruments*, vol. 81, no. 8, 2010. [Online]. Available: http://scitation.aip.org/content/aip/journal/rsi/81/8/10.1063/1.3476317

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

[78] G. Hurkx, D. Klaassen, and M. Knuvers, “A new recombination model for device simulation including tunneling,” *Electron Devices, IEEE Transactions on*, vol. 39, no. 2, pp. 331–338, 1992.

[79] E. A. Webster and R. K. Henderson, “A TCAD and spectroscopy study of dark count mechanisms in single-photon avalanche diodes,” *Electron Devices, IEEE Transactions on*, vol. 60, no. 12, pp. 4014–4019, 2013.

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

[81] H. Finkelstein, M. Hsu, and S. Esener, “STI-bounded single-photon avalanche diode in a deep-submicrometer CMOS technology,” *Electron Device Letters, IEEE*, vol. 27, no. 11, pp. 887–889, 2006.

[82] H. Dautet, P. Deschamps, B. Dion, A. D. MacGregor, D. MacSween, R. J. McIntyre, C. Trottier, and P. P. Webb, “Photon counting techniques with silicon avalanche photodiodes,” *Appl. Opt.*, vol. 32, no. 21, pp. 3894–3900, Jul 1993. [Online]. Available: http://ao.osa.org/abstract.cfm?URI=ao-32-21-3894

[83] C. Niclass, A. Rochas, P.-A. Besse, and E. Charbon, “Design and characterization of a CMOS 3-d image sensor based on single photon avalanche diodes,” *Solid-State Circuits, IEEE Journal of*, vol. 40, no. 9, pp. 1847–1854, 2005.

[84] D. Bronzi *et al.*, “Large-area CMOS SPADs with very low dark counting rate,” pp. 86 311B–86 311B–8, 2013. [Online]. Available: http://dx.doi.org/10.1117/12.2004209

[85] J. Wu, L. S.H., H. F.Z., and S. Lin, “Two-dimensional mapping of photon counts in low-noise single-photon avalanche diodes,” in *International Image Sensor Workshop 2013 (IISW)*, 2013.

[86] R. McIntyre, P. Webb, and H. Dautet, “A short-wavelength selective reach-through avalanche photodiode,” *Nuclear Science, IEEE Transactions on*, vol. 43, no. 3, pp. 1341–1346, Jun 1996.

[87] P. Webb and R. McIntyre, “Large area reach-through avalanche diodes for x-ray spectroscopy,” *Nuclear Science, IEEE Transactions on*, vol. 23, no. 1, pp. 138–144, Feb 1976.

[88] P. Webb and A. Jones, “Large area reach-through avalanche diodes for radiation monitoring,” *Nuclear Science, IEEE Transactions on*, vol. 21, no. 1, pp. 151–158, Feb 1974.

[89] D. Grubisic and A. Shah, “New silicon reach-through avalanche photodiodes with enhanced sensitivity in the DUV/UV wavelength range,” in *Information Communication Technology Electronics Microelectronics (MIPRO), 2013 36th International Convention on*, May 2013, pp. 48–54.

[90] C. Schow, J. Schaub, R. Li, J. Qi, and J. Campbell, “A 1-Gb/s monolithically integrated silicon NMOS optical receiver,” *Selected Topics in Quantum Electronics, IEEE Journal of*, vol. 4, no. 6, pp. 1035–1039, Nov 1998.

[91] L. Garrett, J. Qi, C. Schow, and J. Campbell, “A silicon-based integrated NMOS p-i-n photoreceiver,” *Electron Devices, IEEE Transactions on*, vol. 43, no. 3, pp. 411–416, Mar 1996.

[92] B. Ciftcioglu, L. Zhang, L. Zhang, J. Marciante, J. Zuegel, R. Sobolewski, and H. Wu, “Integrated silicon PIN photodiodes using deep n-well in a standard 0.18- μ m CMOS technology,” *Lightwave Technology, Journal of*, vol. 27, no. 15, pp. 3303–3313, Aug 2009.

[93] C. Veerappan and E. Charbon, “A low dark count p-i-n diode based SPAD in CMOS technology,” *Electron Devices, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2015.

