



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

Time Resolved Single Photon Imaging in Nanometer Scale CMOS Technology

Justin A. Richardson



A thesis submitted for the degree of Doctor of Philosophy
The University of Edinburgh
25 May 2010

Supervised by: Dr. Robert Henderson
Dr. David Renshaw

The School of Engineering,
The University of Edinburgh,
The Kings Buildings,
Mayfield Road,
Edinburgh.

Examined by: Dr. Renato Turchetta
Science and Technologies Facilities Council,
Rutherford Appleton Laboratory,
Harwell Science and Innovation Campus,
Didcot, Oxfordshire.

Prof. Alan Murray
The School of Engineering,
The University of Edinburgh,
The Kings Buildings,
Mayfield Road,
Edinburgh.

Abstract

Time resolved imaging is concerned with the measurement of photon arrival time. It has a wealth of emerging applications including biomedical uses such as fluorescence lifetime microscopy and positron emission tomography, as well as laser ranging and imaging in three dimensions. The impact of time resolved imaging on human life is significant: it can be used to identify cancerous cells *in-vivo*, how well new drugs may perform, or to guide a robot around a factory or hospital.

Two essential building blocks of a time resolved imaging system are a photon detector capable of sensing single photons, and fast time resolvers that can measure the time of flight of light to picosecond resolution. In order to address these emerging applications, miniaturised, single-chip, integrated arrays of photon detectors and time resolvers must be developed with state of the art performance and low cost. The goal of this research is therefore the design, layout and verification of arrays of low noise Single Photon Avalanche Diodes (SPADs) together with high resolution Time-Digital Converters (TDCs) using an advanced silicon fabrication process.

The research reported in this Thesis was carried out as part of the E.U. funded Megaframe FP6 Project. A 32x32 pixel, one million frames per second, time correlated imaging device has been designed, simulated and fabricated using a 130nm CMOS Imaging process from ST Microelectronics. The imager array has been implemented together with required support cells in order to transmit data off chip at high speed as well as providing a means of device control, test and calibration. The fabricated imaging device successfully demonstrates the research objectives.

The Thesis presents details of design, simulation and characterisation results of the elements of the Megaframe device which were the author's own work. Highlights of the results include the smallest and lowest noise SPAD devices yet published for this class of fabrication process and an imaging array capable of recording single photon arrivals every microsecond, with a minimum time resolution of fifty picoseconds and single bit linearity.

Declaration of Originality

I hereby declare that the research recorded in this thesis and the thesis itself originated with and was composed entirely by myself.

Other people who have contributed to this work are acknowledged appropriately within.

Justin A. Richardson

Acknowledgements

During the course of this research I have had the pleasure of working with many highly skilled, professional people who have helped and supported me throughout. Without them, the implementation of the fourteen devices which were created over the last three and a half years would not have been possible.

I would like to express my gratitude particularly towards my academic supervisors at The University of Edinburgh, Dr. Robert Henderson and Dr. David Renshaw, whose experience, knowledge and sound advice was, and always will be the underlying foundation of this work. Outside of the electronic engineering environment, the solid support of my family and friends has been vital and has given me the strength to see it through to a logical conclusion. However, I am thankful that it is by no means the end of the work, and that further ongoing development activities are already commencing.

I have also benefited greatly from my exposure to the European funded Megaframe Project and people involved in it. Working on this project has contributed greatly to my professional knowledge of electronic engineering as well as my own personal development. In their own unique way and in no particular order, I will always be indebted to Prof. Edoardo Charbon, and Drs. Claudio Bruschini, David Stoppa, Fausto Borghetti, Marek Gersbach, Cristiano Niclass, Day-Wei Li and Eric Webster. However, of particular importance to me has been the friendship, freely granted knowledge and engineering skill of key University colleagues Dr. Bruce Rae and Richard Walker, both of whom helped immeasurably to make this research become a physical reality. All these individuals will always be my friends.

The support of STMicroelectronics over the past three years has also been very significant. I would like to thank my supervisor, Lindsay Grant, and long standing colleague and thesis proof reader Dr. Andrew Holmes. Drew's frank and honest advice on all PhD matters has been invaluable. In addition, ST colleagues Dr. Robert Nicol, Bob Stevenson and David Poyner have all contributed their valuable time and advice, and for that I will always be grateful.

This work is dedicated to the memory of my Grandfather, John Alfred (Jack) Hampton, 1908-1994, one of the finest photographers in the world.

Table of Contents

1	Introduction.....	16
1.1	A New Class of Time Correlated Imager	16
1.2	3D Imaging	17
1.3	Fluorescence Lifetime Microscopy.....	17
1.4	Positron Emission Tomography.....	20
1.5	Time Correlated Imaging System Components.....	22
1.5.1	Detection	23
1.5.2	Chronometry	24
1.5.3	Illumination.....	25
1.6	Project Aims and Benefits	26
1.7	Beneficiaries of the Research.....	27
1.8	Methodologies.....	27
1.9	Project Context.....	28
1.10	Contribution to Knowledge.....	28
1.11	Time Correlated Imaging System Metrics	28
1.12	Summary of Thesis Structure.....	29
2	Background	31
2.1	Overview	31
2.2	SPAD Performance Parameters	31
2.2.1	Dark Count Rate	33
2.2.2	Afterpulsing	34
2.2.3	Photon Detection Efficiency	35
2.2.4	Timing Resolution	36
2.2.5	Dead Time.....	38
2.2.6	Crosstalk	39
2.2.7	Active Area, Fill Factor and Breakdown Homogeneity	39
2.3	SPAD Literature Review	41
2.3.1	Background	41
2.3.2	SPAD State of the Art.....	41
2.3.3	Process Choice	42
2.3.4	Detector Construction	44
2.3.5	Quenching	52
2.4	Detector Conclusions	59
2.5	TDC Performance Parameters	61
2.5.1	Time Resolution.....	61
2.5.2	Dynamic Range.....	61
2.5.3	Accuracy	62
2.5.4	Precision.....	62
2.5.5	Conversion Rate.....	62
2.5.6	Power Consumption.....	63
2.5.7	Compensation Capability.....	63

2.5.8	Silicon Area	63
2.6	TDC Literature Review.....	64
2.6.1	Background	64
2.6.2	Clocked Delay Lines.....	64
2.6.3	Vernier Delay Lines	66
2.6.4	Pulse Shrinkers.....	67
2.6.5	Passive Interpolators	69
2.6.6	Time Stretchers	70
2.6.7	Distributed Clock Structures.....	71
2.6.8	Time to Amplitude Approaches.....	73
2.6.9	Gated Ring Oscillators	74
2.7	Integrated Detector-Converter Arrays	76
2.7.1	Photodiode-ADC Arrays for Bioluminescence Detection.....	76
2.7.2	Medipix High Energy Particle Detector Arrays.....	77
2.7.3	SensL Digital APD Arrays.....	77
2.7.4	SPAD-TDC Arrays	78
2.8	Conclusions.....	79
3	SPAD Design	80
3.1	Target Specification	80
3.2	Processing Considerations	81
3.2.1	Introduction.....	81
3.2.2	Designing SPADs in the IMG175 Process	82
3.2.3	Process Implant Toolbox	83
3.3	Gettering	86
3.4	Physical Optimisation	86
3.4.1	Optical Stack Considerations.....	87
3.4.2	Detector Shape	89
3.4.3	Anode Contact Positioning	89
3.5	Prior Art Implementation.....	90
3.5.1	Device Parameter Calculations	90
3.5.2	Process Simulation.....	93
3.5.3	Results Summary	96
3.6	New Detector Constructions	96
3.6.1	PPLUSPSTI_NISO SPAD.....	96
3.6.2	PPLUSPSTI_NWELLNISO SPAD.....	102
3.6.3	PPLUSPWELL_NISO_EIPOLY SPAD	107
3.7	New Detector Physical Optimisation.....	112
3.7.1	Shape.....	112
3.7.2	Scaling.....	114
3.7.3	Active Region Enhancements	116
3.8	Quenching	117
3.8.1	Optimised Passive Quench	117
3.8.2	Thyristor Active Quench.....	120
3.9	Host Integrated Circuit.....	123
3.10	Conclusions.....	124
4	Detector and Quench Characterisation	127
4.1	Introduction.....	127
4.2	Characterisation Procedure	127
4.2.1	Reverse Breakdown Voltage.....	127
4.2.2	Dark Count Rate	127

4.2.3	Photon Detection Efficiency	129
4.2.4	Dead Time.....	130
4.2.5	Afterpulsing	130
4.2.6	Timing Resolution	130
4.3	SPAD Characterisation Results	132
4.3.1	PPLUSPSTI_NISO SPAD Characterisation Results.....	132
4.3.2	PPLUSPSTI_NWELL SPAD Characterisation Results	138
4.3.3	PPLUSPWELL_NISO SPAD Characterisation Results.....	144
4.4	Scaling and Shaping Observations.....	151
4.4.1	Impact on I-V Response.....	151
4.4.2	Impact on Dark Count.....	152
4.4.3	Impact on Timing Resolution	154
4.5	Active Quenching Characterisation	156
4.6	Comparison of Results	157
4.7	Conclusions.....	160
5	Time to Digital Converter Array Design	162
5.1	Target Specification	162
5.2	Description of Chosen TDC Architecture.....	163
5.3	GRO TDC Timing	165
5.4	GRO TDC Block Descriptions	166
5.4.1	Logic Block.....	166
5.4.2	Ring Oscillator	168
5.4.3	Coder.....	170
5.4.4	Ripple Counter	171
5.4.5	Memory.....	172
5.5	Quench Cell	172
5.6	Calibration/Compensation Scheme.....	173
5.7	GRO TDC Pixel Layout.....	174
5.8	GRO TDC Array Data Readout.....	177
5.8.1	Overview	177
5.8.2	System Synchronisation and Device Clocking.....	179
5.8.3	Data Serialiser	181
5.8.4	Row Y-Decoder	181
5.9	Design for Test and Characterisation.....	183
5.10	Top Chip Floorplan.....	184
5.11	Device Packaging.....	185
5.12	Hardware Platform.....	186
5.13	Conclusions.....	187
6	TDC Characterisation	188
6.1	Introduction.....	188
6.2	TDC Characterisation Results.....	188
6.2.1	Photon Counting Mode	189
6.2.2	Time Correlated Single Photon Counting Mode	192
6.3	Conclusions.....	208
7	Conclusions and Outlook.....	210
7.1	Summary	210
7.2	Achievements.....	210
7.3	Improvements and Future Work	213
7.4	Final Remarks	214
Appendix A:	List of Publications	16

Appendix B:	Patent Applications	219
Appendix C:	SPAD Design Additional Data	220
Appendix D:	TDC Design Additional Data.....	224
Appendix E:	References	227

Table List

Table 1: SPAD Metric/Construction Summary	60
Table 2: TDC Metric/Construction Summary	75
Table 3: SPAD Design Performance Goals	81
Table 4: PPLUS_NWELL SPAD Results Comparison.....	96
Table 5: PPLUSPSTI_NISO SPAD Expected Results.....	101
Table 6: PPLUSPSTI_NWELLNISO SPAD Expected Results.....	106
Table 7: PPLUSPWELL_NISO SPAD Expected Results.....	111
Table 8: SPAD Shaping Fill Factor Comparison.....	114
Table 9: Fill Factor Comparison Data	115
Table 10: Scaled Ideal Dark Count Rate.....	115
Table 11: Optimised Passive Quench Summary.....	119
Table 12: Thyristor Active Quench Summary	122
Table 13: Simulated SPAD Performance Comparison.....	125
Table 14: Light Illuminance.....	129
Table 15: PPLUSPSTI_NISO SPAD Summary	137
Table 16: PPLUSPSTI_NWELL SPAD Summary	143
Table 17: PPLUSPWELL_NISO SPAD Summary	150
Table 18: Thyristor Active Quench Performance.....	156
Table 19: SPAD Performance Summary	157
Table 20: SPAD DCR Summary	157
Table 21: GRO TDC Performance Target	163
Table 22: Metal Routing Capacitance, Sheet Resistance and Layout Rules	174
Table 23: Metal Usage and Direction Rules	175
Table 24: TDC Mode Synchronisation	180
Table 25: Test Mode Access.....	183
Table 26: GRO TDC Array Expected Results	187
Table 27: GRO-TDC Time Resolution.....	192
Table 28: TDC Line Rejection Performance	206
Table 29: GRO-TDC Performance Summary.....	208
Table 30: Power Consumption Summary	208
Table 31: Fabrication Process Construction	220
Table 32: TDC Input-Output Signal Definition.....	224
Table 33: 32x32 TDC Array I2C Register Map	225

Table of Figures

Figure 1: Direct ToF 3D Imaging	17
Figure 2: Jablonski Energy Diagram	18
Figure 3: Fluorescence Lifetime Summary.....	19
Figure 4: FLIM Image of Hydrophobic Cancerous Liver Cells	19
Figure 5: Positron Emission Tomography Summary.....	20
Figure 6: CT, PET and Combined Images.....	21
Figure 7: TCSPC System Components.....	22
Figure 8: Basic Geiger Mode SPAD and Passive Quench Circuit	24
Figure 9: ‘Forward Mode’ Time Measurement	25
Figure 10: SPAD Energy Band Diagram.....	32
Figure 11: SPAD Dark Count Sources	33
Figure 12: SPAD Research Areas Classified by Process.....	42
Figure 13: Diffused Guard Ring SPAD.....	45
Figure 14: Reach Through SPAD	46
Figure 15: Enhancement Mode SPAD.....	47
Figure 16: Merged Implant Guard Ring SPAD	48
Figure 17: Gate Bias and Floating Guard Ring Constructions	49
Figure 18: Timing Optimised SPAD	50
Figure 19: Shallow Trench Guard Ring SPAD.....	51
Figure 20: SPAD Cycle of Operation	52
Figure 21: SPAD Passive Quench Model.....	53
Figure 22: Active Quench Operating Phases	55
Figure 23: Example Active Quench Implementation	56
Figure 24: SPAD Orientation Options.....	57
Figure 25: SPAD Arming Options.....	58
Figure 26: Clocked Delay Line TDC.....	65
Figure 27: Delay Locked Loop Calibration Concept.....	66
Figure 28: Vernier Delay Line TDC.....	67
Figure 29: Pulse Nibbler TDC	68
Figure 30: Time Interpolation Concept.....	69
Figure 31: Differential Interpolated Delay Line	70
Figure 32: Time Amplification Concept.....	71
Figure 33: Distributed Clock Concept	72
Figure 34: Time-Amplitude-Digital Conversion Principle.....	73
Figure 35: Ring Oscillator TDC	74
Figure 36: Integrated System Error Sources	76
Figure 37: Process Layer Key.....	83
Figure 38: N-Implant Depth.....	84
Figure 39: N-Implant Doping Concentration.....	84
Figure 40: P-Implant Depth	85
Figure 41: P-Implant Doping Concentration	85

Figure 42: Optical Stack Constructions	87
Figure 43: Use of Optical Stack Features	88
Figure 44: Detector Shape Options.....	89
Figure 45: Anode Contact Positioning.....	90
Figure 46: PPLUS_NWELL SPAD Doping Profile.....	93
Figure 47: PPLUS_NWELL SPAD Electric Field Profile	93
Figure 48: PPLUS-NWELL SPAD I-V Response.....	94
Figure 49: PPLUS-NWELL SPAD 2D Doping Simulation.....	95
Figure 50: PPLUS-NWELL SPAD 2D E-Field Simulation.....	95
Figure 51: PPLUSPSTI_NISO_EPIPOLY SPAD Cross Section.....	97
Figure 52: PPLUSPSTI_NISO SPAD Doping Profile	98
Figure 53: PPLUSPSTI_NISO SPAD Electric Field Profile.....	98
Figure 54: PPLUSPSTI_NISO SPAD I-V Response	99
Figure 55: PPLUSPSTI_NISO SPAD 2D Doping Simulation.....	99
Figure 56: PPLUSPSTI_NISO SPAD E-Field Simulation.....	100
Figure 57: Cadence Layout for PPLUSPSTI_NISO SPAD	100
Figure 58: PPLUSPSTI_NWELLNISO SPAD Cross Section	102
Figure 59: PPLUSPSTI_NWELLNISO SPAD Doping Concentration.....	103
Figure 60: PPLUSPSTI_NWELLNISO SPAD Electric Field.....	103
Figure 61: PPLUSPSTI_NWELLNISO SPAD I-V Response	104
Figure 62: PPLUSPSTI_NWELLNISO SPAD 2D Doping Simulation.....	104
Figure 63: PPLUSPSTI_NWELLNISO SPAD 2D E-Field Simulation.....	105
Figure 64: PPLUSPSTI_NWELLNISO_STI SPAD Layout.....	105
Figure 65: PPLUSPWELL_NISO_EPIPOLY SPAD Cross Section.....	107
Figure 66: PPLUSPWELL_NISO SPAD Doping Concentration	108
Figure 67: PPLUSPWELL_NISO SPAD Electric Field	108
Figure 68: PPLUSPWELL_NISO SPAD I-V Response	109
Figure 69: PPLUSPWELL_NISO SPAD 2D Doping Simulation.....	109
Figure 70: PPLUSPWELL_NISO SPAD 2D E-Field Simulation.....	110
Figure 71: PPLUSPWELL_NISO_EPIPOLY SPAD Layout	110
Figure 72: ‘Make Quadrant’ Script Result.....	112
Figure 73: Fermat Shaped PPLUSPWELL_NISO SPAD	113
Figure 74: Square Shaped PPLUSPWELL_NISO SPAD	113
Figure 75: SPAD Family Plot.....	114
Figure 76: Fermat SPAD Anode Contact Position	116
Figure 77: Passive Quench Optimisation.....	118
Figure 78: Thyristor Active Quench Circuit Diagram.....	120
Figure 79: Slow Rise-Time Current Starved Inverter Schematic	121
Figure 80: ThyristorActive Quench Simulation Results.....	121
Figure 81: Thyristor Active Quench Layout.....	122
Figure 82: ‘IMNSTEST’ Device Micrograph.....	123
Figure 83: Reverse I-V Test Mode Capability.....	124
Figure 84: SPADDEVELA Device Micrograph.....	124
Figure 85: SPAD Simulated E-Field Comparison	126
Figure 86: SPAD Jitter Test Setup.....	131
Figure 87: PPLUSPSTI_NISO SPAD I-V Response	132
Figure 88: PPLUSPSTI_NISO SPAD Dark Count Rate	133
Figure 89: PPLUSPSTI_NISO SPAD Photon Detection Efficiency.....	134
Figure 90: PPLUSPSTI_NISO SPAD Afterpulsing Probability	135
Figure 91: PPLUSPSTI_NISO SPAD Jitter	136

Figure 92: PPLUSPSTI_NWELL SPAD I-V Response.....	138
Figure 93: PPLUSPSTI_NWELL SPAD Dark Count Rate	139
Figure 94: PPLUSPSTI_NWELL SPAD Photon Detection Efficiency	140
Figure 95: PPLUSPSTI_NWELL SPAD Afterpulsing Probability.....	141
Figure 96: PPLUSPSTI_NWELL SPAD Jitter.....	142
Figure 97: PPLUSPWELL_NISO SPAD I-V Response	144
Figure 98: PPLUSPWELL_NISO SPAD Dark Count Rate	145
Figure 99: PPLUS_PWELL SPAD DCR Distribution.....	146
Figure 100: PPLUSPWELL_NISO SPAD Photon Detection Efficiency.....	147
Figure 101: PPLUSPWELL_NISO SPAD Afterpulsing Probability	148
Figure 102: PPLUSPWELL_NISO SPAD Jitter	149
Figure 103: Impact of Diameter on I-V Response.....	151
Figure 104: 4 μ m Diameter SPAD DCR Distribution	152
Figure 105: Round SPAD Dark Count versus Area	153
Figure 106: Round SPAD DCR versus Drawn Diameter.....	154
Figure 107: Impact of Diameter on Timing Resolution.....	155
Figure 108: Active Quench Micrograph	156
Figure 109: SPAD I-V Response Comparison	158
Figure 110: SPAD PDE Comparison Plot	158
Figure 111: SPAD Jitter Comparison	159
Figure 112: GRO TDC Block Diagram	164
Figure 113: GRO TDC-Pixel IO.....	164
Figure 114: GRO TDC Timing.....	165
Figure 115: Logic Block Schematic	167
Figure 116: Differential Gated Ring Oscillator	169
Figure 117: Differential Inverter.....	169
Figure 118: Differential Ring Oscillator Layout	170
Figure 119: Coder Block Schematic	171
Figure 120: Ripple Counter Schematic.....	171
Figure 121: Memory Unit Cell and Timing	172
Figure 122: GRO TDC SPAD Quench Cell	173
Figure 123: PLL Global Calibration Architecture	174
Figure 124: GRO TDC Pixel Layout	175
Figure 125: GRO TDC Pixel TDC Micrograph	176
Figure 126: TCSPC System Block Diagram	177
Figure 127: GRO TDC Pixel Array Data Readout Structure	178
Figure 128: Region of Interest Selection	179
Figure 129: System Clocking and Synchronisation Diagram.....	180
Figure 130: Data Serialiser Block Diagram.....	181
Figure 131: Row Y-Decoder Block Diagram	182
Figure 132: Top Chip Micrograph	184
Figure 133: 32x32 GRO TDC Pixel Array in 180pin CPGA Package.....	185
Figure 134: GRO TDC Array Hardware Platform	186
Figure 135: Photon Counting Mode Transfer Function Profile.....	189
Figure 136: Photon Counting Mode Array Uniformity Plot.....	190
Figure 137: Photon Counting Mode Power Consumption Profile.....	191
Figure 138: Slow TDC Transfer Profile	193
Figure 139: Fast TDC Transfer Profile	193
Figure 140: Code Occurrence DNL Computation Method.....	194
Figure 141: GRO-TDC DNL	195

Figure 142: GRO-TDC INL.....	196
Figure 143: GRO-TDC Code Probability	197
Figure 144: GRO-TDC Output Correlation Analysis	198
Figure 145: GRO-TDC Jitter Analysis	199
Figure 146: Slow GRO-TDC Jitter Statistics.....	199
Figure 147: Fast GRO-TDC Jitter Statistics	200
Figure 148: TCSPC Mode Array Mean Uniformity	201
Figure 149: Mean Code Probability: Slow TDC	201
Figure 150: Mean Code Probability: Fast TDC	202
Figure 151: Row (Vertical) Signatures	203
Figure 152: Column (Horizontal) Signatures	204
Figure 153: Tracking of TDC Resolution to the Control PLL.....	205
Figure 154: TDC Stability to Core Voltage Variation.....	206
Figure 155: TCSPC Mode Power Consumption Profile.....	207
Figure 156: Example Bio-analysis Images	212
Figure 157: Example 3D Imaging Bench Setup and Output Image	212
Figure 158: Implant Energy vs Depth in Silicon	223

Acronyms & Abbreviations

APD	- avalanche photodiode
CCD	- charge coupled device
CFD	- constant fraction discriminator
CMOS	- complimentary metal oxide semiconductor
CT	- computed tomography
CW	- continuous wave
DCR	- dark count rate
DLL	- delay locked loop
DNL	- differential non-linearity
FCS	- fluorescence correlation spectroscopy
FDG	- fludeoxyglucose
FLIM	- fluorescent lifetime imaging microscopy
FPS	- frames per second
FWHM	- full width half maximum
GFP	- green fluorescent protein
INL	- integral non-linearity
LED	- light emitting diode
LOR	- line of response
LPF	- low pass filter
LSB	- least significant bit
MCP	- micro channel plate
MRI	- magnetic resonant imaging
PDE	- photon detection efficiency
PET	- positron emission tomography
PFD	- phase frequency detector
PLL	- phase locked loop
PMT	- photo-multiplier tube
PVT	- process, voltage, temperature
RCA	- Radio Corporation of America
SiPM	- silicon photo-multiplier
SPAD	- single photon avalanche diode
TAC	- time to analogue/amplitude converter
TCAD	- technology computer aided design
TCSPC	- time correlated single photon counting
TDC	- time to digital converter
ToF	- time of flight
TXPC	- time uncorrelated photon counting
VCO	- voltage controlled oscillator
VLSI	- very large scale integration
ZIF	- zero insertion force

1 Introduction

1.1 A New Class of Time Correlated Imager

The detection of light using photographic film was pioneered by several inventors starting from the 1820s. High quality silver-halide film was the result of the following one hundred and sixty years progress of materials research. However, the appeal of chemical-free instant display of images pushed forward novel early solid state electronic based solutions in the 1960s such as the custom process 50x50 pixel array by Schuster & Strull [1] the charge coupled devices (CCD) of the 1970s (Boyle and Smith [2]), followed in the late 1980s and early 1990s by the system on chip (SoC) complimentary oxide semiconductor (CMOS) process based image sensor (Renshaw et al [3], Fossum et al [4]). These silicon image sensors are now commonly used in everyday life for a wide variety of uses. Importantly, the past decade has witnessed significant advances in CMOS cameras in terms of manufacturing technology, image quality and miniaturisation in line with Moore's law process feature size reduction. It is now possible to have an affordable five megapixel, high quality camera complete with auto-focus capability integrated into the case of a slim mobile phone. Systems for automotive, military or security platforms may contain ten or more miniaturised sensors. Collectively these uses have driven an exciting, innovative market place for CMOS cameras. However, all of these applications are concerned with measuring the number of incident photons for each photosensitive site (or 'pixel') within a defined exposure time, i.e. *photon intensity* images. However, photons exhibit other properties that contain important information about a subject under study, such as wavelength, polarisation, and time of arrival.

There is a class of image sensors in which the time of arrival of a single photon with respect to a datum point in time is measured. These are called time-correlated imagers, and are usually one element of a larger system which extracts cumulative statistical information from many sequential measurements in order to, for example, measure distance to or volume of an object (3D), perform Fluorescence Lifetime Microscopy (FLIM), or identify coincident photon arrivals for Positron Emission Tomography (PET). Today's systems use established technology such as PMTs, micro-channel plates or custom-processed devices such as SiPMs.

Leveraging from the progress of the mainstream CMOS camera fabrication technology, the research reported in this Thesis aims to implement both a high performance single photon detector and time resolver together in an array format complete with integrated data readout and support circuitry on the same silicon chip to create a new class of time correlated imagers. Firstly it is necessary to understand the basic principles of the three main targetted time-correlated imaging modalities.

1.2 3D Imaging

The range to an object may be evaluated by timing the duration taken by of a pulse of photons to leave the illumination source, strike an object and return to a suitable photon detector. The constant speed of light ($3 \times 10^8 \text{ ms}^{-1}$) is then used to compute the flight path length. This *direct*, non-inferred computation method is known as time-of-flight (ToF) computation. The basic principle is summarised in Figure 1.

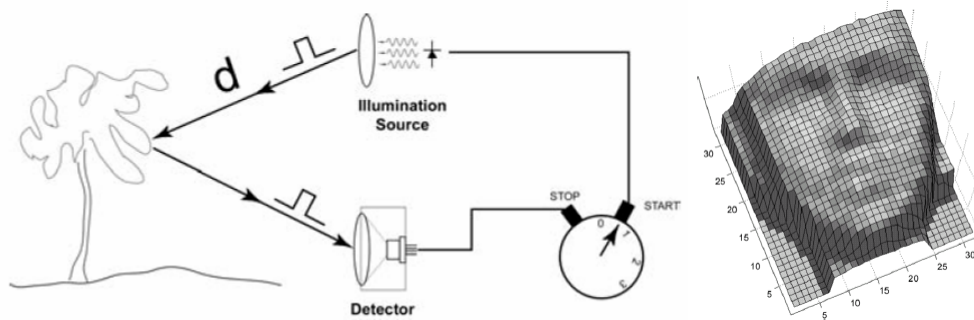


Figure 1: Direct ToF 3D Imaging

(Diagram courtesy of Dr. Cristiano Niclass, EPFL.)

A 3D image may be assembled from many individual range computations using scanned laser systems or by employing arrays of individual detectors with time resolving capability. Alternatively, distance may be computed *indirectly* by analysis of the phase shift of a modulated light source.

1.3 Fluorescence Lifetime Microscopy

Biological analyses based on fluorescence are well established. Fluorescence provides more detailed information about the location and environment of cells or fluorescent marker molecules than other methods which are affected by test setup and environmental effects such as illumination intensity, polarisation, path length and

light source homogeneity. This yields images with better spatial contrast than when using reflected light microscopy. When applied in a wide field format the technique is known as Fluorescence Lifetime Microscopy or FLIM. The molecular principle of fluorescence is described using the Jablonski energy diagram shown in Figure 2.

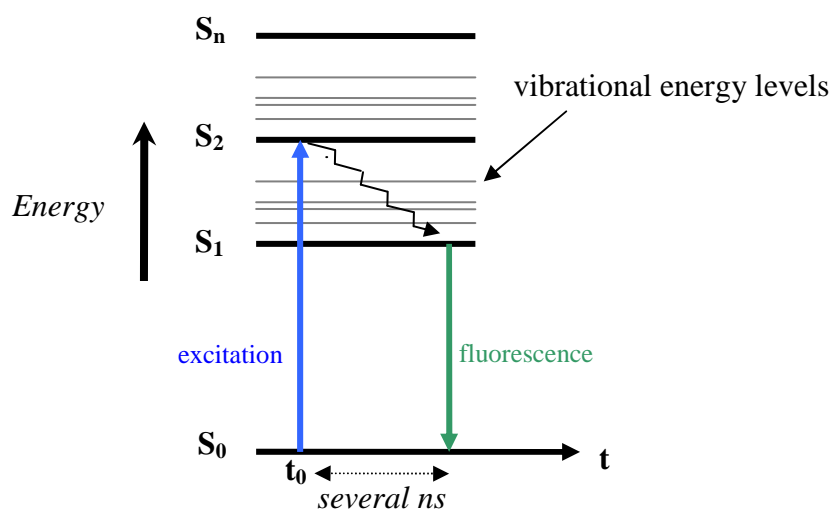


Figure 2: Jablonski Energy Diagram

This shows that by raising the energy level of a molecule via excitation such as a pulsed laser, the emission of a lower energy photon can result. The wavelength difference between excitation and emission photons is caused by energy lost as the molecule relaxes over a period of time through interstitial vibrational energy levels. This phenomenon is known as the *Stokes Shift*. Plotting a histogram of repeated measurements of this period of time for a certain fluorophore molecule reveals the fluorescence lifetime, τ , shown in Figure 3.

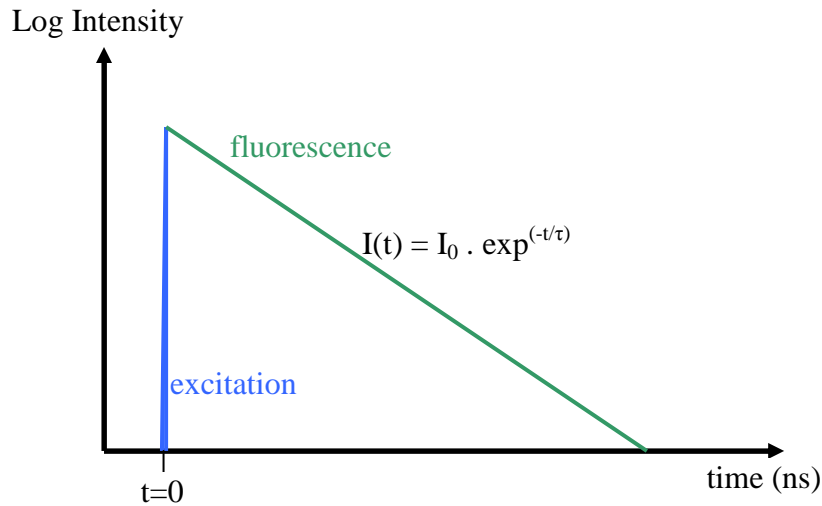


Figure 3: Fluorescence Lifetime Summary

The fluorescence decay lifetime τ represents the mean time the molecules stay in their excited state. This permits biological analyses based on excited state reactions and free from light intensity based variances. This enables scientists to monitor biological reactions deep within cells, as demonstrated in Figure 4. This shows a comparison of intensity (left) and FLIM images of hydrophobic, cancerous liver cells side by side.

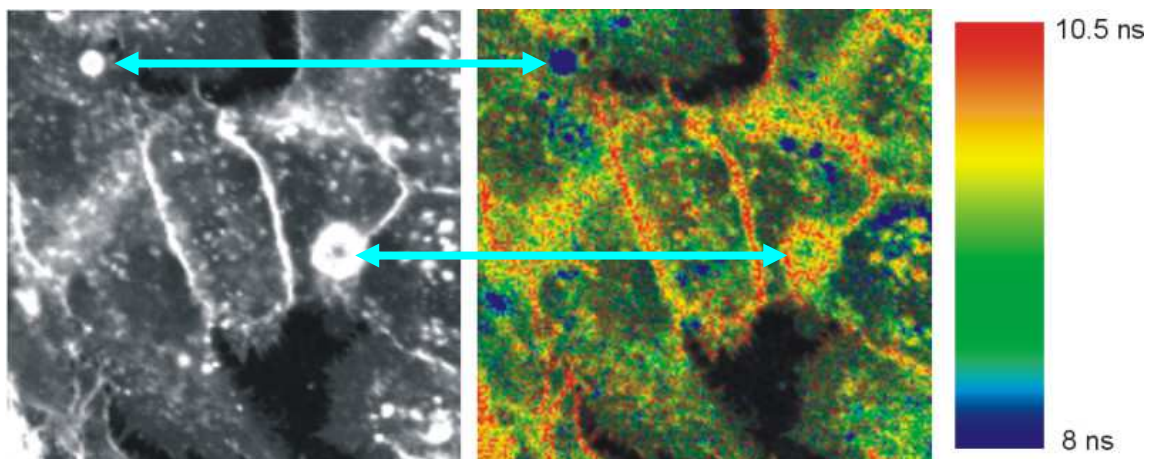


Figure 4: FLIM Image of Hydrophobic Cancerous Liver Cells

The arrows pick out key differences between the images. The intensity image shows two equally bright circular spots but FLIM reveals that the spots have very different lifetimes. Therefore quite different biochemical reactions are taking place at these locations.

1.4 Positron Emission Tomography

Positron Emission Tomography (PET) is a nuclear imaging technique which enables the creation of a 3D representation of biological processes taking place within a living body. Commonly proposed for integration with computed tomography (CT X-ray) or magnetic resonance imaging (MRI), it enables scientists to view zones of biological activity simultaneously alongside an internal structural representation of the subject. This enables greater understanding and detail of events taking place within a living human patient, allows in-vivo animal studies without loss of life and removes the requirement for control subjects. The basic physics of PET are shown in Figure 5.

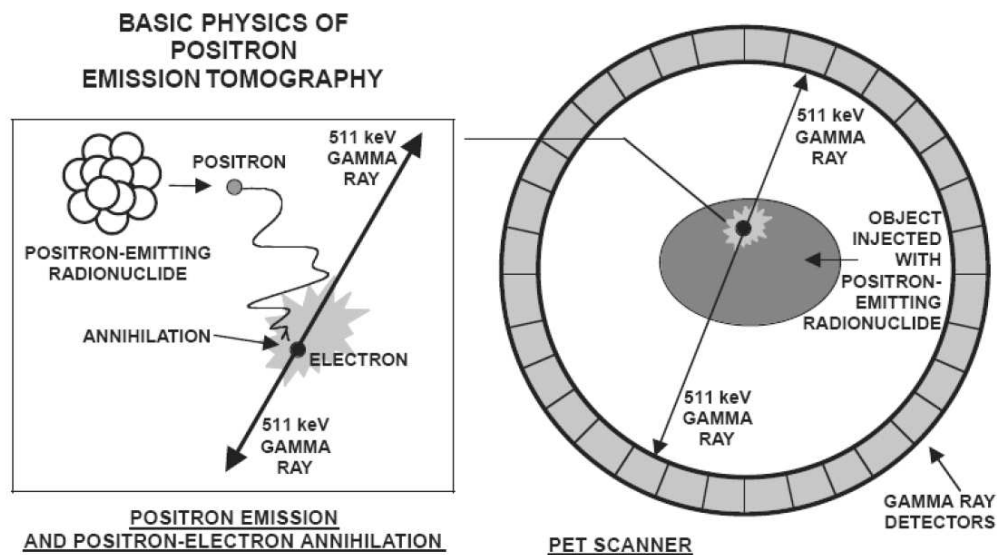


Figure 5: Positron Emission Tomography Summary

A positron emitting radionuclide bio-tracer such as fludeoxyglucose (FDG, a glucose analogue) is introduced into the live subject. The tracer uptake is higher in regions of increased metabolic activity such as those affected by cancerous cell concentrations. The positron annihilates with a nearby electron, resulting in the diametric emission of two gamma photons. The detection of the arrival of the two largely coincident gamma photons by a ring shaped detector allows the plotting of a line-of-response (LOR). A tomographical slice through the body is then reconstructed via the detection of many LORs. The subject can then be traversed through the detection ring to build a full 3D image of biological activity. Figure 6 shows a reconstruction of CT, PET and combined images of FDG within a human subject.

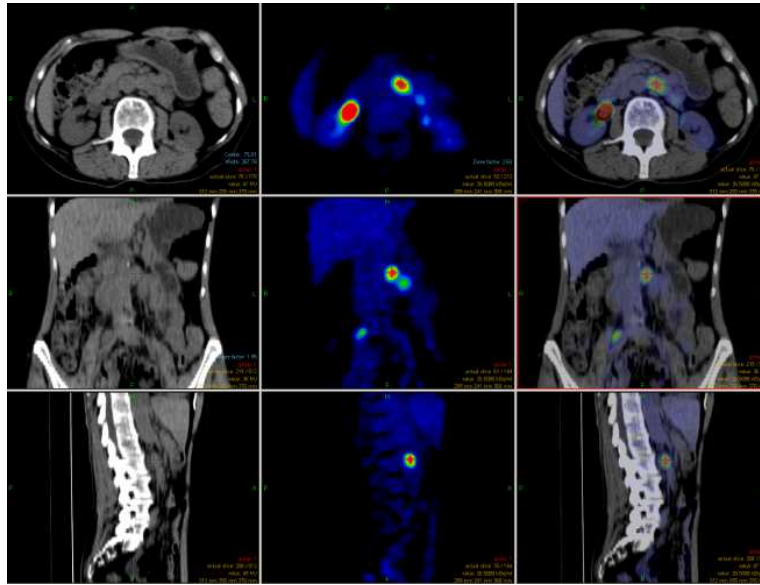


Figure 6: CT, PET and Combined Images

The first column of images shows the structural density of material in the body via conventional CT scan. The middle column shows zones of high metabolic activity via PET. The right column is an overlay of CT and PET images, permitting a greater understanding of the FDG uptake within a human structural context.

This demonstrates the key differentiator of PET, i.e. being able to map zones of defined biological activity. In addition to fludeoxyglucose, there exists a range of radionuclides which each mark a significantly different bio-process, e.g. technetium⁹⁹ is used in the study of brain activity for studies of Alzheimer's disease. Since PET systems have generally used photo-multiplier (vacuum) tubes as the detector, it means that the scanner cannot be easily integrated with other medical imaging modalities which interfere with the operation of the PMTs, such as MRI. CMOS SPADs exhibit a key performance differentiator in that their performance is not similarly impacted, and therefore provide the potential for integration of imaging types in the same apparatus as well as system miniaturisation and cost reduction.

1.5 Time Correlated Imaging System Components

Time Correlated Single Photon Counting (TCSPC) imaging systems such as FLIM generally require the synchronisation of an excitation or illuminator source with a light detector capable of resolving single photons, and a high temporal resolution timer such as a Time to Digital Converter (TDC). When these functions operate together as shown in Figure 7 it is possible to measure the arrival times of photons that are being emitted or reflected by the subject with respect to the initial illumination event.

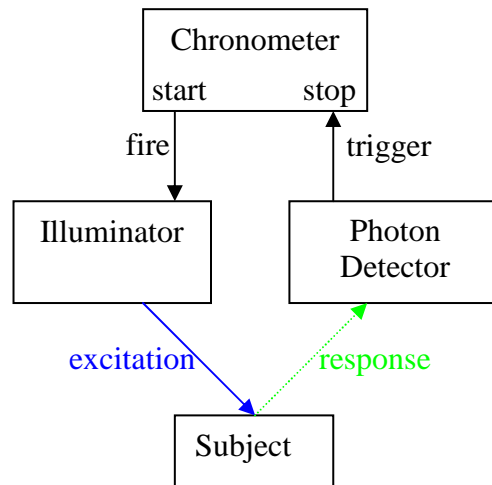


Figure 7: TCSPC System Components

It can be seen that the TCSPC system consists of three main elements: a single photon detector, a chronometer, and illumination source. Such a single detector system can be mechanically scanned over the area of interest to create a larger resolution image. A simpler and cheaper alternative system with no moving elements can be achieved with an *array* of many detector-chronometer channels.

1.5.1 Detection

The detector of choice in terms of *single photon* detection sensitivity, bandwidth and noise performance remains the Photo-Multiplier Tube (PMT). This established technology dates back to work conducted by RCA in the 1930s. The excellent performance of the PMT comes with the limitations that they are bulky and fragile with poor fill factor and spatial resolution. They also require high operating voltages in order to function at the very high gain of input photon flux to output voltage that they achieve, and suffer from aging effects. Their vacuum tube, electron steering based design makes them unsuitable for operating near strong magnetic fields commonly found in some medical equipment such as Magnetic Resonant Imagers (MRI).

A large area multi-element implementation of a PMT with improved spatial resolution is called a micro-channel plate (MCP). Commonly used in scientific instrumentation, MCPs are costly, fragile and suffer from noise and saturation effects.

The modern solid-state Silicon Photo-Multiplier (SiPM) addresses some of the weaknesses of the PMT and is a viable direct replacement for some applications (Burr [5]). However, in order to achieve adequate performance from the Avalanche Photo-Diode (APD) building block element the manufacturing process must be highly optimised. This prevents the possibility of large-scale integration with other circuitry such as time resolvers on the same substrate.

An alternative detector is the silicon Single Photon Avalanche Diode. SPADs can be traced back to the deep planar/reach through structures created by McIntyre [6], Ruegg [7] and Haitz [8] in the 1960's at RCA, Stanford and Shockley Laboratories respectively. These large, deep junction devices and their subsequent developments required large reverse bias voltages and were stand-alone structures incompatible with other circuit elements. Perkin-Elmer, Rockwell Science Center and Russian research groups have all since contributed to the development of these devices. The history is well captured by Renker in [9].

As well as silicon, SPADs have also been developed in III-V classified materials such as Indium-Gallium-Arsenide (InGaAs), or cooled Mercury-Cadmium-Telluride (HgCdTe) for military, security and search and rescue applications where infrared sensitivity is a key requirement. However, as well as providing single photon sensitivity for visible wavelength light (without cooling) the progressive

photolithographic reduction of CMOS processes have provided scope for reduction in detector sizes, improvement in fill factor and provided the possibility of integration of avalanche quenching, time conversion and data processing functions.

The basic circuit showing detector, bias supplies and passive quench component is shown in the following diagram.

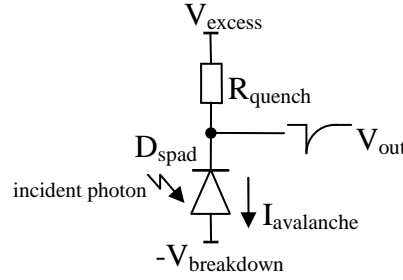


Figure 8: Basic Geiger Mode SPAD and Passive Quench Circuit

A simple inverter circuit may be integrated into the circuit shown above in order to threshold V_{out} and so output detected events straight into the digital domain. The output pulses contain the vital temporal information regarding the photon's arrival time. Alternatively, pulses may be digitally counted during a defined exposure time to generate a measure of photon intensity.

1.5.2 Chronometry

Chronometry is the science of the measurement of time. Whether the measurement is done via marks on candles or incense sticks, by the shadow the sun casts on a sundial, by clockwork or by atomic clock, man has strived over many centuries to constantly improve accuracy and resolution of this important measurement. The advent of electronics has resulted in ever more compact and accurate time measuring devices. For today's TCSPC systems an accurate, high resolution timing resolver circuit is required which measures time between two relative events; i.e. from the illumination event to the subject's response or reflected light, to tens or hundreds of picoseconds resolution. Ideally the result is output as a digital word for ease of downstream data processing.

One of the first implementations of such a function is considered to be the 'Digital Time Intervalometer', reported by Nutt [10]. This was a bulky, time-amplitude-digital implementation for atomic energy research. Digital oscilloscopes and plug-in

computer card based TDCs have lead the way in terms of performance but tend to be large and power hungry. The dramatically smaller, monolithic silicon TDCs of today can be classified into two groups: those which are based on the minimum process gate delay (e.g. Arai [11]), and those which achieve sub-gate delay (e.g. Dudek et al [12]). The design challenge is the trade off between time resolution, complexity, word width, accuracy, area and power consumption bearing in mind the constraints of a particular application.

The basic principle of a time to digital conversion is shown in the figure below.

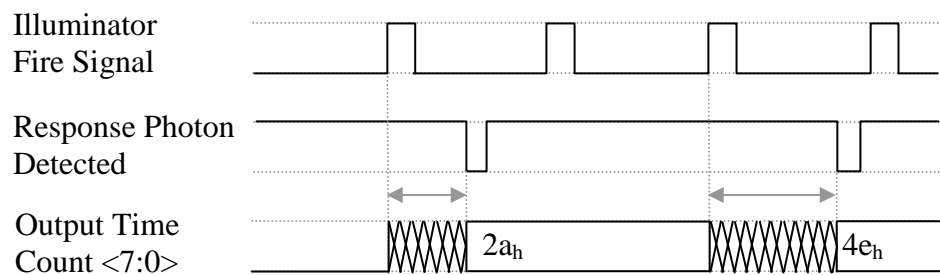


Figure 9: 'Forward Mode' Time Measurement

This shows a forward mode measurement; i.e. from illumination event to first returned photon. Reverse mode may also be employed, in which case it is the time from first photon to the next illumination event that is measured (Kinoshita [13]). This offers improved system power consumption since the data conversion is only done when a photon is received.

1.5.3 Illumination

The illumination or excitation source of a TCSPC system is application dependent. For FLIM a pulsed laser source is normally used as an excitation source. Ranging, or 3D imaging may use a laser or LED, which may be pulsed or a modulated continuous wave (CW). In PET the illumination happens as a result of positron-electron annihilation within the subject emitting diametrically opposite gamma photons or rays. The gamma rays strike a ring of photon emitting scintillation-detectors whose outputs are used to determine photon coincidence.

The illuminator is normally a separate piece of equipment, although there have been recent efforts by Rae et al [14] to implement a miniaturised illumination source on chip alongside the detector element for FLIM. For long range 3D image capture or

ranging, high transmission power is required in order to negate the effects of transmission losses, scene reflectivity and photon detector quantum efficiency.

1.6 Project Aims and Benefits

The aim of this research is the creation of a single photon avalanche diode plus time resolver in an arrayable pixel structure using a modern, unmodified nanometer scale CMOS process with a performance capability adequate for high frame rate (1Mfps) FLIM. The potential for large-scale integration of signal processing circuitry introduces a reduction in size, power consumption, cost and maximises possible manufacturing volume, when compared with existing solutions. It also allows the possibility of bringing together advances in photon detection structures with that of time measurement circuitry.

If these goals are met it could lead to the creation of affordable FLIM instrumentation for point-of-care medical equipment.

To achieve this aim, three main developments are required:

1. A single photon detector with low noise, high quantum efficiency and process compatible breakdown voltage.
2. A time resolver with picosecond level time resolution, integrated into an array with the detector.
3. A data readout sub-system for pipeline mode operation of one million frames per second of a 32x32 pixel array.

These are the topics that are discussed in detail in the relevant sections of the thesis. This constitutes the main body of the research.

1.7 Beneficiaries of the Research

The beneficiaries of the research, how it could affect them and longer term impact are important factors to consider when evaluating the possible impact of this research.

Users of ranging systems such as building surveyors, robotics builders, disabled people, the military and computer gamers would benefit from miniaturisation, cost reduction, faster data acquisition, increased range accuracy and enhanced spatial resolution. This could bring about improvements in survey efficiency, gesture recognition and may contribute to the proliferation of immersive gaming.

Scientific instrumentation users such as life science researchers would benefit from increased accuracy, parallel channel acquisition and fast fluorescence lifetime computation. This could promote more detailed understanding of intra cellular processes progress leading to new drug discoveries. Improvements in system performance also mean earlier detection of cell abnormalities, leading to better patient survival rates.

Similarly in the field of high energy particle detection, particularly positron and single photon emission tomography, medical practitioners could benefit from images of the inside of the body with increased contrast, as well as the cost reduction of the machines themselves helping to promote their availability. Better contrast images results in the earlier detection of abnormalities such as cancerous cells or reduced neural activity in key parts of the brain.

1.8 Methodologies

As this research had a fairly wide scope, several methodologies were employed during the course of the project. The main methodologies are listed below:

- Analogue integrated circuit design methodology using a mixed Cadence/Mentor Graphics tool set, allied to the ST Microelectronics process technology used for silicon fabrication.
- Test and characterisation of devices was performed using a range of electronic and optical test bench/ dark room equipment. Test methods related to key metrics are described in section 4.2.
- Data analysis and documentation was performed using standard desktop computer software.

1.9 Project Context

The research reported in this thesis has been supported by the European Community within the Sixth Framework programme IST FET (Information Society Technologies, Future and Emerging Technologies) Open Megaframe Project, Million Frame Per Second, Time-Correlated Single Photon Camera, contract No. 029217-2. Within this project the author has contributed to the work packages of The University of Edinburgh and ST Microelectronics R&D Ltd, in the period from 2006-2010.

1.10 Contribution to Knowledge

To the author's knowledge, this thesis describes the first implementation of an integrated array of low noise single photon detectors with calibrated, high temporal resolution time to digital converters in a modern nanometer scale CMOS process, capable of being scaled to large array formats. A family of new SPAD constructions called 'retrograde well' detectors have been demonstrated that exhibit low dark count and high PDE compared to other devices fabricated in comparable CMOS technologies. The design can use standard process implants and allows the implementation of large integrated arrays of detectors with fast quenching, readout and time resolving circuitry. The TDC design balances performance constraints such as area, time resolution and power consumption to create a scalable architecture that is power efficient and is robust to process, supply and environmental effects. When implemented as an array (in this initial case 32x32 pixels) converted data for every pixel may be read out every 1 μ s (1Mfps), offering time correlated ten bit data at more than a thousand individual sites in parallel.

1.11 Time Correlated Imaging System Metrics

The reader of this thesis will find many different performance parameters being discussed. This section serves as a brief introduction to the main metrics which are used to evaluate the implementations of the main TCSPC functional blocks.

Since TCSPC systems are frequently based on an interconnection of separate pieces of equipment, the metrics of detector, time to digital converter and illuminator are often grouped separately. The performance of the single photon detector used is often expressed in terms of parameters such as photon detection efficiency, timing resolution, dead time, operating voltages and dark count (noise). The time resolver or

chronometer can be described in terms of accuracy (INL, DNL), time resolution, output word length and dynamic range. The illuminator (not addressed by this work) is often described by parameters such as output wavelength, emission power, modulation frequency, pulse widths and pulse range. For all of these TCSPC system elements, area and power considerations apply. A full description of all parameters and their interdependencies is given in section 2.5.

1.12 Summary of Thesis Structure

The following chapters introduce background information, describe the implementation of the design and provide a discussion of the characterisation results.

Chapter 2: Background

This chapter describes the present state of the art of single photon detectors and time to digital converters, the metrics used to measure them, and introduces the concepts plus design challenges of integrated SPAD-TDC arrays.

Chapter 3: SPAD Design

This chapter describes the detailed design of a number of new single photon detectors, both in terms of device and circuit modelling.

Chapter 4: SPAD Characterisation

This chapter summarises the characterisation (methodology and results) of the detectors introduced in the previous chapter.

Chapter 5: TDC Design

This chapter describes the detailed design of a new time to digital converter architecture, both in terms of device and circuit modelling. Also included is a description of the data readout approach used, as well as an introduction to the hardware sub-system into which the 32x32 array device is embedded.

Chapter 6: TDC Characterisation

This chapter summarises the characterisation (methodology and results) of the converters introduced in the previous chapter. A summary of the performance of the imager integrated circuit is also provided.

Chapter 7: Conclusions and Outlook

This chapter draws overall conclusions and key findings for the research project, including a critical evaluation of what is new from the work performed. Additionally it proposes programmes of work to take the research forward whilst outlining areas for improvement.

2 Background

2.1 Overview

In the previous chapter the key elements required in a time correlated imager system were introduced: the single photon detector and the time to digital converter.

In this chapter more details of the performance metrics of these two elements are discussed in order to enable comparison of the system described in this thesis with existing ones. A literature review is given for each element. The chapter concludes with a review of existing TCSPC systems that have integrated photon detectors with time processing circuitry.

2.2 SPAD Performance Parameters

The primary role of a SPAD in a time correlated imaging system is to detect, with a high degree of precision, single photon arrivals that are correlated to a system illuminator excitation event. The effectiveness with which the detector performs this role is evaluated through a set of metrics introduced in the following section. These performance parameters are often interrelated; changing operating conditions to promote performance in one area has a detrimental effect in another.

A set of metrics which describe SPAD performance have been reported by Rochas in [15], where he discusses SPAD device operation in detail. Later, Niclass in [16] focuses on deep sub-micron implementation for 3D image sensors. The purpose of this section is threefold: to provide necessary background, discuss key points that are thought to have become particularly relevant recently, and introduce some fresh, detector oriented, observations encountered during the research.

The circuit shown in Figure 8 on page 24 shows the main control bias voltages. These are the reverse breakdown voltage, V_{bd} , and the excess bias voltage, V_{eb} . Together the magnitudes of these potentials define how far beyond the breakdown region the device is operated.

An avalanche breakdown is induced when a photon is absorbed in the high field space-charge (depletion) region of the detector, as shown in Figure 10. The absorption creates an electron-hole pair which is accelerated via the high reverse bias electric field, ξ . This carrier movement may in turn generate further electron-hole pairs, leading to a high gain full avalanche multiplication being initiated.

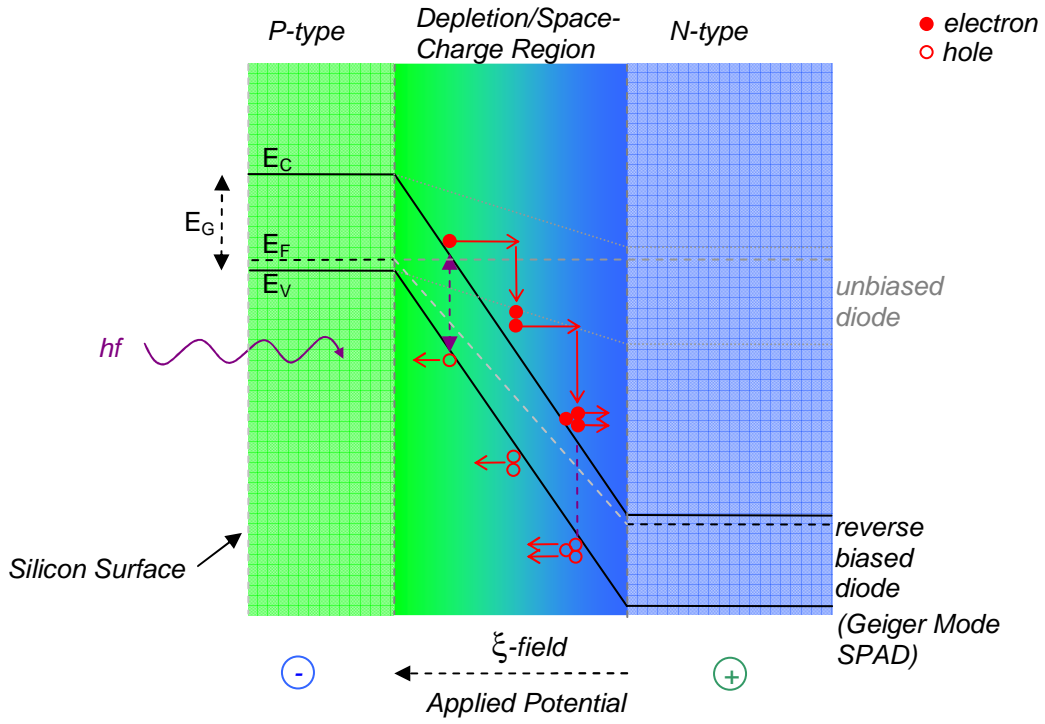


Figure 10: SPAD Energy Band Diagram

The efficiency of this absorption-multiplication process, like a standard silicon photodiode, is limited by the photon absorption coefficient, which is dependent on wavelength. This defines that incident photons with shorter wavelength (and so higher energy) have a higher probability of creating an electron-hole pair by the photoelectric effect. In silicon, at 1100nm wavelength there is only just enough photon energy to transfer an electron across the energy bandgap. This process is represented by the purple arrows in Figure 10. The photon penetration depth is essentially the inverse of the absorption exponential profile, and is generally defined as the mean depth in silicon at which 37% of photons of a specific wavelength are absorbed.

The main performance parameters for single SPADs are dark count rate, afterpulsing, photon detection efficiency, dead time, and timing resolution. When considering arrays of SPADs, detector crosstalk and fill factor are also particularly important. These are detailed in the following sections, assuming room temperature unless otherwise stated.

2.2.1 Dark Count Rate

The dark count rate (DCR) of a SPAD can be thought of being analogous to the dark current of a conventional pixel or APD. A dark count is caused when a non photo-generated carrier enters the high field region of the detector and causes an avalanche event to occur. Non photo-generated carriers may be generated by four factors: diffusion, thermal activity, band to band tunnelling or by release from a charge trap as reported by Haitz in [17]. These factors are illustrated in the silicon diode energy band diagram below.

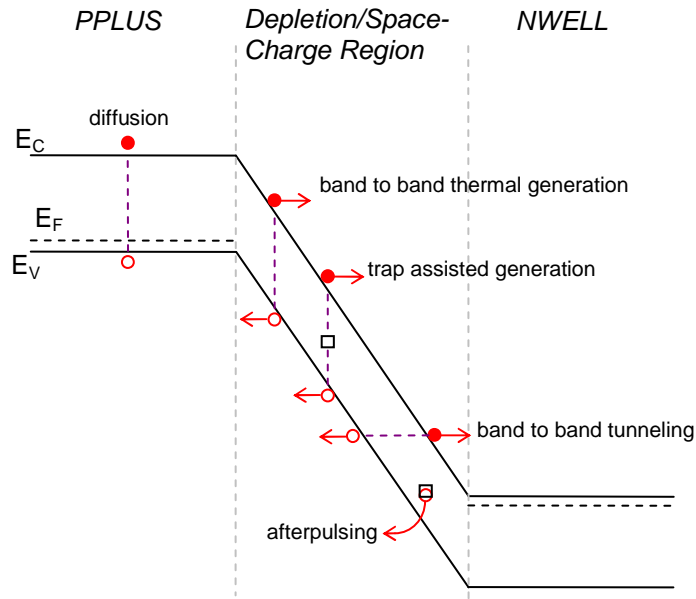


Figure 11: SPAD Dark Count Sources

Carriers generated by any one of these means may result in a non-photon induced avalanche. The per-second rate at which dark count occurs is the DCR in Hertz. The mean level of DCR sets the lower limit of the photon arrival rate that the SPAD is able to detect, i.e. the noise floor. If measured by gathering enough counts in a dark environment, the mean level may be easily calibrated out. Therefore, it is the statistical variability of DCR that is the SPAD's main (gaussian) noise source, Niclass [16]. Dark count would preferably scale with area (i.e. squared with respect to device radius), but in real implementations is generally reported as exhibiting a steeper than expected exponential profile. This is reported by Zanchi et al to be due to the tendency for trapping defects to be centrifugally drawn toward the device perimeter,

[18], in a predictable manner yet peculiar to the manufacturing process employed. Zanchi also reports on the benefit of phosphorous implant based gettering DCR improvements.

DCR increases exponentially with temperature and therefore may be reduced by using cooling methods such as thermoelectric Peltier elements or by forced air-cooling. DCR also varies linearly with the electric field strength [17] so may be reduced by constructing the detector with implants of lower doping concentration or configuring the detector with lower overall bias operating conditions. Whilst effective, the latter method also reduces the overall photon detection efficiency. The expression for the total DCR due to thermal and tunnelling generation is provided by Rochas in [15]. The reduction of charge traps within the silicon crystal lattice structure may be implemented by methods such as using gettering implants or by lengthening process annealing steps.

The goal of this research was to create a detector with average DCR equivalent to that of state of the art custom SPADs of $\leq 100\text{Hz}$. Recent detector implementations in sub 250nm CMOS processes (Finkelstein [19], Niclass [20], Faramarzipour [21] exhibit high dark count rates (60KHz-1MHz). Fabrication processes at this node target the implementation of small, high-speed transistors for large-scale integration of digital systems. This requires shallow implants with high doping concentrations. When these implants are used to implement SPADs the result is a junction with a high electric field, and the result is high DCR dominated by band-band tunnelling. The impact of high dark count rate in a TCSPC system is an unacceptable level of incorrect time measurements. Therefore many averaged measurements must be done to remove the impact of this component, which in turn extends the overall analysis time. It can be appreciated that for high sample rate, low light applications such as FLIM, DCR should ideally be as low as possible.

2.2.2 Afterpulsing

Impurities unintentionally introduced into the wafer manufacturing process can cause trap states in the silicon crystal lattice structure. These result in generation-recombination (GR) centres that can exist in the forbidden zone between valence and conduction bands. The high current peak through the junction during an avalanche breakdown introduces a probability that the trap will be filled by a carrier which is then later released, initiating a second, follow-on Geiger ‘after-pulse’ [17].

Trap occupancy has an associated lifetime. Traps located toward either valence or conduction band have a short occupancy lifetime and therefore do not contribute heavily to afterpulsing. However, traps located halfway between the two bands, called ‘deep traps’ have longer lifetime and are therefore a major contributor. For this reason manufacturing processes should be kept as clean as possible, as well as minimising charge flow via careful detector design and the use of active quench circuits. Minimising charge flow has the knock-on undesired effect of limiting photoelectric gain, which should be maximised for applications that utilise the non-buffered version of the detector output pulse (such as analogue output silicon photo-multipliers with typical gains of 10^5 - 10^6) with a downstream ADC.

2.2.3 Photon Detection Efficiency

A SPAD’s photon detection efficiency (PDE) is the equivalent to the quantum efficiency (QE) of a conventional photodiode. The QE of a photodiode is the efficiency with which the structure creates (and stores for later readout) photo-generated charge with respect to the incoming photon flux, over an incident light bandwidth. For a SPAD, the PDE is the percentage of incoming photons that create an output pulse, again over an incident light bandwidth. An output pulse is initiated by a charge carrier being photo-generated in the high field active region of the SPAD. The carrier is accelerated through the junction, causing secondary carriers to be created by impact ionisation. This multiplication process may continue to create a full discharge event and output pulse.

The probability of a photon arrival causing an output pulse is reduced by three main factors: reflectance, absorption in the optical stack, and self-quenching.

Firstly, an incoming photon may be reflected at the surface of the device or at the interface between the many layers that constitute the optical stack of the detector. These layers can sometimes have quite different refractive index (RIs), which exacerbates the problem.

An antireflection top coating is a common technique used to reduce surface reflections. To maximise photon transmission through the optical stack, process developments such as stack height reduction and layer RI optimisation have been done by vendors of CMOS imaging processes (Cohen et al [22]).

Secondly, a photon may be absorbed above the SPAD within the optical stack materials, just at the surface of the active region, or too deep within the silicon in order to initiate an avalanche.

Thirdly, an avalanche event may be initiated but stall, becoming self quenched. Such an event may not yield enough potential difference in order to trigger an output pulse. In this case these small pulses are not visible externally. This can be addressed partly by careful readout design, positioning the detection threshold just at the onset of avalanche. Self-quenching can be minimised by ensuring a high enough electric field is applied. This ensures that a photo-generated charge has high enough energy to be accelerated through the device structure, resulting in increased chance of impact ionisation taking place.

The shallow implants of nanometer scale CMOS processes result in peak efficiencies in the blue zone of the visible light spectrum (Niclass [20]). Whilst non-ideal for applications necessitating NIR (IR-A band, 700-1400nm) wavelengths such as ranging and 3D cameras, this is well suited to the emission wavelength of commonly used scintillator materials for PET such as '*Cerium doped Lutetium Yttrium Orthosilicate*' (LYSO) which peaks at 420-445nm. This indicates a key reason that SPADs are well suited to the replacement of PMTs, which are commonly used with scintillators for high-energy particle detection.

Previously published SPADs and those reported in this body of work are sensitive to NIR wavelengths, albeit at a reduced PDE of <10 % (e.g. Niclass [20]). Therefore it is commonplace to see IR-A based applications using ultra narrow pass band filters in order to suppress background ambient light.

For FLIM, commonly used green fluorescing proteins emit around 560nm, where the same previously reported detectors have ~25% PDE. With the arrival of synthetic Quantum Dot (QD) based fluorophores, the system may be tailored to be excited by and emit photons at defined wavelengths suited to the detector being used.

2.2.4 Timing Resolution

When a SPAD is struck repetitively with a time accurate photon source, the position in time of the resulting avalanche breakdown pulse has a statistical variation. The timing resolution, or 'jitter' of the detector is the full-width, half-maximum (FWHM) measure of this temporal variation. The factors contributing to timing jitter were a focal point of Ghioni-Cova in [23].

The timing resolution of the detector itself has two main components. The first component is the variation caused by the generated carrier transit time from depletion layer to multiplication region, which is dependent on the depth of absorption of the incident photon. The guideline commonly used to estimate the transit time at carrier saturation velocity is 10ps per micron.

The second, larger component is the statistical build up of the avalanche current itself. This is impacted by the electric field strength, and so jitter may be minimised by employing high overall bias conditions. In some structures there may also be timing uncertainty introduced by lateral propagation, which has been reported as exacerbated in larger area SPADs (Lacaita et al [24]). However, in other structures, notably the STI guard ring structure by Hsu-Finkelstein et al [25], there is no reported area dependent component to the timing resolution. This implies that in the cases where quasi-neutral field areas have been minimised lateral propagation does not affect timing resolution.

The shape of the histogram of avalanche events in response to a time accurate photon arrival provides information regarding the location and speed of avalanche build up. A predominantly Gaussian shape indicates that the bulk of photon initiated avalanches occur in the high field active region of the detector. The timing response of several published structures often exhibits a long tail, indicating that photon-generated carriers diffusing into the high field region of the detector after a short delay initiate a proportion of avalanche events. For this reason this part of the response is called a diffusion tail, and means that measuring the timing response at full-width 100th of maximum is a valid point of measurement.

Several attempts have been made to minimise detector diffusion tail, most notably Lacaita-Cova [24], Ghioni-Cova [23]. It should be noted that the incident light wavelength also has a bearing on diffusion tail, i.e. longer wavelength photons can also take longer to enter the high field region due to deeper absorption.

The method employed for detecting the onset of an avalanche event is of high importance. This may be done by external thresholding just on the edge of breakdown (e.g. by constant fraction discriminator, or CFD) in which case there is no time walk. However, routing the SPAD's moving node directly off chip introduces increased parasitic capacitance. This increases the charge flow during operation of the detector, leading to increased charge trap filling and so increased afterpulsing probability and dark count. The integration of a readout element such as a buffer, inverter or source

follower with the detector obviates these problems. However, the buffer and inverter solutions whilst delivering an output pulse direct into the digital domain have a fixed threshold, defined by the relative sizes of the PMOS and NMOS transistor elements within. This requires careful optimisation so as to deliver CFD level performance. Without optimisation of this cell, the measured jitter can be much higher than the detector element itself ($>4\times$, reported by Hsu et al [25]). The problem is further exacerbated by driving SPAD outputs off chip via wire bond pads, requiring multi-stage buffers capable of driving the external load capacitance of a bench characterisation setup. Indeed, the different test bench configurations reported may themselves play a part in the variation of reported results.

From an application viewpoint, high contrast FLIM images demand better timing resolution. The timing resolution should clearly be as small as possible, of a Gaussian shape, free of diffusion tail effects and any impact of readout mechanisms.

