PhD Dissertation

International Doctorate School in Information and

Communication Technologies

DISI - University of Trento

MODELING, DESIGN AND CHARACTERIZATION OF A MULTI-PIXEL DIGITAL SIPM FOR PET APPLICATIONS

Leo Huf Campos Braga

Advisor:

Dr. David Stoppa

Fondazione Bruno Kessler (FBK)

April 2014

Abstract

Positron Emission Tomography (PET) scanners provide functional three-dimensional images of the body that are extremely useful in cancer and brain research. The goal of this work is the modeling, design and characterization of a CMOS-based photodetector for PET. To this aim, first a model for the energy resolution and coincidence resolution time (CRT) for digital, SPAD-based detectors is developed.

Then, a top-to-bottom detector architecture is proposed, containing an innovative in-pixel compression technique that allows for high fill-factor (FF) and efficient readout. At the top-level of the architecture, an integrated discriminator monitors the photon flux for incoming gamma events, enabling an event-based readout scheme. The first complete implementation of this architecture is described, the SPADnet-I sensor, which is composed by an 8×16 pixel array, each of around 0.6 × 0.6 mm^2 with 720 SPADs, resulting in a pixel FF of 42.6%. The sensor can obtain the discrete photon flux estimation at up to 100 Msamples/s, which are used by the discriminator and also output at real-time.

The complete characterization of the sensor is presented, and the best sensor configuration was found to be at 84% of the SPADs enabled (disabled starting with the highest DCR one), with 2 V SPAD excess bias and 150 ns integration time. This configuration results in an energy resolution of 10.8% and a CRT of 288 ps, the latter which was obtained with a new, hardware-friendly time of arrival (ToA) estimation algorithm, also described in this thesis.

Finally, the sensor model, validated by the experimental results, is used to predict the performance of possible modifications in the sensor, and some design improvements are suggested for a future implementation of the architecture.

Keywords

Digital Silicon Photomultiplier (SiPM), Positron Emission Tomography (PET), spatial and temporal compression, mini-SiPM, Single-Photon Avalanche Diode (SPAD), CMOS, image sensors, biomedical sensors.

Acknowledgements

First of all, I would like to thank my advisor David Stoppa for welcoming me into his research group in 2010 and guiding me throughout all these years. Without his experienced and intelligent technical and personal support I could not have achieved my goals.

I would also like to thank all research personnel at the SOI group, namely Matteo Perenzoni, Nicola Massari, Leonardo Gasparini, Lucio Pancheri, Daniele Perenzoni and Massimo Gottardi for their everyday advices and helpful discussions.

My colleagues at the research group SRS are also recipients of my truest appreciation, you were also instrumental in the success of my work. I personally name Claudio Piemonte, Alberto Gola, Nicola Serra, Alessandro Tarolli, Alessandro Ferri, Nicola Fronza and Gabrielle Giacomini for their support.

All partners in the SPADnet project were also fundamental in the completion of my PhD work. I wish to specifically acknowledge Robert Henderson and Richard Walker from the University of Edinburgh for their close cooperation in the design of the SPADnet-I sensor; Gabor Nemeth and Peter Major from MEDISO for sharing their abundant knowledge about PET systems; and again Richard Walker, Leonardo Gasparini and Zoltan Papp for the development of the firmware for the sensor characterization.

I express my gratitude also to Prof. Gian-Franco Dalla Betta from the University of Trento, for his help in all things related to the PhD, from constructive work goals to everyday problem solving.

Many thanks also to my PhD colleagues, Elisabetta Mazzuca, Michelle Benetti, Marina Repich and Hesong Xu, with whom I could share the troubles (and occasional joys) of doing a PhD.

Finally, I would also like to give special thanks to my wife, Suzana Domingues, for her everyday support and understanding during the long hours, far away travels and challenging times of my PhD. Her deep patience and gifted advices were absolutely invaluable to me.

Contents

CHAPTER 1 INTRODUCTION 1

1.1. POSITRON EMISSION TOMOGRAPHY 1

1.2. THE PROPOSED SOLUTION 4

1.3. STRUCTURE OF THE THESIS 5

CHAPTER 2 STATE OF THE ART 7

2.1. SILICON PHOTOMULTIPLIERS 7

CHAPTER 3 GAMMA DETECTION IN PET SCANNERS 12

3.1. GAMMA DISCRIMINATION 12

3.2. ENERGY ESTIMATION 14

3.3. TIME OF ARRIVAL ESTIMATION 18

3.3.1 Timestamp Model 19

3.3.2 Methods for Timing Resolution Assessment 23

3.3.3 Analysis 25

3.4. POSITION ESTIMATION 30

3.5. OTHER FEATURES 33

3.6. SUMMARY 34

CHAPTER 4 THE SPADNET SENSOR 36

4.1. SPAD-BASED SENSORS: DESIGN CONCEPTS AND COMPROMISES 36

4.2. SENSOR ARCHITECTURE 39

4.2.1 Counting photons: the mini-SiPM 39

4.2.2 Gamma detection: the top-level 44

4.2.3 Meeting in the middle: the pixel 46

4.2.4 Timing Diagram *49*

4.3. FIRST IMPLEMENTATION: SPADNET-I 51

4.3.1 Hierarchy sizing *51*

4.3.2 Schematics *55*

4.3.3 Layout floorplans *59*

4.3.4 Chip micrograph *62*

CHAPTER 5 PROCESSING SPADNET DATA 65

5.1. ENERGY 65

5.2. TIME OF ARRIVAL 68

5.2.1 From codes to timestamps *68*

5.2.2 Finding the first photon *70*

5.2.3 ToA Estimators *72*

CHAPTER 6 EXPERIMENTAL RESULTS 78

6.1. ELECTRO-OPTICAL CHARACTERIZATION 79

6.1.1 Monostable pulse width *79*

6.1.2 TDC linearity and resolution *80*

6.1.3 SPAD DCR and timing resolution *82*

6.1.4 Pixel jitter and skew *83*

6.2. SCINTILLATION CHARACTERIZATION 86

6.2.1 System functionality *87*

6.2.2 Performance optimization *93*

6.2.3 Suggested sensor configuration results *100*

CHAPTER 7 CONCLUSIONS 104

7.1. FUTURE WORK 106

BIBLIOGRAPHY 111

APPENDIX A 121

List of Tables

Table 1: comparison of the PET performance of state-of-the-art SiPMs using LYSO crystals of approximately 3×3×5 mm³. 10

Table 2: specifications and dimensions of all hierarchy levels of SPADnet-I. 54

Table 3: summary of SPADnet-I performance and main characteristics. 105

List of Figures

Figure 1: PET working principle. 2

Figure 2: Scintillation light pulse hitting the photosensor and its respective outputs. 3

Figure 3: analog (a) and digital (b) SiPM architectures. 8

Figure 4: expected flux (photon and noise) at the photosensor when an event occurs at time Θ. 13

Figure 5: expected counting output based on the flux of Figure 4, for a gamma event arriving at two different times, Θ_1 and Θ_2. 14

Figure 6: Typical energy spectrum obtained in a PET scanner, with an energy resolution of about 20% 15

Figure 7: Contour plot of the expected energy resolution versus the number of photon and noise counts. 17

Figure 8: Comparison of the photon timestamp PDF with and without detector jitter. 21

Figure 9: Comparison of the photon timestamp PDF with and without detector jitter and with and without temporal noise. 23

Figure 10: CRT considering only the photon emission process with 1000 total detected photons. 25

Figure 11: CRT with the first photon and the CRB considering only the photon emission process with a varying number of total detected photons. 26

Figure 12: CRT of the single timestamps (a) and of the CRB (b) considering the photon emission and the system jitter with 1000 total detected photons and different values of jitter. 26

Figure 13: CRT of the best single timestamp considering the photon emission and the system jitter versus the number of detected photons and the jitter FWHM. 27

Figure 14: order of the timestamp used in Figure 13. 28

Figure 15: CRT of the Cramér-Rao bound considering the photon emission and the system jitter versus the number of detected photons and the jitter FWHM. 29

Figure 16: CRT of the Cramér-Rao bound considering the complete model (photon emission, jitter and noise) with 1000 total detected photons and different values of jitter. 29

Figure 17: CRT of the Cramér-Rao bound considering the complete model (photon emission, jitter and noise) versus the number of detected photons and the noise rate, with a fixed jitter FWHM of 150 ps. 30

Figure 18: operation principle of SPADs in CMOS technology. 37

Figure 19: energy resolution versus SPAD active diameter for round SPADs based on [Ric+11]. Detection of 1000 and 1500 photons is assumed in an area containing 50k SPADs with 15 μm active diameter. 38

Figure 20: sensor hierarchy overview. 39

Figure 21: simplified schematic of the mini-SiPM architecture, highlighting the two compression schemes: spatial and temporal. 40

Figure 22: example of a mini-SiPM with 32 SPADs and a timing diagram. 43

Figure 23: examples of mini-SiPM floorplans with (a) round and (b) square SPADs. 43

Figure 24: block diagram of the sensor top-level. 45

Figure 25: state diagram of the discriminator. 45

Figure 26: block diagram of the pixel, also showing the required modifications at the mini-SiPM. 47

Figure 27: average probability of the TDC being triggered by a dark count before a gamma event arrives. 49

Figure 28: timing diagram of the sensor. 50

Figure 29: spatial compression loss estimation versus number of SPADs. 52

Figure 30: temporal compression loss estimation versus number of SPADs. 53

Figure 31: mini-SiPM schematics. 55

Figure 32: pixel schematics 56

Figure 33: TDC block diagram. 57

Figure 34: adder tree schematics. 58

Figure 35: top-level block diagram. 58

Figure 36: mini-SiPM electronics floorplan. 59

Figure 37: pixel floorplan. 60

Figure 38: floorplan of the first three levels of the adder tree. 61

Figure 39: floorplan of the last four levels of the adder tree. 61

Figure 40: SPADnet-I micrograph. 62

Figure 41: SPADnet-I chip with backside ball grid array. 63

Figure 42: example of SPADnet-I fitted response curve (compression curve). 67

Figure 43: raw count spectrum used to obtain the data points in Figure 42 and the resulting gamma energy spectrum. 67

Figure 44: example of a 3-bit TDC response and its possible fits. 69

Figure 45: cumulative distribution of timestamps obtained from a typical gamma event, with a rough indication of the gamma ToA. 70

Figure 46: intervals between the timestamps shown in Figure 45. 71

Figure 47: coincidence time distribution using the first photon as the gamma ToA, for three different parameter combinations of the first photon finder algorithm. 72

Figure 48: illustration of estimator steps with a typical Monte Carlo simulation-generated data: (a) comparison of expected and simulated CDF; (b) difference vector between the two. 75

Figure 49: Monte Carlo simulation results comparing the CRT of single-timestamp estimators with the proposed multiple-timestamp estimators. 76

Figure 50: sensor characterization assembly. 78

Figure 51: monostable pulse width versus control voltage. 79

Figure 52: TDC characterization: (a) histogram of all TDCs LSB resolution; (b) and (c) minimum, maximum and typical DNL and INL values, respectively. 81

Figure 53: characterization results of the SPAD DCR: (a) population distribution; (b) median and average values. 82

Figure 54: SPAD-only timing resolution for different excess bias voltages: (a) histogram of timestamps; (b) obtained FWHM versus excess bias. 83

Figure 55: jitter of a typical pixel for different SPAD excess bias voltages: (a) histogram of timestamps; (b) obtained FWHM versus excess bias. 84

Figure 56: histogram of all pixels jitter FWHM. 85

Figure 57: map of pixel skew (a) and jitter FWHM (b) at 3 V SPAD excess bias. 85

Figure 58: setup for gamma coincidence measurements. 86

Figure 59: examples of typical 511 keV, 1275 keV and pile-up events: (a)-(c) real-time energy output and (d)-(f) pixel counts. 87

Figure 60: pixel maps averaged for 75k events with energy around 511 keV: (a) pixel counts and (b) TDC triggering probability. 88

Figure 61: total sensor counts histogram for different discriminator thresholds. 89

Figure 62: measured distribution in time of 511 keV gamma events. 90

Figure 63: spatial and temporal compression loss assessment. 92

Figure 64: parameters of the sensor response fits for different SPAD excess bias and number of SPADs enabled: (a) k; (b) DC. Highest DCR SPADs disabled first. 94

Figure 65: DC of the sensor response equation fits for different SPAD excess bias and number of SPADs enabled. Lowest DCR SPADs disabled first. 95

Figure 66: energy resolution for different SPAD excess bias and number of SPADs enabled, with highest DCR SPADs disabled first. 96

Figure 67: energy resolution versus number of SPADs enabled at 2 V SPAD excess bias, with two sorting orders for disabling the SPADs: highest DCR first and lowest DCR first. 97

Figure 68: CRT versus number of SPADs enabled at 2 V SPAD excess bias, with highest DCR SPADs disabled first. 98

Figure 69: energy spectrum obtained with a ^{22}Na source, highlighting a 10.8% energy resolution at the 511 keV photo peak. 100

Figure 70: CRT obtained with the single- and multiple-timestamp estimators using different photon orders/different number of timestamps. 101

Figure 71: CRT histogram for the best single- (a) and multiple-timestamps (b) estimators. 102

Figure 72: Preliminary flood map of a 4×4 LYSO matrix with 1.12×1.12×13 mm^3 needles. No particular pixel to crystal alignment. 107

Figure 73: Energy resolution prediction with different sensor characteristics. 108

Figure 74: CRT prediction with different sensor characteristics. 109

List of PhD Publications

Patent applications

1. **L. H. C. Braga, M. Perenzoni, N. Massari, L. Gasparini, D. Stoppa, R. Walker, R. K. Henderson, "Dispositivo sensore fotonico" (Photon-sensing device), Italian Patent Application TO2013A000128, February 2013.**
2. **L. H. C. Braga, L. Pancheri, L. Gasparini, D. Stoppa, "Photodetector", European Patent Application EP2541219, priority date June 2011.**

Journals

1. **L. H. C. Braga, L. Gasparini, L. Grant, R. K. Henderson, N. Massari, M. Perenzoni, D. Stoppa, R. Walker, "A fully digital 8×16 SiPM array for PET applications with per-pixel TDCs and real-time energy output," IEEE Journal of Solid-State Circuits, January 2014.**
2. C. Bruschini, E. Charbon, C. Veerappan, L. H. C. Braga, N. Massari, M. Perenzoni, D. Stoppa, R. Walker, A. Erdogan, R. K. Henderson, S. East, L. Grant, B. Jatekos, F. Ujhelyi, G. Erdei, E. Lorincz, L. Andre, L. Maingaultg, V. Rebound, L. Verger, E. G. d'Aillon, P. Major, Z. Pepp, G. Nemeth, "SPADnet: Embedded coincidence in a smart sensor network for PET applications," Nuclear Instruments & Methods In Physics Research Section A: Accelerators, Spectrometers, Detectors And Associated Equipment, January 2014.
3. E. Gros-Daillon, L. Maingault, L. André, V. Reboud, L. Verger, E. Charbon, C. Bruschini, C. Veerappan, D. Stoppa, N. Massari, M. Perenzoni, L. H. C. Braga, L. Gasparini, R. K. Henderson, R. Walker, S. East, L. Grant, B. Jatekos, E. Lorincz, F. Ujhelyi, G. Erdei, P. Major, Z. Papp, G. Nemeth, "First characterization of the SPADnet sensor: a digital silicon photomultiplier for PET applications," Journal of Instrumentation, vol. 8 (12), December 2013.

Conference Proceedings

1. H. Xu, L. Pancheri, L. H.C. Braga, D. Stoppa, "Characterization of Single-Photon Avalanche Photodiodes in CMOS 150nm Technology," Proceedings of SPIE Photonics Europe 2014, April 2014 [in press].

2. E. Charbon, C. Bruschini, C. Veerappan, L. H. C. Braga, N. Massari, M. Perenzoni, L. Gasparini, D. Stoppa, R. Walker, A. T. Erdogan, R. K. Henderson, S. East, L. A. Grant, B. Játékos, F. Ujhelyi, G. Erdei, E. Lörincz, L. André, L. Maingault, V. Reboud, L. Verger, E. Gros d'Aillon, P. Major, Z. Papp, G. Nemeth, "SPADnet: a fully digital, scalable and networked photonic component for time-of-flight PET applications," Proceedings of SPIE Photonics Europe 2014, April 2014 [in press].
3. **L. H. C. Braga, M. Perenzoni, D. Stoppa, "Effects of DCR, PDP and Saturation on the Energy Resolution of Digital SiPMs for PET," Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2013 IEEE, October 2013 [in press].**
4. **L. H. C. Braga, L. Gasparini, L. Grant, R. K. Henderson, N. Massari, M. Perenzoni, D. Stoppa, R. Walker, "Complete characterization of SPADnet-I – a digital 8×16 SiPM array for PET applications," Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2013 IEEE, October 2013 [in press].**
5. L. Gasparini, L. H. C. Braga, M. Perenzoni, D. Stoppa, "Characterizing Single- and Multiple-timestamp Time of Arrival Estimators with Digital SiPM PET Detectors," Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2013 IEEE, October 2013 [in press].
6. E. Charbon, C. Bruschini, C. Veerappan, L. H. C. Braga, N. Massari, M. Perenzoni, D. Stoppa, R. Walker, A. Erdogan, R. K. Henderson, S. East, L. Grant, B. Jatekos, F. Ujhelyi, G. Erdei, E. Lorincz, L. Andre, L. Maingaultg, V. Rebound, L. Verger, E. G. d'Aillon, P. Major, Z. Pepp, G. Nemeth, "SPADnet: A Fully Digital, Networked Approach to MRI Compatible PET Systems Based on Deep-Submicron CMOS Technology," Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2013 IEEE, October 2013 [in press].
7. R. Walker, L. H. C. Braga, A. T. Erdogan, L. Gasparini, L. A. Grant, R. K. Henderson, N. Massari, M. Perenzoni, D. Stoppa, "A 92k SPAD Time-Resolved Sensor in 0.13µm CIS Technology for PET/MRI Applications," Image Sensor Workshop, 2013 International, June 2013.
8. **L. H. C. Braga, L. Gasparini, L. Grant, R. K. Henderson, N. Massari, M. Perenzoni, D. Stoppa, R. Walker, "An 8×16-pixel 92kSPAD time-resolved sensor with on-pixel 64ps 12b TDC and 100MS/s real-time energy histogramming in 0.13µm CIS technology for PET/MRI applications," Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2013 IEEE International, pp. 486-487, February 2013.**

9. **L. H. C. Braga, L. Gasparini, D. Stoppa, "A time of arrival estimator based on multiple timestamps for digital PET detectors," Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2012 IEEE , pp.1250-1252, October 2012.**
10. **L. H. C. Braga, L. Pancheri, L. Gasparini, M. Perenzoni, R. Walker, R. K. Henderson, D. Stoppa, "A CMOS mini-SiPM detector with in-pixel data compression for PET applications," Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2011 IEEE, pp.548-552, October 2011.**
11. **L. H. C. Braga, L. Pancheri, L. Gasparini, R. K. Henderson, D. Stoppa, "A mini-SiPM array for PET detectors implemented in 0.35-µm HV CMOS technology," Ph. D. Research in Microelectronics and Electronics (PRIME), 2011 7th Conference on, pp.181-184, July 2011.**

Abbreviations

A/D	Analog-to-Digital
APD	Avalanche Photodiode
CDF	Cumulative Distribution Function
CMOS	Complementary Metal-Oxide-Semiconductor
CRB	Cramér–Rao bound
CRT	Coincidence Resolving Time
CT	Computed Tomography
DNL	Differential Nonlinearity
DOI	Depth of Interaction
FF	Fill-factor
FOM	Figure of Merit
FPGA	Field-Programmable Gate Array
FWHM	Full Width at Half Maximum
i.i.d.	Independent and identically distributed
INL	Integral Nonlinearity
LOR	Line of Response
LSB	Least Significant Bit
MLE	Maximum-Likelihood Estimator
MRI	Magnetic Resonance Imaging

PCB	Printed Circuit Board
PDF	Probability Density Function
PET	Positron Emission Tomography
PMT	Photomultiplier Tube
SiPM	Silicon Photomultiplier
SNR	Signal-to-Noise Ratio
SPAD	Single Photon Avalanche Diode
SSD	Solid-state Detector
TCSPC	Time-Correlated Single Photon Counting
ToA	Time of Arrival
ToF	Time of Flight
TSV	Through-Silicon Vias

Chapter 1 Introduction

The work presented in this thesis is inserted in the context of the European project SPADnet [Bru+14], which aims to develop a new generation of smart, large area networked photonic modules, primarily aimed at Positron Emission Tomography (PET) applications. The key component of the photonic modules is an array of fully digital photodetectors, which are then connected to a per-module FPGA for control and readout. These FPGAs also take care of the networking between the modules in the detector rings used in PET tomographers.

The scope of this thesis is the modeling, design and characterization of the photodetector for the SPADnet project. This detector is fully digital so as to directly communicate with the FPGA –with no need for external electronics –, and thus uses CMOS technology. Moreover, the photon detection device of choice is the Single Photon Avalanche Diode (SPAD) [Cov+96], which is not only CMOS-compatible, but also provides, as will be explained later in this thesis, the required sensitivity and timing resolution for the target application.

In the following sections, first the working principle of PET is explained, focusing on the requirements that a PET photodetector must meet. Then, a brief summary of the work in this thesis is presented and, finally, the organization of this thesis is described.

1.1. Positron Emission Tomography

Positron Emission Tomography (PET) is a nuclear imaging technique that utilizes annihilation gamma photons from positron decay to generate three dimensional functional images of the body. Its main applications are pre-clinical research, clinical oncology and brain function analyses [Wer+04].

PET is based on the tracer principle, which conveys the fact that radioactive compounds (tracers) take part on body processes in the same way as their non-radioactive counterparts do. This means that the radioactive emission from tracers can be used to image tissues where a specific cell function is occurring as, for instance, the elevated glucose metabolism in cancer cells [Wer+04]. Thus, PET is fundamentally different from other body imaging techniques such as Computed Tomography (CT) and Magnetic Resonance Imaging (MRI), which mainly provide body anatomic information.

The working principle of PET is briefly illustrated in Figure 1:. First, the tracer is injected into the subject, where the blood flow distributes it through the body according to its biochemical properties. Then, when a radioactive atom of the tracer decays, a positron is emitted from the nucleus and, after travelling a short distance (typically between a few tenths of a millimeter up to several millimeters [Phe06]), it combines with an electron. The process that follows is known as annihilation, in which both the positron and the electron are annihilated and a pair of 511 keV gamma photons is emitted in opposite directions (180^{o} apart).

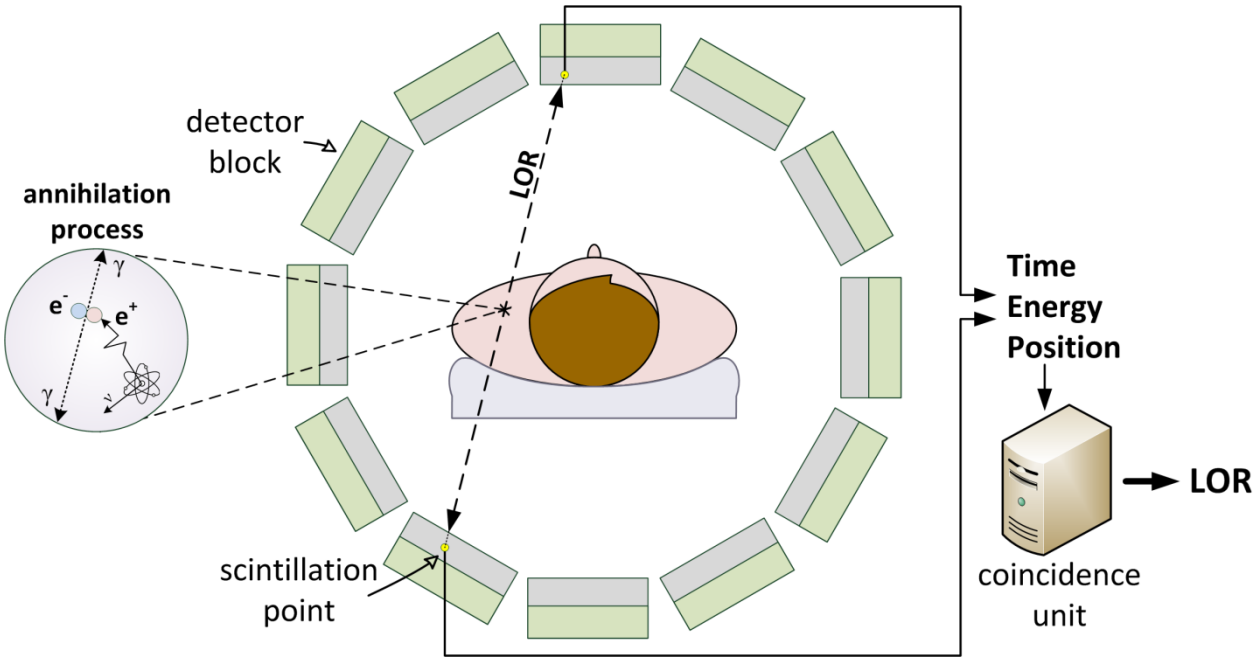


Figure 1: PET working principle.

The PET scanner needs to detect both emitted photons of the pair to establish the line of response (LOR) along which the annihilation took place. After millions of LORs are acquired, a tomographic 3D image of the subject can finally be formed, revealing the places where annihilations occurred (i.e. where the tracer concentration was higher).

To enable the detection of the photon pairs, PET scanners are normally constructed in the form of a ring of detectors, each of which needs to determine the energy, position and time of arrival (ToA) of the incoming gamma photons. This data is then fed to a coincidence unit, which is responsible for determining if any two detected photons are from a unique annihilation process. This is done by first selecting the photons with the correct energy, and then employing a coincidence timing window, usually a few nanoseconds wide [Lew08]. Finally, the LORs are generated based on the photons position information.

The detectors most widely used in PET scanners are scintillation detectors [Phe06], which are composed by a dense crystalline scintillator material coupled to a photodetector. The scintillator is a material that absorbs the incoming high-energy gamma photons and emits low-energy photons (light) as a result. The scintillation light is emitted isotropically in a short pulse in time, typically a couple hundred nanoseconds long [Phy11], as illustrated in Figure 2.

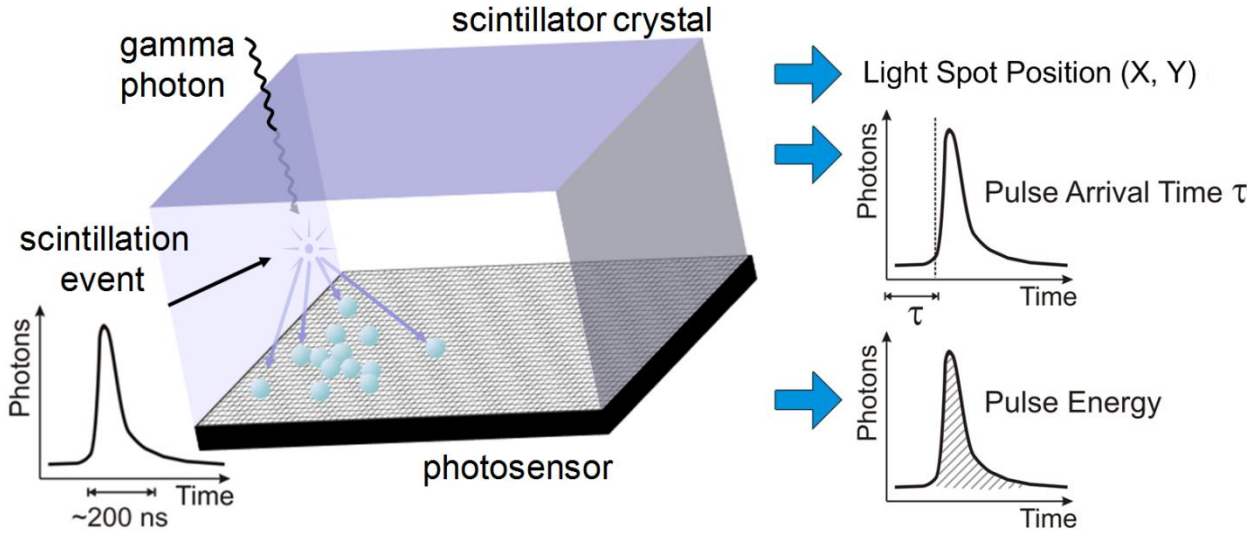


Figure 2: Scintillation light pulse hitting the photosensor and its respective outputs.

The amount of light photons that is emitted from a single 511 keV gamma absorption is typically very low, varying between 1k to 30k photons depending on the scintillator material [Phy11]. Therefore, the first requirement for PET photosensors is to possess a very high sensitivity in order to achieve a good signal-to-noise ratio (SNR).

Another important requirement for the photosensor concerns its timing performance. The recent development of bright and fast scintillators such as LSO, LYSO and $LaBr_3$ has enabled the usage of Time of Flight PET (ToF-PET), which explores the difference between the arrival times of the gamma pair to estimate the position along the line-of-response (LOR) where the annihilation took place. Therefore, to actually improve the signal-to-noise ratio (SNR) and image contrast with ToF-PET, the employed detectors must feature sub-ns timing performance [Mos07].

Moreover, as PET detectors can be up to tens of cm in size [Wer+04], the photosensors must also provide spatial information, so as to localize the scintillation point inside the crystal. The actual performance requirement in this case will depend on the crystal geometry, which is usually a matrix of small needle-sized crystals or a continuous crystal block.

Finally, it is important to note that PET is usually performed alongside CT for body anatomy information. A recent goal of biomedical imaging research is, however, the PET-MRI integration, as MRI, with respect to CT, offers better soft tissue differentiation and does not incur an additional dose of radiation to the patient [Pic+08]. This goal brings an additional requirement for PET photosensors: the compatibility with the magnetic fields generated by MRI.

1.2. The proposed solution

The PET detector developed in SPADnet is scintillator-based, and thus the requirements briefly summarized in the previous section for the light sensor all apply to the solution presented in this thesis. Moreover, the scintillator material selected to form the PET detector for SPADnet was LYSO. LYSO has several advantages that make it a popular scintillator in PET applications: high stopping power (density), high light yield and fast decay time, among others [Phy11]. LYSO is also non-hygroscopic, making its manipulation during experimental measurements much easier than with its hygroscopic counterparts, such as $LaBr_3$(Ce). On the other hand, LYSO contains the radioactive isotope ^{176}Lu, and thus emits background radiation which must be taken into account when designing the sensor architecture.

To best meet the requirements for a PET detector, a comprehensive modeling of the energy and timing performance of single-photon sensors for PET is initially performed. This modeling allows the definition of guidelines for the sensor design, aiming at the ideal parameter compromises. Based on the defined guidelines and requirements, a top-to-bottom architecture using CMOS technology and SPADs is proposed. The architecture incorporates in-pixel spatio-temporal compression of SPAD pulses for increased fill-factor, per-pixel timestamping of photons for improved timing resolution and top-level monitoring of the photon flux for efficient scintillation detection.

The first implementation of this architecture is done in the form of the SPADnet-I, which is a 8×16-pixel sensor fabricated in 0.13 μm 1P4M CMOS imaging technology [Bra+14]. The SPADnet-I is also able to offer a real-time output of the total detected energy that can be used for pile-up rejection and scintillator decay time estimation.

Given some of the innovative aspects of the proposed architecture, new concepts for data post-processing are developed, specially with regards to gamma time of arrival estimation. Using the

developed techniques, the complete characterization of the sensor is performed, where the best performance values obtained are an energy resolution of 10.8% and a coincidence resolution time of 288 ps.

1.3. Structure of the Thesis

The remainder of this thesis is organized as follows: the next chapter describes the state-of-the-art of the light sensors used in scintillator-based detectors for PET applications. Chapter 3 focuses on the PET detector requirements and how they translate into guidelines for the design of the SPADnet sensor, for which some performance models are presented. Chapter 4 details the sensor architecture and its first implementation, the SPADnet-I sensor. Chapter 5 discusses the challenges in processing the innovative data flow from the SPADnet sensor and the proposed solutions. Chapter 6 then presents the characterization results, both electro-optical and in gamma-detection, and, finally, Chapter 7 lays the conclusions and discusses possible future developments of this work.

Chapter 2 State of the Art

Historically, the most commonly used light sensors in PET scanners were photomultiplier tubes (PMTs) [Del+09]. A typical PMT is formed by a vacuum tube containing a photocathode, which emits an electron for each incoming photon, followed by electron multipliers, which multiply each electron up to millions of times [Phe06], and an anode, which is the collector electrode. The very high gain provided by the electron multipliers equates to a very high sensitivity, along with low noise and fast response.

However, since PMTs are composed of vacuum tubes, they are somewhat bulky and fragile. In addition, they also require power supplies of many hundred volts and are sensitive to magnetic fields – meaning their use in PET-MRI scanners is difficult. Due to these disadvantages, solid-state detectors (SSDs) have long been proposed as an alternative to PMTs [Lig+86].

SSDs are intrinsically compact and rugged, besides being insensitive to magnetic fields and usually requiring lower operating voltages. One of the first SSDs to be proposed as a light sensor in PET scanners was the avalanche photodiode (APD). APDs provide reasonable timing resolution and gain, which are, however, substantially worse than in PMTs [Ber+08]. As such, research in the field of SSDs for PET has been very active in the last years, and a new type of SSD recently suggested for PET has been showing promising results: the Silicon Photomultiplier (SiPM) [Ott+04]. As this is the type of detector targeted in this thesis, a detailed discussion about its state-of-the-art is given in the next section.

2.1. Silicon Photomultipliers

SiPMs are formed by a densely packed array of APDs working in Geiger mode (i.e. biased above their breakdown voltage) and connected in parallel, as schematically shown in Figure 3(a). When a single photon is absorbed by a Geiger mode APD, a very fast avalanche is triggered, generating a current pulse. Due to this characteristic, Geiger mode APDs are also known as single-photon avalanche diodes (SPADs). Moreover, as the SPADs are connected in parallel, when a scintillation occurs, a current signal builds up at the SiPM output proportional to the number of SPADs triggered, resulting in an output pulse similar to the one seen with PMTs [Lew08]. This pulse then needs to be processed through external electronics for the estimation of the gamma time, energy and position information.

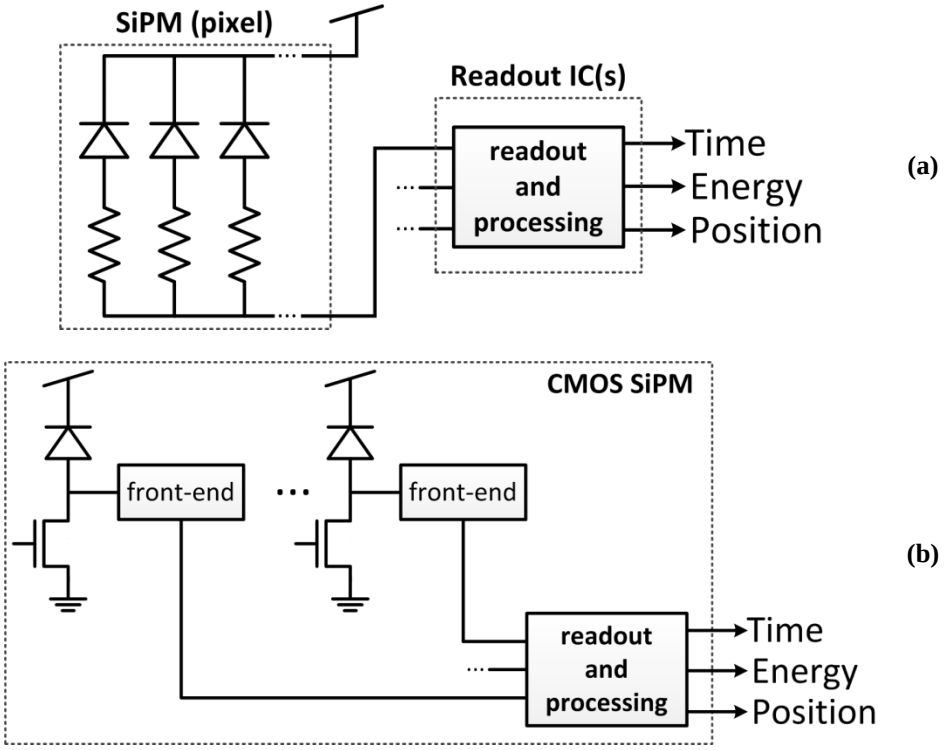


Figure 3: analog (a) and digital (b) SiPM architectures.

The performance of PET detectors heavily depends on the type and dimension of the scintillator crystal used in the measurements. Therefore, for an unbiased figure-of-merit (FOM) comparison between SiPM-based PET photosensors, an LYSO crystal with 3×3×5 mm³ size will be used as a standard. Focusing first on the detectors coincidence resolving time (CRT, also known as timing resolution), [Sei+12a] reports a CRT of 138 ps using Hamamatsu SiPMs, while [Yeo+12] reports 183 ps using SensL devices and [Gol+13] obtained 186 ps with FBK-SRS SiPMs. Other works have focused on energy resolution characterization, another important FOM for PET, with [Ser+13] reporting 10.2% also with FBK-SRS SiPMs, and [Szc+13] reporting 10.5% with Hamamatsu sensors (with a 5×5×5 mm³ crystal, however).

Other companies are also working on SiPM development, such as Excelitas Technologies [Exc14], or KETEK [Ket14]. In general, though, the performance of the various SiPM manufacturers in PET applications is relatively similar, and the above comparison of SiPMs in similar measurement conditions is very representative of the state-of-the-art of the technology.

Still, the intrinsic photon counting capability of SPADs is not fully exploited with SiPMs, as the analog-to-digital (A/D) conversion is only performed on the final summed current output, through external electronics, and is therefore subject to electronic noise. Since the SPAD output

is only able to distinguish between a photon and no photon (i.e. it is an intrinsically binary output), performing the A/D conversion at each individual SPAD can significantly improve the noise performance of the system. This approach has been recently pursued in [Fra+09], with the so-called “digital SiPM”, schematically shown in Figure 3(b).

The digital SiPM takes advantage of CMOS technology to perform a 1-bit A/D conversion per SPAD and to integrate an on-chip digital accumulator that produces the sensor energy output. In addition, the timing information is also generated on-chip, by a time-to-digital converter (TDC), and there are per-SPAD memories that can disable noisy devices, further improving performance and device yield. One disadvantage of the digital SiPM is that the fabrication process cannot be fully customized for optimum SPAD performance, as is the case of the dedicated analog SiPM, since the digital SiPM requires CMOS technology.

Up to now, only one group has successfully developed and characterized a digital SiPM for PET, reporting a CRT of 153 ps and an energy resolution of 10.4% [Hae+12], also with a 3x3x5 mm^3 LYSO crystal. Other groups have also been pursuing the digital SiPM approach [Man+12], [Bér+12] without, however, having reported PET characterization results yet. Finally, CMOS SiPMs have also been reported for different applications, such as fluorescence lifetime imaging [Tyn+12].

In Table 1, the aforementioned performance numbers of SiPM detectors are summarized. As is clear from this comparison, the performance of the various SiPM manufacturers is not that disparate, and in fact the similarity between them indicates that SiPMs may be approaching the physical limits of the PET working principle and of the LYSO scintillator, for instance. Nonetheless, the digital SiPM technology offers interesting opportunities for the development of features that go beyond typical SiPM design: with the possibility of on-chip electronics integration, system-level improvements can be developed that reduce the system complexity and cost, or enable increased data collection from scintillation events, for instance. The availability of this new data flow may then foster the progress of alternative processing techniques that can bring future breakthroughs to the PET field.

Table 1: comparison of the PET performance of state-of-the-art SiPMs using LYSO crystals of approximately 3×3×5 mm³.

Manufacturer	Refs.	Analog (A)/ Digital (D)	CRT	Energy resolution
Hamamatsu	[Sei+12a], [Szc+13]	A	138 ps	10.5%
SensL	[Yeo+12]	A	183 ps	-
FBK-SRS	[Gol+13], [Ser+13]	A	186 ps	10.2%
Philips PDPC	[Hae+12]	D	153 ps	10.4%

Chapter 3 Gamma Detection in PET Scanners

The goal of a PET photodetector is to sense the arrival of a gamma photon, and then estimate three of its features: energy, time of arrival and incident position. Given the structure of scintillation detectors, these tasks are all appointed to the light sensor, which must perform them based on the light incoming from a scintillation event. In the following sections, the main issues regarding each of these tasks will be analyzed, along with the resulting requirements for the sensor. The main goal of these analyses will be to define a set directions for the sensor architecture and design, which will then be discussed in Chapter 4.

3.1. Gamma Discrimination

As gamma photons arrive randomly in time, in a completely asynchronous fashion, it is crucial that the PET photodetector be event-driven so as to only provide data to the system when an actual 511 keV scintillation occurs. This ensures that the next level of the system hierarchy (in the case of SPADnet, the module FPGA) is not overflown with data. Moreover, depending on the sensor architecture, the sensor readout operation may result in a detection dead time, which can be further detrimental to the PET system performance. To identify the best strategy for gamma discrimination, in the next paragraphs a model for the scintillation event will be defined.

The light pulse emitted from a scintillator when a gamma photon is absorbed can usually be described as the convolution of two exponential functions [Hym65]: one representing the gamma energy transfer – which translates into the pulse rise time – and another representing the crystal radiative decay – which translates into the pulse decay time. The equation for the photon flux $F_{scin}(t)$ reaching the photosensor can then be written as in (1), where Θ is the time of absorption of the gamma photon, τ_r is the scintillator rise time, τ_d is the scintillator decay time and N_{ph} is the total number of detected photons after a sufficiently long integration time (i.e. much larger than τ_d).

$$F_{scin}(t) = N_{ph} \cdot \frac{e^{-\frac{t-\Theta}{\tau_d}} - e^{-\frac{t-\Theta}{\tau_r}}}{\tau_d - \tau_r} \tag{1}$$

At the same time, the photodetectors are constantly subject to noise, both from the readout circuits as from the photodetection devices (e.g. the SPADs) themselves. The combination of these noises can manifest itself as a signal equivalent to that of a few light photons, resulting in overall flux similar to the curve shown in Figure 4.