[94] ——, “CMOS SPAD based on photo-carrier diffusion achieving PDP >40% from 440 to 580 nm at 4 V excess bias,” *Photonics Technology Letters, IEEE*, vol. 27, no. 23, pp. 2445–2448, Dec 2015.

[95] Z. Xiao, D. Pantic, and R. Popovic, “A new single photon avalanche diode in CMOS high-voltage technology,” in *Solid-State Sensors, Actuators and Microsystems Conference, 2007. TRANSDUCERS 2007. International*, 2007, pp. 1365–1368.

[96] C. Niclass, “Single-photon image sensors in CMOS: picosecond resolution for three-dimensional imagaing,” Ph.D. dissertation, Ecole Polytechnique Federale de Lausanne, 2008.

[97] D. D. et al, “BackSPAD - back-side illuminated single-photon avalanche diodes: Concept and preliminary performances,” in *IEEE Nuclear science symposium 2012*. IEEE, 2012.

[98] A. Lacaita, S. Cova, A. Spinelli, and F. Zappa, “Photon-assisted avalanche spreading in reach-through photodiodes,” *Applied Physics Letters*, vol. 62, no. 6, pp. 606–608, Feb 1993.

[99] A. Lacaita, A. Spinelli, and S. Longhi, “Avalanche transients in shallow p-n junctions biased above breakdown,” *Applied Physics Letters*, vol. 67, no. 18, pp. 2627–2629, Oct 1995.

[100] G. Ripamonti, M. Ghioni, S. Cova, and M. Mastrapasqua, “Propagating avalanche position-sensitive photon detector with resolution in the micrometer and picosecond range,” *Electron Device Letters, IEEE*, vol. 13, no. 1, pp. 35–37, Jan 1992.

[101] M. Fishburn, Y. Maruyama, and E. Charbon, “Reduction of fixed-position noise in position-sensitive single-photon avalanche diodes,” *Electron Devices, IEEE Transactions on*, vol. 58, no. 8, pp. 2354–2361, Aug 2011.

[102] C. Veerappan, Y. Maruyama, and E. Charbon, “Silicon integrated electrical micro-lens for CMOS SPADs based on avalanche propagation phenomenon,” in *International Image Sensor Workshop*, 2013.

[103] H. Finkelstein, M. Hsu, and S. Esener, “Dual-junction single-photon avalanche diode,” *Electronics Letters*, vol. 43, no. 22, pp. –, Oct 2007.

[104] R. Henderson, E. Webster, and L. Grant, “A dual-junction single-photon avalanche diode in 130-nm CMOS technology,” *Electron Device Letters, IEEE*, vol. 34, no. 3, pp. 429–431, March 2013.

[105] H. Kume, S. Muramatsa, and M. Iida, “Position sensitive photomultiplier tubes for scintillation imaging,” Hamamatsu Photonics KK, 1126-1 Ichino-cho, Hamamatsu, Tech. Rep., 1986.

[106] H. S. Yoon, G. B. Ko, S. I. Kwon, C. M. Lee, M. Ito, I. C. Song, D. S. Lee, S. J. Hong, and J. S. Lee, “Initial results of simultaneous PET/MRI experiments with an MRI-compatible silicon photomultiplier PET scanner,” *Journal of Nuclear Medicine*, vol. 53, no. 4, pp. 608–614, 2012.

[107] J. Mackewn *et al.*, “PET performance evaluation of a pre-clinical SiPM-based MR-compatible PET scanner,” *Nuclear Science, IEEE Transactions on*, vol. 62, no. 3, pp. 784–790, June 2015.

[108] Y. Haemisch, T. Frach, C. Degenhardt, and A. Thon, "Fully digital arrays of silicon photomultipliers (dSiPM)–a scalable alternative to vacuum photomultiplier tubes (PMT)," *Physics Procedia*, vol. 37, pp. 1546–1560, 2012.

[109] C. Veerappan, C. Bruschini, and E. Charbon, "Sensor network architecture for a fully digital and scalable SPAD based PET system," in *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2012 IEEE*, Oct 2012, pp. 1115–1118.

[110] H. Dent, W. Jones, and M. Casey, "A real time digital coincidence processor for positron emission tomography," *Nuclear Science, IEEE Transactions on*, vol. 33, no. 1, pp. 556–559, Feb 1986.