2.2.5 Dead Time

The dead time is the period of time from the moment of impact ionisation, through the avalanche quenching process until the bias conditions are reset to 90% of the final steady state potential as first modelled by Haitz in [26], and discussed further by Rochas in [15] (shown in Figure 21 on page 53). Theoretically the SPAD is not responsive to further incoming photons during this entire period. However, in the case of a passively quenched SPAD this is not strictly the case.

The classic Haitz reverse biased diode Geiger mode model of [26] shows that although the initial breakdown event is a fast build up of current related to the excess bias voltage and internal resistance, the bias potential conditions are reset via the RC time constant of the SPAD capacitance and the large passive quench resistance. This period of recharge can be configured as several tens or hundreds of nanoseconds. As the device is recharged it becomes increasingly more biased beyond its breakdown voltage and so experiences an accompanying increase in junction electric field and photon detection probability. This renders the device able to detect the next photon arrival prior to being fully reset. This behaviour is coupled with a significant fluctuation in the reset waveform. Clearly the dead time should be kept as small, and as consistent as possible in order to achieve the highest possible dynamic range of incident photon flux and least variation in photon count output to a certain photon arrival rate.

For these reasons active quench and recharge circuits are often used. These can be fully active circuits or hybrid implementations of passive and surrounding circuit elements. This topic is discussed in detail in section 2.3.5. Regardless of quenching mechanism employed, the SPAD capacitance should be kept as low as possible in order to reduce the dead time and charge carrier flow. However, short dead times are often accompanied by enhanced afterpulsing probability due to inadequate trap flushing time. The probability of a trap becoming occupied is correspondingly higher with an increased charge carrier flow.

2.2.6 Crosstalk

Crosstalk between adjacent SPADs in an array configuration can occur in two ways. Firstly, a photon absorbed deep in one detector construction may result in a lateral diffusion of carriers to an adjacent device where an avalanche can be initiated.

Secondly, an avalanche event may result in an electro-luminescent emission of photons that are then detected by an adjacent detector. This is captured by Lacaita et al in [27].

Electrical and optical crosstalk can be minimised by detector design. Measures can be taken to limit internal reflections along the optical stack between pixels, as well as taking advantage of SiO₂ shallow trench isolation (STI). Shallow versions of this technique are a feature of modern CMOS processes for the purposes of electro-optical isolation and latch up immunity. Deeper, more effective trenches are a feature of full custom processes such as that used by Sciacca in [28] and by Kindt in [29].

Crosstalk can be measured by performing a correlation analysis of adjacent detector output responses under dark conditions using the mathematical functions available on a digital storage oscilloscope.

2.2.7 Active Area, Fill Factor and Breakdown Homogeneity

The active area of a SPAD is the central photon-sensitive portion of the detector. The diode junction's electric field strength should be consistent across this part of the structure. Zones that exhibit higher field strength compared with the rest of the active area will exhibit locally higher photon detection probability and dark count. Such a zone can exist due to inadequate guard ring design or a physical feature of the structure such as a sharp corner. Therefore the active region must have a homogenous

breakdown probability. The proportion of active region area to total SPAD area is the fill factor and is commonly expressed in percent.

Whilst active area and fill factor are a function of the drawn layout of the structure, the breakdown probability over the active region is normally analysed by capturing the photo-luminescence that is emitted by-product of the charge carrier flow during dark count avalanche breakdown. This is normally captured in a light controlled environment using a high sensitivity CCD camera configured with long exposure time.

One of the goals of this research was to analyse the relationship between increasing active area and the resultant impact on dark count. Clearly increasing active area has an impact on noise, but defect probability is increased accordingly. Understanding the relationship between these parameters has enabled guidelines to be drawn up regarding the trade off between fill factor and noise. This is reported in section 4.4. Additionally, work has been undertaken to maximise fill factor of the overall detector footprint via trials of detector shape and active region area. Part of these trials has been the optimisation of the active region electrical connection, in terms of both minimising active region light blockage and metallisation defectivity introduced during processing of the anode contacts.

2.3 SPAD Literature Review

The aim of this research is to integrate arrays of low noise single photon detectors with time resolving circuitry in a deep sub-micron manufacturing process. The review that follows focuses on previously published state of the art SPADs.

2.3.1 Background

Time correlated, single photon detection systems based on photo-multiplier vacuum tubes, although exhibiting very low dark count rate and high sensitivity tend to be bulky, fragile, have poor spatial resolution and require high DC voltages for operation. Multi-Channel PMT based devices such as Micro-Channel Plates (MCPs) are also generally very costly items.

Time gating of CCD cameras is a commonly used alternative technique but suffers from temperature dependent dark current effects, is wasteful of photons, results in long acquisition times and suffers from resulting sample bleaching. For high accuracy resolution of multi-exponential sample lifetimes, simultaneous gate TCSPC is required (W. Becker [30]). SPADs are well suited to low photon flux level, visible wavelength applications such as FLIM and offer a silicon based solid-state solution.

To challenge the established technologies, SPADs with low dark count rate, high quantum efficiency and high fill factor are required in a modern, dense, nanometer scale CMOS process.

2.3.2 SPAD State of the Art

SPADs are a specially constructed avalanche photodiodes capable of operation beyond their reverse breakdown voltage. Operated in this mode they behave as photon arrival trigger mechanisms via photon induced impact ionisation. This is called ‘Geiger Mode’ operation due to the counter-like behaviour.

A guard ring structure prevents premature edge breakdown of the active region, allowing the detector to operate with single photon sensitivity for a period of time when operated beyond breakdown, as well as promoting homogenous breakdown probability over the device photosensitive area.

Operated in Geiger mode they require a suitable *quenching* circuit for detecting and halting avalanche current flow, and the resetting of bias conditions. This may be done

passively by a resistor or linear mode transistor, or actively via a dedicated companion circuit.

It may be considered that SPAD prior art may be described in terms of:

1. Manufacturing process employed.
2. Constructions used.
3. Quenching methodology.

The following sections discuss each of these factors in turn.

2.3.3 Process Choice

The family tree diagram shown below represents the evolution and different branches of SPAD research, classified by manufacturing process.

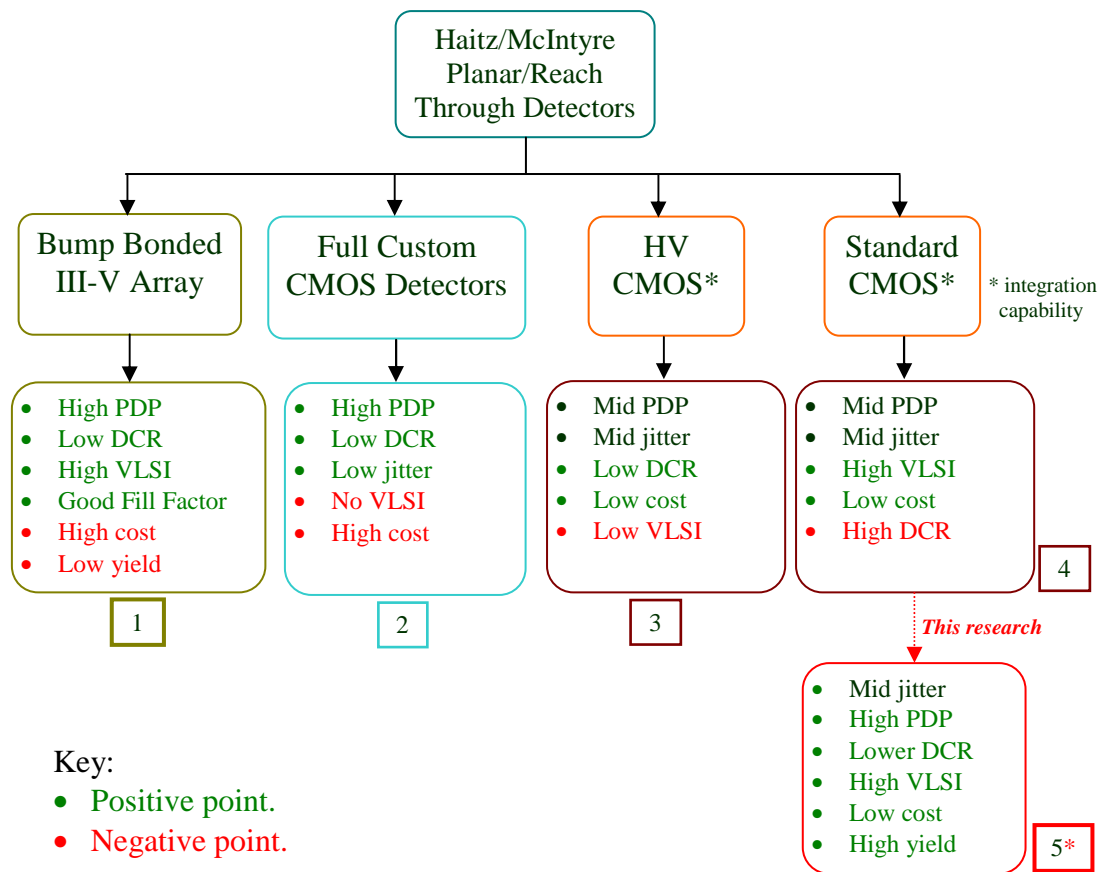


Figure 12: SPAD Research Areas Classified by Process

There exists a distinct group of sensors (track 1) implemented in non-CMOS ‘III-V’ materials (i.e. elements from groups 3 and 5 in the periodic table, e.g. InGaAs). These devices tend to be used for high end, sometimes cryogenically cooled applications

such as bio-med/phys high-energy particle detection [31],[32] and military applications [33, 34] requiring good IR sensitivity. A composite approach of optimised detector array on top of a CMOS backplane has been used with success by SensL/Caeleste and the Medipix consortium [35] for high energy particle detection (not employing SPADs) such as X-ray. It is thought that this approach whilst having very good fill factor may be comparatively costly to produce and would be difficult to yield well in high volumes.

The second approach (track 2, e.g. Prof. Cova of Politecnico di Milano) has been the optimisation of a CMOS manufacturing process to yield the best possible performing single detector element (e.g. Lacaita, Cova et al [24]). This is done via the use of customised, low implant doping concentrations and slow diffusion and annealing steps to minimise silicon lattice damage and hence reduce the quantity of charge traps. Gettering phases and embedded constructions have also been developed in a bid to minimise impurity concentrations hence controlling non-photon induced triggering (dark count) and detector optical crosstalk. The nature of these implementations means that very large-scale integration (VLSI) of on chip circuitry such as time resolvers is not possible.

In the third track, high voltage CMOS processes have been used with success to implement both single detectors, and arrays together with quench circuitry (Rochas et al [36]) and single channel TCSPC systems (e.g. Tisa et al [37]) but the scope for large scale integration is limited compared with track 4.

Track 4 (e.g. Prof. E. Charbon of EPFL) has been to utilise commercially available CMOS processes without any modifications to the layers normally available to the designer. This clearly provides a large potential for large-scale integration and economic system on chip (SoC) manufacture, thus enabling new applications. The potential for arrays of detectors each with dedicated avalanche quenching and readout circuitry is evident. However, the limitations imposed by the shallow implant depths, high doping concentrations and design rule restrictions in advanced manufacturing processes have led to narrow depletion width devices being reported with high DCR (Niclass, Gersbach et al [38, 39]) most likely due to band-band tunnelling. Similar detectors implemented by other groups exhibit low photon detection efficiency (PDE) (Faramarzpour, Marwick [21, 40]), possibly due to the lack of optical stack optimisation in the chosen foundry.

Shallow Trench Isolation (STI) is also a feature of modern CMOS processes; typically below $0.25\mu\text{m}$. STI is an oxide filled trench that is etched from the wafer surface for the purposes of enhancing electrical isolation between transistors and minimising latch up probability. It also improves transistor packing density when compared with older techniques such as ‘locos’. Fill factor and optical isolation improvements via the use of STI based guard rings have been the research focus of UCLA (Finkelstein et al [19]). Whilst this work has successfully addressed pixel pitch, the devices still suffer from dark count rate as high as 1MHz, possibly caused by etching-induced crystal lattice stresses and charge trap quantity as well as tunnelling.

One of the main elements of this work is to address the detector performance issues suffered by previous authors whilst enduring minimal penalty on other device metrics. This intention is represented by track 5, outlined in red in Figure 12.

2.3.4 Detector Construction

This section introduces the construction of state of the art SPADs in CMOS technologies. Constructions suited to array implementation in nanometer scale fabrication processes are of particular relevance to this work.

Single photon avalanche diodes were born out of the early work done by Haitz and McIntyre. Their research focussed on the study of avalanche photodiodes. More specifically, the well-referenced models of an APD were proposed in [6] [26], Geiger mode pulses were observed, and proposals for sources of dark count and afterpulsing were reported [17]. Premature edge breakdown was addressed by an implant positioned at the edge of the junction active region, and hence the first SPAD guard ring was created.

The essential construction features of a SPAD are the method of formation of the guard ring, the overall shape (i.e. circular, square, others), active area diameter and the diode junction itself. Today’s nanometer scale processes provide features such as deep well implants and shallow trench isolation that may be utilised in detector design. Additionally some custom processes provide features such as deep trench isolation, buried implants, and scope for optical stack optimisation.

State of the art SPAD construction can be grouped according to the method of implementation of the guard ring structure, as shown in the following sections.

2.3.4.1 Diffused Guard Ring Structure

The first diffused guard ring SPAD was implemented by Haitz et al for the purposes of investigating microplasmas in p-n junctions under avalanche conditions [41]. Microplasmas are micro-defects that cause high field concentration. The design intention was to reduce the high probability of microplasmas existing at the APD periphery by implanting a lower doped, deeper implant in that area, so reducing the local electric field strength. Although the p-n junctions studied were not intended as photo-detectors, the addition of this guard ring structure had the additional effect of enabling Geiger mode operation, also observed by Haitz. This construction has since been implemented by several research groups, notably by Cova et al [42], Kindt [43], Rochas et al [44] (well illustrated), and Niclass et al [20].

The construction cross section is shown in Figure 13:

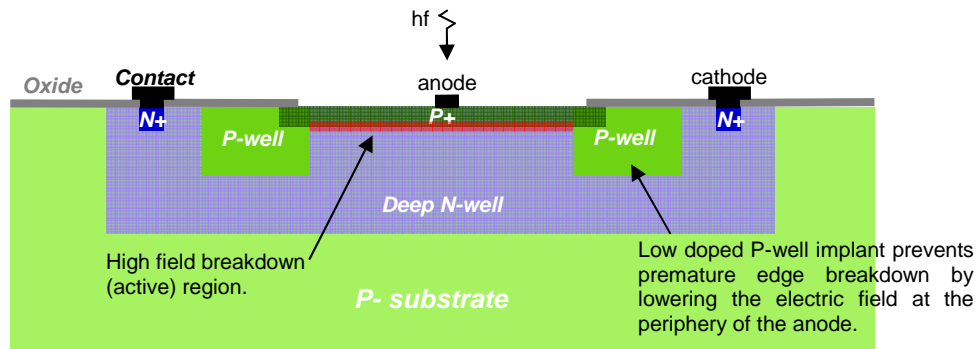


Figure 13: Diffused Guard Ring SPAD

Whilst enabling a low breakdown voltage SPAD using implants that are commonly available in most CMOS processes, this structure has several limitations.

Firstly, when implemented in a modern nanometer scale process, the associated high doping concentration, shallow implants lead to a high electric field structure. This results in a high dark count, band-band tunnelling dominated structure as predicted by Haitz [17], confirmed by Lacaita, Cova et al [24] and also reported recently by Niclass et al in [20] in a manufacturing process similar to that proposed for this research.

Secondly, if long thermal anneal times are employed in relation to the guard ring implant, the resultant field curvature around this key feature creates a non-uniform, dome-shaped electric field profile, peaking at the centre of the device. This in turn

implies a breakdown voltage variation across the active region, which strongly affects the homogeneity of the photon detection efficiency, Ghioni et al [45].

Thirdly, the increase of the quasi-neutral field region at the detector edge promotes late diffusion of minority carriers into the central high field region, resulting in a long diffusion tail in the timing resolution characteristic, first reported by Ghioni, Cova et al in [23], and a long term focus of this particular group's research activities (see following section on timing-optimised structure designs). This behaviour is also evident in the 130nm implementation of [20], Niclass.

Fourthly, this structure has a minimum diameter limitation due to merging of the guard ring depletion region as the active region is reduced, illustrated by Faramarzpour et al in [21]. This limits the scalability of the structure for array implementation purposes.

2.3.4.2 Reach-Through, Backside Illuminated Structure

The reach-through structure (R-APD) was implied by the work done by McIntyre on microplasmas and device modelling [6], explicitly by Ruegg [7] & by Petrillo & McIntyre in [46]. A cross section diagram is shown in Figure 14.

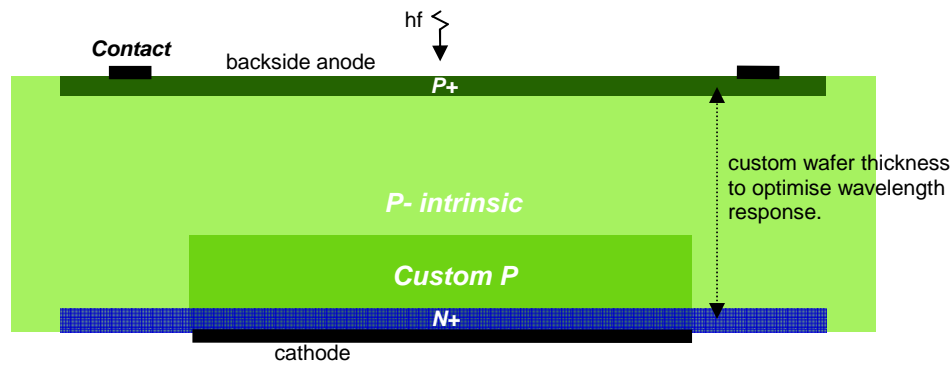


Figure 14: Reach Through SPAD

Implemented in a full custom technology, the junction is formed between a backside anode and front-side cathode, with a P-type field enhancement implant. The width of the depletion region is large and therefore the breakdown voltage is very high, typically hundreds of volts. The broad drift region provides high photon detection efficiency over a large incident wavelength bandwidth. The wide depletion region extending vertically through the device results in minimised neutral field regions and

therefore this structure exhibits no significant timing diffusion tail, with no requirement for lateral breakdown protection (i.e. guard ring).

Although this structure has been implemented commercially, the large breakdown voltage means that the current discharge during a breakdown event is high, resulting in high power consumption and localised heating. Therefore the devices tend to be operated with thermo-electric cooling elements and can be fragile, leading to unreliability. Further, the non-planar custom process and high operating voltages prevent co-integration of other desirable components such as quenching circuitry.

One of the key points regarding the reach-through structure is the first introduction of a field enhancement implant.

2.3.4.3 Enhancement Mode Structure

The enhancement mode structure has its roots in the enhancement implant employed by Petrillo & McIntyre in the reach-through structure of [46], and the hybrid diffused guard ring/enhancement structure of Ghioni, Cova et al in [23]. In 1989 Lacaita, Cova et al then removed the diffused guard ring structure in [24], relying on a single central active region enhancement implant (with doping polarities reversed, and embedded in a dual layer P-epistrata). This is referred to as a ‘virtual’ guard ring structure and is shown in Figure 15.

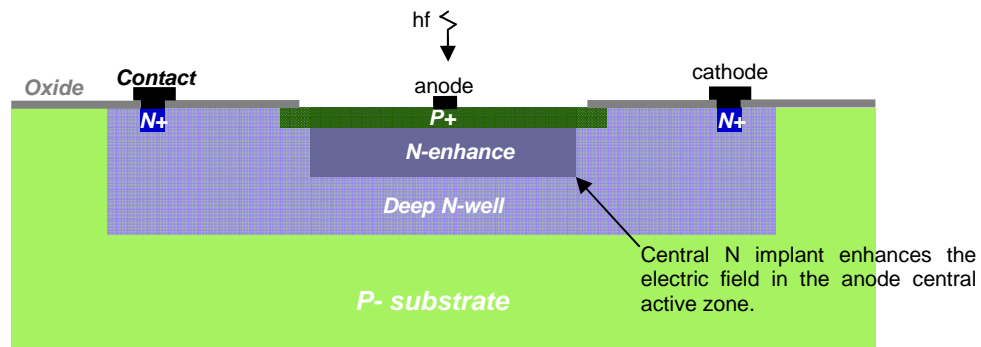


Figure 15: Enhancement Mode SPAD

The benefits of this structure are significant. Firstly, the quasi neutral regions surrounding the guard ring are removed, and therefore the minority carrier diffusion tail is reduced accordingly resulting in improved timing resolution. Secondly the device does not suffer from depletion region merging when scaling down the active region diameter, easing the prospect of array implementations with fine spatial

resolution. Both versions were repeated by Pancheri but with dual orientation (refer to section 2.3.5.3), range of active areas and active quench circuits in a high voltage technology in [47].

2.3.4.4 Merged Implant Guard Ring

An alternative implementation to the diffused guard ring is the idea of the implant merged guard ring introduced by Pauchard et al in [48] and characterised in Geiger-mode by Rochas et al in [49], shown in Figure 16. The lateral diffusion of two closely spaced N-well regions positioned at the active region periphery creates a localised low field region, preventing edge breakdown. This technique was used to ensure both device terminals were isolated from the substrate using only the standard layers available in the process.

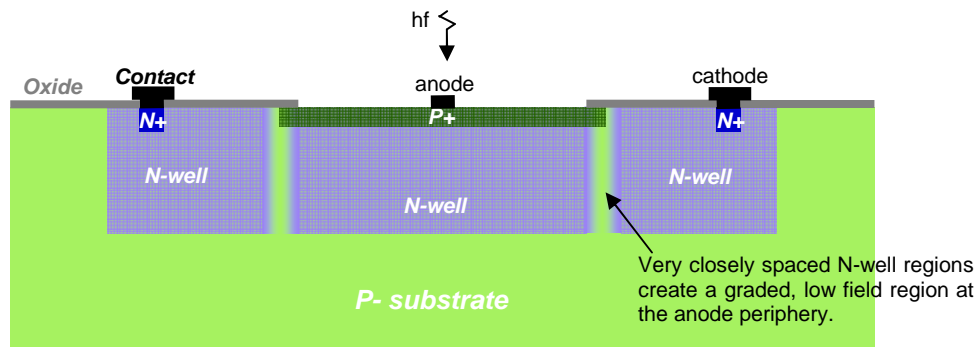


Figure 16: Merged Implant Guard Ring SPAD

Whilst successful with 50Hz DCR, >20% PDE and 50ps timing resolution being reported, the design normally violates the standard design rules and is difficult to implement. Nevertheless the authors were successful co-integrating quenching circuitry and forming small arrays in a 0.8 μ m CMOS process.

2.3.4.5 Gate Bias and Floating Guard Ring Structures

Two further guard ring ideas have been reported that are worthy of mention. Firstly, a metal or polysilicon control ‘gate’ can be added and biased appropriately to control the depth of the depletion region in the zone immediately beneath, discussed by Rochas in [15]. Alternatively an implant may be inserted near the edge of the active region in order to lower the electric field around the anode periphery. This implant is

not driven to a defined potential and therefore this construction is referred to as a ‘floating guard ring’ construction, reported by Xiao et al in [50], shown in Figure 17.

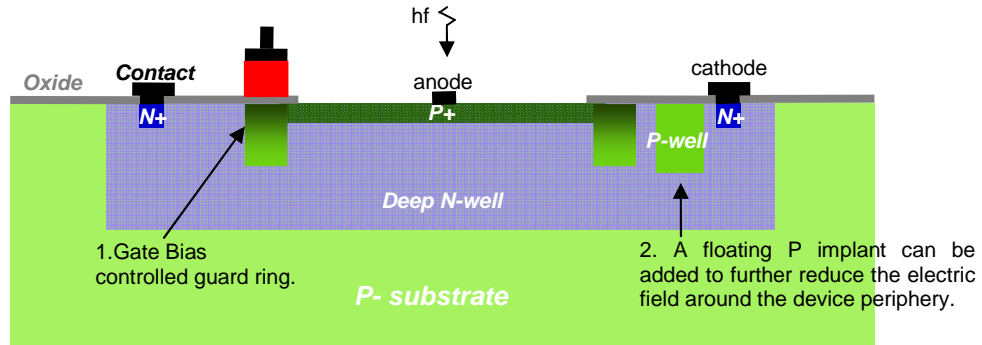


Figure 17: Gate Bias and Floating Guard Ring Constructions

The gate bias method is relatively unproven compared to the more common guard ring implementations. This is an interesting solution due to the common use of polysilicon to move STI out of the active region, i.e. the element exists as an unused control element in several structures already, and is not a widely reported construction despite it’s similarity in operation to a standard MOS transistor.

The floating guard ring construction has parallels with the diffused guard ring construction and is published by the same research group at EPFL. It lends itself to be employed with certain anode implant depths, can be tricky to determine the best layout, and can be area inefficient. However, implemented in a high voltage technology, Xiao et al report SPADs with both low DCR and good timing resolution, albeit with a high breakdown voltage and an off-chip active quench circuit.

2.3.4.6 Timing Optimised Structures

The work of the research group at Politecnico di Milano has prioritised timing resolution as the key performance metric since early publications by Cova et al utilising the diffused guard ring structure [42]. This focus was maintained in the progression to devices implemented in a custom epitaxial layer by Ghioni, Cova et al in [23]. It was observed that previously published devices were plagued by a long diffusion tail in the timing response. This was due to minority carriers generated deep in the quasi-neutral regions beneath the SPAD reaching the depletion layer by diffusion. The single epitaxial layer devices helped to greatly reduce the diffusion tail by drawing away deep photo-generated minority carriers via the secondary epitaxial-

substrate diode junction. This technique was taken further by Lacaita, Cova et al in the double-epitaxial structure of [24], and again in the more complex structure by the same author group in [51]. This work is summarised in Figure 18.

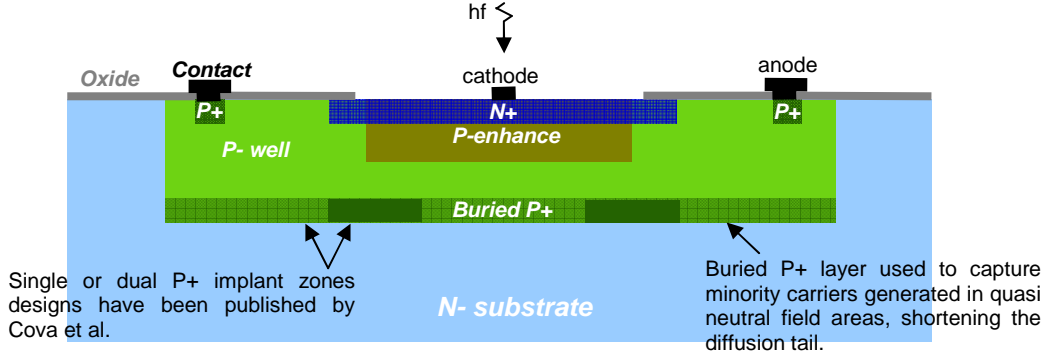


Figure 18: Timing Optimised SPAD

The goal of the ‘double epitaxial’ layer design was to reduce the thickness of quasi-neutral region below the SPAD in order to limit the diffusion tail whilst maintaining a high enough electric field to provide fast response without a high dark count penalty. In the more complex structure published [51] the buried P+ layer was interrupted underneath the active region in order to locally fully deplete the main epitaxial layer by reverse biasing the substrate, for the purposes of eliminating diffusion carriers. Whilst resulting in unprecedented timing performance (35ps FWHM) with DCR of a few hundred Hertz for a 20 μ m diameter structure, the design required full customisation of the manufacturing process, resulting in limited co-integration capability.

2.3.4.7 STI Guard Ring

The shallow trench guard ring structure was first introduced by Finkelstein et al in 0.18 μ m CMOS in [19]. The main goal of this innovation was to increase fill factor and allow fine spatial resolution. The etched, oxide filled trench feature that is a feature of deep submicron processes is used as a physically blocking guard ring, so containing the high field zone in the active region, as shown in Figure 19.

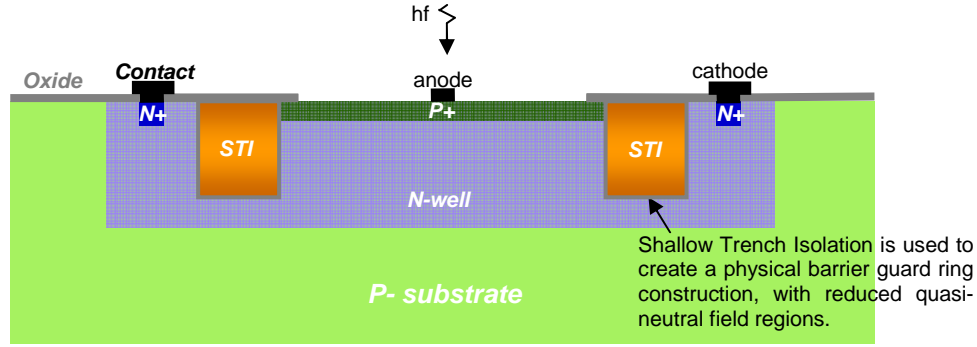


Figure 19: Shallow Trench Guard Ring SPAD

This structure was successful in addressing fill factor and potential pixel pitch, although only single devices were reported. However, the subsequent publication by the same author group, [52], revealed a very high DCR of 1MHz for a small diameter 7 μ m device. This was possibly caused by etching-induced crystal lattice defects and charge trapping associated with STI, as well as band-band tunnelling through the conventional P+ to N-well diode junction. The same author group noted in [25] that despite the high dark count the timing resolution characteristic was unspoiled by a diffusion tail due to reduced quasi-neutral field regions associated with implanted guard rings. Further, it was observed that increasing the active region diameter had no effect on the 27ps timing resolution, suggesting that these structures do not suffer from the lateral avalanche build up uncertainty postulated by Cova et al. Additionally, the lower junction capacitance yields reduced dead time.

There are two further related publications associated with the use of STI in SPADs. Firstly in [20] Niclass et al avoided the clash of STI with the sensitive active region by drawing ‘dummy’ polysilicon to move the etched trench to a safe distance away. Secondly, Gersbach et al took this on a stage further, applying a low doped P type passivation implant around the STI interface. However, DCR is still high with these structures at best 80kHz for an 8 μ m active region diameter and 1V of excess bias voltage. Timing performance remains acceptable for many applications at ~140ps.

2.3.5 Quenching

In this thesis ‘quenching’ is the term given to the process of detecting and stopping a SPAD avalanche event, followed by a resetting of bias conditions. The duration of this process is called the detector dead time, discussed in section 2.2.5 . For clarity, the dead time is defined as the period of time from the moment of impact ionisation, through the avalanche quenching process until the bias conditions are reset to 90% of the final steady state potential, Rochas [15]. This cycle of operation is shown in Figure 20.

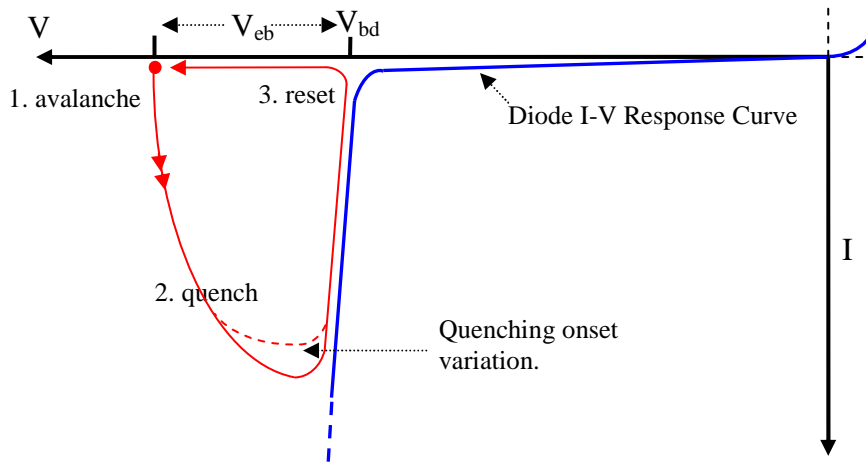


Figure 20: SPAD Cycle of Operation

The quenching function can be implemented using either passive or active components, or a combination of both. Quenching and reset may be performed by the same circuit element, as in passive quenching, or each phase may be performed by separate circuit elements as in active quenching and hybrid implementations. Efficient quenching implementations are those in which the flow of charge through the detector junction is well controlled, yielding short dead time, consistent photon detection probability with minimal impact on afterpulsing and timing resolution.

The following sections discuss the relative merits of passive versus active quenching.

2.3.5.1 Passive Quenching

Passive quenching is the most elementary approach, adding only one additional circuit component. The same passive component is used for both quenching and reset functions. This element is normally integrated on chip as either a doped polysilicon resistor or the non-linear resistance of a gate controlled MOS transistor. The advantages and disadvantages of this method can be demonstrated by analysing the equivalent circuit. The Haitz Geiger mode avalanche diode model of [26] is still valid today, and is shown integrated into the passive quench configuration below.

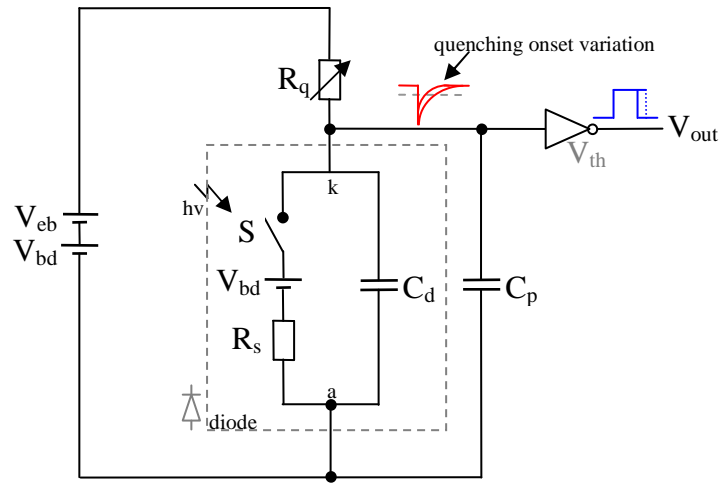


Figure 21: SPAD Passive Quench Model

A photon arrival is modelled by the closing of switch S . It follows that the initial breakdown event is a fast build up of current related to the total bias voltage and internal resistance R_s . This current effectively flows through the junction of the device. The reset phase duration depends on the RC time constant of the total capacitance and the large passive quench resistance R_q . This period can be configured from tens to hundreds of nanoseconds. The resulting waveform at the cathode of the detector is converted directly into the digital domain via a fixed threshold inverter (comparator). This area efficient approach is particularly appealing for arrays of SPADs to ensure maximum overall pixel fill factor. It is also current limiting in nature, so can be configured as power efficient and reliable.

However, passive quenching has several drawbacks. Firstly, during recharge the detector becomes increasingly more biased beyond its breakdown voltage and so

experiences an accompanying increase in junction electric field and so photon detection probability. This renders the device able to detect the next photon arrival prior to being fully reset, with variable timing resolution. This behaviour is coupled with a significant fluctuation in the reset waveform period caused by variation in the avalanche quench point (and noise at the inverter threshold). This is caused by asymptotic recharge current crossing the avalanche latching point of the detector at a shallow angle [53]. The impact of this variation depends on the application but can be minimised by careful quench resistance choice. For example, it may be negligible in applications where the timing of leading edge of the avalanche breakdown is the critical metric.

Additionally, particularly for large area devices with corresponding large total junction and parasitic capacitance, the dead time can be undesirably large. A large junction capacitance also results in a larger volume of charge flow during the breakdown event. This results in a greater probability that a charge trap is filled during this phase and so subsequently leads to increased afterpulsing. For this reason, large diameter devices are best suited to active quench circuits [16]. Conversely, the dead time can be reduced by minimising the detector capacitance as implemented by Rochas in [15]. Rochas reports a small diameter 68fF detector with 32ns dead time using a fully passive quenching/reset approach. In his work he also estimates a full discharge of the detector capacitance in ~ 3 ns. It should be noted that detector orientation also has an impact on device capacitance, as discussed in section 2.3.5.3.

It should be highlighted that the overall timing resolution of the circuit shown in Figure 21 consists of two components; the SPAD itself and the output inverter/buffer. In order to remove the inverter's jitter contribution the passive quench component can be implemented off chip, and the response of the detector monitored in the current domain or at the device periphery. In this case, a constant fraction discriminator (CFD) with variable threshold can be tuned to detect the onset of the avalanche event with high sensitivity [53]. However, this approach dramatically increases the parasitic capacitance on the detector's moving node due to routing and bond pads. This increases dead time, charge carrier flow and afterpulsing. Therefore, an integrated output buffer with unbalanced or tuneable threshold is recommended for passive quenching implementations.

2.3.5.2 Active Quenching

To address the limitations of passive quenching, particularly for large area devices, many active circuit solutions have been proposed originating from Cova et al [42], [53], most often involving a feedback principle. Active quenching circuits aim to detect the onset of a photon arrival event, either preventing a full discharge of the diode's capacitance or redirecting the avalanche charge flow to an alternative path away from the detector's junction. Limiting the charge flow through the junction not only helps reduce afterpulsing probability and localised self heating effects but reduces photo-luminescent emission induced crosstalk, which is important for array implementations. Whilst it may be difficult to better the $\sim 3\text{ns}$ discharge time of a fully integrated passive quench setup without adding significant complexity, active circuitry can address the requirement for fast and well controlled reset of bias conditions. The full operating cycle of an active circuitry approach can be represented by the phases shown in Figure 22.

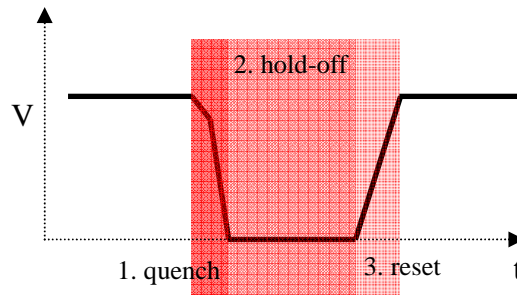


Figure 22: Active Quench Operating Phases

A short hold-off time is mandatory for applications demanding high photon count throughput, and nanosecond time-gated modes. It follows that the use of active quenching is desirable for high background light or DCR conditions so as to maximise throughput of otherwise valid detection events. Active circuit solutions also address the passive quench drawback of slow and non-linear R-C reset by rapidly driving the detector back to photon-ready bias conditions through a low impedance path. This provides a constant photon detection probability outwith the dead time. The dead time is the sum of all three phases of the operating cycle shown above. Cova reports minimum dead times of 10ns for low voltage, planar SPADs in [53].

However, the addition of extra circuitry obviously affects pixel fill factor, which implies that this is an approach best suited to large area detectors. Furthermore it is desirable to include programmability for all three phases of the active quench cycle that adds further complexity.

The use of monostables in the literature is prevalent, such as the classical circuit proposed by Ghioni et al [54], illustrated below.

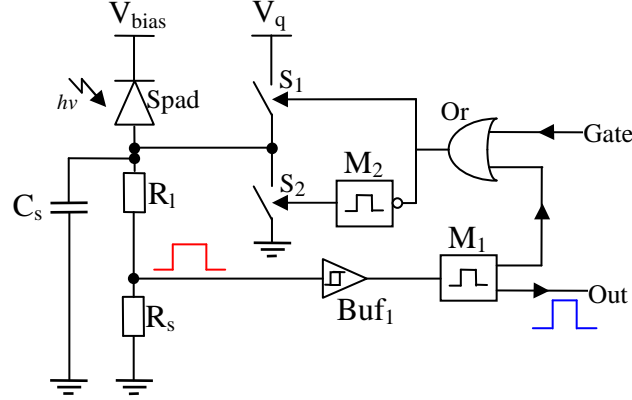


Figure 23: Example Active Quench Implementation

In this circuit, the avalanche induced voltage drop across R_s passively triggers Buf_1 . Quenching is actively completed for a duration set by monostable M_1 , the Or gate, and switch S_1 . Once the M_1 pulse is finished bias conditions are reset via the short pulse duration of negative edge triggered monostable M_2 , and S_2 . Monostables suffer from notoriously poor temperature and supply voltage variation and matching across wafers and are therefore not desirable in timing critical applications.

Hybrid implementations such as the example above are common, i.e. where one operating cycle phase such as the initial quench is addressed passively with hold-off and reset implemented actively.

Finally, it should be noted that the SPAD junction capacitance increases non-linearly with increasing excess bias voltage, resulting in progressively larger numbers of charge carriers flowing during breakdown. Therefore for applications requiring high excess bias conditions and high photon detection efficiency coupled with large active area, active quenching is recommended.

A novel, compact (20T) and monostable free circuit for fully active quenching and reset down to ~ 5 ns dead time is reported as part of this research in [55], and is discussed in section 3.8.

2.3.5.3 SPAD Orientation Options

Since a SPAD is a two terminal diode, operated at an excess bias potential beyond its breakdown voltage, it may be oriented in two ways; either with a negative potential on the anode, or positive on the cathode. As previously discussed, it is desirable to minimise detector dead time and charge flow quantity during an avalanche event. The optimal connectivity to permit this depends on the parasitic capacitances which are specific to the diode construction which has been implemented. To illustrate this point, two possible passive quench configurations are shown in the figure below for a SPAD implemented within a deep NWELL, P-substrate technology such as that by Pauchard et al in [56].

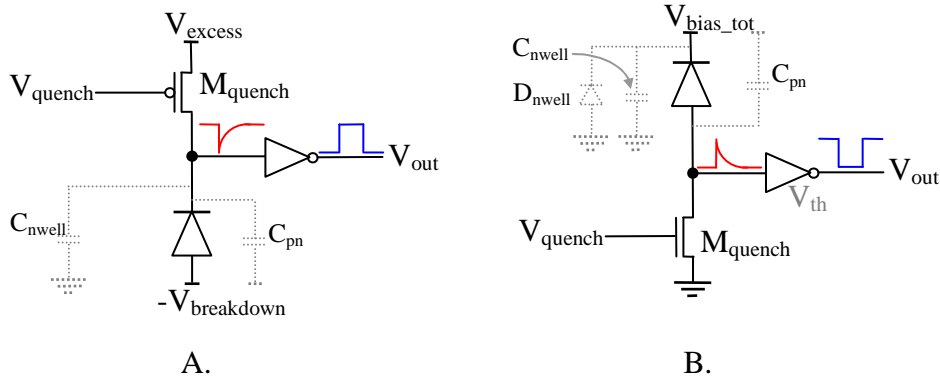


Figure 24: SPAD Orientation Options

Figure 24A shows a large negative voltage applied to the anode, with a much smaller positive excess bias applied via the passive quench PMOS to the cathode. The total bias applied across the diode is $V_{breakdown} + V_{excess}$. Figure 24B employs an NMOS quench component connected with source connected to ground potential, and a large positive bias potential V_{bias_total} applied to the SPAD cathode. In both cases the key moving ‘sense’ node of the circuit is at the buffer/comparator input terminal. The resultant time domain output signal is a thresholded version of the key sense node. However, in configuration A the cathode is the moving node and therefore the additional capacitance of the NWELL to P-substrate parasitic diode (C_{nwell}) must also be charged and discharged during the detector operating cycle. As well as contributing to the lengthening of detector dead time due to the increased R-C load, this adds to the volume of charge flow through the detector, so increasing likelihood of a charge trap becoming occupied and therefore increasing afterpulsing probability.

In the case of the ‘flipped’ configuration (B) the anode of the detector is the moving node, and therefore there is only the charging of the SPAD junction capacitance to consider. Importantly, this permits the sharing of NWELL regions by multiple elements within an array implementation leading to reduced pixel pitches and improved fill-factor. However, it should be noted that it is common for the mobility of an NMOS transistor channel to be higher than PMOS in deep submicron processes (K_p can be typically $\frac{1}{4}$ of K_n) which leads to a larger equivalent area used when compared to a PMOS implementation for the same target quench resistance. Additionally, the breakdown voltage of the NWELL - P-substrate diode must be higher than V_{bias_total} , with low sub-breakdown leakage current in order for this configuration to work. Additionally, regardless of orientation, for reliability reasons it is vital to not exceed the maximum gate oxide potential of the output buffer otherwise permanent damage can occur. For this reason it is common to employ thick gate oxide transistors for this circuit component if available in the target technology.

Both orientations are prevalent in the published literature, for example Ghioni et al [54] implementing configuration B, and Pancheri et al [47], and Niclass [16] discussing both. The two orientations have been implemented as part of this research, the design of which is discussed in section 3.8.

2.3.5.4 SPAD Arming Options

A final detector control and connectivity option to consider is the method used to ‘arm’ the SPAD into a ‘photon ready’ state. Two potential options are shown in the figure below.

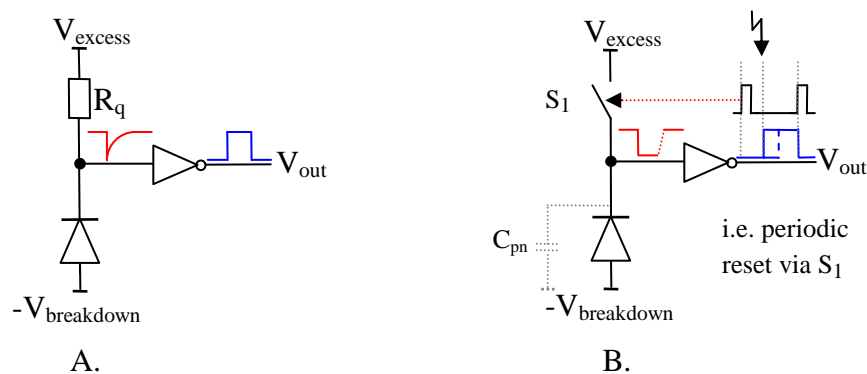


Figure 25: SPAD Arming Options

When a SPAD is reset, the critical moving node may be either left connected to the bias potential via a high impedance passive quench component or alternatively once reset (preferably via a low impedance path) the moving node may be permitted to float at the recharge potential on the diode junction and parasitic capacitance C_{pn} . In this case the SPAD will sit in an ‘armed’ state until either a photon arrival initiates an avalanche event or a dark count event occurs. Once discharged, the SPAD requires active control of S_1 in order to be reset again.

The armed configuration can be useful in applications which make use of the self-latching behaviour of the output signal, particularly in arrayed event counting implementations such as that published by Verghese et al [57].

2.4 Detector Conclusions

The preceding sections of this chapter have introduced SPAD performance parameters, presented a review of SPAD state of the art and discussed the various fabrication processes used for detector implementation. A review of typical SPAD constructions was provided, along with an explanation of quenching and arming options. The requirement for integration of the detector within the same substrate as processing circuitry meant that a SPAD construction which permitted operation independent from the substrate ground potential was mandatory. Additionally, the impact of quasi-neutral field regions associated with implanted guard rings on timing resolution diffusion tail indicated that the enhancement mode structure was a preferred choice. Finally, the simplicity and ease of implementation of this structure in nanometer scale CMOS technology meant that this was the chosen topology for a series of novel detectors, discussed in detail in chapter 3.

The explanation of detector construction choice is summarised in Table 1 below.

SPAD Metric	Diffd/ Implt Guard Ring	Reach- Through	<i>Enhancement Mode</i>	Timing Optimised	STI Guard Ring
DCR	Often high		Generally low	Generally low	Very high
PDE	PDE can be non-planar	Broad PDE	Uniform field		
Timing Resolution	Quasi-neutral region impact.	Reduced diff'n tail	Reduced diff'n tail	Best timing performance	
Area	Scaling limit		Scales well		
Process		Substrate is one terminal.	Standard process	Customised process	

Table 1: SPAD Metric/Construction Summary

2.5 TDC Performance Parameters

The role of a Time-Digital Converter in a time correlated imaging system is to measure the time period between an illumination event to a subject's response or reflected light to a high degree of accuracy. Regardless of architecture, the main parameters used to measure the performance of a TDC are time resolution, dynamic range, linearity (accuracy), precision, speed, area and power consumption. The additional capability of the TDC to maintain its performance whilst enduring operational environmental changes such as temperature and supply voltage variations, as well as coping with manufacturing tolerances is a particularly beneficial feature.

Several of the performance parameters for TDCs are similar to those associated with analogue to digital converters (ADCs). They are described in the published literature (e.g. Mäntyniemi [58]) and describe a clearly defined functional block of circuitry, as opposed to the specific device engineering parameters associated with SPADs.

2.5.1 Time Resolution

The time resolution of the TDC is the minimum value of time span that can be resolved. This is often the key design parameter that allows classification of a TDC architecture as either being interpolative (less than an inverter delay) or based on the minimum process gate delay. The minimum gate delay of a deep sub-micron process is of the order of a few tens of picoseconds.

2.5.2 Dynamic Range

The dynamic range of the converter is the maximum time span that can be converted. Also relevant is the ease with which the dynamic range can be increased (or decreased). For some architectures a doubling of silicon area is required, whereas others may simply require one extra flip-flop register element. This aspect of the converters design flexibility also therefore has an impact on power consumption. Related to dynamic range is the width in number of bits of the converters output data word. For example, a 10 bit TDC with a time resolution of 100ps would have a maximum possible dynamic range of 102.4ns

2.5.3 Accuracy

The accuracy of a TDC is a measure of how closely the converted result matches the true temporal value. The linearity of conversion over the full dynamic range of the converter, expressed in terms of differential and integral non-linearity, demonstrates the deviation from the true value and can be expressed in the same way as for analogue-digital converters.

2.5.3.1 Differential Non-Linearity

Expressed for each output code, DNL is the difference between an actual code step and the ideal code step of 1 LSB. A DNL of less than $\pm 0.5\text{LSB}$ over the full dynamic range of the converter ensures a monotonic transfer function with no missing codes.

2.5.3.2 Integral Non-Linearity

Expressed over the full dynamic range of the converter, INL is the deviation of a converter's transfer function between the actual value and a straight line drawn either as a best fit (preferred) or from the converters beginning to end-points. Any offset and gain errors are also included in the INL profile but are often highlighted separately. Section 6.2.2 contains a detailed discussion of the 'Doernberg method' [59] used during this work for computation of TDC DNL and INL.

2.5.4 Precision

Precision is the standard deviation of a defined number of measurements applied to a certain time measurement. It is thus a measure of the repeatability of a TDC conversion. This is particularly relevant to converter architectures that are susceptible to accumulated jitter effects. This should be quoted for worst-case input conditions with a large percentage of the full-scale dynamic range being exercised.

This parameter is also a key consideration for TDC designs intended for array level implementation, i.e. the uniformity of an array of converters all being stimulated with the same input signals.

2.5.5 Conversion Rate

This is the rate per second at which the converter returns a measured result. For some architectures this parameter can vary dependent on the time duration to be converted. Some TDCs output a data result very quickly after the time period being measured,

sometimes in real time, whereas others always require their full-scale dynamic range to be exercised. This also has an impact on power consumption.

Also relevant is the amount of converter dead time, i.e. when the TDC is not able to be stimulated with valid input conditions. This can occur for some TDCs during data readout or reset periods.

2.5.6 Power Consumption

The power consumption of the converter is normally quoted for a defined input time span, or a percentage of the full-scale dynamic range. This parameter is of particular importance for TDCs intended for array implementation. The power consumption profile can also be very important for some applications, i.e. it can be continuous or exhibiting peaks during a valid conversion cycle. Also relevant is the supply voltage magnitude and the ability of the converter to reject broadband noise on the power supply rails.

2.5.7 Compensation Capability

Ideally the converter should be synchronised to a known, stable, external clock frequency to permit absolute measurements to be made, as well as providing a means of compensation for environmental effects and manufacturing tolerances. Process, supply voltage and temperature variations can seriously impact the converter precision and accuracy. Therefore, designs intended for industrial, scientific or consumer applications should ideally have integrated circuitry that permits a high degree of rejection of or compensation for such effects. Some structures have inherent compensation, such as those based on DLLs, shown in Figure 27 and Figure 33.

2.5.8 Silicon Area

The area in silicon of the TDC is an important parameter for an architecture aimed at array level implementation. Although this metric is heavily influenced by the actual architecture, the minimum feature size of the manufacturing process used has a clear impact.

2.6 TDC Literature Review

The aim of this research is to integrate arrays of low noise single photon detectors with time resolving circuitry in a deep sub-micron manufacturing process. The review that follows focuses on previously published state of the art TDCs.

2.6.1 Background

The purpose of a TDC is to measure the time difference between two signal events, start and stop, and output a digital result. The origins of TDCs can be traced back to US nuclear physics research in the late 1950's such as Ronald Nutt's 'Digital Time Intervalometer' [10]. Since then, many differing TDC architectures have been developed, targeting an array of applications. Each approach may be categorised either as exhibiting a timing resolution based on the minimum gate delay available in the target manufacturing process, or being able to achieve sub gate-delay resolution. Each approach has its unique strengths and weaknesses. Despite the nature of the construction of most of the potential TDC architectures being based on standard cell blocks, the design and layout of the converter must be addressed as a full custom design and cannot be synthesised using automated techniques. This is due to the fact that the internal nodes of the TDC structure become analogue in nature at the propagation speeds involved, and there is an overriding requirement to achieve a fully symmetric and parasitic balanced design in order to achieve a monotonic, linear transfer characteristic. Some of the key architectures are discussed in the following sections.

2.6.2 Clocked Delay Lines

The clocked delay line is the most basic time-digital converter approach. It consists of a delay line of multiple inverter elements into which a 'start' pulse is injected. At some point in time later, the state of the delay line is sampled by a 'stop' clock signal. The output word bus $\langle Q_n:Q_0 \rangle$ represents the time difference between the two signals. This approach is shown in the figure below.

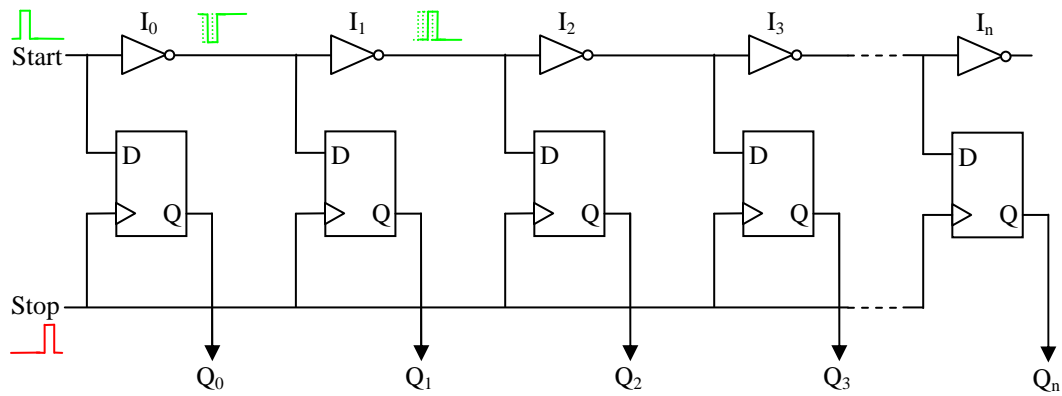


Figure 26: Clocked Delay Line TDC

This is a relatively simple technique with a time resolution equal to the minimum gate delay, and is reported as applied in 90nm CMOS by Staszewski [60]. A recurring theme with flip-flop (FF) based TDC circuits is a critical metastability issue which results in incorrect output coding. Every FF stage has setup and hold time parameters which must be adhered to in order to ensure consistency of operation. The likelihood of data or clock signals changing within these defined parameters in a TDC is very high. The result of such an occurrence is an individual FF output which is skewed in time (see section 2.6.6, time stretching TDCs). The effect on the overall TDC performance depends on when the output data is sampled, although it should be noted that the encoded subsequent result is statically memorised for the duration of power being applied or the circuit reset.

The area occupied by this approach is comparatively high, requiring a FF and inverter pair for every LSB of resolution. Therefore for a 10b TDC, over a thousand FF and inverters are required. The impact on layout is high, requiring very careful clock tree buffering routing and balancing in order to ensure every FF receives coincident clock signal edges. Also, the expansion of the output word by one bit requires a doubling of area, and linearity is heavily impacted by delay line element matching errors. Further, configurations using delay lines such as this tend to always require a minimum conversion time equal to the full dynamic range of the converter, resulting in both lengthy conversion dead times and fixed, relatively high power consumption.

However, the major weakness of this circuit is that the delay through the inverter chain is sensitive to process, supply voltage and temperature dependent variation. Many novel TDC implementations include additional compensation circuitry to

address this issue. The process of correcting for these variations is referred to as ‘calibration’. A typical technique for delay line based implementations is to embed the chain of elements within a delay locked loop (DLL) structure as shown in Figure 27, derived from standard phase-lock techniques as presented in [61].

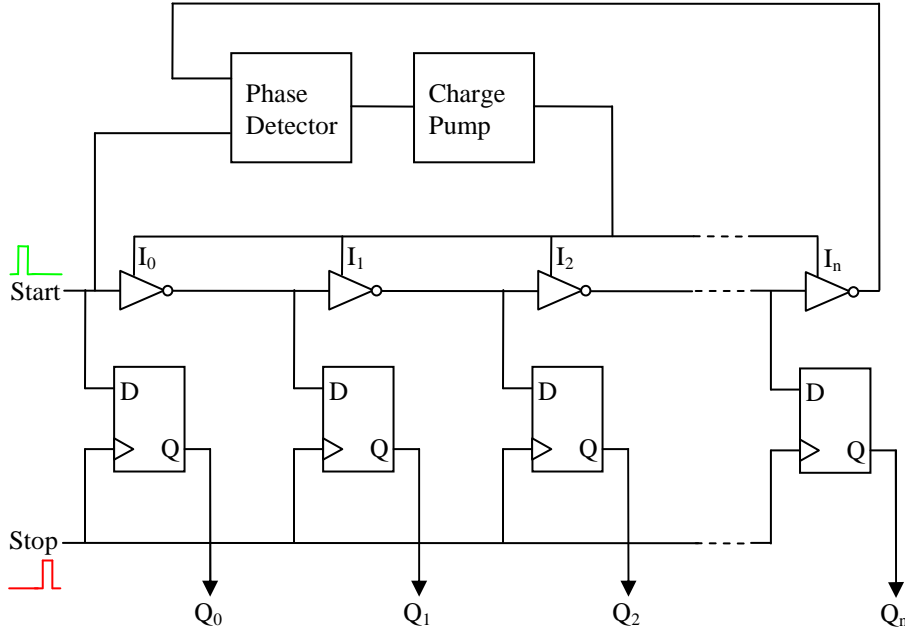


Figure 27: Delay Locked Loop Calibration Concept

Lastly, the digital output from the FF chain shown requires additional binary encoding due to the fact that the delay chain consists of inverting elements at each stage.

2.6.3 Vernier Delay Lines

A variation of the clocked delay line is the vernier delay line (VDL, or ‘pulse chaser’). In this configuration, two *buffer* based delay lines are used with slightly differing unit delays. With the start signal edge already transiting the upper delay line, the stop signal is presented to the lower, faster delay line. The stop signal progressively catches up with the start signal in a Vernier Gauge like manner, clocking the state of the start delay line as it goes. The resulting static thermometer output code shows where the stop and start signals coincided. The stage number at which the start pulse is caught represents the number of LSBs difference between start and stop signals.

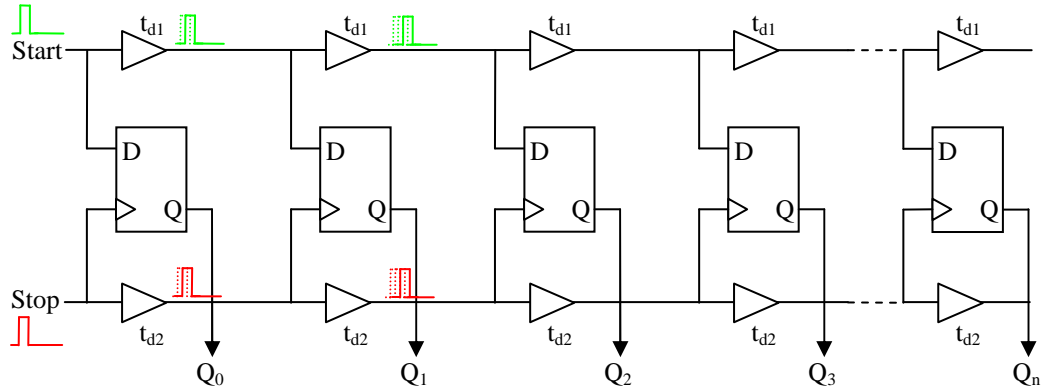


Figure 28: Vernier Delay Line TDC

Although this architecture exhibits some of the undesirable traits of the clocked delay line, such as large area, the time resolution of the VDL is equal to the difference between the unit delays of the two delay lines. This dramatically enhances the resolution which can be as low as $\sim 5\text{ps}$, as discussed by Dudek et al in the $0.7\mu\text{m}$ implementation of [12]. Dudek attempts to address the PVT variation issue of the clocked delay line by embedding both Vernier delay lines in a combined DLL. This requires that sufficient headroom is required on the voltage controlled delay elements in order to allow variability in both speed directions. In this implementation a ‘coarse’ counter is added in order to reduce the delay line lengths. The problem of conversion dead time is addressed by dividing the VDL into sub-sections, each of which are read in a pipelined manner. This allows fresh data to be presented to the TDC as soon as the first sub-section has been read, resulting in several valid data conversions transiting the structure simultaneously. However, reported accuracy is slightly disappointing in [12] at $\pm \sim 25$ LSB INL. Layout related effects on linearity are also reported by Rahkonen et al in [62]. Additional calibration and counting circuitry such as that shown in Figure 27 takes up further silicon area. A thermometer-binary decoder is also required, resulting in high area usage which precludes this architecture from being used for array implementations.

2.6.4 Pulse Shrinkers

The next class of TDC to be discussed is that referred to a pulse shrinkers, as reported by Räisänen-Ruotsalainen et al in [63] and Karadamoglu et al later in [64]. This class is a variation on the VDL, in that it is a sub gate-resolution implementation but with

time resolution equal to the difference between the rise and fall times of the unit element of a bias controlled delay line. The construction is shown in the figure below.

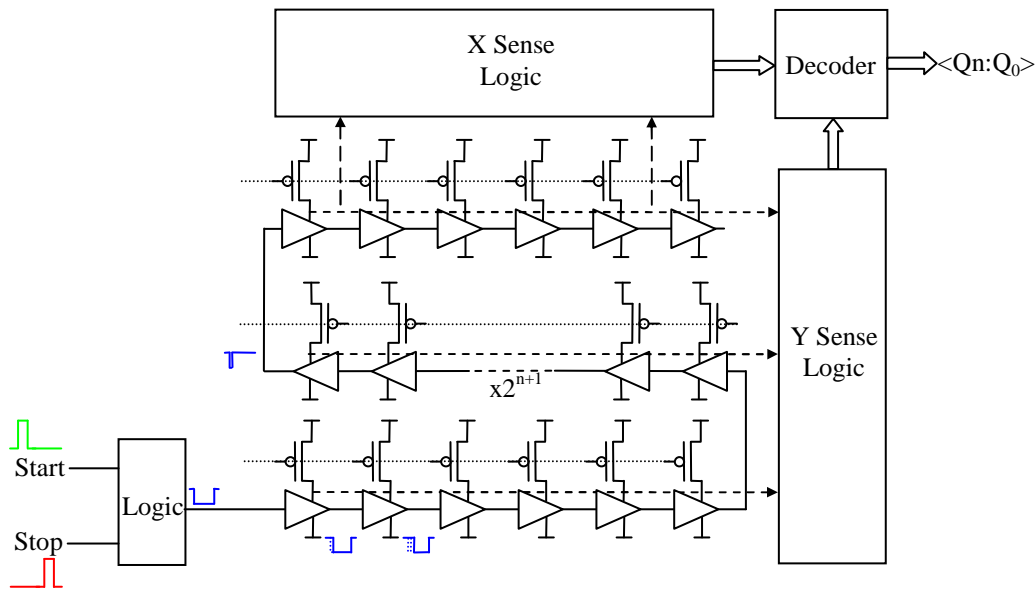


Figure 29: Pulse Nibbler TDC

Start and stop signals are combined to create a single pulse which is fed into the beginning of the pulse shrinking delay line, the unit buffer element of which has unbalanced rise and fall times. In the figure above, the PMOS controlled elements are slow to respond to a low going input, and therefore the leading (falling) edge of the input pulse is gradually moved toward the trailing (rising) edge as it transits the delay line. The role of the X and Y sense logic blocks is to detect where the input pulse becomes annihilated. This position is decoded to represent the binary output.

Whilst the same calibration techniques used for VDLs may again be employed (e.g. embedding the delay line in a DLL) this is an area hungry implementation. The delay line is large, consisting of 2^n elements. Several logic functions are also required in order to process valid input conditions and perform output encoding. The structure suffers from lengthy dead time, as the pre-processed input pulse must be subsequently allowed to transit the full length of the delay chain prior to encoding and final data readout. Again, this structure is not suitable for array implementations.

2.6.5 Passive Interpolators

Passive interpolators borrow the idea of interpolation from voltage interpolating ADCs. This resolution enhancing approach relies on the rise/fall time of a buffer stage being longer than the propagation delay. When this condition is met, intermediate signals may be synthesised which subdivide the intrinsic gate delay time. These are called ‘interpolated’ signals. The computation of these is demonstrated in the figure below.

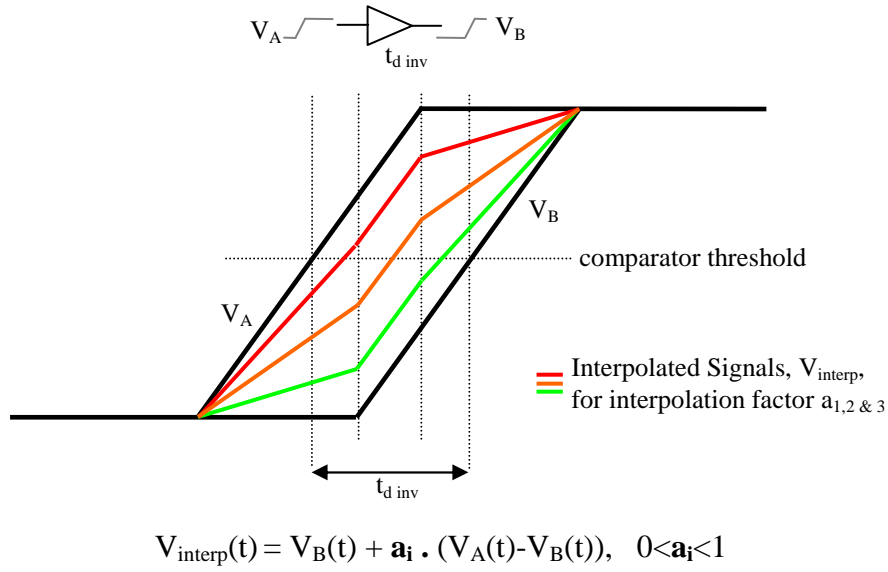
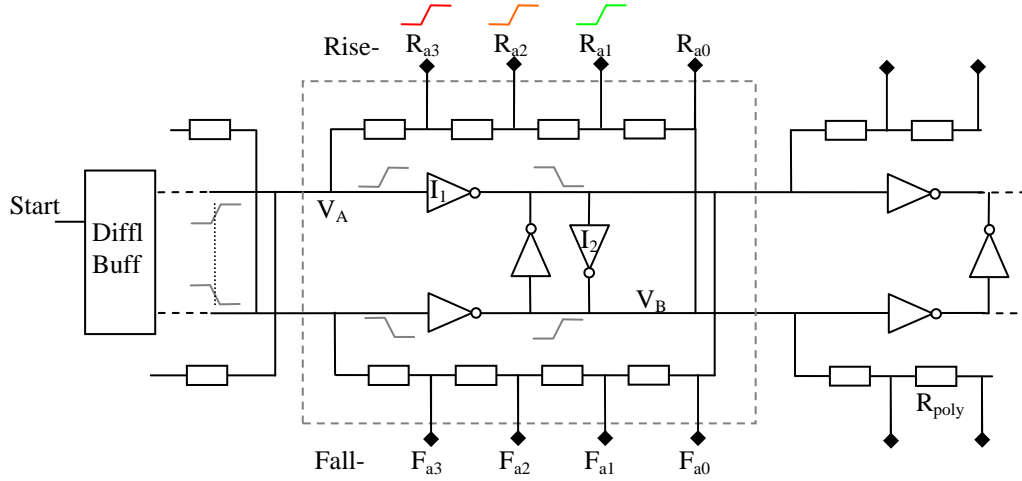


Figure 30: Time Interpolation Concept

It can be seen that the interpolation factor \mathbf{a}_i permits synthesis of a user defined range of additional signals which may be used within a TDC architecture to improve upon the basic gate delay available in the target process, as discussed by Henzler et al in the 90nm CMOS implementation of [65]. The building block of the delay line developed is shown in the figure below. The figure shows the detail of a single interpolation stage within the ‘start’ delay line, with three synthesised intermediate signals which subdivide the combined propagation delay of inverters $I_1 + I_2$.



$$V_{\text{interp}}(t) = V_B(t) + \mathbf{a}_i \cdot (V_A(t) - V_B(t)), \quad \mathbf{a}_i = 0.25, 0.5, 0.75$$

Figure 31: Differential Interpolated Delay Line

This inherently monotonic architecture successfully enhances time resolution without adding to latency and dead time, but has the added complexity of passive elements at every delay unit cell of each delay line. Furthermore, additional readout cells such as voltage sense amplifiers are required to complete the addition computation at each stage which adds to the area overhead. Although the resistive elements are small in value which aids overall power consumption, the desire to use well matched unsalicyded polysilicon elements further adds to the total area used.