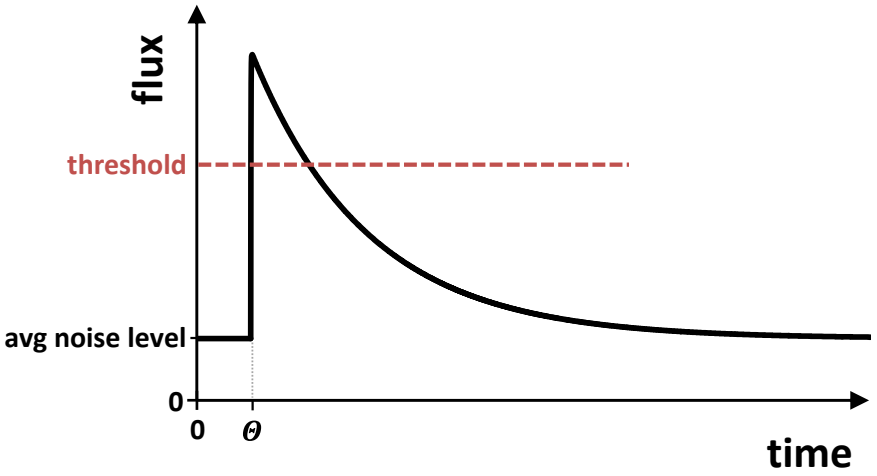


Figure 4: expected flux (photon and noise) at the photosensor when an event occurs at time Θ.

The first point to notice from this graph is that the gamma discrimination function may be a simple threshold, as long as the scintillation flux is sufficiently stronger than the noise level. In other words, the first requirement of the sensor is that its signal-to-noise ratio must be high enough so that the random variations in the noise level do not generate false positive events.

However, intrinsic to the discrimination concept described above is that the sensor must be able to detect the flux of incoming photons. This is a not obvious feature in typical image sensors (e.g. standard CMOS image sensors [ElG+05] or SPAD pulse counters [Sto+09a], [Pan+11]), which are integrating sensors, that is, they contain an analog integrator or digital counter at their output. These sensors would have an output with the form of Figure 5 (i.e. the integral of Figure 4). As the arrival time Θ occurs randomly in time, one does not know when to "start counting" (or, more precisely, when to reset) so that a threshold can be efficiently compared.

Finally, depending on the crystal size and in the optical coupling between sensor and crystal, the scintillation photons will be spread in a relatively large area in the sensor. Therefore, the discrimination of a gamma event requires a sensor with (1), a high SNR, (2), photon flux monitoring, and (3), that this monitoring occurs on a relatively large area.

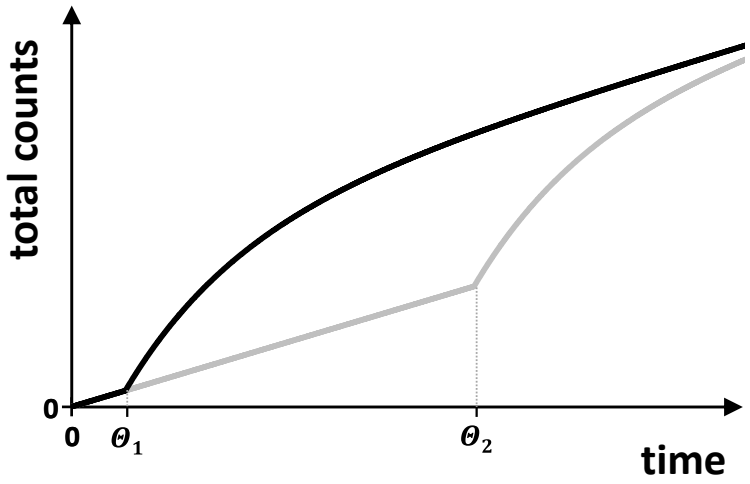


Figure 5: expected counting output based on the flux of Figure 4, for a gamma event arriving at two different times, Θ_1 and Θ_2.

3.2. Energy Estimation

Although identifying the gamma energy may not seem crucial to PET systems, since all gamma rays emitted from the annihilation process have 511 keV, gamma rays can also interact with matter through Compton scattering, which results in the gamma photon losing part of its energy and changing its travel direction. This means that there are a few possible scenarios for scintillation events at a PET detector:

(a) an unscattered gamma photon is fully absorbed by the scintillator through the photoelectric effect;
(b) a previously scattered (e.g. at the body) gamma is absorbed by the scintillator;
(c) a gamma photon goes through Compton scattering in the scintillator, and then escapes it;
(d) a gamma photon goes through Compton scattering in the scintillator and then is fully absorbed by it.

As should be expected, events of type (a) are the ideal ones, enabling the maximum SNR and a correct reconstruction of the LOR. Type (b) events must absolutely be discarded, since they changed direction along their path to the scintillator and would provide an incorrect LOR. Type (c) events could be used to reconstruct an LOR, even if their SNR would be lower due to the smaller deposited energy. However, there is no way to distinguish (c) events from (b) ones, and thus (c) events must also be discarded. Finally, for (d)-type events, the correct LOR could be re-

constructed if the position of the first deposition site could individualized. Given the speed of light of the gamma photons, though, current state-of-the-art sensors do not feature the required timing resolution for this.

The physics behind the Compton scattering process that occurs in scintillators results in scattered events of up to 340 keV [Wer+04]. Additionally, when using a scintillator with intrinsic radioactivity, the scintillator-emitted gamma photons will also generate events that need to be discarded. In the case of LYSO, its intrinsic radiation will emit photons with 88, 202 or 307 keV [Pre08].

Therefore, in PET systems, low-energy events must be distinguished from unscattered, 511 keV gamma absorptions and then discarded. It should be noted that (d)-type events actually cannot be distinguished through their energy, as the full 511 keV were deposited in the scintillator. Therefore, the scintillation position information, which will show two separate deposition sites, must be used to discard these events.

A typical energy spectrum obtained in a PET system is shown in Figure 6 [Med10]. As can be observed in the graph, both the scattered range upper-limit of 340 keV as well as the 511 keV peak are not very sharp, and are actually merged. To describe this energy estimation uncertainty, the energy resolution figure-of-merit (FOM) is typically used, which is obtained by dividing the full-width-at-half-maximum (FWHM) of the 511keV peak of the energy spectrum by the peak value itself. As such, the smaller is the detector energy resolution, the better is its energy estimation.

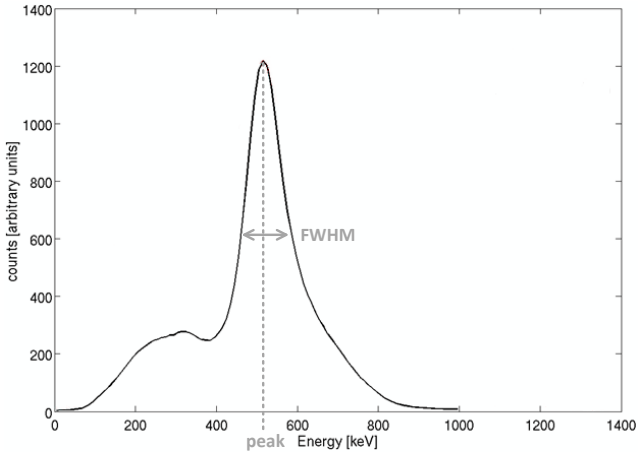


Figure 6: Typical energy spectrum obtained in a PET scanner, with an energy resolution of about 20%.

The energy estimation uncertainty in a PET detector is the result of the many stochastic processes present between the emission of light photons by the scintillator and their detection by the photosensor. Moreover, as shown by the merging of the peaks in the spectrum above, this uncertainty leads to a non-optimal filtering of low-energy events, possibly leading to a deterioration of the final PET image quality. Therefore, an investigation of these processes is merited, and will be performed next. To simplify this discussion, the photosensor will be assumed fully digital, i.e. it will be considered a digital counter with negligible readout electronic noise.

In a PET detector, the energy estimation comes from integrating the incoming flux shown in Figure 4. Therefore, two main processes will contribute to the final estimation: the photon flux itself and the photodetector noise. Given the digital counter assumption, the main source of uncertainty in both these process will be shot noise, which follows a Poisson distribution. Moreover, the scintillator itself is also a source of uncertainty due to the intrinsic variation in the number of emitted low-energy photons for the same absorbed gamma energy.

As these three processes are independent and uncorrelated, their variances can be summed to obtain the total energy variance. From this, and since in a Poisson distribution the variance is equal to the mean, the energy resolution of a detector can be written as in (2).

$$En_{res} = \frac{2.35}{En_{peak}} \cdot ({\sigma_{sensor}}^2 + {\sigma_{scin}}^2)^{\frac{1}{2}} = \left[\left(2.35 \cdot \frac{\sqrt{N_{ph} + N_{noise}}}{N_{ph}} \right)^2 + (En_{res})_{scin}^2 \right]^{\frac{1}{2}} \quad (2)$$

In the above equation, N_{ph} and N_{noise} are the average number of counts due to photons and photodetector noise for one event integration, respectively, and $(En_{res})_{scin}$ is the intrinsic energy resolution of the scintillator. Moreover, to obtain the 2.35 factor that converts the standard deviation into the FWHM, it is assumed that the average number of counts is high enough so that the Poisson distribution can be approximated by a Gaussian distribution. It is also interesting to note that the sensor energy resolution is equal to the inverse of its SNR, apart from the multiplying 2.35 constant. Therefore, minimizing the energy resolution means increasing the SNR, which also improves the gamma discrimination function, as described in the previous section.

Equation (2) can be used to estimate the achievable energy resolution using N_{ph} and N_{noise} as input parameters. This is shown in the contour plot in Figure 7, where the scintillator was considered LYSO, with an intrinsic energy resolution of 8% [Nas+07].

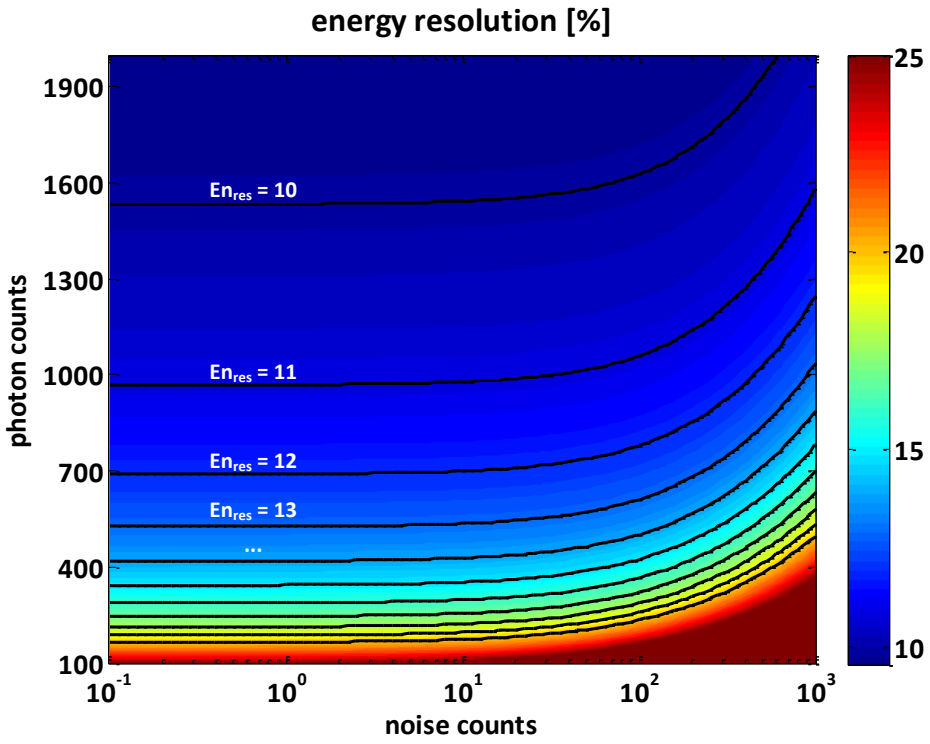


Figure 7: Contour plot of the expected energy resolution versus the number of photon and noise counts.

The graph shows that up to about 10 noise counts per event, the energy resolution is not affected by photodetector noise. From this point onwards, the resolution starts worsening relatively fast with the noise counts in log scale, in accordance with the square root relation of equation (2). On the other hand, with respect to photon counts, the energy resolution changes very slowly for high photon counts, and then worsens somewhat exponentially as photon counts go down in linear scale (this is shown by the constant resolution black lines getting closer to each other).

These considerations make it clear that the intrinsic scintillator resolution is by far the most limiting factor for the detector energy resolution, at least when considering state-of-the-art SiPMs such as the ones referenced in Chapter 2. Nonetheless, to achieve an energy resolution at least as good as the state-of-the-art (i.e. slightly above 10%), the digital counter sensor must detect around 1500 photons with not much more than 100 noise-generated counts.

3.3. Time of Arrival Estimation

Obtaining the time of arrival (ToA) of a gamma photon is crucial to the workflow of a PET scanner, since the coincidence unit of the system uses this information to match two detected gammas in opposite detectors of the ring, as explained in Chapter 1. Similarly to the energy resolution, the timing resolution of a detector indicates the uncertainty with which the time of arrival information is estimated. Therefore, the better is the timing resolution of the system, the smaller can the coincidence window be made, thus reducing the amount of false coincidences and improving the image quality of the system.

From the perspective of the coincidence window alone, i.e. with standard PET scanners, even a relatively low resolution of a few nanoseconds can provide a reasonably good imagine quality [Lew08]. However, in ToF-PET the detector timing resolution has a much greater importance, as the time difference between the pair of detected gammas is used to estimate the position of the annihilation process along the LOR. This way, the detector timing resolution directly impacts the final image quality.

For an effective ToF-PET detector, the timing resolution must be at most a few hundred picoseconds [Mos07], since, for instance, a resolution of 500 ps results in an uncertainty (FWHM) of about 7.5 cm, which is still an order of magnitude higher than the resolution achieved by PET systems through the processing of several LORs [Spa+10]. Therefore, in recent years there has been a strong research push towards improving timing resolution in PET scanners.

Before modeling the ToA estimation, an important distinction must be made between the system (or coincidence) timing resolution and the single detector resolution. The final desired figure, i.e. the time difference between the pair of detected gammas, is obtained from the subtraction of the ToA estimations of the two gammas in coincidence. As such, the system timing resolution – also known as the coincidence resolving time, the CRT – is composed by a linear combination of two stochastic processes. As the two ToA estimations are completely independent, assuming also that the two opposite detectors are identical, the CRT can be written as in (3).

$$CRT = 2.35 \cdot \delta_{system} = 2.35 \cdot \left({\delta_{det}}^2 + {\delta_{det}}^2\right)^{\frac{1}{2}} = 2.35 \cdot \sqrt{2} \cdot \delta_{det} \tag{3}$$

This again assumes that the underlying distribution is Gaussian, and thus the 2.35 factor is used to convert the standard deviations δ_{system} and δ_{det} into the FWHM, which describes the CRT. It should be noted that this conversion is not usually necessary during characterization or in simulations, as the actual CRT distribution can be obtained and the FWHM taken directly. Nonetheless, as will be shown later, this distribution is anyway typically very well fitted by a Gaussian.

Given the above considerations, in the next subsections the single detector timing resolution, described by δ_{det}, will be modelled and analyzed, so as to identify the best compromises in the detector design. This modelling will focus on three processes: the emission of photons by the scintillator, the detection of the photons by the sensor and the sensor temporal noise, which may generate noise timestamps that are indistinguishable from photons ones.

3.3.1 Timestamp Model

The first and most important process that affects the timing resolution in PET detectors is the emission of photons by the scintillator. This process was already described in terms of the flux that reaches the sensor, in section 3.1. To model the uncertainty in the timing estimation, equation (1) can be rewritten as the probability density function (PDF) of the emission of each photon $p_{em}(t|\Theta)$, with each emission being independent and identically distributed (i.i.d.) [Fis+10].

$$p_{em}(t|\Theta) = \frac{e^{-\frac{t-\Theta}{\tau_d}} - e^{-\frac{t-\Theta}{\tau_r}}}{\tau_d - \tau_r} \tag{4}$$

The corresponding cumulative distribution function (CDF) can be written as:

$$P_{em}(t|\Theta) = \int_{-\infty}^{t} p_{em}(\hat{t}|\Theta)d\hat{t} = 1 - \frac{\tau_d \cdot e^{-\frac{t-\Theta}{\tau_d}} - \tau_r \cdot e^{-\frac{t-\Theta}{\tau_r}}}{\tau_d - \tau_r} \tag{5}$$

After the photon is emitted, it will travel through the scintillator crystal until it reaches the photosensor. For simplification purposes, in this work the photon transport time in the crystal will be considered negligible. Next, the photon will be detected by the photosensor. Similarly to the model in section 3.2, the sensor model will comprise an ideal digital counter that is also able to timestamp all detected photons. Between the photon absorption by the sensor and its timestamp-

ing, different processes will occur at the sensor (e.g. SPAD avalanche, signal transmission, TDC triggering) that will contribute to jitter in the final timestamp. As these processes are generally random and independent, for the proposed model they will be approximated by a single Gaussian distribution $p_{jitter}(t)$, with mean μ_{jitter} and standard deviation σ_{jitter}, as written in equation (6) [Sei+12a]. Considering that the photon is absorbed at $t = 0$, to maintain causality, it is assumed that $\mu \gg \sigma$ (so that $p_{jitter}(t < 0) = 0$).

$$p_{jitter}(t) = \frac{1}{\sigma_{jitter}\sqrt{2\pi}} \cdot e^{-\frac{(t-\mu_{jitter})^2}{2\sigma_{jitter}^2}} \tag{6}$$

Since the photon timestamp is the sum of the photon emission with the detector jitter, the resulting distribution is given by the convolution of the two underlying distributions. Therefore, the PDF $p_{pt}(t|\Theta)$ and CDF $P_{pt}(t|\Theta)$ of the photon timestamp can be written as shown below, with the full mathematical development of these equations provided in Appendix A.

$$p_{pt}(t|\Theta) = \int_0^{t-\Theta} p_{em}(t-\tau|\Theta) \times p_{jitter}(\tau)d\tau \tag{7}$$

$$P_{pt}(t|\Theta) = \int_0^{t-\Theta} P_{em}(t-\tau|\Theta) \times p_{jitter}(\tau)d\tau \tag{8}$$

The effect of the detector jitter in the photon timestamp PDF is graphically shown in Figure 8. In this figure, the PDF is plotted for three cases: no jitter, a jitter of 200 ps FWHM and a jitter of 400 ps FWHM. As can be observed, the overall shape of the PDF hardly changes, however the rising time is made progressively slower. As will be shown in the analysis section below (3.3.3), this causes a significant worsening of the system CRT.

The last process that can influence the timing resolution is the detector temporal noise, which has already been discussed in sections 3.1 and 3.2. Observing the expected flux at the sensor as shown in Figure 4, and given the assumption that the sensor can timestamp every single photon that it detects, it is expected that noise-generated timestamps can also occur.

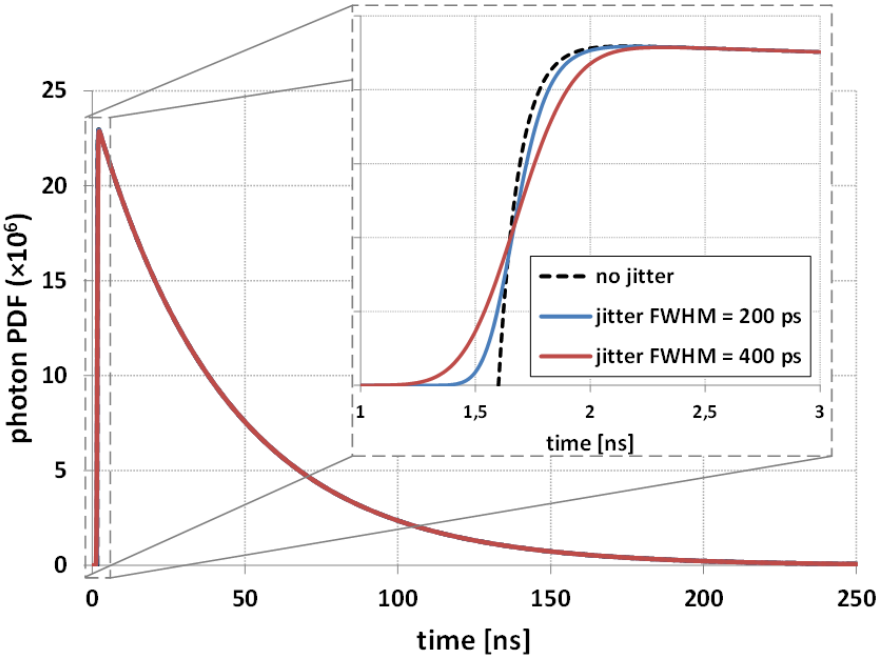


Figure 8: Comparison of the photon timestamp PDF with and without detector jitter.

This temporal noise manifests itself uniformly over time, and thus is described by the uniform distribution. However, for the PDF of the uniform distribution to have a non-zero value, it should be limited to a finite time interval. As such, for the proposed model, noise-generated timestamps will be assumed to occur only during a finite integration time t_{int}, that starts together with the scintillation and finishes after effectively all photons have been detected (i.e. $t_{int} \gg \tau_d$). This implies that, for the actual ToA estimation procedure, it is possible to remove any noise timestamp that occurred before the scintillation itself, while the ones generated during the scintillation event will be undistinguishable from photons. As will be shown later in the experimental results chapter, this is not an unrealistic hypothesis. The resulting noise PDF equation is shown in (9).

$$p_{noise}(t|\Theta) = \begin{cases} \frac{1}{t_{int}}, & \Theta \leq t \leq \Theta + t_{int} \\ 0, & elsewhere \end{cases} \tag{9}$$

From a system perspective, the obtained timestamps may now come from two independent components, that is, the system timestamps are randomly selected between photon and noise timestamps. The resulting distribution is defined as a mixture distribution [Frü06], with a PDF as shown in equation (10), where w_i are the weights of each underlying distribution $p_i(x)$.

$$p_{mix}(x) = \sum_{i=1}^{n} w_i \cdot p_i(x) \tag{10}$$

Given that the proposed model defines a fixed integration time for the noise distribution, the weights can simply be taken as the ratio between the number of photon and noise counts [Man+13]. Moreover, for simplification reasons, the mixed distributions are considered to be $p_{pt}(t|\Theta)$ and $p_{noise}(t|\Theta)$ directly, meaning that the noise timestamps are not convolved with the system jitter. This should not affect the final result, as the noise timestamps are already completely random in time, and additionally, it will allow for an interesting simplification of the final PDF equation. The photon plus noise timestamp distribution can then be written as:

$$p_{pnt}(t|\Theta) = \frac{N_{ph}}{N_{ph} + N_{noise}} \cdot p_{pt}(t|\Theta) + \frac{N_{noise}}{N_{ph} + N_{noise}} \cdot p_{noise}(t|\Theta) \tag{11}$$

It is interesting to note that, since N_{noise} is simply the integral of the noise flux during the defined integration time, a noise occurrence rate r_{noise} can be defined so that $N_{noise} = r_{noise} \cdot t_{int}$. Since $p_{noise}(t|\Theta) = 1/t_{int}$ during the integration time, equation 10 can be rewritten, for this same interval, as:

$$p_{pnt}(t|\Theta) = \frac{1}{N_{ph} + N_{noise}} \cdot \left(N_{ph} \cdot p_{pt}(t|\Theta) + r_{noise}\right) \tag{12}$$

Therefore, as should be expected, the final distribution does not directly depend on the integration time, except in the form of a normalizing constant that ensures that the PDF integrates to unity. Finally, to graphically illustrate the effect of the temporal noise, Figure 9 shows the resulting PDF with the gradual inclusion of the modelled effects: first for the photon emission only, then emission with detector jitter and, finally, emission with detector jitter and temporal noise. As previously discussed and evident in the graph, the temporal noise becomes a baseline for the emission with jitter distribution, which has its peak reduced so that the PDF equation still integrates to unity.

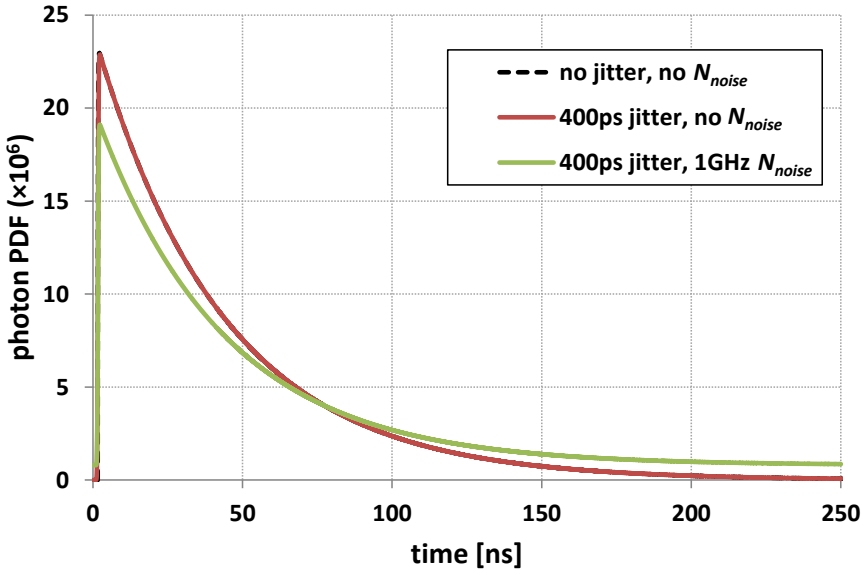


Figure 9: Comparison of the photon timestamp PDF with and without detector jitter and with and without temporal noise.

3.3.2 Methods for Timing Resolution Assessment

The model described in the previous subsection only provides a distribution for the timestamps that will be obtained by the system. Therefore, a method it is still required to assess how this distribution actually translates into the system timing resolution. Two different methods will be considered depending on how the detector estimates the gamma ToA. In the first scenario, the detector uses only a single timestamp for estimation, while in the second all timestamps can be obtained and then combined to provide an improved gamma ToA estimation.

Even if the second scenario may seem the best, in real systems, providing all timestamps can heavily degrade other system parameters, such as the photon detection efficiency (PDE) due to degraded fill-factor (FF). As this compromise completely depends on the actual implementation of the photosensor, it is out of the scope of this chapter and will be left for later discussion. The goal of this subsection is to provide the methods for assessing the timing resolution with respect to the various parameters present in the model described above, such as the jitter standard deviation and the noise occurrence rate, and not to compare the merits and disadvantages of different ToA estimation techniques.

Starting with the scenario of single-timestamp estimation, the theory of order statistics [Arn+92] can provide the tools to quantify the uncertainty that each ordered timestamp will exhibit. Considering the set of N i.i.d. timestamps, if these timestamps are then put in ascending order, the

PDF of the *i*-th order timestamp is given by equation (13), where $p(t)$ and $P(t)$ are the PDF and the CDF of the unordered timestamps.

$$p_{i|N}(t) = \frac{N!}{(i-1)!\,(N-i)!}\{P(t)\}^{i-1}\{1-P(t)\}^{N-i}p(t) \tag{13}$$

Considering a system where a single timestamp is used as the gamma ToA, but where this timestamp can be selected to be a specific order of the incoming flux of timestamps, the variance of the distribution in (13) can be used as an estimate of the detector timing resolution [Fis+10].

For a detector that can timestamp all its detected photons, the Cramér–Rao bound (CRB) can express the lower limit on the variance of an unbiased estimator of the gamma ToA Θ [DeG86]. In other words, given an unbiased estimator for the gamma ToA that combines all detected timestamps, its variance will be equal to or higher than the CRB. The CRB is defined as in equation (14), where $\widehat{\Theta}$ is the unbiased estimator of Θ, and $I_N(\Theta)$ is the Fisher information regarding Θ of all N timestamps.

$$var(\widehat{\Theta}) \geq \frac{1}{I_N(\Theta)} \tag{14}$$

The Fisher information of all N timestamps, in turn, is defined as in equation (15) [Sei+12b].

$$I_N(\Theta) = N \cdot \int_{-\infty}^{\infty}\left[\frac{\partial}{\partial\Theta}p(t|\Theta)\right]^2 \cdot \frac{1}{p(t|\Theta)} \cdot dt \tag{15}$$

It is interesting to note that it has been shown [Sei+12b] that the CRB can already be achieved with only a relatively small subset of the timestamps, namely the first dozen or so, depending on the detector parameters. Therefore, the CRB can be considered a reasonable assessment of a detector timing resolution even if the detector is not able to timestamp all incoming photons, but only the first ones in each scintillation event.

3.3.3 Analysis

Using the model and methods described in the previous subsections, it is now possible to investigate how the detector parameters will influence its timing resolution. Namely, the number of detected photons, the detector noise rate and the detector jitter will be analyzed with regards to their impact in the resolution. For all analyses, a LYSO scintillator will be assumed, featuring a rise time (τ_r) of 90 ps and a decay time (τ_d) of 43 ns [Sei+12b]. In addition, all resolution values will be reported in terms of CRT assuming two identical detectors so as to allow for easier comparison with experimental results both from this thesis as from the literature.

To first analyze the photodetector PDE (i.e. the number of detected photons N_{ph}) isolated from other parameters, the scintillation emission distribution $p_{em}(t|\Theta)$ described in (4) can be applied to the order statistics and CRB methods. In Figure 10, the FWHM of the first 30 orders along with the CRB translated to the FWHM assuming a Gaussian distribution is plotted for a total number of detected photons of 1000.

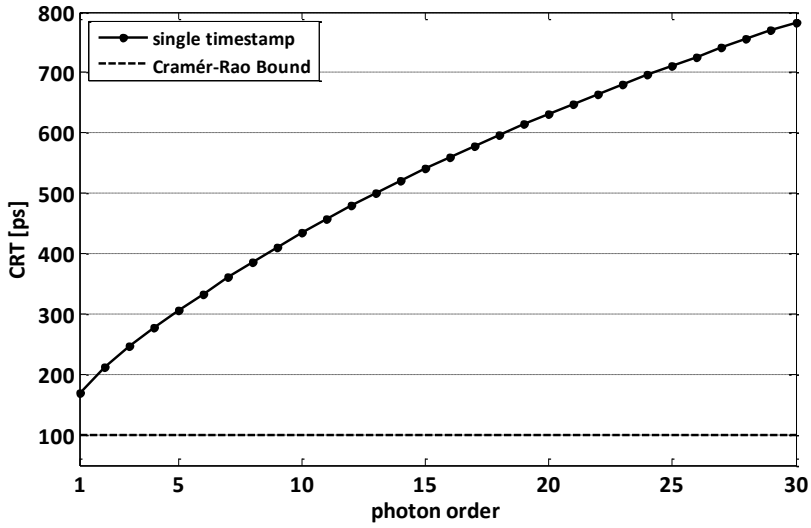


Figure 10: CRT considering only the photon emission process with 1000 total detected photons.

It can be observed that the first ordered statistic features the best single timestamp resolution, a behavior that can be shown to occur for any value of N_{ph}. Moreover, the first photon resolution is about 67% higher than the CRB, a behaviour which, however, changes depending of Nph. Thus, in Figure 11, the resolutions obtained with the CRB and with the first photon are plotted versus a varying number of total detected photons. In this graph, the first photon can be from up to 120% worse than the CRB, at the beginning of the *x*-scale ($N_{ph} = 100$), down to only 60%

worse at the maximum plotted number of total photons ($N_{ph} = 2000$). In any case, both curves present a plateauing effect towards high photon counts, indicating that gains in detector PDE after a certain point have diminishing returns.

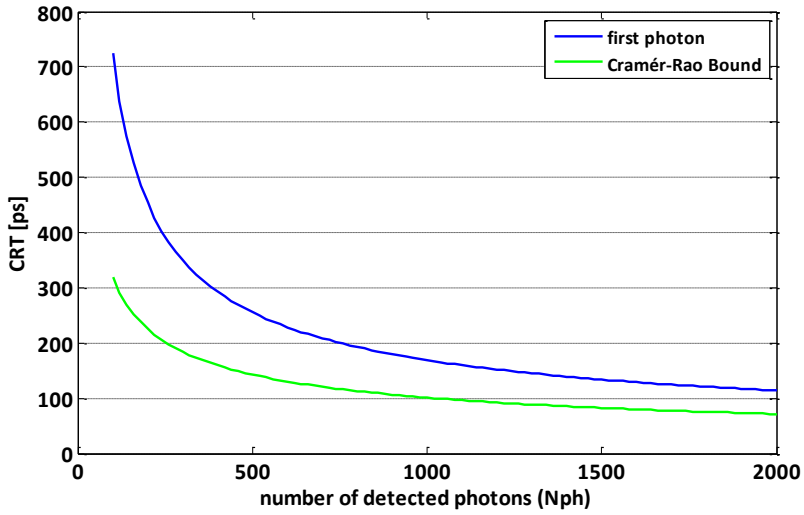
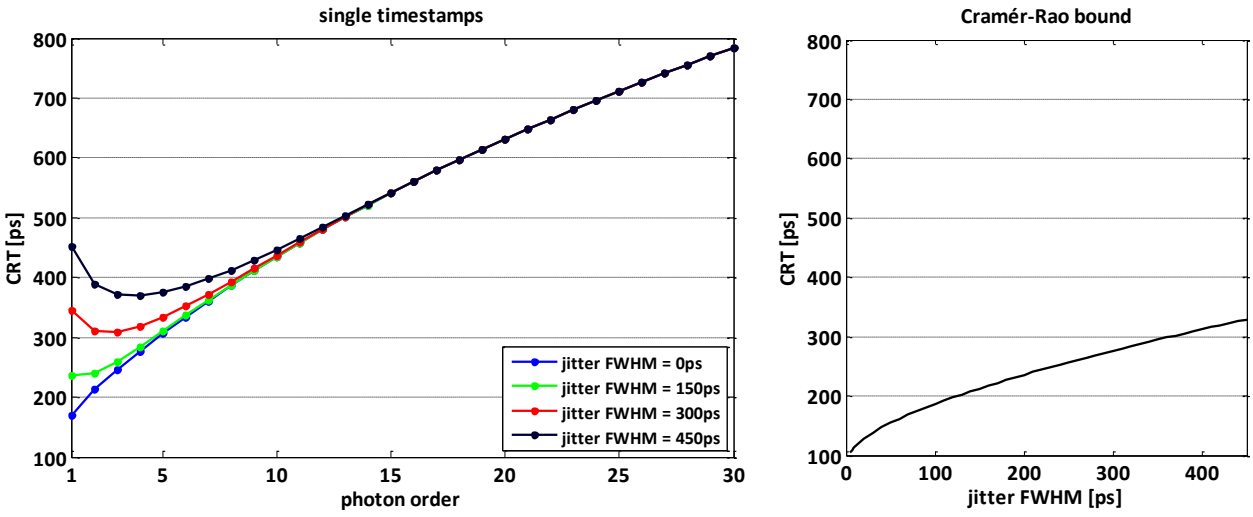


Figure 11: CRT with the first photon and the CRB considering only the photon emission process with a varying number of total detected photons.

Next, the photon with jitter distribution, $p_{pt}(t|\Theta)$, can be used to analyze the impact of the system jitter in the ideal detector CRT discussed above. Similarly to what was done in Figure 10, the graph below plots the single timestamps CRT versus photon order, as well as the CRB, with 1000 detected photons, but now for different jitter values.



(a) **(b)**

Figure 12: CRT of the single timestamps (a) and of the CRB (b) considering the photon emission and the system jitter with 1000 total detected photons and different values of jitter.

With respect to the single timestamps, the addition of jitter into the model results in not only the worsening of the best CRT, but also that the first photon is no longer the best, with the best order increasing as the jitter increases. Interestingly, though, the jitter only affects the CRT of the initial timestamps, with the resolutions from the 10^{th} order and above being approximately equal. Regarding the CRB, the resolution worsens less than linearly with jitter, but still, a jitter of around 400 ps FWHM can worsen the CRB resolution more than three times.

To obtain a complete picture of the CRT versus both the detector jitter and PDE, Figure 13 shows a contour plot of the single timestamp CRT versus these two parameters. To improve visualization, only the CRT of the best order is used, and thus, for reference, Figure 14 shows the order of the best timestamp CRT used in each data point.

As can be observed, the weight of the detector jitter in the CRT increases rapidly as the number of detected photons increase: for very low N_{ph}, the CRT is almost constant with respect to jitter, while at high photon levels the CRT quickly degrades with jitter values as low as 50 ps. Moreover, the trend shown with the photon-emission-only model in Figure 11 is also clear in Figure 13, where, for a constant jitter value, the CRT worsens approximately exponentially when N_{ph} decreases. To complete the analysis of the detector PDE versus jitter, Figure 15 plots the same contour plot as in Figure 13, but now using the CRT obtained through the CRB method.

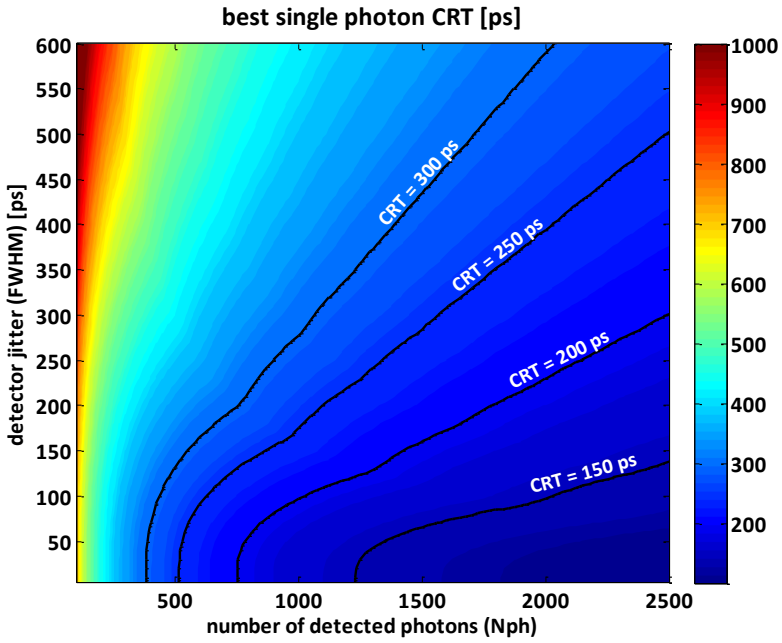


Figure 13: CRT of the best single timestamp considering the photon emission and the system jitter versus the number of detected photons and the jitter FWHM.

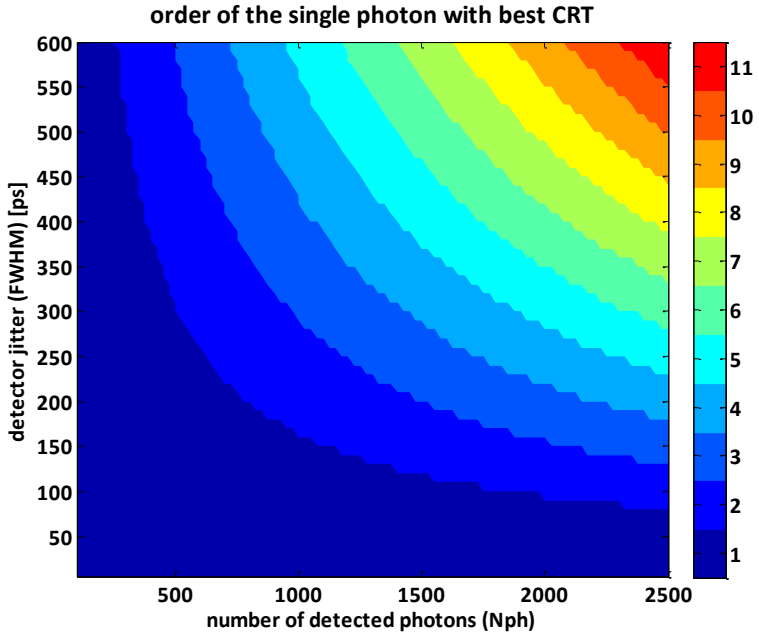


Figure 14: order of the timestamp used in Figure 13.

The CRB behavior differs from the best single photon one in two significant ways: first, as expected, a given CRT value can be obtained with a lower PDE or higher jitter, when compared to Figure 13. Secondly, for low values of jitter, the trend changes completely, with the CRB CRT continuously improving down to zero jitter. Of most importance, though, is that both analysis methods clearly show that, to obtain a CRT comparable with the state-of-the-art, not only is a reasonable PDE required – providing around 1500 detected photons – but also a low detector jitter, near or below 100 ps FWHM, should be targeted.

Finally, to analyze the effect of the temporal noise rate, r_{noise}, in the timing resolution, the distribution model $p_{pnt}(t|\Theta)$ is used. Due to the way that the noise uniform distribution was defined in this model, with a fixed starting point, the ordered timestamp method cannot be used to estimate the CRT, as the timestamp distributions will be strongly biased towards this starting point. The CRB method, on the other hand, evaluates the overall uncertainty that having noise during a scintillation event would add to an unbiased estimator, and thus does not depend on the noise starting point. As such, in Figure 16 the CRB CRT is plotted versus r_{noise} for 1000 detected photons and different values of the detector jitter.