[111] D. Newport, H. Dent, M. Casey, and D. Bouldin, "Coincidence detection and selection in positron emission tomography using VLSI," *Nuclear Science, IEEE Transactions on*, vol. 36, no. 1, pp. 1052–1055, Feb 1989.

[112] M.-A. Tetrault, M. Lepage, N. Viscogliosi, F. Belanger, J. Cadorette, C. Pepin, R. Lecomte, and R. Fontaine, "Real time coincidence detection system for digital high resolution APD-based animal PET scanner," in *Nuclear Science Symposium Conference Record, 2005 IEEE*, vol. 5, Oct 2005, pp. 2849–2853.

[113] E. Kim, P. Olcott, and C. Levin, "A new data path design for a PET data acquisition system: A packet based approach," in *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2011 IEEE*, Oct 2011, pp. 3871–3873.

[114] D. Newport, S. Siegel, B. Swann, B. Atkins, A. McFarland, D. Pressley, M. Lenox, and R. Nutt, "Quicksilver: A flexible, extensible, and high-speed architecture for multi-modality imaging," in *Nuclear Science Symposium Conference Record, 2006. IEEE*, vol. 4, Oct 2006, pp. 2333–2334.

[115] B. Atkins, D. Pressley, M. Lenox, B. Swann, D. Newport, and S. Siegel, "A data acquisition, event processing and coincidence determination module for a distributed parallel processing architecture for PET and SPECT imaging," in *Nuclear Science Symposium Conference Record, 2006. IEEE*, vol. 4, Oct 2006, pp. 2439–2442.

[116] S. D. Chawade, M. A. Gaikwad, and R. M. Patrikar, "Article: Review of XY routing algorithm for network-on-chip architecture," *International Journal*

of Computer Applications, vol. 43, no. 21/973-93-80867-69-8, pp. 20–23, April 2012, full text available.

[117] W. Dally, “Virtual-channel flow control,” *Parallel and Distributed Systems, IEEE Transactions on*, vol. 3, no. 2, pp. 194–205, Mar 1992.

[118] W. Dally and C. Seitz, “Deadlock-free message routing in multiprocessor interconnection networks,” *Computers, IEEE Transactions on*, vol. C-36, no. 5, pp. 547–553, May 1987.

[119] C. Veerappan, C. Bruschini, and E. Charbon, “Distributed coincidence detection for multi-ring based PET systems,” in *Real Time Conference (RT), 2014 19th IEEE-NPSS*, May 2014.

[120] Xilinx. [Online]. Available: http://www.xilinx.com/

[121] M. Bijwaard, “Scalable network based clock synchronization for digital PET system,” Master’s thesis, TU Delft, Delft University of Technology, 2015.

[122] M. Bijwaard, C. Veerappan, C. Bruschini, and E. Charbon, “Fundamentals of a scalable network in SPADnet-based PET systems,” in *Nuclear Science Symposium Conference Record, 2015. IEEE*, Oct 2015.

[123] C. Veerappan, E. Venialgo, C. Bruschini, and E. Charbon, “SPADnet network modeling, simulation and emulation,” in *Real Time Conference (RT), 2014 19th IEEE-NPSS*, May 2014.

Samenvatting

In kankercelonderzoek en in kankerdiagnose-apparatuur, zoals positron emissie tomografie (PET) en enkele-foton emissie berekende tomografie (SPECT), zijn fotonische sensoren nodig met een gevoeligheid voor individuele fotonen. Tot voor kort werden hier standaard foto-vermenigvuldigingsbuizen (PMTs) voor gebruikt. Hoewel een PMT de vereiste gevoeligheid heeft, beperkt zijn afmeting en de benodigde spanning (typisch rond de 1kV) de toepassing in arrays met hoge dichtheid. Deze beperking is recent overwonnen doordat enkele-foton lawinediodes (SPADs) geïntegreerd zijn in CMOS technologie. CMOS SPADs zijn gerealiseerd in massaal-parallele arrays met enkele-foton detectors in hoge dichtheden. Inmiddels hebben SPAD arrays PMTs voorbijgestreefd met betrekking tot resolutie en de mogelijkheid om de aankomsttijden van individuele fotonen te registreren. In combinatie met de compatibiliteit voor magnetische resonantie beeldvorming (MRI), de lage benodigde spanning (rond 25V), en de kleine afmeting, zijn hierdoor nieuwe mogelijkheden onstaan in kankeronderzoek. Hoewel SPAD arrays potentieel bruikbare technologie oplevert voor toekomstige kankerdiagnose-apparatuur en kankercelonderzoek, heeft de lage gevoeligheid, hoge ruis in het donker en hoge generatiesnelheid van gegevens de toepassing beperkt. Dit proefschrift behandelt enkele van deze uitdagingen.