2.6.6 Time Stretchers

Instead of following the path of minimising the time delay through a delay element in order to enhance TDC time resolution, an equivalent effect may be obtained by amplifying or stretching the input start-stop time difference by a known factor prior to conversion. The principle of time amplification was first reported by Abas et al in [66] and then applied to TDC design by Minjae and Abidi in [67].

The time stretching circuit schematic and timing response is shown in Figure 32.

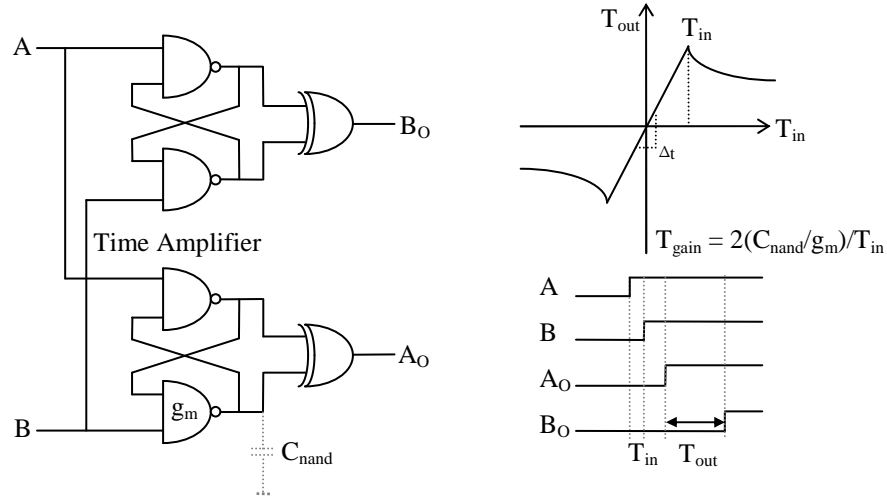


Figure 32: Time Amplification Concept

The result of study on edge coincidence and metastability, the time amplifier is built from two SR latches with Xor gates on the outputs. The timing diagram and transfer characteristic shown demonstrate that when almost coincident edges are presented to the A and B terminals the output result is a linearly expanded pair of outputs with time amplification equal to T_{gain} . The expansion factor is dependent on the transconductance and capacitive load of the NAND gate building block. The resultant output A_o , B_o can then be fed into a TDC structure such as a VDL etc, or more complex realisations such as that reported in [67]. This 90nm CMOS, high resolution approach utilises the metastability characteristic which causes such problems for clocked delay lines. The implemented structure was aimed at PLL designs and exhibits good linearity for such high resolution but occupies a prohibitive area of silicon when considering array implementations. By nature it is an architecture which converts its input in non real-time although it has such fine time resolution this may not be an issue for many applications. Time stretchers (like many TDCs) are thought susceptible to power supply borne disturbances causing single shot acquisition error.

2.6.7 Distributed Clock Structures

The coarse-fine TDC approach embodied by Nutt in [10] can naturally be applied to modern, CMOS implementations. The elementary addition of a digital counter to perform a coarse level of time encoding can be applied to many TDC implementations, for example as reported by Dudek in the VDL [12]. The key architectural issue, particularly for array implementation, is what signal is used to

increment the coarse counter? Commonly this is a stable clock signal common to many TDCs such as the approach reported by Niclass et al in [68]. In this case two further levels of encoding are also applied. A medium resolution conversion is performed by using the coarse input clock to feed a DLL structure to create a 16x division of the primary global clock. Lastly a fine conversion is performed by utilising a differential clocked delay line, with time resolution τ_{fine} equal to the process unit gate delay. An example conversion of time 't_{sample}' is shown broken into its constituent parts in the figure below.

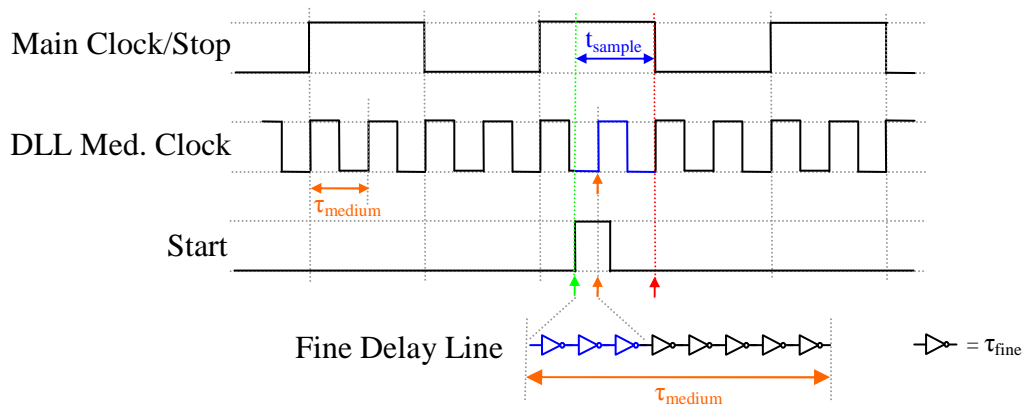


Figure 33: Distributed Clock Concept

This figure shows an asynchronous start, with synchronous stop, as is common in time correlated imaging applications such as 3D imaging and Fluorescent Lifetime Microscopy (FLIM) which both use a synchronous illuminator. For other applications the stop signal may also be asynchronous, resulting in a second 'fine' period which must be computed and factored in to the final result.

Whilst potentially simple, the main issues with this particular real-time conversion approach are the high power consumption associated with clock distribution (particularly for arrays) and the timing delays incurred when routing high frequency signals over long routes on a silicon chip. However, in [68] the embodiment of a DLL structure helps to limit PVT variation, and the sequential line nature of the output data readout suits video rate 3D image display. Timing delay related image droop may also be characterised and corrected via a look up table and linearity is good. However, for simultaneous conversion for arrays there are approaches which may be more power and area efficient in nanometer scale CMOS processes.

2.6.8 Time to Amplitude Approaches

Alternatively, time to amplitude techniques may be utilised in a time-amplitude-digital conversion approach. The transfer of the time computation into the voltage domain allows the application of established ADC design methods. The entire time period to be converted may be processed in the voltage domain, or a coarse-fine approach may be implemented, as reported by Räisänen-Ruotsalainen et al [69], and Swann et al [70]. The basic theory of the approach is explained in Figure 34.

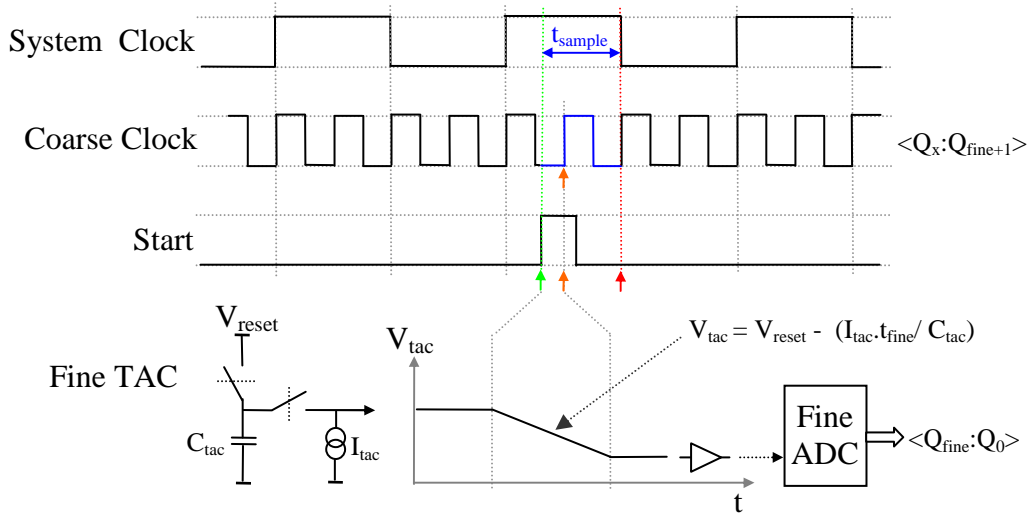


Figure 34: Time-Amplitude-Digital Conversion Principle

In most TACs, charge is either drained from a capacitive store, or accumulated on to it for a certain period of time. The resultant voltage is then fed to, for example, a flash ADC. In the case above, a fine TAC is combined with a coarse clock counter to create the full output word. Whilst TAC operating voltage range is progressively limited in line with the lowering of supply amplitudes and can be susceptible to the impact of kT/C or power supply borne noise, TACs boast some interesting performance advantages. For example, they can be designed in a compact and power efficient manner which is appealing for array implementation ($C_{tac} \approx$ a few pF). By nature they also tend to exhibit static memory traits which can be useful for pipelined read out schemes. Real-time conversion is still possible, with fine being converted whilst coarse clock periods are being counted in the example above. However, their performance is generally limited by the LSB voltage resolution of the chosen ADC architecture.

2.6.9 Gated Ring Oscillators

The final approach to be considered is referred to as ‘gated ring oscillator’. In this architecture, instead of a long delay line with a large number of elements (as in a clocked delay line approach), a reduced number of inverters are configured as a ring oscillator, i.e. with the final stage output fed back to the input. The ring oscillator is seeded with an initialisation pulse/condition, which is allowed to start transiting down the inverter. Normally a coarse counter is incremented when the pulse completes a full circuit of the ring. When a stop condition is received, the output word is a combination of the coarse counter and the decoded state of the internal nodes of the ring.

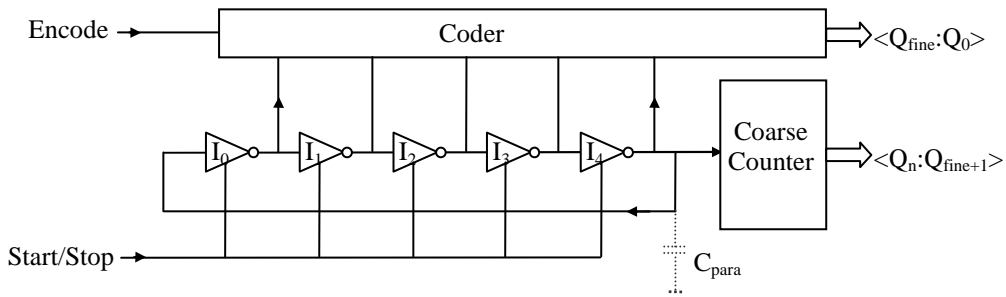


Figure 35: Ring Oscillator TDC

Examples of this approach are Nissinen et al [71, 72], as well as Arai, Ikeno [73], Helal et al [74], and De Heyn et al [75].

Ring oscillator TDCs have several appealing performance characteristics. Unlike architectures which involve a combination of delay lines and flip-flops, ring oscillators have fewer issues with metastability, and perform the conversion in real time. The TDC can be easily calibrated for PVT by using DLL approaches. However, to achieve the smallest gate delay and so the best time resolution, the use of single ended inverters is desirable. Doing so results in an odd number of ring oscillator elements to ensure ring instability, which in turn greatly complicates the encoding of the fine state of the ring into a binary code space. In [73], Arai works around this problem by using balanced look ahead logic. Alternatively, differential buffers may be used instead of single ended, which allows simplified coding and ring instability with an even number of elements, as reported by Nissinen [71, 72] and De Heyn [75].

The power consumption of a ring oscillator based TDC is also appealing for multiple converter systems coping with sparse events, as it tends to peak only when the ring oscillator is in operation. For the rest of the time, consumption is dominated by leakage only. This provides a low average consumption profile which scales with overall system activity.

This approach yields a compact layout when compared to delay line based designs, mainly due to the fact that the delay line is greatly shortened. However, long term jitter and stability under power supply transients is a concern requiring careful design. Similarly, achieving good linearity requires careful balancing of parasitic capacitances between ring oscillator elements, particularly at the feedback point.

This was the architecture chosen for this research activity. The rationale behind the choice and detailed design is summarised below in Table 2 and discussed at length in section 5.

TDC Architecture	Resolution/ Dynamic	Accuracy/ Precision	Conversion Rate	Power Consumption	Area/ Scalability
Clk'd Delay Line/ VDL	calibration possible	metastability issue	fixed at full dynamic range	high dynamic, constant	large and does not scale well
Pulse Shrinker	calibration possible	relies on matching	fixed at full dynamic range	high, constant	v-large, does not scale well
Passive Interpolator	high	high	fixed at full dynamic range	constant	v-large
Time Stretcher	v-high	acceptable for resolution	non-real time	acceptable for resolution	large, does not scale well.
Distributed Clock	min gate delay	jitter a concern	real time	v-high	hard to scale
Gated Ring Oscillator	min gate delay	jitter a concern	real time	low average, peaky nature	small, scales well.

Table 2: TDC Metric/Construction Summary

2.7 Integrated Detector-Converter Arrays

In the preceding sections single photon detectors and time-digital converters have been introduced alongside a discussion of their main performance metrics. This section introduces particularly relevant examples of integrated detector-converter arrays and technology developments. Reducing the risk of integration can be done via an assessment of the potential paths of noise introduction, as shown in the Figure 36.

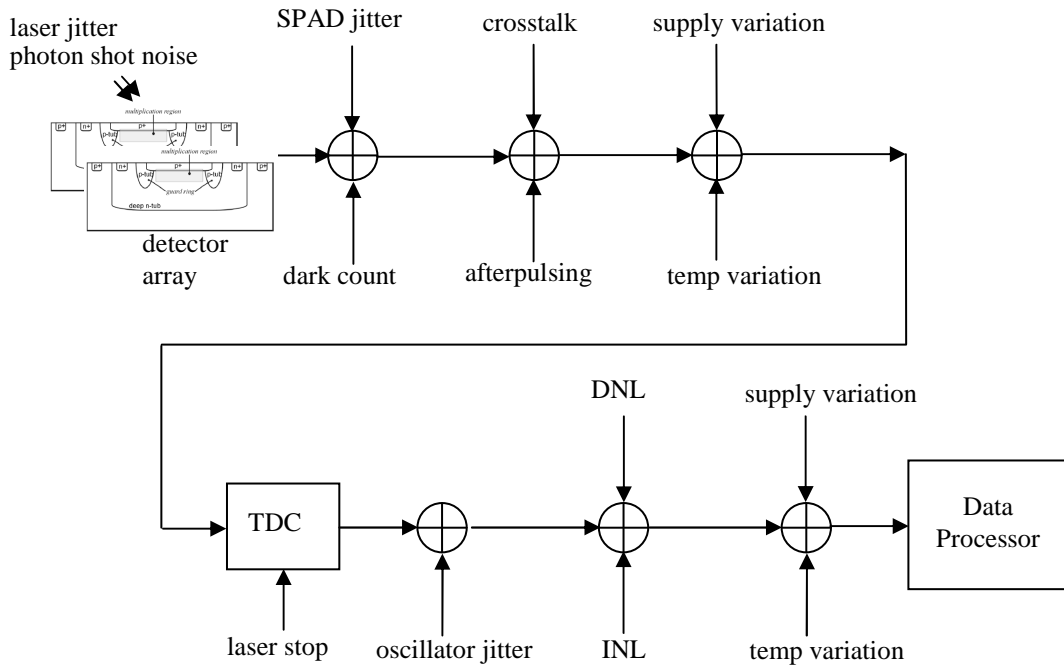


Figure 36: Integrated System Error Sources

The use of this methodology also allows a structured review of the state of the art of integrated solutions and helps to form an understanding of the matching of sensor-detector types with particular applications. Included are standard ‘3T’ photodiode-ADC arrays, X-ray sensors as well as existing SPAD-TDC arrays.

2.7.1 Photodiode-ADC Arrays for Bioluminescence Detection

The photodiode-ADC array designed for bioluminescence detection reported by Eltoukhy et al in [76] sets a challenging benchmark for low noise, low light detection using a comparatively conventional pixel approach. In this $0.18\mu\text{m}$ CMOS approach an 8×16 , $230\mu\text{m}$ assay pitch-matched pixel is mated to a 128 channel, dual-slope ADC matrix. Whilst this is not a time correlated imager, a pseudo differential design,

CDS and multiple sample averaging results in an overall light equivalent noise level of 0.22μ lux. This publication is particularly relevant for comparing the pixel performance with SPAD based imager DCR levels, as well as the area utilisation of 25mm^2 ($\sim 0.5\text{M}$ transistors) for the modest total pixel count of 128.

2.7.2 Medipix High Energy Particle Detector Arrays

A series of interesting multi-pixel, arrayed detectors have been developed by the Medipix Consortium, for use in the Large Hadron Collider and other applications. These devices have their foundations in earlier research at LEPSI/CERN [32], and take the form of an array of high energy particle detecting pixels with associated readout electronics. The implementations are consistently a hybrid of silicon or gallium arsenide detectors bump bonded to a CMOS backplane of processing electronics. The detector array, whose material selection depends on the nature of the radiation of interest, directly senses charge generated by the incident particles, which is then passed through a pre-amplification stage prior to threshold detection and data readout. This high fill factor approach is particularly suitable for this application and others where microlenses cannot be used.

Several developments were introduced through out this specific body of work during a decade long research programme from approximately 1998 onwards. ‘Medipix1’ was a 64×64 array implemented in $1\mu\text{m}$ CMOS, with $170\mu\text{m}$ pitch 400T pixels, each with a single threshold comparator/event counter [35, 77]. In ‘Medipix2’ the spatial resolution was improved using $55\mu\text{m}$ pixels in a 256×256 array format, fabricated in $0.25\mu\text{m}$ CMOS [78]. Each 500T pixel had dual threshold comparators, but charge spread between pixels was evident. The ‘Timepix’ radiation dosimetry modes allowed measurement of time to threshold and time over threshold environmental parameters. Lastly the 130nm ‘Medipix3’ 8×8 pixel test chip incorporated a pixel charge sharing mode between 4 adjacent $55\mu\text{m}$, 1100T, dual counter pixels [79]. The x and γ ray images obtained have greatly improved spatial resolution, contrast and sharpened detail when compared to CCD plus scintillator competing technology.

2.7.3 SensL Digital APD Arrays

Sensl is a company specialising in low light sensing technology for 3D camera, scientific, medical and security applications. Based in Ireland with strong links to local academia their product portfolio ranges from single detector devices and photon

counting modules to full imaging systems. The academic group has published widely their interest in the field of the creation of arrays of enhancement mode single photon Geiger-mode detectors, e.g. Jackson et al [80]. SensL now advertise datasheets for full camera system products with embedded backside illuminated sensor arrays bump bonded to a CMOS read-out IC, complete with peltier cooler, e.g. [81].

The company has also been at the forefront of the push to replace vacuum tube technology with silicon based photo-multipliers, for combined PET-MRI medical imaging systems.

2.7.4 SPAD-TDC Arrays

Research groups at Ecole Polytechnique Fédérale de Lausanne (EPFL) have published widely for many years in the field of SPADs, TDCs and integrated array solutions. The ground breaking work of Rochas, Besse, Pauchard, Popovic et al [48, 49] in SPADs and later Niclass, Charbon et al [68], [82] in SPADs, TDCs and 3D imagers constitutes a significant contribution to the state of the art. The implementations of [68], [82] are of particular relevance to this section.

The time of flight (ToF) mode, pulsed laser illuminated 3D imager of [68] combines the distributed clock, ‘coarse-medium-fine’ TDC introduced in section 2.6.6 with a 128x128 NMOS passive quench SPAD configuration such as that shown in Figure 24B. Each event driven, 100ps resolution TDC is shared by 4 columns of SPADs, with PVT calibration being performed via a DLL. Implemented in 0.35 μ m CMOS, (complete with 8Gbps readout mechanism) despite the 40mm² die floor-plan in which the shared TDCs are located adjacent to the SPAD array, the 25 μ m diameter detectors had ~6% fill-factor, demonstrating a requirement for pitch matched, integrated microlenses. Additionally, the detectors had limited PDE at >800nm wavelength, a bandwidth which is very desirable for 3D cameras.

To address the limitations of this ToF solution, an alternative was sought, leading to the publication of a phase extraction ranging method called ‘Single Photon Synchronous Detection’ shortened to ‘SPSD’ [82]. The 60x48 pixel array was implemented in the same 0.35 μ m CMOS process, but this time the data computation logic was incorporated within each 85 μ m pixel, resulting in a 36mm² die size, and a system capable of outputting combined range and intensity data at 22 frames per

second. An array of modulated IR LEDs formed the illuminator for this complete LIDAR system.

Two other SPAD-Converter arrays have been implemented as part of the Megaframe project. Gersbach, the author et al in [83] presented an alternative, distributed clock-delay line hybrid TDC based array. Whilst attempting to address the same target specification and exhibiting good linearity, this topology was more difficult to scale to larger arrays and had problems with high overall power consumption.

In [84], Stoppa, Borghetti, the author et al implemented a Time-Amplitude-Converter (TAC) array version of the same device. This device was also fully functional, but suffered from higher than desired quiescent power consumption, and as a result the topology implemented as part of this research activity was chosen for the final larger format Megaframe array.

2.8 Conclusions

This chapter has introduced the main performance metrics for both SPADs and TDCs. A review of the state of the art for both SPADs and TDCs has also been presented. The use and application of standardised metrics is essential in order to perform qualitative analyses of the various detector constructions and time-digital converter topologies, and build an understanding of the tasks that each of these functional blocks must perform. The SPAD metrics of photon detection efficiency, timing resolution, dark count rate, dead time and fill factor are particularly relevant when considering a monolithic, nanometer scale, arrayed implementation. The TDC metrics of time resolution, accuracy, precision, conversion rate, power consumption and area are similarly key considerations. It has also been demonstrated that the silicon manufacturing process chosen also plays an important overall role in a technology development such as that embodied by this work. Finally a review of relevant, existing integrated detector-converter technologies has been provided.

The following chapters describe both the novel SPAD and TDC architectures developed as part of this research, as well as their respective characterisation results. The optimisation of both passive and active quenching approaches is presented. Key design differentiators which permit the construction of detector-converter pixel arrays are highlighted along with any resulting compromises. The integration of these two key elements is described and results are discussed.

3 SPAD Design

In this chapter a SPAD detector target specification is presented. The relevance of key features of the ST Microelectronics' IMG175 130nm Imaging CMOS process are then discussed with regard to SPAD design. Following from this the physical implementation options for various SPAD detectors are reviewed. An earlier implementation of a high DCR SPAD in a similar technology is analysed and then compared with the range of new detectors, based on the enhancement mode construction discussed in section 2.3.4.3. Finally, performance expectations for the new detectors are derived and discussed and conclusions are drawn.

3.1 Target Specification

The goal of the SPAD design part of this research was the creation of a low dark count rate, single photon detector which had good timing resolution and high detection efficiency over a broad wavelength range. The detector was to have a low reverse breakdown voltage and low dead time, but this was to be achieved without a dramatic reduction in the device drawn active area. The detector had to be implemented in an unmodified manufacturing process, without the added complexity and cost of SPAD specific implants with low doping concentrations such as that reported by Gersbach with the author et al in [85]. The new detectors were required to be afterpulsing free and compatible with on chip quenching and data processing electronics. The chosen architecture also had to be scalable, without the minimum diameter and quasi-neutral zone timing limitation imposed by implanted guard rings as discussed by Faramarzpour et al [21] and Cova et al [86]. For the above reasons the enhancement mode structure (Figure 15) was chosen for this work. Realistic targets based on state of the art detectors previously created in AMS 0.35 μ m high voltage technology were set for each SPAD metric, summarised in Table 3.

Metric	Quantity*	Units
Reverse Breakdown Voltage	-10 to -20	V
Excess Bias Voltage Range	1.2 ± 0.6	V
Average Dark Count Rate	<100	Hz
Peak Photon Detection Efficiency	≥ 25	% @ ~500nm
Minimum Dead Time	< 50	ns
Afterpulsing	<0.5	%
Timing Resolution @ FWHM	<200	ps
Active Area	~50	μm^2
Homogeneity	invariant active region	-

Table 3: SPAD Design Performance Goals

* - target values defined at room temperature.

3.2 Processing Considerations

3.2.1 Introduction

CMOS processes targeting image sensor implementations have undergone significant development in the past decade. Competition and the desire for ever greater image quality and spatial resolution has driven the semiconductor manufacturers down a path of continued optimisation. A major driver has been the high volume camera phone sector. This development now means that instead of the original intention of utilising digital processes which were imaging capable, there now exists customised imaging processes which are digital capable. Economies of scale still exist despite the comparatively high mask costs with foundries now using 12" manufacturing lines. These processes provide an opportunity to apply imaging optimised performance, small feature size, modern packaging techniques such as through-silicon-via (TSV) and economies of scale to TCSPC systems. It should also be noted that process selection for SPAD test chip implementation is also partly driven by multi-project wafer (MPW) scheduling, a factor particularly relevant in the Megaframe Project planning.

3.2.2 Designing SPADs in the IMG175 Process

ST Microelectronics' IMG175 CMOS Imaging process, reported by Cohen et al in [22], was chosen for the fabrication of the devices defined in this research. This proven technology is based on a 130nm, triple well, twin tub donor process, with a 90nm copper-aluminium metallization-dielectric stack optimised for imaging. Due to the doping concentrations and implant depths for such processes which are required to implement fast, dense digital circuitry, the depletion regions of implemented SPADs have narrowed and junction electric field strengths have increased. Thus a major contributor to DCR has become band to band tunnelling. As well as the side effects of high doping levels, the use of shallow trench isolation (STI) in such processes is known to increase stress and charge traps thus increasing dark count and afterpulsing, Finkelstein et al [52]. The benefit of the use of STI is increased electrical and optical isolation between detector elements, so enhancing an array's spatial resolution.

Another feature of ST's IMG175 process is the presence of an optional deep N implant formed by a high-energy ion implantation step before n-well formation. This deep N-implant is contacted by a ring of n-well and is normally used to completely enclose p-well regions in order to isolate NMOS transistors from the remainder of the substrate. This implant feature, 'NISO' (N-isolation), has been used to minimise noise-coupling issues and improve latch-up immunity.

A major benefit of the ST process is the availability of additional imaging specific implants, intended for conventional photodiode implementation but providing scope for potential innovative uses in SPAD designs. The process's reduced height optical stack with anti-reflective coating provides inherently good photon transmission, enhancing photon detection efficiency.

In order to build a detailed and clear understanding of the construction of the P and N type implants available in the process, as well as the order in which they are applied, a table of process steps was compiled. This is shown in Appendix C for reference, and is a key contributor to choosing the correct layers for building the component parts of a SPAD design: namely the anode, cathode and guard ring. This table permits the simultaneous viewing of the chronological order of implants, as well as doping element, concentration and depth. This also allows the designer to easily build an understanding of which implants are masked by deposition steps such as oxide, nitride and polysilicon. Further, some masks used in the fabrication process are not explicitly

created by the designer, but are generated from a Boolean computation of other drawn layers, also shown in Appendix C.

In the sections which follow, a common colour key is used throughout, and is shown in Figure 37 .

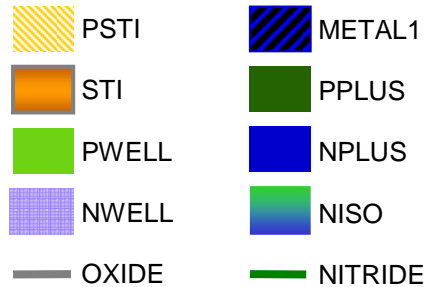


Figure 37: Process Layer Key

As a general guide, green is used for P type implants, blue for N type. Implants which have not been utilised during the course of the research have been coloured grey. Where possible, layer fill usage is aligned with the ST-Cadence design kit layout tools.

3.2.3 Process Implant Toolbox

In preparation for the SPAD design phase, the depth and doping concentration for each implant had to be known. A guide to implant depth can be taken from Figure 158 in Appendix C which is referenced from [87]. Knowing the implant element and energy permits a first order ranking of available implant depths. Once tabulated alongside doping concentration (in atoms/cm³), potential candidates for anode, cathode and guard ring become clearer, and conversely which implants are not suitable. This information is shown for IMG175 N-implants in Figure 38 and Figure 39, and for P-implants in Figure 40 and Figure 41.

The key design decision of terminal orientation can then be considered. In this processing technology all active devices are implemented in a P-type epitaxial layer (P-EPI) which extends some μm from the surface into the silicon substrate. In order to create a device without one of the terminals permanently hard-connected to the substrate potential, the SPAD should be implemented in its own deep NWELL region. This is desirable due to the requirement to have a grounded, non-moving substrate potential, permitting the integration of other active circuitry. The design of the

detector must then always have the breakdown voltage of the SPAD anode to NWELL junction to be safely below that of the deep NWELL to P-EPI junction. Additionally, implementation of the detector in its own N-implant deep well permits inherent gettering, as discussed in section 3.3.

With these established guidelines, the potential construction of the cathode can be considered using the N-implant tables.

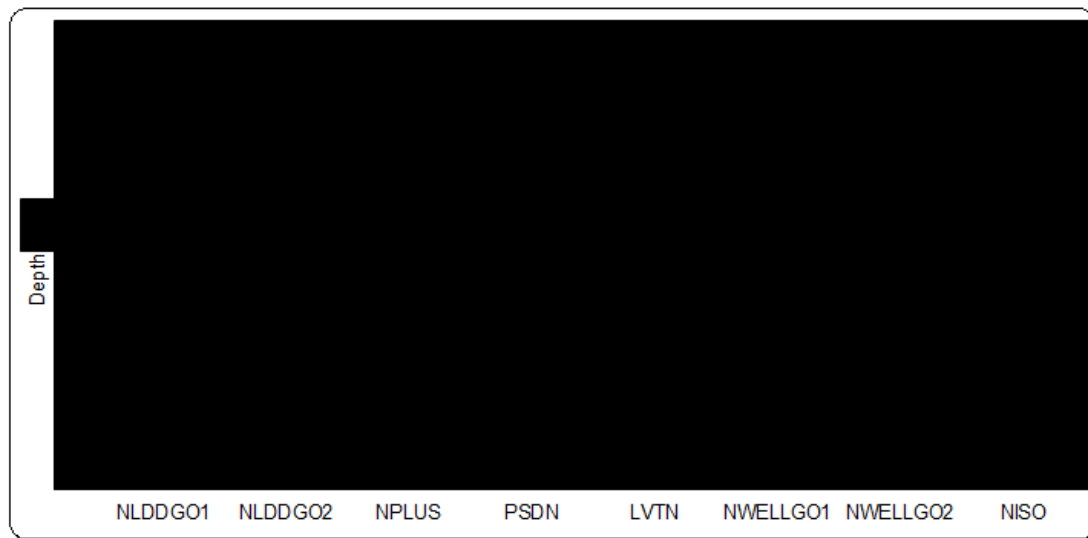


Figure 38: N-Implant Depth

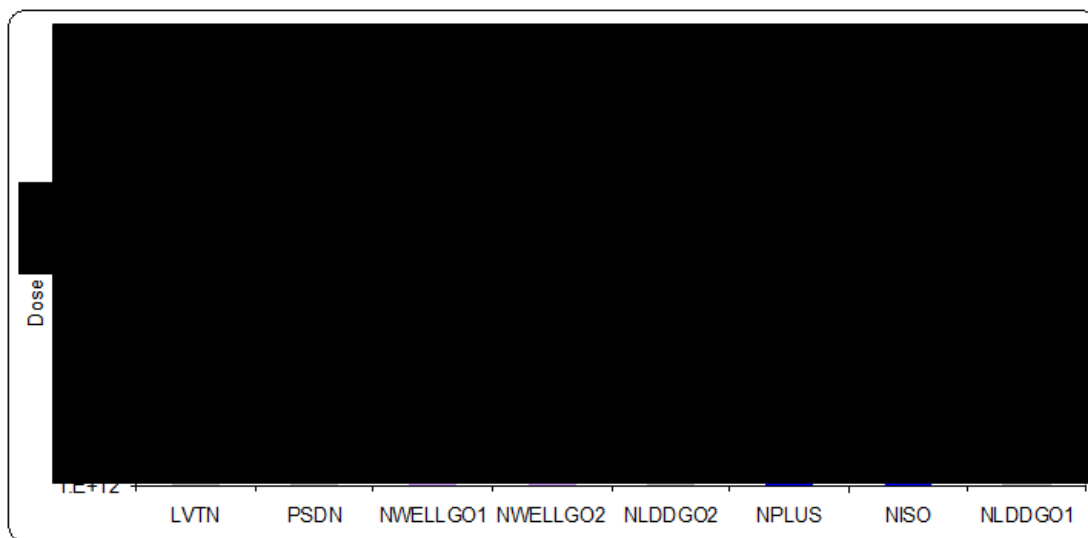


Figure 39: N-Implant Doping Concentration

The cathode side of the junction is required to be within the P-EPI depth, but extending as deep as possible, with a general rule of having a doping concentration lower than that of the anode. Therefore, in this process NWELL and NISO are possible candidates, with connection down to the deep cathode being performed conventionally by a combination of NPLUS-NWELL implants.

The anode terminal can then be similarly assessed using the P-implant tables.

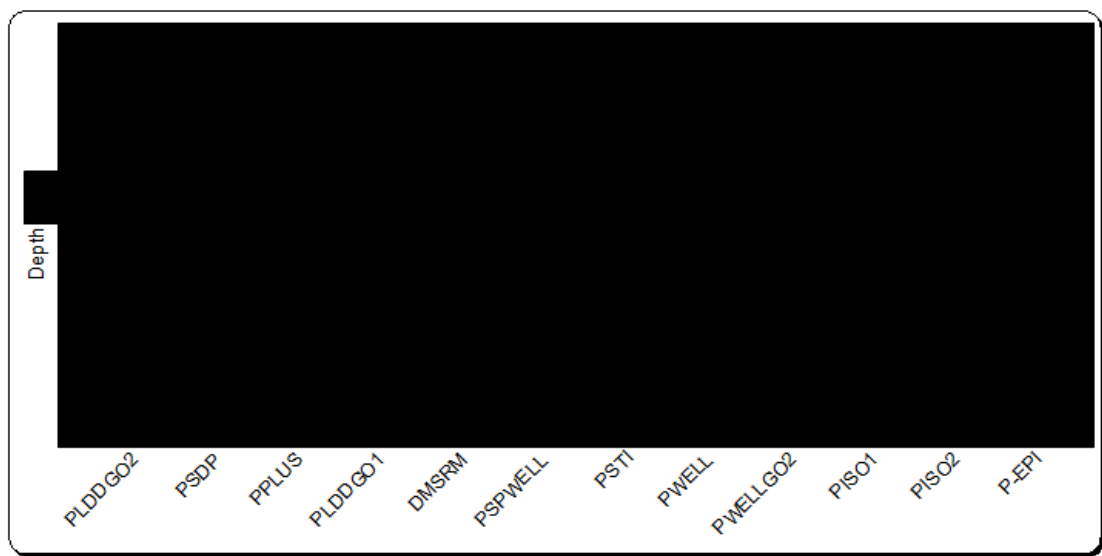


Figure 40: P-Implant Depth

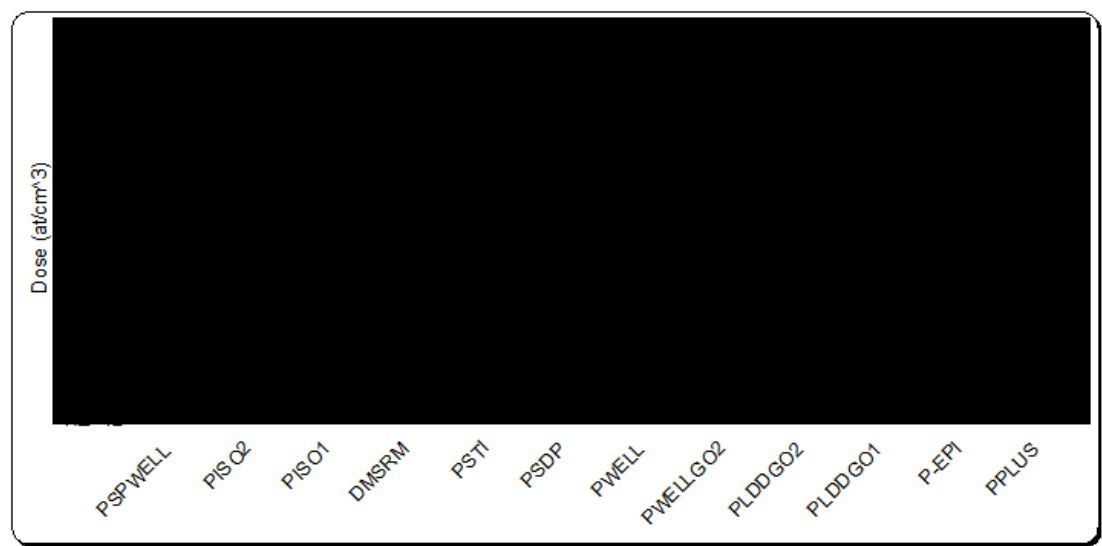


Figure 41: P-Implant Doping Concentration

Considering the likely use of NWELL/NISO for the cathode, the anode depth is required to be significantly shallower, with likely higher total doping concentration. This indicates implants with depths from and including PISO1 (indicated by dashed red arrow) as well as PWELL, PSTI. Doping concentration is always considered as total atoms per cm³ indicating that several implants may be used to increase overall concentration, noting that the overall design goal is the implementation of enhancement mode structures.

The guard ring implementation for each structure is presented in section 3.6.

3.3 *Gettering*

Gettering, in the context of silicon fabrication, is the removal of an unwanted element from a specific area via the use of a localised sacrificial implant (normally phosphor based), or other material such as surface oxides or polysilicon. This is done to enhance the purity of a certain key area, such as the active region of a SPAD. Impurity clusters are known to cause localised areas of high field within SPADs, resulting in premature avalanche ignition without the involvement of a photon arrival. The contribution of certain implants and their physical location during the gettering process is a key design consideration. This is because the use of gettering structures is known to significantly reduce the DCR whilst also enhancing yield, as discussed by Zanchi et al in [18] and Sciacca in [88]. In [18] Zanchi et al demonstrate that phosphor based N-implants make particularly effective gettering agents. In this publication, a highly doped phosphor implant ring is drawn around a SPAD structure. This resulted in a centrifugal attraction of impurities toward the gettering ring placed $\sim 20\mu\text{m}$ from the device periphery. As a consequence a dramatic DCR reduction by at least an order of magnitude was reported.

In the target process for this research, NPLUS, NWELL, NLDD and NISO have a high phosphor doping concentration. This provides an opportunity to either implement gettering inherently within the SPAD device itself or as a stand-alone external structure. The use of these implants with respect to the gettering function is discussed for each structure in section 3.6.

3.4 *Physical Optimisation*

In addition to the formation of the SPAD terminals and guard ring features, other physical aspects of the detector must also be considered. Three key aspects are the optical stack above the diode, the overall SPAD shape and the position of device connections. Each of these can significantly affect the performance of the SPAD in unique ways and are discussed in the following sections.

3.4.1 Optical Stack Considerations

The metallisation layers immediately above the SPAD should be drawn so as to maximise photon transmission into the active region, and minimise reflections. The 90nm optical stack in the IMG175 process has been optimised for imaging applications and so has certain in-built features which help to achieve this. Firstly, restricting the number of metal layers implemented over the detector (i.e. using an optical cavity) reduces the chance of reflection and so boosts the PDE. The standard optical stack versus optical cavity stack height is shown in Figure 42.

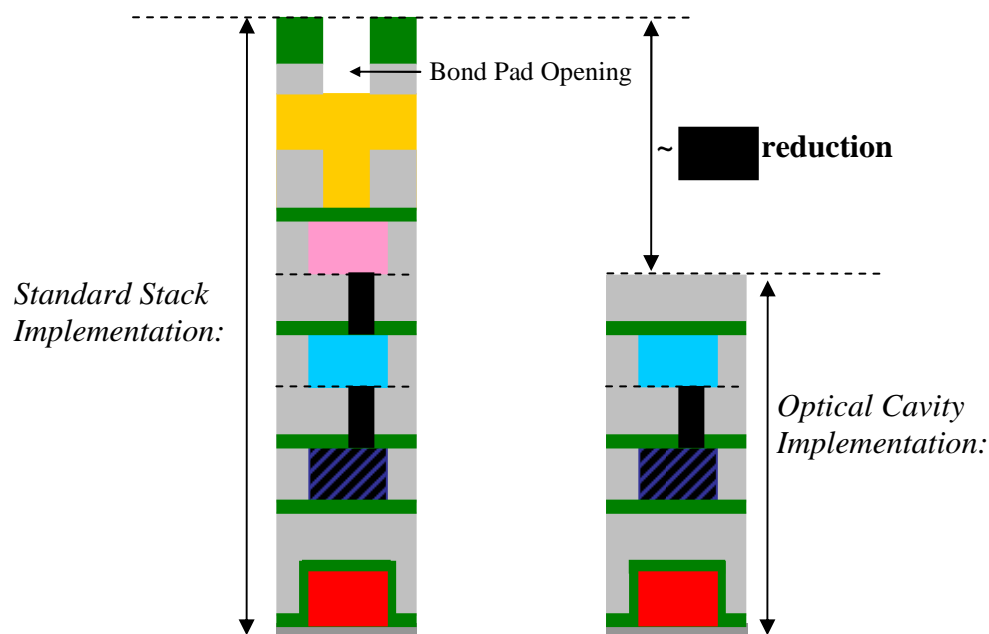


Figure 42: Optical Stack Constructions

It can be seen that using cavity reduces the optical stack by $\sim X$ μm .

The second imaging enhancement feature which can be taken advantage of is the ‘porting’ of the nitride barrier layers. These layers are required to maximise copper metallisation reliability and are normally practically planar across the wafer, sandwiching every copper metal layer (M1-3). However, they have a high refractive index and so can cause a small amount of reflection of incoming light. To prevent this, a layer may be drawn which causes the nitride barrier to be removed, for example above the SPAD active region.

Microlens may also be used to steer photons into the active region rather than striking a non-sensitive part of the detector. However, the microlens resist material available in the process is designed for small ($<2\mu\text{m}$) pixels, and so is not suitable for larger dome heights which would be required for SPADs of say $15\mu\text{m}$ outer diameter. To summarise, it may be feasible to use the standard lens resist to achieve small benefits, apply non-standard resist materials, or use a form of printed microlens in order to maximise overall device fill factor. Figure 43 shows a full cross section, indicating the combined possible use of the optical stack features above an example detector.

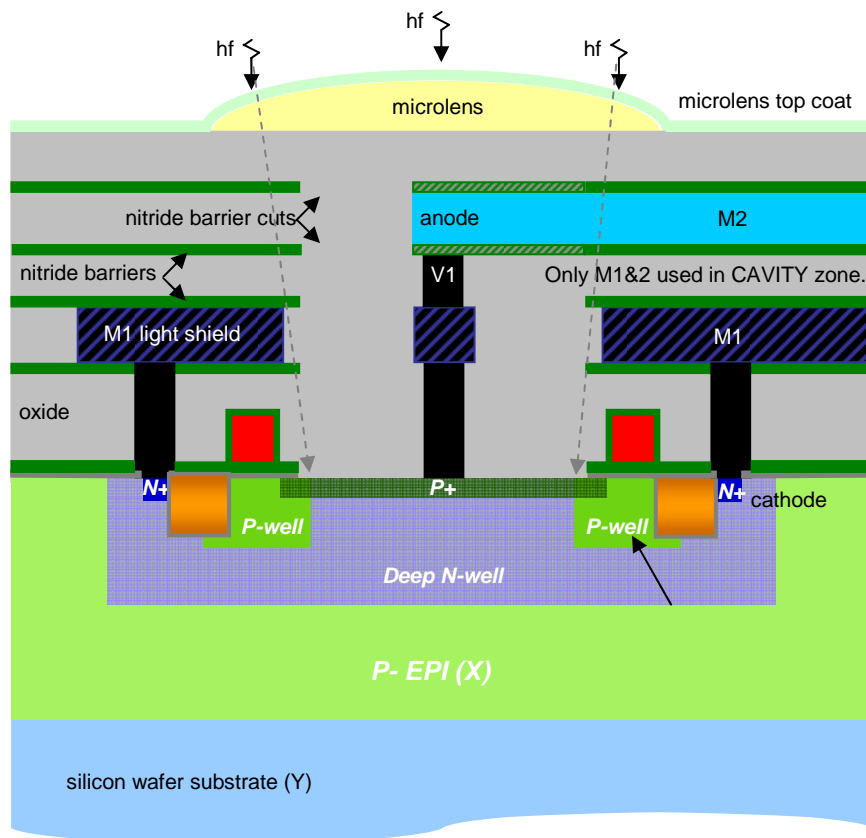


Figure 43: Use of Optical Stack Features

The microlens top coat material is made from SiON and mainly aids in rejecting the attachment of debris to the chip top surface. The refractive index of the SiON is engineered to allow it to act as an anti-reflective coating between air and microlens material.

The measured impact of the application of these features is discussed in section 4.

3.4.2 Detector Shape

SPADs have been implemented in a range of shapes, as shown in Figure 44. The green shaded area represents each SPAD's active region. The choice of shape used is a trade off between fill factor and the minimisation of corner fields.

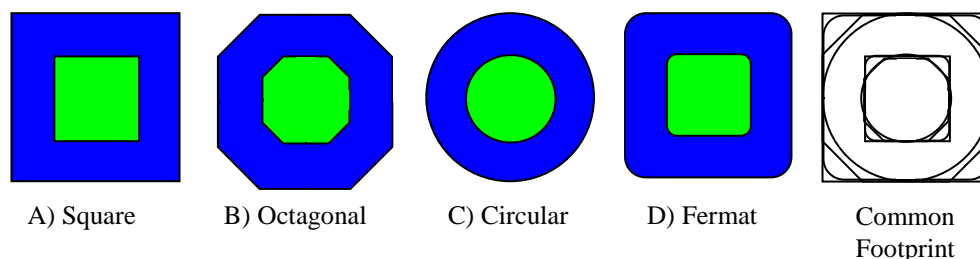


Figure 44: Detector Shape Options

A square shape (A) maximises fill factor for an overall usable footprint area, but sharp corners can result in localised high electric fields which cause increased DCR [19]. The octagonal shape (B) provides a reduction in the severity of angles and relies on lithographic smoothing to ensure low fields at the vertices [20]. The 45° angle is also normally legal in most processes' design rule checks (DRC). The circular shape (C) ensures that there are no sharp vertices but is not normally DRC legal due to the drawn 'conics' polygons becoming detached from the mandatory design grid [85]. The relatively rarely used 'Fermat' shape (D) is more complex to draw but maximises fill factor without having sharp vertices [89], but again like the circular shape can suffer from DRC violations. However, circular and Fermat shapes can be configured as DRC legal, a technique which is presented in section 3.7.

To summarise, many other variations of the above shapes are quite possible, such as the rounded edge pseudo-triangular shape shown in [9], which allows for efficient placement of quenching and readout electronics and so enhances overall fill factor. However, during the course of this research the shapes shown above have all been implemented. The measured impact of the application of these features is discussed in section 4.

3.4.3 Anode Contact Positioning

It can be seen from Figure 43 that the connection to the anode is made by means of a metal track and stack of via-contact down to the PPLUS implanted active region. This

structure blocks incident photons from striking the active region. Therefore the anode contact may be moved toward the periphery, as long as a uniform field can be formed over the active region. This technique is shown in Figure 45 using the round SPAD shape as an example.

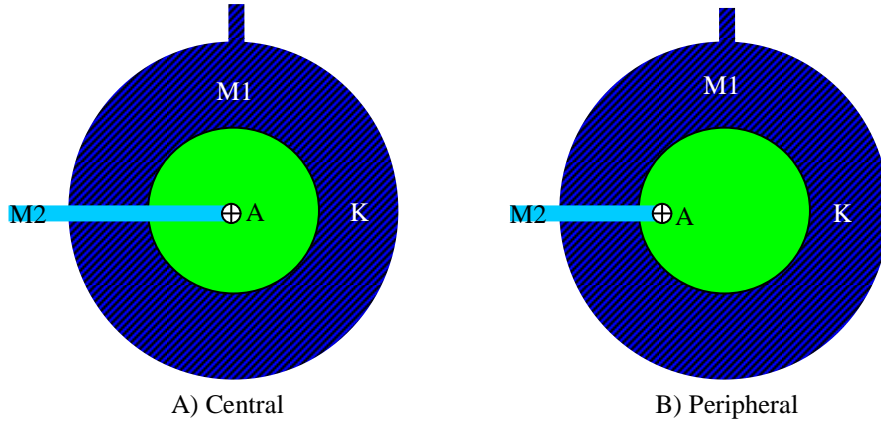


Figure 45: Anode Contact Positioning

The result of this layout technique is enhanced PDE. Section 3.7.3 shows the application of this to a Fermat shaped SPAD.

3.5 Prior Art Implementation

The implanted guard ring detectors reported by Niclass et al [20] and Gersbach et al [38] have been implemented in a similar, unmodified ST Imaging CMOS fabrication process. The detectors implemented were hexagonal and round shaped respectively with centrally positioned anode connections. However, both report prohibitively high DCR of >80KHz. In both cases this is thought to be primarily tunnelling induced, caused by high electric field strength due to the high doping concentrations of the PPLUS and deep NWELL implants used. Other reported metrics of these devices are comparatively good, e.g. 40% PDE, ~140ps timing resolution and low afterpulsing. To implement the more desirable enhancement mode structure whilst addressing the high DCR problem without impacting other metrics, first order calculations were performed to ascertain terms of reference. These are shown in section 3.5.1.

3.5.1 Device Parameter Calculations

Assuming the approximation of an abrupt PPLUS-NWELL junction, first order calculations can be performed in order to determine approximate junction electric

field strength, capacitance, breakdown voltage and avalanche electron flow. These calculations are presented below. Once the key parameters have been estimated, TCAD Process modelling can be used to confirm the results and perform cross section analysis.

Firstly the thermal voltage, V_t , is calculated:

$$V_t = \frac{k \cdot T}{q} \approx 26mV @ 300^\circ K$$

Where k is Boltzmann's constant, T is (room) temperature, q is the charge on an electron (1.602×10^{-19} coulombs)

The built-in potential ϕ_0 is:

$$\phi_0 = V_t \cdot \ln \left(\frac{Na \cdot Nd}{n_i^2} \right),$$

Where $Na=(X) \times 10^{17}$, $Nd=(Y) \times 10^{17}$, $n_i=(Z) \times 10^{10}$ are the doping concentrations in atoms/cm³ for acceptor side, donor side and intrinsic silicon at 300°K.

$$\therefore \phi_0 = 0.906V$$

Note: Na and Nd doping concentrations of PPLUS and NWELL respectively are non linear with depth, and are determined from TCAD process simulation.

Next, the depletion widths can be calculated (at the approximate breakdown voltage):

$$W_1 = \sqrt{\frac{2 \cdot \epsilon_{si} \cdot (\phi_0 - V_d) \cdot Nd}{q \cdot Na \cdot (Na + Nd)}}$$

$$\therefore W_1 = 1.331 \times 10^{-5} cm$$

$$W_2 = \sqrt{\frac{2 \cdot \epsilon_{si} \cdot (\phi_0 - V_d) \cdot Na}{q \cdot Nd \cdot (Na + Nd)}}$$

$$\therefore W_2 = 9.181 \times 10^{-6} cm$$

Where ϵ_{si} is the permittivity of silicon ($\sim 1.04 \times 10^{12}$).

The peak electric field strength is then determined (from the integral of Poisson's equation) as:

$$E_{\max} = \frac{q \cdot Na}{\epsilon_{si}} \cdot W_1 \quad (\text{i.e. E-field is at a maximum at the junction centre, } x=W_1.)$$

$$\therefore E_{\max} = 775 \times 10^3 V/cm$$

Thus to address the tunnelling induced DCR issue for a new detector design, the target for the maximum electric field should be less than 775×10^3 V/cm.

The reverse breakdown voltage V_{BD} can then be evaluated to ensure process compatibility (target <20V).

$$V_{BD} = \frac{\epsilon_{si} \cdot (Na + Nd)}{2 \cdot q \cdot Na \cdot Nd} \cdot E_{\max}^2$$

$$\therefore V_{BD} = 8.72V \text{ (process compatible)}$$

The device capacitance is a key parameter that affects dead time, afterpulsing and discharge characteristics.

The junction capacitance for a $50 \mu\text{m}^2$ area PPLUS-NWELL device ($8 \mu\text{m}$ ϕ) is calculated with the breakdown voltage applied across the diode:

$$C_j = A \cdot \sqrt{\frac{\epsilon_{si} \cdot q \cdot Na \cdot Nd}{2 \cdot (Na + Nd)}} \cdot \frac{1}{\sqrt{\phi_0 - Vd}}$$

$$\therefore C_j = 23 \text{ fF}$$

This figure does not include metallisation parasitic capacitances, which must be extracted from the device layout using a suitable extraction software toolset such as Synopsys ‘Star-RCXT’. (Typical value for compact layout ~25fF.)

For an excess bias voltage (V_{eb}) of 1.2V a full discharge electron flow through the junction is therefore:

$$e_{\text{discharge}} = (C_j \cdot V_{eb}) / q$$

$$\therefore e_{\text{discharge}} = 173 \text{ ke}^-$$

This figure is also essentially the photoelectric gain of the detector.

The basis of the calculations made can then be checked via 2D TCAD modelling*, as shown in section 3.5.2 which follows. 2D simulation has limitations when compared to volumetric 3D, such as when modelling current flow paths.

(* performed by Eric Webster, The University of Edinburgh, using Synopsys ‘Sentaurus’.)

3.5.2 Process Simulation

TCAD simulation of the PPLUS-NWELL junction confirms the first order approximation calculations. Firstly the doping concentrations through the centre of the active region are simulated as a function of depth, as shown in Figure 46.

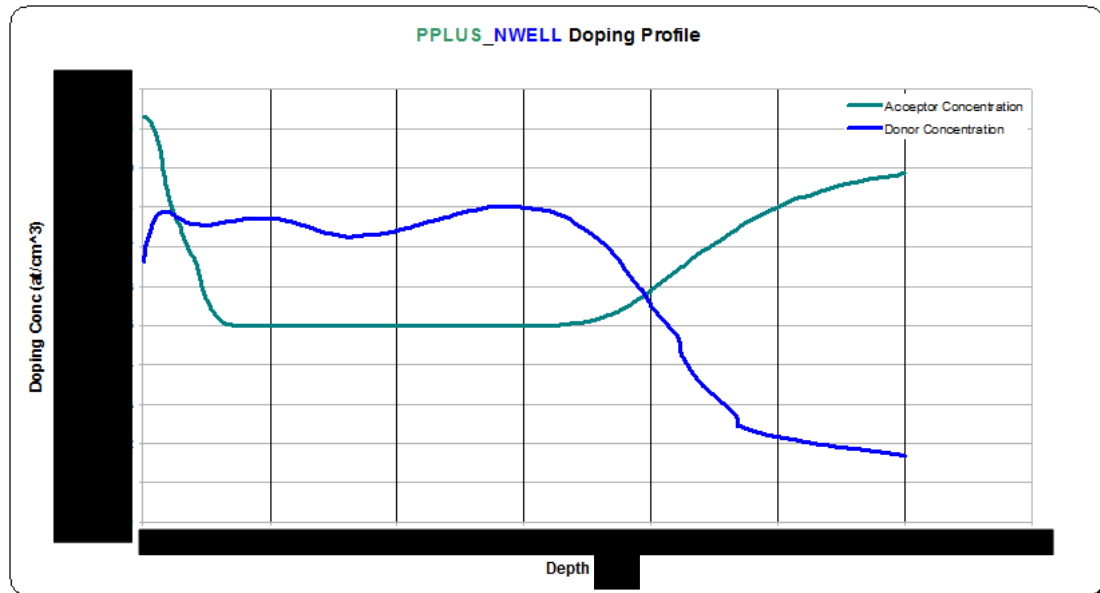


Figure 46: PPLUS_NWELL SPAD Doping Profile

The peak electric field can then be determined shown in Figure 47.

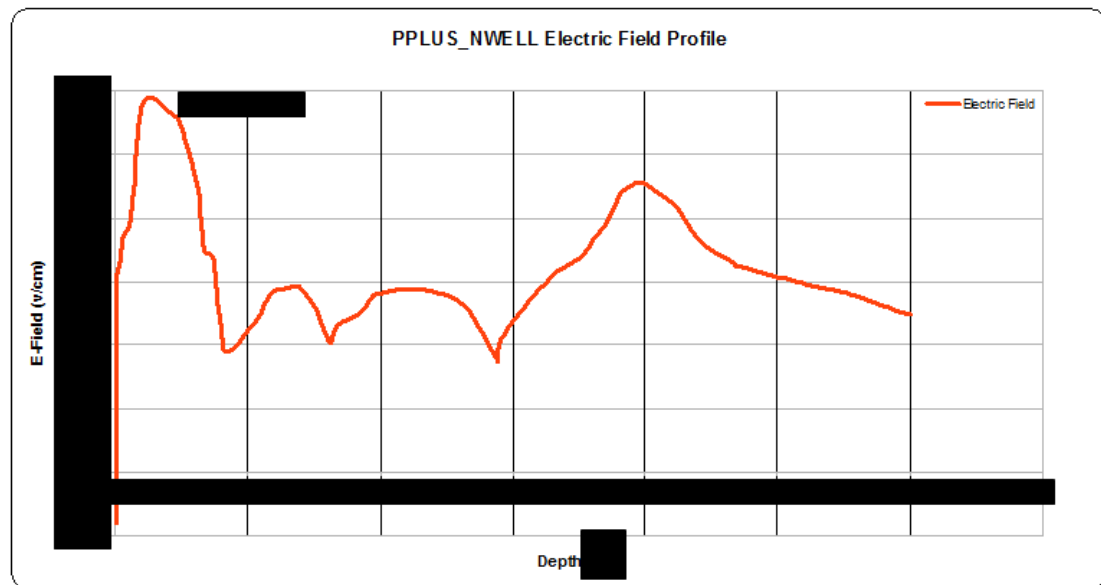


Figure 47: PPLUS_NWELL SPAD Electric Field Profile

As well as comparing favourably with the estimated result of 775kV/cm, this simulation also shows that the peak field occurs at [XX]nm depth (shallow), concurrent with the reported good short wavelength response and timing resolution of this device.

The simulated breakdown voltage also agrees with the calculated value and is shown in Figure 48.

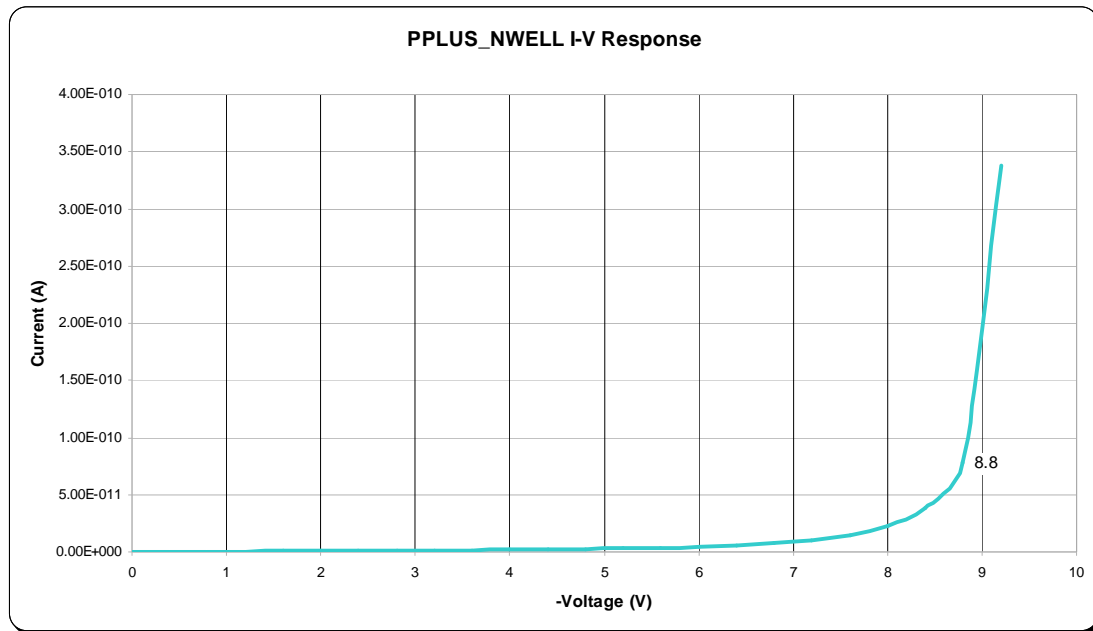


Figure 48: PPLUS-NWELL SPAD I-V Response

A TCAD cross section simulation of this SPAD can then be used to confirm prevention of edge breakdown. In the case of this structure a PWELL implanted guard ring is used for this purpose and is clearly seen in Figure 49.

In the TCAD simulation figures, a dark red line indicates the P-N junction, with the thin white lines denoting the depletion region boundary.

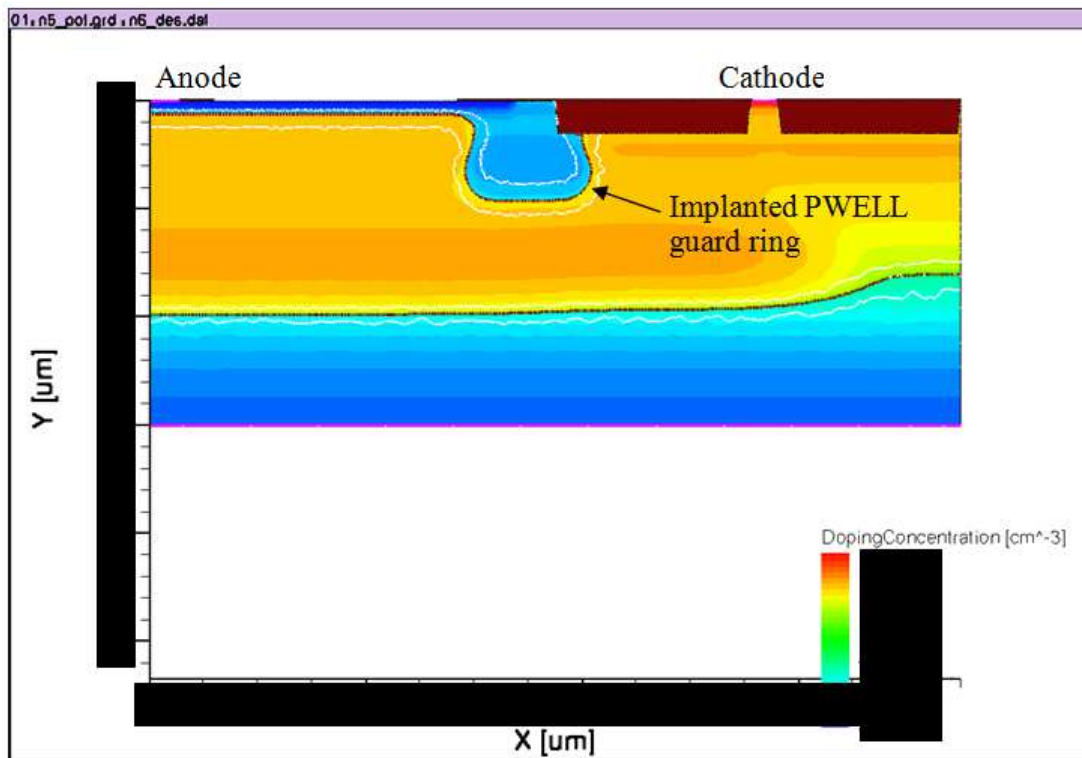


Figure 49: PPLUS-NWELL SPAD 2D Doping Simulation

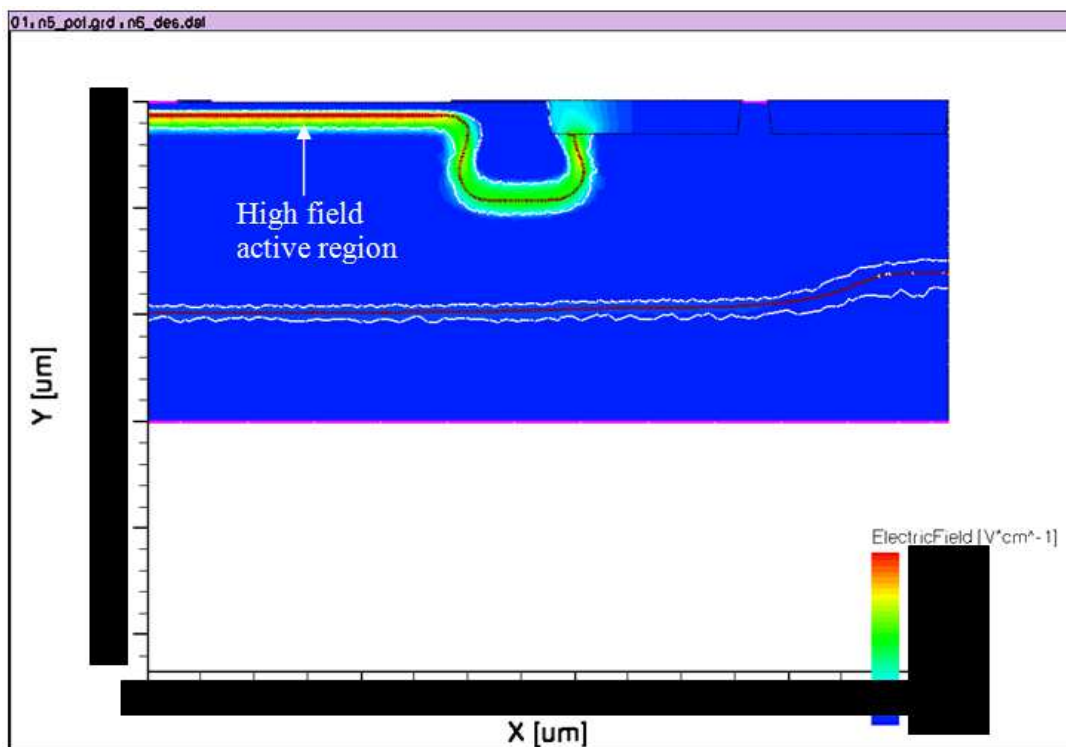


Figure 50: PPLUS-NWELL SPAD 2D E-Field Simulation

The E-field simulation shown in Figure 50 confirms the calculated junction maximum ξ , and the successful lowering of ξ at the device periphery.

3.5.3 Results Summary

Comparing the calculated with the simulated values indicates good correlation, shown in Table 4.

PPLUS_NWELL SPAD	E-Field (kV/cm)	Pk Depth (nm)	V _{bd} (V)	C _j (fF)	Gain (Aval ke ⁻)
Calculated	775	-	8.72	23	173
Simulated	753		8.8	25	187

Table 4: PPLUS_NWELL SPAD Results Comparison

As discussed in section 2.2, the existing implementations of SPADs in nanometer scale CMOS processes have resulted in high tunnelling induced DCR (Niclass et al [20], Finkelstein et al [19], Faramarzpour et al [21], Gersbach et al [38]) and/or low PDE (Marwick et al [40]). To lower the DCR of new devices without impacting other metrics heavily, the peak E-field should be reduced. Section 3.6 demonstrates how this can be done by use of alternative implants and guard ring structures to create several novel detectors.

3.6 New Detector Constructions

In this section new SPAD constructions are presented, along with associated simulation data. Conclusions are drawn for each structure. Characterisation results are provided in chapter 4. The naming convention used to identify each detector design uses the nomenclature of ‘anode_cathode’.