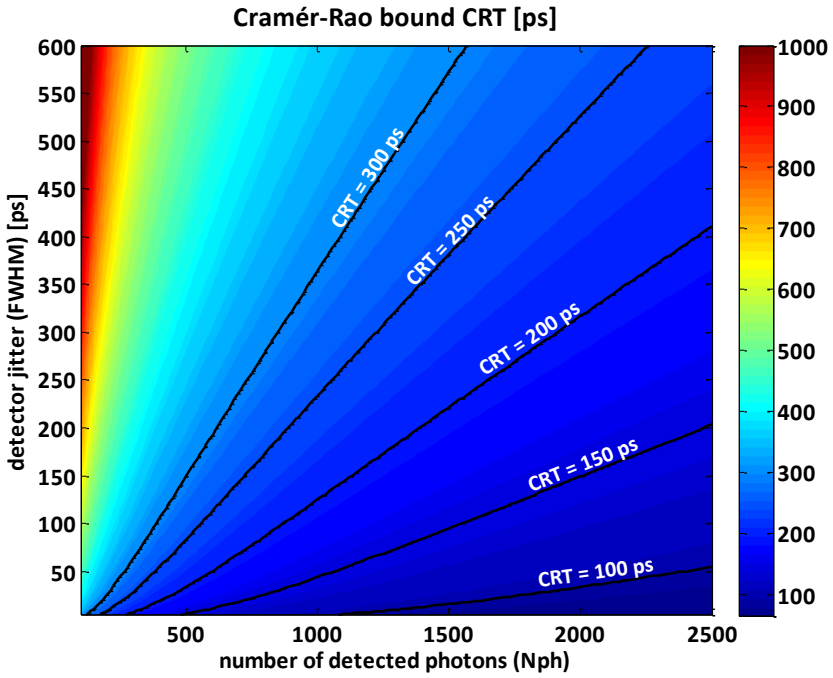


Figure 15: CRT of the Cramér-Rao bound considering the photon emission and the system jitter versus the number of detected photons and the jitter FWHM.

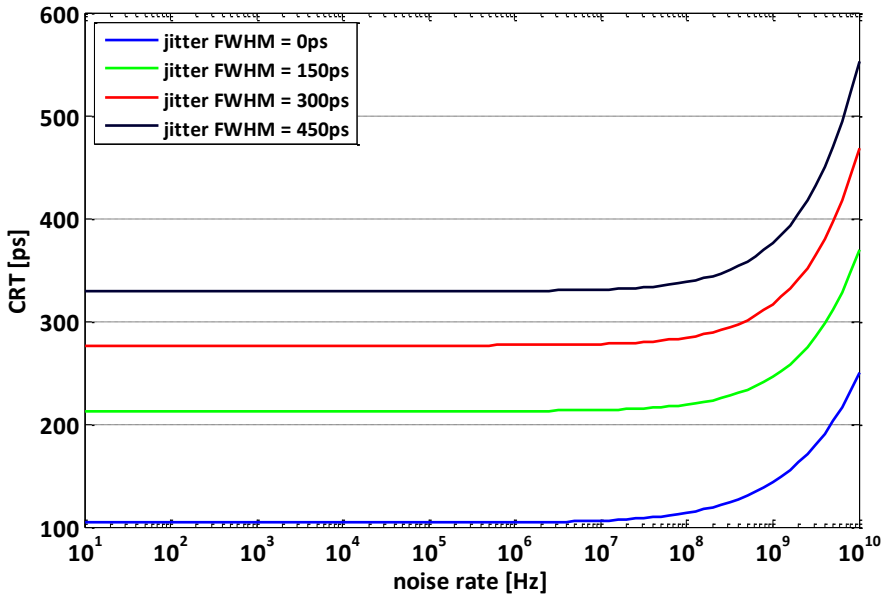


Figure 16: CRT of the Cramér-Rao bound considering the complete model (photon emission, jitter and noise) with 1000 total detected photons and different values of jitter.

The curve clearly shows that up to a noise rate of about 100 MHz, the temporal noise has little effect on the CRT. However, from this point onwards, the relationship is close to an exponential. This trend is repeated for all jitter values, with each curve differing only by its starting level.

Therefore, the next and final analysis plots the contour plot of the CRT versus the number of detected photons and the noise rate, with a fixed jitter FWHM value of 150 ps.

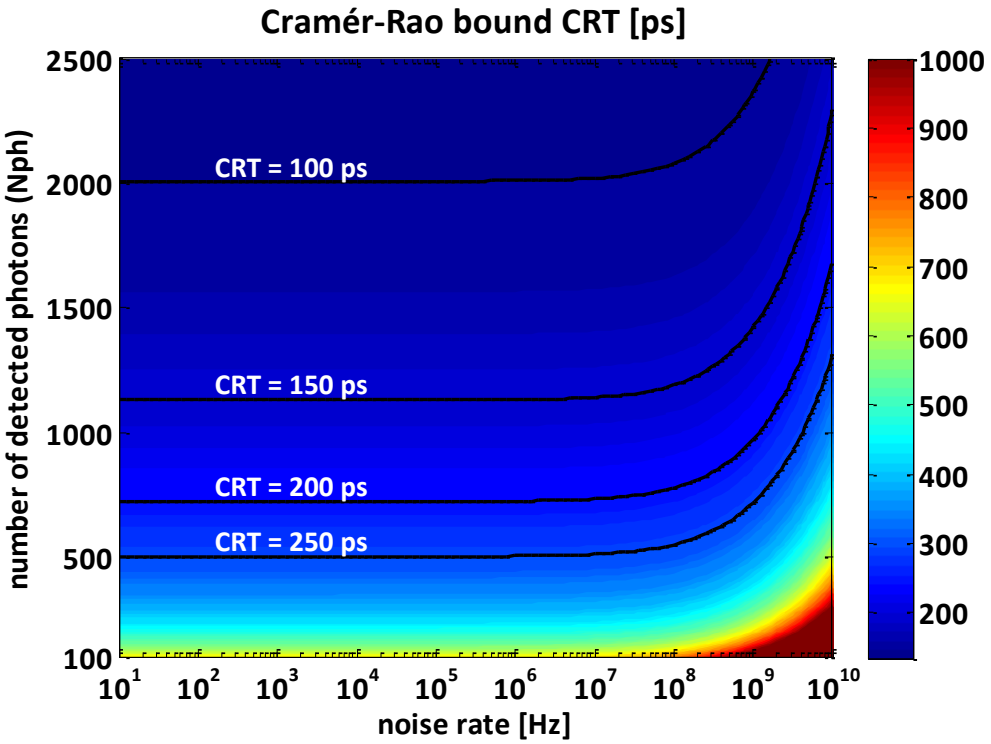


Figure 17: CRT of the Cramér-Rao bound considering the complete model (photon emission, jitter and noise) versus the number of detected photons and the noise rate, with a fixed jitter FWHM of 150 ps.

Figure 17 confirms the previously observed trend that the noise only starts affecting the CRT after around 100 MHz for any number of detected photons, although, for high photon counts, this trend is much less steep. It is also interesting to compare the above plot with the equivalent one for the energy resolution in Figure 7. Both are relatively similar, taking into account that, with a typical integration time of 150 ns, the top of the x-scale for the energy resolution, 1000 counts, equals a 7 GHz rate in the CRT x-scale. However, at high noise levels, the noise contribution to the CRT increases much more rapidly than to the energy resolution.

3.4. Position Estimation

The final piece of information that the photosensor needs to provide about the scintillation is the position. In a typical PET detector ring, as the one shown in Figure 1, each detector block can be as large as 25 cm on each side [Wer+04]. Therefore, knowing the scintillation X and Y position – i.e. the position with respect to the sensor plane – is essential to correctly build the LOR. The

precision of the LOR will be mainly given by the detector spatial resolution, which is then one of the key parameters determining the scanner final image quality. However, there are two physical limitations in the LOR reconstruction that need to be addressed first: positron range and gamma emission non-colinearity.

Positron range accounts for the small distance that a positron travels between its emission and is annihilation. As the gamma pair is emitted from the annihilation site, the LOR is actually indicating this position as opposed to the positron decay one. Even though the positron undergoes multiple direction changes until it is annihilated, the positron range is defined as the Euclidean distance between the decay and the annihilation sites, as the error in the LOR reconstruction is a function of this distance. Moreover, as it is a stochastic process, the value usually quoted as positron range is the FWHM of its distribution [Phe06]. Finally, the positron range depends on the radionuclide used, and the spatial resolution loss that it causes in the final PET image is usually between a few tenths of a millimeter up to several millimeters [Phe06].

The non-colinearity, on the other hand, accounts for the fact that the pair of gamma photons will not be emitted at exactly 180^o, and will, in fact, be emitted with a distribution of angles with a mean of 180^o and a FWHM of 0.5^o [Phe06]. Since the LOR reconstruction assumes perfectly opposite gammas, this effect will also produce an intrinsic spatial resolution error. This error depends on the diameter D of the PET ring, and can be approximated as $0.0022 \cdot D$ [Phe06]. Thus, for clinical PET systems with diameters in the order of 90 cm, the error becomes close to 2 mm, while for pre-clinical PET, with diameters closer to 20 cm, it is smaller than 0.5 mm.

In light of these effects, it is clear that the goal for the detector spatial resolution is considerably loosened depending on the PET application. Nonetheless, the factors influencing the detector spatial resolution will be discussed below, assuming a pre-clinical PET system, which presents a smaller intrinsic spatial error.

With respect to the photosensor in a PET detector, the main parameter that can affect the spatial resolution is the pixel size. In this case, the pixel is defined not as a single photodetection device (i.e. a SPAD), but as the smallest area of the sensor from which independent energy information can be obtained to form a two-dimensional energy map of the event. Additionally, the sensor PDE also affects the spatial resolution, as the higher the number of detected photons, the better the statistics to provide the position estimation.

However, the spatial resolution of scintillation detectors depends even more strongly on the crystal geometry, which can generally be categorized into two main groups: needle or continuous crystals. In the needles configuration, an array of small crystals (typically around 1 × 1 mm² in pre-clinical applications [Wer+04]) is tightly packed together and placed above the photodetector, with each needle surrounded with reflective material so there is no light sharing between needles. In the continuous configuration, a large, continuous crystal is coupled to the photodetector to form the detector block. In both cases, a light guide layer is usually inserted between the crystal and the photodetector with the aim of decreasing the photon density on the sensor surface [Lew08].

Due to the reflective coating in typical needle configurations, any X and Y intra-needle position information is lost. As such, the spatial resolution of a needle-based PET detector is a direct function of the size of the individual crystals. Additionally, the addition of a light guide results in the light from a single crystal spreading over an area much larger than the crystal itself. Therefore, the sensor pixels can be made quite large (e.g. 3.4 mm pixel pitch with 0.8 × 0.8 mm^2 crystals in a detector prototype [Son+10] or 3 mm pixel pitch with 1.1 × 1.1 mm^2 crystals in a commercial scanner [Nag+13]), while still providing enough information to correctly distinguish scintillations from different crystals. Alternatively, one-to-one coupling between crystal and pixel can be used, in which case the pixel pitch will match the crystal pitch [Lev+07].

On the other hand, the spatial resolution in continuous crystal configurations heavily depends on the photosensor ability to recognize the light spatial distribution on its surface. Therefore, for continuous crystal detectors, the pixel size is a much more critical parameter. Still, recent developments have demonstrated continuous crystal PET detectors with a spatial resolution of 0.9 mm with pixel pitch of 1.5 mm [Llo+10], or a resolution of 1.6 mm with a pixel pitch of 3.3 mm [Sch+09]. Therefore, also in the continuous crystal case, a pixel pitch in the order of one to few millimeters is sufficient for the application.

Finally, the scintillation position in the third dimension (Z) of the crystal also affects the precision of the LOR. Crystals used in commercial clinical PET scanners can have up to 30 mm thickness [Wer+04], and thus gamma photons arriving at oblique angles can cause parallax errors in the LOR reconstruction if the depth of interaction (DOI) information is not obtained. Still, many PET systems do not provide this information and simply assign the scintillation point source to a small distance from the front edge of the crystal [Lew08].

Many different methods have been conceived along the years to obtain the DOI in PET scanners. For instance, two different scintillators can be placed on top of each other, and their different decay times used to identify in which layer the scintillation occurred, in what are usually called phoswich detectors [Sao+99], [Jun+07]. Alternatively, two crystal arrays of the same type can be arranged on top of each other with a small offset, and thus the DOI can be obtained from the X and Y position [Liu+01]. Another approach is to place photosensors in more than one crystal surface, in so called double-sided readout detectors, and use the photon ratio between the sensors to estimate the DOI [Del+10].

Finally, in continuous crystal configurations, the light spatial distribution depends on the DOI, and thus can be used to estimate it, with no changes in the actual detector construction [Lin+08], [Ler+05]. Nevertheless, in all these DOI-capable examples, the sensor pixel pitch was in the same range as the ones previously described in this section, indicating that DOI capability does not incur a stricter requirement for the pixel pitch.

3.5. Other features

A few other sensor features can be valuable in a PET system. For instance, as mentioned in the previous section, obtaining the scintillator decay time can be an important tool for DOI determination in some crystal configurations. This requires a sensor that can monitor the incoming flux, which was actually already required for gamma discrimination (see section 3.1).

Another feature that can improve the final PET image quality is the ability to detect pile-up events. Pile-up events occur when two scintillations overlap in time, with the first one being from a low-energy gamma, so that the total detected energy during the integration time is similar to that of a 511 keV event. As such, the initial low-energy event cannot be discarded through energy windowing as discussed in section 3.2. The best alternative to distinguish these types of events from true 511 keV events is again through the incoming flux, which will have a second sharp rising edge during the integration time. Alternatively, also the spatial information can be used for this task, with the limitation, however, that the two overlapped events are sufficiently spaced from each other.

Finally, an additional very important parameter for a PET sensor is its efficiency in detecting gamma photons. Ideally, the sensor needs to be fully ready for a new event as soon as it finishes integrating the previous one, i.e. there should be no dead time due to data readout or processing.

Moreover, also the discrimination function should not incur any dead time, specially deriving from sensor noise. In this way, the sensor is always ready to receive an event, and is only busy when there is an actual scintillation occurring. This would maximize the sensor efficiency, allowing a higher gamma count rate to be detected, and thus enabling, for instance, a reduction in the PET exam time or an improvement in the statistics acquisition.

3.6. Summary

This chapter laid the main requirements of PET detectors and how they translate into sensor design parameters. The key conclusions were:

- Detecting the incoming flux at the sensor is critical, not only for gamma discrimination, but also for pile-up detection and even DOI estimation in phoswich detectors.
- The PET performance, both in terms of energy and timing resolution, showed a plateauing after a high enough number of detected photons per event. This means that there is room to add innovative features to the sensor, as their area occupation (and resulting reduction in PDE) will not necessarily have a strong negative impact on the performance. From a different perspective, this means that increasing the PDE after a certain point has diminishing returns, and should not be pursued at all costs.
- Similarly to the PDE discussion above, the sensor temporal noise only starts affecting performance after a given level, so even if it needs to be minimized, eliminating it completely should not be one of the main goals of the design.
- The sensor jitter, on the other hand, can significantly affect the timing resolution down to very low levels, and thus minimizing it should be high on the sensor design priorities.
- The intrinsic limitations of spatial resolution in PET systems considerably lessen the requirements on the sensor pixel pitch, and thus a pitch similar to the current available sensors is sufficient.
- Finally, minimizing any dead time that may exist in the sensor operation flow is also very important, as it will translate into improved PET system characteristics.

Chapter 4 The SPADnet sensor

As mentioned in the first chapter of this thesis, the SPADnet sensor uses CMOS technology with the main goal of simplifying the PET detector module construction, with the sensors directly providing the scintillation data in digital form. While this can reduce the total system costs and complexity, using CMOS technology also opens the door for new features to be added at the sensor level. In fact, the recent development of deep submicron CMOS nodes, which have been demonstrated to be compatible with SPADs [Nic+10], [Far+08], [Ric+09a], [Kar+10], enables the integration of processing circuits into sensors with minimum area overhead.

Therefore, the main objective during the design of the SPADnet sensor was to not only fulfill all the requirements presented in the previous chapter, but also to improve the state-of-the-art through the inclusion of innovative features in the sensor die. To this aim, the following sections will describe, first, the concepts and compromises pertaining to the design of SPAD-based sensors, then, the sensor architecture will be presented and, finally, a section will be dedicated to the first complete implementation of this architecture, the SPADnet-I chip.

4.1. SPAD-based sensors: design concepts and compromises

The typical operation of SPADs implemented in CMOS technology is briefly described in Figure 18. The SPAD anode is connected to a passive quenching transistor and to an inverter, while the cathode is biased with a voltage above the breakdown. When a photon is detected, an avalanche is triggered and a voltage pulse builds up on the anode, which is then digitized by the inverter, and finally results in a digital pulse with a given duration at the output. This pulse duration is commonly referred as the SPAD dead time (t_{dead}), since any further avalanches generated during this time will not produce a new output pulse.

Furthermore, avalanches can also occur due to thermally generated carriers or band-to-band electron tunneling, which will generate digital pulses indistinguishable from the photon-generated ones. These pulses are the main source of noise in SPAD-based sensors, and are characterized by their occurrence rate, known as the dark count rate (DCR) [Cov+96]. In fact, in a SPAD-based fully digital detector, the DCR is the sole responsible for the temporal noise modeled in Chapter 3, since the high gain yielded by the avalanche and the immediate digitization of the signal mean that the system is free from readout electronic noise.

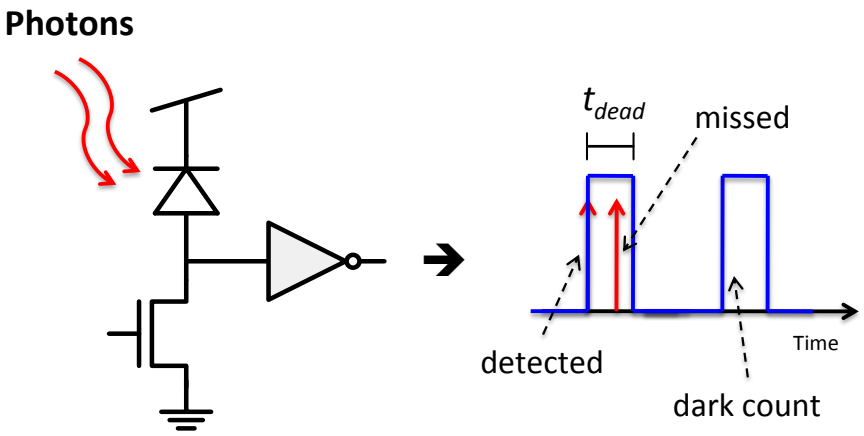


Figure 18: operation principle of SPADs in CMOS technology.

In a large set of equally-sized SPADs (e.g. a SiPM), not all devices will show the same DCR, as the random distribution of defects in a silicon wafer strongly affects the DCR distribution. As such, a small percentage of SPADs – the ones with defects in their active area – exhibit DCRs much higher than the median, usually up to two orders of magnitude higher [Gul+05], [Pan+09]. This phenomenon is quantified by the SPAD yield, which is the percentage of SPADs that have DCRs below a certain desired threshold. In a large SiPM, these high-DCR SPADs can become the detector main source of noise.

The SNR in SPAD-based sensors is strongly impacted by the SPAD characteristics, as the number of detected photons depends on the SPAD PDE, while the detector noise comes from the DCR (see equation (2)). Therefore, maximizing the SNR is strongly correlated to finding the best trade-off in the SPAD size, as enlarging the SPAD increases the FF (the guard-ring thickness is kept constant while the active diameter increases), but it also increases the average DCR and reduces the SPAD yield. Moreover, these two noise characteristics scale more-than-linearly with area [Pan+07], [Ric+11], resulting in practical SPAD diameters being limited to a few dozen micrometers ([Hae+12] is one of the largest CMOS SPAD reported, with 59.4 x 64 um^2 size with 78% FF).

For the SPADnet sensor, the SPAD structure and technology will be derived from previous works in the targeted 0.13 µm 1P4M CMOS process [Ric+09a], [Ric+11]. These works provide an excellent starting point for the design of a SPAD-based sensor, as different SPAD structures and sizes have been fabricated and tested. The results of these tests showed that circular-shaped

SPADs provide the best yield, with up to 90% of SPADs close to the median value of the population. As such, this structure was selected for the SPADnet sensor.

Moreover, the data presented in [Ric+11] can be used to draw a sensor-specific line on the surface graph shown in Figure 7. That is, from the specific relation between DCR and FF provided by the selected structure, a single energy resolution value can be obtained corresponding to a given SPAD size. However, to do that, a few other assumptions still need to be made: first, as the number of detected photons also depends on the optical coupling between the sensor and crystal, two scenarios will be considered: 1000 and 1500 photons with a SPAD of 15 μm active diameter. Secondly, the integration area will be defined so that it contains about 50k SPADs, also of 15 μm active diameter. From this, the number of detected photons and of dark counts can be calculated for each SPAD size, and the energy resolution obtained through equation (2).

Figure 19 plots the results, showing that, for both optical coupling scenarios, the energy resolution shows little variation for SPAD active diameters between about 16 and 34 μm. It should be noted, however, that this model does not take into account second order effects such as optical crosstalk, which occurs when photons are emitted in the avalanche process and subsequently detected by neighbouring SPADs. This effect has been shown to increase with the SPAD size [Pie+13], [Gol+14], and should be taken into consideration when defining the final SPAD diameter.

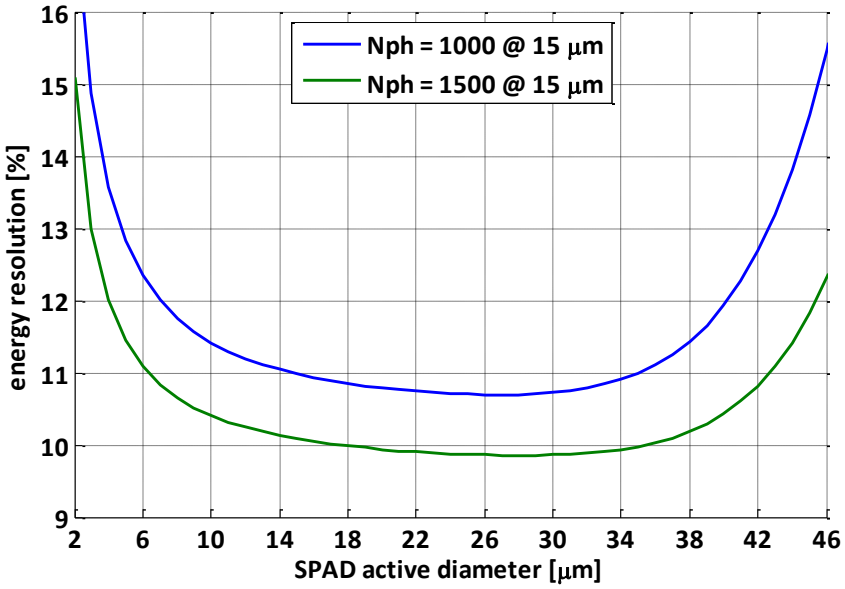


Figure 19: energy resolution versus SPAD active diameter for round SPADs based on [Ric+11]. Detection of 1000 and 1500 photons is assumed in an area containing 50k SPADs with 15 μm active diameter.

4.2. Sensor Architecture

As has been discussed in this thesis, there are two main topics that steer the design of the SPAD-net sensor: SPAD-based design issues and gamma detection requirements. Corresponding to these two points, the sensor design was approached from two perspectives: bottom-up and top-down. In the first, the typical SPAD front-end was taken as a basis for a new detector cell, aiming at maximizing the achievable FF, while still providing digital counting and photon timestamping. The top-down methodology, on the other hand, focused on coping with the gamma detection requirements, such as monitoring the photon flux and minimizing the dead time.

These two approaches will be discussed below, organized according to the resulting sensor hierarchy (shown in Figure 20): the detector cell is the lowest level, consisting of many SPADs connected to a counter. On the opposite end, the top-level is responsible for discriminating gamma events and controlling the sensor exposure. As will be made clear from this discussion, a middle hierarchy level is also required, putting together the results of the bottom level with the expected inputs of the top, which takes form in the pixel. An overview of the architecture will also be discussed through its complete timing diagram.

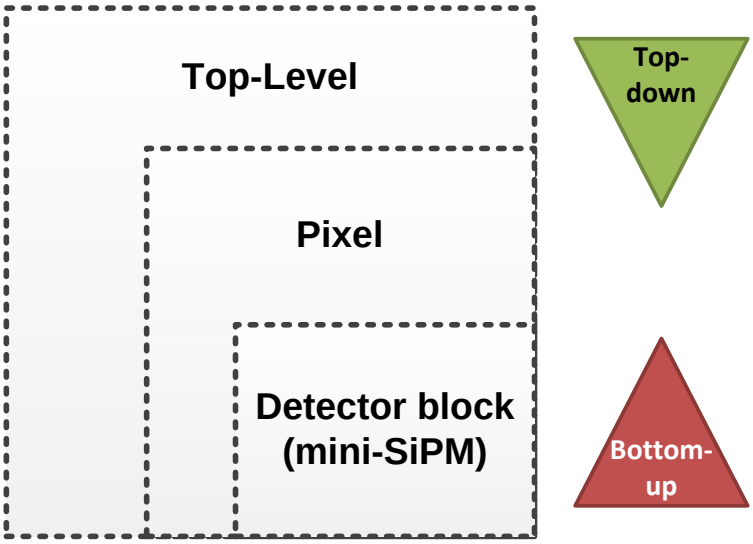


Figure 20: sensor hierarchy overview.

4.2.1 Counting photons: the mini-SiPM

The component that has the highest impact on the detector cell characteristics is, of course, the SPAD. In turn, the SPAD has many of its characteristics defined by its size, which, as shown in the previous section, has a functional range orders of magnitude smaller than the required pixel pitch for PET detectors. This small size of the SPAD has two important consequences for the

sensor design: first, the probability of a SPAD detecting a photon during one scintillation event is quite low – for instance, with 50k SPADs under uniform illumination for a total of 1500 detected photons, the Poisson distribution of counts results in each SPAD having only a 3% probability of detecting one or more photons in an event. Secondly, due to the large number of SPADs, the total area required for the SPADs readout electronics can also strongly impact the sensor FF.

Therefore, to improve the DCR vs. FF compromise while at the same time reducing the required electronics for reading out the SPAD array, two data compression techniques were developed: spatial and temporal compression [Bra+11]. These two techniques are implemented in tandem in a fully digital, small area SiPM (henceforth called mini-SiPM), which form the basic building block of the SPADnet sensor architecture.

At a high-level, the mini-SiPM is simply composed of many SPADs connected to a single digital ripple counter, with the goal of minimizing the electronics per SPAD. However, there is one main challenge with this approach: how to efficiently connect the SPADs outputs to the digital counter. To solve this, the spatial and temporal compression techniques were conceived, which, together, produce the counter clock signal. A simplified schematic of the mini-SiPM architecture is shown in Figure 21.

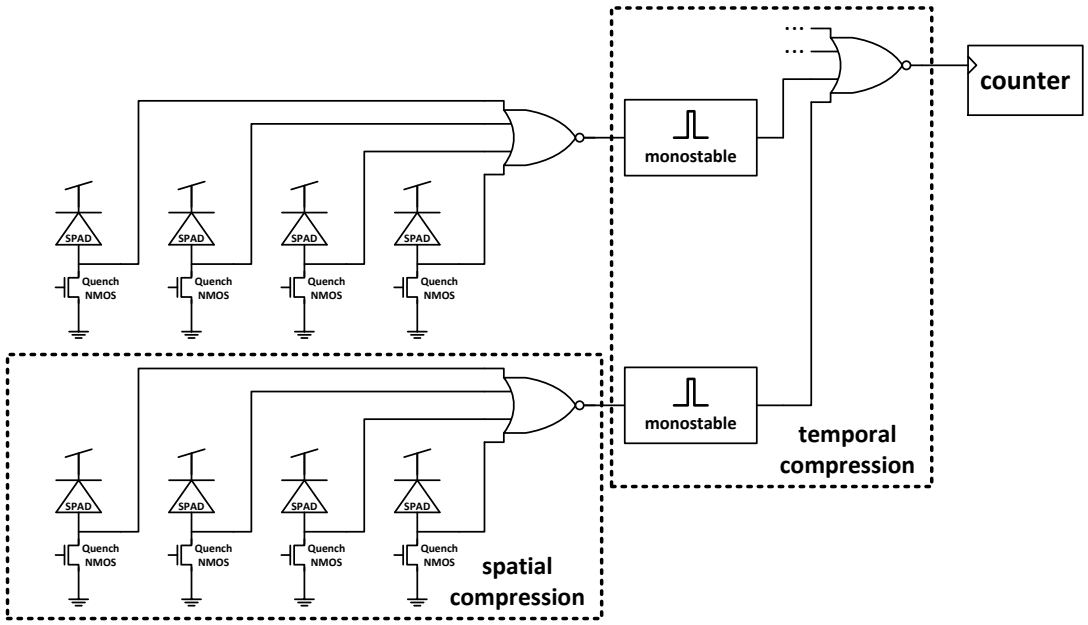


Figure 21: simplified schematic of the mini-SiPM architecture, highlighting the two compression schemes: spatial and temporal.

The spatial compression technique consists of a group of SPADs connected to an OR-function gate, which may also perform the digitization of the signals, as shown in the figure. This means that SPADs connected to the same gate which trigger during each other's dead time are compressed so that only one is counted. As such, for all practical aspects, the spatially compressed group works as a single large SPAD, but with a key advantage: the group's average DCR scales linearly with active area – since the compressed SPADs are of the same size – while the yield is kept constant.

The spatial compression technique can significantly improve the previously mentioned compromise in SPAD size, even if the FF of the spatially compressed SPADs is the same as the one of the single SPAD. On the other hand, the possible disadvantage of compression loss is minimized by the combination of the SPADs small size with the low photon density in PET scintillator detectors.

As a final note on spatial compression, one could also replace the OR gate with a XOR gate in applications where the event duration is sufficiently longer than the SPAD dead time and the photon density is higher. This would allow the counting of up to all SPADs in the group and effectively eliminate any loss due to spatial compression, while only marginally increasing the electronics area occupation.

Whereas the spatial compression explores the relatively faint distribution of photons in space, the temporal compression takes advantage of the distribution of photons in time. By adding a monostable circuit with a short pulse width after the spatial compression, it is possible to effectively mask the SPAD dead time and then compress many more SPADs together.

Depending on the CMOS process node used, the monostable pulse width can easily be made sub-nanosecond, essentially creating a single-wire GHz channel where the SPAD triggers are transmitted. Since the data in this channel is flowing in the form of a train of pulses, the output of the final OR gate can be directly connected to the counter clock input. This provides a substantial area gain compared to other digital summing solutions as, for instance, a full parallel adder. Moreover, each monostable pulse maintains the photon time of arrival information in its rising edge, and thus the timestamping capability of the system is preserved. Finally, due to the binary coding intrinsic to digital counters, another advantage of this architecture is that it can signifi-

cantly reduce the number of bits to be readout when compared to a one-bit per SPAD configuration.

As with spatial compression, the potential disadvantage of count loss with temporal compression is minimized due to the relatively low photon arrival rate with respect to the high speed of the digital blocks in deep-submicron CMOS. It is also interesting to note that the benefits of temporal compression make it attractive even for applications where spatial compression is not desired (i.e. applications with higher photon density). In these cases, the monostable could simply be connected to the SPAD output itself, which would, however, increase the electronics area occupation.

To illustrate the mini-SiPM operation, Figure 22 shows a mini-SiPM with 32 SPADs, divided in 8 spatially compressed groups, and a corresponding timing diagram. For easier understanding, photons arrive sequentially in groups 1, 2, 4, 5 and 8, generating, first, a long pulse at each group OR output, and then a short pulse at their monostables. A second photon at group 4 is also shown, highlighting how compression loss can occur. The final output of the structure is shown at the last line of the timing diagram, demonstrating the train of pulses, which can be directly connected to a counter clock input.

Additionally, the FF of the mini-SiPM can be further optimized when implementing n-well sharing SPADs, as demonstrated in [Pan+07], or in [Wal+12] in the targeted technology. This way, all SPADs can be tightly packed together, while all required electronics can be kept in a separate strip, as shown in the two floorplan examples in Figure 23. This arrangement also allows for maximum packing of different SPAD shapes, such as the honeycomb-like for round SPADs illustrated in the figure.

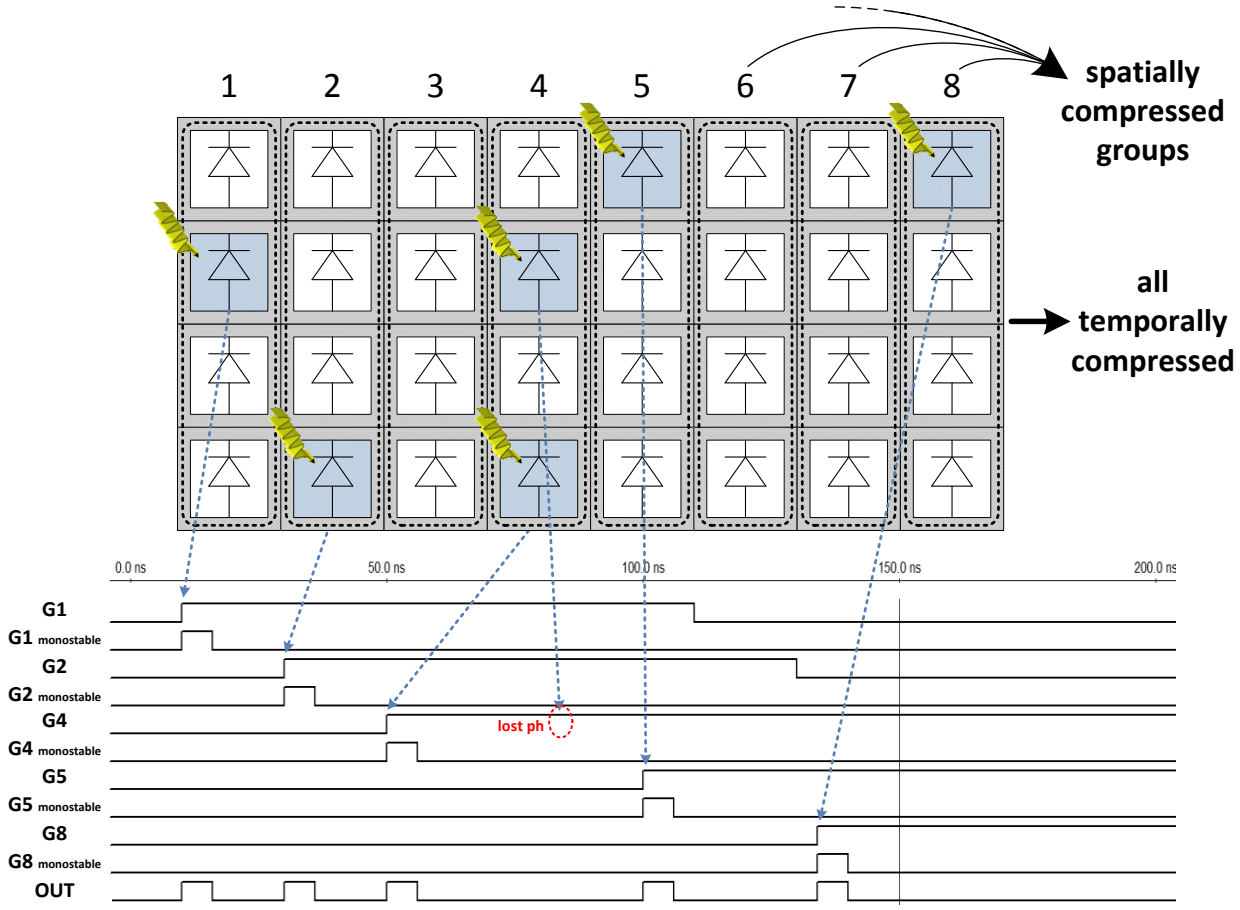
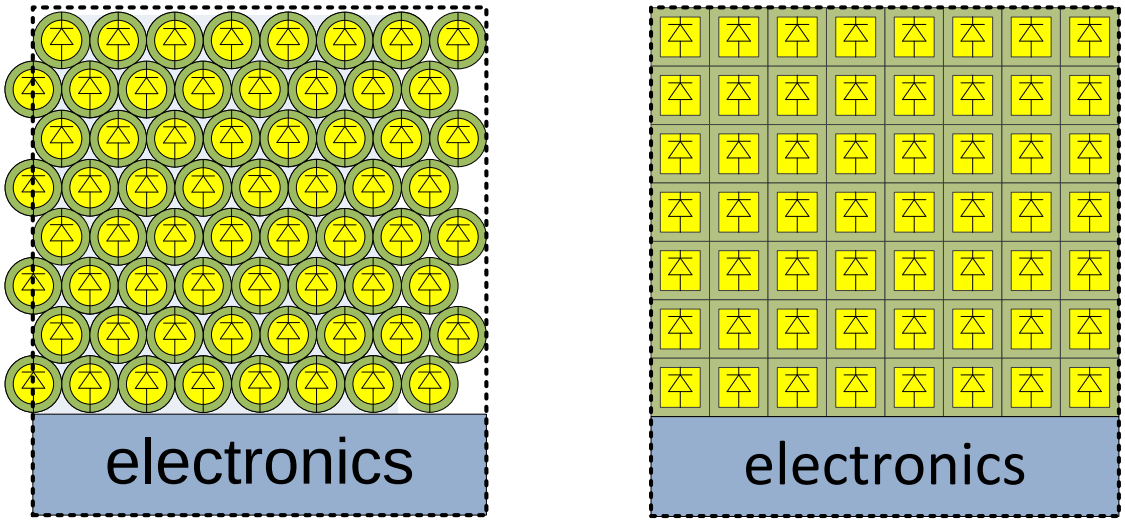


Figure 22: example of a mini-SiPM with 32 SPADs and a timing diagram.



(a) **(b)**

Figure 23: examples of mini-SiPM floorplans with (a) round and (b) square SPADs.

4.2.2 Gamma detection: the top-level

As mentioned in Chapter 3, scintillation events arrive at the sensor randomly in time, and thus must be discriminated, ideally through the monitoring of the incoming flux. On the other hand, the detector cell proposed above, as well as SPAD-based sensors in general, are counting devices which provide an integral of the flux as output. This apparent conflict in the detector design can be solved by dividing the counting operation in short, consecutive time bins, which will provide a discrete approximation of the flux.

To accomplish this, the sensor will feature a synchronous architecture, with a clock signal defining the time bins for the counting operation. A gamma discriminator can then monitor this stream of data to detect the occurrence of a scintillation event. When this happens, an integration period is started, where the counts are actually accumulated to provide the gamma energy estimation, and the timestamps of the initial photons – which were shown to be the best for the gamma ToA estimation in Chapter 3 – are saved.

Since the light photons emitted from the scintillation reach the sensor in a short burst in time, but spread in a relatively large area, the count stream provided to the discriminator must be gathered from many mini-SiPMs up to the top-level, in real time. This can be achieved with a distributed adder in an H-tree-like topology superimposed over the mini-SiPM array, where each node performs the addition of the counts from its leaf cells (the mini-SiPMs). This topology provides equalized propagation delay from all nodes, thus improving the maximum attainable clock speed. Moreover, since all mini-SiPMs must be synchronized for this scheme to work, the clock signal is distributed through the tree in the opposite direction of the count data flow. At the top of the tree, a discriminator can then monitor the total chip count value to determine when an event occurred. The block diagram of the top-level in shown in Figure 24, illustrating an 8x8 mini-SiPM array and a resulting discrete flux estimation.

From the count stream provided by the adder tree, the discrimination logic can use a simple thresholding scheme to distinguish a scintillation event, as discussed in Chapter 3. However, due to the asynchronous nature of the incoming events, a scintillation may start close enough to the end of a bin that not enough photons are detected to surpass the discriminator threshold in this first bin. As such, the discriminator actually uses two thresholds, comparing them to two consecutive photon flux samples to detect an event, as shown in the state diagram in Figure 25. With

this algorithm, the first threshold can be made small, slightly above the noise level, so that even in the case of an event arriving very close to the end of the time bin, the photons on this bin will be taken into account both for the energy estimation and for the timestamps. The second threshold, on the other hand, can be made large, so as to clearly distinguish between the desired gamma events and noise/scattered events.

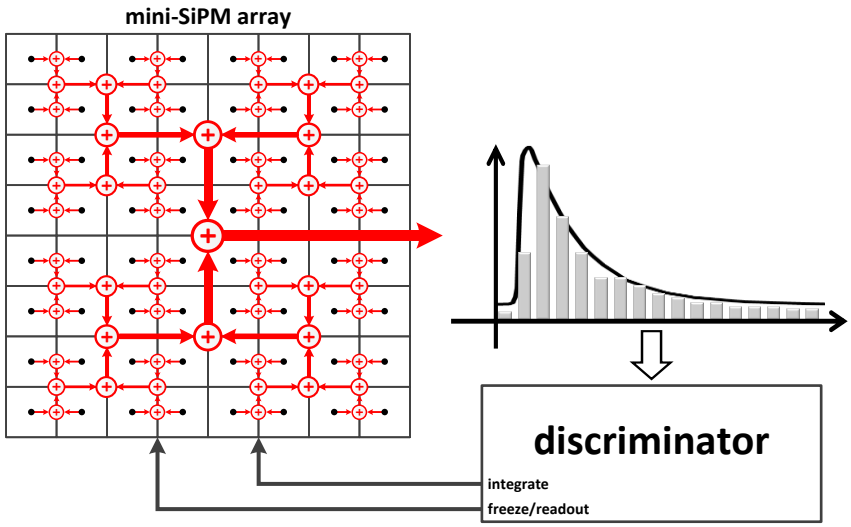


Figure 24: block diagram of the sensor top-level.

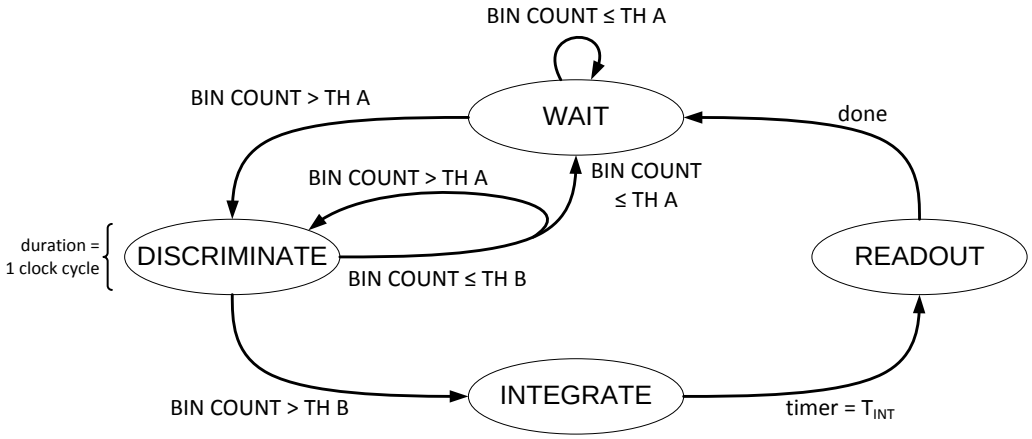


Figure 25: state diagram of the discriminator.