In dit proefschrift hebben we, om de circuit-integratie te vergemakkelijken en de overspraak tussen SPADs te verminderen, vooral gekeken naar het verbeteren van de gevoeligheid voor enkele fotonen van SPADs met een losgekoppeld substraat. In Hoofdstuk 3 zijn twee ontwerptechnieken gepresenteerd om een bredere spectrale gevoeligheid te verkrijgen. De eerste techniek is gebaseerd op de brede depletieverbinding en de tweede op de smallere depletieverbinding maar met een breder gebied om fotonen te verzamelen. Voor de eerste techniek zin drie ontwerpen voorgesteld, gebruikmakend van de p^{+}/deep nwell, pwell/deep nwell en pwell/p-epitaxy/begraven-n (p-i-n) verbindingen. De ontwerpen haalden betere foton-detectie-waarschijnlijkheden (PDP) dan 40% op het interval 440nm tot 620nm, met een extra voorspanning hoger dan 8V. De behaalde spectrale gevoeligheid is hoger en breder dan voor de andere bekende substraat-geïsoleerde CMOS SPADs. Bij gebruik van brede depletieverbindingen werd ook de ruis ten gevolge van de tunnelbijdrage verminderd, resulterend in een lage foton-telfrequentie in het donker (DCR). Bijvoorbeeld gaf een SPAD gebaseerd op p-i-n diodes een DCR van 1.5 cps/μm^2 bij 11V extra voorspanning. Voor de tweede techniek is een ontwerp gemaakt met een pplus/nwell verbinding waarbij deep nwell en begraven-n gebruikt zijn voor de foton-verzamelgebieden. Bij 4V extra voorspanning gaf dit ontwerp bijna dezelfde PDP als de brede depletie verbinding-gebaseerde SPAD. De voor- en nadelen van deze twee ontwerptechnieken zijn ook besproken in dit Hoofdstuk.

Verder is in dit proefschrift een gedetailleerde studie gedaan naar de invloed van het beveiligingsgebied en de SPAD omgeving op de ruis in het donker. De studie is gepresenteerd in Hoofdstuk 4. De resultaten suggereren dat de ruis ook kan ontstaan in het beveiligingsgebied, ofwel door het falen van de beveiligingsverbinding en/of door zijn beperkte effectiviteit bij lage voorspanningen. In het geval van de omgeving van de SPAD bleek het falen van de parasitische verbinding tussen de omgeving en een van de hoofdverbindingen een van de oorzaken van de ruis in het donker. Hoofdstuk 4 bespreekt ook in detail ontwerptechnieken om de ruis ten gevolge van deze effecten te verminderen.

Een van de beperkingen van SPADs is de lage sensordichtheid ten gevolge van het inactieve oppervlak bezet door het beveiligingsgebied, de omgeving, en de bedrading. In dit proefschrift zijn twee verschillende SPAD ontwerpen voorgesteld om de inactieve oppervlakte te verminderen. De ontwerpen zijn gepresenteerd in Hoofdstuk 5. Het eerste ontwerp emuleert optische microlenzen, waarin de SPAD wordt ontworpen om dichtbij de grenzen van het falen van de veiligheidsverbinding te werken. Onder deze condities worden sommige van de fotodragers die in de veiligheidsverbinding zijn ontstaan gemeten door de hoofdverbinding door middel van de zijdelingse lawinepropagatie. Omdat het voorgestelde ontwerp ook transistors bovenop de veiligheidszone kan hebben zonder dat de omgeving nodig is, denken we dat het voorgestelde ontwerp kan leiden tot een nieuwe generatie van pixel arrays met hoge dichtheden. Het tweede ontwerp is een uitbreiding van de eerste, waarin zowel de hoofdverbinding als de veiligheidszone boven de faalgrens werken. Het voorgestelde ontwerp resulteerde in een verbetering van 22.2% in de opvulfactor vergeleken met het gebruikelijke ontwerp gepresenteerd in Hoofdstuk

3.