3.6.1 PPLUSPSTI_NISO SPAD

In this enhancement mode structure the anode is formed from PPLUS and PSTI. PSTI is an imaging process specific implant which is normally used for passivation of traps associated with the formation of shallow trench isolation (STI). Figure 40 and Figure 41 indicate that PSTI is both deeper and has a much lower doping concentration respectively than the PPLUS implant recipe, desirable for potentially lowering E-field and DCR.

The cathode *and* a new guard ring structure are then formed simultaneously by the use of drawing NISO *without* NWELL. This latter technique is performed by using the drawn layer ‘P-EPI’, which is used as a well blocking implant as indicated by the process generated mask boolean. This construction is shown in Figure 51.

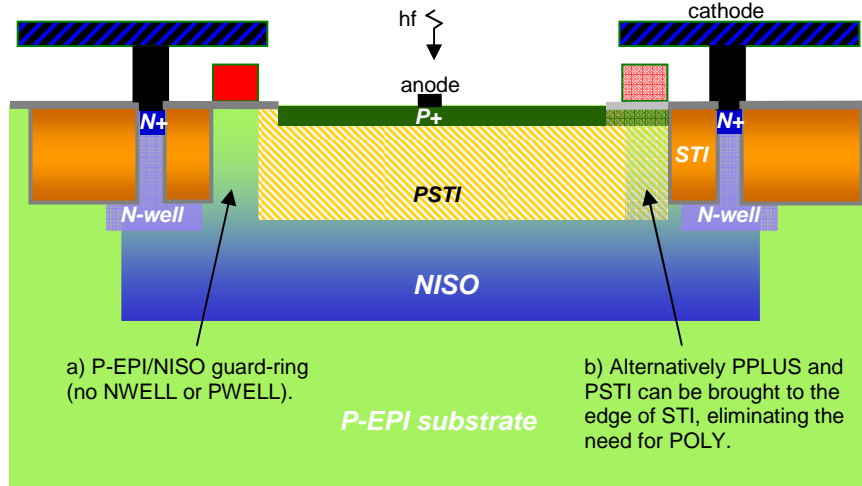


Figure 51: PPLUSPSTI_NISO_EPIPOLY SPAD Cross Section

The use of NISO without NWELL is unconventional, and creates a novel form of guard ring. The result is a progressively graded doping profile in the guard ring region, reducing in concentration near the substrate surface, as indicated by the shading of the NISO zone in Figure 51. This results in a lowered electric field at the periphery in comparison to the main P-N junction, the result of which is the desired enhancement mode structure.

The guard ring zone can be kept free of STI by use of a POLY deposition technique. This is shown in Figure 51 as guard ring (a). STI formation is known to introduce defects and crystal lattice stresses which cause high DCR so it is normally important to move the trench away from the main diode P-N junction. However, the passivation properties of the boron only based PSTI implant allow guard ring option (b), where the PSTI is brought to the STI boundary, negating the requirement for the use of POLY.

Finally, the connection to the deep NISO cathode is implemented by contacting to drawn NPLUS and NWELL at the outer edge.

TCAD simulation confirms that the doping concentrations and electric field have been sufficiently lowered as shown in Figure 52 and Figure 53 respectively.

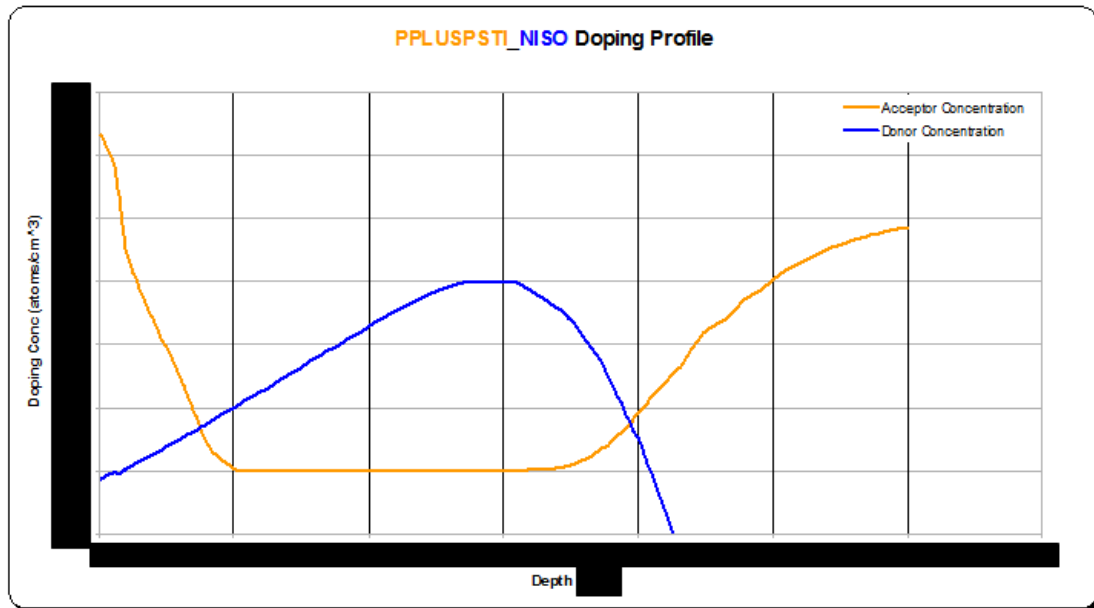


Figure 52: PPLUSPSTI_NISO SPAD Doping Profile

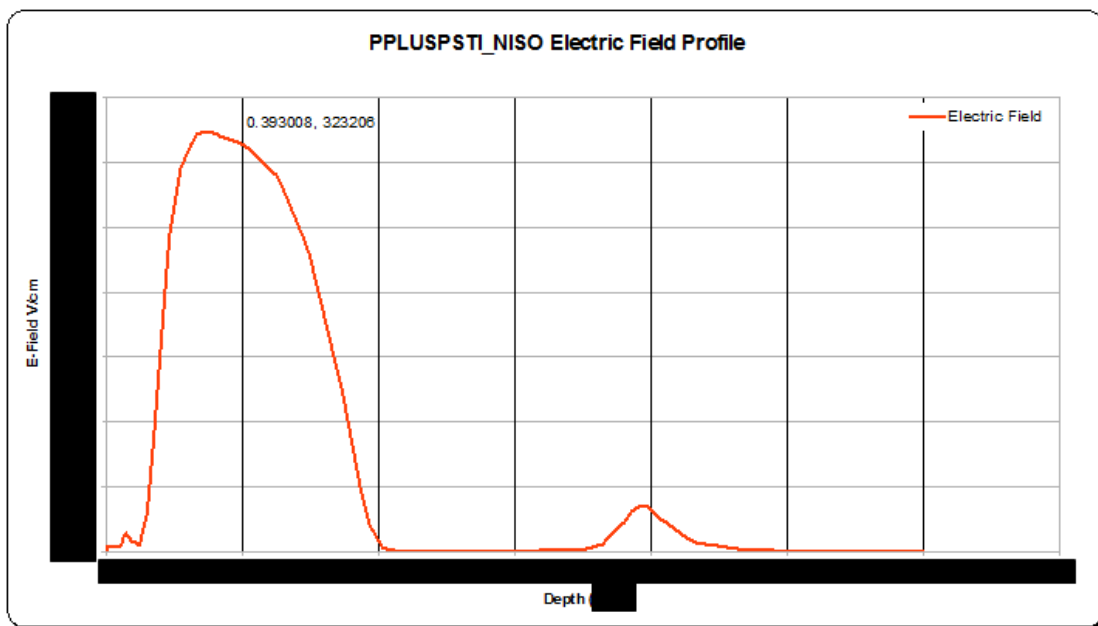


Figure 53: PPLUSPSTI_NISO SPAD Electric Field Profile

The doping profiles of this new SPAD yield a device with much lower electric field with the peak deeper in the substrate. Additionally, the lower doping concentration yields an increased breakdown voltage of ~17.5V (still within target specification) as shown in Figure 54.

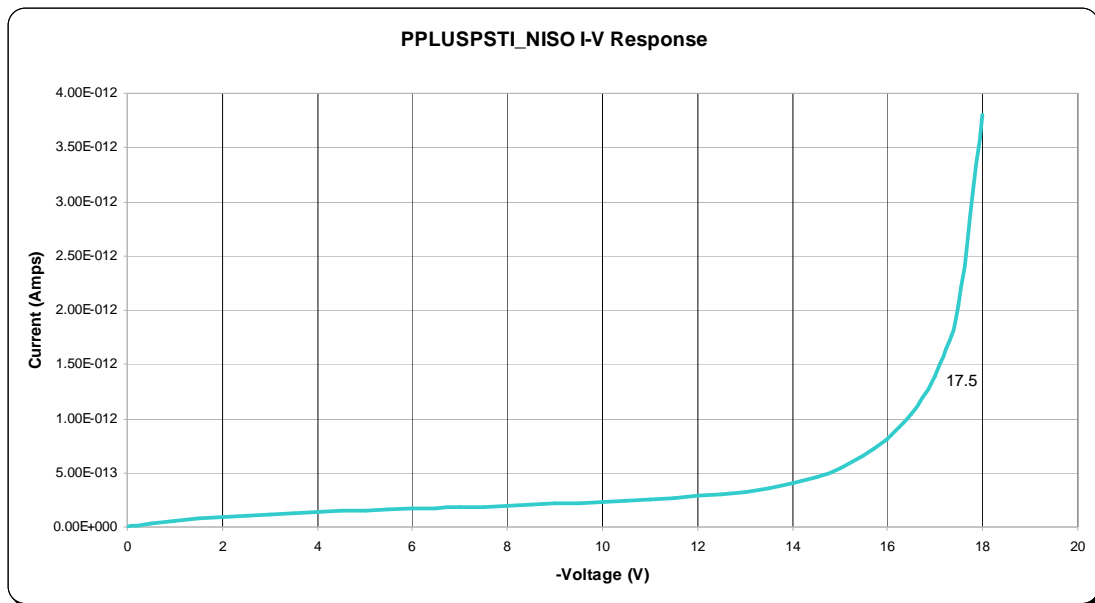


Figure 54: PPLUSPSTI_NISO SPAD I-V Response

The TCAD simulation of Figure 55 reveals some clear differences with the previous high DCR structure of Figure 49. The implanted guard ring has been replaced by an enhancement mode structure, with lower active and peripheral zone doping concentration as well as a significant reduction in the magnitude of the junction E-field. The TCAD simulation of guard ring option (a) is shown in Figure 55.

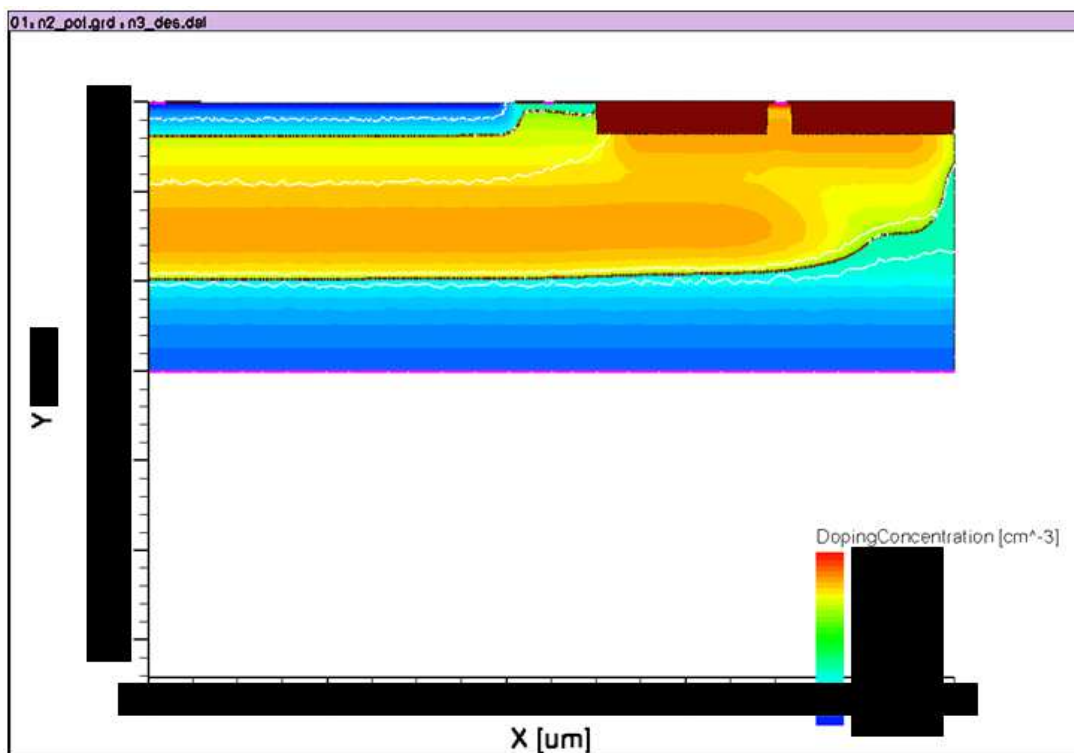


Figure 55: PPLUSPSTI_NISO SPAD 2D Doping Simulation

The electric field profile for this structure is shown in Figure 56.

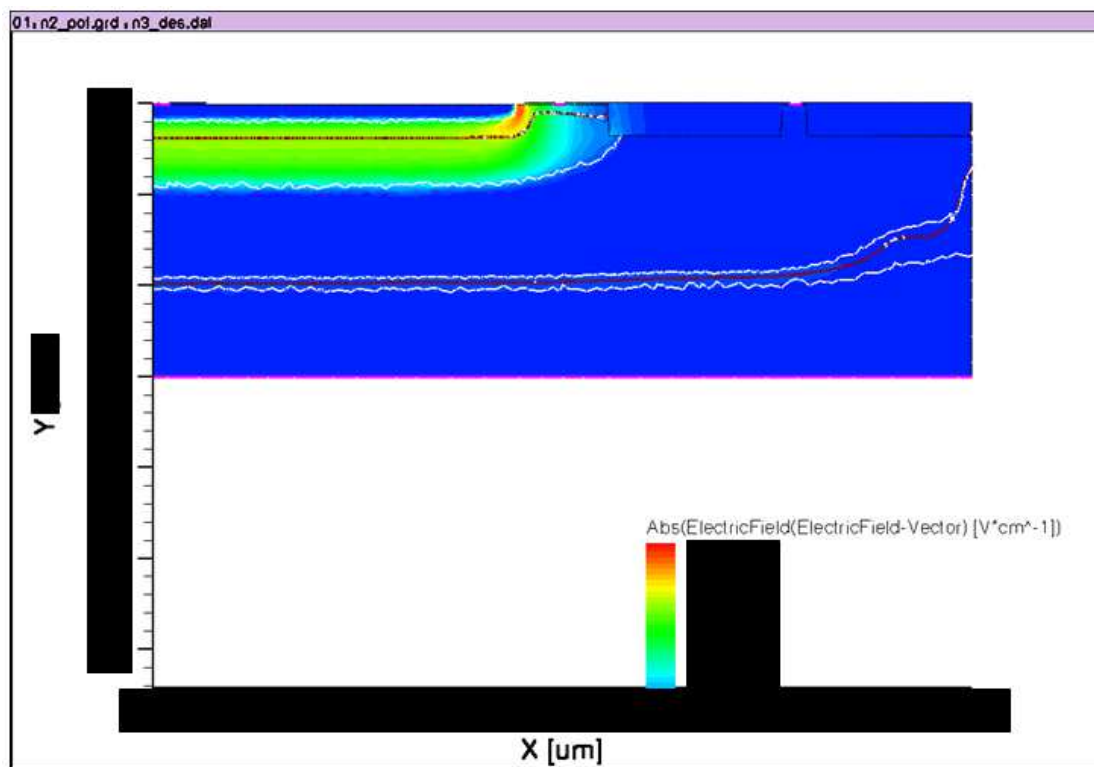


Figure 56: PPLUSPSTI_NISO SPAD E-Field Simulation

This shows increased field strength at the active region periphery, a phenomenon not clearly evident in the characterisation results for this detector presented in section 4.3.1. The device cross section can be implemented as the round, central anode Cadence layout which is shown in Figure 57.

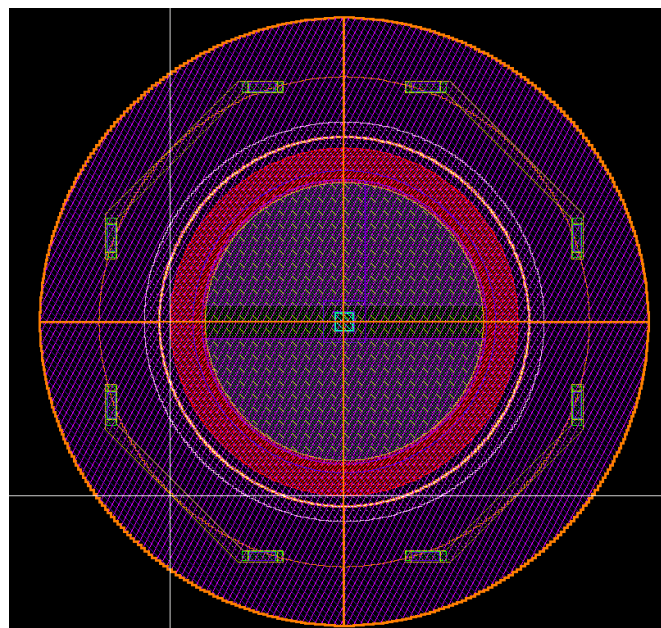


Figure 57: Cadence Layout for PPLUSPSTI_NISO SPAD

This detector has an 8 μ m active region diameter and an overall outer diameter of 17 μ m. The version shown has the EPI-POLY guard ring option. The nitride cuts for optimising photon transmission can be seen in the central active region.

3.6.1.1 Results Summary

Expected results for the device can be compared with the high DCR prior art, shown below in Table 5.

	E-Field (kV/cm)	Pk Depth (nm)	V _{bd} (V)	C _j (fF)	Gain (Aval ke ⁻)
PPLUS_NWELL	775		8.72	23	173
PPLUSPSTI_NISO	323		17.5	10	70
Comment	decreased	deeper	increased	decreased	decreased

Table 5: PPLUSPSTI_NISO SPAD Expected Results

It can be predicted that the reduction in E-field would have an appropriately favourable impact on tunnelling induced DCR. The deeper depletion region should favour a broadening of PDP wavelength response, and the reduced junction capacitance should further reduce afterpulsing probability. This is achieved at the expense of increased breakdown voltage, although this is still within the target of <20V. The other likely penalty to be incurred with a reduced peak E-field is that of degraded timing resolution.

The junction of the detector presented in section 3.6.1 can be modified by the addition of NWELL in the active region as shown below in Figure 58. This enhances the electric field in the centre, further promoting avalanche events there rather than at the device periphery. Like the previous design which utilises PSTI, two guard ring options are available.



102

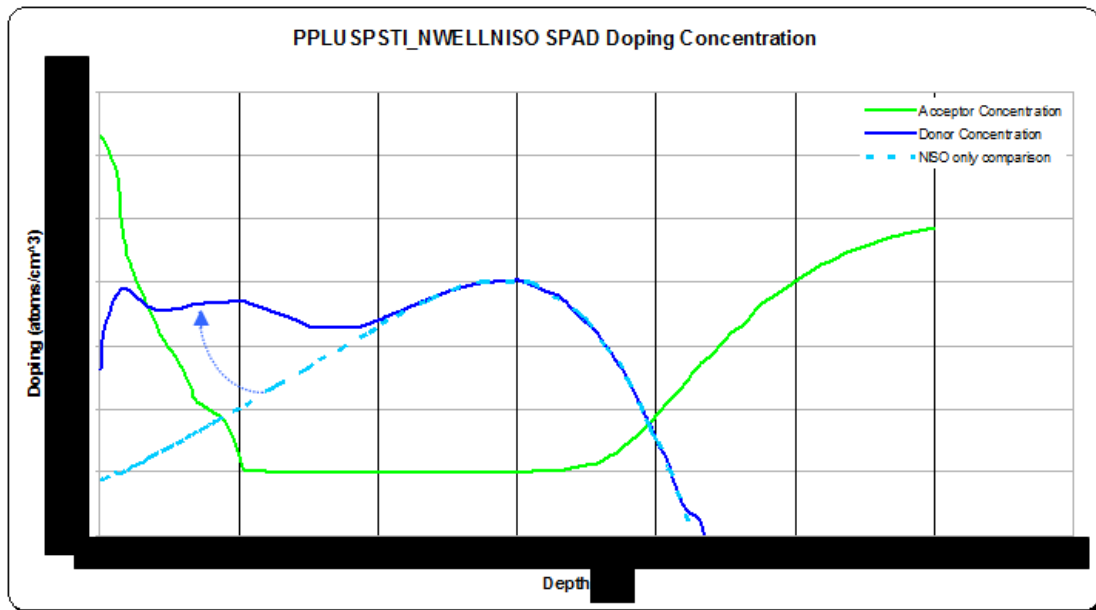


Figure 59: PPLUSPSTI_NWELLNISO SPAD Doping Concentration

The E-field simulation shown in Figure 60 indicates a higher peak response than the previous design, but deeper than that of the high DCR SPAD of section 3.5.

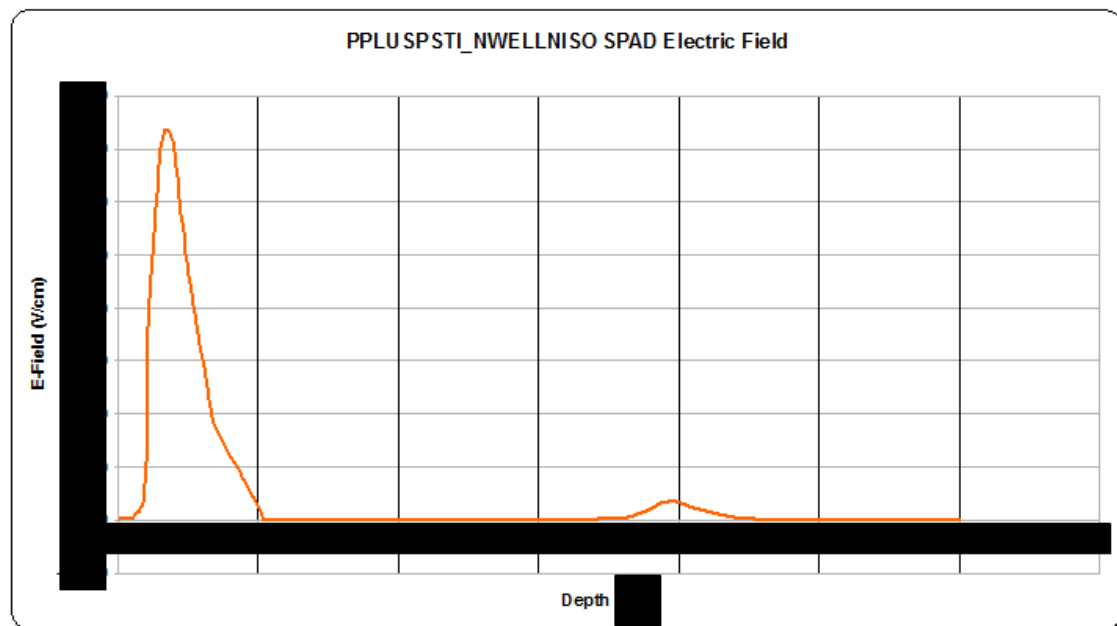


Figure 60: PPLUSPSTI_NWELLNISO SPAD Electric Field

The breakdown voltage of this detector is simulated as ~11.5V, shown in Figure 61.

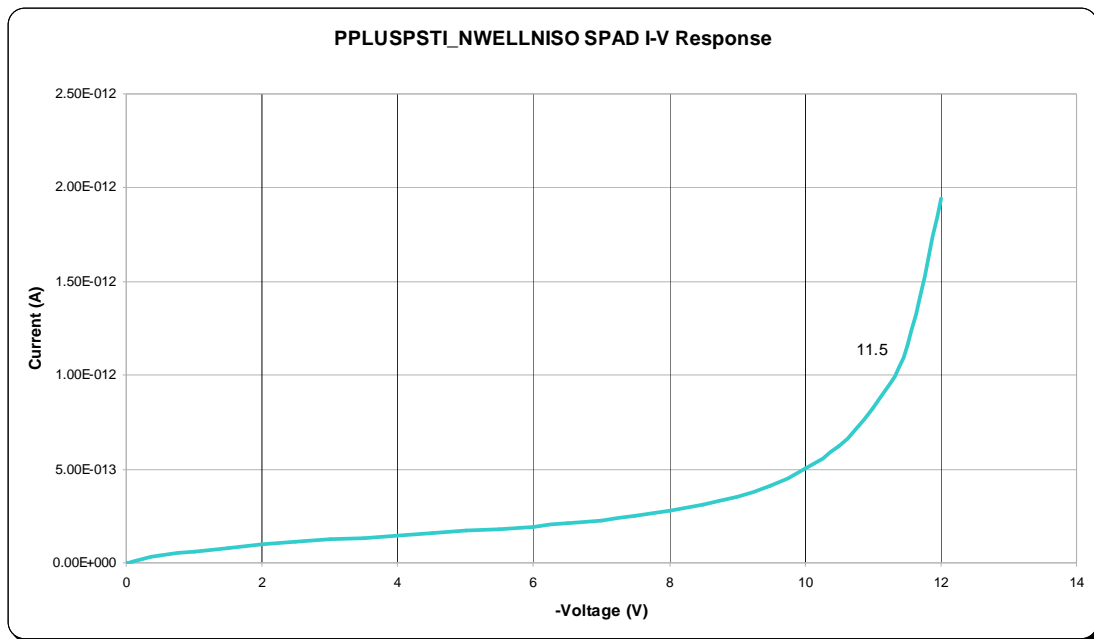


Figure 61: PPLUSPSTI_NWELLNISO SPAD I-V Response

The 2D simulation confirms the efficacy of the guard ring and the calculated E-field strength as shown in Figure 62 and Figure 63 respectively.

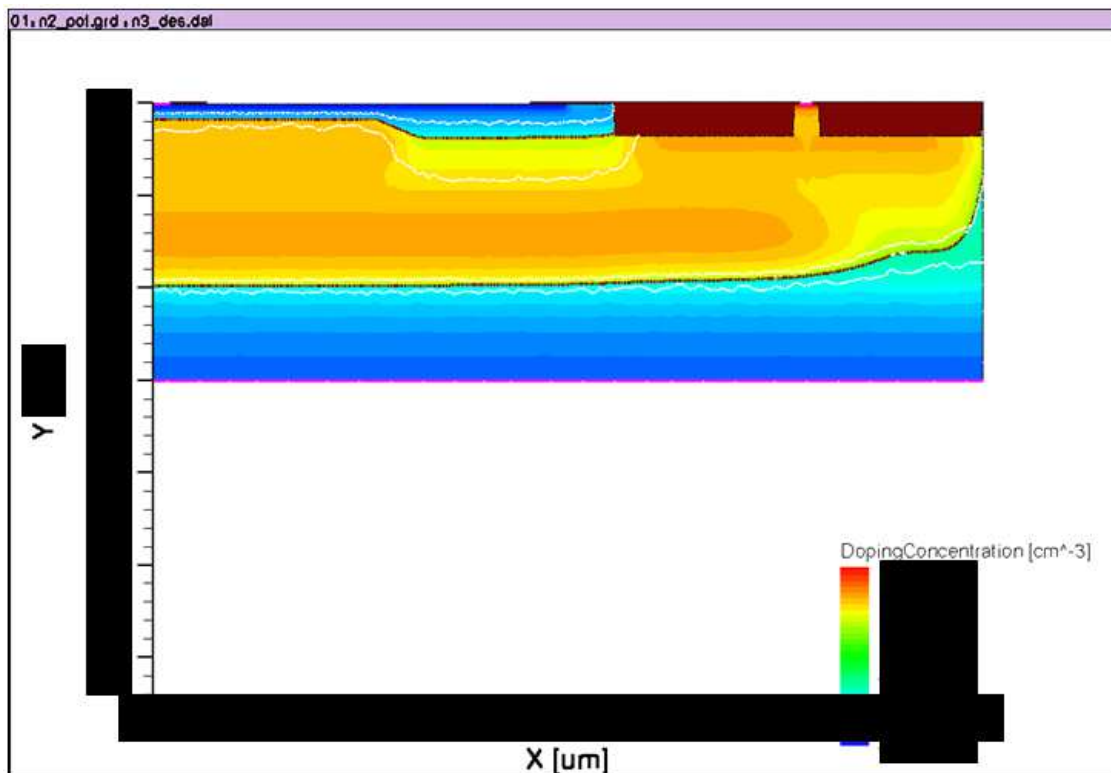
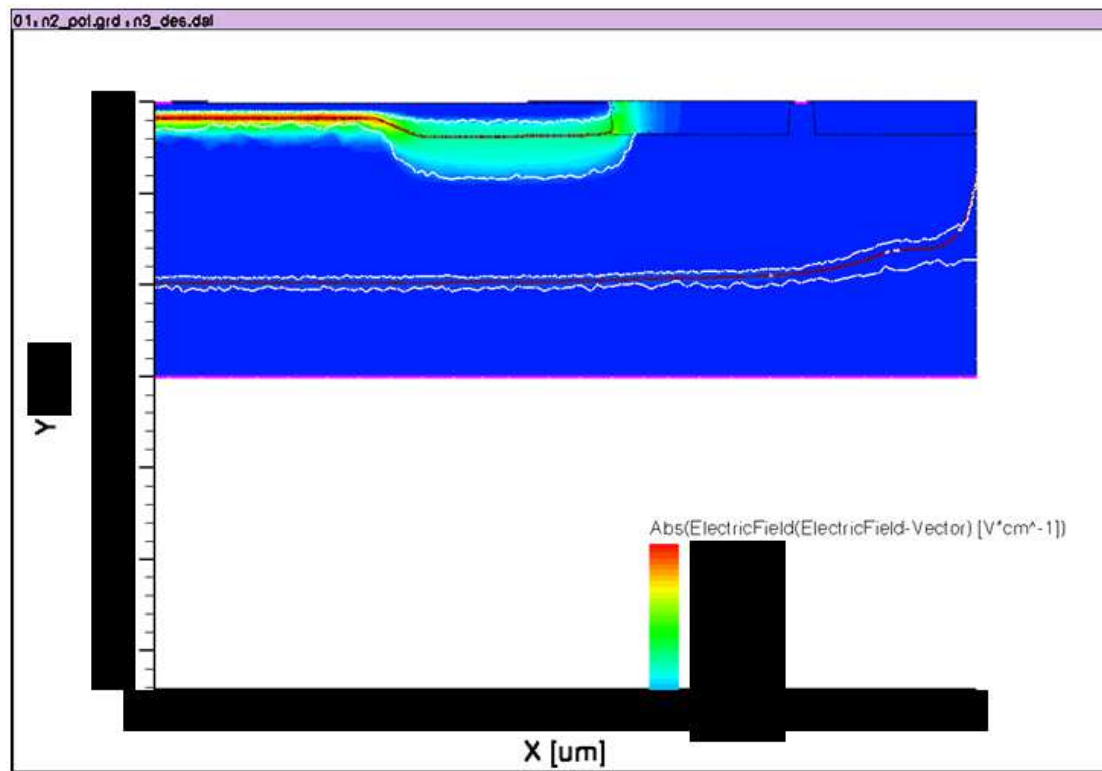


Figure 62: PPLUSPSTI_NWELLNISO SPAD 2D Doping Simulation



The device cross section can be implemented as the round, central anode Cadence layout which is shown below in Figure 64.

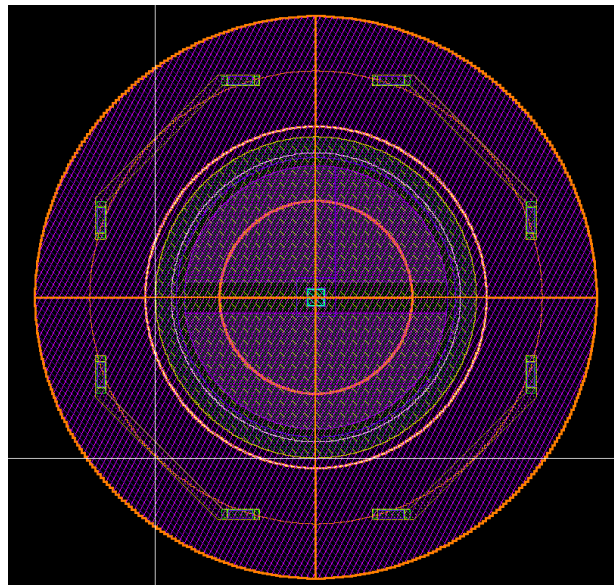


Figure 64: PPLUSSTI_NWELLNISO_STI SPAD Layout

This detector again has an 8 μ m active region diameter and an overall outer diameter of 17 μ m. The version shown has the guard ring option with PSTI coincident with the edge of STI. The NWELL enhancement can be seen in the central active region.

3.6.2.1 Results Summary

Expected results for the device can be compared with the high DCR prior art, shown below in Table 6.



	E-Field (kV/cm)	Pk Depth (nm)	V _{bd} (V)	C _j (fF)	Gain Aval' ke ⁻
PPLUS_NWELL	775		8.72	23	173
PPLUSPSTI_NWELLNISO	735		11.5	21	158
Comment	decrease	deeper	increase	minimal change	decrease

Table 6: PPLUSPSTI_NWELLNISO SPAD Expected Results

It can be predicted that the more modest reduction in E-field would have the appropriate impact on DCR. The deeper depletion region should favour a more red PDP response, but with less of a timing resolution penalty. This is achieved at the expense of slightly increased breakdown voltage.

3.6.3 PPLUSPWELL_NISO_EPIPOLY SPAD

The designs presented thus far both use imaging option implants. In contrast the following detector design uses solely standard implants as shown below in Figure 65. The deep anode is constructed from PPLUS and PWELL (with optional additional P-implant called PISO). The graded NISO-EPI guard ring construction of the detector presented in section 3.6.1 is again used. Both anode and guard ring can be formed by only two drawn implants. Similarly NPLUS-NWELL forms the connection down to the deep NISO cathode.

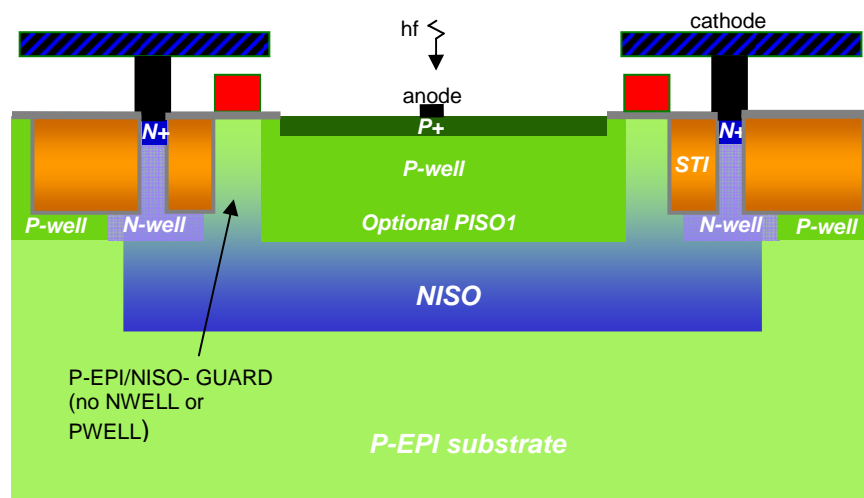


Figure 65: PPLUSPWELL_NISO_EPIPOLY SPAD Cross Section

It can be seen that in this construction, the PSTI implant of Figure 51 has been replaced with the higher doped and deeper PWELL. Therefore it can be expected that the E-field peak will be comparatively higher and deeper, with an overall reduced breakdown voltage. This is confirmed by TCAD simulation, the results of which are shown in Figure 66 (doping), and Figure 67 (E-field).

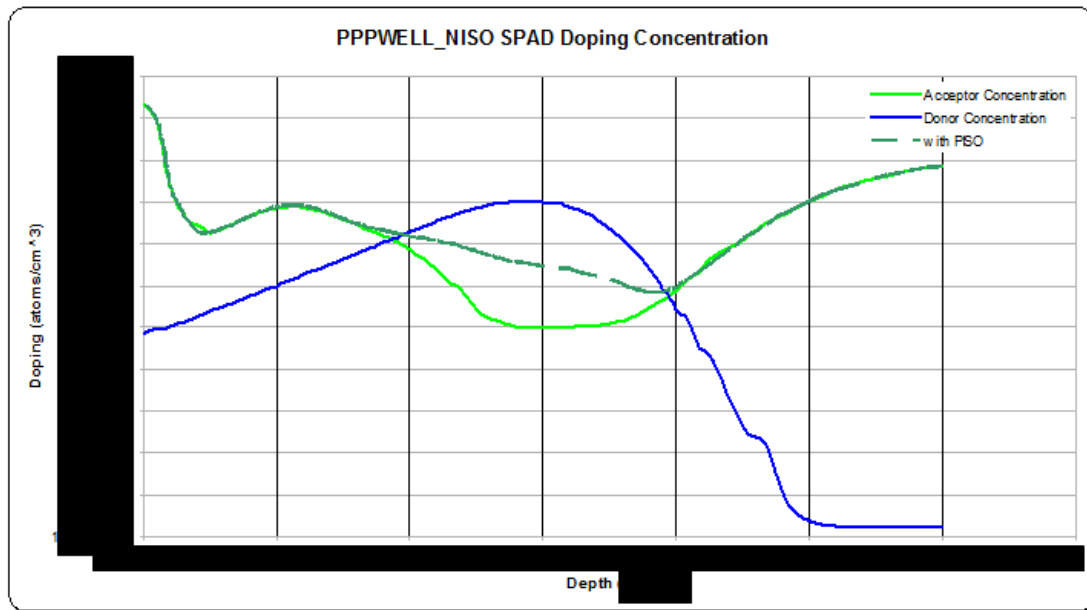


Figure 66: PPLUSWELL_NISO SPAD Doping Concentration

It can be seen that the addition of PISO to the anode has the effect of slightly extending the junction and E-field peak deeper into the substrate.

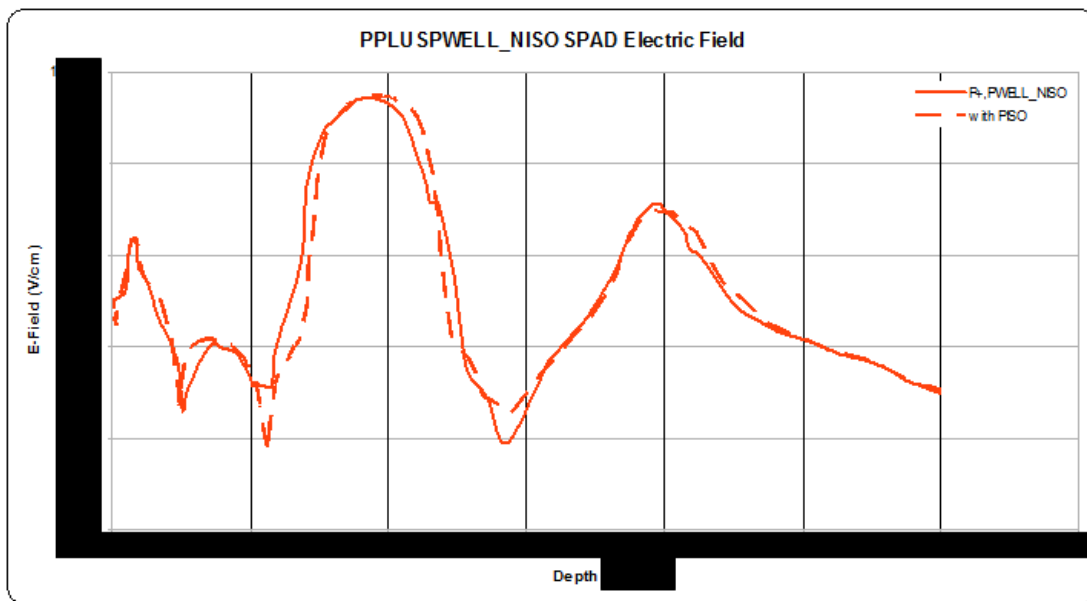


Figure 67: PPLUSWELL_NISO SPAD Electric Field

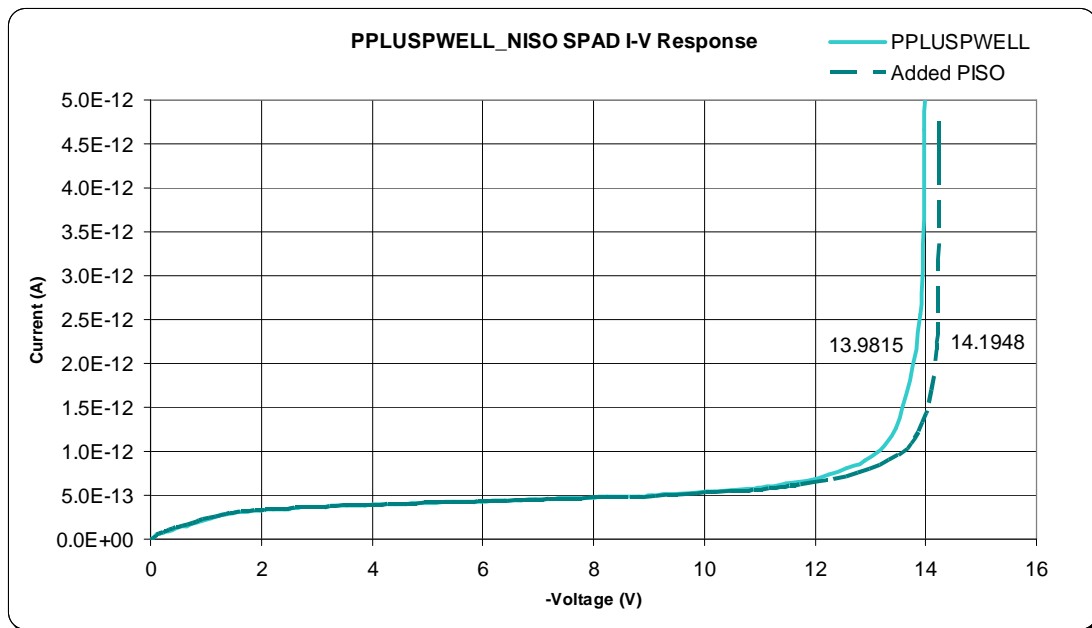


Figure 68: PPLUSPWELL_NISO SPAD I-V Response

Simulation demonstrates that the overall breakdown voltage is reduced compared to the PSTI anode SPAD as shown in Figure 68. The addition of PISO has a small impact.

2D TCAD simulation again confirms the efficacy of the guard ring, shown below in Figure 69 (doping) and Figure 70 (E-field).

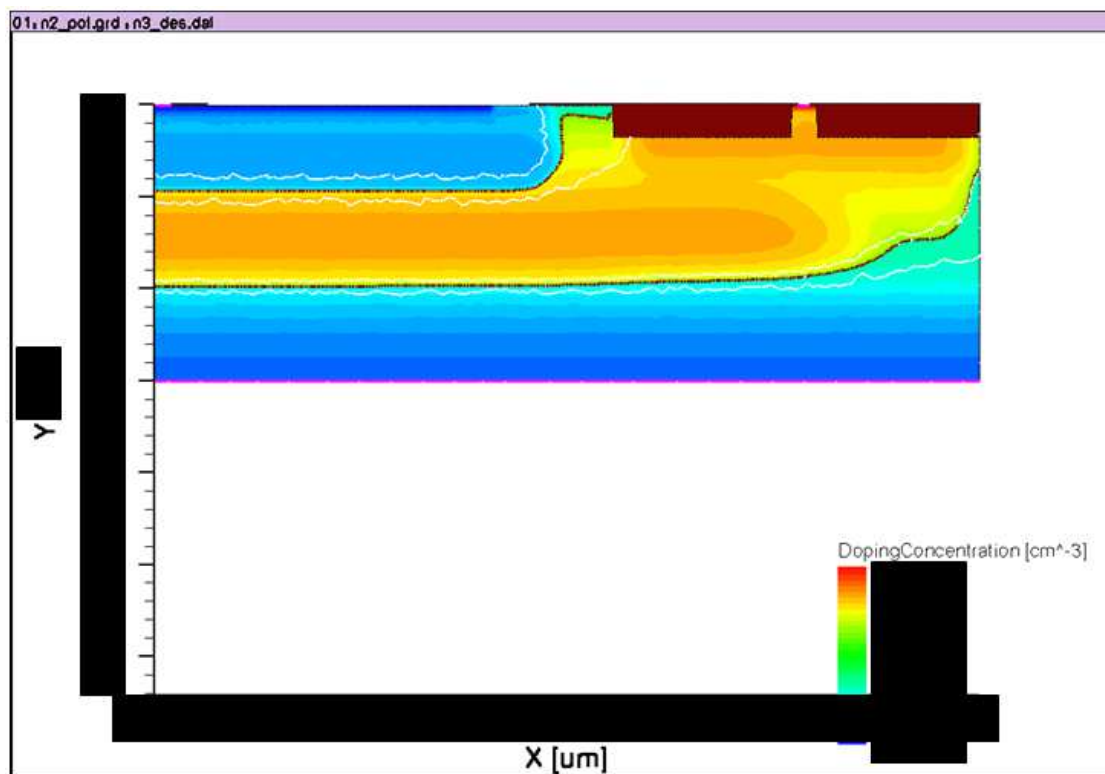


Figure 69: PPLUSPWELL_NISO SPAD 2D Doping Simulation

The simulated and calculated values for peak junction E-field compare very favourably as being in the region of $\sim 540\text{KV/cm}$, a figure which lies mid-way between the high DCR prior art and the PSTI anode SPAD of section 3.6.1.

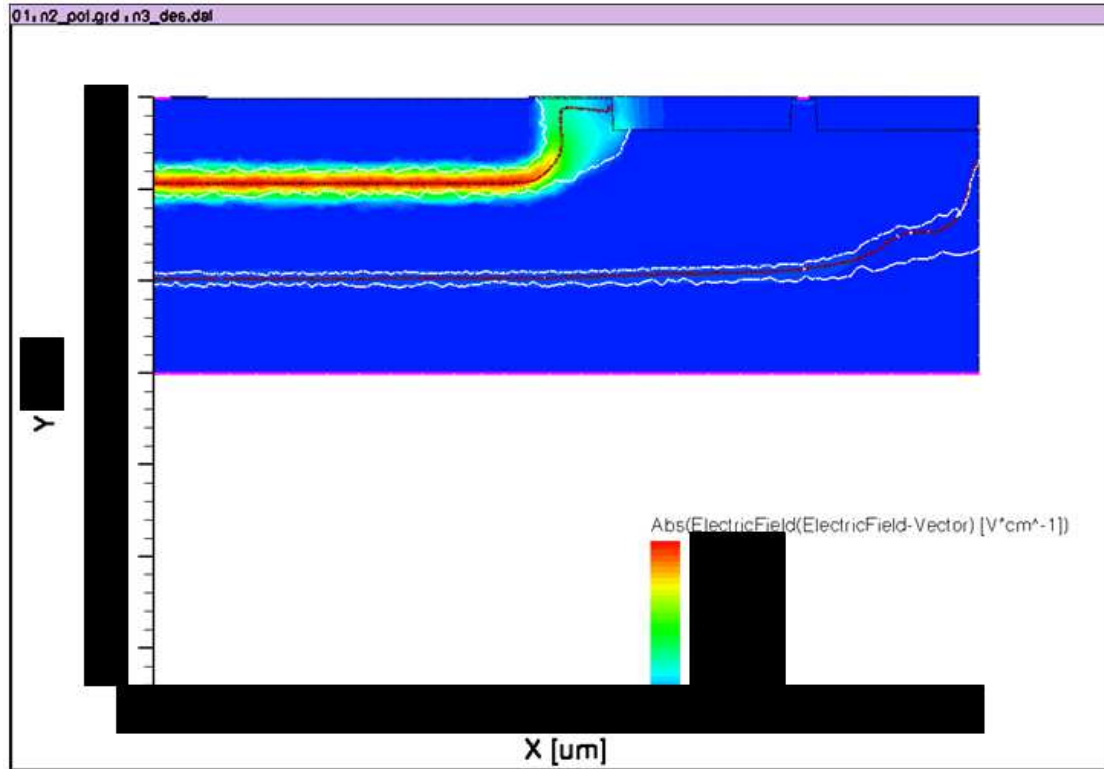


Figure 70: PPLUSPWELL_NISO SPAD 2D E-Field Simulation

The device cross section can be implemented as the round, $8\mu\text{m}$ diameter active region, central anode contact, Cadence layout which is shown below in Figure 71.

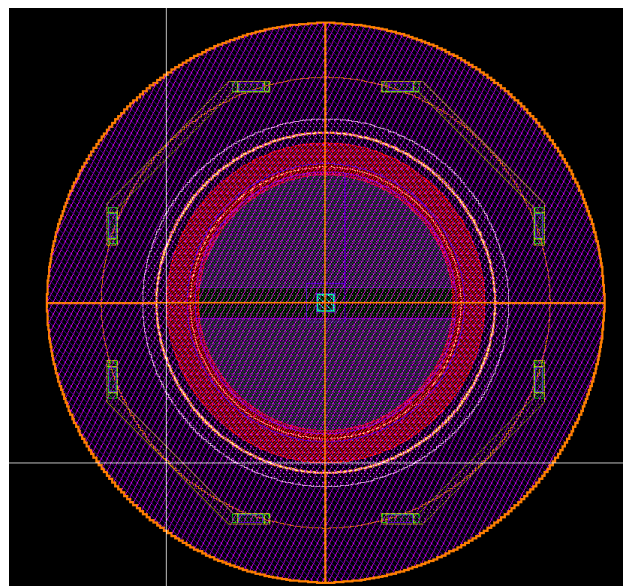


Figure 71: PPLUSPWELL_NISO_EPIPOLY SPAD Layout

This device construction has been implemented in round, square and ‘Fermat’ shapes, in a range of active region sizes. Conclusions from this aspect of the research are discussed in section 3.7.

3.6.3.1 Results Summary

Expected results for the device can be compared with the high DCR prior art, shown below in Table 7.

	E-Field (kV/cm)	Pk Depth (nm)	V _{bd} (V)	C _j (fF)	Gain (Aval ke ⁻)
PPLUS_NWELL	775		8.72	23	173
PPLUSPWELL_NISO	536		14.4	9	72
+ added PISO	551		14.3	10	75
Comment	decreased	deeper	increased	decreased	decreased

Table 7: PPLUSPWELL_NISO SPAD Expected Results

It can be predicted that the reduction in E-field would again have the desired effect of reducing DCR. The deeper depletion region should also favour a higher red PDP response. This is achieved at the expense of increased breakdown voltage which is still within the target specification of <20V. Junction capacitance is more than halved which should lower afterpulsing further.

An attractive feature of this particular detector design is that the PWELL-NISO diode structure is a legitimate design kit device, recognisable and extractable by the CAD tool suite. This is due to the conventional use of NISO-NWELL to fully isolate zones of PWELL for noise limitation reasons. This means that for this particular SPAD, there is access to the full design verification software suite (Calibre DRC-LVS) with the SPAD in situ alongside any number of associated circuit elements such as those required for quenching and time-digital conversion. Furthermore, since LVS is possible, full parasitic extraction is enabled. The impact of this improved modelling capability on the optimisation of quenching circuitry is presented in section 3.8.

3.7 New Detector Physical Optimisation

In this section a series of experiments regarding optimisation of the physical properties of the SPAD are discussed. The use of optical stack features, different shapes, active region diameters and anode contact positions is investigated. These features were first introduced in section 3.4. A discussion of the characterisation results for these experiments is provided in section 4.

3.7.1 Shape

Conventionally, detectors have been implemented as round shapes to prevent localised high E-field regions at sharp corners, known to be a cause of high DCR. Polygons drawn using the conics tools within CAD software suites such as Cadence Virtuoso introduce problematic off-grid design rule violations. To avoid this problem a PERL script has been created which quantises the vertices of a $\frac{1}{4}$ circular shape to force it on to the design grid. An example of the drawn result is shown in Figure 72.

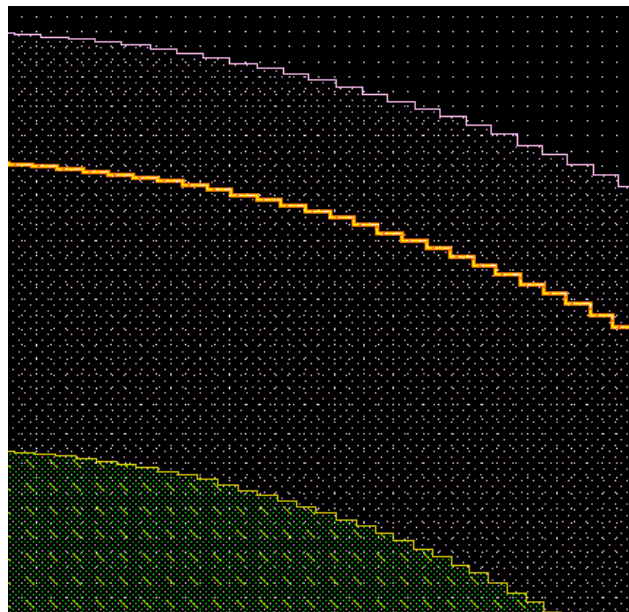


Figure 72: 'Make Quadrant' Script Result

This 'Make Quadrant' script can be applied to all polygons which go to make up the construction of the detector. Examples of the final round detector assembly are shown throughout section 3.6.

It should be noted that the round SPAD shape does not give the best achievable overall fill factor, which would be the case for a square shaped detector. Therefore, to

improve fill factor without creating any sharp vertices which increase DCR, the ‘Make Quadrant’ script was further developed to create what is called a ‘Fermat’ shape, named after the 17th century mathematician and geometrist Pierre de Fermat. By passing this new script an additional ‘power’ argument the shape shown below in Figure 73 can be drawn, allowing a significant improvement in SPAD active region fill factor whilst avoiding sharp vertices.

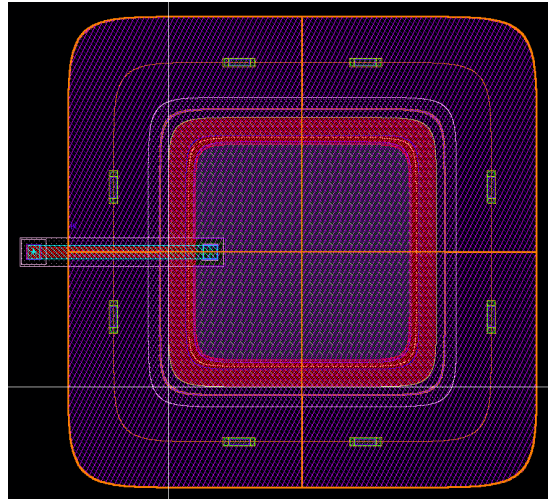


Figure 73: Fermat Shaped PPLUSWELL_NISO SPAD

The final shape used for the shaping experiments was a square version. This was implemented to enable performance comparison with the smooth shaped profiles of the round and Fermat devices.

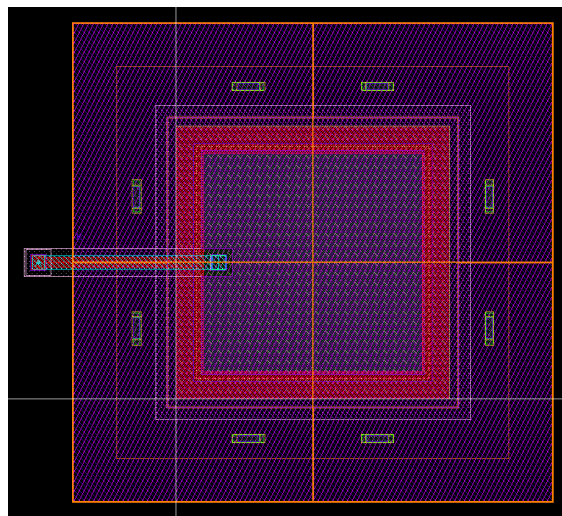


Figure 74: Square Shaped PPLUSWELL_NISO SPAD

The table below shows a comparison of the different shapes fill factor for an $8\mu\text{m}$ active region width dimension, based on a $17\mu\text{m}$ /side footprint.

	Round	Fermat	Square
Active Area, μm^2	50	63	64
Fill Factor %	16	20.5	21

Table 8: SPAD Shaping Fill Factor Comparison

The generally low fill factor, even for the square SPAD, strengthens the case for the application of microlenses for further fill-factor recovery, as well as the reduction of overall footprint caused by the adherence to conservative design rules in the guard ring periphery zone.

3.7.2 Scaling

A range of different diameters of the PPLUSWELL-NISO SPAD, introduced in section 3.6.3, have been implemented. Diameters of 2, 4, 8, 16 and $32\mu\text{m}$ detectors were designed in round, Fermat and square shapes. This enabled an analysis of the impact of improved fill factor on SPAD performance. The detectors were instantiated alongside integrated PMOS passive quench and output buffer elements. The family of SPADs is shown below in Figure 75.

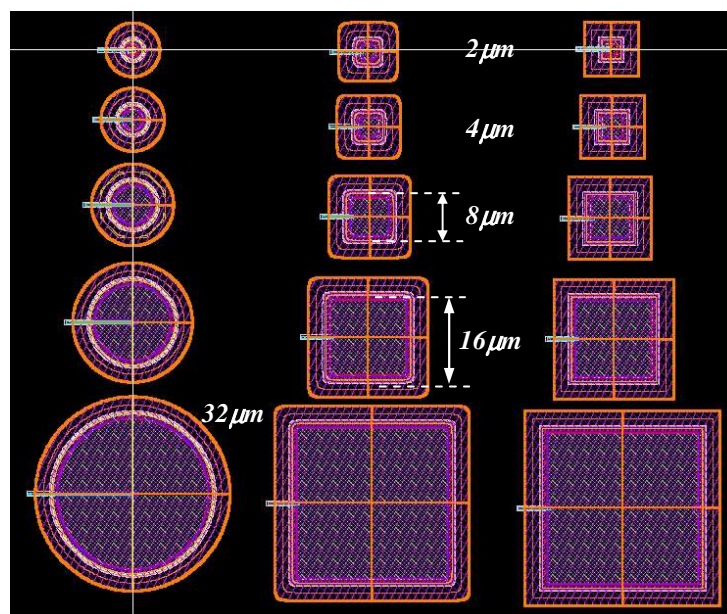


Figure 75: SPAD Family Plot

Table 9 below illustrates for each diameter of detector the active area and fill factor (FF) expressed as a percentage of the overall total footprint occupied by the SPAD.

Active Zone ϕ μm (area μm^2)	Total sqr Footprint ϕ & \square area μm^2	Round Active Area & FF %	Fermat Active Area & FF %	Square Active Area & FF %
2 ϕ	11.5 ϕ , 132.2	3.14, 2.4%	3.94, 3%	4, 3.0%
4 ϕ	13.5 ϕ , 182.2	12.56, 6.9%	15.77, 8.7%	16, 8.8%
8 ϕ	17.5 ϕ , 306.2	50.26, 16.4%	63, 20.6%	64, 20.9%
16 ϕ	25.5 ϕ , 650.2	201, 30.9%	252.3, 38.8%	256, 39.4%
32 ϕ	41.5 ϕ , 1722.2	804, 46.7%	1009, 58.6%	1024, 59.4%

Table 9: Fill Factor Comparison Data

The data reveals that the Fermat shape improves FF by a factor of 1.25, with the square shape improving FF by 1.275.

Using the round, 8 μm device with 47Hz DCR as a datum at room temperature with moderate excess bias, expected DCR can be tabulated.

Table 10 illustrates expected DCR if scaled with area.

SPAD Shape	SPAD Diameter (μm)				
	2	4	8	16	32
Round	3	12	47	188	752
Fermat	4	15	59	236	944
Square	5	16	60	239	957

Table 10: Scaled Ideal Dark Count Rate

However, Rochas [15], Pancheri et al [47] and Zanchi et al in [18] all report that DCR does not in practice follow scaling with area. This is also the case with the experiment reported here. Measured results for the above detectors are compared with the data from Table 10, in section 4.4.2.

3.7.3 Active Region Enhancements

The Fermat and square shaped detectors have been optimised in three ways so as to maximise the number of photons able to enter the active region and minimise the introduction of defects.

Firstly, the approach of using a peripheral anode contact position has been adopted, as discussed in section 3.4.3 so as to maximise fill factor. A zoomed in view of the anode contact of a Fermat detector is shown in Figure 76.

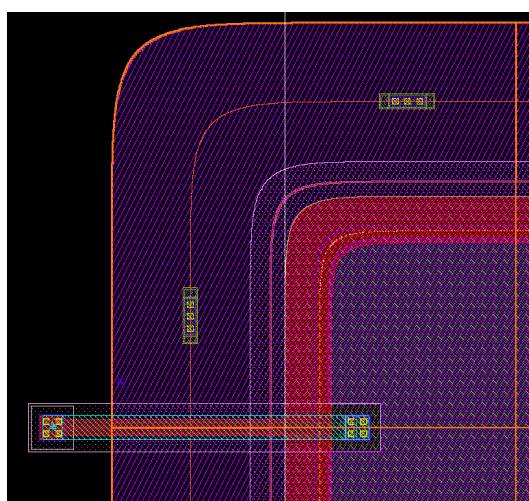


Figure 76: Fermat SPAD Anode Contact Position

Secondly, the saliciding process normally applied to contacts has been intentionally prevented from occurring in the anode region by using the ‘silicide protection’ mask. Saliciding is intended to reduce the resistance of contacts, and is normally a desirable process feature. However, it is a tungsten based phase of the process which is known to introduce metallisation induced lattice defects, a cause of high dark current in conventional photodiodes and so potentially DCR in SPADs. The increase of the resistance of unsalicyded contacts is small when compared to a typical passive quench resistance ($\sim 300\text{k}\Omega$), and can therefore be discounted.

The final improvement implemented is the adoption of nitride barrier cuts over the active region, as discussed in section 3.4.1, Figure 43. These cuts are intended to reduce the reflections caused at material boundaries which exhibit large refractive index steps.

3.8 Quenching

As discussed in section 2.3.5, ‘quenching’ is the term given to the process of detecting and stopping a SPAD avalanche event, followed by a resetting of bias conditions. During this work, two key areas of study have been performed related to quenching: accurate modelling of passive quenching and novel active quench techniques.

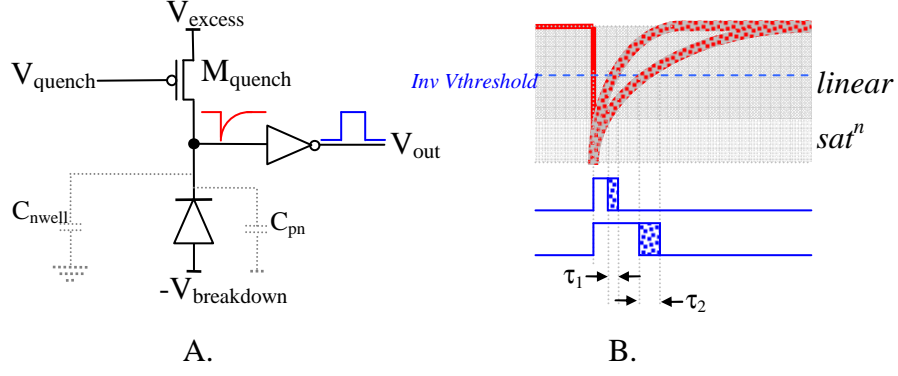
Firstly, due to both the calculation of junction capacitance, plus the LVS extraction capability of the PPLUS-PWELL_NISO SPAD of section 3.6.3, both junction and parasitic capacitive elements of a conventional passive quench configured SPAD circuit can be predicted. This means that the passive quench MOS element can be sized appropriately for linear mode operation, and accurate simulations performed. This optimisation process is presented in section 3.8.1.

Secondly, a novel active quench circuit has been designed and fabricated. This is based on a CMOS thyristor configuration, allows very high dynamic range, and is presented in section 3.8.2.

3.8.1 Optimised Passive Quench

Incorrectly sized passive quench circuitry can result in undesirable circuit operation, such as lengthy and inconsistent dead times. This introduces noise, imposes severe dynamic range limitations, particularly impacting photon counting applications and those employing gated counters such as ranging, 3D cameras and FLIM.

Despite the limitations of this means of quenching, introduced in section 2.3.5.1, passive elements are efficient, compact, and inherently current limiting. These features are highly desirable when considering large SPAD arrays for reasons of simplicity, fill factor and power consumption. Figure 77 illustrates the optimisation concept and procedure for a PMOS passive quench setup.



$$\text{Recharge: } V_{spad} = V_{eb} \left(1 - e^{-t/R_{quench} \cdot C_{tot}} \right) \quad (i)$$

$$R_{quench} = \frac{V_{eb}/2}{I_{d_{Mquench}}} \quad (ii)$$

$$\text{Avalanche: } I_{d_{Mquench}} = K'_p \frac{W}{L} \left[(V_{gs} - V_t) - \frac{V_{ds}}{2} \right] \cdot V_{ds} \quad (iii)$$

Figure 77: Passive Quench Optimisation

Figure 77(A) shows the basic passive quench circuit, parasitic elements and waveforms. Figure 77(B) shows how noise can cause increased dead time variation in a non-optimised system. In an optimised configuration the crossing of the inverter threshold is performed at a steep gradient. Noise on the moving node results in a dead time variation τ_1 at the inverter output. If the passive quench element W/L ratio is too small, the extended RC recharge time results in a shallower angle of attack as the moving node crosses the inverter threshold, impacting noise immunity and resulting in a higher dead time variation τ_2 . In the most basic configuration the gate of the PMOS passive quench element is simply grounded. Under the aforementioned non-optimal circumstances a small negative voltage can be applied to the gate of the PMOS transistor but this is considered an undesirable complication. Similarly, in the case of an undersized PMOS element, a positive voltage must be applied to its gate to ensure an output pulse is generated.

In a low voltage process, with comparatively low overall excess bias conditions, for the bulk of the quenching and reset cycle the PMOS element is mainly operating in linear mode, with a brief entry into saturation.

Thus, equations (i), (ii), and (iii) can be applied to optimise the passive quench element. For a parasitic capacitance $C_{\text{tot}} = 50\text{fF}$, $V_{\text{eb}}=1.2\text{V}$ and inverter $V_{\text{threshold}}=0.6\text{V}$, for a dead time of 20ns and a PMOS passive quench element in IMG175 technology (where $K'_p = 30 \mu\text{A/V}$):

$$R \approx 400\text{k}\Omega$$

$$\therefore Id = 0.6\mu\text{A}$$

$$\Rightarrow \frac{W}{L} = \frac{1}{8}$$

Therefore for a $0.35\mu\text{m}$ width, grounded gate PMOS transistor, L should be $\sim 2.7\mu\text{m}$ to meet the above specification.

The optimised passive quench configuration is summarised below in Table 11.

	Magnitude	Unit
Transistor Count	3	T
Area ¹	329	μm^2
Supply Voltage	1.2 - 3.6	V
Power Consumption ²	137	μW
Quench duration	30 - 400	ns

Table 11: Optimised Passive Quench Summary

Notes:

¹ - this figure is calculated from a passively quenched pixel layout capable of being instantiated in an array format without design rule violation, so includes overhead to allow appropriate spacing between detectors. Figure does not include the SPAD itself.

² - this is a measured figure for 10MHz count rate, equivalent to ambient office lighting conditions at room temperature and typical supply voltage levels: $1.23\mu\text{A}$ from the combined SPAD bias supply of 14V and $100\mu\text{A}$ from the output buffer supply of 1.2V. With zero photon flux the consumption for a single passively quenched device is negligible.

3.8.2 Thyristor Active Quench

Active quenching, introduced in section 2.3.5.2, has several performance advantages even over the optimised passive configuration of the last section. Dead time can be significantly reduced, with a more predictable hold-off time (see Figure 22), enhancing the consistency and overall magnitude of dynamic range and photon detection efficiency. The main drawbacks with existing active quench implementations have been high area utilisation and complexity. The novel thyristor approach presented below in Figure 78 attempts to address these limitations using only MOS elements.