When the discriminator goes into the *integrate* state, a corresponding signal will be fed back into the array, indicating that the flux must now be integrated, starting from the time bin compared to the first threshold, and also that the timestamps from these initial bins must be saved. Then, after a configurable integration time t_{int} is finished, the discriminator goes into the *readout* state and signals the array to stop integration.

Due to the synchronous nature of the architecture, it is straightforward to implement a pipeline between photon counting and discrimination, so that both operations can occur in parallel. This means that while the mini-SiPMs are counting in a given time bin, the discriminator can compare the previous bins to the desired thresholds. As such, the system does not incur any dead time due to the discrimination operation. Depending on the implementation, the array readout operation, which is required to obtain the spatial information, can also be pipelined, at the expense of area occupation for the data buffers.

Finally, another important feature of this architecture is that the ToA estimation is completely decoupled from the discrimination function, as the relevant event timestamps are saved for post-processing. This means that the discriminator can be configured with the sole purpose of increasing the sensor efficiency, i.e. of reducing the number of triggers due to noise or scattered events, without regarding the effects on the timing resolution.

4.2.3 Meeting in the middle: the pixel

Between the mini-SiPM and top-level architectures described above, a few required tasks are still missing in the system, namely: photon timestamping, event integration management and data retention for readout. These tasks could be directly implemented at the mini-SiPM level, however, there is a key limiting factor on the mini-SiPM size: the compression losses. If the number of SPADs spatially or temporally compressed is too large, the assumption of low photon detection probability in the groups may no longer hold true, and a significant number of photon counts may be lost.

As such, the optimal mini-SiPM size may not match the desired pixel pitch for the application, and thus the pixel was introduced as a mid-level in the hierarchy. The pixel is composed by an array of mini-SiPMs, and can be ideally placed at one of the adder tree node levels. It is responsible for accumulating the counts when an event starts, timestamping photons and saving data for readout. Its block diagram is shown in Figure 26, along with the required modifications in the mini-SiPM schematic.

The first thing of notice in the pixel diagram is that the critical resources for energy and timestamping, the mini-SiPM counter and the pixel TDC, are duplicated. This was done since, at the end of each time bin, these blocks need to be frozen, read out into a memory, and finally reset so as to be ready for the next time bin. These operations take a non-negligible amount of

time, and thus, to avoid any dead time due to them, the blocks were duplicated so that while one is working, the other one is going through readout and reset.

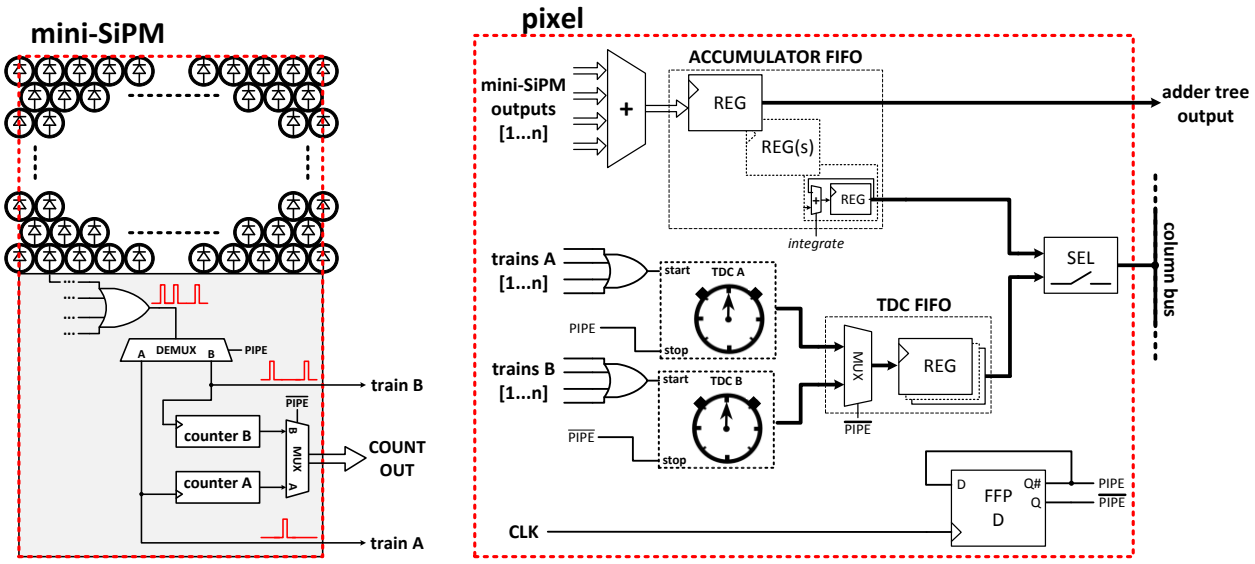


Figure 26: block diagram of the pixel, also showing the required modifications at the mini-SiPM.

The selection between working and stopped block is done through the *PIPE* signal, which is simply a frequency division of the clock (i.e. it stays high for a whole clock period, and then low in the next, and so on). This selection actually occurs in two places: at the inputs of the counters/TDCs, *PIPE* is used to direct the incoming data (i.e. the mini-SiPM pulses) to the working block. On the other hand, at their outputs, an inverted *PIPE* signal is used to multiplex the stopped block output to the next stage of the architecture. The detailed timing diagram of these operations will be described in the next subsection.

The overall pixel architecture contains two main data paths: counts (energy) and timestamps. The energy data comes from the mini-SiPM count outputs, which are summed together (as part of the previously described adder tree) and then fed to a FIFO register for storage. This FIFO has two duties: first, it stores the pixel counts while waiting for the top-level integration control (*integrate*). Then, when *integrate* is set high, the last register starts acting as an accumulator so that, when the pixel is readout, the total energy accumulated during the exposure period is provided in a single register. As such, the FIFO depth (i.e. the number of FIFO register stages) is defined by the delay of the feedback loop with the discriminator. It should be noted that the first FIFO stage is already used for pipelining the adder tree data propagation (or part of it), with the mini-SiPM

counter readout and summing, since the pixel energy is fed to the adder tree only after the first register.

For the timestamping subsystem, the mini-SiPM pulse trains, already divided by the *PIPE* signal, are OR'd together and connected to the two TDCs, with each one being stopped by a different phase of *PIPE*. This way, in each time bin, the enabled TDC is able to timestamp the first pulse that reaches it. This architecture demands for a careful layout of the mini-SiPM towards minimizing the skew added by the compression tree, since, at the pixel level, there is no way of knowing from which SPAD the timestamped pulse came from (so as to correct its skew).

A possible disadvantage of the proposed structure is that timestamping only the first pulse in each time bin may deteriorate the gamma ToA estimation. However, even if more TDCs could be added so as to also timestamp the following pulses, this would result in diminishing returns: first, due to the temporal compression of pulses up to the TDC level, the second and following photons could actually be compressed with the first, and thus would not be timestamped. Secondly, these TDCs would reduce the FF, contributing to an overall worsening of the timing resolution as shown in Chapter 3. Finally, there is a probability that these subsequent photons will impinge as the first photons on different pixels, and will therefore be timestamped anyhow. This means that the trade-off in the number of timestamped photons (and their order) is balanced not only by the number of TDCs in a pixel, but also by the pixel size. For PET applications, the one active TDC per pixel figure was deemed as a good compromise.

Moreover, due to DCR, the TDCs may also be triggered when there are no impinging photons. Therefore, to minimize the probability that the TDC was already triggered by a dark count when a photon arrives, the time window defined by PIPE must be relatively short or, more specifically, the clock frequency should be sufficiently higher than the total DCR of the pixel. This probability can be plotted versus the ratio between DCR and clock frequency according to equation (16), which considers that as gamma events occur randomly in time, their average time of arrival in one clock cycle is at half the cycle. The results are shown in Figure 27.

$$P(x > 0) = 1 - e^{-\frac{DCR}{2 \cdot CLK_{freq}}} \tag{16}$$

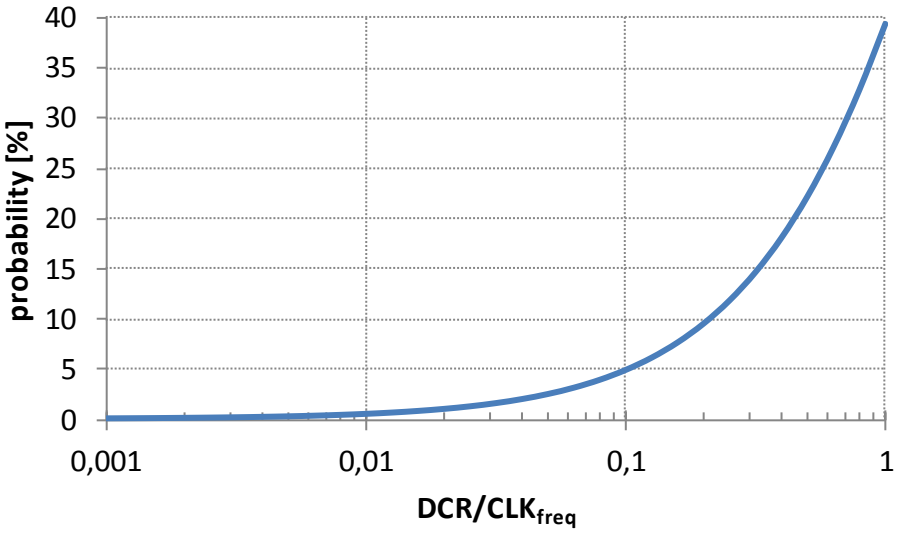


Figure 27: average probability of the TDC being triggered by a dark count before a gamma event arrives.

As can be observed in the figure, with a DCR of 10% or less of the clock frequency, the probability of the TDC being occupied by a dark count when the gamma event occurs becomes negligible, at below 5%. Moreover, even with a DCR equal to the clock frequency, this probability rises to only 40%, indicating that even with a high DCR, if the photons are spread across a sufficient number of pixels, there will anyway be some TDCs available for photon timestamping.

Going back to the pixel block diagram, similarly to the energy FIFO, the TDCs FIFO is responsible for storing the timestamps until readout. However, since only the initial timestamps of the event are required, once the *integrate* signal arrives at the pixel, the TDCs FIFO will be frozen, and only the timestamps of the first few time bins (the same number as the FIFO depth) will be available for readout.

Finally, it is interesting to note that the pixel architecture is independent of the TDC internal architecture; however, to minimize power consumption, it is preferred that the TDC is started only when a pulse arrive from the mini-SiPMs, being stopped by the *PIPE* signal, and not the other way around.

4.2.4 Timing Diagram

To provide an overview of the full sensor architecture, the timing diagram of the sensor is shown in Figure 28. To simplify its presentation, it is assumed that there is a single pipeline register stage inside the adder tree, and that the feedback signals can travel from the discriminator to the pixels in half clock-cycle.

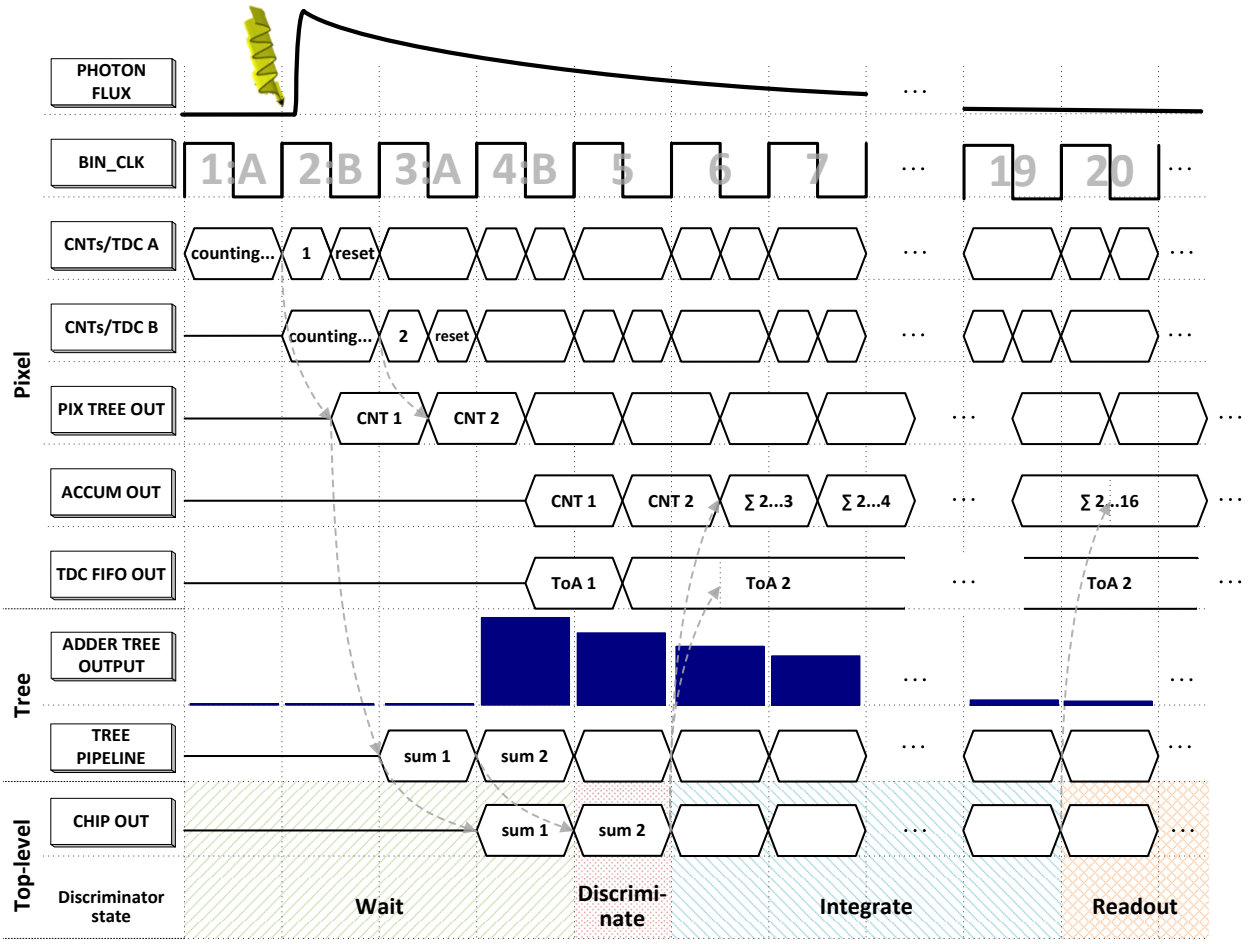


Figure 28: timing diagram of the sensor.

In time bin "1", the pulses generated by all mini-SiPMs of the chip are being counted in their respective *A* counters, while the timestamp of the first pulse in each pixel will be saved by TDC *A*. During the next time bin, the *B* counters and TDCs will start operating, while the *A* blocks will be frozen for half clock cycle. After this half cycle, the FIFOs are clocked, thus saving the sum of the *A* counters and the timestamp of TDC *A*, and the *A* blocks are then reset. From this moment on, the pixel counts will flow through the adder tree, reaching the top-level output after one and a half clock cycles (half cycle to the tree pipeline stage, and then the assumed single stage up to the top-level).

As an example, the diagram shows a gamma event arriving during bin "2". The counts will propagate to the top-level such that while bin "5" is being counted at the pixel level, the discriminator will process the total array count value from bin "2" (which is above the first threshold) and enter the *discriminate* state. Following this, in bin "6" the discriminator will enter the *integrate* state (i.e. bin "3" counts above second threshold), sending the corresponding *integrate* signal to

the pixel, which will both freeze the TDC FIFOs and start integration in the energy FIFO. This way, the timestamps generated at bin "2" are stored at the last TDC FIFO register, while the sum at the accumulator is started also from bin "2". When the integration time (comprised by the *discriminate* and *integrate* states) is finished, the sensor will go into the *readout* state, stopping the pixel accumulators and then waiting to be readout by an external controller.

4.3. First implementation: SPADnet-I

One of the key issues when implementing the architecture described above is sizing all the hierarchy levels. This will be discussed in the next subsection, with the subsequent ones presenting the detailed schematics of the chip, the layout floorplans and, finally, the chip micrograph.

4.3.1 Hierarchy sizing

Analogously to the architecture conception, the process of sizing the different sensor blocks is started by looking at the two extremes of the hierarchy concurrently.

The target for the SPADnet module dimensions is 50 × 50 mm^2, with all module photosensors connected to a single FPGA for readout. As a compromise between board layout complexity, required FPGA input pins and sensor yield, a sensor pitch of 10 mm in both directions was chosen. Other requirements for die sawing and placement further reduce the sensor usable area to 9.85 × 9.85 mm^2. However, this first implementation is designed to be only an evaluation vehicle of the architecture, and is not intended to go into the final SPADnet module. As such, it was designed with one of the dimensions halved, which allows for the connectivity and data propagation at the full diameter to still be tested, while at the same time reducing the cost and risk of the design.

On the other end of the hierarchy, section 4.1 provided an assessment of the optimum SPAD size, which ranged from about 16 to 34 μm active diameter. However, the energy resolution model used to obtain this range does not take into account some effects which are known to worsen with SPAD size, such as optical crosstalk, and thus the SPAD size was chosen to be at the lower end of the range, i.e. circa 16 μm. Further SPAD specifications are a circular shape and a guard-ring thickness of 1.5 μm, which results in a SPAD pitch of 19 μm. This value, however, will still be optimized for maximum FF after all the other hierarchy levels are defined.

Using the initial value of 16 μm for the SPAD active diameter, the mini-SiPM compression losses can be estimated. For the spatial compression, it is assumed that the SPAD dead time is similar to the event integration time, and thus the SPAD can count a maximum of one photon per event, and further that, due to spatial compression, only one photon can be counted per spatially compressed group. From the Poisson distribution, the average count error due to spatial compression, err_{spt}, can then be expressed by equation (17), where D is the detected photon density, A_{SPAD} is the active area of one SPAD and N_{spt} is the number of SPADs spatially compressed.

$$err_{spt} = 1 - \frac{1 - e^{-D \cdot A_{SPAD} \cdot N_{spt}}}{D \cdot A_{SPAD} \cdot N_{spt}} \tag{17}$$

Figure 29 plots the resulting count error versus the number of SPADs spatially compressed. For this graph, an average of 1500 detected photons was assumed, uniformly distributed in two light spot sizes: 3 × 3 and 4 × 4 mm^2, resulting in density values of 166.7 and 93.75, respectively. These spot sizes were chosen as a good approximation based on the typical needle sizes used in pre-clinical and clinical applications, considering also the expansion of the spot due to light guides. Furthermore, the DCR was assumed sufficiently low so as to be considered negligible with respect to the photon density (for instance, with 150 ns integration time and 2000 SPADs/mm^2, each with 1 kHz average DCR, the dark count density in one integrated event would be only 0.3/mm^2).

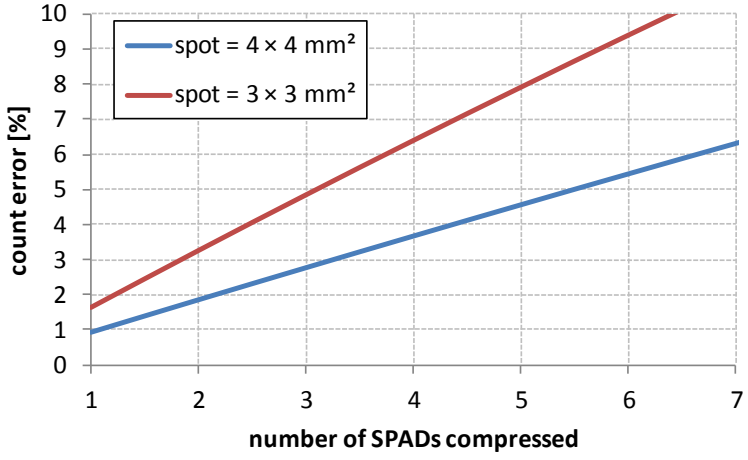


Figure 29: spatial compression loss estimation versus number of SPADs.

A very similar estimation can be made for the temporal compression, simply replacing the density in equation (17) with the flux, and the SPAD area with the monostable pulse width. Since the scintillation light flux is continuously changing during an event, its maximum value will be used as a worst-case estimation, which, from equation (1) with LYSO timing constants, becomes $N_{ph} \cdot 0.023$. The number of photons (N_{ph}) in this case will be the expected number of counts in the temporally compressed group, thus finally resulting in the count error due to temporal compression, err_{tmp}, shown in equation (18).

$$err_{tmp} = 1 - \frac{1 - e^{-D \cdot A_{SPAD} \cdot N_{tmp} \cdot 0.023 \cdot T_{mst}}}{D \cdot A_{SPAD} \cdot N_{tmp} \cdot 0.023 \cdot T_{mst}} \tag{18}$$

To plot the temporal compression count error, the same assumptions as for the spatial compression were made, that is, an average of 1500 detected photons, uniformly distributed in two light spot sizes: 3 × 3 and 4 × 4 mm^2. Moreover, the monostable pulse width, T_{mst}, was estimated at half nanosecond, which is a realistic value given the 0.13 µm CMOS process node. The results are shown in Figure 30.

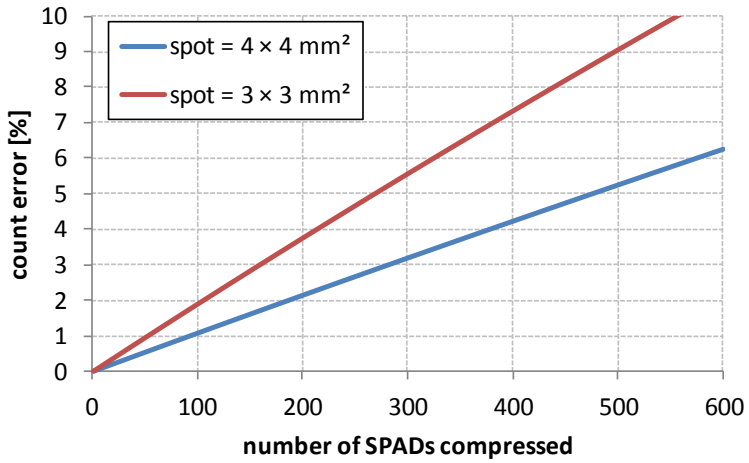


Figure 30: temporal compression loss estimation versus number of SPADs.

Targeting a maximum total loss below 10%, and equally dividing this allowance between the two compression techniques, results in a spatially compressed group with 3 SPADs and around 200 temporally compressed SPADs. Based on this, the mini-SiPM SPAD array dimensions can be estimated, assuming, for instance, a 15 × 13 array with the honeycomb packing shown in Figure 23(a), which results in about 250 × 250 µm^2.

To define the pixel size, the most significant factor is the area occupation of the required electronics. An estimation of this area in the targeted technology resulted in an occupation of only 3% in a 0.5 × 0.5 mm^2 pixel or 1% in a 1 × 1 mm^2 pixel. Counterbalancing the FF, however, are the many advantages that having a small pixel brings: for ToA estimation, for instance, smaller pixels mean more timestamps available in each event and a smaller DCR at the TDC input; for the energy and spatial estimation, it means that a more fine-grained information is available, allowing for a better optimization of the integration area, for instance. Therefore, the pixel pitch was chosen to be at the vicinity of 0.5 mm. From this number, and taking into account the mini-SiPM sizing described above, the pixel was defined as an array of 2×2 mini-SiPMs.

Defining all the exact sizes of the full hierarchy now becomes an exercise in iterative optimization of the sensor FF. For this, the final factors that need to be taken into account are the spacing required at the pixel array periphery for power distribution and the fact that the adder tree propagation delay is minimized with a power of 2 number of pixels. Table 2 shows the results of this optimization process, with all sizes and specifications of the SPADnet-I sensor.

Table 2: specifications and dimensions of all hierarchy levels of SPADnet-I.

SPAD	Diameter (active/total)	16.27/19.27 μm
Mini-SiPM	Spatial compression	3 SPADs
	Structure (temporal comp.)	12 × 15 SPADs
	Dimensions (with electronics)	270 × 290 μm^2
Pixel	Structure	2 × 2 mini-SiPMs
	Dimensions	570 × 610 μm^2
	Fill-factor	42.6%
SPADnet-I	Structure	8 × 16 pixels
	Dimensions	9.85 × 5.45 mm^2
possible SPADnet module version	Composition	16 × 16 pixels
	Dimensions	9.85 × 9.85 mm^2
	Fill-factor (with 10 mm pitch)	38.3%

One final architecture element that needs to be defined is the number of pipeline stages between the pixel and the discriminator. This depends heavily on the desired clock speed, which was targeted at 100 MHz. From propagation delay simulations, it was defined that a single register stage at 5th level of the tree was sufficient.

4.3.2 Schematics

The schematics of the mini-SiPM are shown in Figure 31, illustrating the SPAD front-end, the compression electronics and the counting block, with inputs highlighted in red. All SPAD cathodes are connected to a common bias line (V_{SPAD}) and their anodes individually fed into passive quenching transistors. The front-end circuit is further composed by a Schmitt trigger inverter, which digitizes the SPAD pulse and prevents the slow recharge of the SPAD from affecting the compression circuit, and by a 6T SRAM, which allows disabling high-DCR SPADs. There is also a level shifter which is used to pull the SPAD anode into a high voltage to guarantee it is biased below breakdown voltage when disabled.

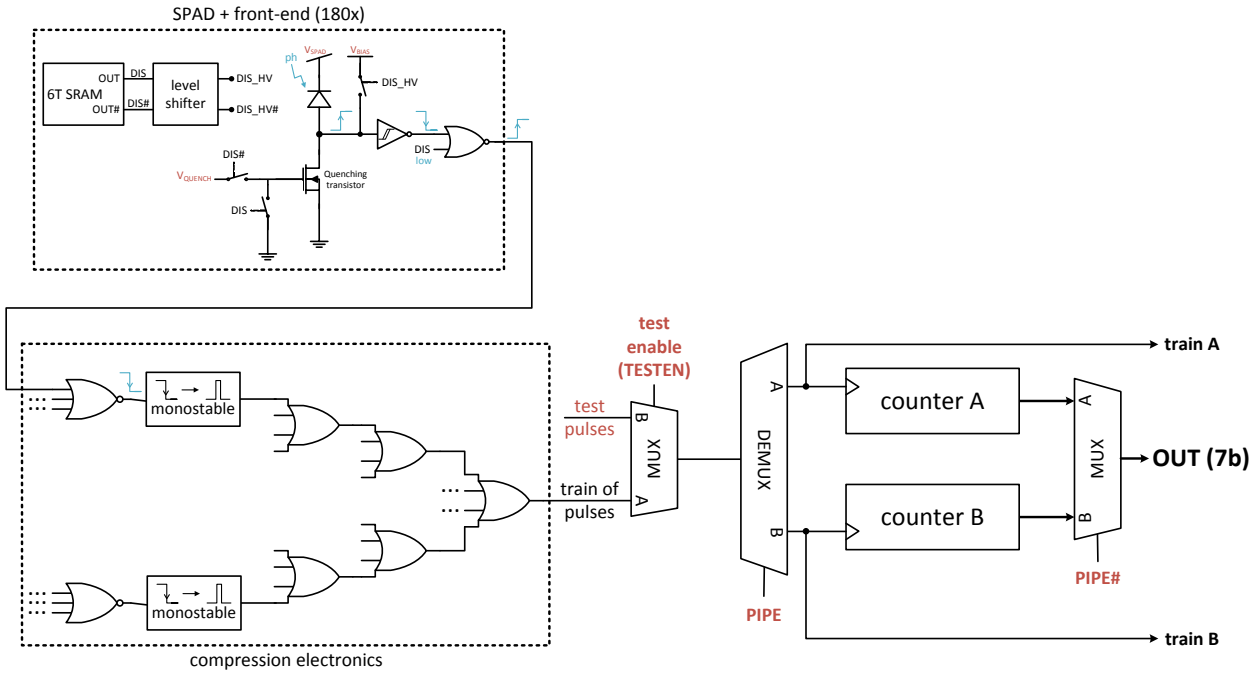


Figure 31: mini-SiPM schematics.

Next, as previously defined, 3 SPADs are spatially compressed, and then the full 12 × 15 array (180 SPADs) is temporally compressed in a final train of pulses. The implemented monostables are current-starved, and their current can be controlled through an external control voltage. All digital ports of the compression tree were optimized for equalized delays between their inputs.

Finally, the two ripple counters of the mini-SiPM are implemented with 7 bits as a safety margin in case the targeted 100 MHz clock frequency cannot be achieved. With this depth, it is expected that even with a clock period matching the integration time, the counters will not saturate. Also implemented in the mini-SiPM is a test mode, selected by *TESTEN*, which switches the counters inputs to an external test pulse input.

The pixel schematics are shown in Figure 32. The *pixel logic* generates the pipeline select signal *PIPE* and the TDCs/counters reset signals according to the timing diagram in Figure 28. Shown in the *TDC trigger logic* is the option to only start the TDC on the second incoming photon in each time bin. This option was added as a failsafe in case the pixel DCR is excessively high with respect to the achievable clock frequency (i.e. if the probability of the TDC being occupied by a dark count, as shown in Figure 27, is too high). Figure 32 also shows the TDC FIFO, with 3×12-bit registers which are frozen when the discriminator enters the *integrate* state with the signal *ACCUMULATE*, and the accumulator/energy FIFO, which starts summing the pixel counts at the *integrate* state, and is frozen at the *readout* state.

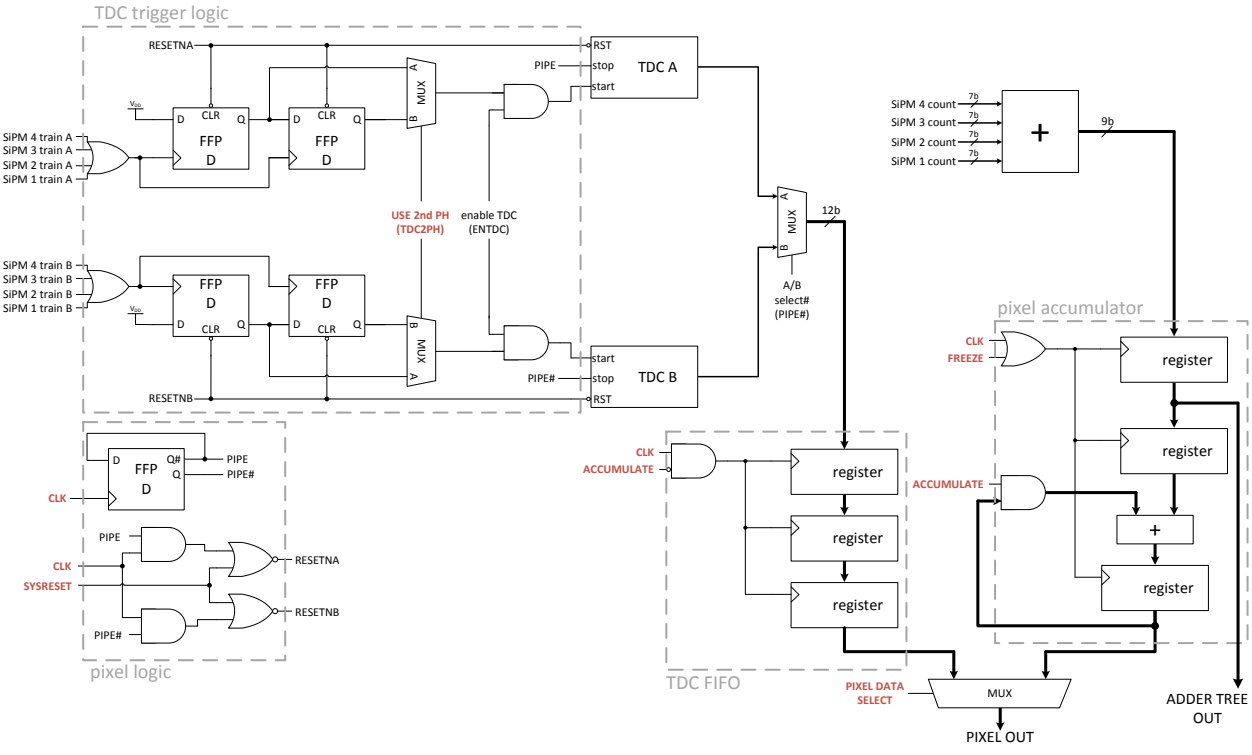


Figure 32: pixel schematics.

The TDCs were implemented based on a ring oscillator architecture [Ric+09b], which was selected due to its relative low power consumption and low area occupation. Moreover, the resolu-

tion of 50 ps achieved in [Ric+09b] in the same technology is already much lower than the scintillation timing resolution limit described in Chapter 3, and thus should not significantly affect the total system CRT.

The block diagram of the TDC is shown in Figure 33: the ring oscillator has four pseudo-differential stages, and it triggers a 9-bit ripple counter at each period, providing the coarse output, while the 3-bit fine output is obtained by encoding its four internal nodes. As previously suggested, to minimize the TDCs power consumption, the ring oscillator is started only when a pulse arrives from the mini-SiPMs, while being stopped by the pipeline select signal. Similarly to the mini-SiPM counters depth, the 12 bits of the TDCs were chosen so that their range could cope with a full integration time window. Lastly, the TDCs are connected to a separate, regulated supply line on the chip, so that, first, the TDC speed and power consumption can be tuned, and, second, their peak power consumption do not affect the other sensor circuits.

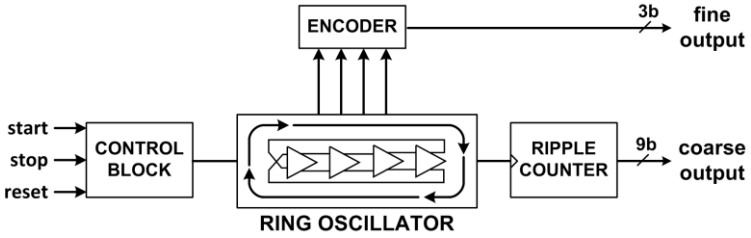


Figure 33: TDC block diagram.

Figure 34 shows the schematics of the adder tree, together with the clock tree propagating in the opposite direction. The pixels are the tree leaves, and, as previously mentioned, already include two internal nodes for summing the mini-SiPM counters. The clock tree, on the other hand, actually contains four signals that need to arrive at all pixels synchronously and with the same delay as the clock: the clock, the FIFO control signals and the test pulses input for the mini-SiPMs.

Finally, Figure 35 shows the block diagram of the chip top-level. Each pixel energy and timestamp information is readout by accessing the sensor in a standard row-wise mode, resulting in a full chip readout time of 2.84 µs at a 100 MHz clock, during which the sensor cannot detect further events. Moreover, during all discriminator states before *readout*, the sensor is constantly outputting the adder tree result, i.e. the discrete photon flux, which can be monitored externally in real-time, at the same sampling frequency as the clock (i.e. 100 MSamples/s). As a final back-up feature, the internal discriminator can be completely bypassed in favour of an external one, since both the *integrate* and *readout* state signals can be provided externally.

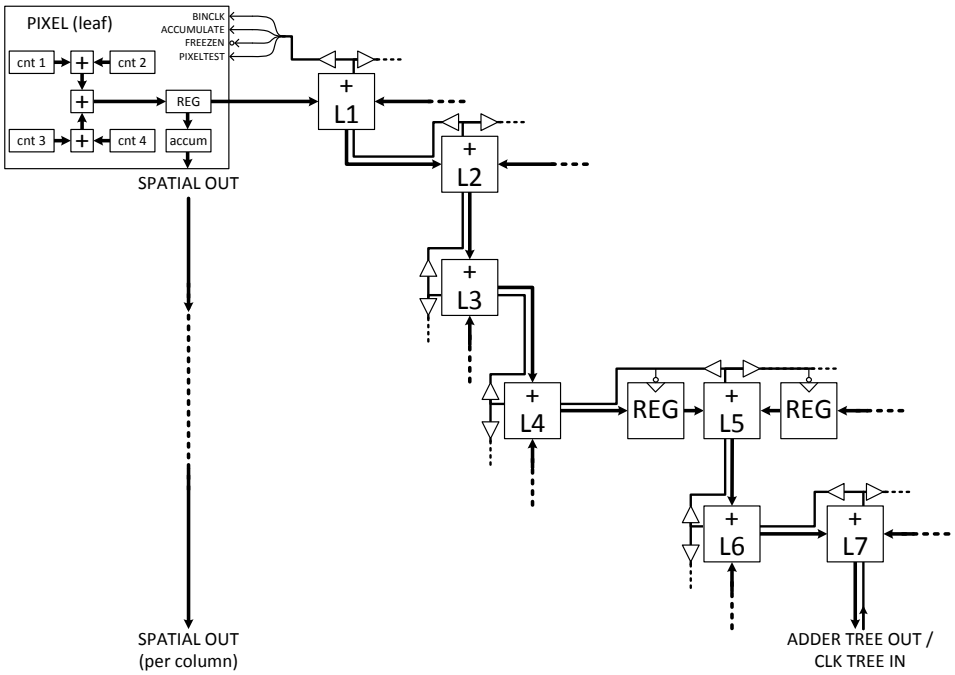


Figure 34: adder tree schematics.

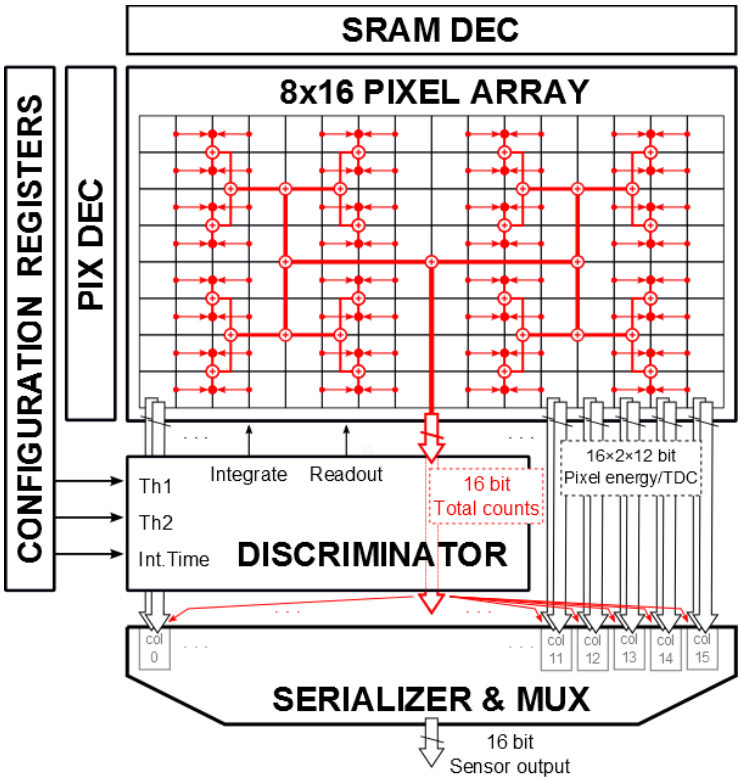


Figure 35: top-level block diagram.

4.3.3 Layout floorplans

Since having a high PDE, and thus a high FF, is one of the main goals of the SPADnet architecture, defining a careful floorplan for the sensor layout is crucial. Moreover, due to the synchronous architecture and the high timing resolution requirements, the high FF has to be balanced with equalized signal distribution. Therefore, the critical floorplans of the sensor are detailed below: the mini-SiPM, the pixel and the adder tree.

For the mini-SiPM floorplan, besides the shared-n-well honeycomb packing structure for the SPAD array, shown in Figure 23(a), the electronics also plays a major role in the attainable FF and timing resolution. Figure 36 shows the floorplan for the SPAD front-ends and compression electronics. Each 12-SPAD column is divided in 4 spatial compression groups with 3 SPADs each. The columns are then horizontally flipped so that the high voltage transistors in the SPAD front-ends are close to one another. The paths in the OR tree from each monostable up to the final compression OR are shown as the coloured lines. This arrangement was chosen so as to minimize the path differences and thus the added skew. Finally, all empty spaces in the layout can be filled with decoupling capacitors for increased immunity to power supply noise.

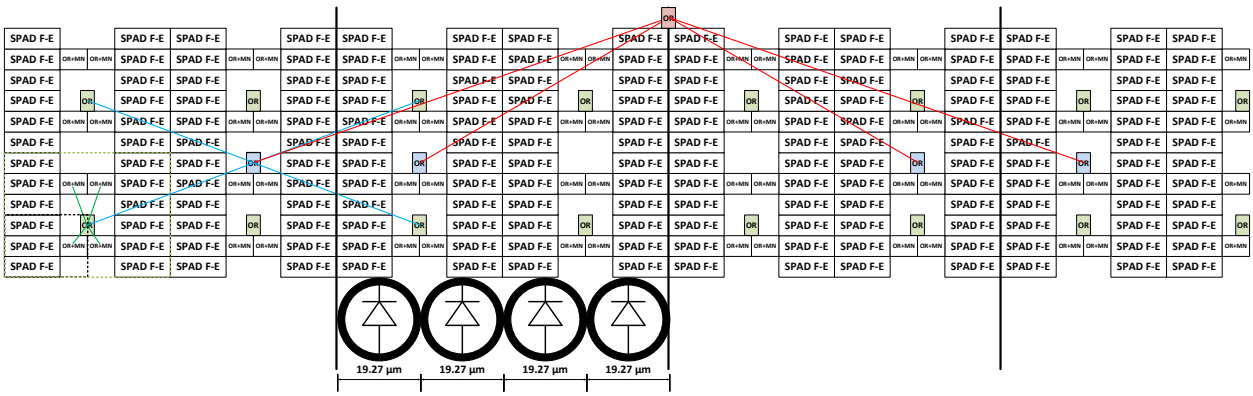


Figure 36: mini-SiPM electronics floorplan.