Bij het construeren van een systeem voor beeldvorming gebruikmakend van SPAD sensoren is de hoge generatiesnelheid van gegevens tengevolge van de pixel granulariteit een uitdaging voor de signaalverwerking. In toepassingen zoals PET moeten honderden SPAD arrays samenwerken. In Hoofdstuk 6 is een schaalbaar en flexibel meetsysteem voorgesteld, gebaseerd op de sensor netwerk architectuur van het PET systeem. In de voorgestelde aanpak bestaat elke foton-module uit 25 SPAD-gebaseerde sensoren inclusief de dataverwerking en communicatie-unit. In deze configuratie is een gegevensnetwerk opgezet tussen de modules om de datastroom gedistribueerd te verwerken, om de hoeveelheid gegevens in-situ in het systeem te verminderen. De voorgestelde aanpak is effectief voor pre-klinische, brein- en klinische PET systemen.

Acknowledgments

I take this opportunity to thank all those people who made this day possible for me.

First of all, I would like to thank my supervisor and promoter Prof. dr. Edoardo Charbon in providing me an opportunity to work in this exciting field of research. This thesis would not have taken the shape that it is today, without his guidance and support. Many a times I have benefitted from his ability to foresee problems and from his desire to be at the pinnacle of his field. Opportunities, when working with him were never limited - be it a tapeout, measurement setups or a collaboration. His doors were always open to discuss the weirdest of ideas, the exciting to bad results, or even to share my frustrations. On a personnel front I would like to thank him for letting me take a long leave especially when my family needed the most.

Second to my promoter, I would like to thank Dr. Claudio Bruschini, he was the project manager for the FP6 (Megaframe) and the FP7 (SPADnet) project that I worked on during my Phd days. As a student, I did learn a lot from his meticulous reviews and discussions. Today the confidence that I have in project management and in event organization is all because of him. The multi-ring network architecture presented in Chapter 6 is an outcome of many brain storming sessions that I had with him over the years

Next to them, I would like to thank Prof. dr. Chris van Hoof, Prof. dr. David Cumming, Prof. dr. Arokia Nathan, Dr. Ivan Rech, Dr. Erik Jan Lous, Prof. dr. Albert Theuwissen and Prof. dr. Paddy French for accepting to be in my thesis defense committee. I am grateful to have experts of their kind, reviewing my thesis. I am thankful for all their time, review and painstaking effort in traveling to Delft just for my defense.

During my Phd I did get lot of opportunities to collaborate with some of the leading researchers in the Netherlands and abroad. I am grateful to all of them who made me richer by sharing their experience and knowledge in various platforms. I thank all consortium members of SPADnet and Megaframe projects including Prof. dr. Robert Henderson, Dr. David Stoppa, Steve East, Gabor Nemeth, Prof. dr. Emoke Lorincz, Dr. Eric Gros D'Aillon, Zoltan Papp, Dr. Richard Walker, and Dr. Leonardo Gasparini for their help, support and time in providing me with valuable feedback. I would like to thank Frank Thus, and again Erik Jan Lous, and all of their group members for their generous help in providing me with tapeouts and even hosting me at their office when I needed it the most. I am thankful for all their support and encouragement during the various stages of my research work. Although I could not have Dr. Hein Valk in my committee today, I would like to point out that the results on the SPAD design presented in this thesis is an outcome of the initial discussions that I had with him. I am grateful to have worked with a knowledgeable person of his caliber during my Phd.