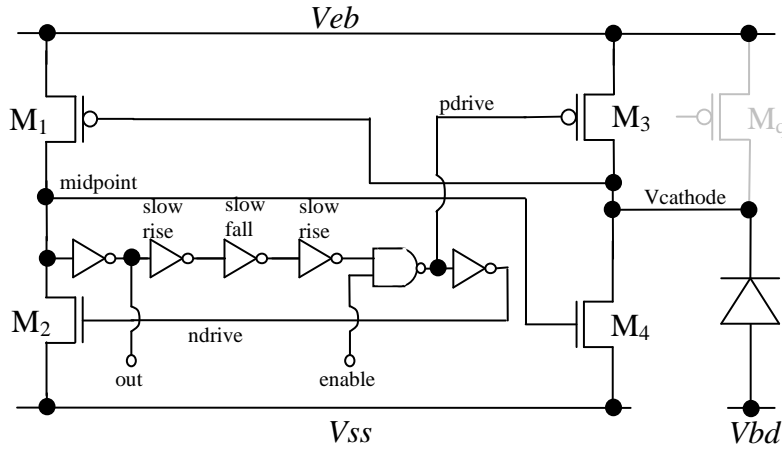


Figure 78: Thyristor Active Quench Circuit Diagram

Under quiescent conditions *midpoint* and *Vcathode* nodes are high impedance and power consumption is minimised. When the detector fires the positive feed-forward loop of M_1 and M_4 promotes the discharge and clamps the SPAD cathode to 0V. A delay time later, set by an unbalanced threshold inverter delay chain, the latched condition is reset via the dominant drive strength PMOS M_3 and loop disabled through NMOS M_2 . The passive quench element M_q , shown in grey, is an optional backup. Dominant sub-threshold leakage of PMOS M_3 over NMOS M_4 is required to maintain the SPAD at a high potential between avalanche events ($\sim 8:1$ width). Similarly, the M_2/M_1 sizing relationship ($2:1$ width) contributes to correct quiescent potential on node *midpoint* so avoiding erroneous leakage induced triggering. The charge injection on the *Vcathode* and *midpoint* nodes via control signals *pdrive*/*ndrive* respectively is also in the correct sense to provide gate overdrive on the M_4 pull-down

and M_1 pull-up devices to minimize sub-threshold leakage in those devices and thus promoting the correct dominant current path.

This architecture creates a well defined hold-time time duration and fast reset. Utilising such a delay chain provides opportunity for integrating enable/disable and hold-off pulse duration adjustment functions via the introduction of appropriate logic and current starved inverters respectively. The basic slow-rise, fast-fall cell used in the delay generator is shown in Figure 79.

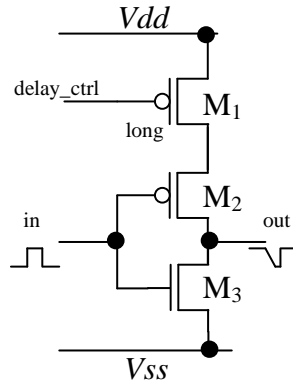


Figure 79: Slow Rise-Time Current Starved Inverter Schematic

The control signal *delay_ctrl* is used to vary the hold-time duration and hence can be used to set the dynamic range of photon counts. It is also known that quench duration and SPAD afterpulsing are related and so this may be traded off against count rate dynamic range. Simulation results showing two different dead times is shown below in Figure 80.

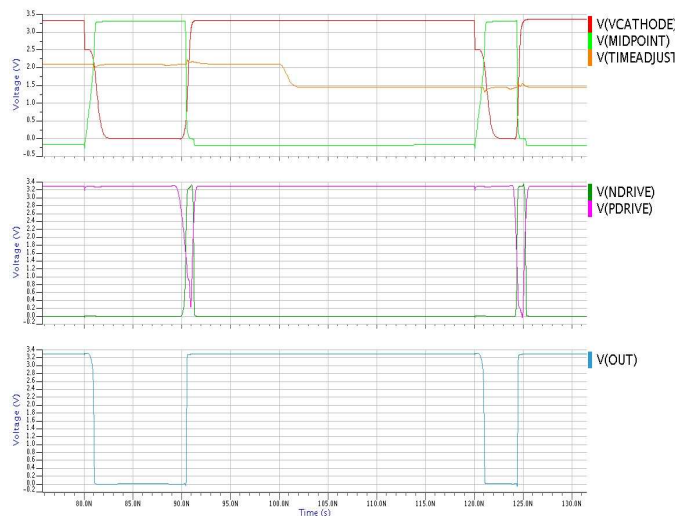


Figure 80: ThyristorActive Quench Simulation Results

The layout of this circuit is $\sim 17\mu\text{m} \times 8\mu\text{m}$ and is shown below in Figure 81. The layout is kept compact by the implementation of full custom transistor layout, without the use of standard cell library elements. Note that the output buffer of a passive quench setup is an integral part of this circuit so is not an additional requirement.

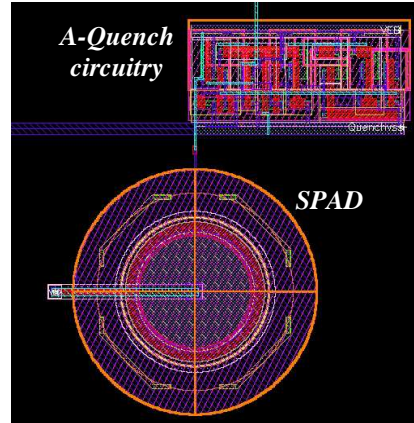


Figure 81: Thyristor Active Quench Layout

In summary, this active quench configuration enables sub ten nanosecond level dead times, enabling a very high maximum count rate of 120MHz. This is achieved in an area less than half that of the footprint of a $17\mu\text{m} \varnothing$ SPAD. Conventionally, due to high area usage and large avalanche current magnitudes, active quenching was best suited to larger SPADs. The area of the new approach permits the application of active quenching to smaller diameter SPADs and enables high count rate, fully actively quenched arrays. The performance is summarised below in Table 12.

	Magnitude	Unit
Transistor Count	20	T
Area	130	μm^2
Supply Voltage	1.2 - 3.6	V
Max Count Rate	120	MHz (measured)
Power Consumption	60	μW average @ 25MHz
Quench duration	2 – 30	ns
Arming (enable) Time	300	ps
Feed fwd trigger time	100 - 500	ps

Table 12: Thyristor Active Quench Summary

Characterisation results of the fabricated circuit, implemented together with a PPLUSPWELL_NISO SPAD are discussed in section 4.5.

Since this part of the research was performed, Gronholm et al have reported a ring oscillator based active quench circuit [90], referencing the design presented here.

3.9 Host Integrated Circuit

The detector designs presented in section 3.6 were implemented on the ‘IMNSTEST’ device, spread over two test ‘columns’. A micrograph of this device is shown below alongside an inset showing the layout of the contents within a section of a column.

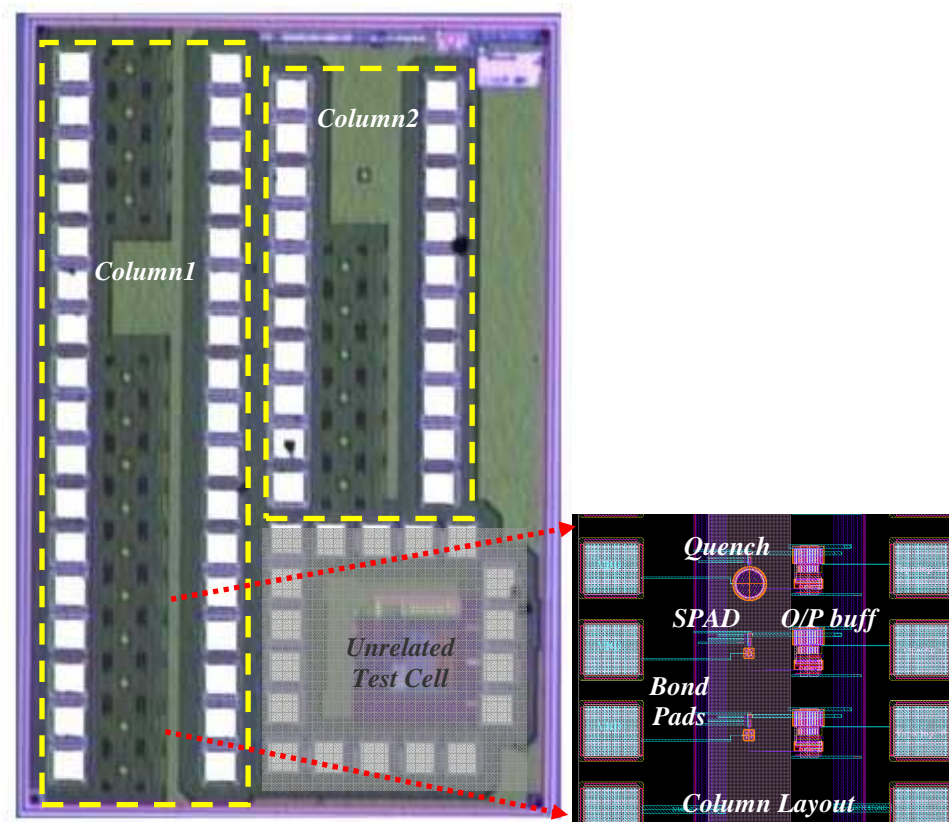


Figure 82: ‘IMNSTEST’ Device Micrograph

Normally instantiations of detectors separate from their quench and readout electronics are required in order to obtain control of the nodes concerned. In the case of the devices instantiated in the columns above, testability has been improved via the integration of a thick oxide, ESD proofed, NMOS transistor which enables both I-V test mode and conventional operating modes using the same SPAD device. This is shown as M_{test} in Figure 83.

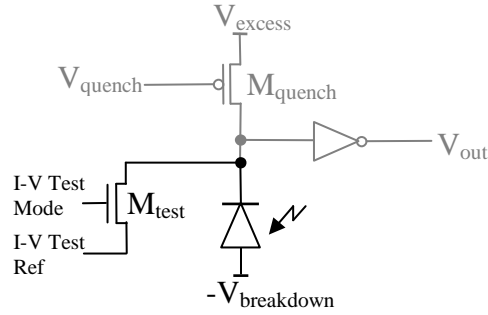


Figure 83: Reverse I-V Test Mode Capability

This circuit was again utilised for the SPAD scaling, shaping and optimisation experiments implemented on the device ‘SPADDEVELA’, along with optimised passive and thyristor active quench circuits.

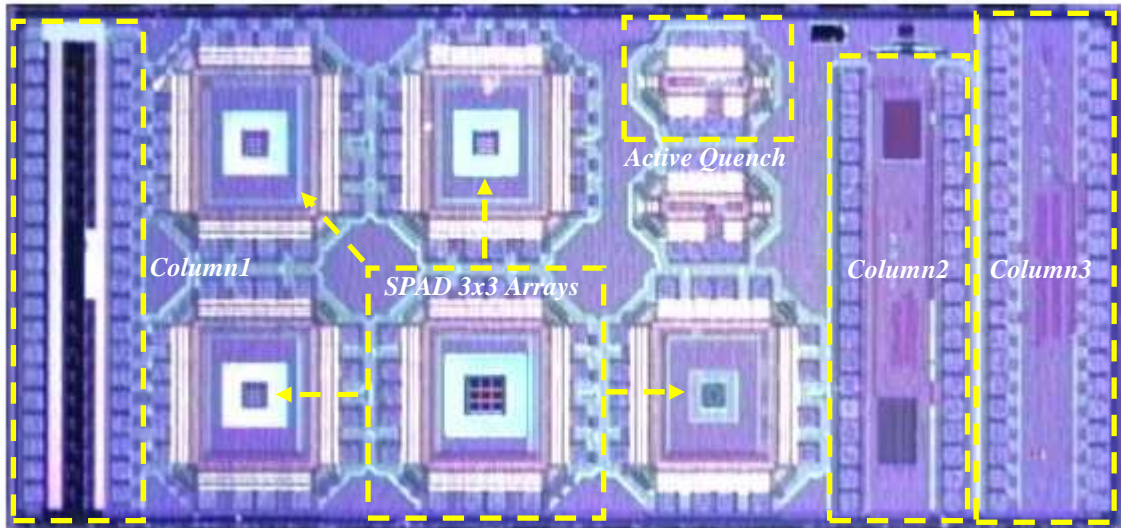


Figure 84: SPADDEVELA Device Micrograph

The SPADDEVELA device also includes a selection of 3x3 arrays of detectors of varying sizes and pitch, allowing the evaluation of optical and electrical crosstalk versus different fill factors.

3.10 Conclusions

In this chapter a target design specification for SPADs in nanometer scale CMOS was determined. Processing and physical implementation considerations were discussed. A series of new SPAD designs were presented, with calculated and simulated results compared to a key example of prior art implemented in a related technology. Performance expectations for scaling and shaping experiments were discussed, as

well as guidelines for optimising passive quench components. A novel active quench circuit based on a CMOS thyristor which ensures maximum possible photon detection performance was also introduced.

The design of new nanometer scale CMOS SPADs based on the understanding of detailed process parameters, alongside the ability to perform calculation and simulation of the main performance metrics is an important development. The use of a lower doped, deeper p implant such as PWELL or PSTI in conjunction with the standard PPLUS implant to form a deep breakdown junction of the SPAD with NWELL or with NISO alone is a development over existing SPADs which conventionally use PPLUS implant to NWELL as the active region P-N junction.

Furthermore when NISO is used on its own without NWELL, a novel structure is created with a lower doped, graded donor-region allowing a wider depletion region, so reducing peak E-field and band-band tunnelling induced DCR. This feature is compatible with most standard CMOS technologies.

The use of NISO without an accompanying well implant is extended into the guard ring zone where a novel enhancement mode structure is created using only two drawn implants. The resultant graded virtual guard ring promotes higher field strength at depth dissuading surface or peripheral breakdown. The guard ring is compatible with STI and allows dense, compact and highly scalable SPAD layout. STI induced defectivity is avoided either via implant passivation or movement of STI away from the active region, using drawn polysilicon techniques.

Furthermore, analysis of high DCR prior implementation allowed the comparison of metrics, permitting conclusions to be drawn during the silicon characterisation phase. Expected results for the four different SPAD designs are shown together in Table 13.

SPAD Type	E-Field (kV/cm)	Pk nm Depth	V_{bd} (V)	C_j (fF)	Gain (Aval ke)
PPLUS_NWELL (prior art)	775	■	8.72	23	173
1.PPLUSPSTI_NISO	323	■	17.5	10	70
2.PPLUSPSTI_NWELLNISO	735	■	11.5	21	158
3.PPLUSPWELL_NISO	536	■	14.4	9	72
Design Intention wrt P.A.	decrease	deepen	increase	decrease	decrease

Table 13: Simulated SPAD Performance Comparison

Indications of SPAD performance may also be drawn by comparing the different SPAD electric field profiles side by side, as shown in Figure 85.

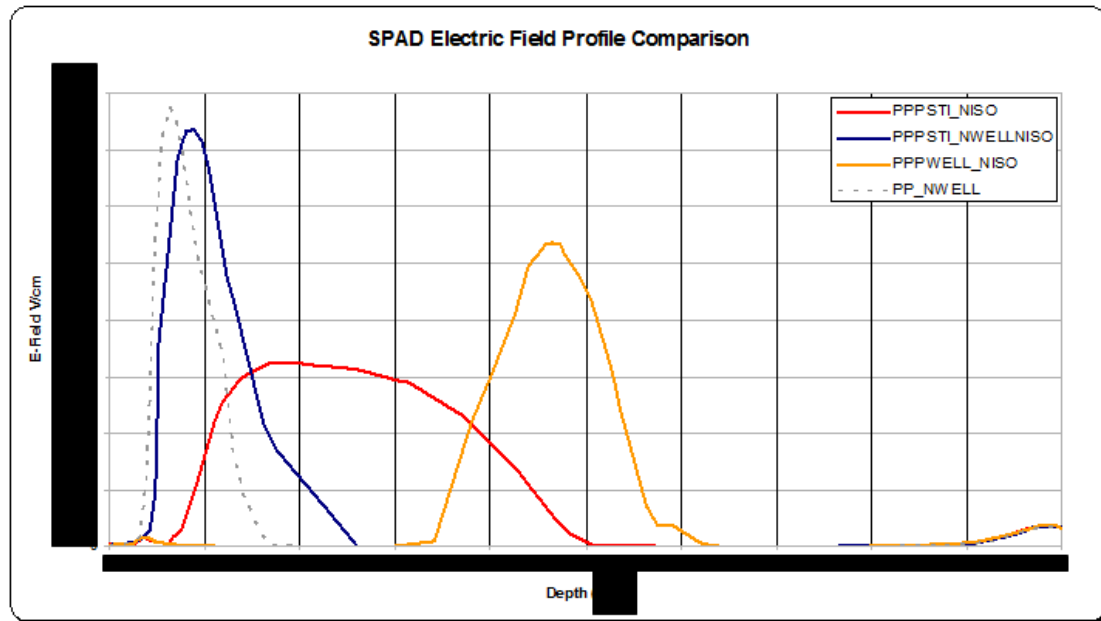


Figure 85: SPAD Simulated E-Field Comparison

The dotted trace indicates the high DCR prior art. The data indicates that the expected performance parameters agree with the design intention of reducing the maximum electric field from this reference point, as well as promoting a broader wavelength response into the red region. By minimising the junction capacitance the afterpulsing probability should also be correspondingly lower.

The following chapter describes the test methodology and results for the silicon evaluation phase of the proposed detector designs 1,2 and 3 listed in Table 13. The integration of the new detectors with both passive and active quenching circuits is evaluated and compared. The behaviour of one of the designs when it is scaled is presented, as well as the impact on performance of three different shapes of detector; round, square and ‘Fermat’.

The key performance metrics for the three detector designs are finally compared and ranked, enabling conclusions to be drawn.

4 Detector and Quench Characterisation

4.1 Introduction

In this chapter the characterisation procedure for determining the main SPAD metrics is described, namely reverse breakdown voltage, dark count rate, photon detection efficiency, dead time, afterpulsing probability and timing resolution. This is followed by a discussion of the test results obtained for the new detectors presented in section 3.6 and an assessment of the results versus expectations.

4.2 Characterisation Procedure

4.2.1 Reverse Breakdown Voltage

The reverse breakdown I-V profile of a SPAD is obtained by the sweeping of an applied voltage source and measurement of current drawn, using a semiconductor parameter analyser. Referring to Figure 83, the parameter analyser is connected between the nodes 'I-V Test Ref' and '-Vbreakdown' whilst the node 'I-V Test Mode' is overdriven high. The analyser is always configured with a safe compliance current level to avoid damage to the device under test.

4.2.2 Dark Count Rate

Dark count rate is a SPAD's main noise source, and is determined using a 40Gs/s digital oscilloscope fitted with high bandwidth active probes. The bias conditions and dead time are firstly configured and then the detector is placed under a blackout cloth in a dark room environment. The oscilloscope sample rate is fixed (~100Msamples/sec) so that no pulses are missed when the oscilloscope time-base is extended to enable the counting of the total number of pulses in 100ms. The count rate measured using the oscilloscope's in-built signal analyses features represents $1/10^{\text{th}}$ of the per-second DCR. This permits count rate measurements from around 1Hz to hundreds of MHz, using high speed oscilloscope statistical analysis tools.

Once obtained by the above measurement technique, the DCR may be extrapolated as a light level equivalent figure. This enables appreciation of the minimum light level which is detectable by the SPAD, and is a common way of expressing imager noise

level, enabling comparison with other pixel types. For example, a detector with a dark count equivalent to 20 photons per second (i.e. 20Hz), we can say:

$$\Rightarrow \#Photons / active_area = 20 = Incident_radiant_flux / photon_energy$$

Assuming mid-range wavelength of 550nm, the energy of a green photon is $3.62 \times 10^{-19} \text{J}$.

$$\therefore Incident_radiant_flux = 20 \times 3.62 \times 10^{-19} = 7.24 \times 10^{-18} \text{W}$$

The Irradiance, or radiant flux density, can then be calculated for the detector with a given active area of $5.06 \times 10^{-11} \text{m}^2$:

$$\Rightarrow Irradiance = \frac{1}{5.06 \times 10^{-11}} \times 7.24 \times 10^{-18} \text{W} = 0.143 \mu\text{W} / \text{m}^2$$

Expressing this figure in lux (i.e. Illuminance) so that it can be appreciated as a real world light level, we can say:

$$10^{-4} \text{W} / \text{m}^2 = 6.83 \times 10^{-6} \text{lux} (@ 555 \text{nm})$$

$$\therefore 0.143 \mu\text{W} / \text{m}^2 = 97.7 \mu\text{lux}$$

Table 14 (edited from Beynon, Lamb in [91]) shows that this noise floor is equivalent to ‘moonless overcast night sky’.

Illuminance	Abbreviation	Example
0.00005 lux	50 μ lx	Starlight.
0.0001 lux	100 μ lx	Moonless overcast night sky.
0.001 lux	1 mlx	Moonless clear night sky.
0.01 lux	10 mlx	Quarter Moon.
0.25 lux	250 mlx	Full Moon on a clear night.
1 lux	1 lx	Moonlight at high altitude at tropical latitudes.
50 lux	50 lx	Family living room.
400 lux	4 hlx	A brightly lit office.
400 lux	4 hlx	Sunrise or sunset on a clear day.
1000 lux	1 klx	Typical TV studio lighting.
32000 lux	32 klx	Sunlight on an average day (min.)
100000 lux	100 klx	Sunlight on an average day (max.)

Table 14: Light Illuminance

4.2.3 Photon Detection Efficiency

The photon detection efficiency is determined by evaluating the ratio of the number of output pulses versus active-region incident photons over the light bandwidth 350-1150nm (dictated by monochromator capability). The number of output pulses is measured in a similar fashion to the DCR measurement. The test is performed in a controlled light environment with a Sciencetech 9055 monochromator and Newport 1830C/818-UV reference photodiode. The latter devices are used to sweep the incident wavelength and permit calculation of the number of striking photons using the equation shown.

$$\#striking_photons = \frac{\left(irradiance / m^2 \right) \cdot active_area(m^2)}{photon_energy_{nm}}$$

The PDE is then calculated for each monochromator setting:

$$PDE_{nm} = \frac{output_count_rate}{\#striking_photons}$$

The PDE can then be plotted versus wavelength for each monochromator setting.

4.2.4 Dead Time

The dead time is dependent on the excess bias voltage magnitude and the quench component configuration. When evaluating a SPAD with an integrated readout buffer the output pulse width between 90% maximum points is used as an indication of actual dead time. In [15] Rochas defines dead time as between the 90% points of the non-threshold, analogue waveform as discussed in section 2.2.5. In the case of a PMOS passive quench element the minimum dead time is normally obtained with the gate grounded and a small magnitude of excess bias. In some cases the gate can safely be brought slightly negative to further reduce the channel resistance and dead time. Minimum dead times are typically reported in the prior art as in the range ~20-40ns. The dead time may be significantly extended by raising the gate voltage of the passive quench MOS toward the excess bias level. Maximum dead times of hundreds of nanoseconds are normally achievable.

4.2.5 Afterpulsing

Afterpulsing can be measured using a high sample rate oscilloscope such as the LeCroy WP735Zi Wavepro. The built in channel autocorrelation function permits the direct plotting of afterpulsing probability versus time of a potentially repeating pulse shape. This technique provides an output graph which can be difficult to interpret and compare with other similar traces. Alternatively the unit under test may be placed in the dark, and the oscilloscope configured with conditional triggering. Under these conditions the oscilloscope will only trigger when a second SPAD pulse is observed within a defined time window after the primary pulse. The average number of dark counts which happen prior to oscilloscope triggering enables the calculation of afterpulsing likelihood. The measurement is repeated many times and an average result is taken.

4.2.6 Timing Resolution

The timing resolution or ‘jitter’ of a SPAD avalanche event is analysed using a high sample rate oscilloscope, picosecond laser and laser driver module as shown in Figure 86.

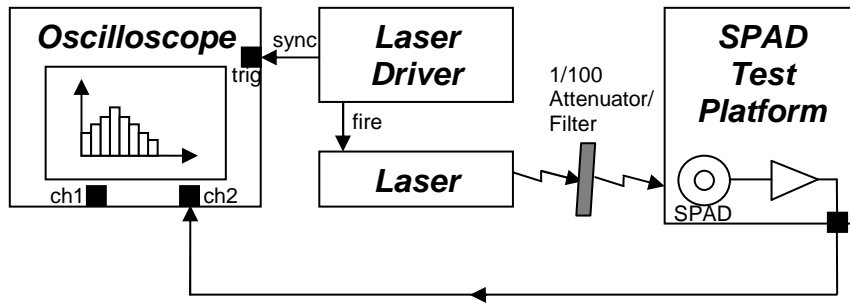


Figure 86: SPAD Jitter Test Setup

In the case of this research, the equipment used was a LeCroy WP735Zi Wavepro Oscilloscope, a Picoquant PDL 800-B laser driver and Picoquant LDH-P-C-470 pulsed mode laser. The laser driver is configured in a free running mode, with the oscilloscope triggered from the driver sync output. The time between the driver sync pulse and the SPAD output pulse is then continuously measured and a histogram is compiled. The timing resolution of the SPAD is the FWHM of the resulting histogram. The jitter of the laser driver ($<20\text{ps}$) is removed by means of oscilloscope trigger tracking. The quoted laser pulse FWHM of $\sim 68\text{ps}$ for the 470nm laser head must be subtracted in quadrature from the measured histogram result.

4.3 SPAD Characterisation Results

In this section the silicon characterisation results for each of the detectors presented in section 3.6 are provided (in order). For each of the round shaped detectors, I-V response, DCR, PDP and jitter is shown for a device which exhibits typical performance. Shaping and scaling observations are provided. Test results for the quenching approaches defined in section 3.8 are discussed. Summary comparison tables are then shown and conclusions drawn.

Measurements are performed at room temperature unless otherwise stated.

4.3.1 PPLUSPSTI_NISO SPAD Characterisation Results

4.3.1.1 Reverse I-V Response

The reverse I-V response was obtained using the procedure defined in section 4.2.1.

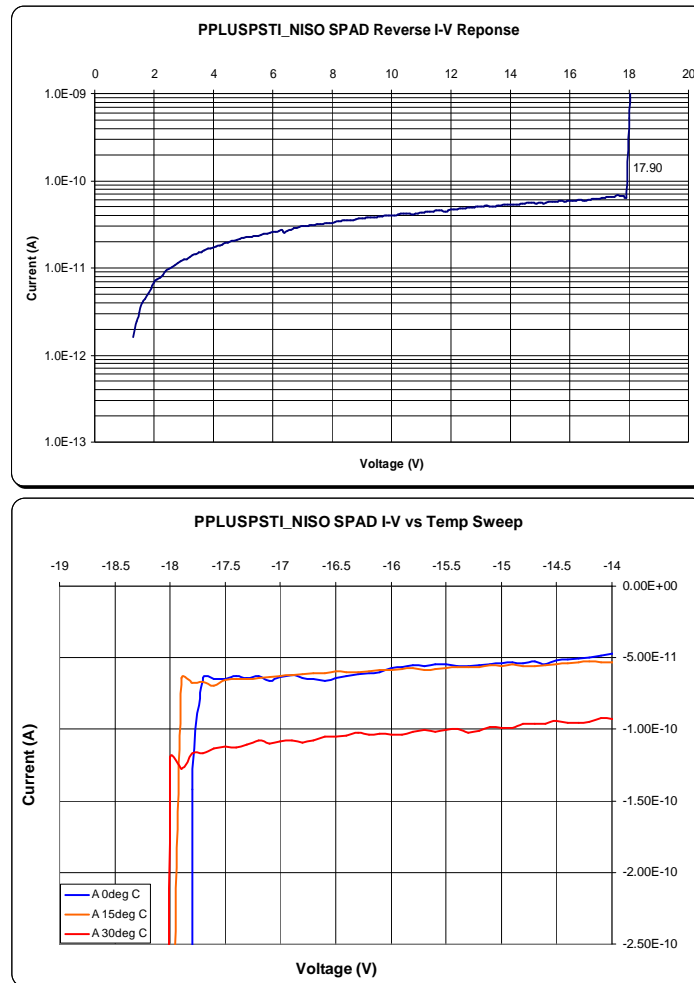


Figure 87: PPLUSPSTI_NISO SPAD I-V Response

The first graph shows the breakdown knee occurring at 17.9V with only 64pA of dark current. The measured result compares favourably with the simulated result of 17.5V. The second graph shows the variation of the breakdown knee with temperature as $\sim 6.7\text{mV}/^\circ\text{C}$. It can be seen that an increase of temperature also results in the expected elevation of dark current level.

4.3.1.2 Dark Count Rate

The dark count rate was obtained using the procedure defined in section 4.2.2.

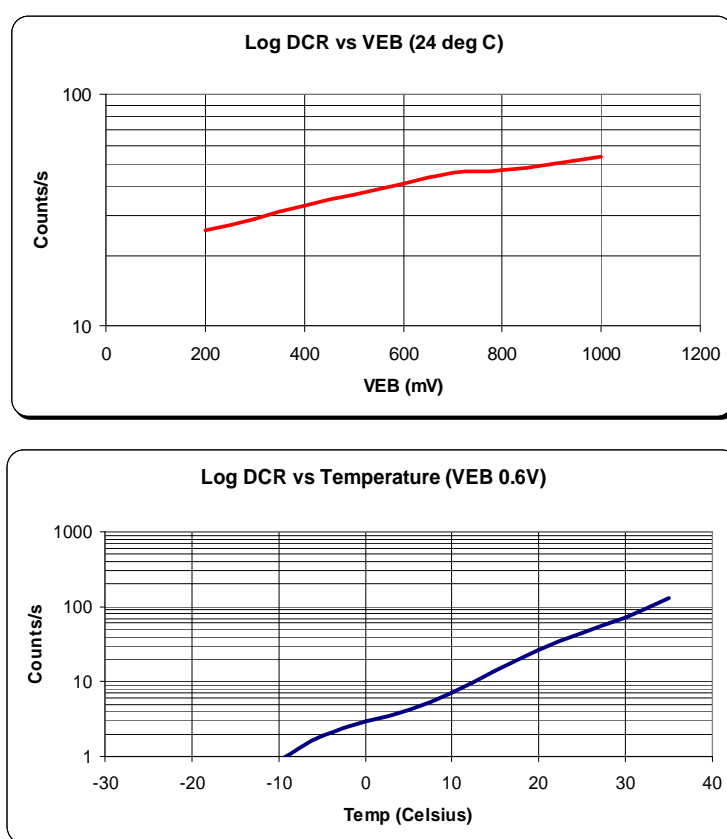


Figure 88: PPLUSPSTI_NISO SPAD Dark Count Rate

The upper graph shows the expected linear relationship between DCR and excess bias. The lower graph shows that this detector has a very low, thermal carrier dominated dark count level of $\sim 44\text{Hz}$ at room temperature and moderate excess bias, approximately doubling every 8°C . This is line with the intended reduction of peak junction E-field.

4.3.1.3 Photon Detection Efficiency

PDE was determined using the procedure defined in section 4.2.3.

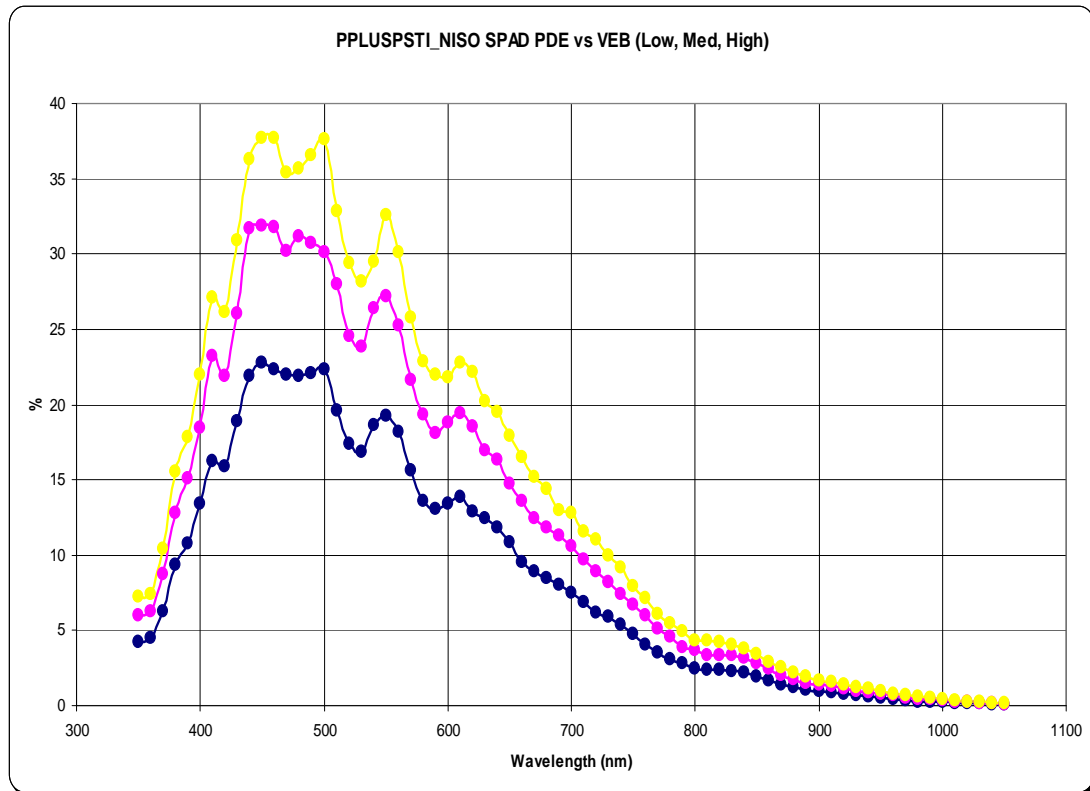


Figure 89: PPLUSPSTI_NISO SPAD Photon Detection Efficiency

The graph shows that this detector has the expected peak response in blue but with a reasonable extension into mid region wavelengths, commensurate with a deepening of the peak E-field location. This extends up to 800nm where it still has ~5% PDE at higher excess biases. The perturbations in the response are due to constructive/destructive interference patterns caused by the dielectric stack above the detector, which consists of several different materials with varying refractive indices.

4.3.1.4 Afterpulsing

Afterpulsing performance was determined using the procedure defined in section 4.2.5.

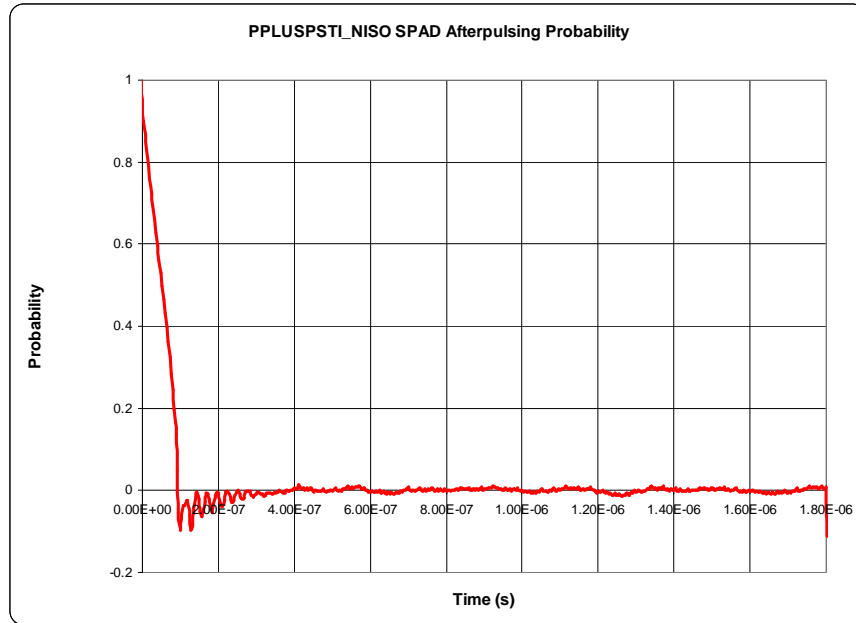


Figure 90: PPLUSPSTI_NISO SPAD Afterpulsing Probability

The autocorrelation graph shows that afterpulsing is very low or not present in this structure. This is in line with the reduction of junction capacitance and photoelectric gain.

4.3.1.5 Timing Resolution

The timing resolution for the SPAD was determined using the procedure defined in section 4.2.6. The almost symmetric gaussian response indicates efficient photon absorption and avalanche seeding in the high field active region, as discussed by Hsu, Finkelstein et al in [25].

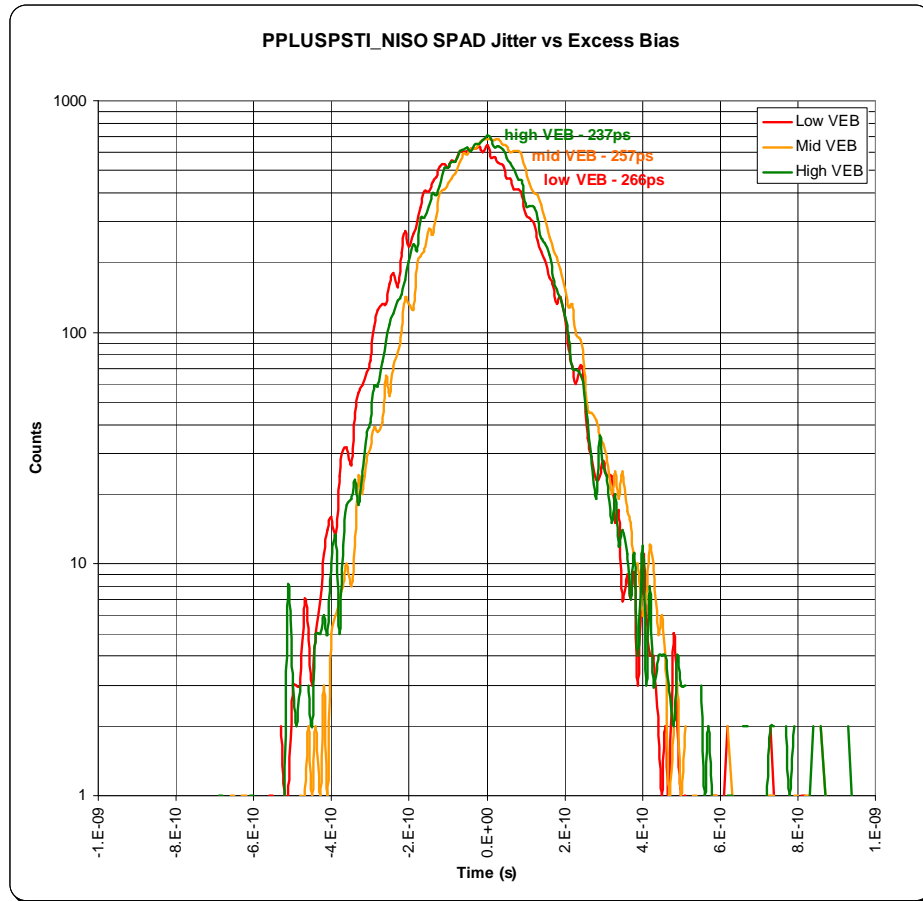


Figure 91: PPLUSPSTI_NISO SPAD Jitter

The graph shows five responses at different excess bias levels. The E-field increase at higher excess bias is sufficient to improve the measured timing resolution to a minimum of 237ps at the higher excess bias level, although this figure is slightly outside of the target specification. The lack of the normally problematic diffusion tail, characteristic of implanted guard-ring architectures, is due to limiting the volume of low field zones mainly via use of an enhancement mode structure.

4.3.1.6 Summary

The performance of the PPLUSPSTI_NISO SPAD is summarised in Table 15.

PPLUSPSTI_NISO SPAD		
Reverse Breakdown Voltage	-17.9V	target met
Excess Bias Range	1.275V	target met
Dark Count Rate	40Hz, 215 μ lx, 13.9pA/cm ²	target met
Photon Detection Efficiency	37% at 500nm	target met
Afterpulsing	0.02%	target met
Jitter	~237ps FWHM	target just missed
Active Area	50.26 μ m ²	target met

Table 15: PPLUSPSTI_NISO SPAD Summary

4.3.2 PPLUSPSTI_NWELL SPAD Characterisation Results

4.3.2.1 Reverse I-V Response

The I-V response was obtained using the procedure defined in section 4.2.1.

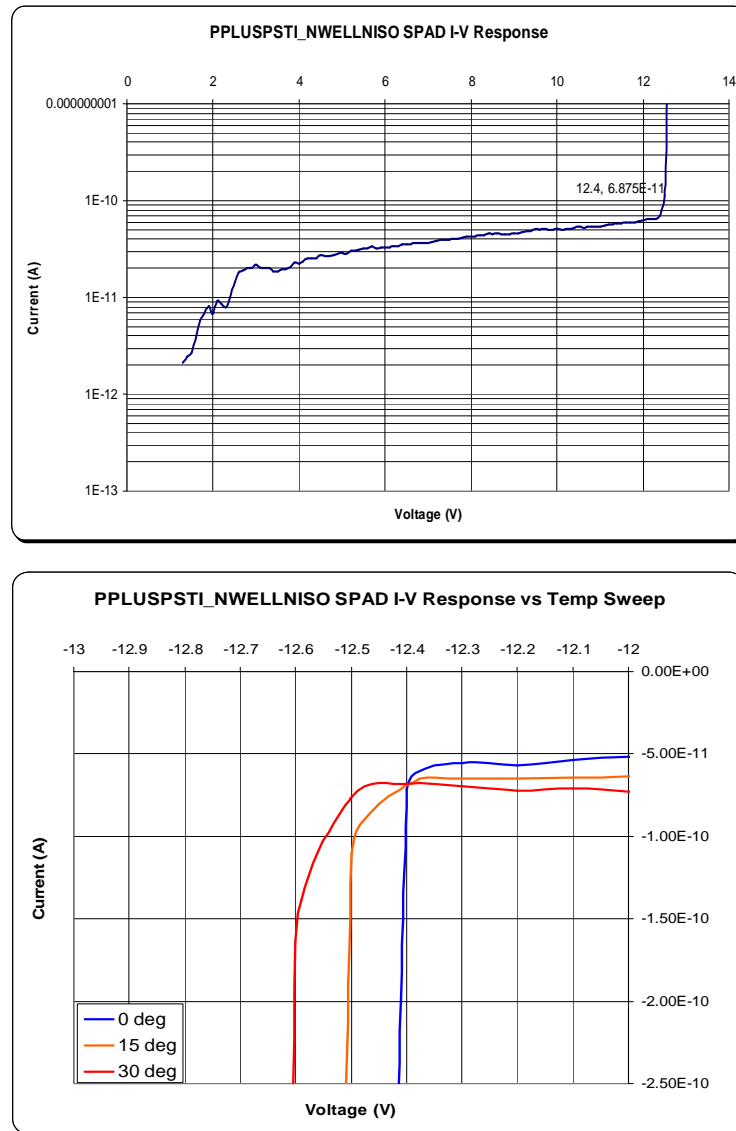


Figure 92: PPLUSPSTI_NWELL SPAD I-V Response

The first graph shows the breakdown knee occurring at 12.4V with only 69pA of dark current. This is in line with the reduction predicted by simulation (11.5V predicted), caused by the addition of NWELL doping to the junction. The second graph shows the variability of the breakdown knee over temperature as $\sim 6.7\text{mV}/^\circ\text{C}$.

4.3.2.2 Dark Count Rate

The dark count rate was obtained using the procedure defined in section 4.2.2.

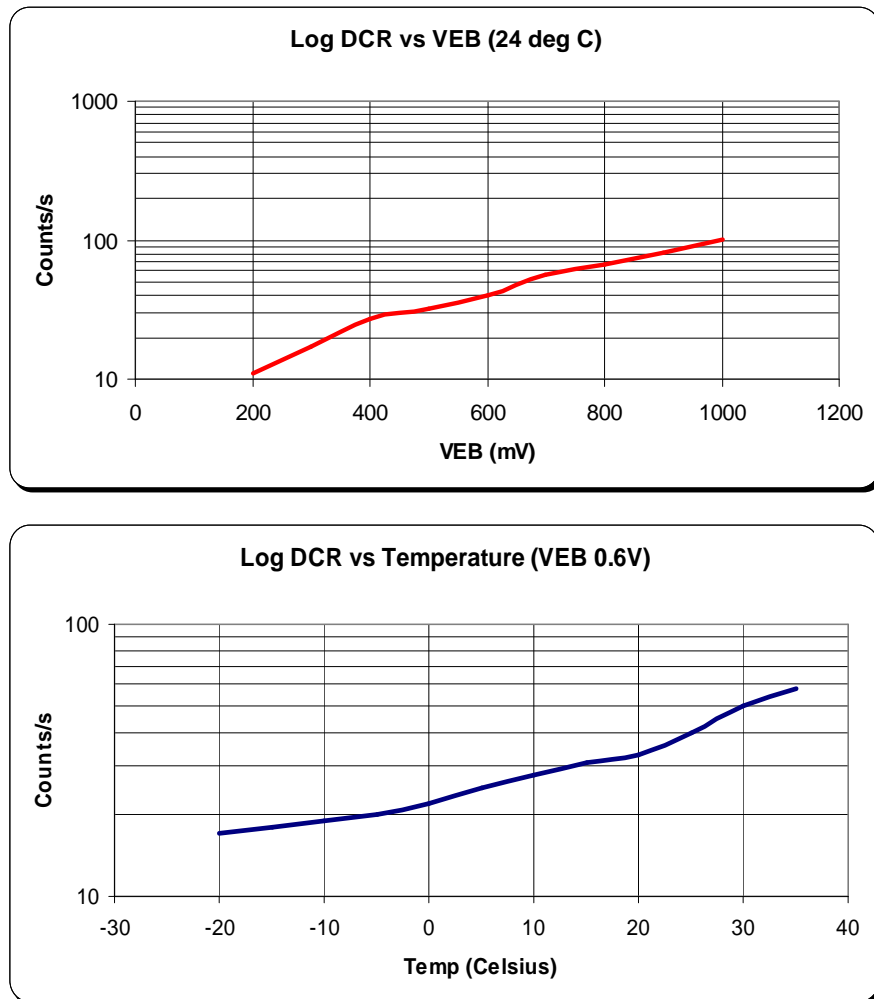


Figure 93: PPLUSPSTI_NWELL SPAD Dark Count Rate

The first graph shows the expected ~linear relationship between DCR and excess bias. The second graph shows that this detector has a low dark count level with less of a temperature dependent profile. This is to be expected with an increase of peak E-field due to addition of NWELL doping.

4.3.2.3 Photon Detection Efficiency

PDE was determined using the procedure defined in section 4.2.3.

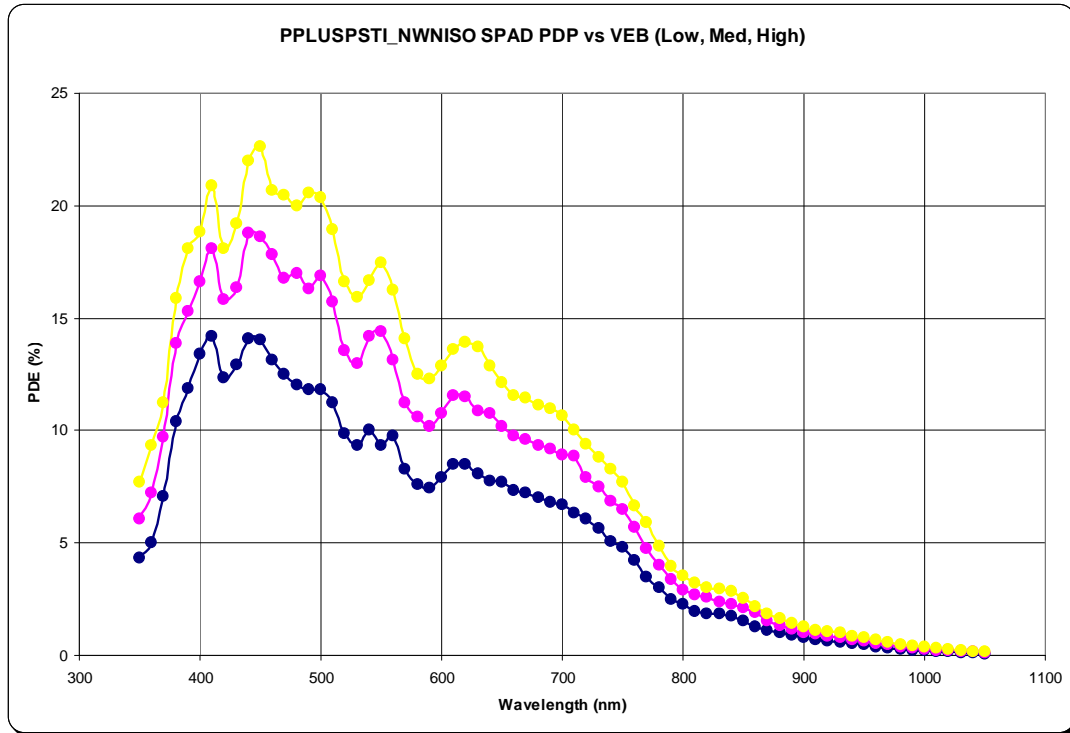


Figure 94: PPLUSPSTI_NWELL SPAD Photon Detection Efficiency

The graph shows that this detector has a comparatively lower peak response in blue and reduced extension into mid region wavelengths. At 800nm it has ~3% PDE at higher excess biases.

4.3.2.4 Afterpulsing

Afterpulsing performance was determined using the procedure defined in section 4.2.5.

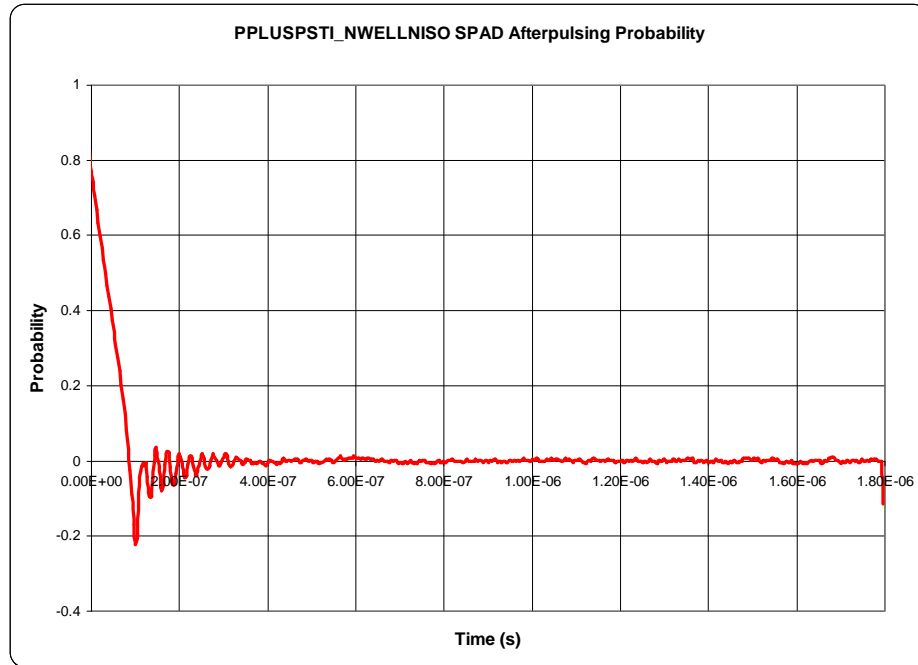


Figure 95: PPLUSPSTI_NWELL SPAD Afterpulsing Probability

The autocorrelation graph shows that afterpulsing is very low to the extent it can be largely discounted in this structure.

4.3.2.5 Timing Resolution

The timing resolution for the SPAD was determined using the procedure defined in section 4.2.6.

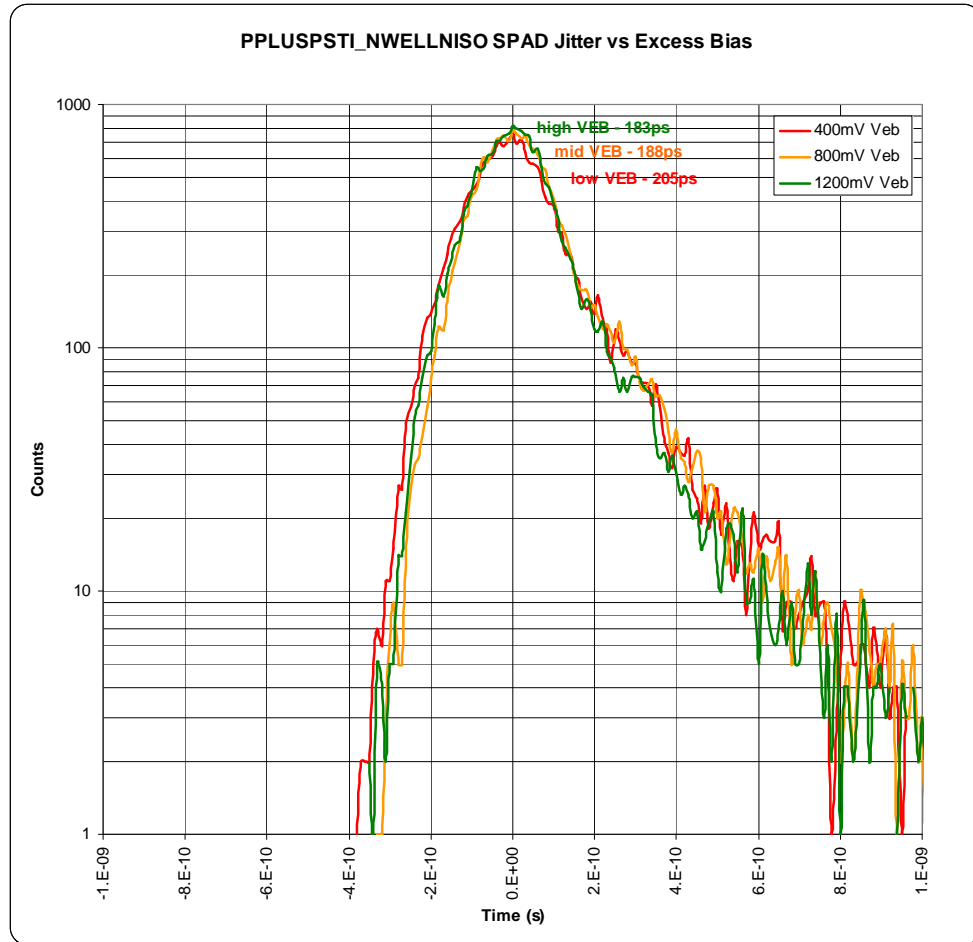


Figure 96: PPLUSPSTI_NWELL SPAD Jitter

The graph shows responses at different excess bias levels. Minimum jitter is measured as ~183ps at FWHM with an approximately Gaussian profile but more prominent diffusion tail than the other detectors. This characteristic is in line with the trend seen in detectors with higher junction electric field strength.

4.3.2.6 Summary

The performance of the PPLUSPSTI_NWELL SPAD is summarised in Table 16.

PPLUSPSTI_NWELL SPAD		
Reverse Breakdown Voltage	-12.4V	target met
Excess Bias Range	1.31V	target met
Dark Count Rate	40Hz, 349 μ lx, 22.7pA/cm ²	target met
Photon Detection Efficiency	22% at 450nm	target met
Afterpulsing	0	target met
Jitter	~183ps FWHM	target met, with tail
Active Area	50.26 μ m ²	target met

Table 16: PPLUSPSTI_NWELL SPAD Summary

4.3.3 PPLUSWELL_NISO SPAD Characterisation Results

4.3.3.1 Reverse I-V Response

The I-V response was obtained using the procedure defined in section 4.2.1.

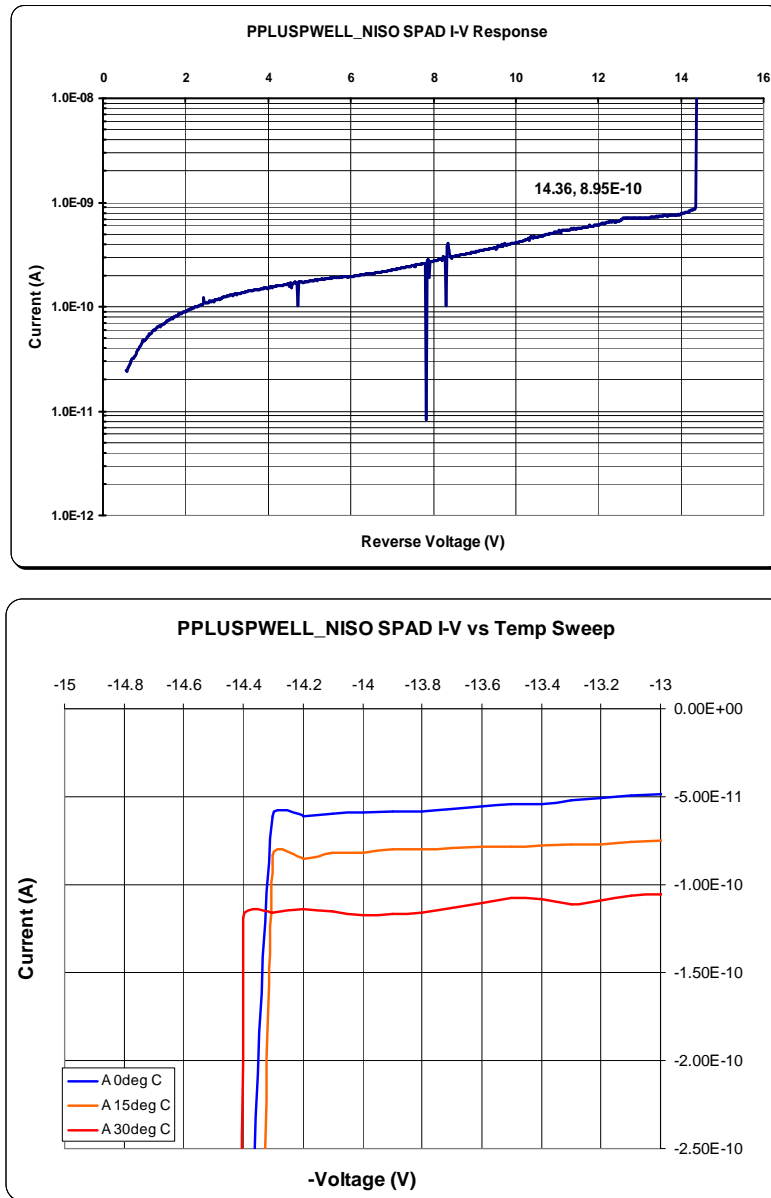


Figure 97: PPLUSWELL_NISO SPAD I-V Response

The first graph shows the breakdown knee occurring at 14.3V with 83pA of dark current, in very close agreement with simulation (14.4V). The second graph shows the variability of the breakdown knee over temperature as $\sim 6.7\text{mV}/^\circ\text{C}$.

4.3.3.2 Dark Count Rate

The dark count rate was obtained using the procedure defined in section 4.2.2.

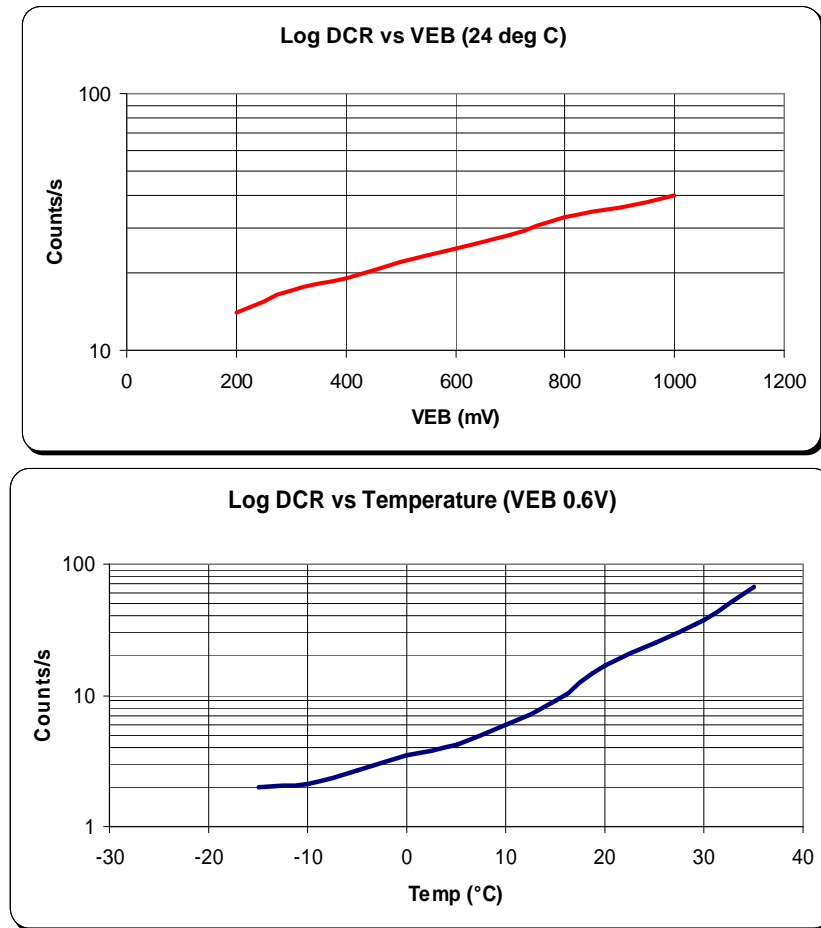


Figure 98: PPLUSPWELL_NISO SPAD Dark Count Rate

The first graph set shows the expected linear relationship between DCR and excess bias. The second graph set shows that again this detector has a very low, thermal carrier dominated dark count level.

Furthermore, the implementation of this design of detector in a 32x32 array as part of the Megaframe Project provided the opportunity for analysis and plotting of DCR population distribution for 1024 elements, shown below in Figure 99. The active region in this case was 7 μ m diameter.

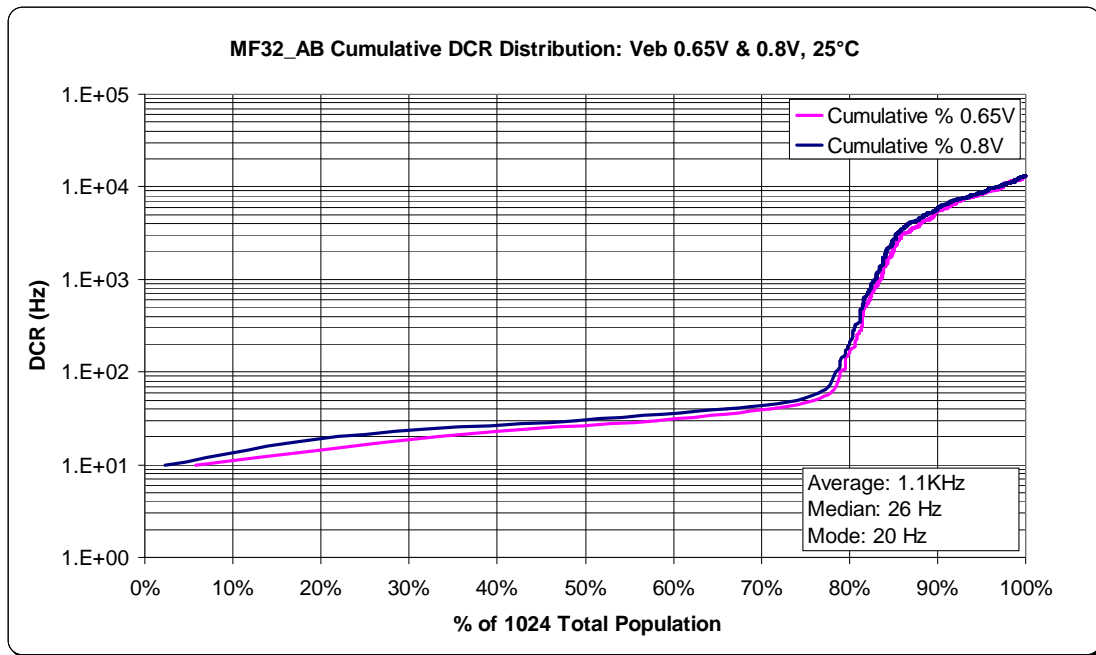


Figure 99: PPLUS_PWELL SPAD DCR Distribution

It can be seen that the population splits roughly into two groups: those with low DCR of <100Hz, and those with higher DCR up to 10 KHz. The split ratio is ~80:20. All 1024 SPADs were functional. The impact on DCR by increased excess bias is evident from the two traces.

4.3.3.3 Photon Detection Efficiency

Photon detection efficiency was determined using the procedure defined in section 4.2.3.

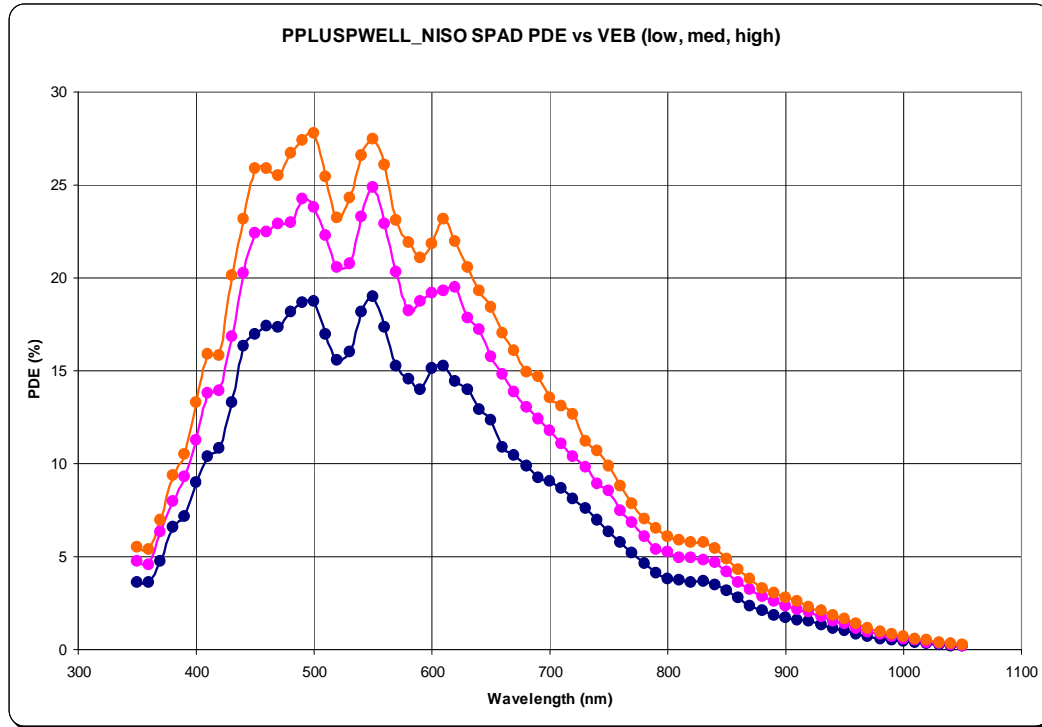


Figure 100: PPLUSPWELL_NISO SPAD Photon Detection Efficiency

The graph shows that this detector has the expected peak response in blue but with a reasonable extension into mid region wavelengths, commensurate with the distinct deepening of peak E-field in this structure. This extends beyond 800nm where it still has ~6% PDP at higher excess bias.

4.3.3.4 Afterpulsing

Afterpulsing performance was determined using the procedure defined in section 4.2.5.

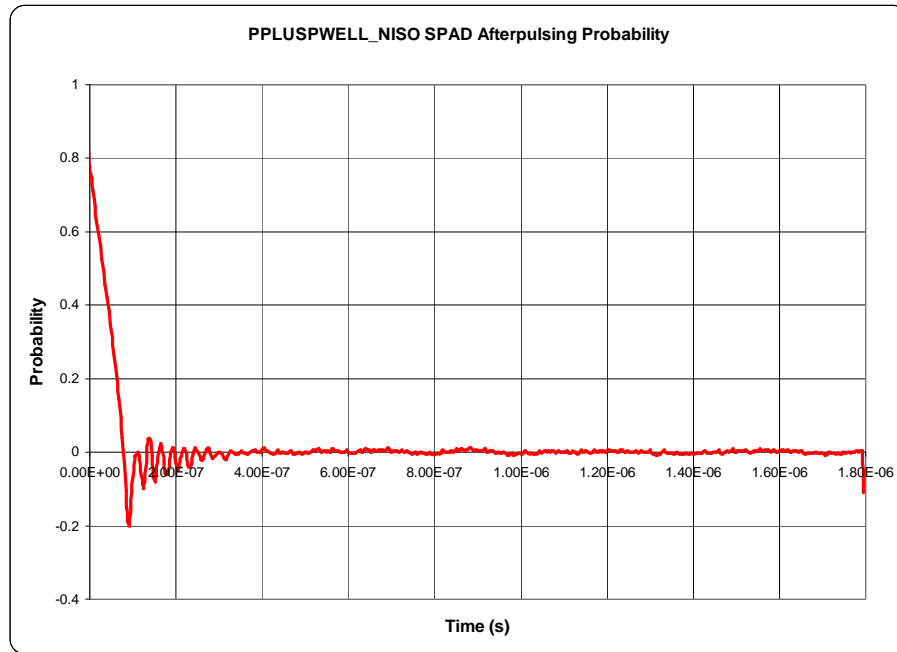


Figure 101: PPLUSPWELL_NISO SPAD Afterpulsing Probability

The autocorrelation graph shows that afterpulsing is very low to the extent it can be discounted in this structure (0.02%).