The pixel floorplan, although relatively simple, is also crucial for the sensor FF and timing resolution. It is shown in Figure 37, illustrating the vertical mirroring of the mini-SiPMs, with the pixel electronics in the middle. This allows for the SPAD arrays in adjacent pixels to be placed together, with no extra spacing with respect to intra-pixel SPADs. Moreover, the paths from the mini-SiPMs pulse trains and counters to the TDCs and adders are fully balanced.

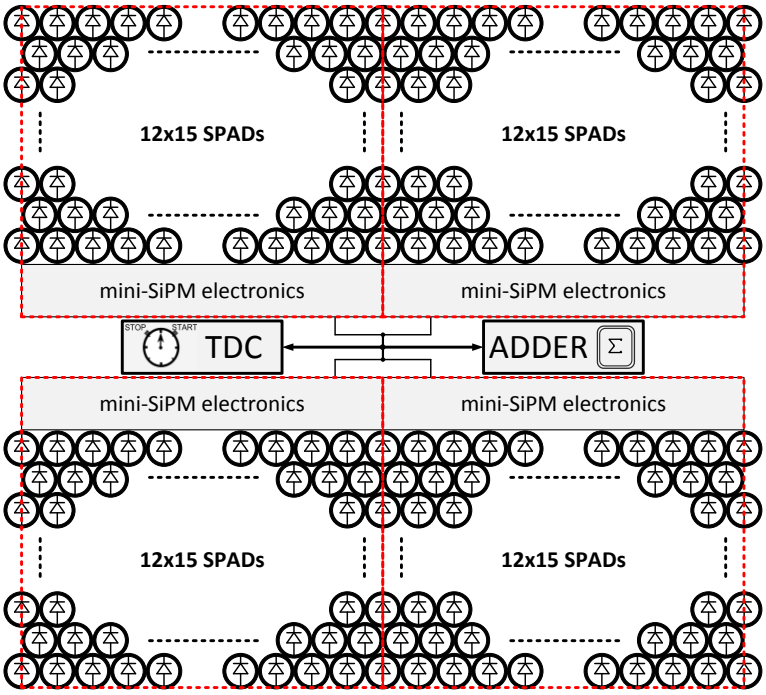


Figure 37: pixel floorplan.

Even though the floorplan of the adder tree no longer influences the timing resolution of the system, it is crucial for the FF and the clock speed of the system. Therefore, the goal when floorplanning the adder tree levels was, first, to keep a balanced propagation delay so as to maximize the achievable clock speed, and second to minimize the FF loss. This was achieved with the floorplans presented in Figure 38 and Figure 39. Since the pixel electronics do not take the whole width of the pixel, the area in between pixels can be used for the adders, as shown in the figures. As can be observed, levels 1 and 2 follow a typical horizontal pattern. However, L3 sums every other L2 so that it can be placed on the same strip of electronics, keeping the balanced signal distribution with no loss of FF.

From level 4 and above, shown in Figure 39, this pattern is maintained, resembling two overlapped H-trees with a one pixel offset. Even if the path to L6 is unbalanced in SPADnet-I, this will not be required in the full array planned for the SPADnet module. Finally, L7 is placed at the top of the array, with its output going all the way down to emulate as close as possible a full 16 × 16 array delay.

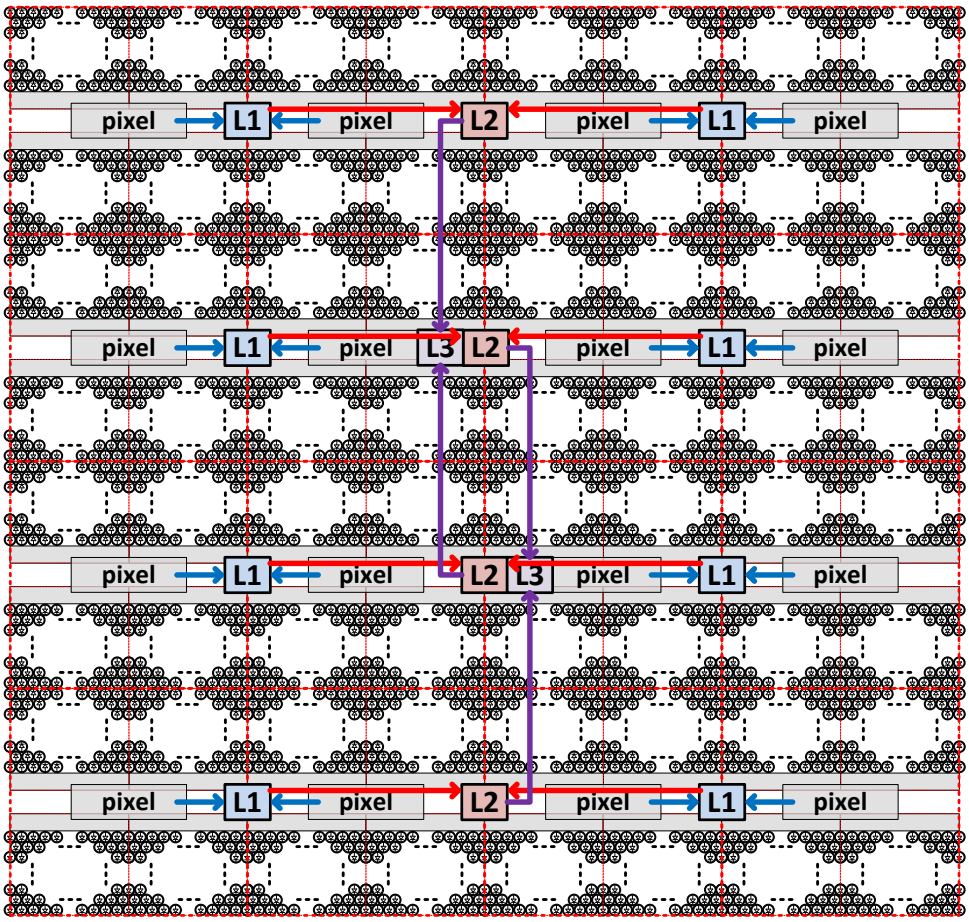


Figure 38: floorplan of the first three levels of the adder tree.

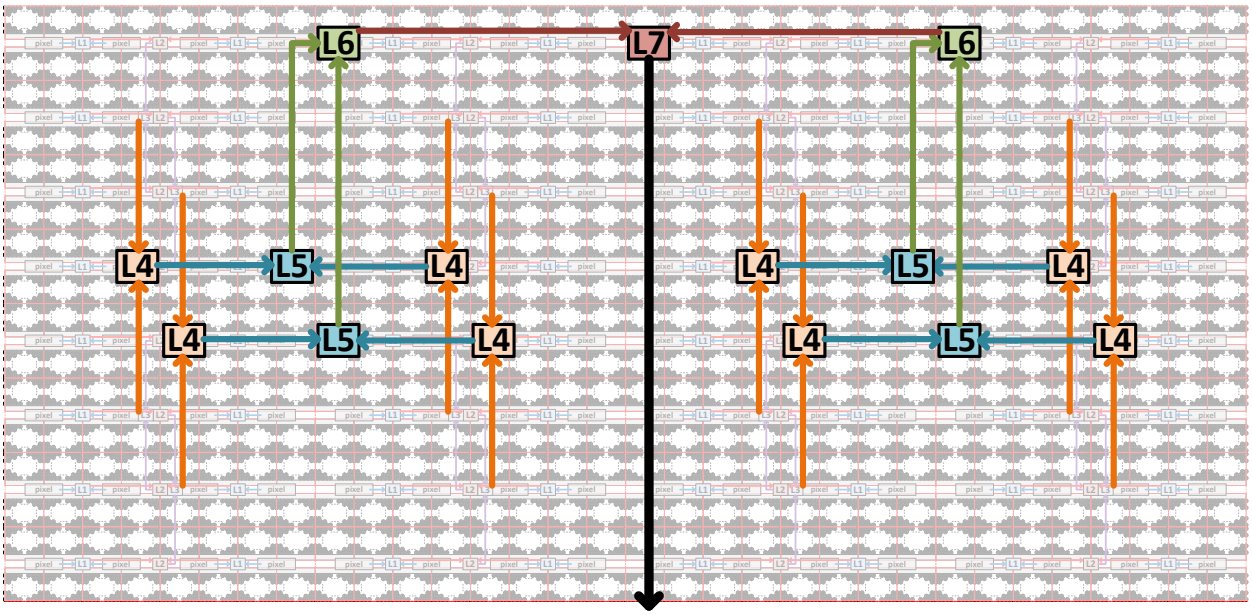


Figure 39: floorplan of the last four levels of the adder tree.

4.3.4 Chip micrograph

As previously mentioned, the sensor is implemented in a 0.13µm 1P4M CMOS imaging technology, achieving a pixel FF of 42.6%. The full die is 9.85×5.45 mm^2 and is shown in the micrograph in Figure 40. The top pads are for testing purposes only, and the full sensor operation can be achieved using only the bottom pads. This way, the full 16 × 16 pixel array required for the SPADnet module can be obtained by simply doubling SPADnet-I with its lower pads, which would result in a FF of 38.3% with a 10 mm sensor pitch. This pitch and resulting small loss in FF is achievable thanks to the process featuring Through-Silicon Vias (TSV), which enable all pads to be connected via a ball grid array on the back of the die, as shown in Figure 41.

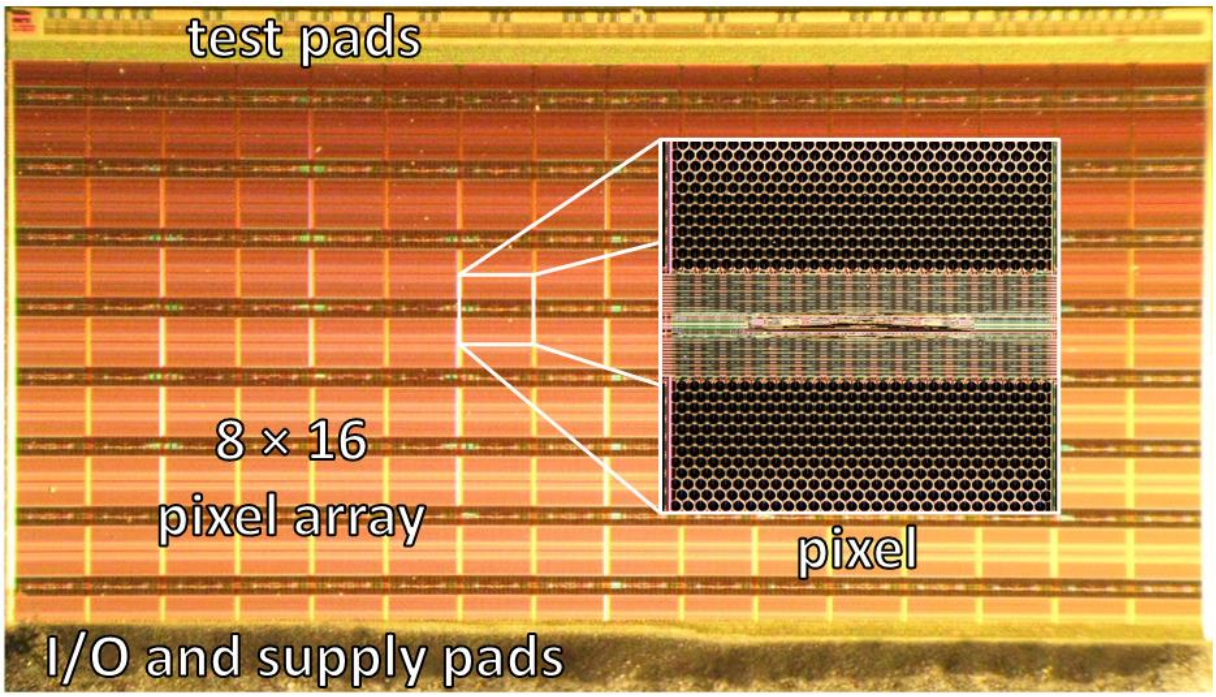


Figure 40: SPADnet-I micrograph.

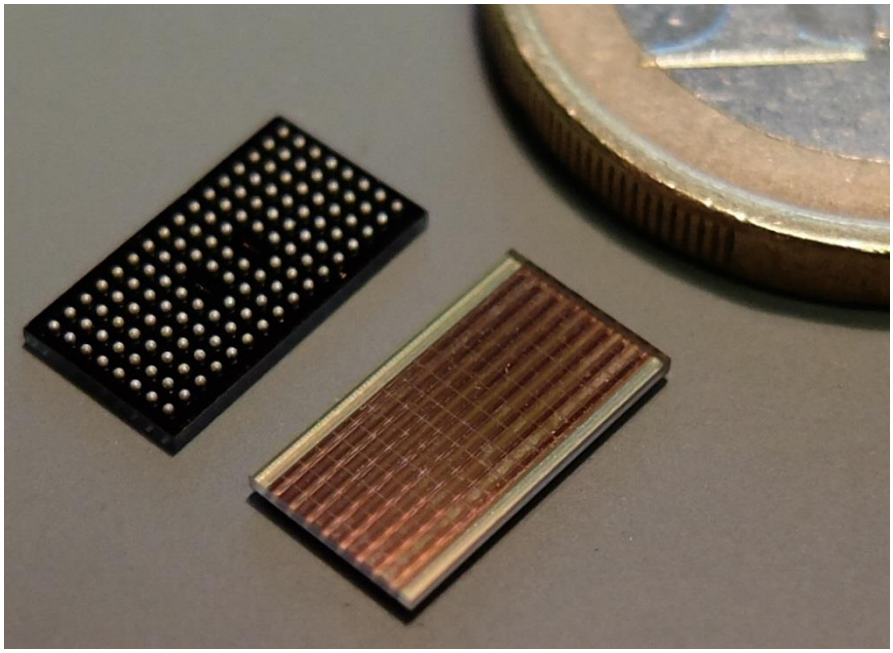


Figure 41: SPADnet-I chip with backside ball grid array.

Chapter 5 Processing SPADnet Data

The SPADnet sensor, or in fact any typical digital PET sensor, will provide two types of raw data for each scintillation events: SPAD trigger counts and TDC codes. Both these values need to be processed to provide the required information about the originating gamma ray, that is, its absorption position, its time of arrival and its energy.

Starting with the later, the gamma energy is directly related to the trigger counts, however dark counts and SPAD saturation need to be taken into account for a correct estimation. Regarding the gamma ToA, the processing of the raw TDC codes must use the response function of the TDCs to generate timestamps, which can then be further refined through previous knowledge of the scintillator and of the sensor characteristics to obtain the final ToA estimation. Finally, the scintillation spatial information can be calculated by processing the distribution of the energy across many pixels. The algorithms and methods for spatial processing, however, strongly depend on the scintillator geometry, and thus are out of the scope of this thesis.

Therefore, the following sections will describe the energy and ToA processing flow and, more specifically, how certain characteristics of the SPADnet sensor architecture can be taken advantage of for improving these estimations.

5.1. Energy

The process of energy estimation is basically a conversion from SPAD trigger counts to gamma energy using the sensor response curve. In an analog SiPM, this conversion is typically modeled on the basis that a SPAD can only detect one photon per event [Ren06], due to the SPAD dead time usually being equal to or longer than the event integration time. This way, the relation between the number of counted SPAD triggers, N_{fired}, the number of absorbed photons N_{ph} and dark counts DC can be expressed by equation (19).

$$N_{fired} = N_{cells} \cdot \left(1 - e^{-\frac{N_{ph}+DC}{N_{cells}}}\right) \tag{19}$$

$$N_{ph} = \gamma_{en} \cdot k \tag{20}$$

As should be expected, the sensor response in equation (19) also depends on the number of SPADs in the area under illumination, N_{cells}. One of the assumptions inherent to this response model is that the light is uniformly distributed throughout these SPADs. Moreover, correlated noise sources, such as optical crosstalk, are considered negligible and thus are not included in the equation. Finally, the relation between N_{ph} and the gamma energy γ_{en} is considered linear, as expressed by equation (20), with k depending on the optical coupling and on the sensor PDE.

In the specific case of the SPADnet sensor, the count to energy conversion is further complicated by the mini-SiPM compression schemes. On the one hand, the principle of spatial compression corresponds to having a larger SPAD with the same dead time, and thus the conversion model from the analog SiPM still applies. On the other hand, the temporal compression depends on the photon flux impinging at the sensor. This flux is continuously changing during a scintillation event, and thus its inclusion on the conversion model is not straightforward.

In a digital SiPM without temporal compression, equation (19) could theoretically be solved for one variable, N_{ph}, as both N_{fired} and N_{cells} would be precisely known. In practice, however, this cannot be done since the light distribution on the sensor is not necessarily uniform. As such, the typical procedure for obtaining the sensor response versus gamma energy is to fit equation (19) to a number of known energy points, which can be obtained with gamma sources with different energies.

For the SPADnet sensor, the same procedure will be adopted, establishing that any loss due to temporal compression will be modeled as a reduced value of N_{cells}. Moreover, the optical cross-talk is considered negligible based on measurements reported in [Wal+12], where a mini-SiPM-like structure using the same SPADs of SPADnet-I showed less than 2% total cross-talk.

Therefore, to obtain the SPADnet response curve, the sensor counts are measured for three known energy values: zero (i.e. in the dark) and the two peaks of a ^{22}Na gamma source, 511 keV and 1275 keV. Then, equations (19) and (20) are fitted to this data with three fitting parameters: N_{cells}, k and DC. The fitting method uses the free and open source Matlab toolbox EzyFit [Ezy10], which performs a nonlinear minimization of the sum of squared residuals to find the best fit. This fitting is performed for each measured sensor configuration (e.g. different SPAD bias, different percentage of SPADs disabled, etc) so as to correctly obtain the energy resolution figure.

As an example of this procedure, Figure 42 shows the resulting fit with the three measured data points, as well as an indication of the theoretical linear response. The data points are obtained from the raw count spectrum shown in Figure 43, which, for comparison purposes, also shows the energy spectrum after conversion with the fitted curve.

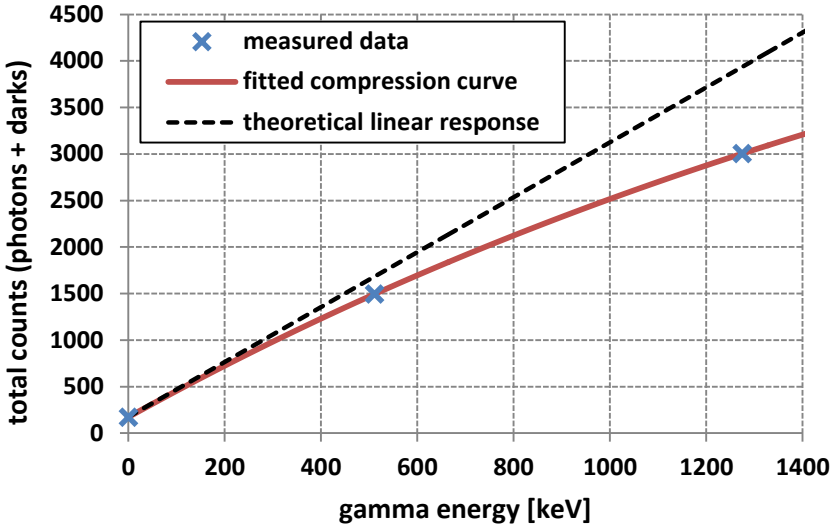


Figure 42: example of SPADnet-I fitted response curve (compression curve).

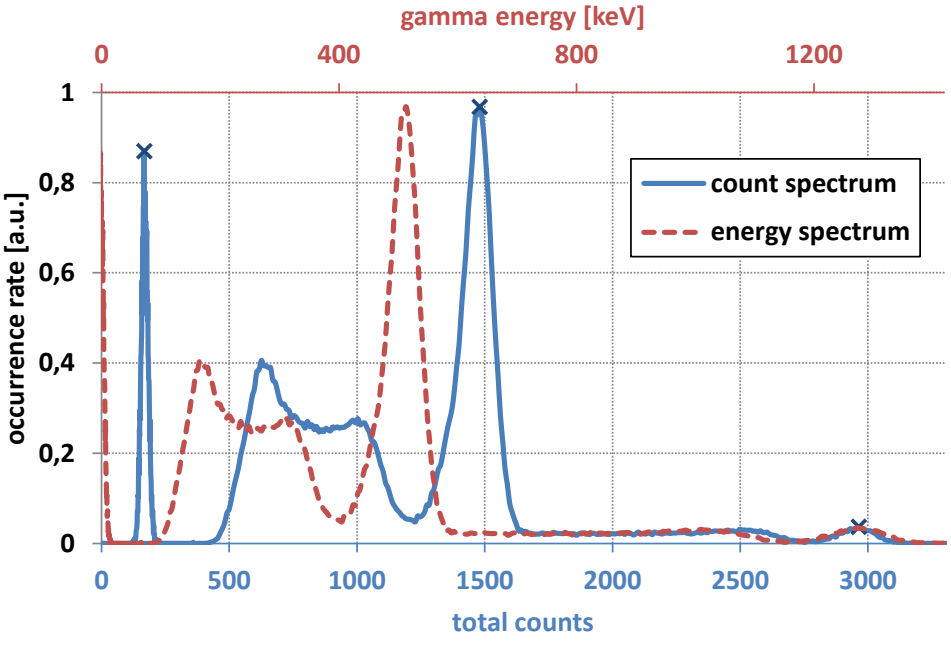


Figure 43: raw count spectrum used to obtain the data points in Figure 42 and the resulting gamma energy spectrum.

5.2. Time of Arrival

Given the small pixel size in the SPADnet sensor, it is safe to assume that light sharing will be applied in almost all crystal configurations, meaning that the light coming from a crystal needle will be spread across more than one pixel. Therefore, multiple TDC codes will be obtained for each scintillation event. These codes can serve two main purposes: one, they enable the system to better identify the gamma arrival in the presence of high DCR, that is, even if a few TDCs may have been already triggered when a gamma arrives, there will still be other TDCs available to timestamps the scintillation photons. Two, the photon-triggered TDC codes can be combined to improve the ToA estimation through multiple-timestamp algorithms [Bra+12], [Dam+13] [Ven+13].

As such, there are three main steps for converting the TDC codes into a final ToA estimation: first, the raw codes are converted into time values (timestamps), based on the TDC response function. Secondly, once an ordered list of timestamps is obtained, any initial dark-triggered timestamp must be identified and removed from that list. Finally, the remaining timestamps are used to obtain the gamma ToA. Even if the last two steps could be combined in a single maximum-likelihood estimator (MLE), these algorithms are computationally intensive, and as such are usually unsuitable for a real-time FPGA implementation.

In the next subsections, the algorithms that perform these three steps are explained in detail. These algorithms have been chosen aiming at maximizing the timing resolution performance of the system while having minimum computational power requirements.

5.2.1 From codes to timestamps

An ideal TDC would have as its response function a perfect staircase, where each bit increment results in the same time-domain step. In this case, it would suffice to know the step value to transform the TDC code into a timestamp. This step could be obtained, for instance, by measuring the maximum TDC code for a given known time window.

In reality, however, there are small variations at every bit increment, which result in a non-ideal response. These non-idealities are usually characterized by the integral nonlinearity (INL) and the differential nonlinearity (DNL) [Doe+84], which can both be measured through a code density test [Pel+97]. In a SPAD-based sensor, this test can be performed by acquiring many samples

of the TDC with a low uniform flux, which can be obtained through simple DCR or with constant low-light illumination [Ric+09b], [Sto+09b], [Ger+09].

The data obtained in this test can be used to build the TDC response function, which is equivalent to the INL curve. The TDC response function will then be of the form of a truth table, where each TDC code has its corresponding time value. However, with a large number of TDCs, storing all the required tables in the FPGA can quickly become impractical and inefficient. As such, a typically implemented alternative is to fit a first degree polynomial (i.e. $y = ax + b$) to the TDC response, and then use the resulting equation to convert the code to timestamp.

Another possible simplification of the conversion method is to use a linear fit ($y = ax$) based only on the full-scale value of the TDC, that is, the time value of the maximum TDC code. This method features an interesting advantage, as the full-scale value can be continuously monitored during the system operation, thus allowing for real-time calibration of the TDCs with no measurement interruption. This becomes particularly important due to temperature and voltage variations that can occur during the system operation. A simple example of a fictitious 3-bit TDC is shown in Figure 44, illustrating an ideal response, the actual TDC response, and the two above-described fits.

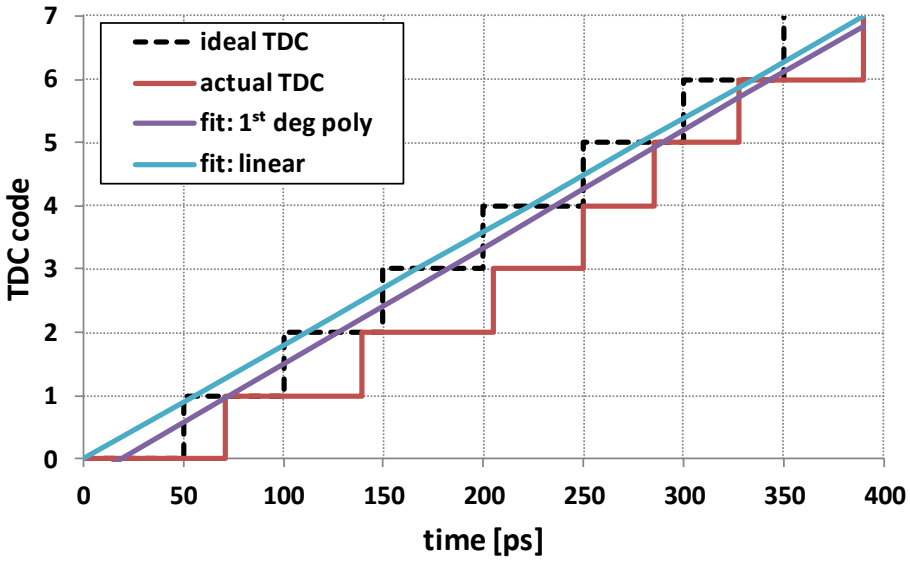


Figure 44: example of a 3-bit TDC response and its possible fits.

For the SPADnet sensor measurements described in the next chapter, due to the demonstrated good linearity of the TDCs (see section 6.1.2), the linear fit method will be used to convert the TDC codes to timestamps. Moreover, due to the two TDCs in each pixel working interchangea-

bly, care must be taken so that the correct TDC response function is applied. This can be controlled by the FPGA, which simply flips a bit at each clock period, indicating which TDC is active.

5.2.2 Finding the first photon

Once the list of timestamps for an event is obtained and then put in ascending order, a cumulative distribution of the timestamps such as the one shown in Figure 45 can be obtained. The plotted data in this example was taken from actual gamma measurements with the SPADnet-I sensor and LYSO that will be described in the next chapter.

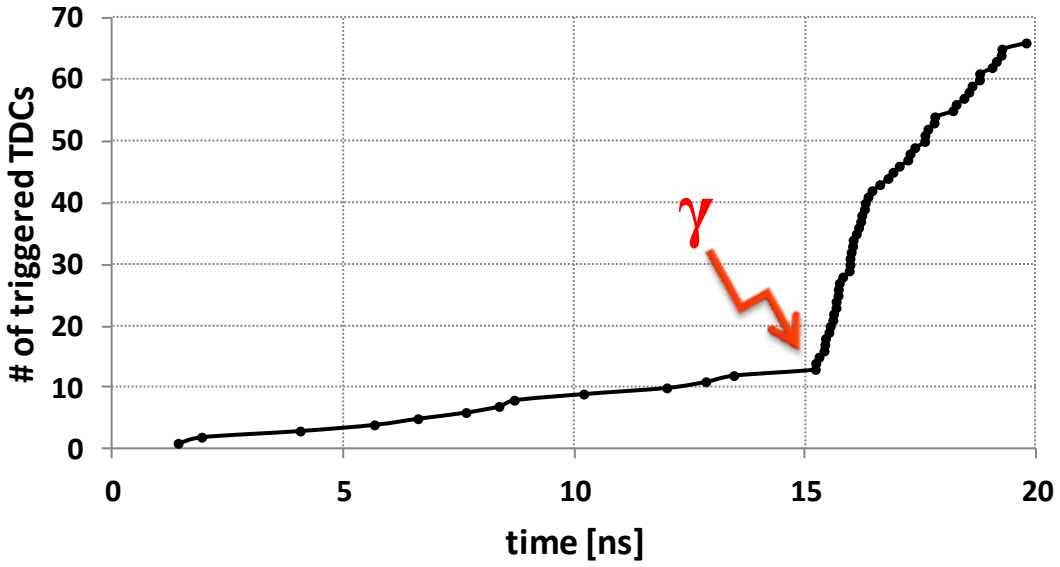


Figure 45: cumulative distribution of timestamps obtained from a typical gamma event, with a rough indication of the gamma ToA.

Three regions can be clearly distinguished in the figure: the first one, characterized by a very slow slope, is due to dark-generated timestamps occurring before the gamma event arrives. The second one, starting slightly after the 15 ns mark and characterized by a very fast slope, is due to the actual photons generated from the gamma scintillation event. The third and final one occurs right after the gamma fast slope, and is characterized by a slowing of the slope due to most of the TDCs being already triggered, and thus the probability of one photon actually generating a timestamp being greatly reduced.

To find the approximate starting point of the scintillation, i.e. the first photon, a threshold on the slope of the curve in Figure 45 can be used. To this aim, the intervals between each timestamp

are obtained, as shown in Figure 46. As can be observed, in this example, the dark-generated intervals have values around 1 ns, while the initial scintillation timestamps are below 100 ps.

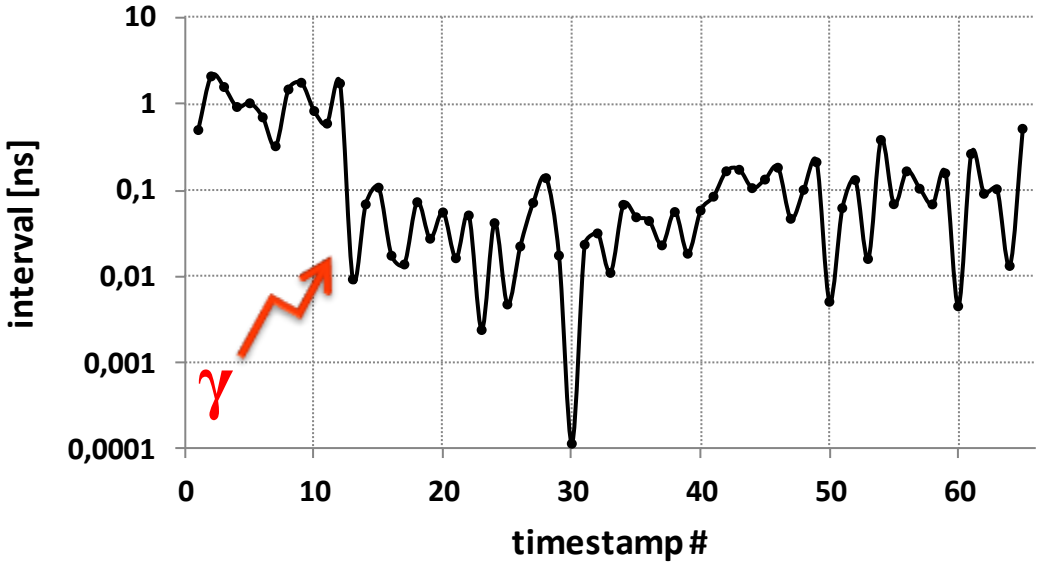


Figure 46: intervals between the timestamps shown in Figure 45.

However, due to the random nature of DCR, and furthermore due to optical cross-talk, there is a non-negligible probability that dark-generated timestamps also occur very close to each other, creating intervals similar to the scintillation-derived ones. Therefore, the first photon finder algorithm was specified to compare a number of consecutive timestamp intervals to a threshold, defining the first timestamp of the first interval group that is below the threshold as the first scintillation photon. For example, with the data shown in Figure 46 and the algorithm set at three intervals below 200 ps, the timestamp #13 would be correctly chosen, as all intervals before it are above 200 ps, while the intervals 13-to-14, 14-to-15 and 15-to-16 are the first 3 consecutive intervals that are all below 200 ps.

Both the number of intervals and the actual threshold value are configurable, and will vary according to the sensor configuration. As such, when measuring several different configurations in sequence, an automated method to obtain the best values for these two parameters is required. To illustrate the developed method, Figure 47 shows the coincidence time distribution using the first photon as the gamma ToA, for three different parameter combinations, using the measured data from which the example above was taken.

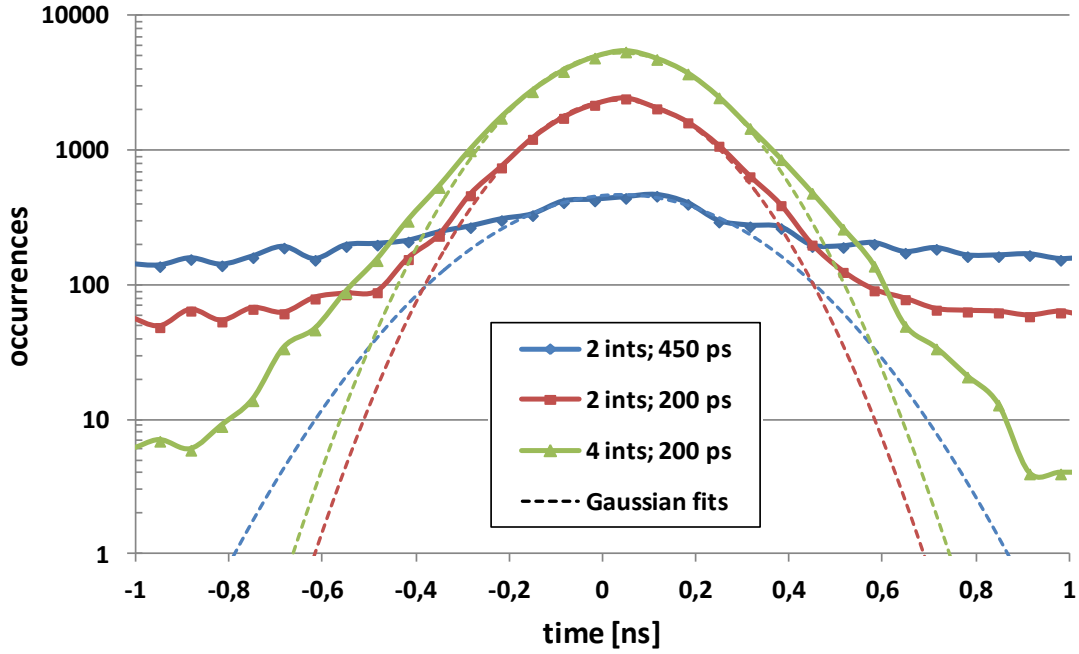


Figure 47: coincidence time distribution using the first photon as the gamma ToA, for three different parameter combinations of the first photon finder algorithm.

As can be observed in the figure, using a relatively high threshold and few consecutive intervals results in a barely perceptible coincidence peak, meaning that the first photon is being selected almost at random between dark timestamps. By reducing the threshold, and then increasing the number of consecutive intervals, the coincidence peak becomes clear, being well fitted to a Gaussian and featuring few occurrences outside the fit. Based on these results, the automated method for selecting the algorithm parameters was defined as finding the parameter combination with the highest peak value and fewer events outside the Gaussian fit.

5.2.3 ToA Estimators

After the first scintillation photon has been defined, the remaining timestamps are assumed to be due to photons and thus appropriate for use in the gamma ToA estimation. However, as illustrated by Figure 45 and Figure 46, a saturation effect starts to take shape after a couple dozen photons. This means that only the first few timestamps in the list will have an order that corresponds to the actual photon arrival order – e.g. the 5^{th} timestamp is likely to be the 5^{th} detected photon, but the 20^{th} timestamp is probably closer to the 25^{th} detected photon. As previously explained, this occurs due to presence of only one active TDC per pixel, meaning that any succeeding photons that hit an already triggered pixel do not have their timestamp registered.

In any case, as demonstrated by [Sei+12b], the initial scintillation photons are the most important for the ToA estimation, and thus this saturation effect may not necessarily hinder the CRT performance of the system. It does mean, however, that these high-order timestamps probably cannot be utilized by a multiple-timestamp estimator, or that at least the estimator must correctly model this saturation in order to utilize all available timestamps.

For the SPADnet characterization, two algorithms to estimate the gamma ToA from the remaining list of timestamps will be used. The first and simplest ToA estimator that can be implemented is a single timestamp estimator, which defines the time of arrival of the gamma as one specific timestamp. This timestamp can be the first, second, third, and so on, from the remaining timestamp list obtained by the sensor. As shown in Chapter 3, the order of the timestamp with the best CRT will depend on the sensor and scintillator characteristics.

Multiple timestamp estimators, on the other hand, combine the time of arrival of the first N photons to get a more precise estimate of the gamma arrival time. This class of estimators includes Maximum Likelihood Estimators (MLE), such as the one described in [Dam+13]. However, a MLE not only requires a very accurate model of the system in order to provide improved results, but also usually has considerable computational complexity, preventing its implementation for real-time ToA estimation in the SPADnet module FPGA.

As such, a hardware friendly, multiple-timestamp algorithm based on order statistics was developed [Bra+12]. This estimator is based on the idea that using multiple timestamps to obtain the gamma arrival time is intuitively better than using a single one since the system noise can be averaged. However, as shown in Chapter 3, each ordered scintillation photon has a different distribution, and thus simply averaging the timestamps will not produce the desired effect of reducing the noise.

To overcome this, the integral of the photon flux – which is simply the scintillation CDF multiplied by the number of detected photons as shown in equation (21) – can be used to obtain the expected timestamps of the ordered photons. These expected values can then be subtracted from the measured ones, in an operation somewhat equivalent to a deconvolution, and which results in values that should contain a measure of the noise in each timestamp. An estimator algorithm implementing this principle was developed, named the inverse CDF-based estimator, and is explained in detail in the next paragraphs.

$$C_{scin}(t|\Theta) = N_{ph} \cdot \left(1 - \frac{\tau_d \cdot e^{-\frac{t-\Theta}{\tau_d}} - \tau_r \cdot e^{-\frac{t-\Theta}{\tau_r}}}{\tau_d - \tau_r} \right) \quad (21)$$

The scinillation expected timestamps values can be obtained numerically through the inverse ${C_{scin}}^{-1}(n)$, where n is the photon order. This numerical calculation only needs to be performed once, offline, since the average number of detected photons can be used for N_{ph}, which, as with τ_d and τ_r, are characteristics known *a priori* to the system. With this pre-built list of expected timestamps and the ordered list of measured timestamps, the inverse CDF-based estimator, then, proceeds as follows:

- First, a constant is added to all expected timestamps so that the first value, ${C_{scin}}^{-1}(1)$, equals the first measured timestamp. This step is necessary since the offline calculation of the expected timestamps assumes $\Theta = 0$.
- Then, the differences between the two vectors – i.e. between each measured timestamp and ${C_{scin}}^{-1}(k)$, where k is the order of the timestamp – are obtained.
- Finally, the average of these differences is obtained and then added to the measured timestamp of the first photon, which is used as the final gamma ToA estimation.

From this description, it can be observed that each estimation requires only some addition and subtraction operations (as many as the number of timestamps used) and one multiplication/division in the last step. As such, it is relatively straightforward to implement this algorithm in the SPADnet module FPGA with the required speed and resource utilization

To analyze the effectiveness of this estimator, a Monte Carlo simulator was built based on the scintillation and sensor jitter models described in Chapter 3. The simulations are performed as follows: first, the number of detected photons is randomly generated according to a Gaussian distribution with mean N_{ph} and FWHM corresponding to a desired energy resolution. Then, each photon has its emission time randomly generated according to the scintillation PDF in equation (4). Next, for each photon, the sensor jitter is randomly generated according to the PDF in equation (6) and then added to the photon emission time, producing the final timestamp. The timestamps are then ordered, representing the sequence in which the TDCs would be triggered in a real event.

For each Monte Carlo iteration, the inverse CDF-based estimator is run on the resulting timestamp list to obtain a single arrival time for the gamma event. After many iterations, the timing resolution of the single detector Δt_{det} is obtained by taking the FWHM of the histogram of the estimated arrival times, and then the CRT is calculated assuming two identical detectors ($\sqrt{2} \cdot \Delta t_{det}$). The simulations were run with LYSO rise and decay constants, a sensor jitter with a FWHM of 250 ps and 1000 average detected photons (N_{ph}) with 12% energy resolution.

The internal data generated in the simulator can be used to demonstrate some of the estimator steps. Figure 48 uses the timestamps generated in a typical simulation iteration to, in (a), illustrate the comparison of the expected CDF with the Monte Carlo-generated one, and in (b), show the difference vector between the two curves for the first 15 timestamps. The CDF starts at a non-zero value due to the mean value of the signal transport in the detector (i.e. μ_{jitter} in equation (6)).

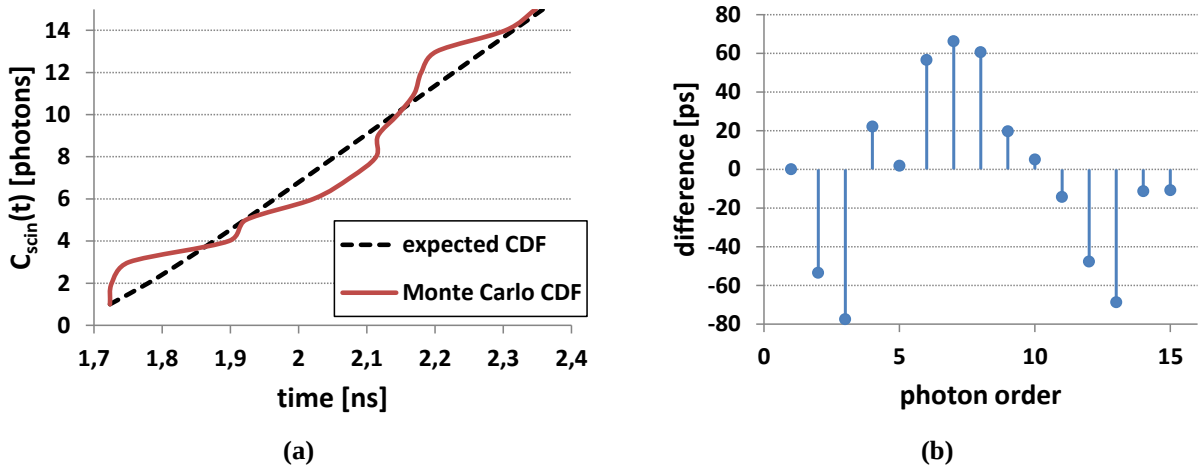


Figure 48: illustration of estimator steps with a typical Monte Carlo simulation-generated data: (a) comparison of expected and simulated CDF; (b) difference vector between the two.