I am deeply indebted to friends and my colleagues Dr. Yuki Maruyama, Dr. Matthew W. Fishburn, Dr. Shingo Mandai, Esteban Venialgo and Augusto Carimatto in helping me with their advice. Be it a technical problem related to my research work or on a personal front, they were some of the few whom I could fall back on. I would like to thank my group mates Tim Gong, Harald Homulle, Ivan Michel Antolovic, Chao Zhang, Dr. Myung-Jae Lee, Mauro Scandiuzzo, Mahdi Aminian, Juan Mata Pavia, Scott Linder, Martijn Bijwaard, Sachin Chadha, Joop Meijlink, Mohammed Azim Karami, Dr. Francesco Regazzoni, Pengfei Sun, Samuel Burri, visiting professors Dr. Masahiro Akiyama and Dr. Yudong Li and many other Msc students with whom I have worked with, in making my Phd days a memorable one.

Focus on my studies and research at TU Delft would not have been possible without Minaksie Ramsoekh, Rosario Salazar Lozano and Joyce Siemers. Though they might know little about my thesis, I can very well say that this thesis would not have been possible without their help and support. Even I could forget my residence permit expiry date but they never did. I am indebted for all their help and care for all these years.

Today when I look at my results one person whom I always think of is Antoon Pieter Frehe, our system administrator, without whom it would have been a nightmare to set up a software environment, experimental setup, simulation platform or a repository to manage my documents. I thank him for all his help, support and time.

Coming to TU Delft for MSc would not have been possible without the scholarship offered by the circuit and systems (CAS) group under the top talent foundation.

Ever since I reached Delft, the help and the support I received from the CAS group did lay the foundation for the seven years to follow. I would like to thank Prof. dr. Nick van der Meijs for his help during the initial days, and for his directions towards research and writing. I would like to thank all CAS members and professors - Alle-Jan van der veen, Geert Leus, Rene van Leuken, and Gerard Janssen, for their help and support.

During my Phd, I did get an opportunity to serve in the prestigious MEST (micro electronic systems and technology) student board taking up various responsibilities. I feel gifted to have a worked with some of the best, in organizing many social and technical events. I learnt a lot from the experiences that we acquired together from our successes and challenges. I will cherish my days in MEST, and will always look forward to contributing to its growth in whatever form I could.

Also I would like to thank my colleagues and my manager Rene Kohlmann from Dialog Semiconductors in understanding and in supporting me when I take days off from work towards the preparation of this thesis.

I wonder if I should thank the following people or not. After carefully thinking it through, I decided not to, as I don't deserve to do so.

Anyone who would know me from India would wonder how I survived away from my family. Few people who have come to my mind are my friends. They were among the few who knew, what my likes and dislikes are, and what I need and don't. Today to have survived in Delft without even knowing how to cook is all because of them. Their influence was significant - be it a brainstorming of my ideas; a presentation that I gave; an article or even in this thesis that I wrote. The cover page of this thesis is an excellent example of their contribution. These are the friends whom I am gifted with: Hrishikesh Patel (Tiggy), Aditya Thallam Thattai, Sachin Navalkar, Rahul Shukla, Yash Joshi, Saket Sakunia, Subramanya Prasad Nageshrao, Venkata Girish, Reshu Gupta, Srijith Menon, Saish Sridharan, Aman Kalsi, Sidharth Mahalingam, Krishna Kowlgi, D H Krishna Murthy, Venkatesh Seshan, Sumeet Kumar, Raj Thilak Rajan, Rajat Bhardwaj, Rohit Gupta, Karthik Chandrasekar, Madhavan Manivannan, Vinoth Krishnan Elangovan, Shenario Ezhil, Vishwas Jain, Venkataraman Krishnaswami and many more. I am eagerly looking forward to continuing to trouble them for many more years to come.

What I am today is what I have learnt and been taught to me by my parents (Appa - Veerappan and Amma - Ravichandra) and brothers (Senthi anna and Arun). My perspective towards life, motivation towards studies, research and work were all indulged in me by them. My sister-in-law (Kannamai Annie and Raje Annie) were a constant source of support for me. I need to accept that the whole of my family is more excited for my Phd defense than I am. Today I stand together with their best gift of my life, my wife Shivakami. I could not have asked her for more, than to accept my situation where I had to stay away for almost 9 months since our wedding, just for me to finish my Phd.

Chockalingam Veerappan
March 2016
Delft, The Netherlands

Publications

Journals

- **C. Veerappan** and E. Charbon, “A low dark count p-i-n diode based SPAD in CMOS technology,” *Electron Devices, IEEE Transactions on (TED)*, vol. 63, no. 1, pp. 65–71, Jan 2016.