4.3.3.5 Timing Resolution

The timing resolution for the SPAD was determined using the procedure defined in section 4.2.6.

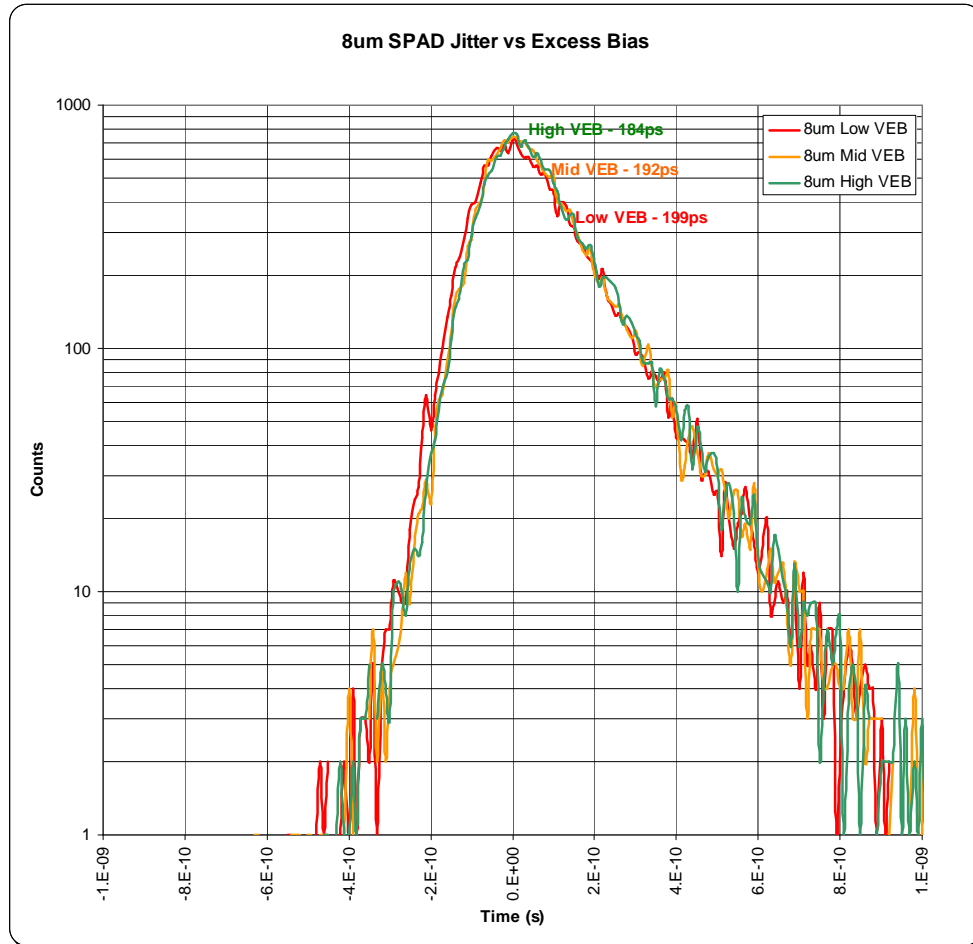


Figure 102: PPLUSWELL_NISO SPAD Jitter

The graph shows three responses at different excess bias levels. Jitter is measured as ~184ps at FWHM, high excess bias with a slightly more noticeable diffusion tail when compared with the other devices.

4.3.3.6 Summary

The performance of the PPLUSWELL_NISO SPAD is summarised in Table 17.

PPLUSWELL_NISO SPAD		
Reverse Breakdown Voltage	-14.36V	target met
Excess Bias Range	1.3V	target met
Dark Count Rate	25Hz, 122 μ lx, 7.91pA/cm ²	target met
Photon Detection Efficiency	28% at 500nm	target met
Afterpulsing	0.02%	target met
Jitter	~184ps FWHM	target met
Active Area	50.26 μ m ²	target met

Table 17: PPLUSWELL_NISO SPAD Summary

4.4 Scaling and Shaping Observations

The measurement of the scaling and shaping experiments outlined in section 3.7 yielded some revealing effects which are presented in this section. All shapes and sizes of detectors in this part of the research were configured with an optimised PMOS passive quench element, as defined in section 3.8.1. In this experiment, a reduced height optical stack was also used to further enhance the PDE figure. The device employed for this phase of research was the PPLUSPWELL_NISO_EPIPOLY structure shown in section 3.6.3.

4.4.1 Impact on I-V Response

Although all diameters of device shared the same P-N junction of PPLUS, PWELL into NISO, the respective measured I-V responses revealed a trend of slightly reducing breakdown voltage with area. This is shown in Figure 103.

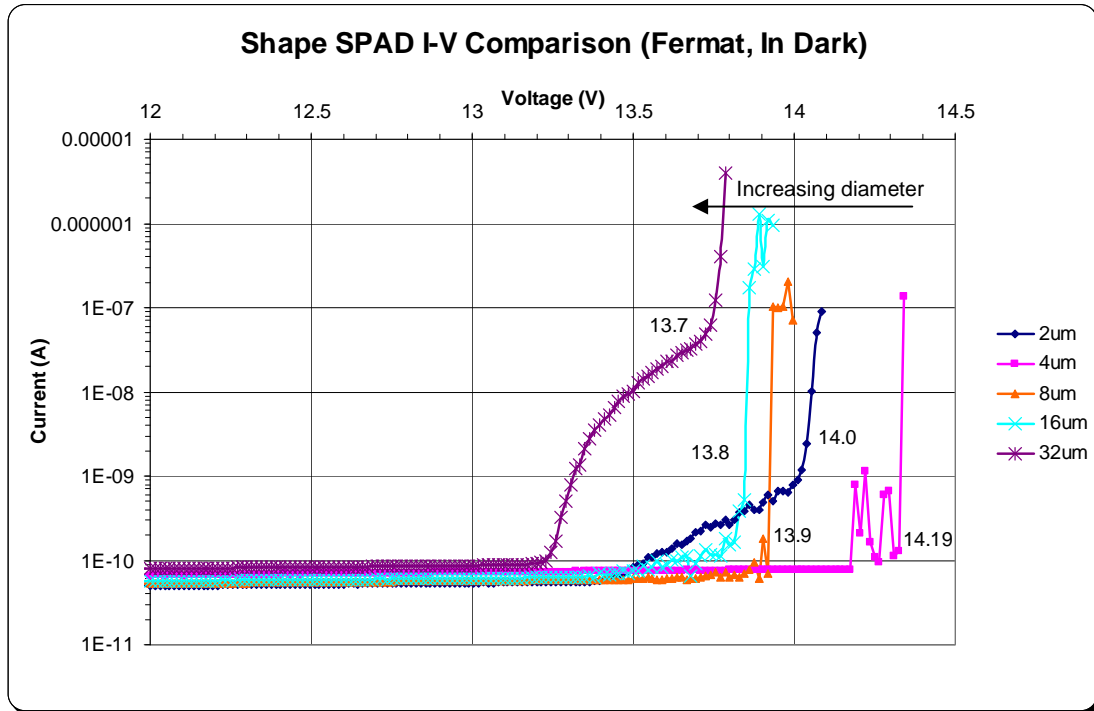


Figure 103: Impact of Diameter on I-V Response

The slight lowering of I-V breakdown knee may be due to the increased chance of encountering a localised region of higher electric field (i.e. a 'microplasma') in the larger area detectors [6, 8]. The smallest SPADs show a flickering behaviour at the breakdown knee, consistent with a low field junction with minimal carrier generation

rates, Zanchi [18]. Conversely, the largest detectors exhibit an early onset of breakdown similar to that observed for a smaller SPAD I-V response when exposed to light Haitz [8]. It is noted that in the chip tested above the particular $2\mu\text{m}$ device tested is out of sequence with the other diameters. The difference in breakdown voltages between different diameter devices must be taken into account when configuring excess bias voltage conditions. The trend is also consistent for circular, Fermat and square shapes.

4.4.2 Impact on Dark Count

The goal of this phase of characterisation was to determine which of the three detector shapes yielded the most consistently low DCR devices for each of the diameters implemented. DCR was measured using the procedure outlined in section 4.2.2 with excess bias conditions yielding a consistent 50ns dead time.

Typically, a loss of yield control for low DCR devices was observed for the Fermat and square devices. This is illustrated below in Figure 104 for a batch of 22 $4\mu\text{m}$ diameter devices of each shape.

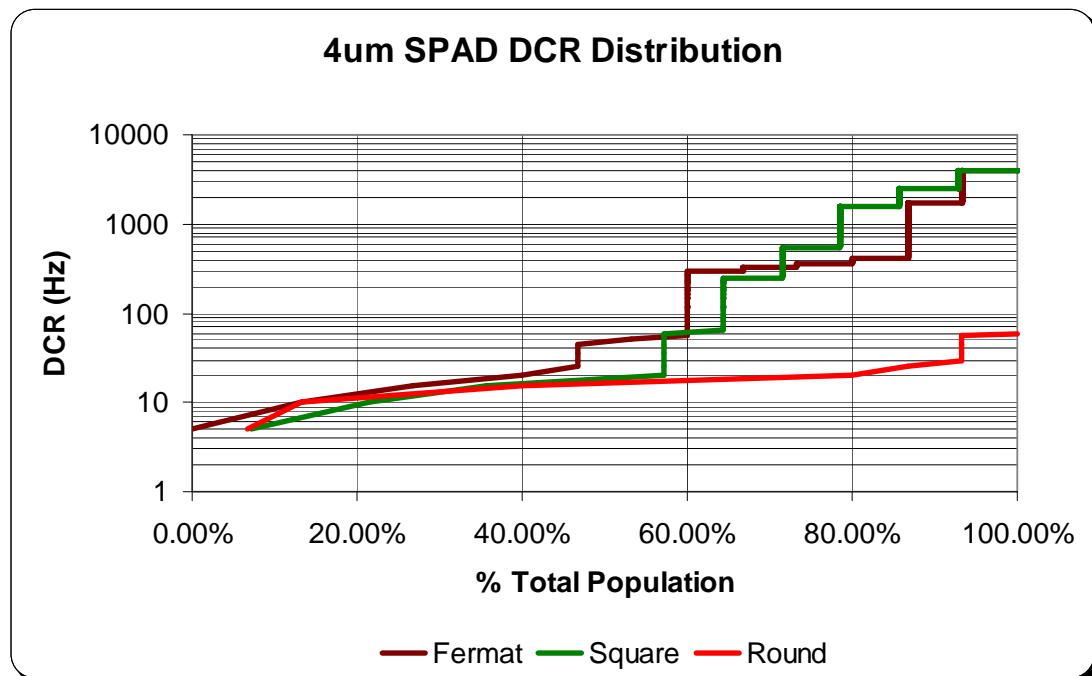


Figure 104: $4\mu\text{m}$ Diameter SPAD DCR Distribution

It can be seen that the round device yielded particularly well, with over 90% of devices below 30Hz. The data implies that independent of detector size, the round shape provides the highest population of low DCR devices.

Analysing the different diameters of round shaped devices, the average DCR of the main population may be plotted against area to enable observation of a trend, shown in Figure 105. The dotted black trace represents the expected DCR if scaled purely by area, using the 8 μ m device as a datum.

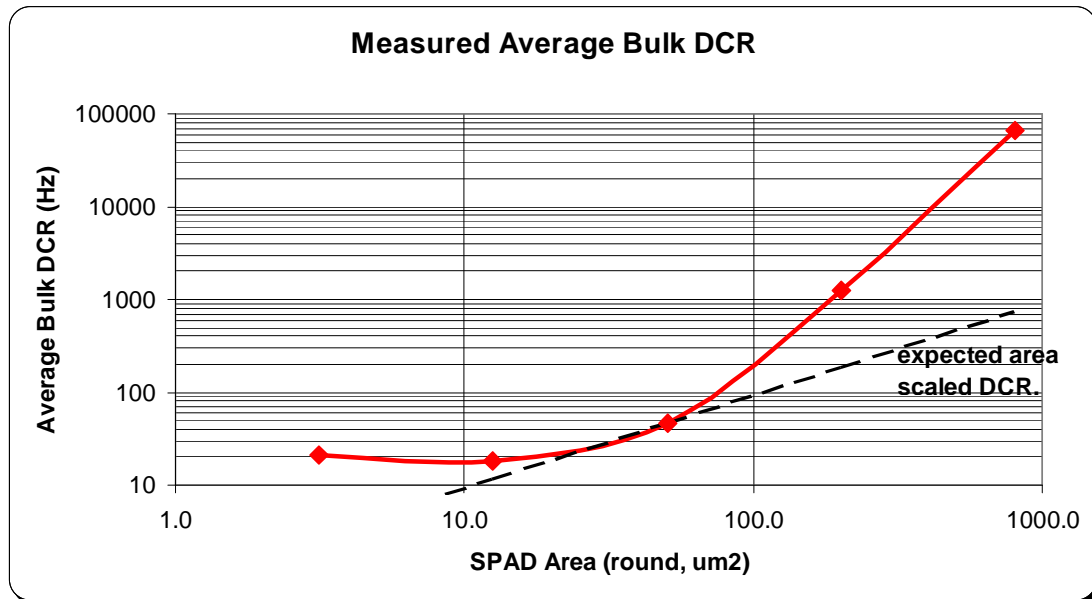


Figure 105: Round SPAD Dark Count versus Area

Thus it is clear that in agreement with Zanchi et al [92], actual DCR does not purely scale with area, and that there is a limit to the lowest DCR device area. If the DCR is plotted against SPAD diameter as shown in Figure 106, a more linear response is evident, permitting ease of predicting DCR for larger diameter devices.

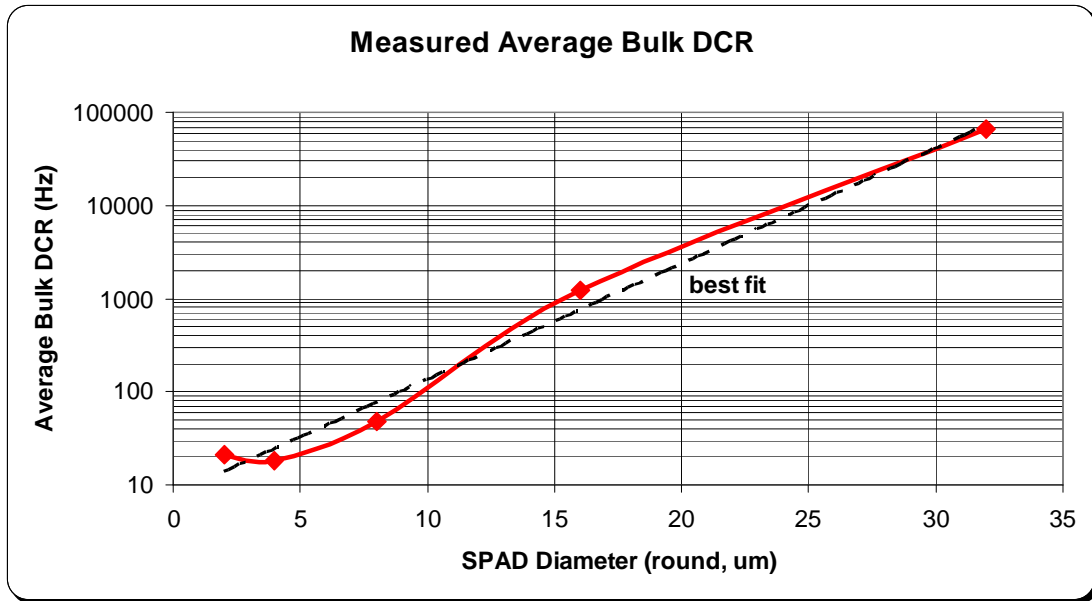


Figure 106: Round SPAD DCR versus Drawn Diameter

The data suggests the presence of a limit to the scalability of the structure, considering the necessity of having a contact to the detector anode.

4.4.3 Impact on Timing Resolution

Using the procedure outlined in section 4.2.6, the timing resolution figures for the different diameters of round devices with optimised PMOS passive quench were obtained for different values of excess bias voltage. The data obtained is shown in Figure 107.

The purple trace shows the data for the $4\mu\text{m}$ device, blue for $8\mu\text{m}$ and green for $16\mu\text{m}$ diameter. The FWHM figures obtained for the high excess bias traces are annotated on the graph.

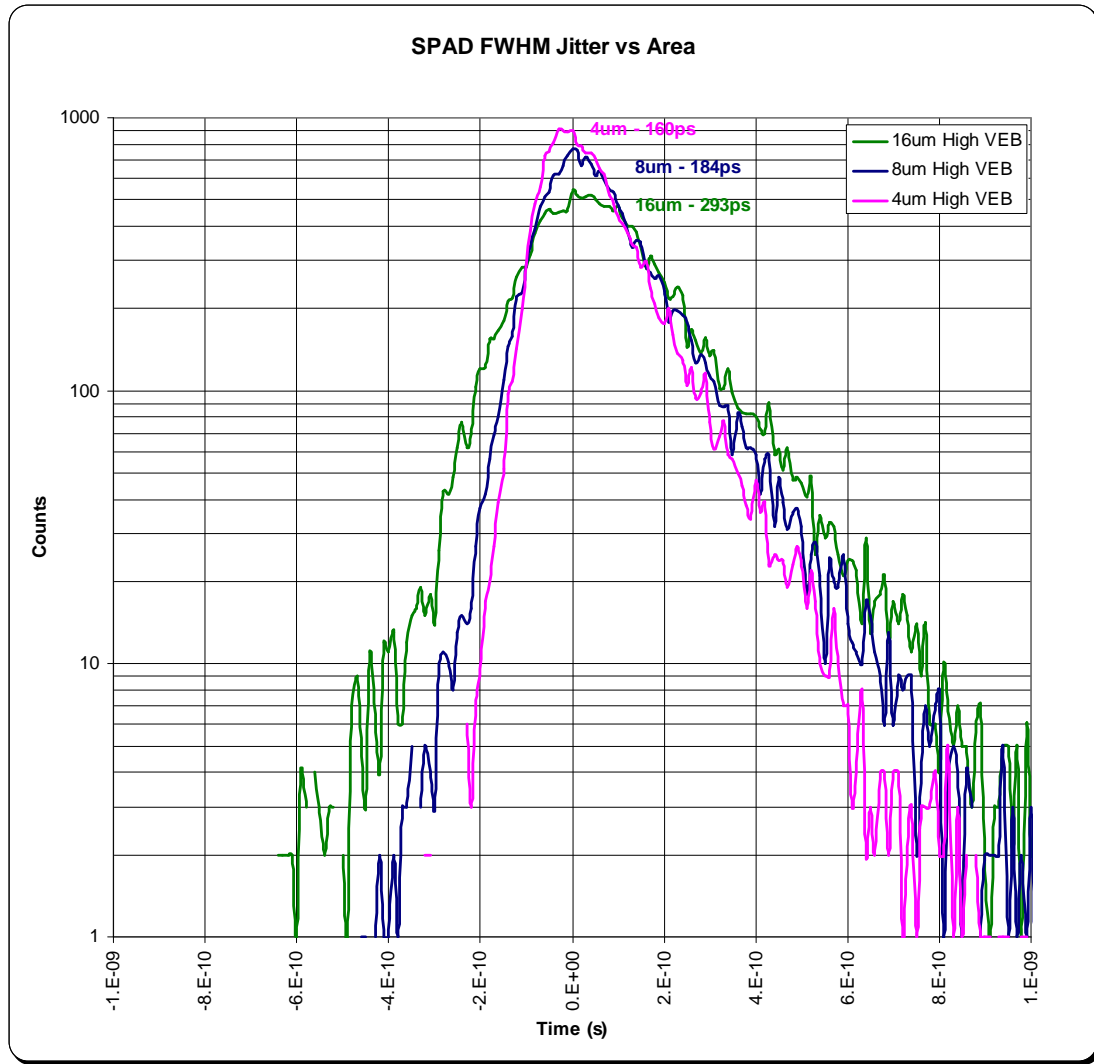


Figure 107: Impact of Diameter on Timing Resolution

The data shows good timing resolution performance for all diameters of detector, compared with the prior art. All devices exhibit improved timing resolution at the higher excess bias level, as expected with a higher electric field across the junction. The STI guard ring structures published by Hsu, Finkelstein et al [25] are reported as having remarkably low and consistent ~27ps FWHM timing resolution for the two diameter devices implemented (2 and 20 μ m). In [51] Lacaita et al report a decrease of timing performance with larger area detectors, due to the larger quasi-neutral field areas characteristic of implanted guard ring structures. For this work, the slightly poorer timing resolution of the larger devices is thought to be due to the increased possibility of slower avalanche propagation for photon arrivals at the centre of the active region.

4.5 Active Quenching Characterisation

The CMOS thyristor active quench circuit described in section 3.8.2 was implemented in silicon along with an accompanying PPLUSWELL_NISO SPAD as defined in section 3.6.3. A photomicrograph of the circuit implemented is shown below in Figure 108. The ‘cavity’ enhancement was not used over this implementation.

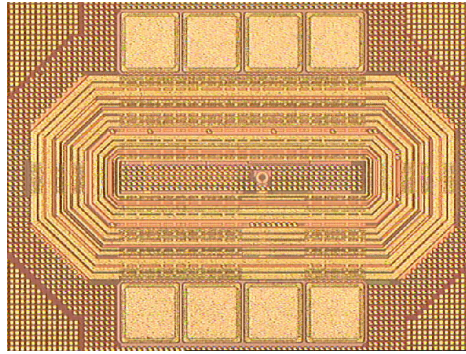


Figure 108: Active Quench Micrograph

The key performance parameters to be evaluated for an active quench circuit are the minimum dead time, and the resulting impact on count rate dynamic range and afterpulsing. The measured parameters are quoted below, compared to the optimised passive quench circuit described in section 3.8.1.

Quench	Min T_{dead} (ns)	Max T_{dead} (ns)	Max Count Rate (Hz)	Dynamic Range dB ¹	Afterpulsing Prob. (%)
Passive	15	400	85M	126.5	0.02
Active	8	300	120M	129.5	0.02 ²

Table 18: Thyristor Active Quench Performance

Notes –

¹: Dynamic range is the counting range from the dark count noise floor to the maximum possible count rate expressed in dB.

²: Afterpulsing probability varies with hold-off time. Configured with similar dead time to the passive quench, the active quench circuit yields comparable afterpulsing probability. However, at minimum hold-off time for the active quench loop (8ns), the afterpulsing probability is markedly increased to approximately 15%.

4.6 Comparison of Results

The performance of the new detectors is compared in Table 19 alongside the high DCR prior art published by Niclass et al in [20].

	PPLUS_NWELL	PPLUSPSTI_NISO	PPLUSPSTI_NWNISO	PPLUSPWELL_NISO
Vbreakdown (V)	-9.8	-17.9	-12.4	-14.36
Veb Range (V)	>2V	1.275	1.31	1.3
DCR (Hz)	100k	40	47	25
Peak PDE (%)	40	37	22	28
Afterpulsing %	not quoted	0.02	not detected	0.02
Jitter (ps FWHM)	144	237ps	183 (with tail)	184
Active Area (μm^2)	87.5	50.26	50.26	50.26

Table 19: SPAD Performance Summary

The average bulk population DCR obtained for each structure type, expressed as a light equivalent noise level is quoted in Table 20 below alongside the high DCR prior art datum (grey text) [20]. All detectors are round $8\mu\text{m}$ diameter unless otherwise stated.

Device	DCR (@ 25°C, & 0.6V V_{EB})	Light equivalent Noise Level (Lux)	Current/unit area equivalent
PPLUS_NWELL (7 μm)	30000	153m lux	13.5nA/cm ²
PPLUSPSTI_NISO	44	215 μ lux	13.9pA/cm ²
PPLUSPSTI_NWELLNISO	40	349 μ lux	22.7pA/cm ²
PPLUSPWELL_NISO	25	122 μ lux	7.91pA/cm ²

Table 20: SPAD DCR Summary

With DCR successfully reduced for all three SPAD junctions, comparison of other parameters is necessary to perform further ranking. Figure 109 shows a comparison of the three SPAD I-V responses.

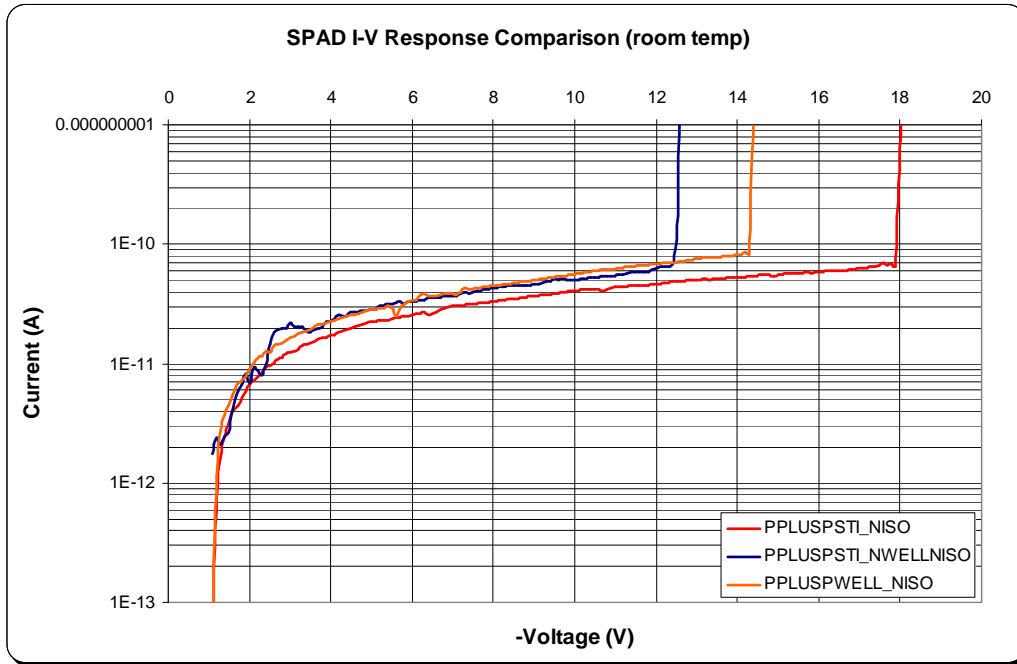


Figure 109: SPAD I-V Response Comparison

This data is relevant when comparing the PDE for each detector type at high excess bias, shown in Figure 110 below. It can be seen that the peak PDP magnitude follows the same trend of increase of breakdown voltage shown above.

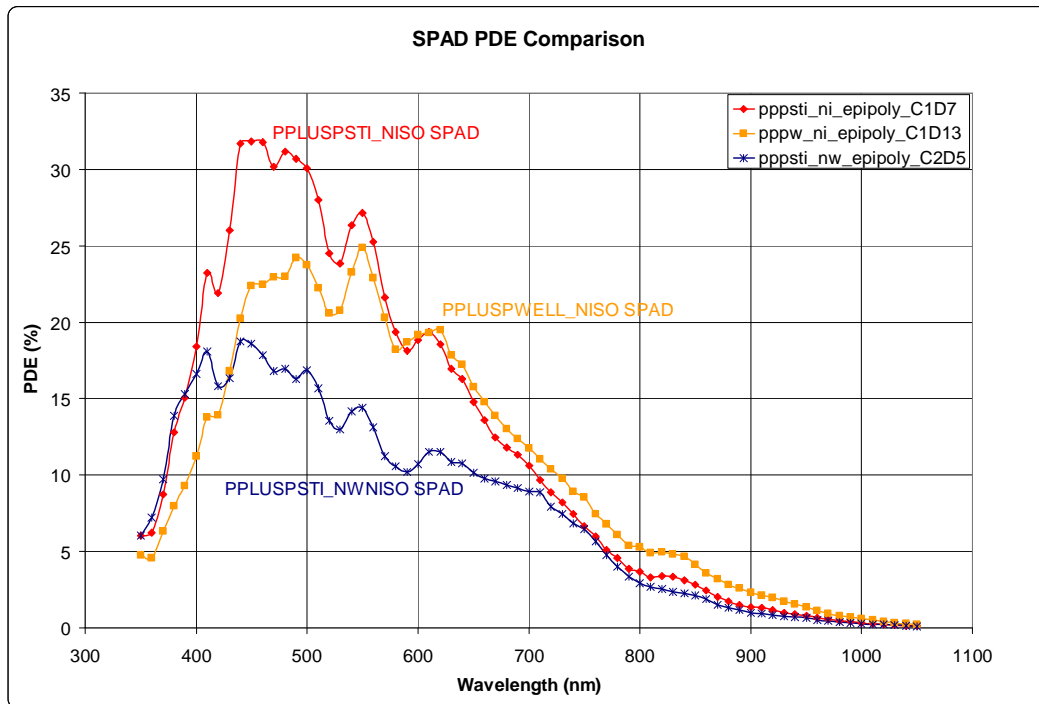


Figure 110: SPAD PDE Comparison Plot

The later use of CAVITY in the shaping and scaling experiments further boosted the measured peak PDE by another 10-15% by a reduction of the optical stack height.

Best case timing resolution was very competitive for all structures at around 200ps. Figure 111 shows a comparison of the timing resolution under best-case high excess bias operating conditions.

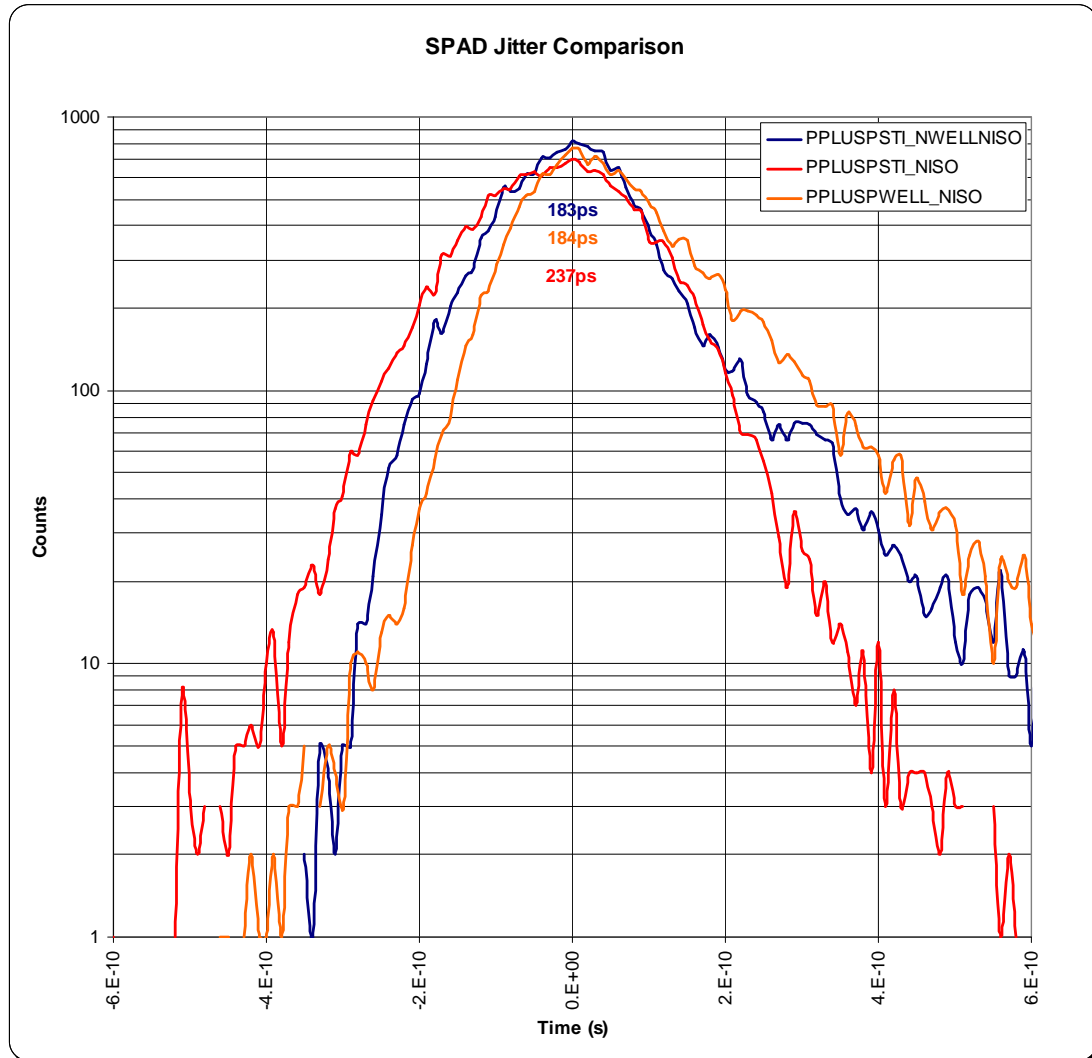


Figure 111: SPAD Jitter Comparison

It is evident that by a small margin the PPLUSPSTI_NWELLNISO device exhibits the best timing performance at FWHM, followed by the other detectors continuing in the order of descending peak electric field strength. The reduction of timing performance versus the 144 ps FWHM of Niclass et al [20] is the penalty paid for lowering the electric field strength to reduce DCR. The detectors exhibit slightly different diffusion tails, indicating that the full-width at say ¼ maximum is relevant for some applications.

These characterisation results enable the following ranking of *junction* type overall performance: (by ‘anode_cathode’ type)

1. PPLUSPSTI_NISO
2. PPLUSPWELL_NISO
3. PPLUSPSTI_NWELLNISO

As the second placed PPLUSPWELL_NISO structure uses standard (non imaging specific) implants only, it stands out from the other junctions. For this reason it was chosen for the shaping and scaling experiments. These experiments suggest that for optimum DCR and jitter, smaller, round devices of 4 μ m diameter exhibit a performance sweet spot. Nevertheless, the larger devices still have very competitive noise levels when compared with the published literature. The non linear increase of DCR with the larger devices has a likely periphery dependent component, possibly a cause of the inherent gettering feature of these detector designs. The impact of microplasmas on the DCR and timing performance of these larger devices could be improved by adding more phosphorous implant cathode contact implant regions. The data gathered permits a data-based trade off between overall fill factor, detection efficiency and noise, a requirement which may vary between applications.

4.7 Conclusions

In this chapter, the characterisation procedures required to measure key SPAD performance metrics were described. The data obtained for the new SPAD devices described in section 3.6 was then presented, along with results obtained from the quench circuits described in section 3.8.

The test results for the SPADs designed as part of this research matched the simulations performed. The goal of reduction of DCR from the high count rate detectors of Niclass et al [20], and Gersbach et al [38] fabricated in related technologies has been successfully achieved by reducing the peak electric field strength. This was achieved using implants already available in the unmodified process flow with suitable doping concentrations, profile and depth.

Both optimised passive quench and CMOS thyristor active quench circuits have boosted maximum possible count rates significantly, whilst improving dead time

consistency. The PMOS based passive quench permits an area efficient, simple approach, whilst the active quench circuit presented offers potential for the implementation of the first large *array* of fully controllable, actively quenched SPADs integrated alongside other data processing circuitry and conversion circuits such as TDCs. The design of a companion TDC is described in the next chapter.

5 Time to Digital Converter Array Design

In this chapter requirements for a TDC suitable for array implementation are presented. The resulting gated ring oscillator based design which was chosen is presented, along with proposed timing for TCSPC at 1Mfps. A detailed review of the design of each of the TDC sub-blocks is provided, with resolution, power consumption and data-path oriented physical implementation aspects as key considerations. The additional design for support of 1Gbps data readout, both on and off-chip is then presented. Performance expectations for the TCSPC system are derived and discussed.

5.1 Target Specification

The goal of the TDC design part of the research was the creation of a converter structure suitable for scaling to large format *arrays*. Such a TDC structure differs from a stand-alone single TDC design, as an arrayable TDC is a trade off between time resolution, word width, complexity, area, power consumption and accuracy. The *Gated Ring Oscillator* (GRO) approach was chosen for the TDC design implemented as part of this research. This was due to the inherent flexibility of this structure in achieving the above requirements whilst minimising the opportunities for introducing system noise as shown in Figure 36. The flexibility of this approach also permitted operation in two modes: photon counting and photon time-of-arrival measurement.

The GRO approach provides an incident light level intensity dependant power-consumption profile. This is a particularly desirable feature for non-cooled arrayed implementations where in reverse start-stop operation (Becker [30]) only pixels stimulated by a photon arrival consume digital core power. Reverse mode TCSPC is a popular mode of operation for low incident photon arrival rate applications such as FLIM, the primary mode for the Megaframe Project camera. Furthermore, the GRO architecture does not require the broadcast of very high speed clock signals over a large distance. The clock trees required in such cases introduce undesirably high quiescent power consumption due to shoot-through currents and image droops due to signal propagation variation. Distribution tree structures also do not scale well with increased array size.

Although oscillator jitter is a concern for the GRO structure, incorporating companion structures in order to bound the long term jitter provides an opportunity to globally

calibrate the ring oscillator array to a reference PLL. This also enhances robustness to power supply noise induced variations.

The 130/90nm hybrid imaging CMOS process includes options for high speed transistors, enabling the possibility to achieve the challenging time resolution target specification of <100ps. An area specification of 50x50 μ m was set for the overall TDC design to be compatible with the SPAD designs outlined in section 3.6.

The overall GRO TDC performance target for this work is summarised in Table 21 and is interpreted from the Megaframe technical deliverable document ‘From Concepts to Requirements’ D2.1, [93].

Metric	Quantity*	Units
Resolution	<100	ps
Conversion Range	~100	ns
Word Length	10	Bits
DNL	< \pm 0.5	Bits
INL	<1	Bits
Uniformity (σ)	<3	%
Jitter (mean)	<3	%
Area	~50x50	μ m
Power Consumption ¹	<50	μ W
Calibration for PVT	Array basis	-
Array Size	32x32 (1024)	Pixels

Table 21: GRO TDC Performance Target

¹ - retrospectively applied.

5.2 Description of Chosen TDC Architecture

The GRO TDC design is based on a high speed ring oscillator and ripple counter. A block diagram representation of this architecture is shown in Figure 112. A photon arrival activates the SPAD whose output is fed to the logic block. The Logic block enables the oscillator, the ripple counter then being incremented every subsequent period of the ring to create a coarse time to digital conversion. The conversion process is halted by a time stamp reference, usually an illuminator activation signal. This power efficient mode of operation is known as TCSPC ‘reverse mode’ start-stop. The

states of the internal nodes of the ring at this precise moment in time are then decoded to extract a fine conversion contribution. Coarse and fine bits are combined to form a 10bit output, which is buffered into memory to allow for pipelined readout, to minimise conversion dead time and maximise potential frame rate. The logic block generates ring oscillator differential control signals using only high-speed combinatorial logic to avoid flip-flop setup and hold violations.

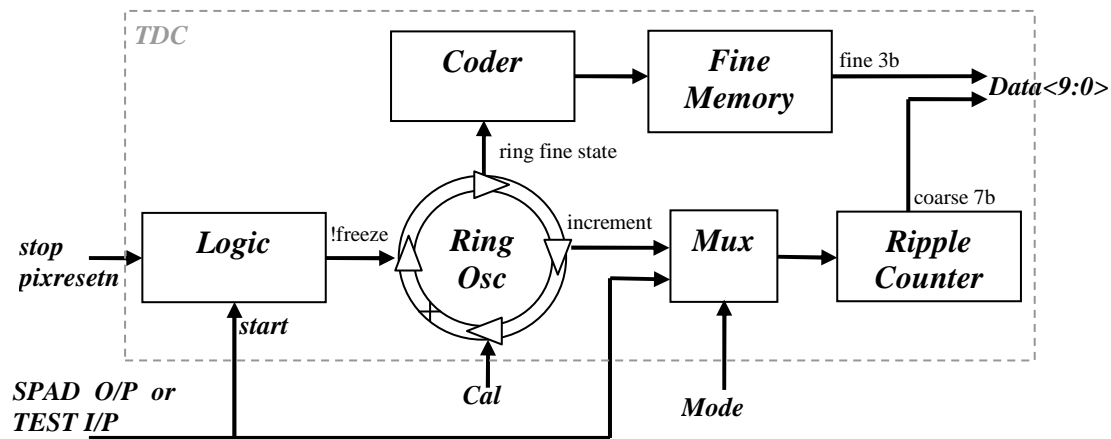


Figure 112: GRO TDC Block Diagram

The multiplexer block allows a photon counting mode where the GRO is bypassed to enable the detector output to directly increment the counter.

The TDC is instantiated along with the SPAD, memory and glue logic to create the pixel which forms the basic building block of an array. A simplified representation of this pixel showing how the TDC is integrated is shown in Figure 113.

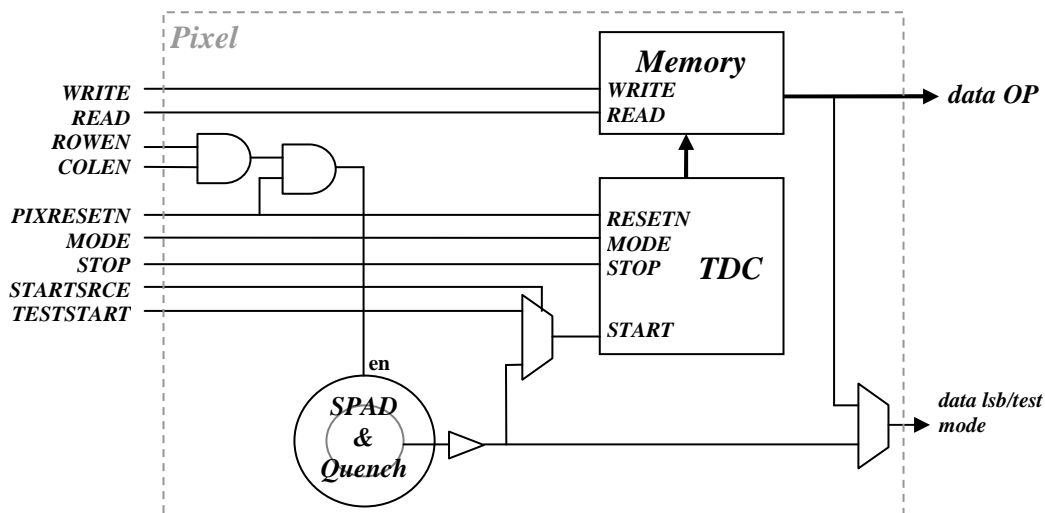


Figure 113: GRO TDC-Pixel IO

Dynamic range can be doubled in an area efficient manner by adding one extra counter flip flop and pipeline memory element. The pixel design supports pipelined data readout and may be instantiated many times to create any desired array format.

5.3 GRO TDC Timing

A simplified timing diagram of the pixel operating in time correlated (TCSPC) reverse mode at the target 1Mfps is shown below in Figure 114. This example timing mode is intended for fluorescence lifetime microscopy (FLIM).

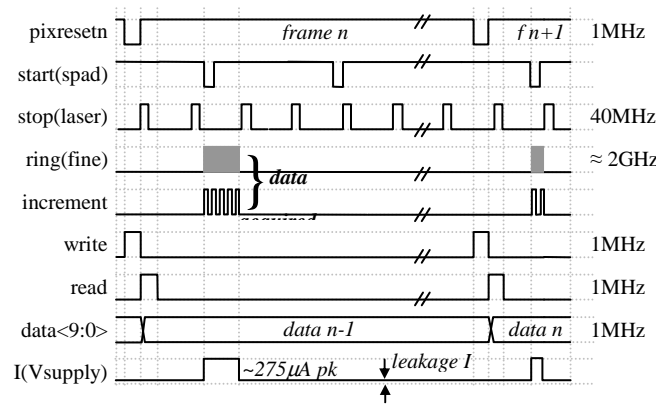


Figure 114: GRO TDC Timing

During each $1\mu\text{s}$ frame n there are multiple laser excitation pulses (40 in this case). This means that there are 40 potential excitation-detection cycles which may yield an emitted photon from the fluorophore or quantum dot marked bio-sample. Low photon yields are common in small volume assays with intentionally low illumination power so as to avoid pile up and sample bleaching.

The TDC must commence a time conversion on the first detected photon and hold the result in memory until it can be read out via a pipeline memory during frame $n+1$. This can be seen in Figure 114, where the second photon arrival during frame n is ignored. Reverse mode operation dictates that upon the first photon arrival the ring oscillator becomes enabled and is permitted to run freely until the next laser fire signal is received. As well as resulting in a light level dependent average power consumption, any quiescent (zero photon) consumption is limited only to that which is the result of signal distribution and leakage.

The *increment* signal is responsible for incrementing the ripple counter, which yields the upper 7 coarse data bits. Following the subsequent laser fire signal, the ring oscillator state is frozen and decoded by the Coder block to provide the lower 3 fine data bits.

At the end of the $1\mu\text{s}$ frame n , coarse and fine bits are written into the buffer memory via the *write* signal. The data readout subsystem then provides the means to extract the data row-wise in a conventional rolling blade readout manner via the signal *read*. Alternatively, the pixel can operate in time-uncorrelated photon counting mode via bypassing of the GRO and simply ripple-counting the number of photon events within a defined, variable exposure time.

5.4 GRO TDC Block Descriptions

5.4.1 Logic Block

The Logic Block forms the interface between the TDC control signals coming from the distribution trees, the photon detector (SPAD), and the ring oscillator.

Figure 112 shows that it is the Logic Block where the oscillator gating function resides, i.e. the generation of the freeze/unfreeze signals which dictate when the ring oscillator is permitted to be active (and consume power). This block has an important role as well as a series of stringent timing requirements. It must provide precise, glitch free control of the ring under the following operating scenarios:

- i. No photon arrival.
- ii. Very early photon arrival.
- iii. Mid-range photon arrivals.
- iv. Very late photon arrival.
- v. Multiple photon arrivals within a single conversion period.
- vi. Reset (low power) conditions.

For these reasons the Logic Block is constructed only from combinatorial, standard cell high-speed logic, so avoiding the issues of setup and hold time violations associated with a flip-flop based implementation. The circuit schematic is shown in Figure 115.

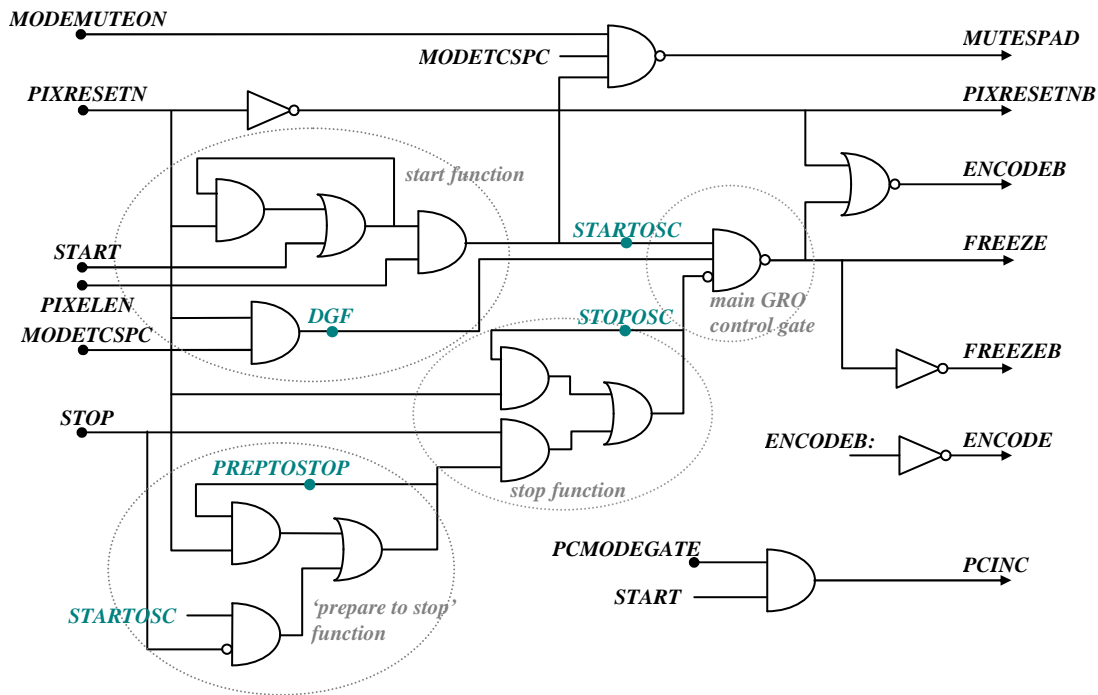


Figure 115: Logic Block Schematic

There are several key sub-functions within the Logic Block, picked out in grey above. All of the main sub-functions contribute to the creation of the ‘*FREEZE*’ signal (and its inverse) which is responsible for permitting the ring oscillator to operate via the *main GRO control gate*.

The *start function* ensures that the *FREEZE* signal is not asserted unless the SPAD has fired and the system is not in a reset condition.

The *stop function* ensures that the *FREEZE* signal is not de-asserted unless a *STOP* signal has been preceded by the correct output from the *prepare to stop* function, outwith a reset state.

The *prepare to stop* function ensures that the *stop function* cannot operate without the ring being first started. This feature removes the potential for incorrect *stop function* level sensitivity to the *STOP* input signal.

All sub-functions are reset by the low going signal *PIXRESETN*.

5.4.2 Ring Oscillator

The ring oscillator is the heart of the TDC. Its operation largely defines the key performance metrics of the converter; time resolution, linearity, power consumption, robustness to supply variations and calibration capability. It is responsible for incrementing the coarse ripple counter, and must hold its frozen dynamic state until the end of the frame when the decoded fine value can be transferred to pipeline memory.

Conventionally built from a chain of inverter stages, the jitter of the oscillator is a primary design consideration, as well as the requirement for integrating start-stop and reset control function elements. In [94], Herzel and Razavi indicate that oscillator jitter is heavily dependent on noise injected via supply and substrate, rather than temporal noise sources of the MOS transistors which make up an inverter stage. Therefore to bound the long term jitter, the ring oscillator is embedded within a PLL structure, with an NMOS regulation element dedicated to each stage. This regulation element provides both a means of speed control as well as rejecting power supply transients.

Furthermore, jitter, matching and linearity performance is enhanced by using a small number of wide-swing, differential buffer stages with zero static power consumption. Critically, the use of differential stages permits the chain to be constructed of a binary-power number of stages by swapping the polarity of feedback at the last element. This results in the fine code space being inherently perfectly mapped to the magnitude of time represented by the coarse LSB, permitting simpler fine state encoding. This is not the case for a single ended based ring oscillator which also requires either an odd number of stages or complex look ahead logic [73] to maintain positive feedback induced oscillation. A differential implementation also results in twice the number of potential per-stage output states compared to a single ended version. This allows the ring to be densely constructed from a reduced number of stages, in this case four. This results in eight potential fine output states.

A simplified schematic of the gated ring oscillator implemented is shown in Figure 116 alongside the Coder block. The role of the transmission gate switches which control ring oscillation, state sampling, initialisation and encoding can also be seen.

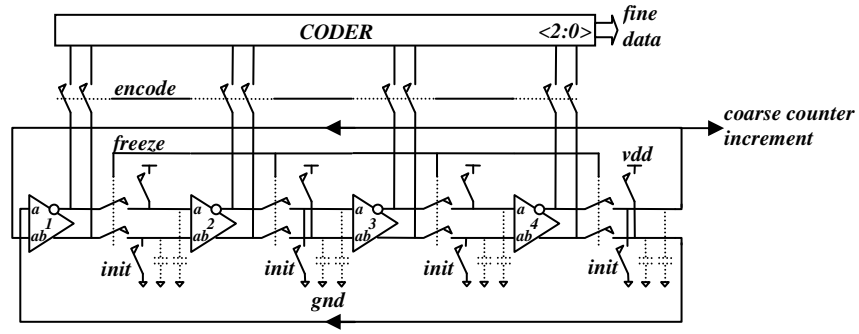


Figure 116: Differential Gated Ring Oscillator

Each buffer stage is constructed from the differential circuit similar to that of Nissinen et al in [72], with W/L chosen as a trade off between operating current and gate capacitance to achieve the shortest propagation delay, as shown in Figure 117. For a 1.2V core digital supply, the minimum propagation delay of this element is ~ 170 ps. A faster, 50ps ring design has also been implemented which trades off better time resolution for increased power consumption. An integrated NMOS speed control and regulation element is included. An NMOS device provides better supply rejection in this configuration when compared with the PMOS equivalent due to the fact that supply borne noise does not impact NMOS gate-source voltage. This device is sized as a 300 μ A current source with 300mV $V_{gs}-V_t$ and body effect impact.

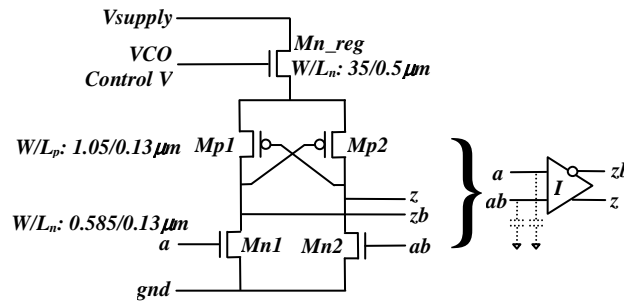


Figure 117: Differential Inverter

For the four stage ring constructed from the above element, the result is in an ‘ON’ state maximum supply current of approximately 275 μ A being drawn. The maximum duration for this is equal to the frequency of the system STOP signal.

The ring oscillator circuit and layout are a critical to the timing resolution and power consumption of the converter. Furthermore, the linearity of the TDC is dependent on

the careful balancing of the parasitic capacitances ($\sim 15\text{fF}$) between each differential stage. For this reason, the oscillator has been laid out in a ring shape as shown in the Cadence layout view of Figure 118. The alternative linear layout results in an inevitably increased, non-matched parasitic capacitance at the last (feedback) stage.

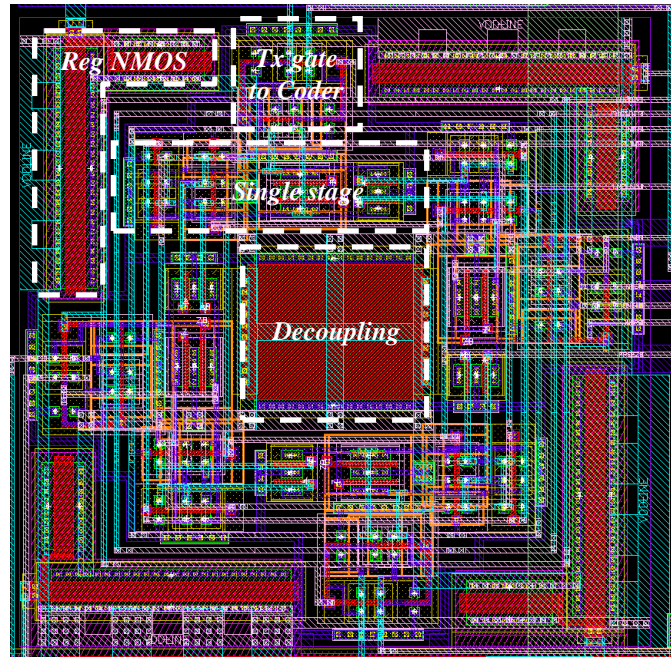


Figure 118: Differential Ring Oscillator Layout

Extracted simulations are necessary to check for balanced parasitic capacitances as well as to ensure that STI induced stress parameters have not significantly affected linearity by the skewing of buffer thresholds. It should be noted that the careful use of STI stress may actually be utilised for speeding up the MOS elements of the buffer if better time resolution is required in future.

The layout also incorporates L-shaped, wraparound NMOS regulation elements and wide metal power connections around the ring periphery with a supply decoupling capacitance provided in the centre. For optimal robustness to supply noise effects (both susceptibility and emission), the ring oscillator is placed in its own isolated well zone.

5.4.3 Coder

The role of the Coder block is to interpret the frozen state of the ring oscillator, and map this to the correct fine state binary code. This process must only occur when the Coder is instructed to do so at the end of an acquisition frame. The Coder block is

constructed from standard cell, combinatorial logic, with the resulting minterms being determined via interpretation of conventional Karnaugh maps. The schematic for the Coder is shown in Figure 119.

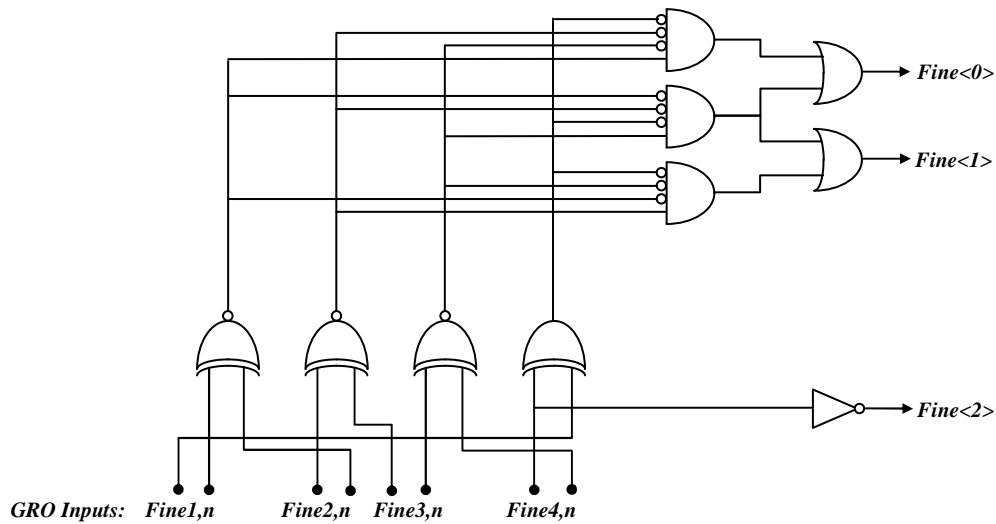


Figure 119: Coder Block Schematic

5.4.4 Ripple Counter

This cell is a conventional, 7 stage ripple counter, responsible for counting the TDC ‘coarse’ events, i.e. the number of full cycles of the ring oscillator. A ripple counter is ideally suited for this role, as it places minimal capacitive load on the last stage of the ring oscillator and its associated multiplexer cell. Furthermore, only the first element of the chain of flip-flops has a high speed requirement. The schematic is shown in Figure 120.

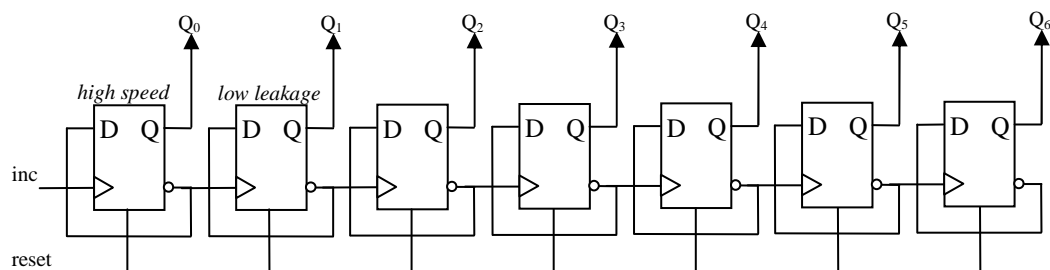


Figure 120: Ripple Counter Schematic

5.4.5 Memory

The memory blocks shown in Figure 112 and Figure 113 are constructed from the simple unit cell structure shown in Figure 121.

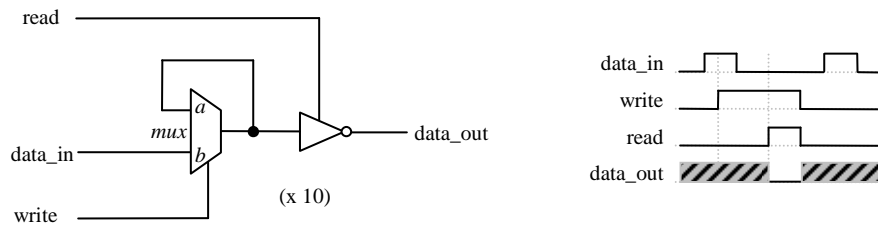


Figure 121: Memory Unit Cell and Timing

The memory block is constructed from standard cell library elements. Assertion of the *write* signal sets the multiplexer into a fed-back state, protecting the initial input data value. The inverted data can then be output on to the system databus via assertion of the *read* signal. The output data lines are otherwise tri-stated. The memory cell may be instantiated as many times as required, in this case 7 for coarse data and 3 elements for fine data.

5.5 Quench Cell

The Quench Cell is an integral part of the GRO TDC pixel, incorporating SPAD enable/disable, output buffer, and optional impedance passive quench. The circuit schematic is shown in Figure 122.

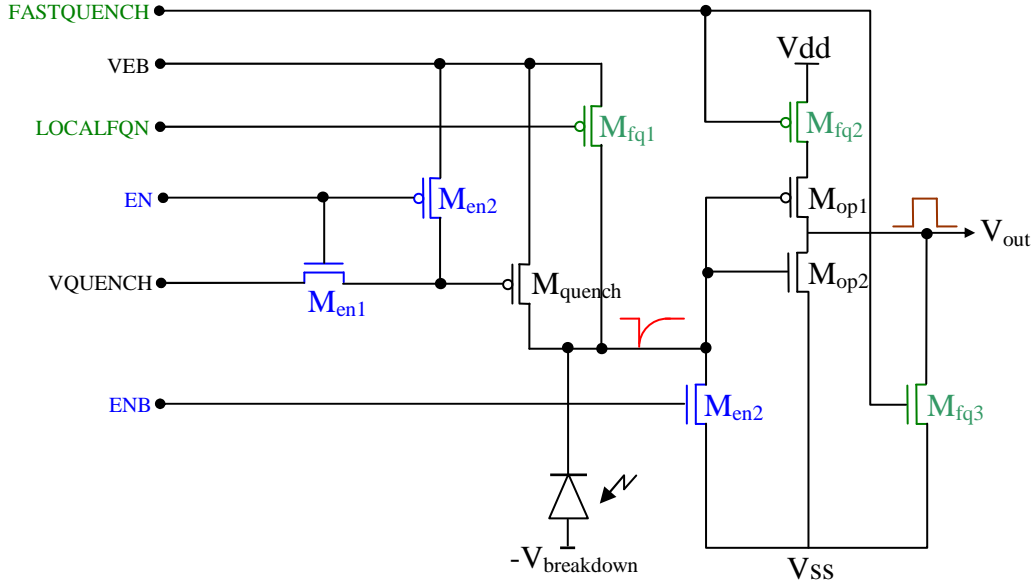


Figure 122: GRO TDC SPAD Quench Cell

The standard PMOS passive quench and output buffer components remain unchanged. The additional components added for SPAD enabling/disabling (i.e. permitting ‘time-gating’) are highlighted in blue, with low impedance ‘fast-quench’ components highlighted green. The fast-quench function allows the SPAD to be rapidly brought out of a disabled state, typically at the start of a timing frame.

5.6 Calibration/Compensation Scheme

A global calibration function is embedded within the TDC array design in order to track process, voltage and temperature (PVT) induced variations. The average oscillation speed of the array of GRO TDCs is controlled by a globally distributed reference calibration voltage which controls the NMOS regulation elements within the GROs. This reference signal is created by embedding an exact copy of the ring oscillator within a single, locally positioned phase locked loop (PLL) and then making the locked-in VCO control voltage available for chip-wide distribution. This block diagram for this feature is shown in Figure 123.

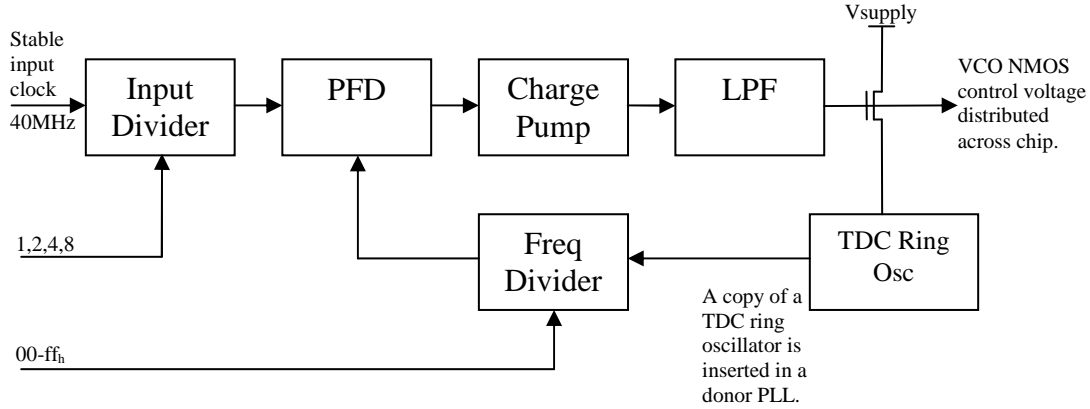


Figure 123: PLL Global Calibration Architecture

Once the PLL is locked to a stable external clock, the average time resolution of the GRO TDC array will track that of the donor PLL. Modulation of the GRO NMOS gate voltage also allows a means of time resolution control without adding area inefficient capacitance to slow the oscillator, as implemented by De Heyn et al in [75].

5.7 GRO TDC Pixel Layout

Considering the integration of the blocks presented which constitute the GRO TDC pixel, the key performance features of the metallisation layers available in the process are important to note. The unit capacitance, sheet resistance and spacing rules are summarised in Table 22 for the target IMG175 process used.

Layer	fF/ μm^2	m Ω/\square	Wid/Space (μm)
M1	0.076	235	0.12/0.12
M2	0.05	195	0.14/0.14
M3	0.037	92	0.14/0.14
Alucap	0.025	60	1.0/2.0

Table 22: Metal Routing Capacitance, Sheet Resistance and Layout Rules

This enables the appropriate allocation of metal layer which best suits the intended design floorplan. Table 23 summarises the choices made for this chip design. The lower capacitance and sheet resistance of metal 3 resulted in it being used for high speed signal distribution, with Alucap being used for low impedance power supply routing.

Metal	Direction	Guidelines
M1	H	Core lib supply routing.
M2	V	Data and counter buses.
M3	H	High speed clock tree signals.
Alucap	V	Power and ancillary routing.

Table 23: Metal Usage and Direction Rules

Applying these guidelines, the Logic, GRO, Coder, Ripple Counter and Memory block layouts were integrated tightly along with the SPAD and Quench to form the Pixel layout shown in Figure 124. Specific zones dedicated to control and data buses, as well as supply decoupling capacitance are highlighted.

The pixel's full list of input and output signals is provided in Table 32 in Appendix D.

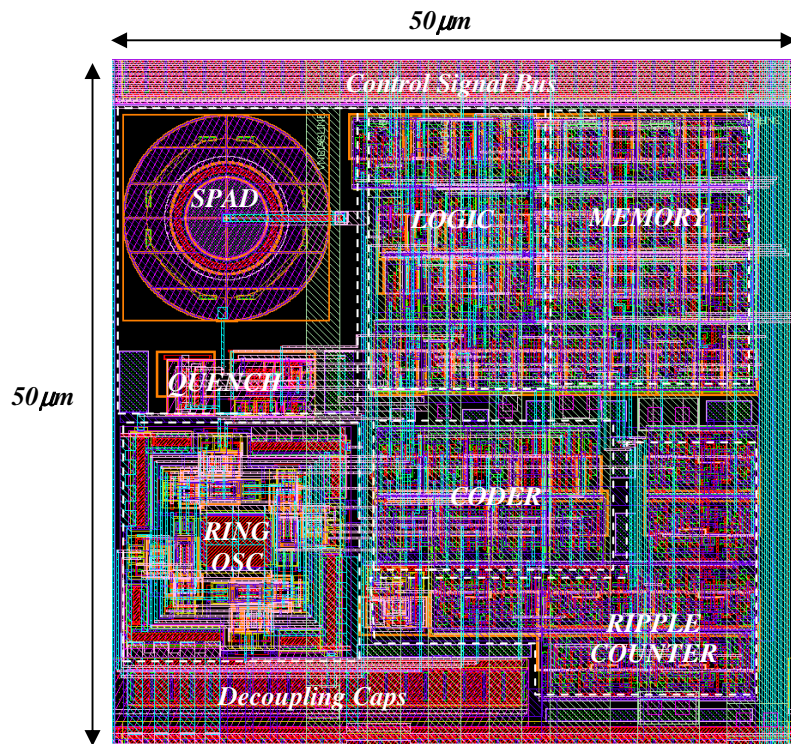


Figure 124: GRO TDC Pixel Layout

The final pixel layout pitch is 50x50µm. The ratio of the SPAD active area to the pixel area results in a low fill factor of ~2%. Due to the requirement to run power supply routes in upper level metals across the pixel array, CAVITY cannot be used.