The simulator was run with 50k iterations and its results analyzed using both the single-timestamp estimator and the inverse CDF-based multiple-timestamp estimator described above. Figure 49 plots the obtained CRTs using different photons as single-timestamp estimators, and using *1:n* photons for the multiple-timestamp estimators. Additionally, the CRT from the Cramér-Rao bound with the simulated system parameters is also plotted for comparison purposes.

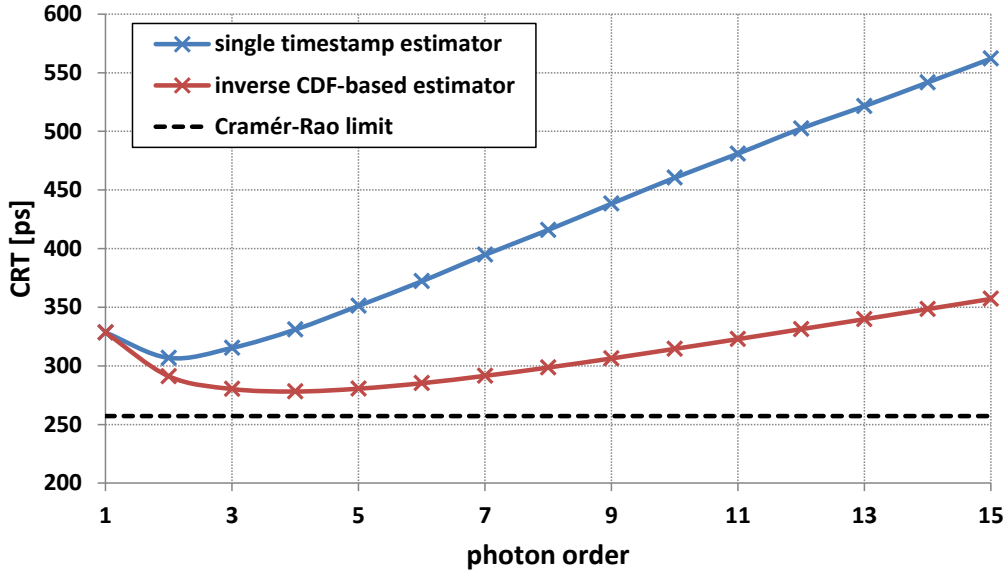


Figure 49: Monte Carlo simulation results comparing the CRT of single-timestamp estimators with the proposed multiple-timestamp estimators.

The results show that the proposed estimator provides a clear improvement over single-photon estimators. With the particular simulated system parameters, the best CRT for the inverse CDF-based estimator was obtained at n=4, at which point it offers around 10% improvement over the best single photon estimator (which occurs at n=2). It should be noted that, as explained in Chapter 3, the photon orders at which the best CRTs occurs will vary according to the number of detected photons and sensor jitter. Nonetheless, in actual experiments these numbers will be known, and thus so will the best photon orders.

Comparing the proposed estimator with the Cramér-Rao bound, it can be noted that the estimator does not converge to the bound, exhibiting an optimum number of timestamps at which point the CRT is about 8% worse than the CRB. This indicates that the proposed estimator does not perform as well as an unbiased Maximum-Likelihood estimator. Nonetheless, for the SPADnet architecture, it does improve the system timing resolution at no extra silicon cost and with low computational power requirements.

Chapter 6 Experimental Results

The SPADnet sensor assembly for experimental characterization is shown in Figure 50. Due to the TSV pads of SPADnet-I, the sensor does not uses a standard packaging: it is directly bonded to a small custom printed circuit board (PCB), referred to as the sensor board, which can then be plugged into a larger test PCB. This second PCB provides digitally-controllable voltage regulators for power supply and circuit bias and has a connector to a third board, the one housing an FPGA for sensor control and readout. For this, the Xilinx SP605 board was selected, which contains a Spartan-6 FPGA and a Gigabit Ethernet port for communication with the PC.

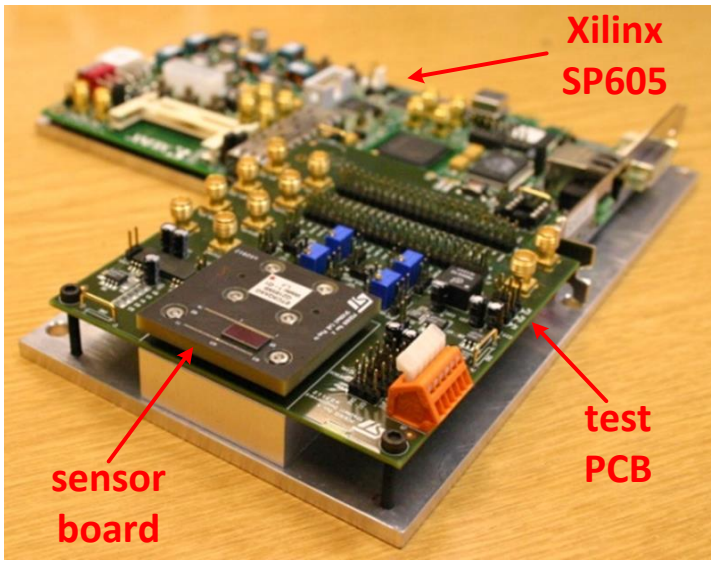


Figure 50: sensor characterization assembly.

The characterization of the sensor was done in two phases: first, the electrical and optical performance of the relevant blocks was evaluated, using, when needed, a pulsed laser together with a Time-Correlated Single Photon Counting (TCSPC) module. Then, scintillator measurements were performed using an LYSO crystal and a ^{22}Na gamma source. All measurements, except the ones requiring the laser, were performed inside a temperature chamber, which was kept at a constant 20^{o} C. The laser measurements were performed outside the chamber due to space constraints, and thus at room temperature.

6.1. Electro-Optical characterization

The first tests to be performed on the SPADnet-I sensor were functional tests to confirm that all digital functionality was operating correctly. To specifically check the feasibility of the 100 MHz clock frequency, the mini-SiPMs *TEST_PULSE* ports were used to input different data patterns into the adder tree and discrimination logic. All tests completed successfully, and thus for all SPADnet-I measurements the targeted clock frequency of 100 MHz was used.

In the following subsections, the key blocks of the sensor will be characterized for their performance metrics, namely: the monostable pulse width, the TDC linearity and resolution, the SPAD DCR and timing resolution and, finally, the full pixel jitter and skew.

6.1.1 Monostable pulse width

Test structures were included on the sensor die that allow the characterization of the current-starved monostable pulse width. This is done through two flip-flops connected to the monostable output, each of which switch on one of the monostable pulse edges. The monostable pulse width can then be computed from the time difference between the switching of the two flip-flops, measured by an oscilloscope. Figure 51 shows the results versus the monostable control voltage, which regulates the current flow. The circuit works correctly up to its maximum current, achieved at zero voltage, and at which the pulse width is in the order of 250 ps. To minimize the temporal compression loss, this minimum pulse width value was used for all gamma measurements.

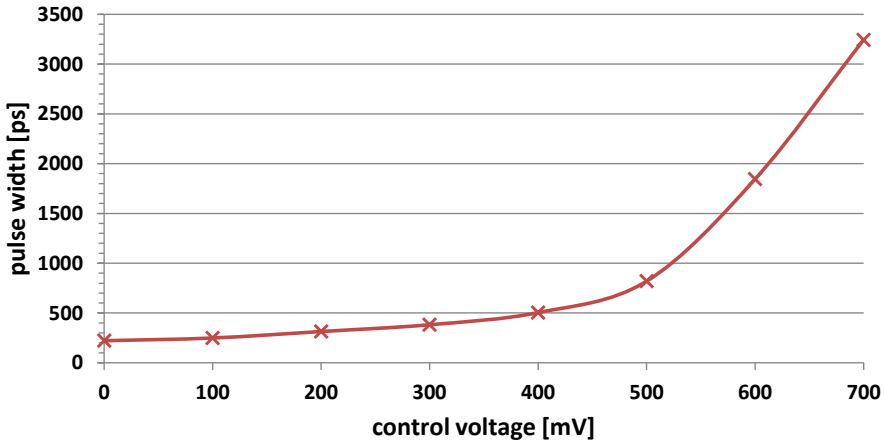


Figure 51: monostable pulse width versus control voltage.

6.1.2 TDC linearity and resolution

The TDC characterization aimed at obtaining three performance figures: INL, DNL and the resolution of the least significant bit (LSB). The INL and DNL performance numbers are obtained through a code density test, which was performed simultaneously in the whole array by enabling a few SPADs per pixel, and thus keeping a low pixel DCR. The pixel DCR value was chosen so that the exponential time distribution until the first trigger inside the 10 ns clock period could be approximated by an uniform distribution with 0.1% precision. This resulted in a DCR of about 100 kHz per pixel. Finally, to obtain the LSB resolution, the clock period is divided by the maximum TDC code obtained in the test.

Figure 52 plots, first, a histogram of the LSB resolutions of all 256 on-chip TDCs. As can be observed, the average TDC resolution is about 66 ps, with a very good uniformity across the array, resulting in a LSB standard deviation below 1 ps. Next, the typical, minimum and maximum INL and DNL values are plotted, showing, however, a significant performance degradation towards the initial and final 3-4 TDC codes. This degradation will heavily affect the timestamps of photons arriving near the edges of the time bin, and thus a more detailed discussion of its causes and consequences is merited.

To fully understand the origin of this INL and DNL behavior, it must first be noted that the TDC codes are inverted with respect to time, since the TDC is started by the photon instead of stopped by it. This means that low-valued codes represent pulses arriving at the end of the time bin, while high-valued codes correspond to pulses that arrive at the beginning of the bin. As such, the degradation at the initial codes was identified as due to a slow starting period of the TDC ring oscillator, as it may take up to 3 or 4 LSBs for the oscillator to reach its steady-state frequency. This results in a reduced occurrence probability for these initial codes, which then produces the steep drop in INL and DNL shown in Figure 52.

For the final TDC codes, there are actually two phenomena that occur. First, there is a slight increase in the occurrence probability, resulting in high DNL values, that happens towards the end of the TDC range. This peak is then followed by a “slow stop”, again for about 3 to 4 codes. The cause of the initial probability increase was identified as likely due to a mismatch between the PIPE signal at the mini-SiPMs, which separates the pulses into the A and B trains, and the PIPE at the pixel, which is used to define the operating window of the TDC. This mismatch can lead to

some pulses reaching the TDC before it is actually enabled, with the TDC then starting together with PIPE instead of the pulse, and thus creating an accumulation of occurrences at the last TDC codes. The slow stop, on the other hand, is believed to occur due to jitter in the PIPE signal at the edge where it is enabling the TDC window. In fact, this jitter, but at the edge where PIPE is stopping the TDC, also contributes to the slow start at the initial TDC codes.

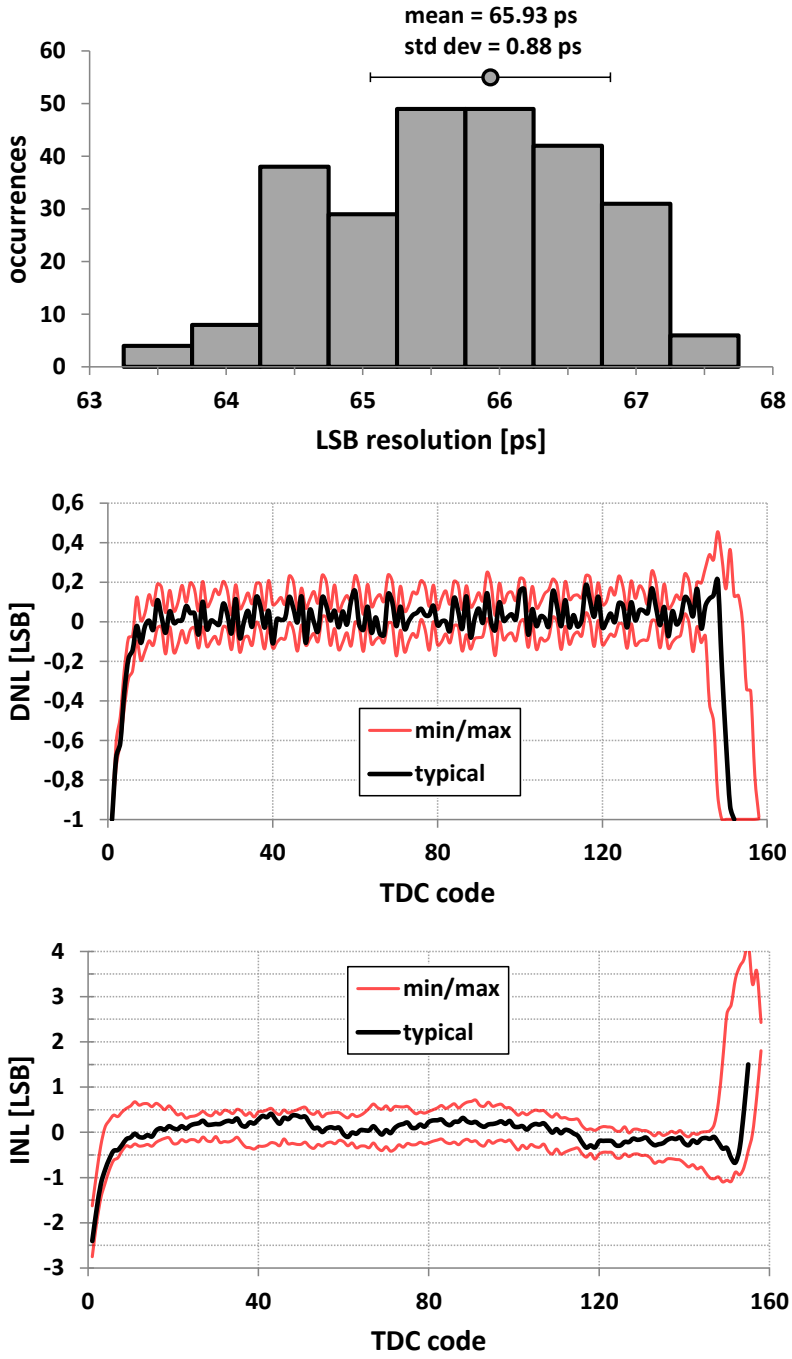
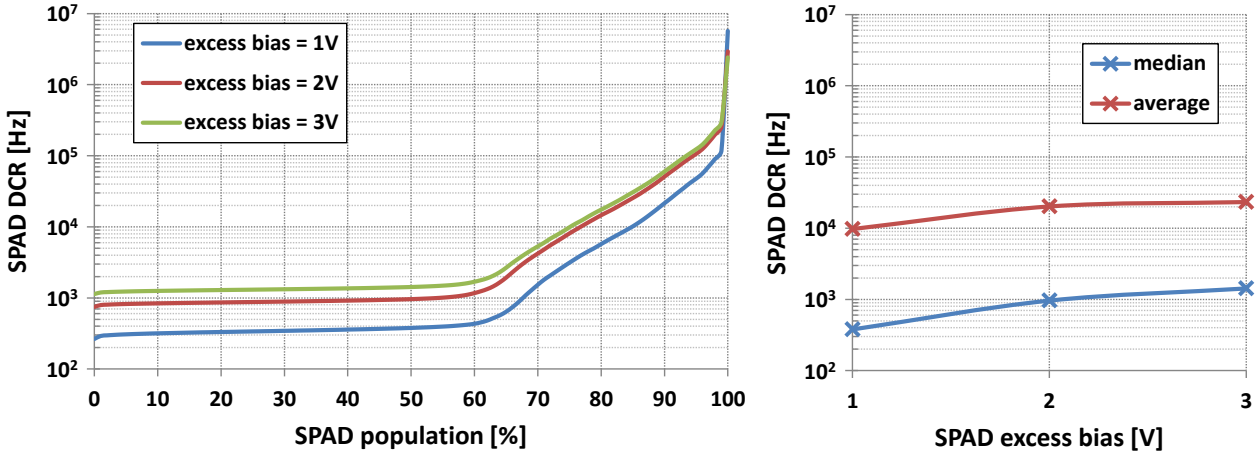


Figure 52: TDC characterization: (a) histogram of all TDCs LSB resolution; (b) and (c) minimum, maximum and typical DNL and INL values, respectively.

Even if the above described issues affect both the INL and DNL curves, the INL curve, as explained in Chapter 5, can be fully corrected for, leaving the DNL as the main linearity figure of the TDC. Its consequences on the gamma ToA estimation will be discussed in section 6.2.1. In any case, considering only the central part of the TDC range, the DNL is actually quite good: it stays between -0.17 and $+0.25$ LSB for more than 90% of the range. Within the same range, the INL is between -0.8 and $+0.7$ LSB.

6.1.3 SPAD DCR and timing resolution

The DCR of each SPAD has been individually measured by enabling a single SPAD per pixel and integrating the resulting counts for a total period of 2 seconds. The distribution of the obtained DCR values at different excess bias voltages is shown in Figure 53, along with the median and average values.



(a) **(b)**

Figure 53: characterization results of the SPAD DCR: (a) population distribution; (b) median and average values.

The measurements show a SPAD yield slightly above 60%, which is significantly worse than expected from [Ric+11] and which results in an average SPAD DCR almost 10× higher than in [Ric+11]. This yield drop has been traced back to fabrication process changes that are out of control for the designers, and that are being worked on by the foundry. Nonetheless, the resulting pixel average DCR, between 15 and 20 MHz, should not hamper the correct operation of the SPADnet sensor. Additionally, the highest DCR SPADs can be individually disabled with the goal of improving the DCR vs. FF compromise. This compromise will be analyzed in section

6.2.2, where also an estimate of the PDE obtained from gamma measurements will be assessed versus the SPAD excess bias voltage.

A few SPADs on the sensor have their Schmitt trigger outputs directly connected to test pads, allowing for the characterization of the SPAD timing resolution isolated from other system blocks, such as the compression tree and the TDC. For this measurement, a 470-nm, 70-ps pulsed laser (PicoQuant LDH-P-C-470) and the TCSPC module PicoHarp 300 were used. Figure 54 shows the results for different excess bias voltages, without discounting the laser or TCSPC module intrinsic timing resolutions.

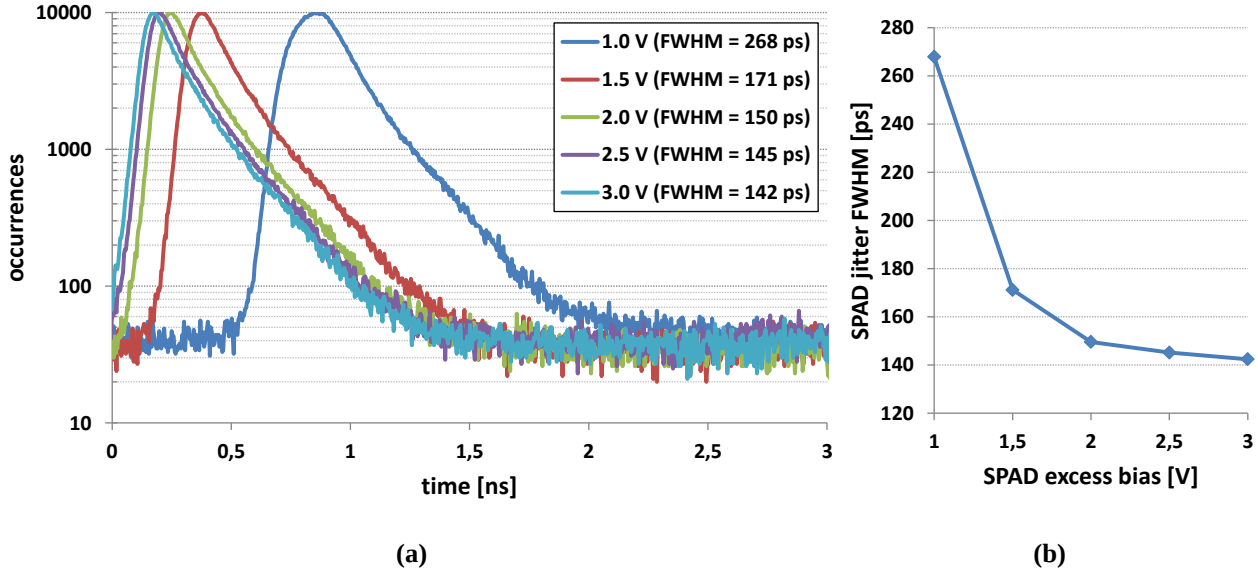


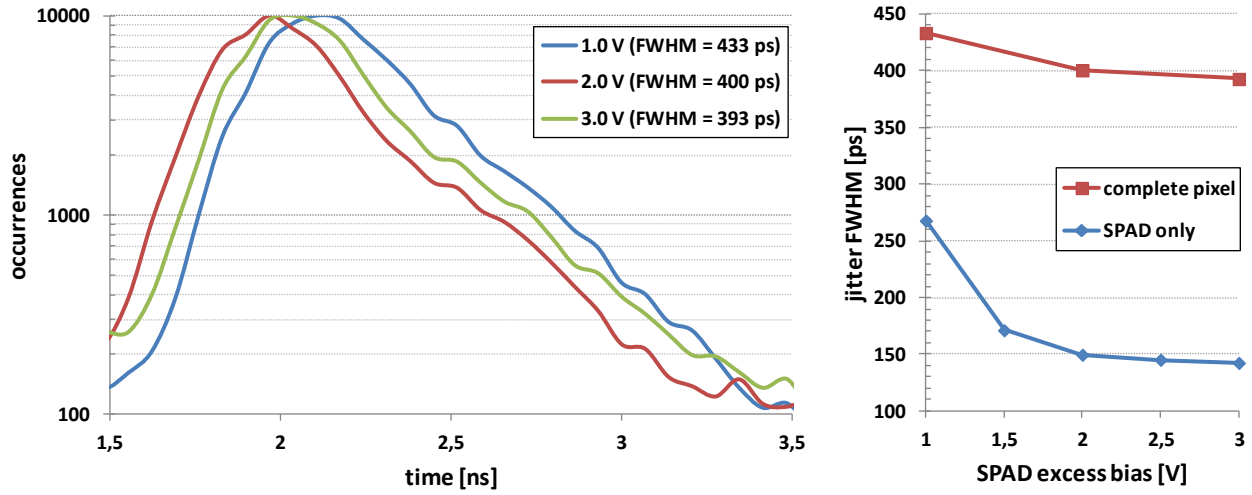
Figure 54: SPAD-only timing resolution for different excess bias voltages: (a) histogram of timestamps; (b) obtained FWHM versus excess bias.

As can be observed, the SPAD timing resolution starts stabilizing after around 2 V excess bias, with the best FWHM value of 142 ps occurring at the maximum tested voltage of 3 V. These results show that low SPAD excess bias voltages should be avoided during gamma measurements, since the resulting high SPAD jitter should significantly worsen the system CRT.

6.1.4 Pixel jitter and skew

Similarly to the SPAD jitter, the pixel jitter was measured with the same 70-ps pulsed laser, taking the TDC timestamps as the measurement for the photons time of arrival. This ensures that the full system jitter is being assessed, including the previously measured TDC resolution and SPAD jitter and also the mini-SiPM compression circuit and the clock jitter. Figure 55 plots the obtained

timestamp histogram and FWHMs for a typical pixel, together with the SPAD-only jitter FWHM for comparison purposes.



(a) **(b)**

Figure 55: jitter of a typical pixel for different SPAD excess bias voltages: (a) histogram of timestamps; (b) obtained FWHM versus excess bias.

The pixel jitter is significantly higher than the SPAD jitter or the TDC resolution, indicating that some other effect is dominating the pixel timing resolution. If independent jitter sources are assumed, due to the ensuing squared sum of the individual jitter values, a component with a very large jitter by itself, very close to the total value, would need to be present to justify the high pixel jitter. Such a high-jitter individual phenomenon is unlikely, and thus the pixel jitter is most probably explained by skew in the paths between the various SPADs in a pixel and the TDCs. Even if the signal path and the gates of the compression tree have been optimized for equalized delays, higher than expected process parameter variations could affect the SPAD front-end enough to generate the skew that results in this high pixel jitter. In fact, per-SPAD breakdown voltage measurements showed a pattern in accordance to the electronics floorplan, corroborating this hypothesis.

Figure 56 plots a histogram of the jitter FWHM of all pixels, at the three measured SPAD excess bias voltages. The graph shows a jitter FWHM standard deviation among the pixels in the order of 40 to 50 ps. More interesting, however, is looking at the mean values obtained from each pixel timestamp histogram (i.e. the mean values of the curves in Figure 55(a)). These values can be used as a measure of the pixel skew, which arises from the clock distribution tree on the sensor.

The map of these mean values for a 3 V SPAD excess bias, along with the map of the FWHM for comparison, is shown in Figure 57.

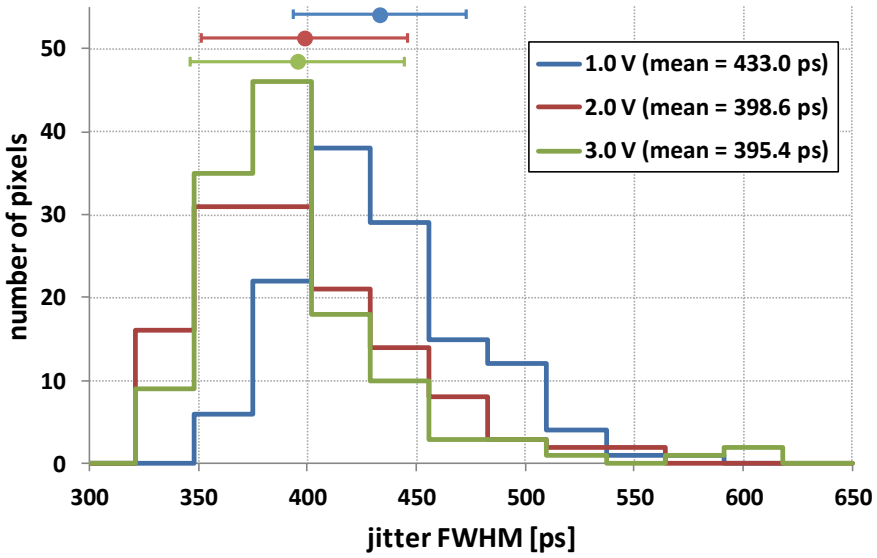


Figure 56: histogram of all pixels jitter FWHM.

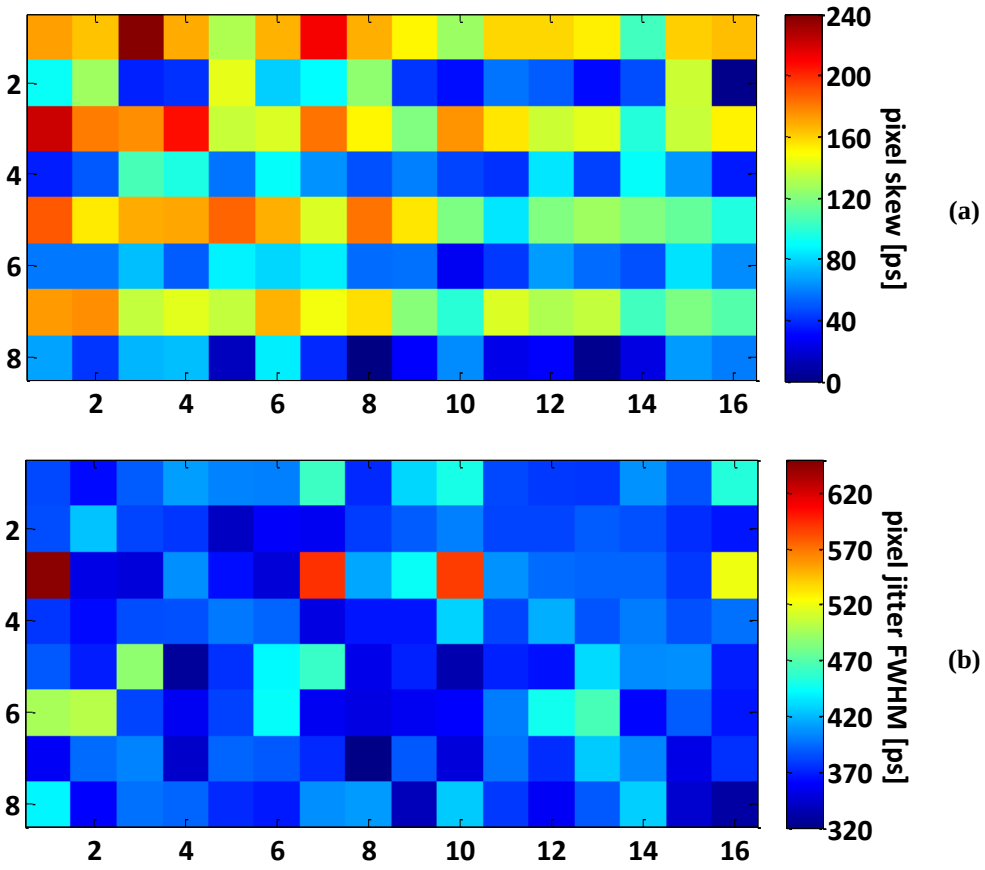


Figure 57: map of pixel skew (a) and jitter FWHM (b) at 3 V SPAD excess bias.

As can be observed, there is a clear pattern in the pixel skew map, with rows interchanging between a higher and lower average skew. The skew values themselves are relatively low, indicating a good clock tree balancing. In any case, the obtained skew map can be used to correct the pixel timestamps during gamma measurements. The pixel jitter FWHM values, on the other hand, are randomly distributed on the map with no clear pattern, as expected.

6.2. Scintillation characterization

For the scintillator-based characterization, two LYSO crystals of 3×3×5 mm^3 in size and wrapped in Teflon were optically coupled to two SPADnet-I samples using Cargille Meltmount glue. To emulate a PET experiment, a Na^{22} gamma source with 370kBq of activity was put in the middle of the two sensors, using the mechanical setup shown in Figure 58. As previously mentioned, this source emits 511 keV photons, as in a PET experiment, and also 1275 keV photons, allowing the calibration of the sensor response.

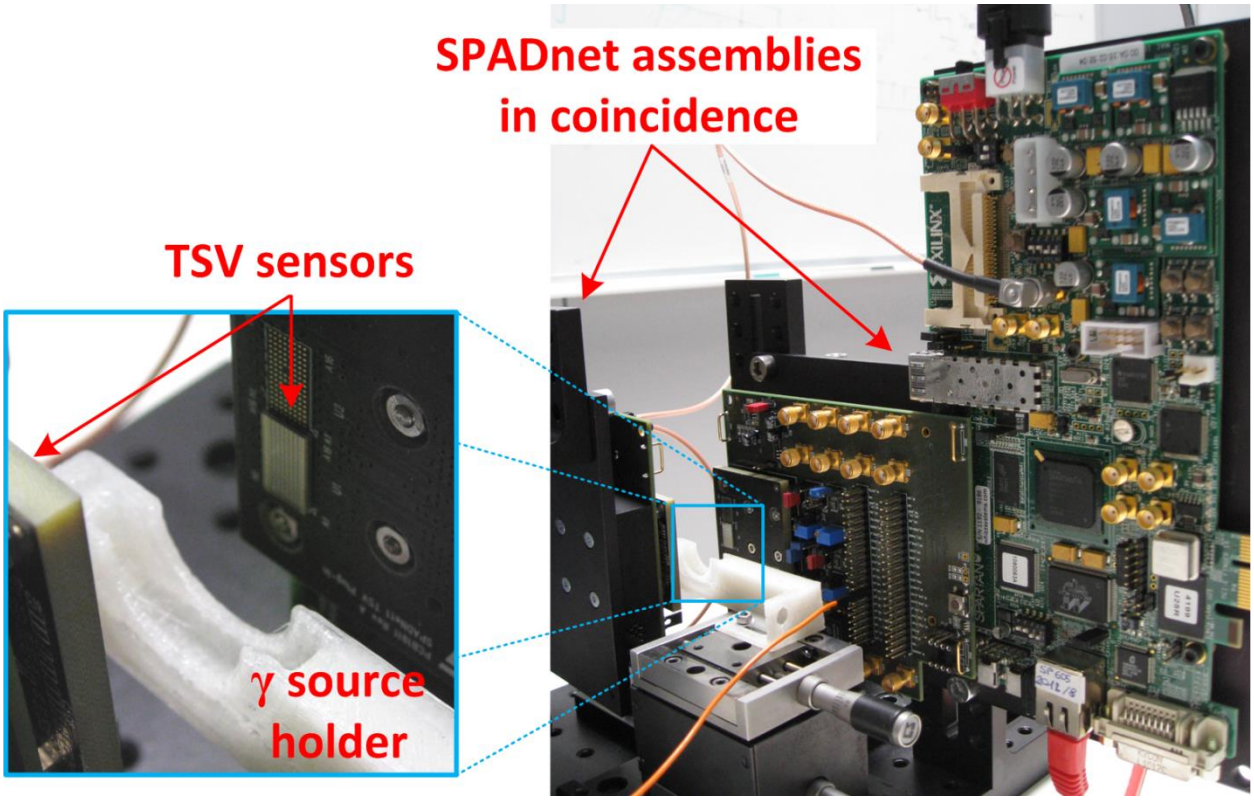


Figure 58: setup for gamma coincidence measurements.

The gamma measurements results are organized as follows: first, a general picture of the sensor functionality will be given, verifying some of the design parameters and confirming that the top-

level requirements were met. Next, a performance optimization measurement campaign will be described, sweeping different sensor configurations and comparing the results to the models described in Chapter 3. Finally, detailed measurement results will be presented for the best performing sensor configuration, which is suggested as a standard for future PET-system implementations using the SPADnet-I sensor.

6.2.1 System functionality

A. Gamma detection overview

The best illustration of the correct operation of the system is given by the real-time energy output from the adder tree and by the pixel map of a gamma event. Figure 59 shows these two outputs for three examples of typical scintillation events: a 511 keV gamma, a 1275 keV gamma and a pile-up event with a total energy similar to that of a 511 keV event. These results were obtained with all sensor SPADs enabled, at 2 V excess bias and with 150 ns integration time.

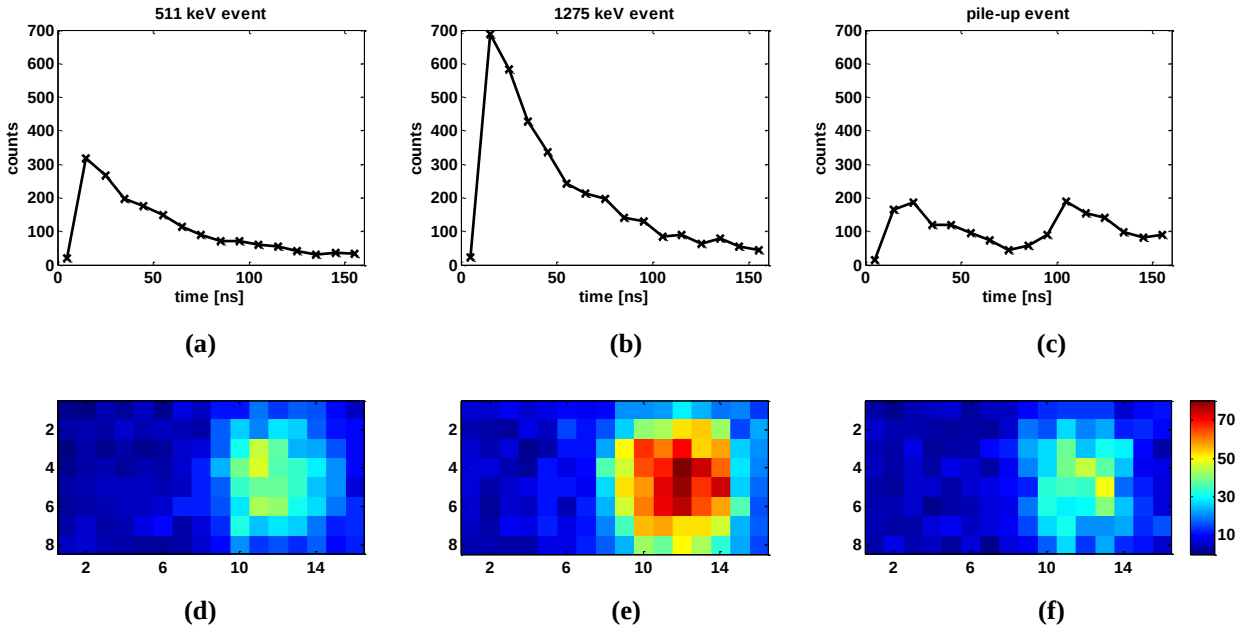


Figure 59: examples of typical 511 keV, 1275 keV and pile-up events: (a)-(c) real-time energy output and (d)-(f) pixel counts.

These examples show, for instance, how the sensor can detect pile-up events such as the one in (c), with the real-time energy output presenting two peaks inside one integration window. Additionally, (d) and (e) show that the photon distribution is actually non-uniform, and illustrate how the fine-grained pixel pitch (~ 0.6 × 0.6 mm^2) of SPADnet can be used to optimize the integra-

tion area. To complement these examples, Figure 60 shows the pixel average count map and the TDC triggering probability map obtained from 75k events with energies around 511 keV, and using the same measurement conditions as above.

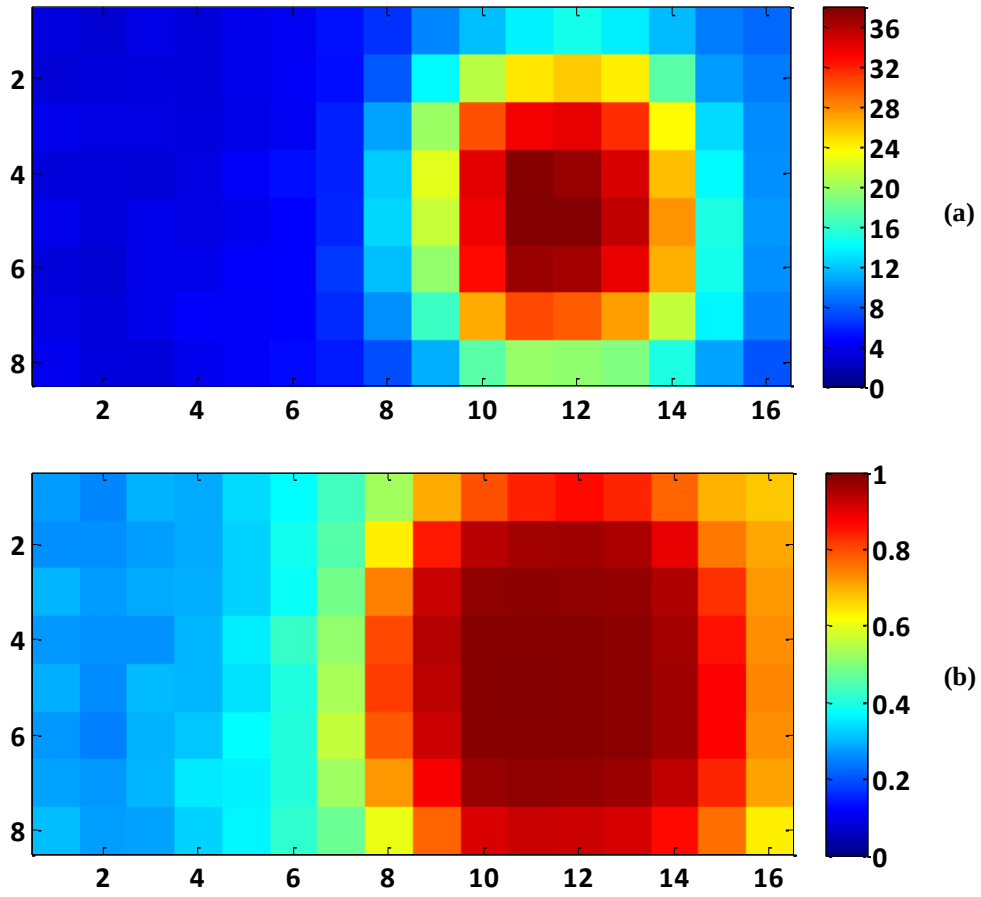


Figure 60: pixel maps averaged for 75k events with energy around 511 keV: (a) pixel counts and (b) TDC triggering probability.

Even in the average pixel count map, the photon distribution is very non-uniform, with only a small area at the middle of the crystal with a relatively uniform, albeit high, count value. Moreover, the photons are actually spread in an area significantly larger than the LYSO crystal itself, of circa 8 × 8 pixels. This can be mainly attributed to a 500 µm thick glass that is attached to the wafer front-side during the TSV fabrication process, which works as a light guide, further increasing the light spot on the sensor surface. Finally, regarding the TDC triggering probability, the area directly below the crystal has close to 100% probability of generating a timestamp, which then gradually reduces outside the crystal, similarly to the count distribution.

B. Discrimination and ToA distribution analyses

The correct operation and the importance of the discriminator block are demonstrated in Figure 61. This figure shows several output spectrums – not converted to gamma energy –, each with a different discriminator configuration. The first curve, marked "th 1", was obtained with both thresholds very close to the DCR level, at 35 counts. In this configuration, most of the events read out by the sensor are purely due to noise, as evidenced by the low energy, out of scale, peak. However, by sequentially increasing the second discriminator threshold, first the noise peak can be greatly reduced ("th 2", 40 counts), then completely removed ("th 3", 45 counts), followed by also filtering out the unwanted scattered events, and finally leaving only the 511 keV peak and above ("th 6", 180 counts). In all cases, the total number of acquired events was the same, and thus the increase in height of the photo peak highlights the increased efficiency. Converting the obtained sensor counts to gamma energy shows that the minimum discerned gamma energy is below 40 keV.