- **C. Veerappan** and E. Charbon, “CMOS SPAD based on photo-carrier diffusion achieving PDP >40% from 440 to 580 nm at 4V excess bias,” *Photonics Technology Letters, IEEE (PTL)*, vol. 27, no. 23, pp. 2445–2448, Dec 2015.

- **C. Veerappan** and E. Charbon, “A substrate isolated CMOS SPAD enabling wide spectral response and low electrical crosstalk,” *Selected Topics in Quantum Electronics, IEEE Journal of (JSTQE)*, vol. 20, no. 6, pp. 299–305, Nov 2014.

- C. Bruschini, E. Charbon, **C. Veerappan**, L. Braga, N. Massari, M. Perenzoni, L. Gasparini, D. Stoppa, R. Walker, A. Erdogan, R. Henderson, S. East, L. Grant, B. Jatekos, F. Ujhelyi, G. Erdei, E. Lrincz, L. Andr, L. Maingault, V. Reboud, L. Verger, E. Gros d’Aillon, P. Major, Z. Papp and G. Nmeth, “SPADnet: Embedded coincidence in a smart sensor network for PET applications,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 734, Part B, pp. 122 – 126, 2014.

- E. Gros-Daillon, L. Maingault, L. Andr, V. Reboud, L. Verger, E. Charbon, C. Bruschini, **C. Veerappan**, D. Stoppa, N. Massari, M. Perenzoni, L. Braga,

L. Gasparini, R. Henderson, R. Walker, S. East, L. Grant, B. Jatekos, E. Lorincz, F. Ujhelyi, G. Erdei, P. Major, Z. Papp and G. Nemeth, "First characterization of the SPADnet sensor: a digital silicon photomultiplier for PET applications," *Journal of Instrumentation*, vol. 8, no. 12, p. C12026, 2013.

Journals under preparation

- **C. Veerappan** and E. Charbon, "A novel high performance CMOS SPAD with dual junction".

- **C. Veerappan** and E. Charbon, "Design and optimization of the guard region and the periphery for CMOS SPAD".

- **C. Veerappan** and E. Charbon, "A scalable data acquisition approach using sensor networks for digital SiPM based PET systems".

Book Chapters

E. Charbon and **C. Veerappan**, "Designing Photon Counting, Wide Spectrum Optical Radiation Detectors in CMOS-Compatible Technologies," from Analog Electronics for Radiation Detection, R. Turchetta, Taylor & Francis 2016.

Conferences

- M. Bijwaard, **C. Veerappan**, C. Bruschini, and E. Charbon, "Fundamentals of a Scalable Network in SPADnet-based PET Systems," in *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2015 IEEE*, Nov 2015.

- **C. Veerappan**, C. Bruschini, and E. Charbon, "Distributed coincidence detection for multi-ring based PET systems," in *Real Time Conference (RT), 2014 19th IEEE-NPSS*, May 2014.

- **C. Veerappan**, E. Venialgo, C. Bruschini, and E. Charbon, "SPADnet network modeling, simulation and emulation," in *Real Time Conference (RT), 2014 19th IEEE-NPSS*, May 2014.

- **C. Veerappan**, Y. Maruyama, and E. Charbon, “Silicon integrated electrical micro-lens for CMOS SPADs based on avalanche propagation phenomenon,” in *International Image Sensor Workshop (IISW)*, 2013.

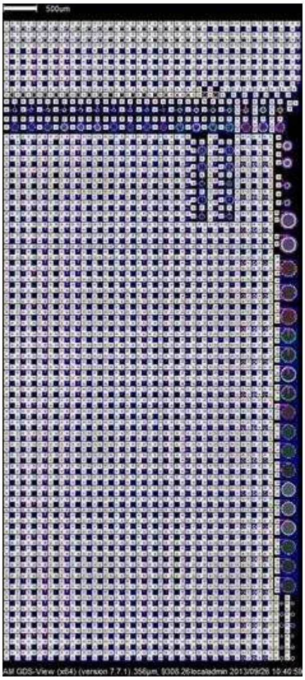
- **C. Veerappan**, C. Bruschini, and E. Charbon, “Sensor network architecture for a fully digital and scalable SPAD based PET system,” in *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2012 IEEE*, Oct 2012, pp. 1115–1118.