Power routing plays an important role in ensuring that the array does not suffer from image bows or droops. Each pixel has a carefully designed lower metal digital core supply grid, with top 'AluCap' low resistivity metal routing responsible for routing all

supplies deep into the pixel array. This feature can be easily seen in the pixel micrograph plot of Figure 125.

It should be noted that to address the fill-factor issue, microlenses are required. This significant optical design topic forms a separate Megaframe Project work package, which does not form part of this Thesis. However, provision for optical concentrator alignment aids has been implemented as a component of the pixel array layout.

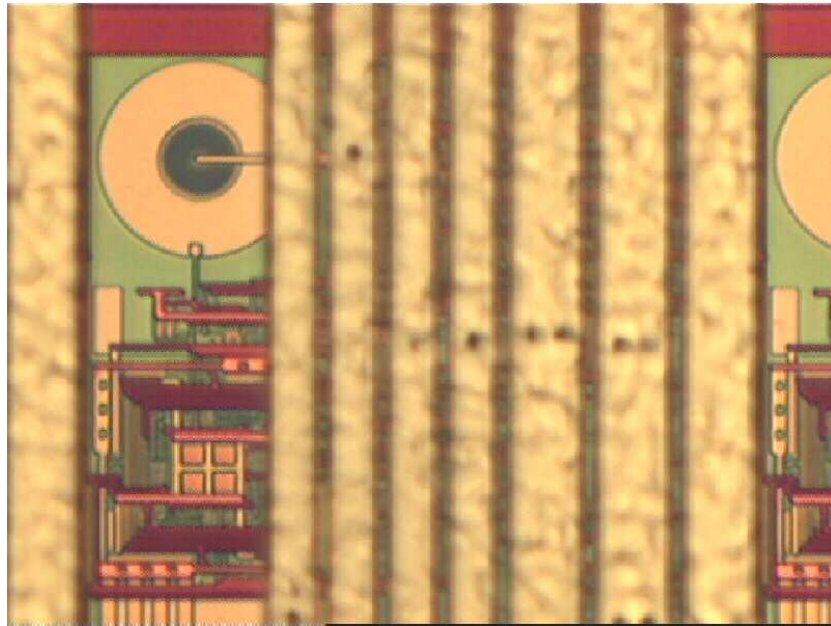


Figure 125: GRO TDC Pixel TDC Micrograph

Alucap metal power routing obscures most of the pixel, apart from the GRO, SPAD and Quench blocks. The challenge of distributing signals to and reading $>10\text{Gbps}$ data out from the 32×32 pixel array at $1\mu\text{s}$ intervals, as well as controlling the many mode and test signals is a significant one. An overview of the design of the related array support and readout cells is presented in the following sections.

5.8 GRO TDC Array Data Readout

5.8.1 Overview

The GRO TDC device is designed to reside within the TCSPC system shown below in Figure 126. The key parts of this system are a processor motherboard, controller PC, pulsed laser illuminator, the GRO TDC chip itself and its associated daughtercard. In this context a full frame of data is that which represents 32x32 pixels (1024) of 10b word width per pixel, generated at 1 μ s intervals. The GRO-TDC device therefore generates TCSPC data at >10Gbps.

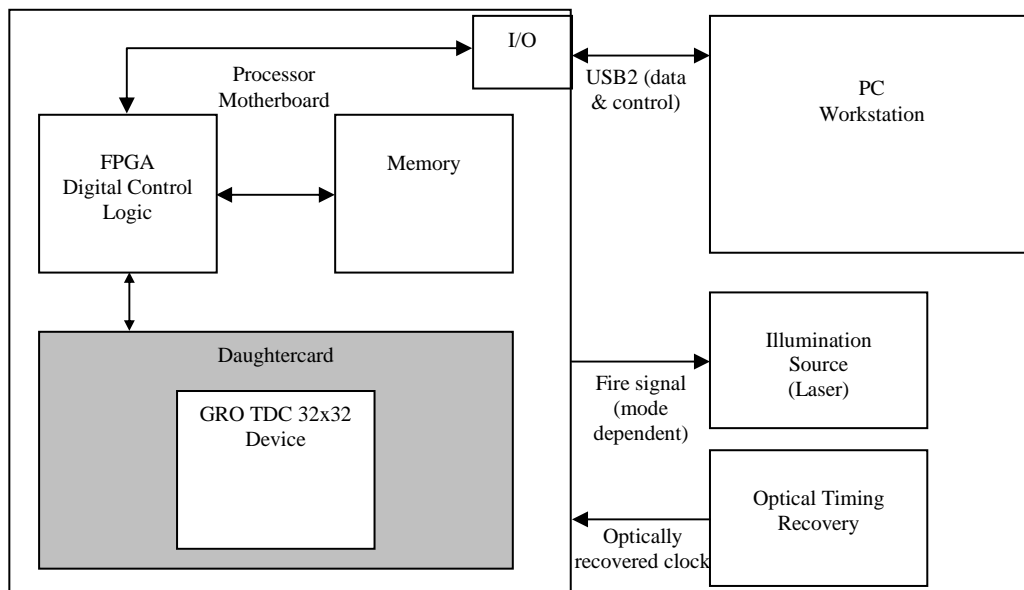


Figure 126: TCSPC System Block Diagram

The Processor motherboard provides the hardware support for controlling and powering the GRO TDC device. It also contains a memory store, enabling the option for recording a continuous sequence of frames generated at 1 μ s intervals which can be subsequently post processed. Alternatively the controller PC user may request a preview mode, where a reduced number of frames are displayed on screen in real time but at a greatly reduced USB2 rate. In this mode, the FPGA and PC engage in a data handshake sequence, with frames of data that were generated at full rate being picked from the data stream and displayed as fast as the USB link can support, typically about 100fps. It can be thus appreciated that in this mode most of the generated frames of data are dropped.

Within the GRO TDC device, the architecture shown in Figure 127 has been implemented. The array has been split into two 32x16 sub-arrays of pixels. Each 16 element column's data is serialised and read out in sequential, 'rolling blade' fashion through a row of single ended I/O pads at the top and bottom of the device. At full data rate, top and bottom data channels generate output data at ~5Gbps each. Control and timing signals are distributed across the pixel array with minimum skew via carefully balanced signal distribution trees.

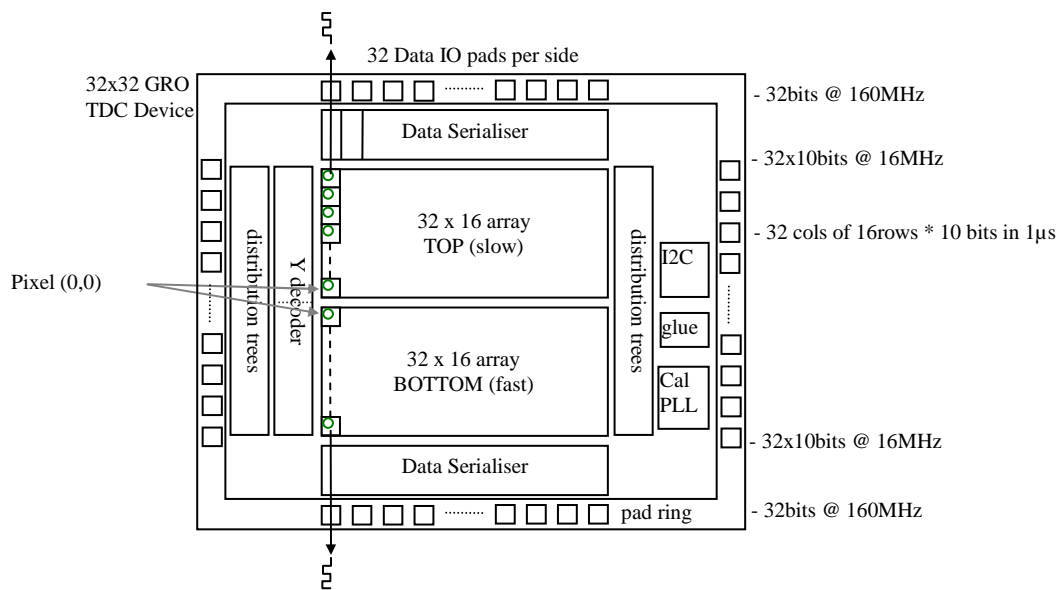


Figure 127: GRO TDC Pixel Array Data Readout Structure

Although the TCSPC data is generated with respect to the laser 'FIRE' global timing reference signal, the rolling blade serialised readout sequence is controlled by a token passing Y-decoder cell resident at the edge of each row. Each row and column has an individual enable signal, programmable via an I2C serial control bus register. An enabled row results in absorption of the Y-decoder token for the duration of a line (16MHz line clock period), and subsequent serialising of the output data from the pixel concerned. If a row is deselected, the readout token passes through the Y-decoder structure to the next enabled pixel. A deselected column results in a shutdown of both the associated column of pixels and serialiser block. Sub-sampling may therefore be performed by selecting alternate rows and columns, with a region of interest being selected by enabling a group of adjacent rows and columns. Both of these readout schemes are demonstrated in Figure 128.

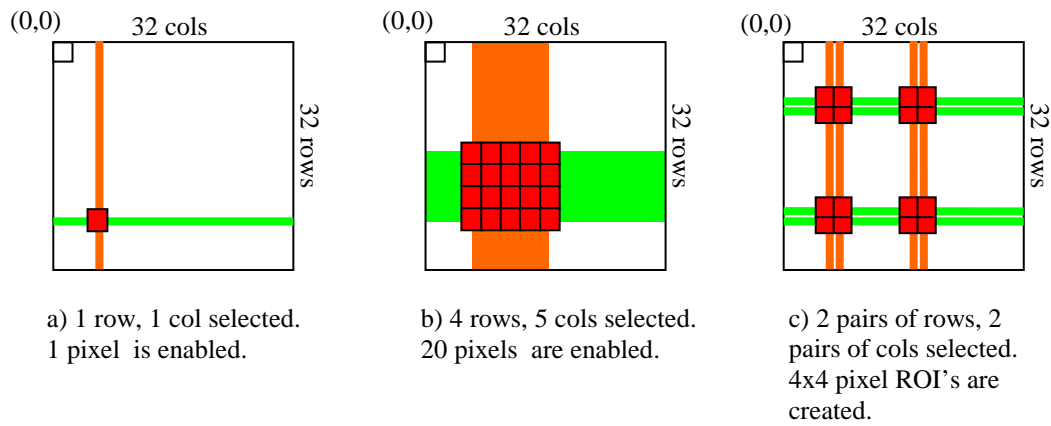


Figure 128: Region of Interest Selection

Region of interest mode is useful in FLIM where only the lifetime of a specific region of a cell or assay is of interest. Sub-sampling is a useful feature for FCS where it is the movement of a fluorophore marked cell type from one region of the image to another which is important.

Other device level blocks required in the device level floorplan are the Calibration scheme PLL and glue logic associated with the support of different system synchronisation modes. The I2C module is also shown, with the full register map listed in Appendix D, Table 33.

5.8.2 System Synchronisation and Device Clocking

A TCSPC system must be synchronous, ensuring that the elements shown in Figure 126 remain temporally locked together. This holds true for two potential scenarios. The first scenario occurs when the firmware resident within the processor motherboard fpga is responsible for driving both the laser illuminator and the timing reference signal for the GRO-TDC device. This is known as 'Master' mode, with the system clocks being generated via an fpga PLL which is locked to a stable local oscillator.

The alternative 'Slave' mode occurs when the laser illuminator is allowed to free run. The digital control sub-system is then locked to a clock signal which is optically recovered using a reference PMT or photodiode. This is commonplace in systems using larger, pumped laser systems which are not capable of responding to a trigger logic input.

These two modes are summarised in Table 24.

Mode	Defined Clock Source		
	FPGA PLL input (6-27MHz)	32x32 PLL input (6-27MHz)	TDC Stop Signal input
<i>Master:</i> (laser being driven by the system)	XTAL ref.	FPGA provides.	FPGA provides programmable delay stop pulse.
<i>Slave:</i> (system listening for the laser)	Pulse picker or photodiode	Photodiode with PLL divider enabled	Pulse picker or photodiode plus long cable, or FPGA provides programmable delay stop pulse.

Table 24: TDC Mode Synchronisation

The clocking and control infrastructure required to support the two synchronisation modes within the GRO-TDC device is shown in Figure 129 along with the interconnect to the motherboard FPGA control logic.

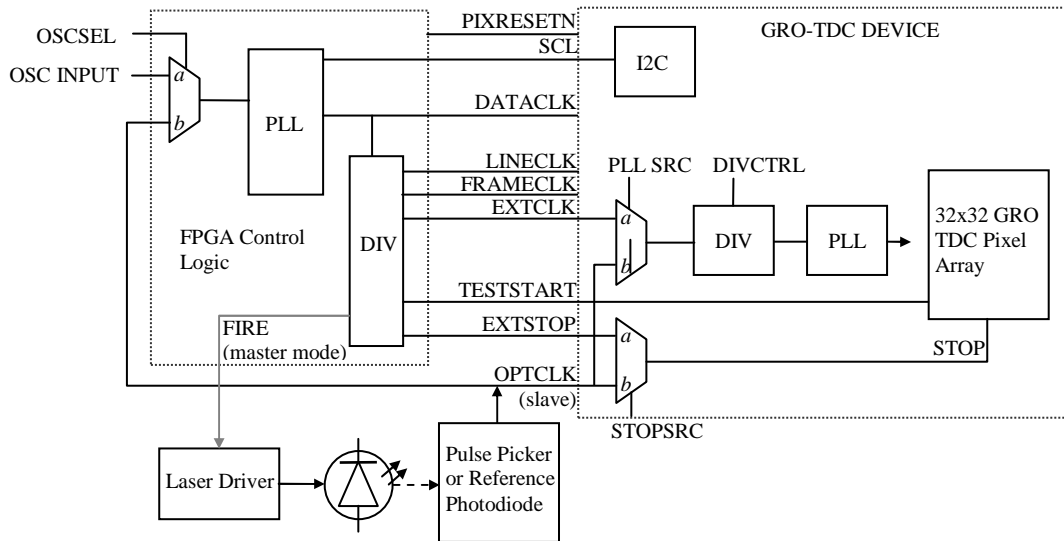


Figure 129: System Clocking and Synchronisation Diagram

The ‘OSCSEL’ multiplexer state configures whether the FPGA timing is synchronised to the stable oscillator (master mode) or the optically recovered clock (slave mode). Once the FPGA based PLL is locked to the chosen system clock, the generated GRO-TDC device control signals will be synchronous, albeit with a small, inherent jitter component.

where the image subject is most likely to be. This is advantageous for future event driven pixel readout schemes.

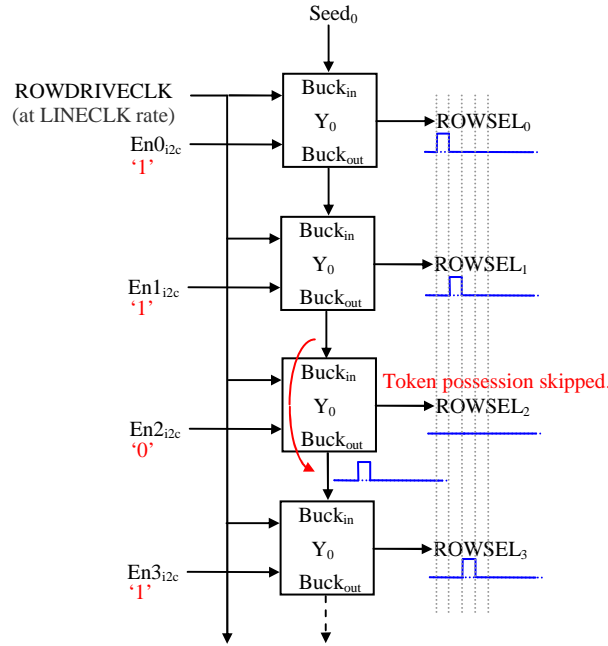


Figure 131: Row Y-Decoder Block Diagram

The token passing scheme permits each enabled row to take ownership of the vertically running parallel data bus for a single 16MHz line period, before the token is passed to the next valid row. During this time the TDC data is driven on to the bus, is serialised and then transmitted to the hardware sub-system. Ownership of the bus is dictated by a Y-Decoder ROWSEL_x output signal.

The simplified schematic shows four rows, row 2 being disabled. The timing diagram inset shows rows 0 and 1 are enabled for a single line period, row2 is skipped, with row 3 picking up immediate bus ownership. This timing scheme ensures that the data from a series of enabled rows is compressed into the shortest possible output data burst so that serialiser clocks may be disabled for power saving reasons.

5.9 Design for Test and Characterisation

The GRO-TDC device has several different operating modes which can be accessed for the purposes of test, debug or characterisation. Test mode access is common to all pixels, configured by the signals STARTSRC and SPADEN as shown in Table 25.

Mode	STARTSRC	SPADEN	Description
1	0	0	SPAD is disabled via gating switch ('gating' mode).
2	0	1	Normal operating mode (Photon Counting or TCSPC)
3	1	0	TESTSTART initiates a conversion, SPAD is disabled. This is 'external stimulus Digitiser test mode'.
4	1	1	This mode is detected by in-pixel logic and the raw SPAD digital output is routed to the data bus LSB. The data serialiser is disabled. This is 'Raw SPAD test mode'.

Table 25: Test Mode Access

Mode 1 allows for a conventional time gated photon counting mode, discussed by Becker in [30].

Mode 2 is the normal operating mode for the device, with either photon counting or TCSPC operation selected via the global control signal 'MODETCSPC'.

Mode 3 is the Digitiser test mode, where a series of test pulses may be injected to test artificial photon counting operation. Alternatively an artificial conversion start signal may be fed to the TDC to test TCSPC operation.

Mode 4 is the Raw SPAD test mode, where the response of a selectable, single row of SPADs may be routed directly out through the IO pad, bypassing the serialiser logic. This permits large scale SPAD characterisation and bias point setup.

5.10 Top Chip Floorplan

The floorplan of Figure 127 results in overall die dimensions of $4578\mu\text{m} \times 3754\mu\text{m}$, a total die area of 17.186mm^2 . The device is heavily pad-limited due to the overhead of having many serialised output channels both top and bottom. This was necessary to achieve the high frame rate specification of 1Mfps. The control signals required for chip operation, programmability and test, access the device on the left and right sides as defined in Table 32, Appendix D. A micrograph of the final chip is shown in Figure 132. The peripheral cells and power routing are obscured by dummy polygons, necessary to achieve the required layer density rules.

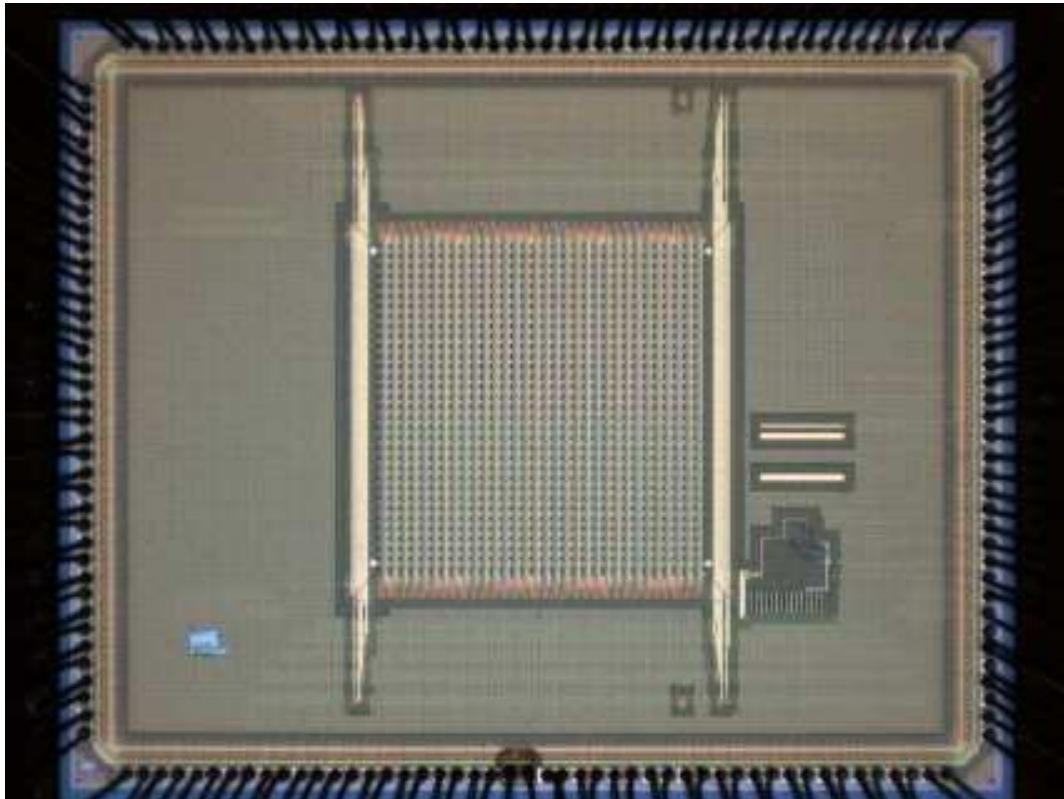


Figure 132: Top Chip Micrograph

5.11 Device Packaging

The non-backlapped ($\sim 750\mu\text{m}$ thick) device was bonded into a 180 pin, ceramic pin grid array (CPGA) package, used for its high IO count, wire bonding simplicity, good thermal properties and ease of use with zero-insertion-force sockets. An example is shown below in Figure 133.

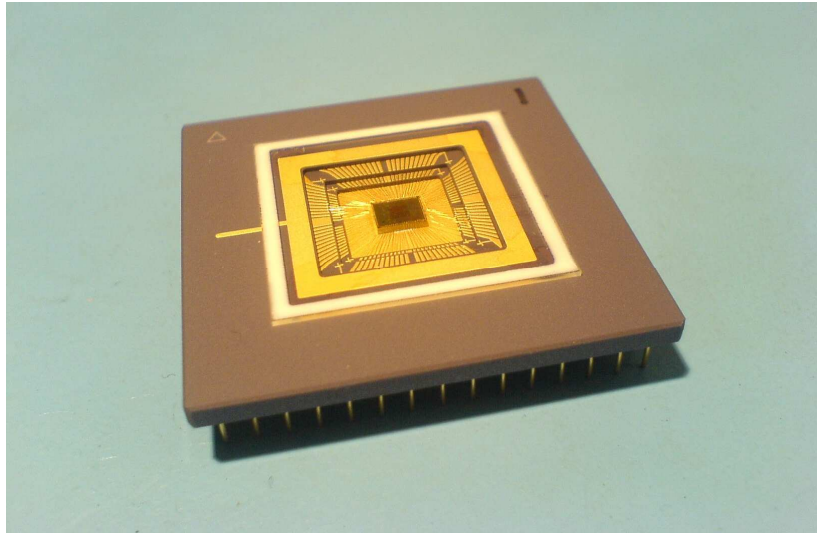


Figure 133: 32x32 GRO TDC Pixel Array in 180pin CPGA Package

The large cavity optical package used was fitted with a clear cover glass. Gold was used throughout for the 172 bond wires.

5.12 Hardware Platform

The hardware support platform consists of a processor motherboard (supplied by EPFL), a daughtercard to interface the GRO-TDC device to an FPGA, and PC running a graphical user interface. Essentially these elements are the physical embodiment of the functions represented in the block diagram of Figure 126. The assembled hardware platform is shown in Figure 134.

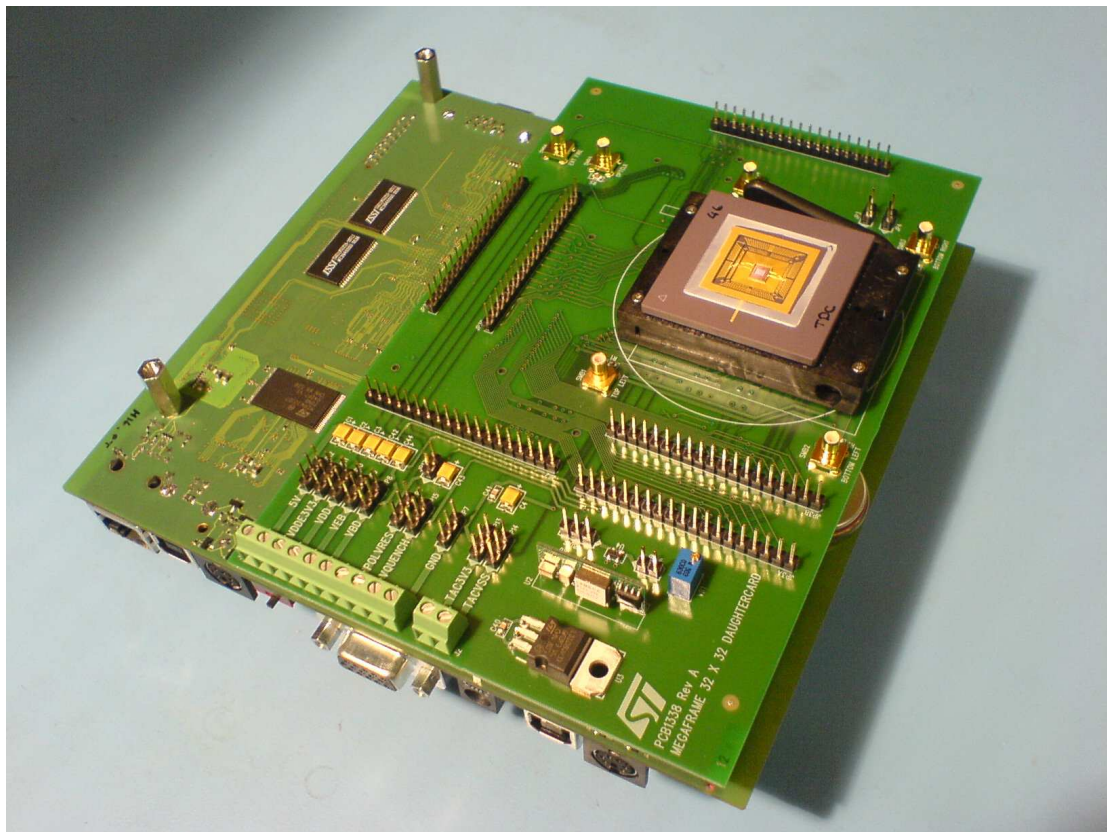


Figure 134: GRO TDC Array Hardware Platform

5.13 Conclusions

In this chapter a target specification for an array of TDCs implemented in a deep sub-micron technology was determined. The chosen gated-ring-oscillator (GRO) architecture was introduced along with a suggested timing scheme for 1Mfps TCSPC operation. A detailed design description was then provided for each of the TDC functional blocks, along with critical performance criteria, and how the blocks fit together with SPAD and Quench cells in a final pixel layout.

In addition, an overview of the necessary data readout support cells was provided, as well as an account of the integrated design for test features. Finally, the topics of chip floorplan and packaging were introduced, and a summary given of how the device fits into the overall hardware support system.

The resulting GRO based 32x32 TDC array design balances area, power consumption and performance constraints to create a scalable architecture that is designed to be power efficient and robust to process, supply and environmental effects.

The expected results are summarised in Table 26.

Metric	Value (slow, fast)		Unit
Technology	130nm, 4M-1P CMOS		
TDC Word Length	10		bits
Array Uniformity (σ)	<3%		LSB
TDC Jitter (mean)	<1		LSB
TDC Area	2200		μm^2
TDC DNL/INL	± 0.5	± 1	LSB
Resolution	50-150		ps
TDC Power Consumption	<50		μW
Chip Data Rate	~10		Gbps
Array Frame Rate	1M		fps

Table 26: GRO TDC Array Expected Results

The GRO TDC array was characterised using the hardware platform shown in section 5.12 and software graphical user interface resident on the controller PC. The results obtained are presented in the following chapter.

6 TDC Characterisation

6.1 Introduction

In this section the silicon characterisation results for the GRO-TDC 32x32 array outlined in chapter 5 are presented. As previously mentioned, the TDC array consists of two halves; the lower half being a TDC with finer time resolution and increased power consumption. For each of the two TDCs the key metrics of time resolution, dynamic range, accuracy, precision, uniformity and power consumption are analysed. The behaviour of the array when operating locked within the integrated PVT calibration loop is also presented, along with a review of power supply noise rejection performance. Summary comparison tables are then shown and conclusions drawn. Measurements are performed at room temperature with typical supply voltage magnitudes unless otherwise stated, the Megaframe host PC GUI being used as the primary method for data gathering. Aspects of characterisation procedures are highlighted within the appropriate section where particular attention or clarification is necessary.

6.2 TDC Characterisation Results

The two main operating modes of the GRO TDC Pixel are Photon Counting mode (PC) and Time Correlated Single Photon Counting (TCSPC). In PC mode the pixel is required to count the number of primary events within a defined time period, or exposure time. The primary events can come from either a test input pulse stream (see section 5.9 for test mode entry procedure) or from the SPAD itself.

In the TCSPC mode the TDC is fully operational. The pixel outputs a digital value representing the time difference between a test input or SPAD event and the system STOP datum signal. The STOP signal is normally associated with the firing of an excitation illumination source.

In this chapter the characterisation results for the two operating modes are presented separately.

6.2.1 Photon Counting Mode

6.2.1.1 Dynamic Range and Accuracy

In photon counting mode, test or SPAD pulses are fed directly to the enabled pixel's ripple counter. For characterisation, test pulses are used. The ripple counter is 7 bits wide, and therefore the dynamic range expected in this mode was $0-2^7$, i.e. 128 counts per frame period ($2\mu\text{s}$ in this case). This was confirmed by analysis of a sweep of input test pulses over the whole 32×32 array. The transfer function for a single pixel is shown in Figure 135.

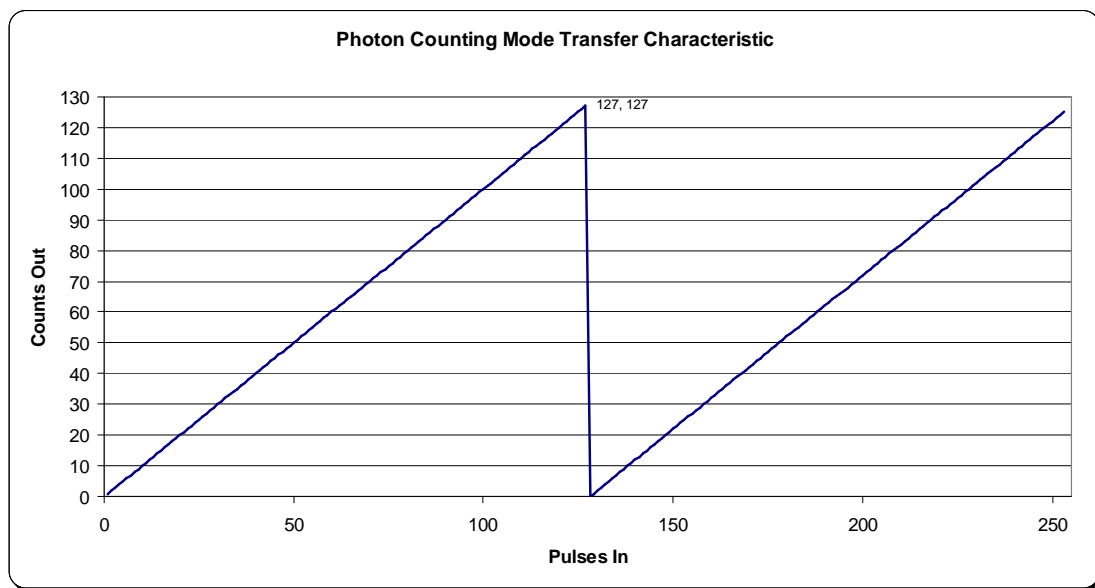


Figure 135: Photon Counting Mode Transfer Function Profile

This confirmed that the dynamic range was the expected 128 codes or 7 bits, with no missing codes. It can be seen that in the event of more than 128 input pulses in a certain integration period the counter overflowed back to zero. The transfer profile demonstrated that there was no test mode offset or gain error.

6.2.1.2 Array Uniformity

For this test every pixel in the array was measured at 90% full dynamic range, i.e. 114 input test pulses per integration period. The data for a single shot image was captured using the Megaframe GUI and is shown in Figure 136.

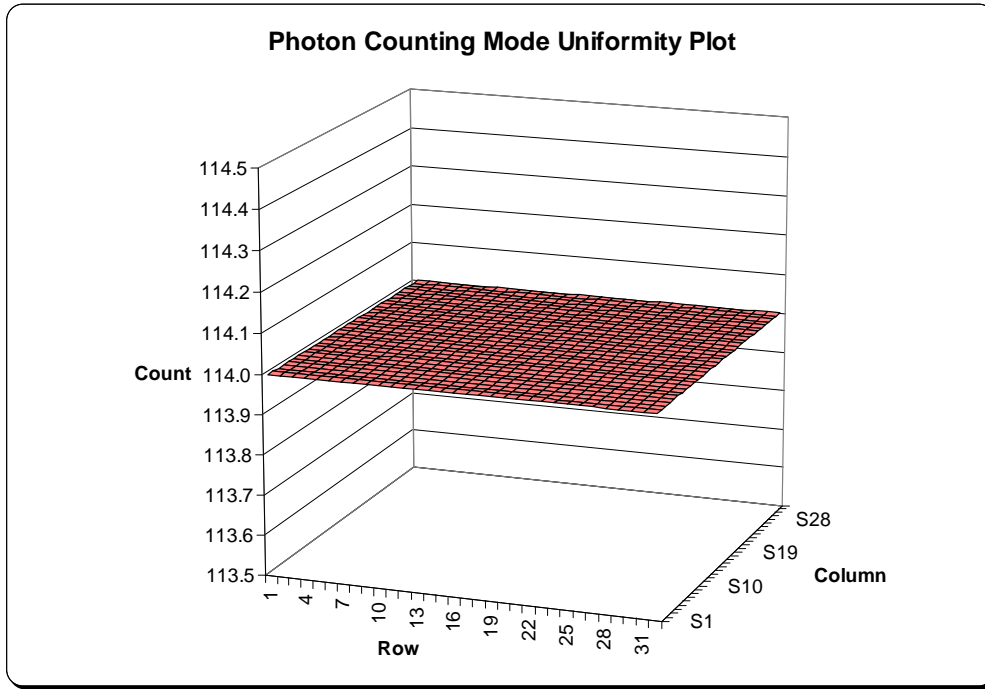


Figure 136: Photon Counting Mode Array Uniformity Plot

This result demonstrates that the array uniformity in photon counting mode exhibits zero error, with all 1024 pixels reporting the correct output code.

6.2.1.3 Power Consumption and Line Rejection

The core power supply consumption and rejection of noise is an important aspect of a TDC-pixel's performance, particularly with respect to array implementations as discussed in section 2.5.6. For the purposes of photon counting mode characterisation, the power consumption was measured for an input test pulse count rate of 10MHz. This translated to 20 test pulses per 2 μ s frame time, i.e. a frame rate of 500kfps. The graph shown in Figure 137 was constructed by progressively increasing the number of enabled rows of pixels. In this configuration the full array device consumes ~23mW in the digital core, 21.6mW of which is required by the ancillary cells and signal distribution trees (i.e. not in the pixels themselves). This corresponds to a single pixel core consumption of only ~1.6 μ W for this defined test pulse count magnitude.

However, it was important to note that 33.6mW was also drawn on the separate supply used by the data output pads, running at a nominal 2.8V.

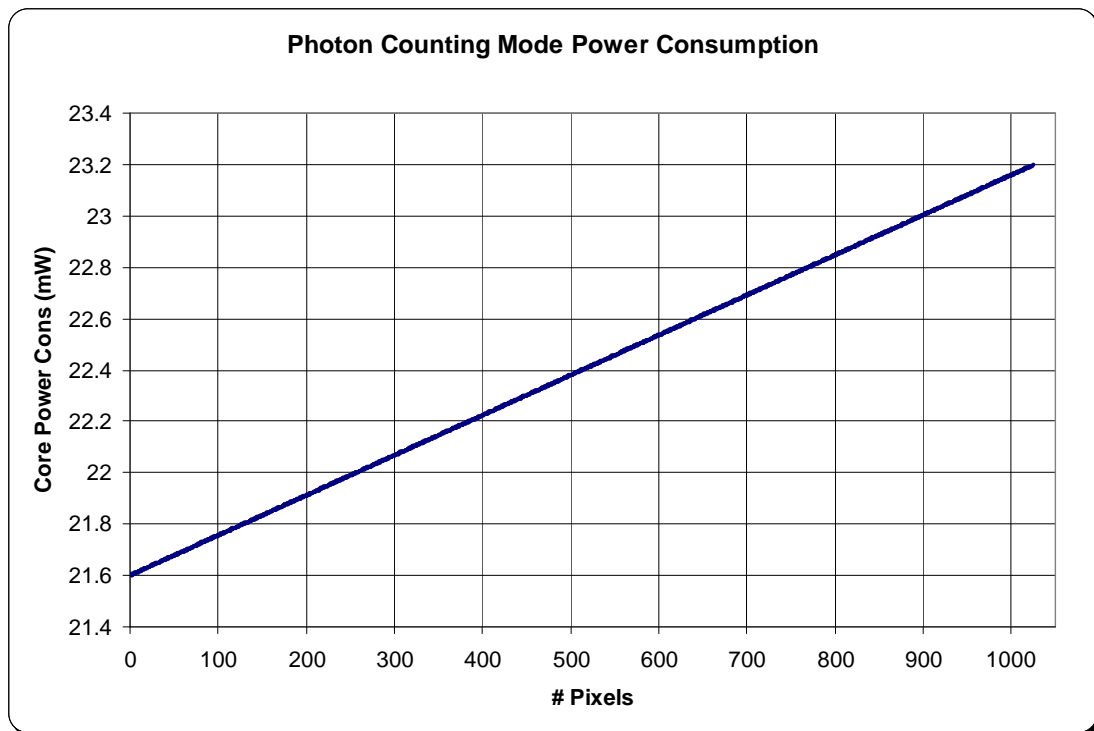


Figure 137: Photon Counting Mode Power Consumption Profile

Confirming the device's robustness to supply variation, the full array was configured to count a 90% dynamic range count sequence. The core supply was swept from 1.05 - 1.35 V in 50mV increments with no observed change in any pixel's output code.

6.2.2 Time Correlated Single Photon Counting Mode

6.2.2.1 Time Resolution and Dynamic Range

The magnitude of the TDC time resolution is one of the most important performance metrics, previously discussed in section 2.5.1. This sets the smallest time period which can be resolved within the TCSPC system. As mentioned in section 5.4.2 two different ring oscillators are used in the top and bottom halves of the array. These are intended to investigate the trade off of speed versus power consumption, and are referred to as SLOW and FAST.

The time resolution was determined by feeding in the TESTSTART and EXTSTOP stimuli, generated via FPGA firmware hosted on the hardware platform (FPGA RTL code written by Richard Walker). The average result obtained for a single pixel is quoted in Table 27. Dynamic range is computed by multiplying the average resolution by 2^N , N being the number of bits of the TDC output word.

<i>GRO-TDC Type</i>	<i>Resolution (ps)</i>	<i>Dynamic Range (ns)</i>
SLOW	178	182.3
FAST	52	53.2

Table 27: GRO-TDC Time Resolution

The time delta between the input stimuli START-STOP signals was swept enabling the TDC transfer profiles to be constructed, shown in Figure 138 and Figure 139 for SLOW and FAST TDC designs respectively. Note that both TDC designs roll over back to code zero when the dynamic range limit is exceeded.

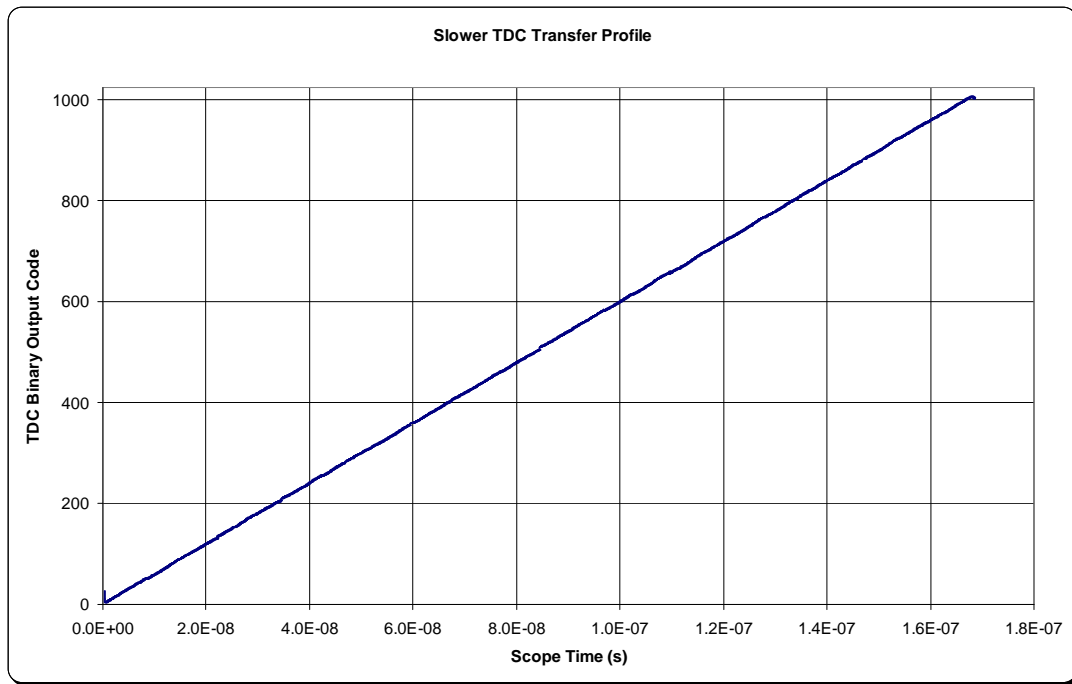


Figure 138: Slow TDC Transfer Profile

Note that since the slower TDC has a larger time resolution it has a correspondingly larger dynamic range. The theoretical maximum dynamic range is not quite reached during these particular data acquisition cycles due to the jitter contribution of the FPGA generated test stimuli waveforms. However, the TDC transfer function profiles are demonstrated over 98% of the full dynamic range.

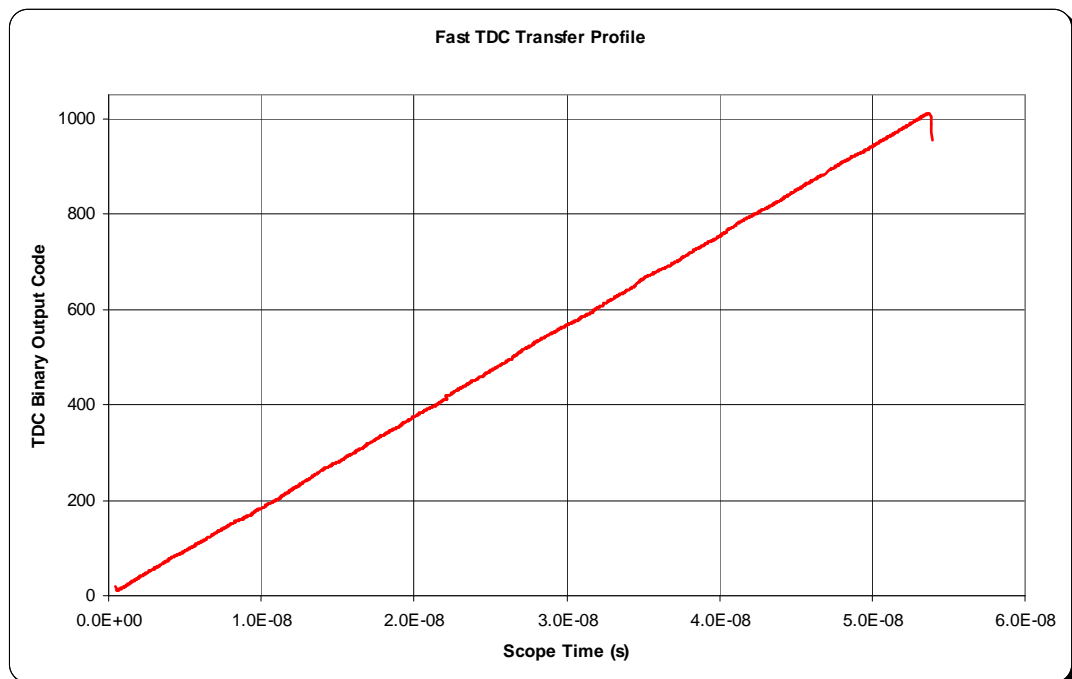


Figure 139: Fast TDC Transfer Profile

6.2.2.2 Accuracy

The accuracy of the TDC, i.e. the measure of how correct the conversion of a defined time interval with respect to the true value is described in terms of integral non-linearity (INL) and differential non-linearity (DNL), previously discussed in section 2.5.3. For this phase of the characterisation process, the white noise stimulation method reported by Doernberg et al in [59] for ADC testing purposes was employed. This method was particularly applicable to the testing of the GRO-TDC implemented as part of this research due to the fact that it had an embedded SPAD. By applying an uncorrelated input photon source to the SPAD, the same number of code occurrences (counts) could be expected for each accumulated TDC output code. Any variance from the mean count level indicates a non-linearity as illustrated in Figure 140. The uncorrelated excitation source used for this part of the characterisation was simply ambient light applied through a diffuser (since SPAD dark count rate was prohibitively low).

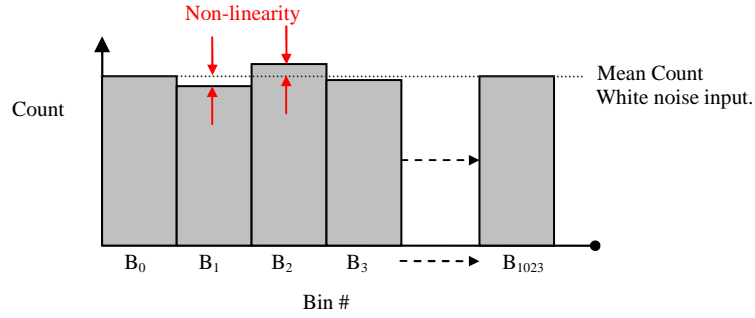


Figure 140: Code Occurrence DNL Computation Method

Once an adequate number of samples was obtained, the per-code differential non-linearity (DNL) was computed using:

$$\Rightarrow DNL_n = \frac{Count_i / \sum_0^{2^n} Count}{P_i} - 1$$

where P_i is the ideal bin width as a fraction of the full scale code range:

$$P_i = 1/2^n$$

For a truly monotonic converter, i.e. with no missing codes, the DNL should not exceed ± 0.5 LSB.

The Integral Non-Linearity is the deviation of the TDC actual transfer function from the ideal transfer function. The TDC actual transfer function can be constructed by successively accumulating the bin widths previously computed above. Subtracting this from a straight-line best-fit (ideal) profile reveals the INL.

The results obtained using this test method are as follows.

6.2.2.3 DNL

The DNL profiles for SLOW (blue) and FAST (red) GRO-TDCs are shown in Figure 141. The figure shows that viewed over the full code range the SLOW TDC has a peak DNL of +0.50, -0.40 LSB, and the FAST TDC +0.40, -0.27 LSB. Both TDCs are within the ± 0.5 LSB target.

The smaller dynamic range of the FAST TDC is also clearly evident.

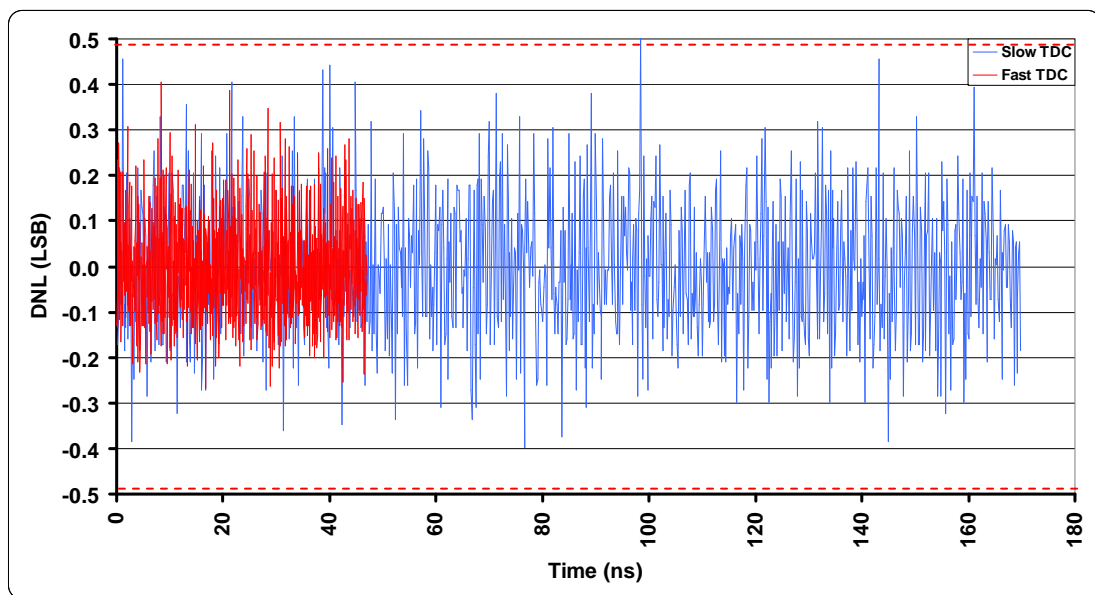


Figure 141: GRO-TDC DNL

6.2.2.4 INL

The INL profiles for SLOW (blue) and FAST (red) GRO-TDCs are shown in Figure 142, expressed over the dynamic range of the converter. The figure shows that viewed over the full code range the SLOW TDC has a peak INL of +1.83, -2.36 LSB, and the FAST TDC +1.24, -1.43 LSB.

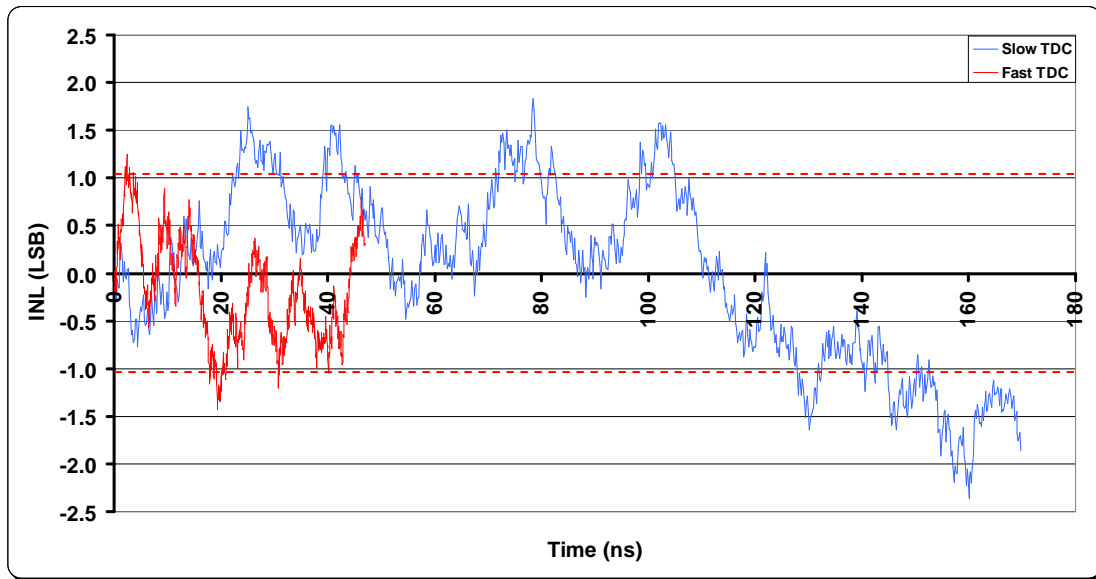


Figure 142: GRO-TDC INL

The INL profiles indicate that there is no obvious systematic offset or gain error, although the peak results fall slightly outside of the target specification for the slower TDC. This could be due to the fact that the extended dynamic range of the slower TDC exposes it to a longer duration of accumulated jitter and noise disturbances.

6.2.2.5 Code Probability

In response to the uncorrelated input test source used for the linearity measurement it was expected that each TDC output code would exhibit the same likelihood of occurring. A magnified view of the normalised code occurrence is shown in Figure 143.

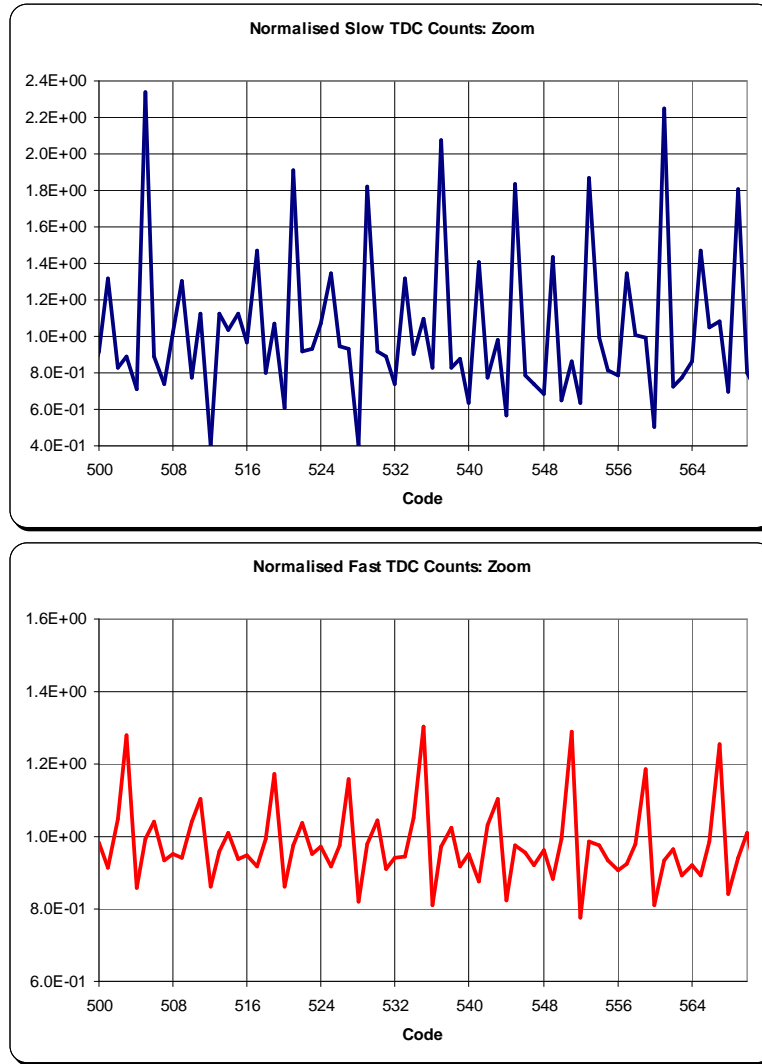


Figure 143: GRO-TDC Code Probability

The data reveals a slightly increased code probability every 8 codes. This is the beat rate of both ring oscillator designs (quad differential stages). It is thought that this ‘code pushing’ is due to the absence of a schmitt-type hysteresis stage at the input to the first element of the ripple counter. This issue has been addressed in a subsequent design.

6.2.2.6 Precision

As discussed previously in section 2.5.4 the precision of a TDC is a measure of the repeatability of a time to digital conversion result. This is particularly relevant when considering structures based on gated ring oscillators which could be susceptible to accumulated jitter effects.

The precision of the TDC was analysed by performing a correlation analysis of three adjacent GRO-TDC pixels constantly converting a time period of 85% of dynamic range. The correlation data obtained is shown in Figure 144. Solid lines shown are Gaussian fits; dashed lines represent the actual data obtained.

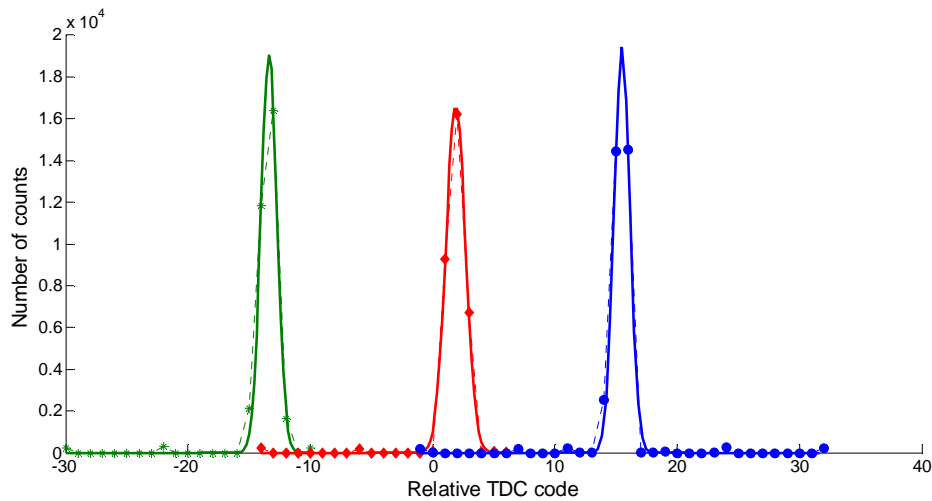


Figure 144: GRO-TDC Output Correlation Analysis

This shows the relative conversion differences between 3 SLOW TDCs, although an equivalent data set was gathered for the FAST TDC. The difference between the peak counts is effectively the gross TDC uniformity mismatch (i.e. ~ 15 codes in ~ 850 , or approximately 1.7%).

The data of interest however for this analysis is how these distributions vary with respect to each other. Therefore, the above computation was repeated for every pixel site in the 32×32 array, allowing a plot of the per pixel standard deviation jitter, shown in Figure 145.

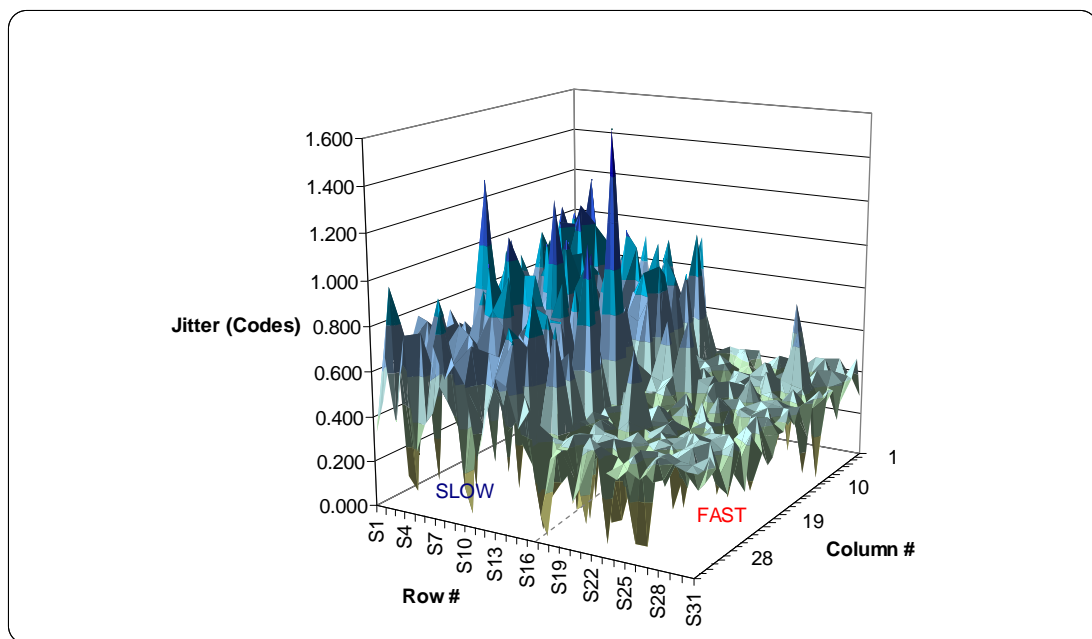


Figure 145: GRO-TDC Jitter Analysis

This shows that the accumulated jitter error is mostly <1 LSB, with the SLOW TDC performing slightly worse than FAST. This difference is most likely due to the FAST TDC accumulating jitter for less accumulated time. Separating SLOW from FAST data, the mean variations for the two halves can be extracted, shown in Figure 146 and Figure 147 respectively.

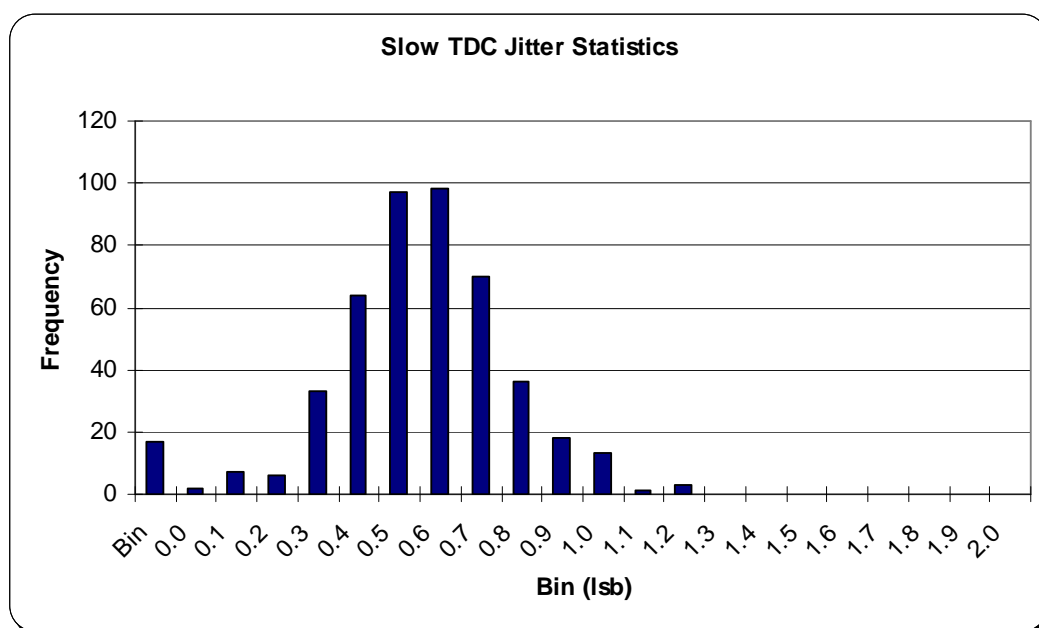


Figure 146: Slow GRO-TDC Jitter Statistics

The mean SLOW TDC standard deviation jitter is 0.602 codes, or ~100ps. The mean FAST TDC standard deviation jitter is 0.361 codes, or ~20ps, with a modal peak of 0.6 codes (~31ps), shown in Figure 147. The FAST TDC exhibits an unexpected distribution characteristic.

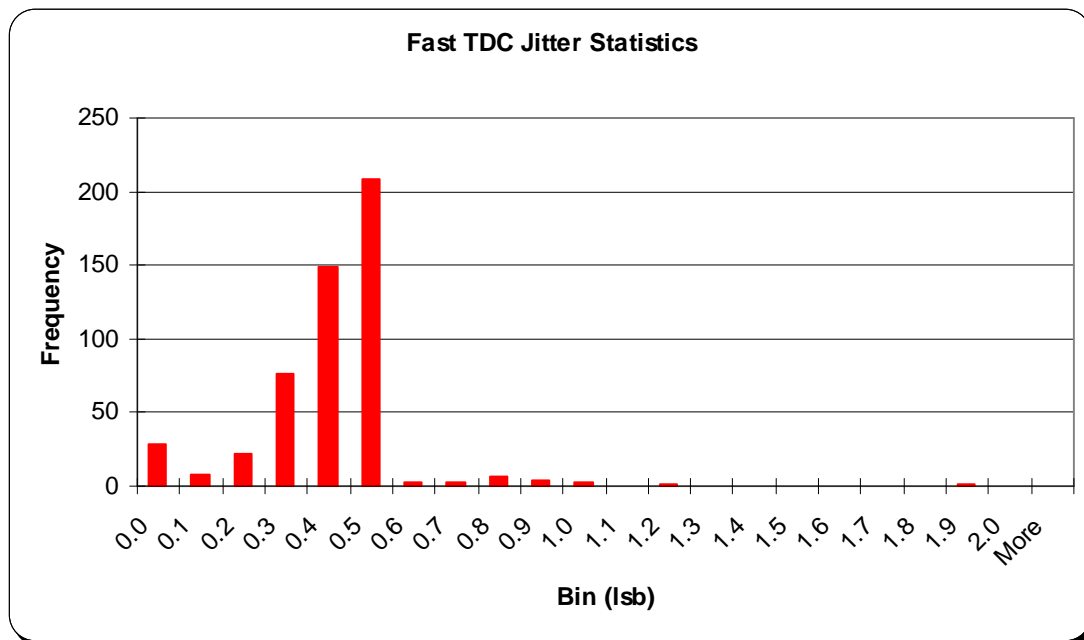


Figure 147: Fast GRO-TDC Jitter Statistics

Since this data is gathered at the upper end of the converter dynamic range, it is assumed that since jitter is a time-accumulated phenomenon it is less than the reported figure for conversion periods less than the reported figure obtained at 85% of full dynamic range.

This analysis confirms that the underlying conversion variations for both TDCs are within the 1 LSB target, indicating measurement results are repeatable for the same consistent input conditions.

6.2.2.7 Array Uniformity

The GRO-TDC array uniformity was assessed by analysing the mean result of time-digital conversion across the array for a constant time interval equivalent to ~90% of the full dynamic range. The average output code for 4096 conversion samples is shown in Figure 148.

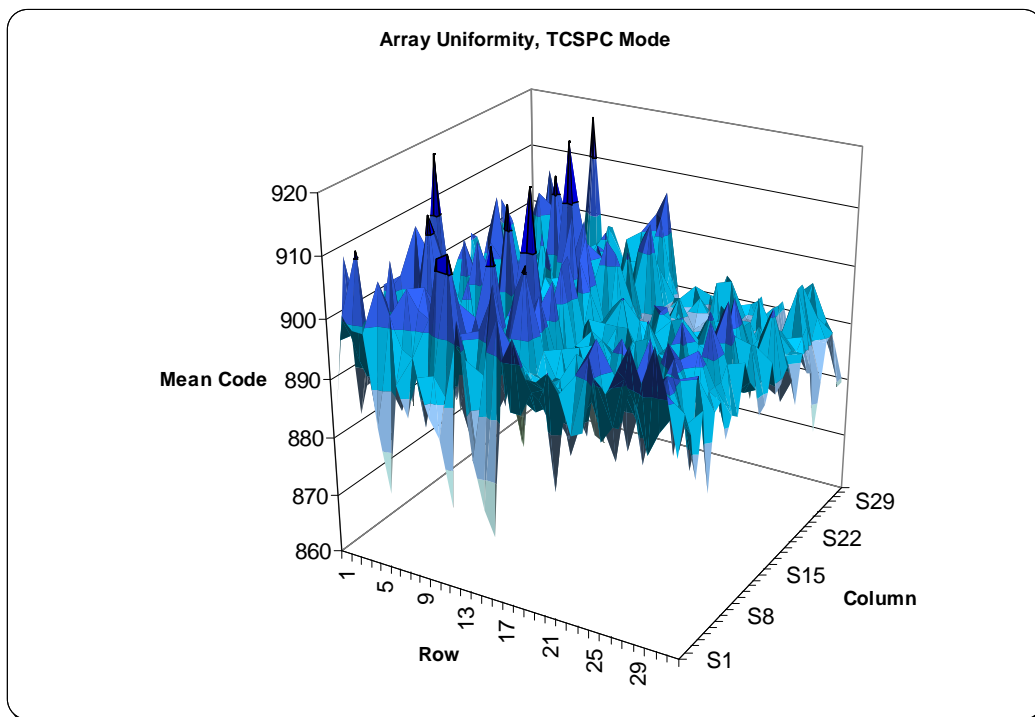


Figure 148: TCSPC Mode Array Mean Uniformity

The acquisition of this data permits the statistical analysis of each half array's data. The SLOW GRO-TDC output data is shown in the histogram of Figure 149.