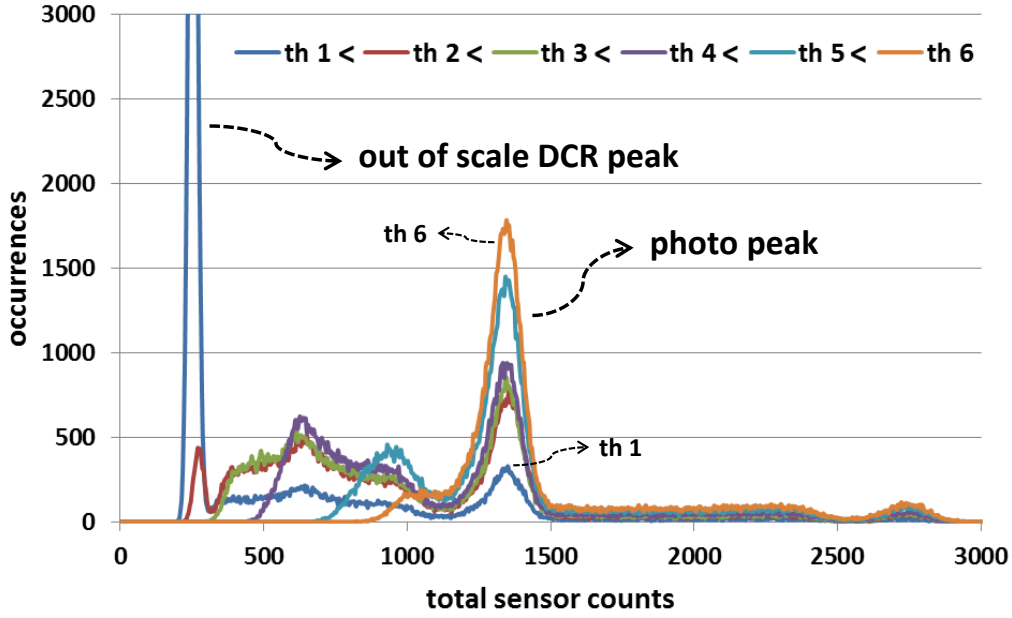


Figure 61: total sensor counts histogram for different discriminator thresholds.

Even if the above experiment was performed with the first threshold fixed at 35 counts, further analysis has shown that the first threshold may be set to zero with no change in the sensor performance. In fact, the important outcome of the double threshold mechanism is that the data from the bin *before* the high-count discriminated bin must also be saved, since there is a non-negligible probability that the event has started close to the end of this previous bin. As such,

keeping the first threshold near the noise level or at zero does not alter the discriminator efficiency, and for all subsequent gamma measurements, the first threshold was kept at zero, while the second threshold was set so as to remove DCR and low-energy events from the detected spectrum.

The uncertainty in which time bin the gamma event has started can be illustrated by plotting the measured distribution of the 511 keV gamma events in time. This graph is obtained by, first, calculating the time of arrival of each detected gamma event inside the photo peak. Then, these ToAs are histogrammed with reference to the first two integrated time bins (i.e. the two bins that were compared to the two discriminator thresholds). The resulting distribution is shown in Figure 62.

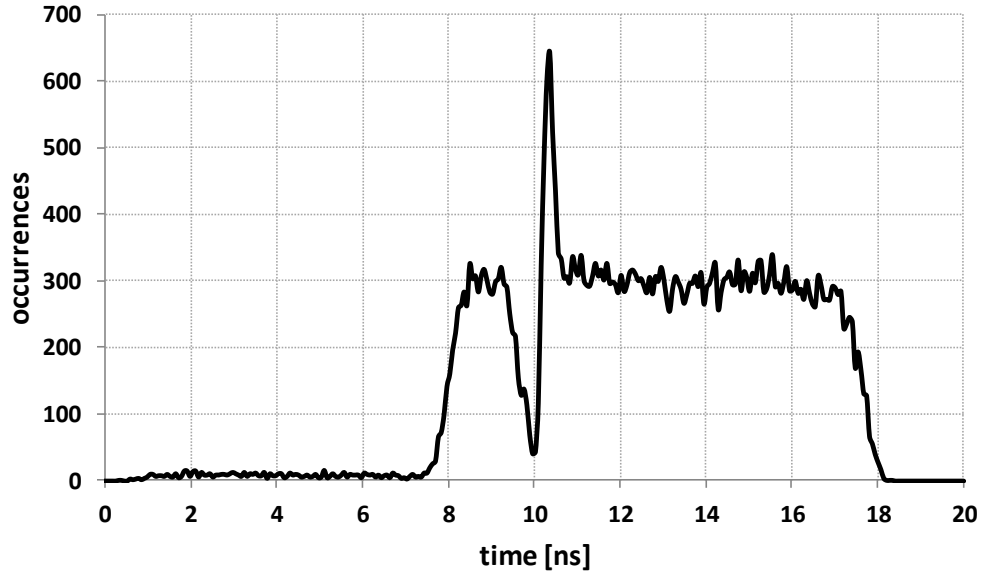


Figure 62: measured distribution in time of 511 keV gamma events.

Ignoring for a moment the valley and peak that occur around the 10 ns mark, the plot shows an approximately uniform distribution from about 8 ns to 18 ns. The uniform shape is a consequence of the random arrival of events in time, while the 10 ns duration of the distribution comes from the clock period of the system, as every event is discriminated at one specific time bin of the system. The time values where the distribution starts and ends, on the other hand, depend on the high-valued (second) discriminator threshold: the higher it is, the *earlier* will be the start of the distribution. This occurs since, as this threshold is increased, the more photons can arrive in the previous bin without triggering discrimination. The distribution example in Figure 62 shows a relatively *low* second threshold being used, as only events arriving closer than about 2 ns to the

end of a time bin (i.e. after about 8 ns in the graph) are discriminated at the second bin. If an event arrives at the first 8 ns of a time bin (i.e. between 10 and 18 ns in the graph), that bin itself will trigger discrimination, and thus the event will be plotted in the second bin of integration. As such, the total length of the distribution will always equate to one clock cycle.

The deep valley followed by a peak at the exact shift from one time bin to the next is a direct consequence of the poor DNL performance at the beginning and end of the TDCs range. As discussed in section 6.1.2, triggers arriving at the very beginning or end of a time bin will have a reduced probability of generating a timestamp. Therefore, if a gamma event arrives during a time bin transition, it will initially generate very few timestamps, and then, as the TDC starts entering its linear domain, many timestamps will finally be generated. The resulting effect on the time of arrival estimation process is that the gamma events that arrive during a time bin transition will have their ToA wrongly estimated later in the second time bin, thus generating the valley-peak behavior shown in Figure 62. Due to these wrongly estimated ToAs, all events in the valley-peak region (i.e. with ToAs roughly between 9.5 and 10.5 ns) will be discarded for the final CRT calculation, as they would heavily distort the coincidence distribution.

As a last note on the time distribution of events in Figure 62, the few events that occur from 0 up to slightly before 8 ns are due to the first photon of these events being wrongly attributed to a dark count. In these cases, a few dark counts occurred in a small interval, probably also due to cross-talk, resulting in the first photon finder algorithm making a mistake. This is anyway the expected behavior of the system, and these events are neither discarded nor specially processed for the final CRT calculation.

C. Mini-SiPM compression losses assessment

The sensor energy response curve was shown in Figure 42, which was also obtained with all SPADs enabled, at 2 V excess bias and with 150 ns integration time. This curve presented about 10% compression loss, that is, at 511 keV the average counts are 10% lower than the theoretical linear response. This value is in very good agreement with the goal defined when sizing the mini-SiPM. To complement this measurement, using the SPADnet sensor configuration capabilities, the individual contributions that result in this 10% compression loss can be independently assessed. As previously mentioned, these contributions are due to three main factors: single SPAD dead time, spatial compression and temporal compression.

To quantify the spatial compression, different spatial compression ratios were emulated by first enabling a single SPAD per spatially compressed group, then two SPADs in a group every two groups, and finally all SPADs in a group, every three groups. This way, the relative PDE between the measurements is kept constant (at 33% of the actual PDE), and the results can be directly compared. For the temporal compression, the monostable pulse width was swept in accordance with the results presented in Figure 51, with all SPADs enabled. Since the actual counts with no temporal compression could only be measured with a theoretical zero width, a spline fit was used to extrapolate the total counts at zero. The results for both techniques are shown in Figure 63, normalized by the no-compression values.

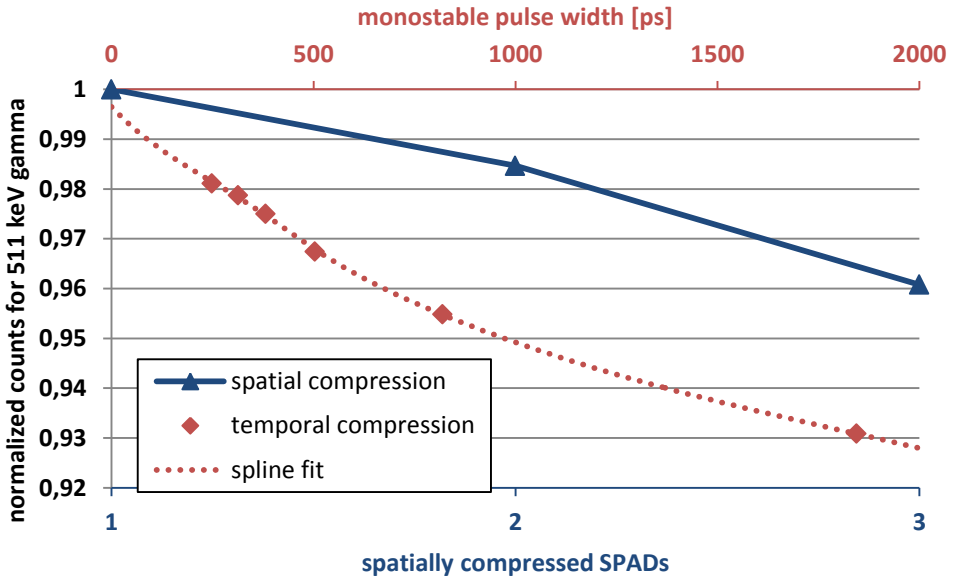


Figure 63: spatial and temporal compression loss assessment.

Regarding the spatial compression, the photon loss with the implemented mini-SiPM, i.e. with 3 SPADs spatially compressed, was measured at around 4%. This number is in very good agreement with the value estimated from section 4.3.1 with a 4×4 mm^2 light spot. Even if the actual integration area, as shown in Figure 60, is larger than this, the photon density is very non-uniform, resulting in a good approximation to a uniform distribution in a 4×4 mm^2 area.

The results for temporal compression indicate around 2% loss at the minimum pulse width of 250 ps. This is a very positive result, which demonstrates the potential of the temporal compression scheme: it allows a significant reduction in electronics area occupation while causing minimum photon loss. Even if this value is also well matched to the estimations in section 4.3.1, this

similarity must be pondered with the fact that the estimation was a really worst case, as only the peak flux was considered, while a 500 ps pulse width value was assumed.

Finally, the above measurements result in that only around 4% of the photon loss (the remainder) is due to the typical saturation cause in SiPMs, i.e. photons hitting a SPAD during its dead time. Moreover, it is important to note that all the above compression results depend on the impinging photon density, and as such will vary with the scintillator dimensions and with the use of an additional light guide.

6.2.2 Performance optimization

The various configuration options of the SPADnet sensor allow for a performance optimization procedure to be performed, sweeping different sensor parameters. In this section, the SPAD excess bias voltage and the number of SPADs enabled parameters will be analysed, with the goal of optimizing both the energy and timing resolutions. It is interesting to note that all these parameters are simultaneously changing both the PDE and the DCR of the sensor, and thus don't have a clearly expected behaviour. Moreover, the obtained results will be compared to the sensor models described in Chapter 3.

A. Energy resolution

Before comparing it to the performance optimization results, the model developed in section 3.2 needs to be slightly improved to better match the measured data. This is required due to the sensor response, which compresses the number of counts with respect to the linear response modeled in equation (2). To this aim, the number of photons N_{ph} and dark counts N_{noise} in equation (2) can be replaced by the actual number of SPAD triggers N_{fired} from equation (19). This results in an energy resolution model as shown in equation (22), where ph_{ratio} represents the proportion of the total counts that are due to photons in a 511 keV event, obtained based on the sensor response curve in equation (19).

$$En_{res} = \left[\left(2.35 \cdot \frac{\sqrt{N_{fired}(511)}}{N_{fired}(511) \cdot ph_{ratio}}\right)^2 + (En_{res})^2_{scin}\right]^{\frac{1}{2}} \tag{22}$$

$$ph_{ratio} = \frac{511 \cdot k}{511 \cdot k + DC} \tag{23}$$

It is important to note that using N_{fired} in equation (22) means that any deviation from the Poisson distribution due to the sensor compression is being considered negligible. Moreover, since ph_{ratio} depends on the fitting of the sensor response, the first step in comparing the measured energy resolution to the model of equation (22) is obtaining k and DC for all measured configurations. This is shown in Figure 64, where SPADs were disabled starting from the highest DCR one and a 150 ns integration time was used.

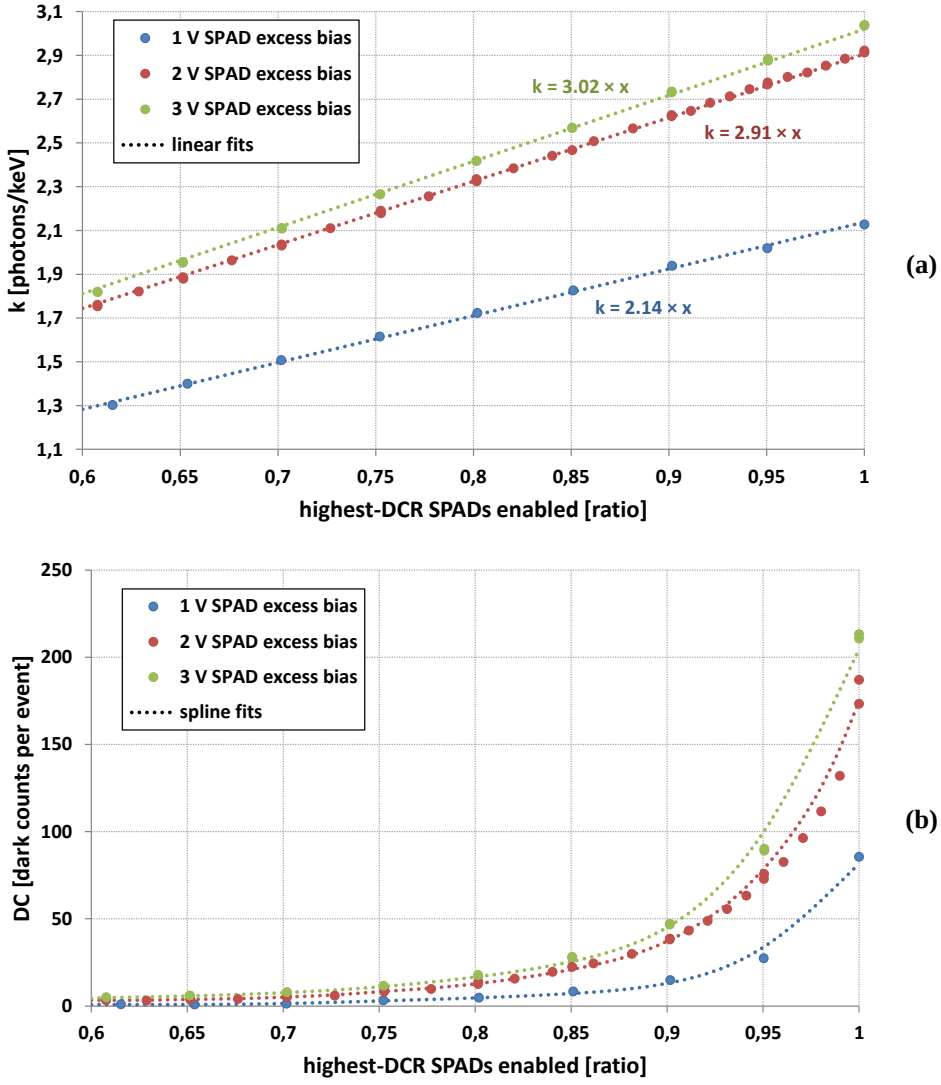


Figure 64: parameters of the sensor response fits for different SPAD excess bias and number of SPADs enabled: (a) k; (b) DC. Highest DCR SPADs disabled first.

Regarding the sensor PDE, the obtained k's for each measurement point fit very well to a linear curve, which is expected since the FF is changing linearly, while all other factors affecting k are kept constant (SPAD PDP and optical coupling). The linear fits also serve to compare the sensor PDE versus the SPAD excess bias: as can be observed, the PDE increases about 25% from 1V to 2V, but then only about 5% from 2 V to 3 V. This second step is even smaller than the DCR increase of about 15%, which was measured in section 6.1.3. Nevertheless, the effects of the SPAD excess bias on the energy resolution still merit investigation, since, as demonstrated in Chapter 3, PDE has a bigger impact on sensor performance than DCR.

The sensor DCR behaviour is also in accordance to expectations, given the SPAD yield shown in Figure 53(a): disabling the highest DCR SPADs results in an exponential decrease in the total sensor DCR. From about 75% of SPADs enabled and below, the sensor DCR results in about 10 dark counts or less per event, which, according to the model in section 3.2, means that the energy resolution will no longer be affected by DCR.

For the investigation of the effects of DCR and PDE on the energy resolution of the sensor, it is also interesting to take another approach to the order with which the SPADs are disabled: disabling the lowest DCR SPADs first. This should result in an almost constant DCR for the entire range of SPADs disabled, allowing the PDE effects to be verified independently. Figure 65 plots the fitted DC parameter, as in Figure 64(b), but now using this new order for disabling the SPADs. As can be observed, the desired effect of an almost constant DCR is achieved.

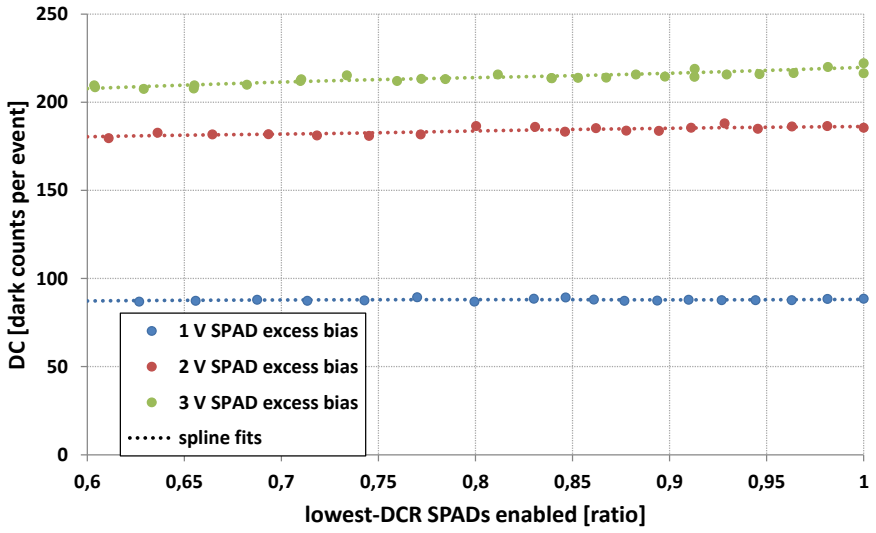


Figure 65: DC of the sensor response equation fits for different SPAD excess bias and number of SPADs enabled. Lowest DCR SPADs disabled first.

Finally, the measured energy resolution versus SPAD excess bias and number of SPADs enabled with 150 ns integration time is shown in Figure 66. For the plotted model estimations, the values of k and DC were taken using the fits shown in Figure 64 and Figure 65, and the LYSO intrinsic resolution was considered 8%, as in Chapter 3.

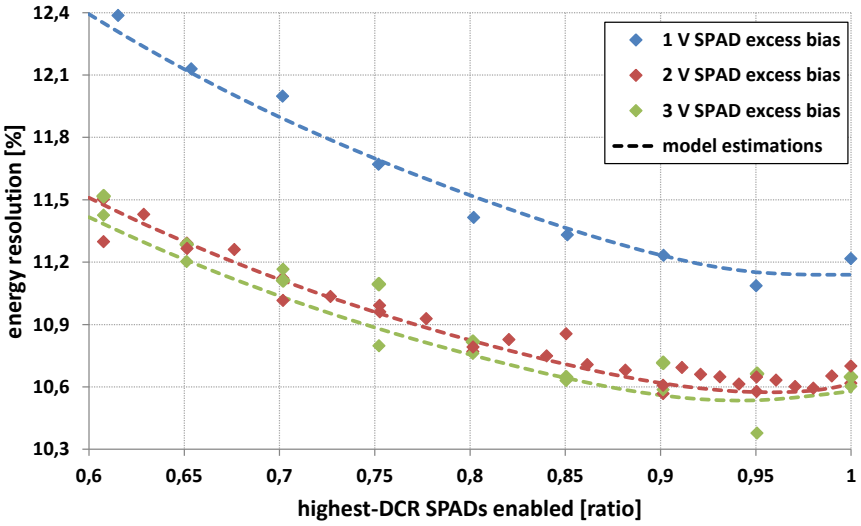


Figure 66: energy resolution for different SPAD excess bias and number of SPADs enabled, with highest DCR SPADs disabled first.

Several interesting observations can be made from these results. First and foremost, a SPAD excess bias of 3 V produces the best energy resolution values, even if by only a negligible margin of 0.04% with respect to the 2 V model, both at their optimum points. These optimum points occur at 94%, 96% and 100% for 3 V, 2 V and 1 V SPAD excess bias, respectively. Moreover, the model was shown to be a very good match to the experimental results for all measured SPAD excess bias voltages. Finally, the improvement from all SPADs enabled to the optimum is negligible, at less than 0.1%, whereas reducing the overall PDE down to 60% worsens the resolution only by less than 1% for the 2 V and 3 V SPAD excess bias cases. This later fact can be explained by the simultaneous strong reduction in DCR.

Therefore, Figure 67 compares the measured energy resolution and its model estimation at 2 V SPAD excess bias for both orders of SPAD disabling: highest DCR first and lowest DCR first. By disabling the lowest DCR SPADs first, the energy resolution worsening is significantly steeper, with a reduction of almost 2% from all SPADs enabled to 60% enabled.

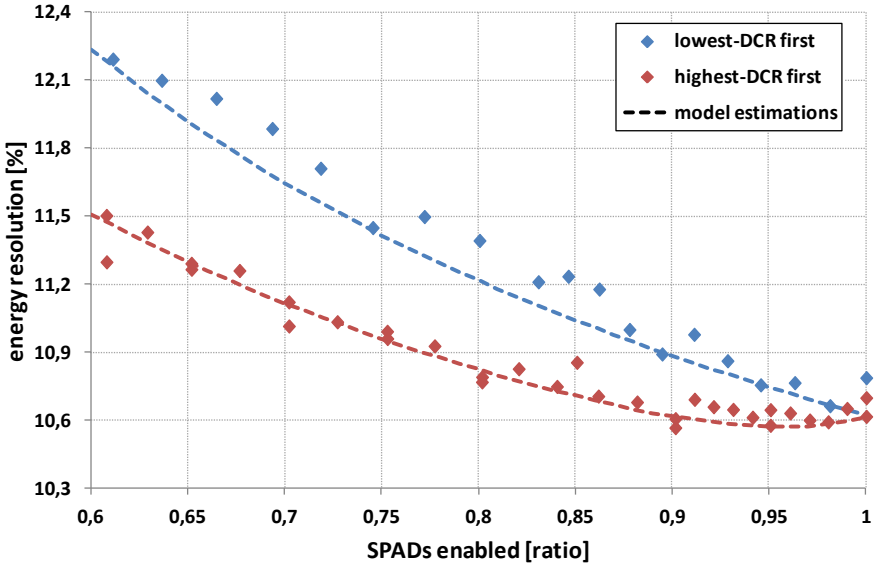


Figure 67: energy resolution versus number of SPADs enabled at 2 V SPAD excess bias, with two sorting orders for disabling the SPADs: highest DCR first and lowest DCR first.

B. Timing resolution

Similarly to the energy resolution analysis above, the CRT optimization can also be compared to the models developed in Chapter 3. For this, the model described by equation (12) will be applied, since it is the more complete one, taking into consideration the scintillation photons distribution, the sensor jitter and the DCR. Furthermore, the CRT will be calculated through the CRB method. Finally, as opposed to the energy resolution model, in the CRT model the number of detected photons and of dark counts will not be replaced with their compressed values. The reason for this is that the implemented ToA estimation algorithms, discussed in Chapter 5, use only the very few first photons of the scintillation, which are not strongly subjected to compression.

The CRT optimization measurements were performed for 2 V and 3 V SPAD excess bias and sweeping the number of enabled SPADs with the highest DCR SPADs disabled first. The integration time value has no effect on the CRT performance, but it was set in any case to 150 ns. Due to the relatively long time required to obtain coincidence data (few hours for each measurement point), and since the data for 2 V and 3 V SPAD excess bias was virtually indistinguishable, the measurement campaign focused on obtaining the most data with 2 V SPAD excess bias.

The results of this campaign are presented in Figure 68, with the CRT values being obtained with the inverse CDF-based estimator described in Chapter 5, selected at the best order for each measurement point. The model estimations, on the other hand, were calculated using the N_{ph} and DC values of that were obtained for the energy resolution model, shown in Figure 64. Addition-

ally, the model used a 400 ps FWHM sensor jitter, and a LYSO rise and decay constants of 90 ps and 43 ns [Sei+12b], respectively.

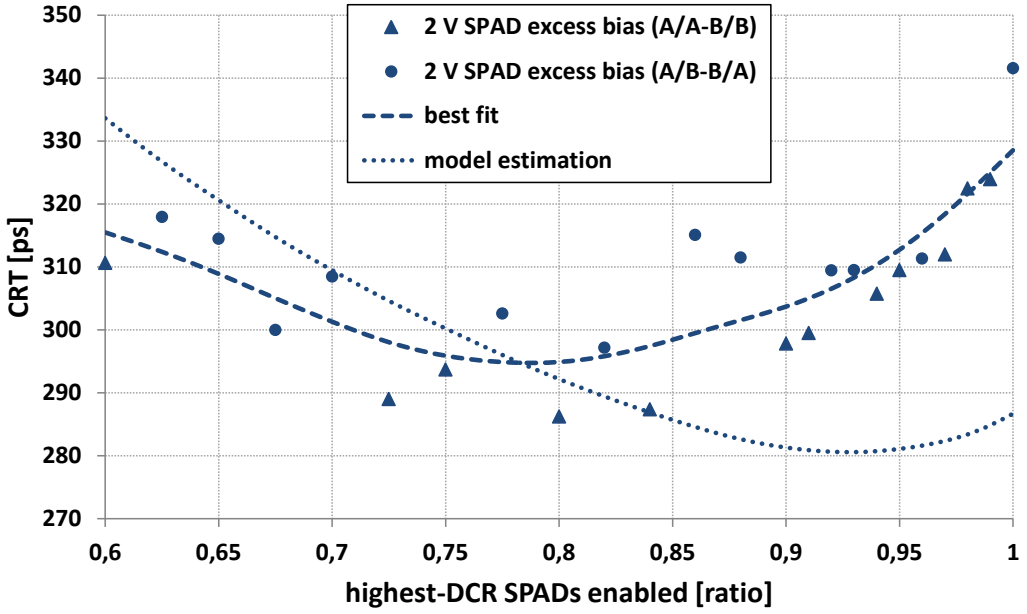


Figure 68: CRT versus number of SPADs enabled at 2 V SPAD excess bias, with highest DCR SPADs disabled first.

Considering all measured points – i.e. ignoring the A/B-B/A and A/A-B/B separation that will be explained in a moment –, the first thing to notice is that the points are much more scattered around the overall data shape than the energy resolution data. This has been traced back not to noise in the measurement, but to a bug in the acquisition firmware that results in one missed clock pulse in one of the sensors during the system start-up. Because of this, when one sensor has TDC A active in its pixels, the other sensor may have TDC B. This condition lasts for a full measurement configuration, and occurs randomly. As such, some data points in Figure 68 had this A/B or B/A TDC coincidence condition (triangular markers), while other points had an A/A or B/B condition (circular markers).

Unexpectedly, though, the A/B-B/A condition has a significantly worse performance than the A/A-B/B condition. Even if the full explanation for this behavior has not been determined yet, a possible hypothesis is that the PIPE signal has a non-ideal duty cycle (i.e. different from 50%), and thus the TDC code post-processing would need to take this into account. However, measuring this duty cycle difference is non-trivial, as the main TDC characteristic – its full scale value for a given clock frequency – depends both on the TDC speed and on the PIPE duty cycle. An alternative measure of this duty cycle can be obtained through the ratio of events that occur at TDCs A and B over the total events. This method, however, yielded only partially satisfactory results: while the data plotted in Figure 68 uses it, a visible performance degradation for the A/B-

B/A cases can still be noted. Not using it, though, resulted in an even greater performance discrepancy between the two conditions.

In any case, in an actual PET system, it may happen that avoiding A/B-B/A coincidences is impossible: due to the large distance between two coincident modules, there may be enough clock skew between them that a significant part of the events will necessarily be in the A/B-B/A condition. As such, the data presented in Figure 68 included both conditions, and moreover the best fit curve that showcases the overall data shape was obtained using all data points. This fit is probably a good approximation of the system performance when the clock skew between the PET scanner modules is in the order of half clock cycle. On the other hand, in a system where this skew can be kept close to zero, the performance curve will go through the A/A-B/B data points only, resulting in about 10 ps better performance.

Another important aspect that can be noted in Figure 68 is that the measured data does not follow the same shape as the model estimations. Two main hypothesis have been formulated to account for this mismatch: first, considering that the model estimations are based on the Cramér-Rao lower bound of an unbiased estimator, it is quite peculiar that the measured data could be below this limit, which occurs for several measured points below 80% SPADs enabled. Combining this with the hypothesis behind the high pixel jitter, that is, that the jitter is mostly due to skew between different SPADs, it becomes clear that the pixel jitter is also changing with the number of SPADs enabled, something that was not accounted for in the model. As such, the fixed jitter value in the model, which was measured with all SPADs enabled, results in an overestimated Cramer-Rao bound when less and less SPADs are enabled. This hypothesis also explains the difference in shape between data and model, since modeling the simultaneous variation in PDE, DCR and jitter could result in the optimum operation point near the 80%, as the data behaves.

The second hypothesis that also contributes to the difference between data and model in Figure 68 is that the Cramér-Rao-based model assumes an unbiased estimator, which is not the case of the inverse CDF-based estimator, as shown in Chapter 5. This could not only explain part of the mismatch in the shape of the two curves, but it also explains the difference between them at 100% SPADs enabled. In this configuration, the modeled pixel jitter does indeed match the one of the actual system, but still the measured performance is about 15% worse than the CRB. This is almost double than what was shown in the simulations in section 5.2.3, which perhaps could be further explained by non-modeled effects such as optical crosstalk.

The verification of these two hypotheses is possible, through measurements of the pixel jitter for the various SPAD enabled ratios, and through the implementation of a maximum-likelihood estimator to calculate the CRT. These two assignments are left as future work in the SPADnet project.

6.2.3 Suggested sensor configuration results

Given the performance optimization results above, the following configuration was chosen as a standard for future measurements in a PET system: 84% of the SPADs enabled (ordered from the highest DCR one), 2 V SPAD excess bias and 150 ns integration time. This parameter setup aims at balancing the best CRT with the best energy resolution, and were chosen based also on the available measurement points. Small variations in their values should not strongly affect the performance, and may actually increase it. Moreover, the optimum performance point may change depending on different system parameters, such as the scintillator type and geometry. As such, this sensor configuration and the ensuing performance numbers should be seen only as a demonstration of the SPADnet-I capabilities, and not as a guarantee of performance in any situation.

The energy spectrum measured with the ^{22}Na source is shown in Figure 69, highlighting a 10.8% energy resolution. Besides the 511 keV peak, the spectrum also clearly illustrates the low-energy events due to gamma scattering, below 400 keV, then, after the photo peak, a second set of events that are mostly due to 1275 keV gamma scattering and, finally, the 1275 keV peak.

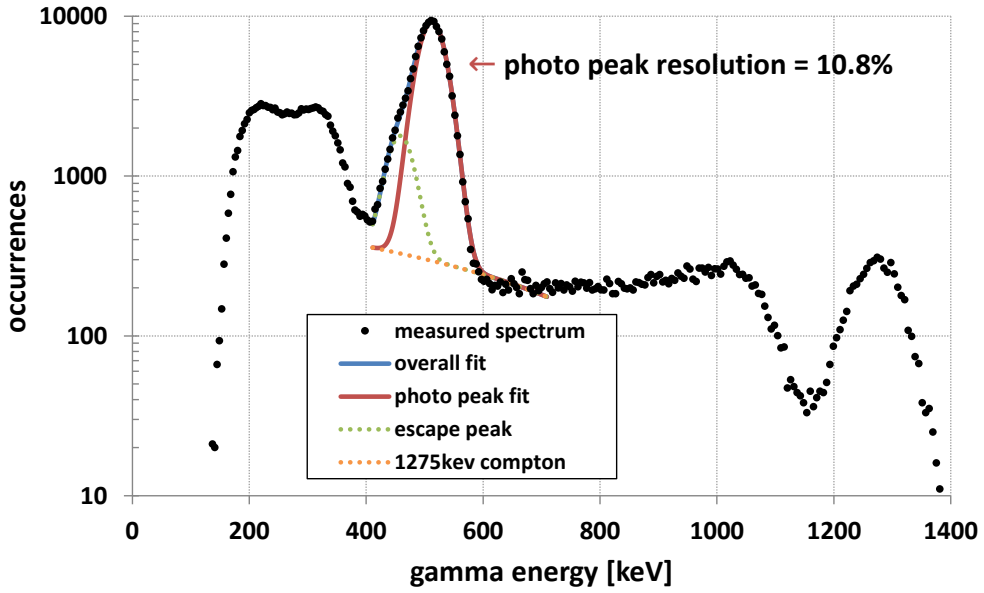


Figure 69: energy spectrum obtained with a ^{22}Na source, highlighting a 10.8% energy resolution at the 511 keV photo peak.

As previously mentioned, the energy resolution is defined as the ratio between the FWHM and the mean of the 511 keV photo peak. However, the actual measured peak has two further components that must be taken into account before calculating the sensor resolution. One is the pile-

up and scattered events that are spread all over the spectrum, while the other is the so-called escape peak. The escape peak is due to events where the 54 keV X-ray that is generated during the photoelectric interaction in LYSO escapes the crystal, thus reducing the detected energy to 457 keV. Due to these factors, the measured peak is fitted with a sum of three functions: a Gaussian for the 511 keV photo peak, another Gaussian for the 457 keV escape peak and a first degree polynomial for the scattered events [Szc+13]. Finally, the energy resolution is obtained taking into consideration the FWHM and mean of the fitted photo peak only. This procedure was applied to all measured configurations to obtain the energy resolution values plotted in the performance optimization section.

For the gamma coincidence resolution time (CRT) calculation, the best first photon finder algorithm parameters were found to be two consecutive intervals below a threshold of 340 ps. These relatively relaxed values are made possible due to the low DCR of the selected sensor configuration, which results in an average number of dark counts per time bin below 1.5. Figure 70 shows the obtained CRT for the two ToA estimation algorithms described in Chapter 5. The single-timestamp estimator is plotted against its timestamp order, while the x-scale for the inverse CDF-based, multiple-timestamp estimator corresponds to the number of timestamps used in the estimation, always taken starting with the first.

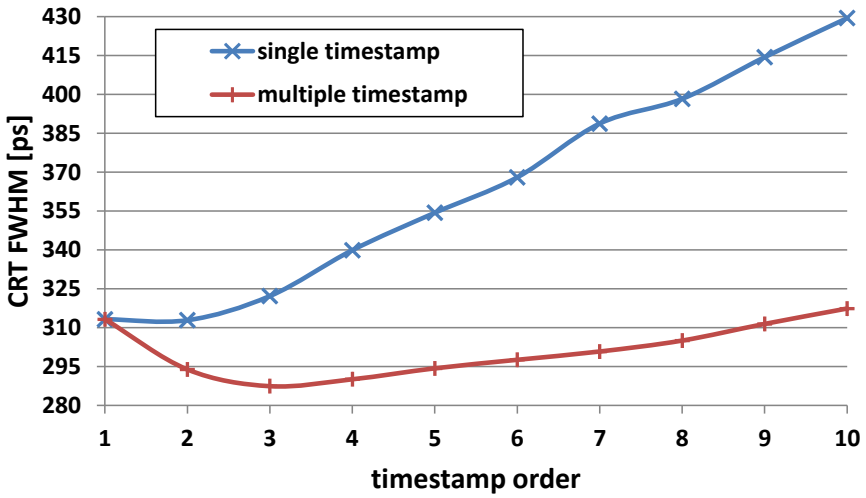


Figure 70: CRT obtained with the single- and multiple-timestamp estimators using different photon orders/different number of timestamps.

The results are in good agreement with the simulations presented in section 5.2.3, with the behavior of both estimators versus the timestamp order as expected. The best single-timestamp estimator occurs at the 2nd timestamp, for which its CRT histogram is shown in Figure 71(a), high-

lighting the full width at half-maximum (FWHM) value of 313 ps. On the other hand, the best inverse CDF-based estimator combines the first three timestamps to obtain a CRT of 288 ps, which represents an improvement of about 8% with respect to the single timestamp estimator. The CRT histogram for this estimator is shown in Figure 71(b).

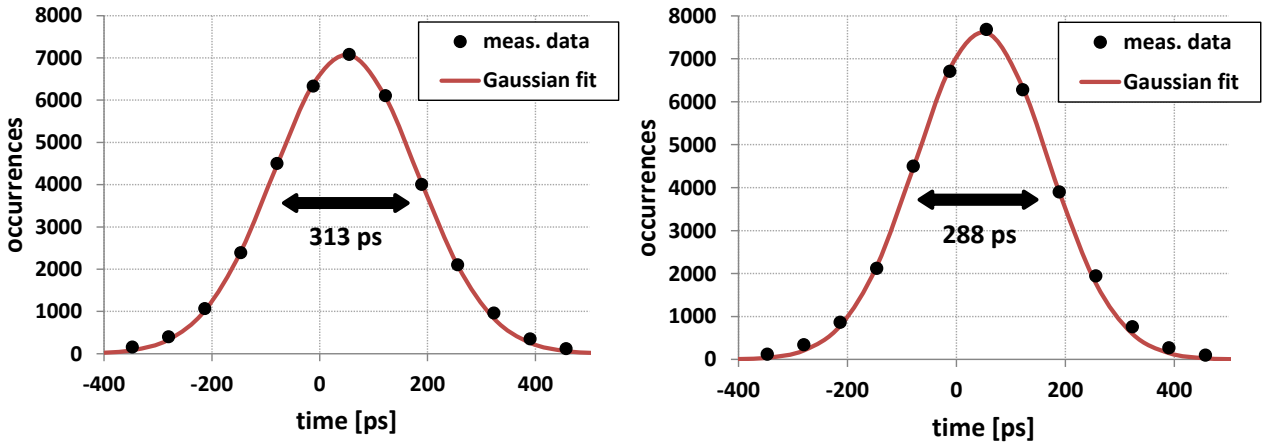


Figure 71: CRT histogram for the best single- (a) and multiple-timestamps (b) estimators.

As a final note, it can be observed in Figure 71 that both histograms are not centred at zero. This is a result of the clock distribution between the two boards, which generates a skew of about 50 ps, and which was not corrected for in post-processing.

Chapter 7 Conclusions

The modelling, design and characterization of a multi-pixel, fully digital SiPM for PET applications have been described. Through modelling, it was shown that the targeted PET application requires a sensor capable of measuring the incoming photon flux with maximum sensitivity and reasonably low temporal noise (i.e. DCR). Moreover, the sensor jitter needs to be strongly minimized when considering a ToF-PET scanner.

To meet these requirements, a new top-to-bottom architecture using CMOS technology was proposed. At the bottom, the architecture is composed of SPAD-based mini-SiPMs with spatial and temporal compression that provide an area- and readout-efficient detection cell. The top-level, on the other hand, deals with the detection of gamma events through a distributed adder, which sums the sensor total counts in real time. These counts are discretised in time so as to provide an estimation of the flux, which is then used by the discriminator to detect gamma events. Finally, the architecture also features per-pixel TDCs, which enabled the development of multiple-timestamp algorithms for the gamma ToA estimation.

The suggested architecture was implemented in the SPADnet-I sensor, which has its main characteristics and performance data summarized in Table 3. The sensor contains 92k SPADs, arranged in 8×16 pixels, each of size around 0.6×0.6 mm^2, with a resulting FF of 42.6%. The in-pixel 12-bit TDCs have an average resolution of 65.9 ps, and a DNL below 0.25 LSB after removing the 3-4 initial and last codes. The SPADs, at their maximum excess bias voltage, have a timing resolution of 142 ps and a DCR median of 1.42 kHz, which grows to an average of 23.27 kHz due to the circa 60% yield. Finally, the mini-SiPM monostables were tested to work with a pulse width down to 250 ps, validating the mini-SiPM compression concepts.

The complete system was first measured for its jitter through single photon timestamping of a pulsed laser, which resulted in an average jitter of 395 ps across all pixels. This value is significantly higher than expected from the SPAD-only jitter, and it has been traced back to process parameter variations that affect the SPAD front-end, resulting in timing skew between the SPADs in each pixel, something that cannot be corrected for externally. The implemented adder tree was verified to provide a real-time output of the detected energy in the array at up to targeted rate of 100 Msamples/s.