- J. Meijlink, **C. Veerappan**, S. Seifert, D. Stoppa, R. Henderson, E. Charbon, and D. Schaart, “First measurement of scintillation photon arrival statistics using a high-granularity solid-state photosensor enabling time-stamping of up to 20,480 single photons,” in *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2011 IEEE*, Oct 2011, pp. 2254–2257.

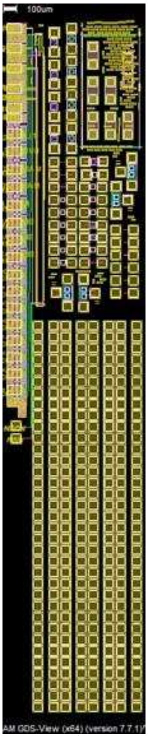
- **C. Veerappan**, J. Richardson, R. Walker, D. Li, M. Fishburn, D. Stoppa, F. Borghetti, Y. Maruyama, M. Gersbach, R. Henderson, and E. Charbon, “Characterization of large-scale non-uniformities in a 20k TDC/SPAD array integrated in a 130nm CMOS process,” in *Solid-State Device Research Conference (ESSDERC), 2011 Proceedings of the European*, Sept 2011, pp. 331–334.

- **C. Veerappan**, J. Richardson, R. Walker, D. Li, M. Fishburn, Y. Maruyama, D. Stoppa, F. Borghetti, M. Gersbach, R. Henderson, and E. Charbon, “A 160x128 single-photon image sensor with on-pixel 55ps 10b time-to-digital converter,” in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2011 IEEE International*, Feb 2011, pp. 312–314.

Chip gallery



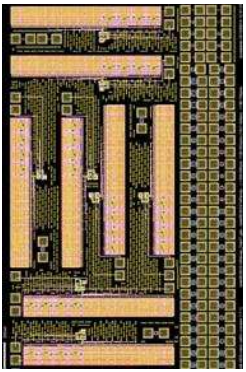
(a) IISW 2013



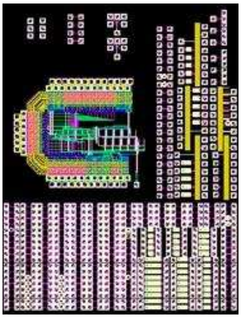
(b)

100um
AM GDS-View (x64

(e) JSTQE 2014



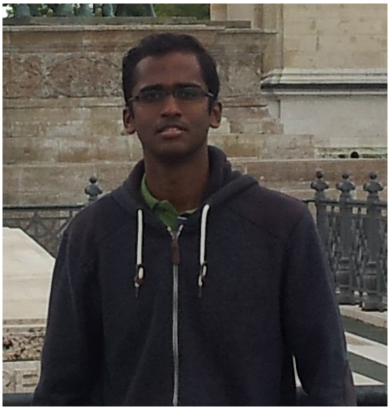
(c)



(d) PTL and TED 2015

The SPAD designs and characterization results presented in the thesis were from chip (a), (d) and (e) fabricated in 180nm CMOS technology. Chip (b) and (c) were designed in another similar CMOS technology node to test some of the designs.

About the author



Chockalingam Veerappan was born in Coimbatore, India on 5th of April 1986. He graduated in Electrical and Electronics Engineering from Amrita Vishwa Vidyapeetham (Amrita University), Coimbatore India. He started his work in center of excellence in computational engineering and networking (CEN) also at Amrita University as a research assistant, where he was involved in the development of software defined radio and in hardware realization of machine translation systems. After a year at CEN, in September 2008, he started his masters in Microelectronics at Delft University of Technology, The Netherlands. During his masters he specialized in digital design and it was during his thesis with Prof. dr. Edoardo Charbon, he got his first exposure to single-photon imaging. Then on he continued on similar lines for his Phd, specializing in the design of Single-Photon Avalanche Diode (SPAD) and sensor networks. Since December 2015 he has started his career in industry as a junior system design engineer at Dialog Semiconductor, The Netherlands.