Measuring the sensor in a PET-like setup demonstrated the operation of the discriminator, which greatly improves the sensor efficiency while at the same time allowing the correct validation of events below 40 keV. Moreover, the total compression loss was shown to be in agreement with

the design, at 10%, of which 4% are due to spatial compression, 2% due to temporal compression and the 4% remaining due to the SPAD dead time.

Table 3: summary of SPADnet-I performance and main characteristics.

Sensor			
Process technology	CMOS 1P4M 0.13μm Imaging	**Maximum clock frequency**	100MHz
Array Size	8×16 pixels	**Output data rate**	1.6Gb/s
In-pixel photodetectors	4 mini-SiPMs of 180 16.27μm SPADs each	**Chip size**	9.85×5.452 mm^2
Array fill factor	42.6%	**Supply voltage**	3.3V for digital 1.2V for core
Pixel pitch	610.5×571.2μm^2	**Chip power consumption**	200 mW (dark) 300 mW (light)
TDC			
TDC range (12b)	269.98 ns	**TDC DNL (central range)**	-0.17 … +0.25 LSB
TDC resolution (1 LSB)	65.9 ps	**TDCs spatial uniformity** (σ across a chip)	0.88 ps
TDC INL (central range)	-0.8 … +0.7 LSB	**TDC current consumption** (for a single running TDC)	0.79 mA
SPAD (@ 3 V excess bias)			
Jitter (incl. 70 ps laser)	142 ps	**Peak PDP**	45%
Median DCR	1.42 kHz	**Average DCR**	23.27 kHz
System jitter (@ 3 V SPAD excess bias)			
Single photon, average of all pixels jitter		395.4 ps (σ = 48.9 ps)	
Gamma detection performance (using 3×3×5 mm^3 LYSO scintillator, @ 20^o C)			
511 keV peak energy resolution		10.8%	
Compression loss with 511 keV gammas		~ 10%	
Best CRT using multiple timestamps		288 ps	

A measurement campaign was performed with the goal of finding the best performance configuration of the sensor. The obtained data was also used to validate the detector models developed in Chapter 3. The energy resolution model resulted in an excellent match to the measured data, while the CRT model had some non-negligible mismatches. These mismatches were hypothesized to be mostly due to the pixel jitter changing with the number of SPADs enabled, something which was not accounted for in the model.

The best sensor configuration was found to be at 84% of the SPADs enabled (disabled starting with the highest DCR one), with 2 V SPAD excess bias and 150 ns integration time. This configuration results in an energy resolution of 10.8%, which is in line with the state-of-the-art SiPMs for PET described in Chapter 2. The best CRT value of this configuration, on the other hand, is significantly worse than the state-of-the-art at 288 ps – obtained with the developed multiple-timestamp ToA estimator combining 3 timestamps –, which can be attributed to the higher than expected pixel jitter. It should be noted, however, that due to the fully digital, integrated nature of SPADnet-I, the conditions utilized in its measurements are very likely to be replicable in a real PET scanner, something that is usually not the case with other systems. For instance, in publications investigating typical analog SiPMs performance, it is usual to acquire the full output waveform of the SiPM with a fast oscilloscope and then post-process it to obtain the energy and ToA values, which allows for improved performance, but obviously cannot be replicated in a real PET system.

7.1. Future work

As mentioned in Chapter 6, a few interesting measurements are still left to be performed with SPADnet-I. For instance, the system jitter could be measured for different number of SPADs enabled, following the same disabling order as in the gamma measurements, so that the resulting data could be fed to the CRT model. This could then validate the CRT model with respect to the measured gamma data. Even more interestingly, perhaps, would be to measure each pixel jitter with only one SPAD enabled, for all SPADs. With this data, the disabling order of the SPADs in the gamma measurements could be set to the ones having its skew farther away from the average, which would result in a faster decrease of the pixel jitter than using the highest DCR order, which likely has SPAD skews ordered randomly. It would then be interesting to see if the gamma CRT results would be significantly improved.

Another way to improve the performance of the SPADnet-I sensor would be to develop further improvements in the data processing algorithms. Regarding the energy resolution, one way of doing this would be to perform a per-pixel decompression, given that the photon density between pixels was shown to be very non-uniform. However, to do this, a measurement would need to be performed where the impinging energy on each pixel was known, and where furthermore the light pulse shape was the same as in real gamma measurement, so that the temporal compression was the same. This, unfortunately, is not straightforward. Regarding the CRT, a relatively simple way to improve performance would be to try a MLE for the gamma ToA, which, however, could be difficult to implement at the FPGA level for real-time PET systems.

Moreover, given the multi-pixel and large area of the SPADnet-I sensor, it is also possible to perform gamma measurements with crystal matrices, aiming at evaluating the achievable spatial resolution as well as the CRT and energy resolution uniformity. To this aim, preliminary results were obtained using a 4×4 LYSO matrix with 1.12×1.12×13 mm^3 needles. The measurements were performed using a single sensor and only the intrinsic, low-energy radioactivity of LYSO, with the resulting crystal flood map shown in Figure 72. The algorithm utilized for position estimation was a simple center-of-gravity algorithm with a single threshold to discard pixels with dark counts only. As can be seen in the figure, the individual crystals can be easily distinguished, and in fact the crystal separation indicates that much smaller needles can be used, if an application requires it. The low-rate background in the map is due to pile-up events, which were not filtered out.

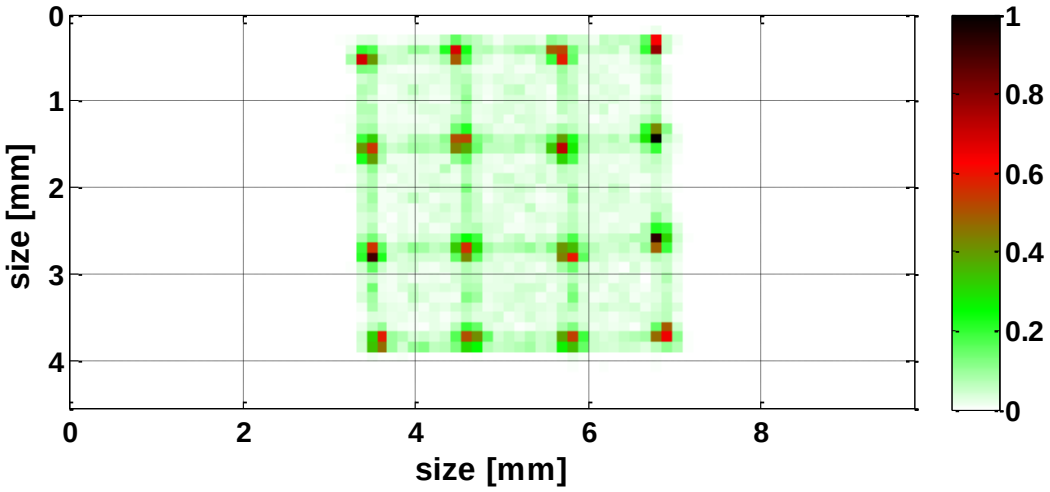


Figure 72: Preliminary flood map of a 4×4 LYSO matrix with 1.12×1.12×13 mm^3 needles. No particular pixel to crystal alignment.

Finally, one can also use the obtained results of this thesis to design an improved version of the sensor. The first are foremost change should be improved robustness to process variations at the SPAD front-end, with the aim of reducing the timing skew between SPADs in the same pixel and thus improving the pixel jitter. Another relatively simple change should be an improvement on the TDC enabling window definition, so as to improve the DNL of the initial and final codes of the TDC. This could be done, for instance, by purposively anticipating the beginning of the TDC window and delaying its end, while keeping the pulse division at the mini-SiPM the same. This would create an offset on the first TDC code – which can be easily removed in post-processing –, while strongly improving the DNL performance.

Another interesting analysis towards future sensor development can be done by using the validated sensor model to predict the performance of different design improvements. To this aim, the parameters that go into the sensor model are scaled in a way that any change in PDE comes from

changing the number of SPADs in the sensor, but not the SPAD dimensions. Two cases of possible improvements in the sensor will be considered: one, removing the SRAMs, and thus increasing the sensor FF to about 50%, and two, reducing the average SPAD DCR. The energy resolution results for both cases, their combination and also the current SPADnet-I characteristics, are shown in Figure 73.

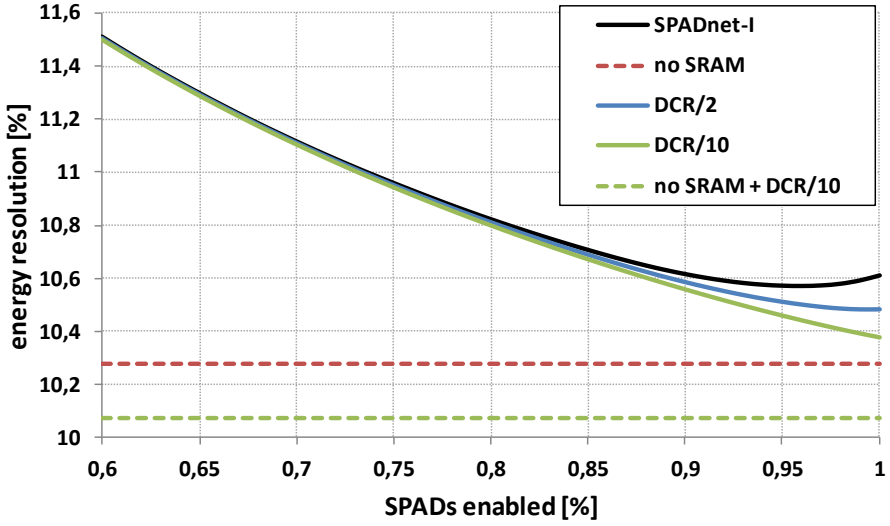


Figure 73: Energy resolution prediction with different sensor characteristics.

A few interesting observations can be made from the results above. First and foremost, including per-SPAD SRAMs is actually disadvantageous in terms of energy resolution, since the loss in FF is not compensated by the optimum operating point. This, of course, depends on how much FF was lost due to SRAMs, but even with a FF value as low as 45% the sensor energy resolution is already better than the current optimum operating point. Regarding DCR, reducing it produces diminishing returns: 0.1% is gained by halving the DCR, but then again only 0.1% is gained by further dividing the DCR by 5. Moreover, as soon as the DCR/2 level, the optimum operating point is shifted to all SPADs enabled, meaning that the SRAMs provide no improvement at all.

To understand how these changes would affect the sensor timing resolution, the CRT model with all SPADs enabled will be assumed as a reasonable estimate of the sensor CRT, even if its correspondence with the measured data in Chapter 6 was not ideal. The two design cases with no SRAMs in Figure 73, along with the SPADnet-I parameters, were swept for different pixel jitter FWHM values, starting from the SPADnet-I value of around 400 ps, down to 100 ps. The results are shown in Figure 74.

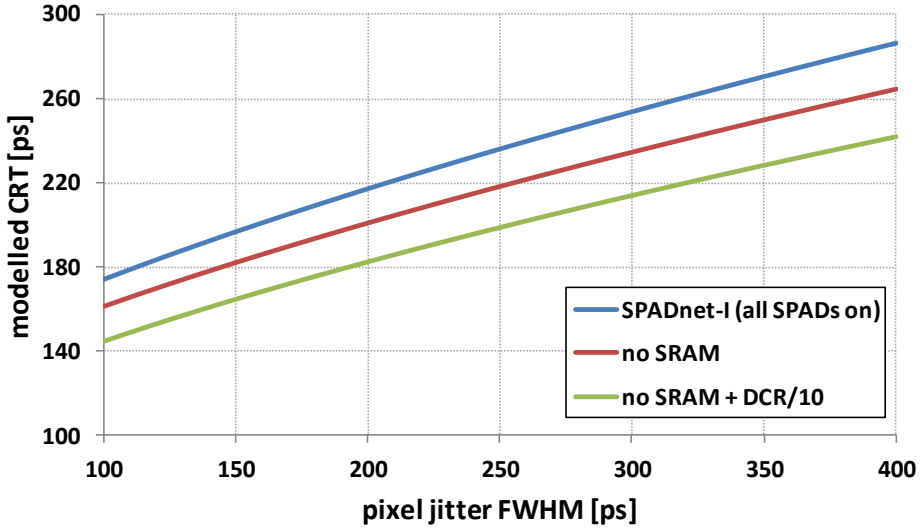


Figure 74: CRT prediction with different sensor characteristics.

As can be seen in the graph, to achieve a CRT similar to that of the state-of-the-art – that is, below 180 ps –, the SPADnet-I sensor would require a pixel jitter FWHM in the order of 100 ps. Given the SPAD-only jitter of circa 140 ps, it is clear that this goal is too ambitious, as it would require design improvements that not only completely remove the SPAD skew, but also significantly improve the SPAD jitter itself, possibly through re-engineering of the device. The supposition of removing the SRAMs and thus improving the FF to 50% would result in the same 180 ps CRT being achieved with a pixel jitter of around 150 ps, meaning that basically the SPAD skew needs to be eliminated, but the SPAD does not need to be re-engineered. Finally, a sensor without SRAMs and with a DCR improved tenfold would further lessen the pixel jitter requirement to around 200 ps, meaning that some SPAD skew would still be allowed.

As such, this final configuration of no SRAMs (and thus 50% FF), a tenfold reduction in DCR through fabrication process improvements and, finally, layout improvements resulting in a pixel jitter of 200 ps, is the suggested design goal for a future implementation of the SPADnet architecture. Such an implementation could keep the high efficiency and ease of use of the SPADnet-I sensor while very likely achieving better than state-of-the-art performance in all PET figures of merit.

Bibliography

[Arn+92] B. C. Arnold, N. Balakrishnan, and H. N. Nagaraja, "A First Course in Order Statistics," 1st ed. Wiley-Interscience, 1992.

[Ber+08] P. Berard, M. Bergeron, C.M. Pepin, J. Cadorette, M. Tetrault, N. Viscogliosi, R. Fontaine, H. Dautet, M. Davies, P. Deschamps, and R. Lecomte, "LabPET II, a novel 64-channel APD-based PET detector module with individual pixel readout achieving submillimetric spatial resolution," Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2008 IEEE, pp. 5457-5462, 2008.

[Bér+12] B.-L. Bérubé, V.-P. Rhéaume, A. C. Therrien, S. Parent, L. Maurais, A. Boisvert, G. Carini, S. A. Charlebois, R. Fontaine, J.-F. Pratte, "Development of a Single Photon Avalanche Diode (SPAD) Array in High Voltage CMOS 0.8 µm dedicated to a 3D Integrated Circuit (3DIC)," IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2012.

[Bra+11] L. H. C. Braga, L. Pancheri, L. Gasparini, M. Perenzoni, R. Walker, R. K. Henderson, and D. Stoppa, "A CMOS mini-SiPM detector with in-pixel data compression for PET applications," in Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2011 IEEE, pp. 548 –552, 2011.

[Bra+12] L. H. C. Braga, L. Gasparini, and D. Stoppa, "A time of arrival estimator based on multiple timestamps for digital PET detectors," in 2012 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), pp. 1250–1252, 2012.

[Bra+14] L. H. C. Braga, L. Gasparini, L. Grant, R. K. Henderson, N. Massari, M. Perenzoni, D. Stoppa, and R. Walker, "A Fully Digital 8x16 SiPM Array for PET Applications With Per-Pixel TDCs and Real-Time Energy Output," *IEEE Journal of Solid-State Circuits*, vol. 49, no. 1, pp. 301–314, Jan. 2014.

[Bru+14] C. Bruschini, E. Charbon, C. Veerappan, L. H. C. Braga, N. Massari, M. Perenzoni, L. Gasparini, D. Stoppa, R. Walker, A. Erdogan, R. K. Henderson, S. East, L. Grant, B. Jatekos, F. Ujhelyi, G. Erdei, E. Lörincz, L. André, L. Maingault, V. Reboud, L. Verger, E. Gros d'Aillon, P. Major, Z. Papp, and G. Németh, "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, Jan. 2014.

[Cov+96] S. Cova, M. Ghioni, A. Lacaita, C. Samori, and F. Zappa, "Avalanche photodiodes and quenching circuits for single-photon detection," Applied Optics, vol. 35, no. 12, pp. 1956-1976, 1996.

[Dam+13] H. T. van Dam, G. Borghi, S. Seifert, and D. R. Schaart, "Sub-200 ps CRT in monolithic scintillator PET detectors using digital SiPM arrays and maximum likelihood interaction time estimation," Phys. Med. Biol., vol. 58, no. 10, p. 3243, May 2013.

[DeG86] M. H. DeGroot, "Probability and Statistics," Addison Wesley Publishing Company, 1986.

[Del+09] A. Del Guerra, N. Belcari, M.G. Bisogni, G. Llosá, S. Marcatili, and S. Moehrs, "Advances in position-sensitive photodetectors for PET applications," Nucl. Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, vol. 604, pp. 319-322, 2009.

[Del+10] E. P. Delfino, S. Majewski, R. R. Raylman, and A. Stolin, "Towards 1mm PET resolution using DOI modules based on dual-sided SiPM readout," in 2010 IEEE Nuclear Science Symposium Conference Record (NSS/MIC), 2010, pp. 3442–3449.

[Doe+84] J. Doernberg, H.-S. Lee, and D. Hodges, "Full-speed testing of A/D converters," IEEE Journal of Solid-State Circuits, vol. 19, no. 6, pp. 820–827, 1984.

[ElG+05] A. El Gamal and H. Eltoukhy, "CMOS image sensors," IEEE Circuits and Devices Magazine, vol. 21, no. 3, pp. 6–20, May 2005.

[Exc14] Excelitas Photon Detection Solutions Catalog, http://www.excelitas.com/Downloads/CAT_PhotonDetection.pdf, retrieved Feb. 2014.

[Ezy10] EzyFit toolbox for Matlab version 2.40, http://www.fast.u-psud.fr/ezyfit/, retrieved July 2010.

[Far+08] N. Faramarzpour, M. J. Deen, S. Shirani, and Q. Fang, “Fully Integrated Single Photon Avalanche Diode Detector in Standard CMOS 0.18-um Technology,” IEEE Transactions on Electron Devices, vol. 55, no. 3, pp. 760–767, Mar. 2008.

[Fis+10] M. W. 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, 2010.

[Fra+09] T. Frach, G. Prescher, C. Degenhardt, R. de Gruyter, A. Schmitz, R. Ballizany, “The digital silicon photomultiplier — Principle of operation and intrinsic detector performance,” IEEE Nucl. Sci. Symp. Conf. Rec. (NSS/MIC), pp. 1959–1965, 2009.

[Frü06] S. Frühwirth-Schnatter, “Finite Mixture and Markov Switching Models,” Springer, 2006.

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

[Gol+13] A. Gola, C. Piemonte, A. Tarolli, “Analog Circuit for Timing Measurements With Large Area SiPMs Coupled to LYSO Crystals,” IEEE Transactions on Nuclear Science, vol. 60 (2), pp. 1296–1302, 2013.

[Gol+14] A. Gola, A. Ferri, C. Piemonte, A. Tarolli, N. Zorzi, “SiPM Cross-talk Amplification due to Scintillator Crystal: Effects on Timing Performance,” Physics in Medicine and Biology, 2014 (accepted for publication).

[Gul+05] A. Gulinatti, I. Rech, P. Maccagnani, M. Ghioni, and S. Cova, “Large-area avalanche diodes for picosecond time-correlated photon counting,” in Solid-State Device Research Conference, 2005. ESSDERC 2005. Proceedings of 35th European, 2005, pp. 355–358.

[Hae+12] Y. Haemisch, T. Frach, C. Degenhardt, 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.

[Hym65] L. G. Hyman, “Time Resolution of Photomultiplier Systems,” Review of Scientific Instruments, vol. 36, no. 2, pp. 193–196, 1965.

[Jun+07] J. H. Jung, Y. Choi, Y. H. Chung, O. Devroede, M. Krieguer, P. Bruyndonckx, and S. Tavernier, “Optimization of LSO/LuYAP phoswich detector for small animal PET,” Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, vol. 571, no. 3, pp. 669–675, Feb. 2007.

[Kar+10] M. A. Karami, M. Gersbach, H.-J. Yoon, and E. Charbon, “A new single-photon avalanche diode in 90nm standard CMOS technology,” Opt. Express, vol. 18, no. 21, pp. 22158–22166, Oct. 2010.

[Ket14] KETEK SiPM product portfolio, http://www.ketek.net/products/sipm/, retrieved Feb. 2014.

[Ler+05] C. W. Lerche, J. M. Benlloch, F. Sanchez, N. Pavon, B. Escat, E. N. Gimenez, M. Fernandez, I. Torres, M. Gimenez, A. Sebastia, and J. Martinez, “Depth of gamma-ray interaction within continuous crystals from the width of its scintillation light-distribution,” IEEE Transactions on Nuclear Science, vol. 52, no. 3, pp. 560–572, Jun. 2005.

[Lev+07] C. S. Levin and H. Zaidi, “Current Trends in Preclinical PET System Design,” PET Clinics, vol. 2, no. 2, pp. 125–160, Apr. 2007.

[Lew08] T.K. Lewellen, “Recent developments in PET detector technology,” Physics in Medicine and Biology, vol. 53, pp. R287-R317, 2008.

[Lig+86] A.W. Lightstone, R.J. McIntyre, R. Lecomte, and D. Schmitt, “A Bismuth Germanate-Avalanche Photodiode Module Designed for Use in High Resolution Positron Emission Tomography,” Nuclear Science, IEEE Transactions on, vol. 33, pp. 456-459, 1986.

[Lin+08] T. Ling, T. H. Burnett, T. K. Lewellen, and R. S. Miyaoka, “Parametric positioning of a continuous crystal PET detector with depth of interaction decoding,” Phys. Med. Biol., vol. 53, no. 7, pp. 1843–1863, Apr. 2008.

[Liu+01] H. Liu, T. Omura, M. Watanabe, and T. Yamashita, “Development of a depth of interaction detector for ?-rays,” Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, vol. 459, no. 1–2, pp. 182–190, Feb. 2001.

[Llo+10] G. Llosá, J. Barrio, C. Lacasta, M. G. Bisogni, A. D. Guerra, S. Marcatili, P. Barrillon, S. Bondil-Blin, C. de la Taille, and C. Piemonte, “Characterization of a PET detector head based on continuous LYSO crystals and monolithic, 64-pixel silicon photomultiplier matrices,” Phys. Med. Biol., vol. 55, no. 23, pp. 7299–7315, Dec. 2010.

[Man+12] S. Mandai, E. Charbon, “Multi-Channel Digital SiPMs: Concept, Analysis and Implementation,” IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2012.

[Man+13] S. Mandai and E. Charbon, “A 4 x 4 x 416 digital SiPM array with 192 TDCs for multiple high-resolution timestamp acquisition,” J. Inst., vol. 8, no. 05, p. P05024, May 2013.

[Med10] Mediso Ltd. (http://www.mediso.com/), internal discussions with, October 2010.

[Mos07] W. W. Moses, “Recent advances and future advances in time-of-flight PET,” Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, vol. 580 (2), pp. 919–924, 2007.

[Nag+13] K. Nagy, M. Tóth, P. Major, G. Patay, G. Egri, J. Häggkvist, A. Varrone, L. Farde, C. Halldin, and B. Gulyás, “Performance Evaluation of the Small-Animal nanoScan PET/MRI System,” J Nucl Med, p. jnumed.112.119065, Oct. 2013.

[Nas+07] A. Nassalski, M. Kapusta, T. Batsch, D. Wolski, D. Mockel, W. Enghardt, and M. Moszynski, “Comparative Study of Scintillators for PET/CT Detectors,” Nuclear Science, IEEE Transactions on, vol. 54, pp. 3-10, 2007.

[Nic+10] C. Niclass and M. Soga, “A miniature actively recharged single-photon detector free of afterpulsing effects with 6ns dead time in a 0.18 um CMOS technology,” in Electron Devices Meeting (IEDM), 2010 IEEE International, 2010, pp. 14.3.1–14.3.4.

[Ott+04] A. N. Otte, J. Barral, B. Dolgoshein, J. Hose, S. Klemin, E. Lorenz, R. Mirzoyan, E. Popova, and M. Teshima, “New results from a test of silicon photomultiplier as readout for PET,” Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2004 IEEE, pp. 3738–3742, 2004.

[Pan+07] L. Pancheri and D. Stoppa, “Low-Noise CMOS single-photon avalanche diodes with 32 ns dead time,” in Solid State Device Research Conference, 2007. ESSDERC 2007. 37th European, 2007, pp. 362–365.

[Pan+09] L. Pancheri and D. Stoppa, “A SPAD-based pixel linear array for high-speed time-gated fluorescence lifetime imaging,” in ESSCIRC, 2009. ESSCIRC ’09. Proceedings of, 2009, pp. 428–431.

[Pan+11] L. Pancheri, N. Massari, F. Borghetti, D. Stoppa, “A 32x32 SPAD Pixel Array with Nanosecond Gating and Analog Readout,” presented at the 2011 International Image Sensor Workshop, Hokkaido, Japan, 2011.

[Pel+97] R. Pelka, J. Kalisz, and R. Szplet, “Nonlinearity correction of the integrated time-to-digital converter with direct coding,” IEEE Transactions on Instrumentation and Measurement, vol. 46, no. 2, pp. 449–453, Apr. 1997.

[Phe06] M.E. Phelps, “PET: Physics, Instrumentation, and Scanners,” Springer, 2006.

[Phy11] Physical properties of Common Inorganic Scintillators, Saint-Gobain Crystals, http://www.detectors.saint-gobain.com (retrieved May 2011).

[Pic+08] B. J. Pichler, M. S. Judenhofer, C. Pfannenberg, “Multimodal Imaging Approaches: PET/CT and PET/MRI”, Molecular Imaging I, vol. 185/1, pp. 109-132, 2008.

[Pie+13] C. Piemonte, A. Ferri, A. Gola, T. Pro, N. Serra, A. Tarolli, and N. Zorzi, “Characterization of the First FBK High-Density Cell Silicon Photomultiplier Technology,” IEEE Transactions on Electron Devices, vol. 60, no. 8, pp. 2567–2573, Aug. 2013.

[Pre08] PreLude420 Scintillation Material datasheet, Saint-Gobain Crystals, 2008, http://www.detectors.saint-gobain.com (retrieved May 2011).

[Ren06] D. Renker, “Geiger-mode avalanche photodiodes, history, properties and problems”, Nuclear Instruments and Methods in Physics Research A, vol. 567, pp. 48–56, 2006.

[Ric+09a] J. A. Richardson, L. A. Grant, and R. K. Henderson, “Low Dark Count Single-Photon Avalanche Diode Structure Compatible With Standard Nanometer Scale CMOS Technology,” Photonics Technology Letters, IEEE, vol. 21, no. 14, pp. 1020–1022, 2009.

[Ric+09b] J. Richardson, R. Walker, L. Grant, D. Stoppa, F. Borghetti, E. Charbon, M. Gersbach, R. K. Henderson, "A 32×32 50ps Resolution 10 bit Time to Digital Converter Array in 130nm CMOS for Time Correlated Imaging," IEEE Custom Integrated Circuits Conference (CICC), pp. 77–80, 2009.

[Ric+11] J. A. Richardson, E. A. G. Webster, L. A. Grant, and R. K. Henderson, “Scaleable Single-Photon Avalanche Diode Structures in Nanometer CMOS Technology,” Electron Devices, IEEE Transactions on, vol. 58, no. 7, pp. 2028 –2035, Jul. 2011.

[Sao+99] A. Saoudi, C. M. Pepin, F. Dion, M. Bentourkia, R. Lecomte, M. Andreaco, M. Casey, R. Nutt, and H. Dautet, “Investigation of depth-of-interaction by pulse shape discrimination in multicrystal detectors read out by avalanche photodiodes,” Nuclear Science, IEEE Transactions on, vol. 46, no. 3, pp. 462–467, 1999.

[Sch+09] D. R. Schaart, H. T. van Dam, S. Seifert, R. Vinke, P. Dendooven, H. Löhner, and F. J. Beekman, “A novel, SiPM-array-based, monolithic scintillator detector for PET,” Phys. Med. Biol., vol. 54, no. 11, pp. 3501–3512, Jun. 2009.

[Szc+13] T. Szczęśniak, M. Kapusta, M. Moszyński, M. Grodzicka, M. Szawłowski, D. Wolski, J. Baszak, N. Zhang, “MPPC Arrays in PET Detectors With LSO and BGO Scintillators,” IEEE Trans. Nucl. Sci., vol. PP (99), 2013.

[Ser+13] N. Serra, A. Ferri, A. Gola, T. Pro, A. Tarolli, N. Zorzi, C. Piemonte, “Characterization of new FBK SiPM technology for visible light detection,” Journal of Instrumentation, vol. 8, no. 03, pp. 3019, 2013.

[Sei+12a] S. Seifert, H. T. van Dam, R. Vinke, P. Dendooven, H. Lohner, F. J. Beekman, D. R. Schaart, “A Comprehensive Model to Predict the Timing Resolution of SiPM-Based Scintillation Detectors: Theory and Experimental Validation,” IEEE Transactions on Nuclear Science, vol. 59 (1), pp. 190 –204, 2012.

[Sei+12b] S. Seifert, H. T. van Dam, and D. R. Schaart, “The lower bound on the timing resolution of scintillation detectors,” Physics in Medicine and Biology, vol. 57, no. 7, pp. 1797–1814, Apr. 2012.

[Son+10] T. Y. Song, H. Wu, S. Komarov, S. B. Siegel, and Y.-C. Tai, “A sub-millimeter resolution PET detector module using a multi-pixel photon counter array,” Phys. Med. Biol., vol. 55, no. 9, p. 2573, May 2010.

[Spa+10] V. C. Spanoudaki and C. S. Levin, “Photo-Detectors for Time of Flight Positron Emission Tomography (ToF-PET),” Sensors, vol. 10, no. 11, pp. 10484-10505, 2010.

[Sto+09a] D. Stoppa, D. Mosconi, L. Pancheri, and L. Gonzo, “Single-Photon Avalanche Diode CMOS Sensor for Time-Resolved Fluorescence Measurements,” Sensors Journal, IEEE, vol. 9, no. 9, pp. 1084–1090, 2009.

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

[Szc+13] T. Szczęśniak, M. Grodzicka, M. Moszyński, M. Szawłowski, D. Wolski, J. Baszak, “Characteristics of scintillation detectors based on inorganic scintillators and SiPM light readout,” Nucl. Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, vol. 702, pp. 91-93, 2013.

[Tyn+12] D. Tyndall, B. R. Rae, D. D. Li, J. Arlt, A. Johnston, J. A. Richardson, R. K. Henderson, "A High-Throughput Time-Resolved Mini-Silicon Photomultiplier With Embedded Fluorescence Lifetime Estimation in 0.13 µm CMOS," IEEE Transactions on Biomedical Circuits and Systems, vol.6 (6), pp. 562–570, 2012.

[Ven+13] E. Venialgo, S. Mandai and E. Charbon, “Time Mark Estimators for MD-SiPM and Impact of System Parameters,” IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2013.

[Wal+12] R. J. Walker, E. A. G. Webster, J. Li, N. Massari, R. K. Henderson, “High Fill Factor Digital Silicon Photomultiplier Structures in 130nm CMOS Imaging Technology,” IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2012.

[Wer+04] M.N. Wernick and J.N. Aarsvold, “Emission Tomography: The Fundamentals of PET and SPECT,” Academic Press, 2004.

[Yeo+12] J. Y. Yeom, R. Vinke, N. Pavlov, S. Bellis, K. O’Neill, C. Jackson, C. S. Levin, “Performance of Fast Timing Silicon Photomultipliers for Scintillation Detectors,” IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2012.

Appendix A

Development of the PDF and CDF equations for the photon timestamp model

The PDF of the photon timestamp model, discussed in section 3.3.1, equation (7), is obtained as follows.

$$p_{pt}(t|\Theta) = \int_0^{t-\Theta} p_{em}(t-\tilde{t}|\Theta) \times p_{jitter}(\tilde{t})d\tilde{t}$$

First, the two underlying PDFs, which are both exponential functions, are multiplied.

$$p_{pt}(t|\Theta) = \int_0^{t-\Theta} \frac{e^{-\frac{t-\tilde{t}-\Theta}{\tau_d}} - e^{-\frac{t-\tilde{t}-\Theta}{\tau_r}}}{\tau_d - \tau_r} \times \frac{1}{\sigma_{trans} \cdot \sqrt{2\pi}} \cdot e^{-\frac{(\tilde{t}-\mu)^2}{2\cdot{\sigma_{trans}}^2}} \cdot d\tilde{t}$$

$$p_{pt}(t|\Theta) = \frac{1}{\tau_d - \tau_r} \cdot \frac{1}{\sigma_{trans} \cdot \sqrt{2\pi}} \cdot \int_0^{t-\Theta} \left[e^{-\frac{t-\tilde{t}-\Theta}{\tau_d}} - e^{-\frac{t-\tilde{t}-\Theta}{\tau_r}} \right] \cdot e^{-\frac{(\tilde{t}-\mu)^2}{2\cdot{\sigma_{trans}}^2}} \cdot d\tilde{t}$$

$$p_{pt}(t|\Theta) = k \cdot \int_0^{t-\Theta} \left[e^{-\frac{t-\tilde{t}-\Theta}{\tau_d} - \frac{(\tilde{t}-\mu)^2}{2\cdot{\sigma_{trans}}^2}} - e^{-\frac{t-\tilde{t}-\Theta}{\tau_r} - \frac{(\tilde{t}-\mu)^2}{2\cdot{\sigma_{trans}}^2}} \right] d\tilde{t}$$

$$p_{pt}(t|\Theta) = k \cdot \int_0^{t-\Theta} \left[e^{\frac{-\tilde{t}^2}{2\cdot{\sigma_{trans}}^2} + \tilde{t}\cdot\left(\frac{\mu}{{\sigma_{trans}}^2} + \frac{1}{\tau_d}\right) - \frac{\mu^2}{2\cdot{\sigma_{trans}}^2} - \frac{t-\Theta}{\tau_d}} - e^{\frac{-\tilde{t}^2}{2\cdot{\sigma_{trans}}^2} + \tilde{t}\cdot\left(\frac{\mu}{{\sigma_{trans}}^2} + \frac{1}{\tau_r}\right) - \frac{\mu^2}{2\cdot{\sigma_{trans}}^2} - \frac{t-\Theta}{\tau_r}} \right] d\tilde{t}$$

The integral of these two exponentials can be solved by making:

$$a = \frac{1}{2\cdot{\sigma_{trans}}^2};\ b_1 = \left(\frac{\mu}{{\sigma_{trans}}^2} + \frac{1}{\tau_d}\right);\ c_1 = -\frac{\mu^2}{2\cdot{\sigma_{trans}}^2} - \frac{t-\Theta}{\tau_d};$$

$$b_2 = \left(\frac{\mu}{{\sigma_{trans}}^2} + \frac{1}{\tau_r}\right);\ c_2 = -\frac{\mu^2}{2\cdot{\sigma_{trans}}^2} - \frac{t-\Theta}{\tau_r};$$

And then completing the squares as in:

$$-ax^2 + bx + c = -a\left(x - \frac{b}{2a}\right)^2 + \frac{b^2}{4a} + c$$

Resulting in:

$$p_{pt}(t|\Theta) = k \cdot \left[\int_0^{t-\Theta} e^{-a\tilde{t}^2 + b_1\tilde{t} + c_1} d\tilde{t} - \int_0^{t-\Theta} e^{-a\tilde{t}^2 + b_2\tilde{t} + c_2} d\tilde{t} \right]$$

$$p_{pt}(t|\Theta) = k \cdot \left[\int_0^{t-\Theta} e^{-a\left(\tilde{t} - \frac{b_1}{2a}\right)^2 + \frac{{b_1}^2}{4a} + c_1} d\tilde{t} - \int_0^{t-\Theta} e^{-a\left(\tilde{t} - \frac{b_2}{2a}\right)^2 + \frac{{b_2}^2}{4a} + c_2} d\tilde{t} \right]$$

$$p_{pt}(t|\Theta) = k \cdot \left[e^{\frac{{b_1}^2}{4a} + c_1} \int_0^{t-\Theta} e^{-a\left(\tilde{t} - \frac{b_1}{2a}\right)^2} d\tilde{t} - e^{\frac{{b_2}^2}{4a} + c_2} \int_0^{t-\Theta} e^{-a\left(\tilde{t} - \frac{b_2}{2a}\right)^2} d\tilde{t} \right]$$

Finally, by replacing $x = \tilde{t} - \frac{b}{2a}$, the integral can be solved with the error function as in:

$$\int e^{-ax^2} dx = \sqrt{\frac{\pi}{4a}} \cdot erf\left(\sqrt{a}x\right)$$

Resulting in:

$$p_{pt}(t|\Theta) = k \cdot \left[e^{\frac{{b_1}^2}{4a} + c_1} \cdot \sqrt{\frac{\pi}{4a}} \cdot \left| erf\left(\sqrt{a}\left(\tilde{t} - \frac{b_1}{2a}\right)\right) \right|_0^{t-\Theta} - e^{\frac{{b_2}^2}{4a} + c_2} \cdot \sqrt{\frac{\pi}{4a}} \cdot \left| erf\left(\sqrt{a}\left(\tilde{t} - \frac{b_2}{2a}\right)\right) \right|_0^{t-\Theta} \right]$$

Which finally gives:

$$p_{pt}(t|\Theta) = k \cdot e^{\frac{{b_1}^2}{4a} - \frac{\mu^2}{2 \cdot {\sigma_{trans}}^2} - \frac{t-\Theta}{\tau_d}} \cdot \sqrt{\frac{\pi}{4a}} \cdot \left[erf\left(\sqrt{a}\left(t - \Theta - \frac{b_1}{2a}\right)\right) - erf\left(-\frac{b_1}{2\sqrt{a}}\right) \right] - k \cdot e^{\frac{{b_2}^2}{4a} - \frac{\mu^2}{2 \cdot {\sigma_{trans}}^2} - \frac{t-\Theta}{\tau_r}} \cdot \sqrt{\frac{\pi}{4a}} \cdot \left[erf\left(\sqrt{a}\left(t - \Theta - \frac{b_2}{2a}\right)\right) - erf\left(-\frac{b_2}{2\sqrt{a}}\right) \right]$$

The CDF in equation (8) can be obtained in a similar way.

$$P_{pt}(t|\Theta) = \int_0^{t-\Theta} P_{em}(t - \tilde{t}|\Theta) \times p_{jitter}(\tilde{t}) d\tilde{t}$$

First, the two exponential-based PDFs are multiplied.

$$P_{pt}(t|\Theta) = \int_0^{t-\Theta} \left(1 - \frac{\tau_d \cdot e^{-\frac{t-\tilde{t}-\Theta}{\tau_d}} - \tau_r \cdot e^{-\frac{t-\tilde{t}-\Theta}{\tau_r}}}{\tau_d - \tau_r} \right) \times \frac{1}{\sigma_{trans} \cdot \sqrt{2\pi}} \cdot e^{-\frac{(\tilde{t}-\mu)^2}{2 \cdot {\sigma_{trans}}^2}} \cdot d\tilde{t}$$

$$P_{pt}(t|\Theta) = \int_0^{t-\Theta} \frac{1}{\sigma_{trans} \cdot \sqrt{2\pi}} \cdot e^{-\frac{(\tilde{t}-\mu)^2}{2\cdot {\sigma_{trans}}^2}} \cdot d\tilde{t} -$$
$$\frac{1}{\tau_d - \tau_r} \cdot \frac{1}{\sigma_{trans} \cdot \sqrt{2\pi}} \cdot \int_0^{t-\Theta} \left[\tau_d \cdot e^{-\frac{t-\tilde{t}-\Theta}{\tau_d}} - \tau_r \cdot e^{-\frac{t-\tilde{t}-\Theta}{\tau_r}}\right] \cdot e^{-\frac{(\tilde{t}-\mu)^2}{2\cdot {\sigma_{trans}}^2}} \cdot d\tilde{t}$$

The integral of the Gaussian function is well-known to be the error function:

$$P_{pt}(t|\Theta) = \frac{1}{2} \cdot erf\left(\frac{\tilde{t} - \mu}{\sqrt{2} \cdot \sigma_{trans}}\right)\Bigg|_{\tilde{t}=0}^{t-\Theta} -$$
$$k \cdot \int_0^{t-\Theta} \left[\tau_d \cdot e^{-\frac{t-\tilde{t}-\Theta}{\tau_d} - \frac{(\tilde{t}-\mu)^2}{2\cdot {\sigma_{trans}}^2}} - \tau_r \cdot e^{-\frac{t-\tilde{t}-\Theta}{\tau_r} - \frac{(\tilde{t}-\mu)^2}{2\cdot {\sigma_{trans}}^2}}\right] d\tilde{t}$$

The remaining integral can be solved in the same way as the integral in the PDF solution above, finally resulting in:

$$P_{pt}(t|\Theta) = \frac{1}{2} \cdot \left[erf\left(\frac{t - \Theta - \mu}{\sqrt{2} \cdot \sigma_{trans}}\right) - erf\left(\frac{-\mu}{\sqrt{2} \cdot \sigma_{trans}}\right)\right] -$$
$$k \cdot \tau_d \cdot e^{\frac{{b_1}^2}{4a} - \frac{\mu^2}{2\cdot {\sigma_{trans}}^2} - \frac{t-\Theta}{\tau_d}} \cdot \sqrt{\frac{\pi}{4a}} \cdot \left[erf\left(\sqrt{a}\left(t - \Theta - \frac{b_1}{2a}\right)\right) - erf\left(-\frac{b_1}{2\sqrt{a}}\right)\right] +$$
$$k \cdot \tau_r \cdot e^{\frac{{b_2}^2}{4a} - \frac{\mu^2}{2\cdot {\sigma_{trans}}^2} - \frac{t-\Theta}{k_\tau}} \cdot \sqrt{\frac{\pi}{4a}} \cdot \left[erf\left(\sqrt{a}\left(t - \Theta - \frac{b_2}{2a}\right)\right) - erf\left(-\frac{b_2}{2\sqrt{a}}\right)\right]$$