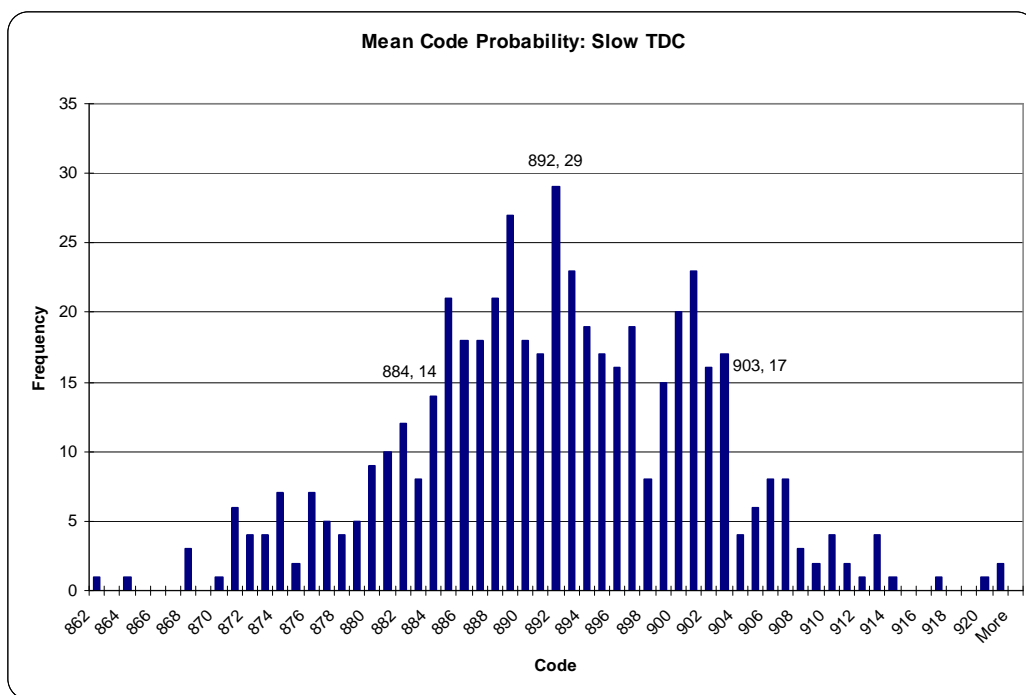


Figure 149: Mean Code Probability: Slow TDC

This gives a FWHM code span of 20 codes, or $20/892 = 2.24\%$.
 The FAST GRO-TDC output data is shown in the histogram of Figure 150.

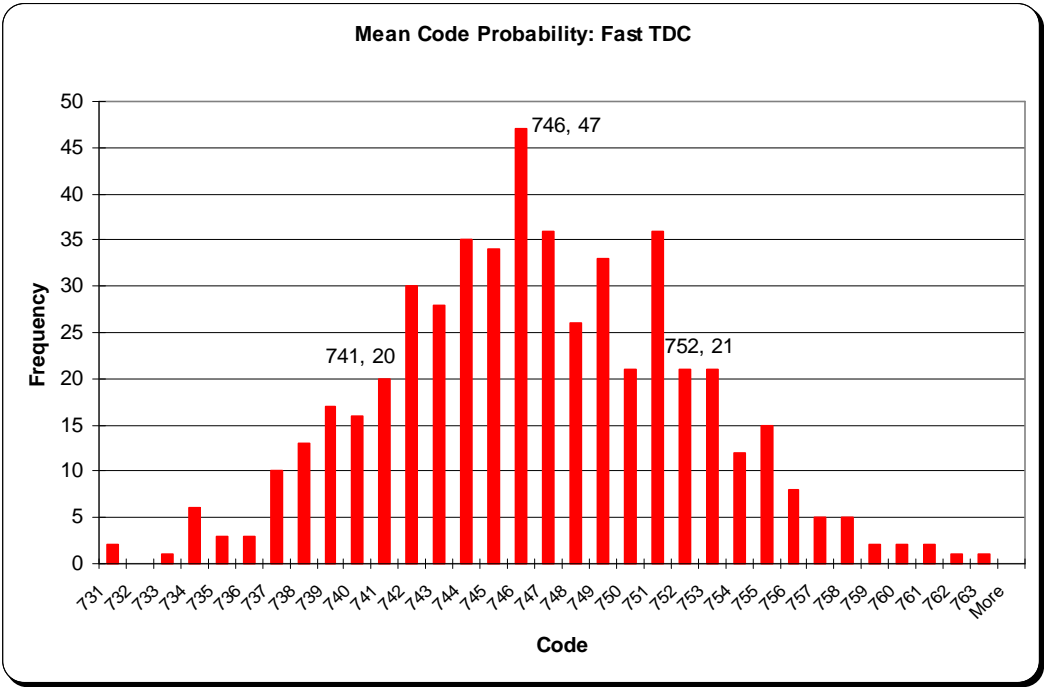


Figure 150: Mean Code Probability: Fast TDC

This gives a FWHM code span of 11 codes, or $11/746 = 1.47\%$. Both TDC designs exhibit array uniformity within the 3% target specification.
 From this data the row signatures can be extracted to highlight if there are any inherent vertical droops within the array.

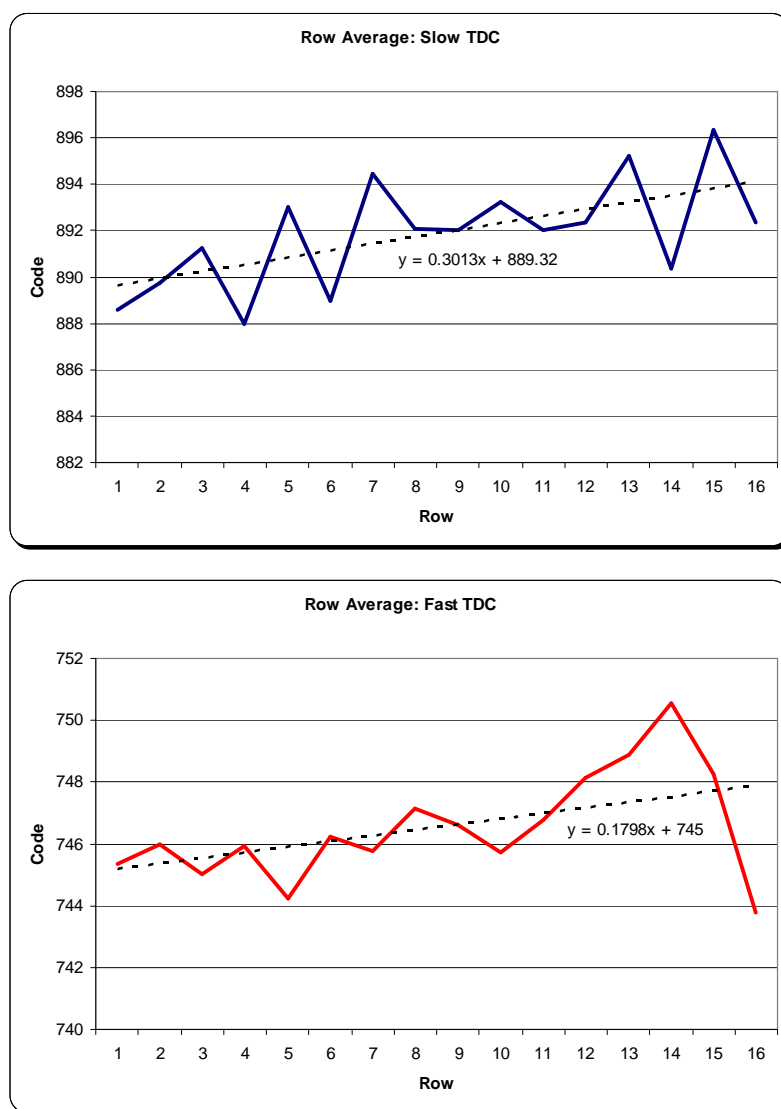


Figure 151: Row (Vertical) Signatures

This shows a small top to bottom droop trend of 4.8 codes for the SLOW GRO-TDC (0.3 codes/row), with 2.9 codes for the FAST TDC (0.18 codes/row). This could be due to power supply voltage drops, as the supplies are routed vertically within the array. Similarly, the horizontal signatures were extracted, shown in Figure 152 for SLOW and FAST TDCs.

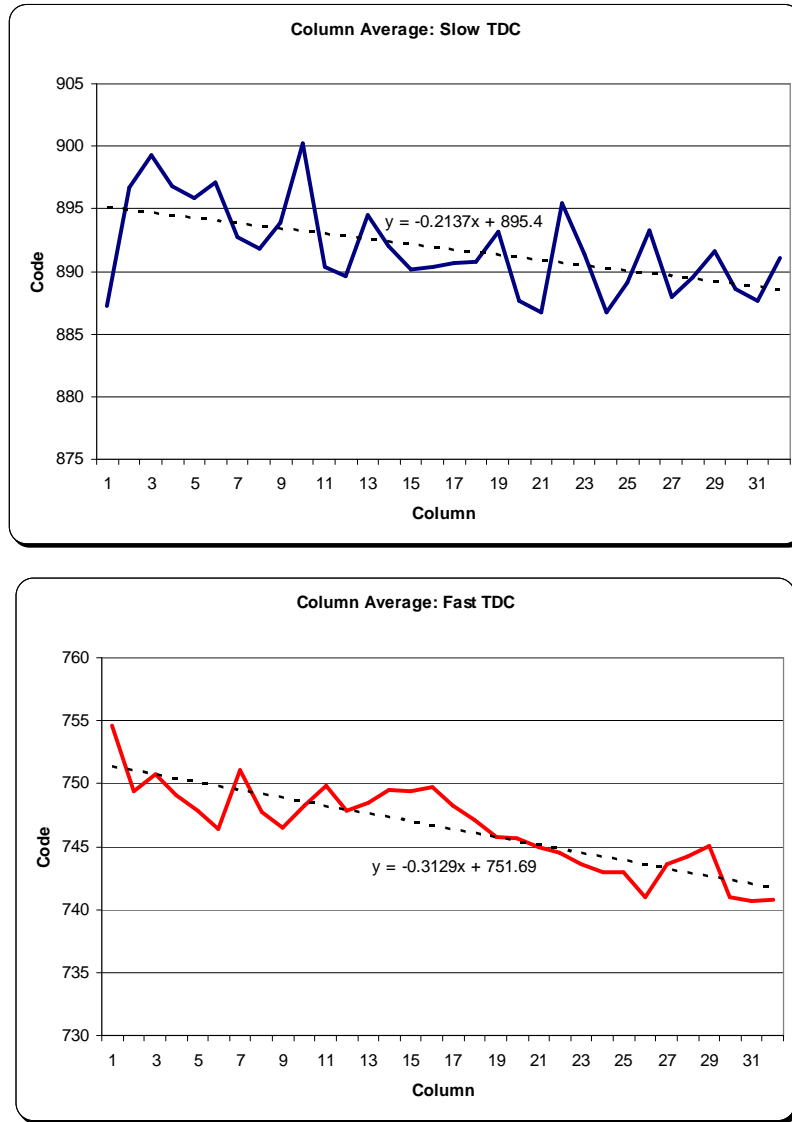


Figure 152: Column (Horizontal) Signatures

This shows a small right to left droop trend for both TDC designs: 6.84 codes for the slow TDC (0.21 codes/column) and 10 codes (0.31 codes/column) for the fast TDC. This is most likely due to diverging delays between key timing signals as they propagate across the array with differing per-signal capacitive loads between the two TDC designs. This data demonstrates subtle differences of TDC resolution which cause an accumulated array-wide variance of dynamic range. In ‘standard’ CMOS imagers such as those for the mobile market, an acceptable level of non-uniformity is deemed 1-2%. The quoted uniformity figures above are within this target whilst containing the small reported row and column gradient effects.

6.2.2.8 Calibration/Compensation and Line Rejection

The role of the embedded calibration/compensation function described in section 5.6 is to enhance the devices robustness to process, voltage and temperature variations (PVT). This function relies on the global distribution of a VCO control voltage, generated by an on-chip PLL whose ring oscillator is an instantiation of exactly the same cell used in the SLOW TDC. Once the PLL is locked to a stable external clock frequency, any variation in the oscillator output caused by PVT should be compensated for by the PLL control loop, and the adjustment made to the TDC array by the appropriate change in VCO control voltage transmission.

The function was firstly enabled and characterisation began with plotting of the GRO-TDC array resolution versus the computed PLL oscillator resolution as shown in Figure 153. This figure shows how the TDC elements in the array track the ring oscillator embedded in the control PLL within the range of loop frequencies that the PLL remains locked.

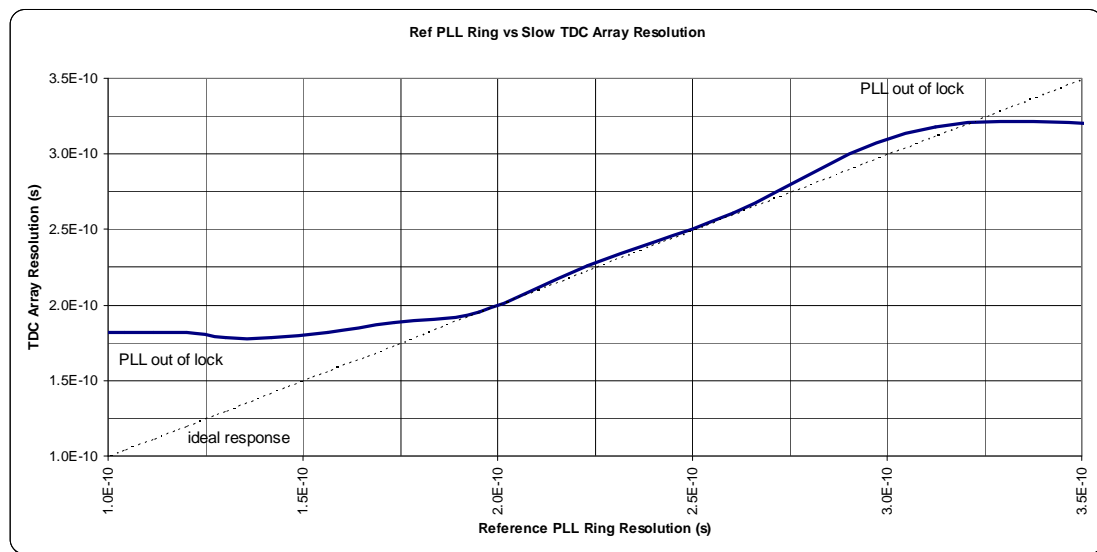


Figure 153: Tracking of TDC Resolution to the Control PLL

The graph demonstrates that PVT compensation is possible between the points where the PLL is configured with 175ps to 325ps resolution.

The sensitivity of the array resolution to core voltage variation was then plotted with resolution configured at the mid point of the PLL-lock range, i.e. ~240ps. The resulting graph is shown in Figure 154.

Traces which show the same response to supply variation without the PLL control loop are also plotted to show the impact on time resolution without having a compensation mechanism.

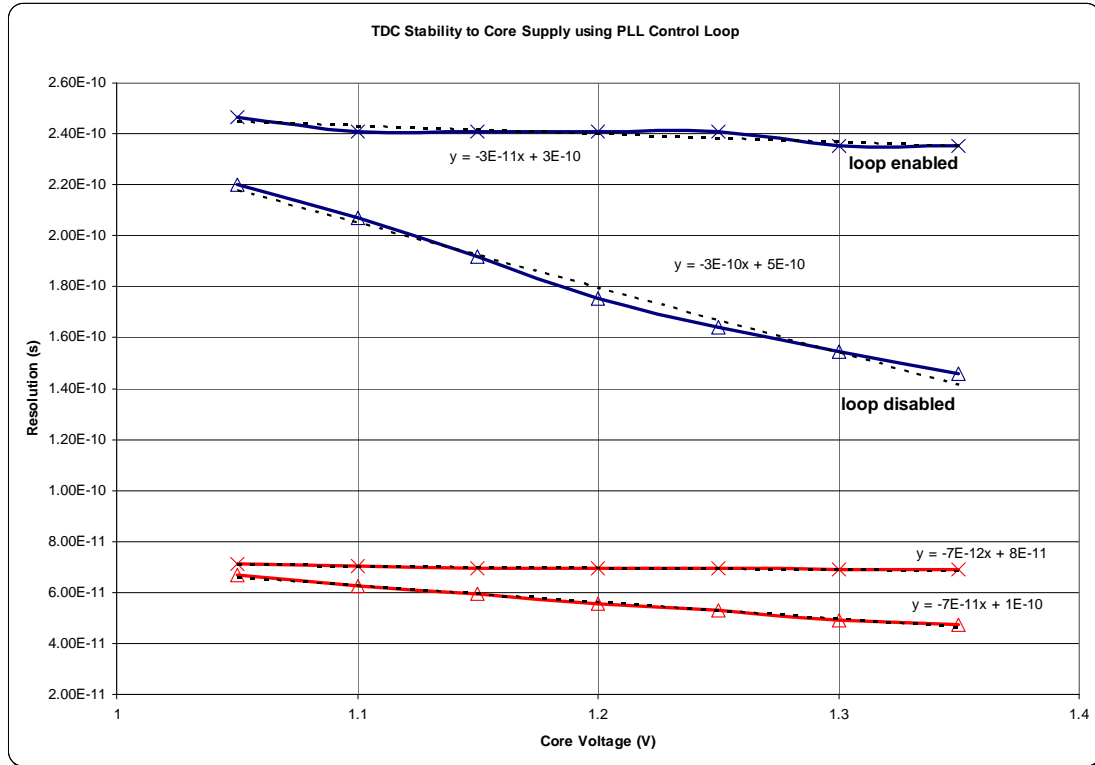


Figure 154: TDC Stability to Core Voltage Variation

Since the ring oscillator design which is embedded in the control PLL is that from the SLOW TDC, the FAST TDC would not normally be expected to be ideally calibrated. This will not be an issue for a single TDC design array.

Comparison of the slopes of the two traces above demonstrates a 20dB power supply line rejection performance when the compensation loop is enabled. The relative sensitivities of SLOW and FAST TDCs to supply level is captured below in Table 28.

<i>TDC</i>	<i>Std Slope</i>	<i>Sensitivity</i>	<i>In-Loop Slope</i>	<i>Sensitivity</i>
SLOW	300e-12	8.4 codes/V	30e-12	0.84 codes/V
FAST	70e-12	6.7 codes/V	7e-12	0.67 codes/V

Table 28: TDC Line Rejection Performance

6.2.2.9 Power Consumption

Similarly to photon counting mode, the main power consumption contributor was established as the I/O data pads, which were found to consume ~530mW at 2.8V. This can be dramatically reduced by running the I/O pad-ring at a lower voltage; for example at 1.8V the I/O power consumption would be ~220mW, 2.4 times less.

For the purposes of TCSPC mode characterisation, the core digital supply power consumption was measured for a test stimuli time interval of 10ns and frame rate of 500 kfps. The number of operating rows of pixels was incremented and the core consumption plotted, as shown in Figure 155.

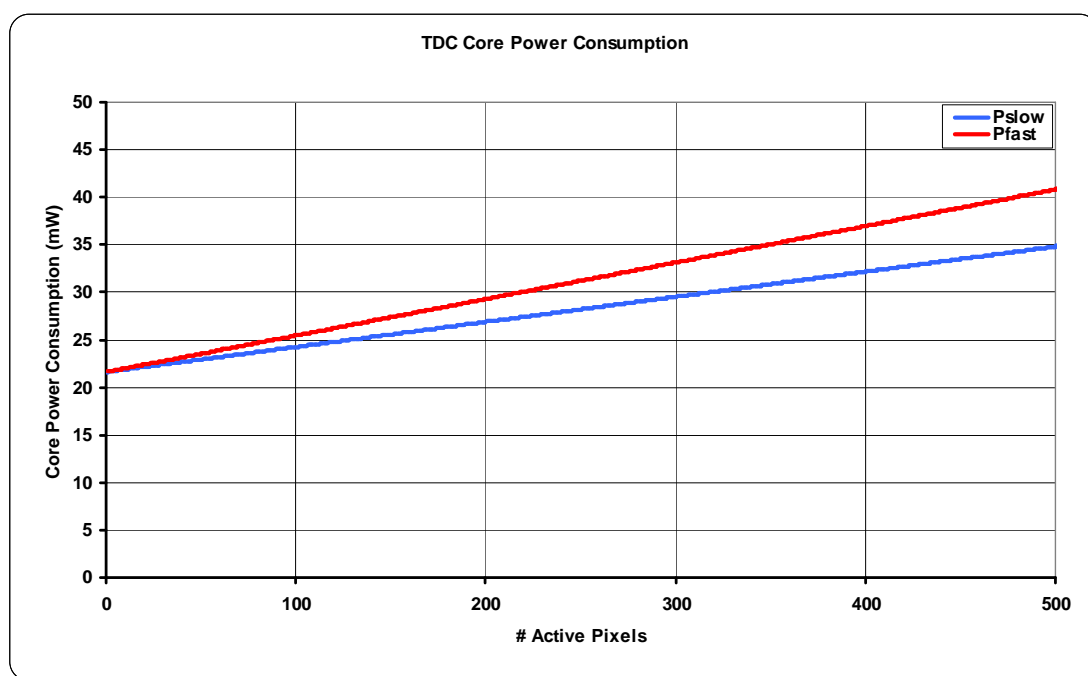


Figure 155: TCSPC Mode Power Consumption Profile

The graph demonstrates that the device exhibits a consumption profile dependent on the number of active pixels, a performance trait which is highly desirable for low photon arrival rate applications such as FLIM. Under such circumstances the array would be operating most commonly toward the lower end of the traces shown. The digital core consumption was measured at 27.6 μ W (per SLOW TDC pixel) and 38.4 μ W (per FAST TDC pixel) with a 21.6mW offset required by the ancillary and signal distribution cells. This data was for a typical 1.2V core supply. It can therefore be extrapolated that for a 32x32 array single TDC design, the consumption would be 48.6mW for the SLOW TDC and 60.9mW for the FAST TDC.

6.3 Conclusions

In this chapter, the characterisation procedures required to measure key TDC performance metrics were described. The characterisation data obtained for the new GRO-TDC based arrays described in chapter 5 was also presented for both photon counting and TCSPC operating modes. The performance obtained matched expectations and is summarised below in Table 29.

Metric	Expected Value	Measured Value (SLOW, FAST)		Unit
Technology	130nm, 4M-1P CMOS			
# Transistors	580			Per pixel
Word Length	10			bits
Uniformity (σ)	<1	0.9% (8 lsb)		%
Jitter (mean)	n/a	0.6		LSB
TDC Area	< 2500 incl SPAD	2200		μm^2
DNL/INL	$\pm 0.5/\pm 1$	$\pm 0.5/2.4$	$\pm 0.4/1.4$	LSB
Resolution	<100	178	52	ps
Power Consumption ¹	<10 μW	28	38	μW
Chip Data Rate ²	10.24	5.12		Gbps

Table 29: GRO-TDC Performance Summary

Notes: ¹:per pixel converting 10ns time period, not including array ancillary cells and chip IO. See following separate power consumption summary table; ²:on evaluation platform at 500 kfps.

The device's power consumption in the two main operating modes can be summarised and compared as shown in Table 30.

Mode	1.2V Domain		2.8V Domain
	Core	Signal Distribution	I/O
Photon Counting	1.4mW	21.6mW	33.6mW
TCSPC (slow/fast)	27.6 μW / 38.4 μW	21.6mW	530mW

Table 30: Power Consumption Summary

This demonstrates that when the device is outputting a single constant code as in photon counting test mode, power consumption is low. However, the dominant consumption of the I/O domain in TCSPC mode demonstrates that for future larger resolution SoC implementations sample lifetime computation would be best done power-efficiently on chip.

Although the characterisation of the GRO-TDC pixel shows that the design meets the overall target specification apart from slightly higher INL than desired, the code probability issue illustrated in Figure 143 reveals that there is still room for the introduction of further design improvements.

7 Conclusions and Outlook

7.1 Summary

The aim of this thesis was to push forward the boundary for time-correlated imaging in nanometer scale CMOS technology. The work embodied in this thesis has successfully demonstrated that by careful assessment of the range of implants available in a target CMOS process the implementation of high detection efficiency, low noise, deep-submicron SPADs is indeed possible with minimal performance compromises. Through this research, the application of such an assessment alongside the application of proven CMOS imager design methodologies has resulted for the first time in high performance detectors capable of resolving single photons being successfully integrated on the same silicon substrate with an array of time to digital converters. The completed imager is capable of simultaneously measuring and capturing the arrival times of *1 Giga-photons per second*. The 32x32 Gated Ring Oscillator TDC-SPAD pixel array device designed as part of this work targeted the application of fluorescence microscopy under the auspices of the Megaframe Project. The novel architecture implemented ensured the future possibility for scaling to larger arrays whilst limiting power consumption to a manageable level which scaled with photon arrival rate. The creation of a custom hardware sub-system has permitted the continued, ongoing development of system features and capabilities. A 160x128 resolution device is now undergoing bench testing.

7.2 Achievements

The implementation of low noise SPADs in STMicroelectronics 130nm Imaging CMOS process is a significant achievement, progressing from the work on detectors in a related technology by Niclass [16]. Several novel ‘retrograde well’ SPAD structures which exhibited good performance were simulated and implemented in silicon. This performance was achieved without process modification and thus permitted the successful co-integration of other data conversion and processing electronics. The resulting SPAD designs are the subject of a patent application, details for which are provided in Appendix B. The detector designs incorporate for the first time the novel use of several deeper P implants for the anode terminal in conjunction with a progressively graded N type cathode region. The graded cathode is formed by

utilising a deep N type implant which is normally intended for isolating regions of P-well. Simultaneously, the novel guard ring region is formed by the use of an implant blocking layer, meaning that the guard ring and active region dimensions are simultaneously defined by a single drawn layer. The result is a family of low noise, enhancement mode single photon detectors, without the area and timing drawbacks of implanted guard ring structures.

The novel, area efficient GRO-TDC design which was implemented was designed to suit low incident light level applications, and be compatible with scaling to large format arrays. This was the first known array of gated ring oscillator based TDCs with accompanying SPADs in an imaging process. The manufacturing process design rules allied to the chosen architecture permitted the implementation of both photon counting and time of arrival modes in a pixel of only $50 \times 50 \mu\text{m}$.

The on-chip data readout scheme functioned alongside the hardware sub-system to enable 10Gbps data transfer, or in other words one million frames per second, hence the naming of the project as ‘Megaframe’.

The key developments of this research, along with the implemented hardware-software user platform have enabled the creation of a new classification of CMOS imager. To demonstrate the use of the imager in a real application environment there follows two examples of the use of the resulting device: (a) for fluorescence microscopy (FLIM), and (b) for 3D imaging.

The imager developed enables detailed spatial and temporal analyses of biological fluorescence lifetime events which can be quickly gathered and processed in a cost effective, miniaturised platform without the need for cryogenic cooling. An integrated photon counting mode permitted ease of microscope use as a viewfinder, as well as being the primary sensing mode for fluorescence correlation spectroscopy (FCS) imaging. Examples of bio-analysis images obtained by users of the Megaframe imager are shown in Figure 156.

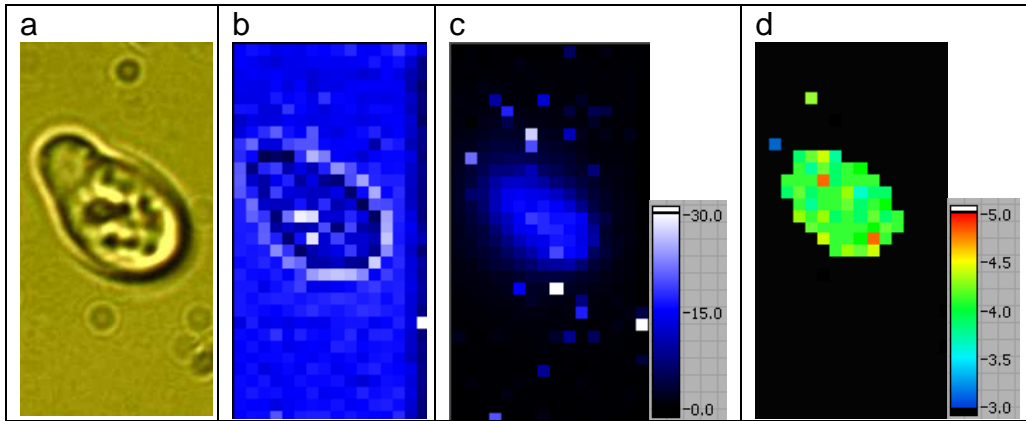


Figure 156: Example Bio-analysis Images

The images shown in Figure 156 were obtained by Day-Uei Li and Jochen Arlt of The University of Edinburgh and are the subject of a forthcoming publication. They show fungal spores (*Neurospora crassa*) which have been genetically modified to express GFP. Brightfield image on (a) standard CCD and (b) 32×16 SPAD array in photon count mode (c) Fluorescence intensity image (scale indicating photon count rate in kHz) and (d) TCSPC mode fluorescence lifetime image indicating a uniform lifetime of about 4ns throughout the spore. The Field of view is approximately $8\ \mu\text{m} \times 16\ \mu\text{m}$.

The use of the Megaframe device as a 3D camera is demonstrated below in Figure 157.

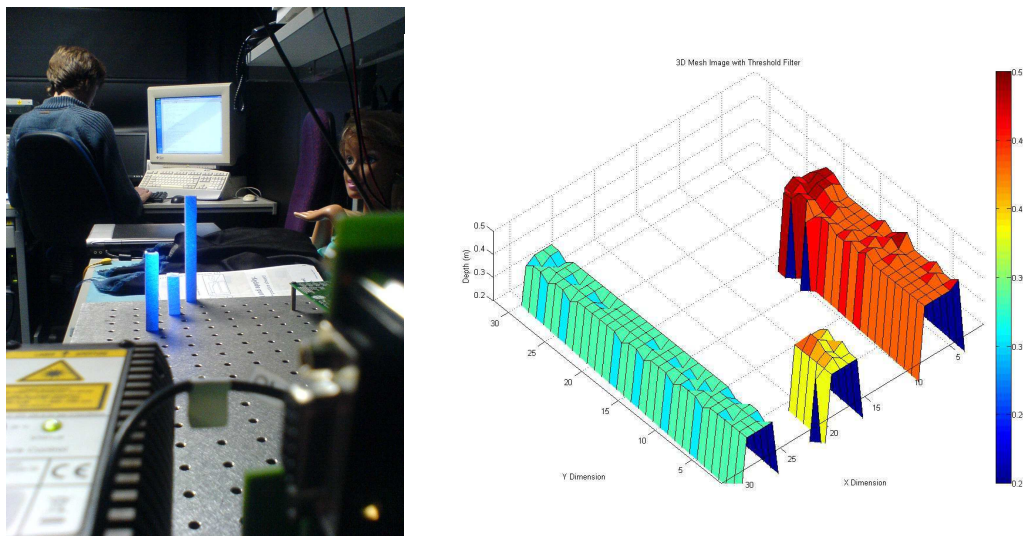


Figure 157: Example 3D Imaging Bench Setup and Output Image

The images obtained by Richard Walker of The University of Edinburgh and STMicroelectronics show the Megaframe 32x32 device ranging under ambient conditions to three metal test posts of varying height and range. The illuminator in this case is a picosecond pulsed laser at 470nm wavelength. The colour rendered mesh output image shows the X, Y and Z dimensions. Depth resolution obtained is approximately 20mm.

7.3 Improvements and Future Work

Following from this work there are several areas to investigate further. The PDE characteristic of the SPAD remains as peaking in the blue region, despite partially successful attempts to boost the longer wavelength red response. However, increasingly popular applications which use scintillators to permit the detection of nuclear particle arrivals require single photon detectors with a response weighted toward the blue, as the most commonly used scintillator material radiates around 420-450nm.

One of the main issues with the pixel layout proposed was the poor fill factor of the SPAD active region in relation to the overall pixel area (~2%) necessitated the use of a method for focussing the incoming light. This low fill factor was exacerbated by the relatively poor quantum efficiency of the SPAD (~28%) when compared to a high QE photodiode (~80%). This was partially due to the necessity of the use of upper level metals within the pixel so that it was not possible to apply the normal approach of reducing the optical stack height in order to maximise photon transmission. The result of the associated Megaframe work package which was concerned with the fabrication and installation of a microlens or microconcentrator array over the silicon die showed that this approach was distinctly challenging. A conclusion of this phase of work is that array lens fabrication is best kept at wafer-scale where it can be implemented with photolithographic accuracy and in a guaranteed clean environment, even if the result is reduced peak concentration factor. Consequently, the experimental use of the Megaframe devices has been largely done without any means of optical concentration, resulting in a system which is inefficient in its gathering of the available biological sample output photons.

An unexpected outcome of working with no optical concentrator was the discovery that the relatively wide spatial distribution of SPADs in the Megaframe array is well

suited to multipoint fluorescent correlation spectroscopy (FCS) and confocal microscopy applications.

As discussed in section 6.2.2.5, the TDC exhibited a non-uniform code probability, with a ‘beat’ of increased code occurrence every eight codes. This was thought to be due to increased susceptibility to the impact of noise at the interface between the ring oscillator and ripple counter. As a result, in the final Megaframe device which has a larger resolution 160x128 pixel array the ring oscillator has been implemented within its own isolated well. Additionally, hysteresis has been added to the buffer element which performs the role of linking the final stage of the ring oscillator to the ripple counter. This device has been fabricated and is now under evaluation.

Although the pixel power consumption was within the required target specification, the Megaframe 32x32 system had an externally FPGA based fluorescent lifetime extraction algorithm. This meant that the overall system power consumption was dominated by high frequency transfer of photon time-of-arrival data in a higher magnitude I-O voltage domain. This issue was to be addressed in the final Megaframe 160x128 device where the lifetime algorithm was implemented on-chip, resulting in a greatly reduced data transmission overhead.

7.4 Final Remarks

The research has generated significant interest in the technical community, in both electronic and biomedical engineering. As well as FLIM, other new, exciting and useful uses for the technology have become evident after discussion with biomedical researchers in a number of different establishments in several European countries. For example, the 32x32 device has already been designed and successfully integrated into another host platform for investigations into multi-point fluorescence correlation spectroscopy. This platform is already producing exciting results which will be published in the coming months.

Additionally, the capability of having detectors and processing electronics on the same substrate enables the creation of monolithic silicon replacements for vacuum tube technologies such as PMTs. The resulting silicon photomultipliers (SiPMs) are key components in scintillator based Positron Emission Tomography (PET) machines, and are robust to the high magnetic fields associated with Magnetic Resonance Imagers (MRI). The detectors developed as part of this research are central to a newly

approved European Commission funded collaborative project aimed at the creation of a network based, combined PET-MRI scanner. This four year, ~5M€ project involves eight partners from Italy, France, Netherlands, Hungary, Switzerland and the United Kingdom. The intention is to ultimately permit the medical community to use cost effective platforms to view high quality 3D images of a subject's physical structure simultaneously alongside the underlying biological processes without the need for invasive surgery. The potential for the early detection of cancers, the monitoring and impact of the treatment of degenerative diseases such as Alzheimer's, along with the enhancement of associated drug discovery research and reduction of the impact of the overall scanning procedure on the patient/subject is unequivocal.

For industrial users, the successful creation of SPADs in the same technology that is already used for a wide range of commercial imaging applications brings new possibilities for mass manufacture. A development programme capable of bringing low cost ranging capability to robotics, construction and manufacturing platforms, as well as automotive and security applications has been initiated, with the first test chips already producing promising results.

This research also has very real potential for manufacturers and users of hand-held mobile platforms. It is envisaged that the results of this research will enable the introduction of new features and capabilities based on short range proximity detection. This could result in new methods for interpreting user commands, for example by hand movement or gesture recognition techniques. The addition of new types of sensors into popular, powerful, and compact processing platforms will significantly expand the possibilities for future novel applications.

Appendix A: List of Publications

1) Justin Richardson^[1,2], Robert K. Henderson^[1], David Renshaw^[1]: ^[1]The University of Edinburgh, Institute for Integrated Micro and Nano Systems, Edinburgh, U.K. ^[2]STMicroelectronics Imaging Division, Edinburgh, UK: ‘Dynamic Quenching for Single Photon Avalanche Diode Arrays’, International Image Sensor Workshop (IISW) June 2007, Ogunquit, Maine, USA.

2) Marek Gersbach^[1], Cristiano Niclass^[1], Edoardo Charbon^[1], Justin Richardson^[2,3], Robert Henderson^[2], Lindsay Grant^[3]: ^[1]Ecole Polytechnique Fédérale de Lausanne (EPFL) Lausanne, Switzerland, ^[2]University of Edinburgh, Edinburgh, Scotland, ^[3]STMicroelectronics Edinburgh, Scotland: ‘A single photon detector implemented in a 130nm CMOS imaging process’, Solid-State Device Research Conference, 2008. (ESSDERC), 38th European, 0270-273, September 2008.

3) Day-Uei Li^[1], Richard Walker^[1], Justin Richardson^[1,2], Bruce Rae^[1], Alex Buts^[1], David Renshaw^[1], Robert Henderson^[1]: ^[1]Institute for Integrated Micro and Nano Systems, School of Engineering, The University of Edinburgh, Edinburgh, Scotland, UK, ^[2]Imaging Division, STMicroelectronics, 33 Pinkhill, Edinburgh, UK: ‘Hardware implementation and calibration of background noise for an integration-based fluorescence lifetime sensing algorithm’, Journal of Optical Society of America (JOSA), Vol. 26, No. 4, April 2009.

4) Day-Uei Li^[1], Richard Walker^[1], Justin Richardson^[1,2], Bruce Rae^[1], Alex Buts^[1], David Renshaw^[1], and Robert Henderson^[1]: ^[1]Institute for Integrated Micro and Nano Systems, School of Engineering, The University of Edinburgh, Edinburgh, UK; ^[2]Imaging Division, STMicroelectronics, Edinburgh, UK: ‘FPGA Implementation of a Video-rate Fluorescence Lifetime Imaging System with a 32×32 CMOS Single-Photon Avalanche Diode Array’. IEEE International Symposium on Circuits and Systems (ISCAS), Taipei, Taiwan, May 24-27 2009, p3082-3085.

5) Justin Richardson^[1,2], Richard Walker^[1,2], Lindsay Grant^[2], David Stoppa^[3],

Fausto Borghetti^[3], Edoardo Charbon^[4,5], Marek Gersbach^[4,5], Robert K. Henderson:
^[1] The University of Edinburgh, Edinburgh, UK; ^[2]STMicroelectronics, Imaging Division, Edinburgh, UK; ^[3]Fondazione Bruno Kessler, Trento, Italy; ^[4]EPFL, Lausanne, Switzerland; ^[5]TU Delft, Delft, The Netherlands; 'A 32x32 50ps Resolution 10 bit Time to Digital Converter Array in 130nm CMOS for Time Correlated Imaging', International Image Sensor Workshop (IISW) July 2009, Bergen, Norway.

6) Justin A. Richardson^[1,2], Lindsay A. Grant^[2], and Robert K. Henderson^[1]: ^[1]The University of Edinburgh, Institute for Integrated Micro and Nano Systems, Edinburgh, U.K. ^[2]STMicroelectronics Imaging Division, Edinburgh, UK: 'A Low Dark Count Single Photon Avalanche Diode Structure Compatible with Standard Nanometer Scale CMOS Technology', International Image Sensor Workshop (IISW), July 2009, Bergen, Norway.

7) Justin A. Richardson^[1,2], Lindsay A. Grant^[2], and Robert K. Henderson^[1]: ^[1]The University of Edinburgh, Institute for Integrated Micro and Nano Systems, Edinburgh, U.K. ^[2]STMicroelectronics Imaging Division, Edinburgh, UK: 'Reduction of Band-to-band Tunneling in Deep-submicron CMOS Single Photon Avalanche Diodes', International Image Sensor Workshop (IISW), July 2009, Bergen, Norway.

8) Justin A. Richardson^[1,2], Lindsay A. Grant^[2], and Robert K. Henderson^[1]: ^[1]The University of Edinburgh, Institute for Integrated Micro and Nano Systems, Edinburgh, U.K. ^[2]STMicroelectronics Imaging Division, Edinburgh, UK: 'A Low Dark Count Single Photon Avalanche Diode Structure Compatible with Standard Nanometer Scale CMOS Technology', IEEE Photonics Technology Letters, July 2009, Vol. 21, Issue 14, p1020-1022.

9) M. Gersbach^[1], J. Richardson^[2,3], E. Mazaleyrat^[2], S. Hardillier^[2], C. Niclass^[1], R. Henderson^[3], L. Grant^[2], and E. Charbon^[1], ^[1]Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, ^[2]ST Microelectronics Imaging Division, Edinburgh, Scotland, ^[3]University of Edinburgh, Edinburgh, Scotland: "A low-noise single-photon detector implemented in a 130 nm CMOS imaging process," Solid-State Electronics, vol. 53, pp. 803-808, Jul 2009.

10) Justin Richardson^[1,2], Richard Walker^[1,2], Lindsay Grant^[2], David Stoppa^[3], Fausto Borghetti^[3], Edoardo Charbon^[4,5], Marek Gersbach^[4,5], Robert K. Henderson: ^[1] The University of Edinburgh, Edinburgh, UK; ^[2]STMicroelectronics, Imaging Division, Edinburgh, UK; ^[3]Fondazione Bruno Kessler, Trento, Italy; ^[4]EPFL, Lausanne, Switzerland; ^[5]TU Delft, Delft, The Netherlands; ‘A 32x32 50ps Resolution 10 bit Time to Digital Converter Array in 130nm CMOS for Time Correlated Imaging’, Custom Integrated Circuits Conference (CICC) September 2009, San Jose, USA.

11) David Stoppa¹, Fausto Borghetti¹, Justin Richardson^{2,3}, Richard Walker^{2,3}, Lindsay Grant², Robert K. Henderson³, Marek Gersbach^{4,5}, Edoardo Charbon^{4,5}
¹Fondazione Bruno Kessler (FBK), Trento, Italy, ²STMicroelectronics Imaging Division, Edinburgh, U.K., ³The University of Edinburgh, Edinburgh, U.K.
⁴TU Delft, Delft, The Netherlands, ⁵Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, ‘A 32x32-Pixel Array with In-Pixel Photon Counting and Arrival Time Measurement in the Analog Domain’, ESSCIRC 2009, Athens, Greece.

12) M. Gersbach^{1,5}, Y. Maruyama⁵, E. Labonne¹, J. Richardson^{2,3}, R. Walker³, R. Henderson³, F. Borghetti⁴, D. Stoppa⁴, and E. Charbon^{1,5}, ¹Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, ²ST Microelectronics Imaging Division, Edinburgh, Scotland, ³University of Edinburgh, Edinburgh, Scotland, ⁴Fondazione Bruno Kessler, Trento, Italy, ⁵TU Delft, Delft, The Netherlands: ‘A Parallel 32x32 Time-To-Digital Converter Array Fabricated in a 130 nm Imaging CMOS Technology’, ESSCIRC 2009, Athens, Greece.

Appendix B: Patent Applications

1) EP08275029, 'Improvements in Single Photon Avalanche Diodes', 10th July 2008, STMicroelectronics R&D Limited, P106923.EP.01, Justin Richardson & Robert Henderson.

2) US12327240 , 'Single Photon Detector And Associated Methods For Making The Same', 3rd December 2008, STMicroelectronics R&D Limited, Justin Richardson, Marek Gersbach, Edoardo Charbon, Cristiano Niclass, Robert Henderson, Lindsay Grant.

Appendix C: SPAD Design Additional Data

Step	Name	CAD-Mask	Data	Usage/Comments
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				
28				
29				
30				
31				
32				
33				
34				
35				
36				
37				
38				
39				
40				
41				
42				
43				
44				
45				
46				
47				
48				
49				
50				
51				
52				
53				
54				
55				
56				
57				
58				
59				
60				
61				
62				
63				
64				
65				
66				
67				
68				
69				
70				
71				
72				
73				
74				
75				
76				
77				
78				
79				
80				
81				
82				
83				
84				
85				
86				
87				
88				
89				
90				
91				
92				
93				
94				
95				
96				
97				
98				
99				
100				

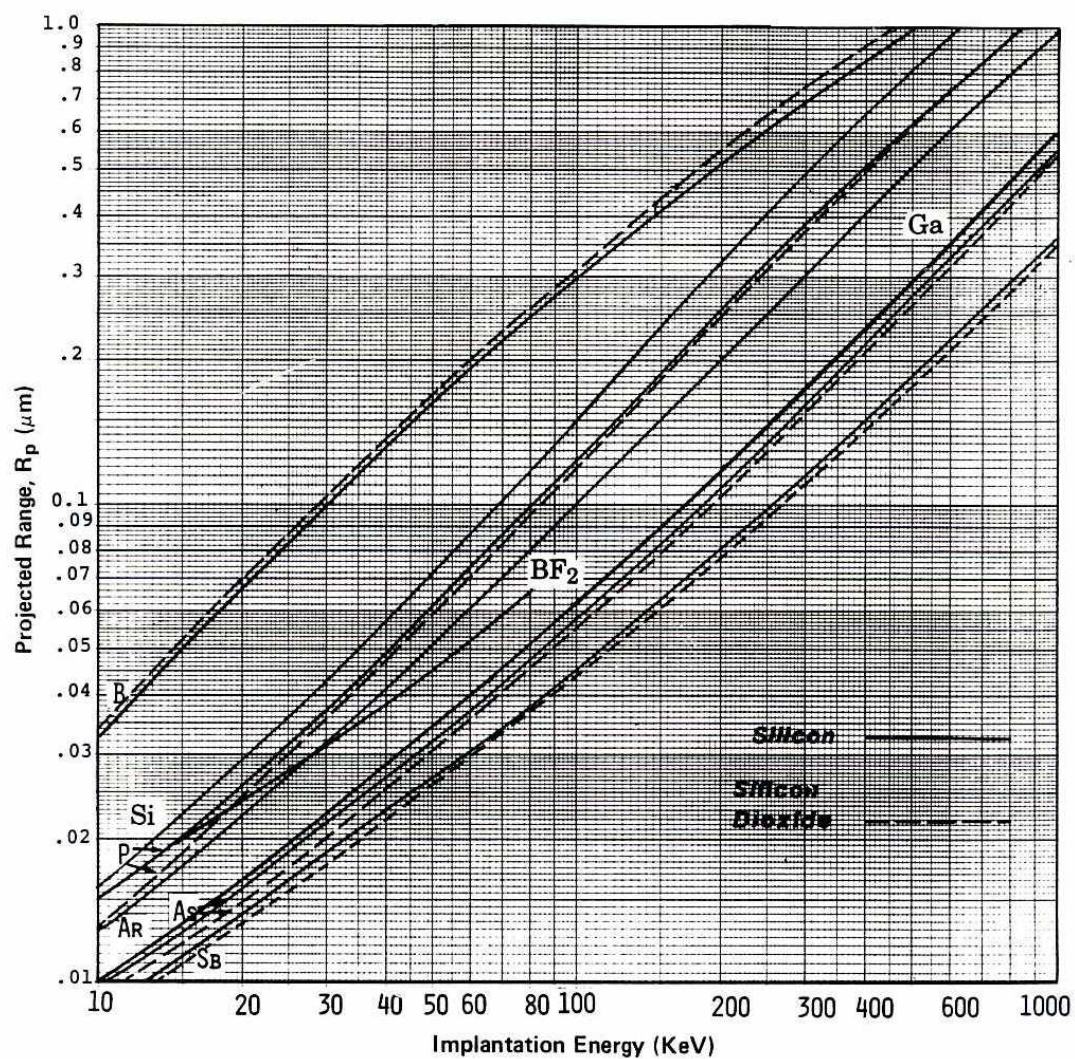


Figure 158: Implant Energy vs Depth in Silicon

Appendix D: TDC Design Additional Data

Table 32: TDC Input-Output Signal Definition

Name	I/O	Dir	Width	Description
VQUENCH	I	V in M4	1	Analogue voltage to control PMOS gate quench resistance/dead time. Access TOP/BOTTOM (power grid)
MODETCSPC	I	H in M3	1	Static global control signal selecting TAC/TDC mode: Requires global buffering. 0: photon count mode 1: TCSPC mode Access RIGHT.
STARTSRC	I	H in M3	1	Static global control signal selecting digitiser start input: (extra logic req'd for TDC cal). Requires global buffering. 0: normal, from the SPAD 1: test mode, from the TESTSTART input. Access RIGHT.
COLEN	I	V in M2	1	Column enable direct from I2C. Active high static control signal. Access TOP/BOTTOM.
ROWEN	I	H in M3	1	Row enable direct from I2C. Active high static control signal. Access RIGHT.
ROWSEL	I	H in M3	1	Row select direct from YDEC – enables output bus drivers. Active high. Access LEFT.
WRITE	I	H _{ct} in M3	1	Write signal for pixel memory. Once per frame 1MHz signal. Global clock tree distribution. Access LEFT.
PIXRESETN	I	H _{ct} in M3	1	Pixel reset (active low), once per frame global digitiser reset (1MHz), variable (X x laser prf) duration. Global clock tree distribution. Access LEFT.
SPADEN	I	H _{ct} in M3	1	SPAD enable/gating signal for mode2 and FCS. Global clock tree distribution. Access RIGHT.
FASTQUENCH	I	H _{ct} in M3	1	SPAD fast quenching signal. Global clock tree distribution. Access RIGHT.
STOP	I	H _{ct} in M3	1	Stop signal for TAC or TDC, active high. (either laser prf or FPGA delayed stop) Global clock tree distribution. Access RIGHT.
DIGITISERCLK	I	H _{ct} in M3	1	64MHz for TAC 512MHz for TDC Global clock tree distribution. Access RIGHT.
TESTSTART	I	H _{ct} in M3	1	Start signal for the digitiser while in test mode. Global clock tree distribution. Access RIGHT.
DATAOUT<7:0>	O	V in M2	8	Connects to column data bus. Must tri-state outputs when row is not selected. Access TOP/BOTTOM.

Table 33: 32x32 TDC Array I2C Register Map

Reg Index	Bit	Name	Default	Description
100d	<7:0>	COLEN<7:0>	00h	Column enable registers
101d	<7:0>	COLEN<15:8>	00h	
102d	<7:0>	COLEN<23:16>	00h	
103d	<7:0>	COLEN<31:24>	00h	
104d	<7:0>	ROWEN<7:0>	00h	Row enable registers
105d	<7:0>	ROWEN<15:8>	00h	
106d	<7:0>	ROWEN<23:16>	00h	
107d	<7:0>	ROWEN<31:24>	00h	
108d	<0>	MODETCSPC	0	Selects TAC/TDC Mode 0 – photon counting 1- TCSPC
	<1>	STARTSRC	0	Selects TAC/TDC start signal source: 0=SPADs; 1=TESTSTART.
	<2>	STOPSRC	0	Selects TAC/TDC stop signal source: 0=OPTCLK; 1=EXTSTOP.
	<3>	PLLSRC	0	Selects device PLL source clock: 0=OPTCLK; 1=EXTCLK.
	<4>	YDECBP	0	Allows Y decoder to be bypassed (for raw spad o/p mode): 0=Y decoder operating; 1=ROWSEL signals driven by ROWEN registers.
	<5>	SERBP	0	Allows serialisers to be bypassed for raw spad op mode: 0=serialisers operating; 1=column LSBs connected to op pads
	<6>	SERGATINGBP	0	Allows the serialisers to be disabled via clock gating when not required: 0=Gated when not required; 1=Permanently enabled.
	<7>	SOFTRESETN	0	Active low soft reset:

Reg Index	Bit	Name	Default	Description
				0=System in soft reset; 1=System operating.
109d (TDC specific register)	<0>	SEROUTALGN (UNIED: COUNTENABLE)	0	TDC specific. Selects serialiser output alignment: 0=photon counting mode; 1=time correlated mode.
	<1>	MANCALIB (UNIED: MODEMUTEON)	0	Selects manual calibration mode
	<5:2>	REF<3:0>	0h	Reference result for calibration
	<7:6>	Unused		
110d PLL Reg 1	<0>	ENABLE	0	PLL enable.
	<2:1>	DIVCTRL<1:0>	00	Sets divider ratio for PLL input clock divider. 00 = No division 01 = divide by 2 10 = divide by 4 11 = divide by 8
	<4:3>	PDIV1<1:0>	00	PLL P1 divider ratio.
	<6:5>	PDIV2<1:0>	00	PLL P2 divider ratio.
	<7>	SSCG_CONTROL	0	PLL spread spectrum control enable.
111d PLL Reg 2	<7:0>	NDIV<7:0>	00	PLL N divider ratio.

Appendix E: References

1. Schuster, M.A. and G. Strull. *A monolithic mosaic of photon sensors for solid state imaging applications*. in *International Electron Devices Meeting*. 1965.
2. Boyle, W.S. and G.E. Smith, *Charge-coupled Semiconductor Devices*. Bell Systems Technical Journal, 1970. **49**: p. 587-593.
3. Renshaw, D., et al. *CMOS video cameras*. in *Euro ASIC '91*. 1991.
4. Fossum, E.R., S. Mendis, and S.E. Kemeny, *CMOS active pixel image sensor*. *Electron Devices, IEEE Transactions on*, 1994. **41**(3): p. 452-453.
5. Burr, K.C. and W. Gin-Chung. *Scintillation detection using 3 mm x 3 mm silicon photomultipliers*. in *Nuclear Science Symposium Conference Record, 2007. NSS '07. IEEE*. 2007.
6. McIntyre, R.J., *Theory of Microplasma Instability in Silicon*. *Journal of Applied Physics*, 1961. **32**(6): p. 983-995.
7. Ruegg, H.W., *An optimized avalanche photodiode*. *Electron Devices, IEEE Transactions on*, 1967. **14**(5): p. 239-251.
8. Haitz, R.H., et al., *Avalanche Effects in Silicon p---n Junctions. I. Localized Photomultiplication Studies on Microplasmas*. *Journal of Applied Physics*, 1963. **34**(6): p. 1581-1590.
9. Renker, D., *Geiger-mode avalanche photodiodes, history, properties and problems*. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 2006. **567**(1): p. 48-56.
10. Nutt, R., *Digital Time Intervalometer*. *The Review of Scientific Instruments*, 1968. **39**(9).
11. Arai, Y. and T. Baba. *A CMOS time to digital converter VLSI for high-energy physics*. in *VLSI Circuits, 1988. Digest of Technical Papers., 1988 Symposium on*. 1988.
12. Dudek, P., S. Szczepanski, and J.V. Hatfield, *A high-resolution CMOS time-to-digital converter utilizing a Vernier delay line*. *Solid-State Circuits, IEEE Journal of*, 2000. **35**(2): p. 240-247.
13. Kinoshita, S., H. Ohta, and T. Kushida, *Subnanosecond fluorescence-lifetime measuring system using single photon counting method with mode-locked laser excitation*. *Review of Scientific Instruments*, 1981. **52**(4): p. 572-575.
14. Rae, B.R., et al. *A Microsystem for Time-Resolved Fluorescence Analysis using CMOS Single-Photon Avalanche Diodes and Micro-LEDs*. in *Solid-State Circuits Conference, 2008. ISSCC 2008. Digest of Technical Papers. IEEE International*. 2008.
15. Rochas, A., *PhD Thesis: Single Photon Avalanche Diodes in CMOS Technology*, in *Section De Microtechnique*. 2003, EPFL: Lausanne. p. 227.
16. Niclass, C., *PhD Thesis: Single-photon image sensors in CMOS : picosecond resolution for three-dimensional imaging*. 2008, EPFL: Lausanne.
17. Haitz, R.H., *Mechanisms Contributing to the Noise Pulse Rate of Avalanche Diodes*. *Journal of Applied Physics*, 1965. **36**(10): p. 3123-3131.
18. Zanchi, A., F. Zappa, and M. Ghioni, *A probe detector for defectivity assessment in p-n junctions*. *Electron Devices, IEEE Transactions on*, 2000. **47**(3): p. 609-616.

19. Finkelstein, H., M.J. Hsu, and S. Esener. *An ultrafast Geiger-mode single-photon avalanche diode in 0.18- μ m CMOS technology*. in *Advanced Photon Counting Techniques*. 2006. Boston, MA, USA: SPIE.
20. Niclass, C., et al., *A Single Photon Avalanche Diode Implemented in 130-nm CMOS Technology*. *Selected Topics in Quantum Electronics, IEEE Journal of*, 2007. **13**(4): p. 863-869.
21. Faramarzpour, N., et al., *Fully Integrated Single Photon Avalanche Diode Detector in Standard CMOS 0.18 μ m Technology*. *Electron Devices, IEEE Transactions on*, 2008. **55**(3): p. 760-767.
22. Cohen, M., et al. *Fully Optimized Cu based process with dedicated cavity etch for 1.75 μ m and 1.45 μ m pixel pitch CMOS Image Sensors*. in *Electron Devices Meeting, 2006. IEDM '06. International*. 2006.
23. Ghioni, M., et al., *New silicon epitaxial avalanche diode for single-photon timing at room temperature*. *Electronics Letters*, 1988. **24**(24): p. 1476-1477.
24. Lacaita, A., M. Ghioni, and S. Cova, *Double epitaxy improves single-photon avalanche diode performance*. *Electronics Letters*, 1989. **25**(13): p. 841-843.
25. Hsu, M.J., S.C. Esener, and H. Finkelstein, *A CMOS STI-Bound Single-Photon Avalanche Diode With 27-ps Timing Resolution and a Reduced Diffusion Tail*. *Electron Device Letters, IEEE*, 2009. **30**(6): p. 641-643.
26. Haitz, R.H., *Model for the Electrical Behavior of a Microplasma*. *Journal of Applied Physics*, 1964. **35**(5): p. 1370-1376.
27. Lacaita, A.L., et al., *On the bremsstrahlung origin of hot-carrier-induced photons in silicon devices*. *Electron Devices, IEEE Transactions on*, 1993. **40**(3): p. 577-582.
28. Sciacca, E., et al., *Arrays of Geiger mode avalanche photodiodes*. *Photonics Technology Letters, IEEE*, 2006. **18**(15): p. 1633-1635.
29. Kindt, W.J., H.W. van Zeijl, and S. Middelhoek. *Optical Cross Talk in Geiger Mode Avalanche Photodiode Arrays: Modeling, Prevention and Measurement*. in *Solid-State Device Research Conference, 1998. Proceeding of the 28th European*. 1998.
30. Becker, W., *Advanced Time-Correlated Single Photon Counting Techniques*. 2005, Berlin: Springer.
31. Pellegrini, S., et al., *Design and performance of an InGaAs-InP single-photon avalanche diode detector*. *Quantum Electronics, IEEE Journal of*, 2006. **42**(4): p. 397-403.
32. Turchetta, R., et al., *High spatial resolution silicon read-out system for single photon X-ray detection*. *Nuclear Science, IEEE Transactions on*, 1994. **41**(4): p. 1063-1068.
33. Aull, B., et al. *Laser Radar Imager Based on 3D Integration of Geiger-Mode Avalanche Photodiodes with Two SOI Timing Circuit Layers*. 2006.
34. McIntosh, K.A., et al. *Arrays of III-V semiconductor Geiger-mode avalanche photodiodes*. 2003.
35. Campbell, M., et al., *A readout chip for a 64 \times 64 pixel matrix with 15-bit single photon counting*. *Nuclear Science, IEEE Transactions on*, 1998. **45**(3): p. 751-753.
36. Rochas, A., et al., *Single photon detector fabricated in a complementary metal--oxide--semiconductor high-voltage technology*. *Review of Scientific Instruments*, 2003. **74**(7): p. 3263-3270.
37. Tisa, S., A. Tosi, and F. Zappa, *Fully Integrated CMOS single photon counter*. *OSA Optics Express*, 2007. **15**(6).

38. Gersbach, M., et al. *A single photon detector implemented in a 130nm CMOS imaging process*. in *Solid-State Device Research Conference, 2008. ESSDERC 2008. 38th European*. 2008.
39. Niclass, C., M. Sergio, and E. Charbon. *A Single Photon Avalanche Diode Array Fabricated in 0.35um CMOS and based on an Event Driven Readout for TCSPC Experiments*. in *SPIE*. 2006.
40. Marwick, M.A. and A.G. Andreou, *Single photon avalanche photodetector with integrated quenching fabricated in TSMC 0.18 μ m 1.8 V CMOS process*. *Electronics Letters*, 2008. **44**(10): p. 643-644.
41. Goetzberger, A., et al., *Avalanche Effects in Silicon p-n Junctions. II. Structurally Perfect Junctions*. *Journal of Applied Physics*, 1963. **34**(6): p. 1591-1600.
42. Cova, S., A. Longoni, and A. Andreoni, *Towards picosecond resolution with single-photon avalanche diodes*. *Review of Scientific Instruments*, 1981. **52**(3): p. 408-412.
43. Kindt, W.J. *A novel avalanche photodiode array*. in *Nuclear Science Symposium and Medical Imaging Conference, 1994., 1994 IEEE Conference Record*. 1994.
44. Rochas, A., et al., *Low-noise silicon avalanche photodiodes fabricated in conventional CMOS technologies*. *Electron Devices, IEEE Transactions on*, 2002. **49**(3): p. 387-394.
45. Ghioni, M., et al., *Progress in Silicon Single-Photon Avalanche Diodes*. *Selected Topics in Quantum Electronics, IEEE Journal of*, 2007. **13**(4): p. 852-862.
46. Petrillo, G.A., et al., *Scintillation Detection with Large-Area Reach-Through Avalanche Photodiodes*. *Nuclear Science, IEEE Transactions on*, 1984. **31**(1): p. 417-423.
47. Pancheri, L., L. Pancheri, and D. Stoppa. *Low-Noise CMOS single-photon avalanche diodes with 32 ns dead time*. in *37th European Solid State Device Research Conference, 2007. ESSDERC*. 2007.
48. Pauchard, A., et al. *Ultraviolet avalanche photodiode in CMOS technology*. in *International Electron Devices Meeting (IEDM)*. 2000.
49. Rochas, A., P.A. Besse, and R.S. Popovic. *A Geiger Mode Avalanche Photodiode Fabricated in a Conventional CMOS Technology*. 2001.
50. Xiao, Z., D. Pantic, and R.S. Popovic. *A New Single Photon Avalanche Diode in CMOS High-Voltage Technology*. in *Solid-State Sensors, Actuators and Microsystems Conference, 2007. TRANSDUCERS 2007. International*. 2007.
51. Lacaita, A., et al., *Single-photon avalanche diode with ultrafast pulse response free from slow tails*. *Electron Device Letters, IEEE*, 1993. **14**(7): p. 360-362.
52. Finkelstein, H., M.J. Hsu, and S.C. Esener, *STI-Bounded Single-Photon Avalanche Diode in a Deep-Submicrometer CMOS Technology*. *Electron Device Letters, IEEE*, 2006. **27**(11): p. 887-889.
53. Cova, S., et al., *Avalanche photodiodes and quenching circuits for single-photon detection*. *Applied Optics*, 1996. **35**(12): p. 1956-1976.
54. Ghioni, M., et al., *Compact active quenching circuit for fast photon counting with avalanche photodiodes*. *Review of Scientific Instruments*, 1996. **67**(10): p. 3440-3448.
55. Richardson, J.A., R.K. Henderson, and D. Renshaw, *Dynamic Quenching for Single Photon Avalanche Diode Arrays*. *IISW*, 2007.

56. Pauchard, A., et al., *Ultraviolet-selective avalanche photodiode*. Sensors and Actuators A: Physical, 2000. **82**(1-3): p. 128-134.
57. Verghese, S., et al., *Arrays of InP-based Avalanche Photodiodes for Photon Counting*. Selected Topics in Quantum Electronics, IEEE Journal of, 2007. **13**(4): p. 870-886.
58. Mantyniemi, A., *PhD Thesis: An Integrated CMOS High Precision Time-to-Digital Converter based on Stabilised Three-Stage Delay Line Interpolation*. 2004, University of Oulu: Oulu, Finland.
59. Doernberg, J., H.S. Lee, and D.A. Hodges, *Full-speed testing of A/D converters*. Solid-State Circuits, IEEE Journal of, 1984. **19**(6): p. 820-827.
60. Staszewski, R.B., et al., *1.3 V 20 ps time-to-digital converter for frequency synthesis in 90-nm CMOS*. Circuits and Systems II: Express Briefs, IEEE Transactions on [see also Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on], 2006. **53**(3): p. 220-224.
61. Gardner, F.M., *Phaselock Techniques*. 2nd ed. 1979: Wiley.
62. Rahkonen, T., J. Kostamovaara, and S. Saynajakangas. *Time interval measurements using integrated tapped CMOS delay lines*. 1989.
63. Raisanen-Ruotsalainen, E., T. Rahkonen, and J. Kostamovaara, *A low-power CMOS time-to-digital converter*. Solid-State Circuits, IEEE Journal of, 1995. **30**(9): p. 984-990.
64. Karadamoglou, K., et al., *An 11-bit high-resolution and adjustable-range CMOS time-to-digital converter for space science instruments*. Solid-State Circuits, IEEE Journal of, 2004. **39**(1): p. 214-222.
65. Henzler, S., et al., *A Local Passive Time Interpolation Concept for Variation-Tolerant High-Resolution Time-to-Digital Conversion*. Solid-State Circuits, IEEE Journal of, 2008. **43**(7): p. 1666-1676.
66. Abas, A.M., et al., *Time difference amplifier*. Electronics Letters, 2002. **38**(23): p. 1437-1438.
67. Minjae, L. and A.A. Abidi. *A 9b, 1.25ps Resolutilon Coarse-Fine Time-to-Digital Converter in 90nm CMOS that Amplifies a Time Residue*. in *VLSI Circuits, 2007 IEEE Symposium on*. 2007.
68. Niclass, C., et al., *A 128 x 128 Single-Photon Image Sensor With Column-Level 10-Bit Time-to-Digital Converter Array*. Solid-State Circuits, IEEE Journal of, 2008. **43**(12): p. 2977-2989.
69. Raisanen-Ruotsalainen, E., T. Rahkonen, and J. Kostamovaara. *A high resolution time-to-digital converter based on time-to-voltage interpolation*. in *ESSCIRC*. 1997. Estoril, Portugal.
70. Swann, B.K., et al., *A 100-ps time-resolution CMOS time-to-digital converter for positron emission tomography imaging applications*. Solid-State Circuits, IEEE Journal of, 2004. **39**(11): p. 1839-1852.
71. Nissinen, I. and J. Kostamovaara. *A temperature stabilized CMOS ring oscillator for a time-to-digital converter of a laser radar*. in *IMTC*. 2004. Como, Italy.
72. Nissinen, I., A. Mantyniemi, and J. Kostamovaara. *A CMOS time-to-digital converter based on a ring oscillator for a laser radar*. in *Solid-State Circuits Conference, 2003. ESSCIRC '03. Proceedings of the 29th European*. 2003.
73. Arai, Y. and M. Ikeno, *A time digitizer CMOS gate-array with a 250 ps time resolution*. Solid-State Circuits, IEEE Journal of, 1996. **31**(2): p. 212-220.

74. Helal, B.M., et al. *A Low Jitter 1.6 GHz Multiplying DLL Utilizing a Scrambling Time-to-Digital Converter and Digital Correlation*. in *VLSI Circuits, 2007 IEEE Symposium on*. 2007.
75. De Heyn, V., et al. *A fast start-up 3GHz;10GHz digitally controlled oscillator for UWB impulse radio in 90nm CMOS*. in *Solid State Circuits Conference, 2007. ESSCIRC 2007. 33rd European*. 2007.
76. Eltoukhy, H., et al. *A 0.18/spl mu/m CMOS 10/sup -6/ lux bioluminescence detection system-on-chip*. in *Solid-State Circuits Conference, 2004. Digest of Technical Papers. ISSCC. 2004 IEEE International*. 2004.
77. Mikulec, B., *Single Photon Detection with Semiconductor Pixel Arrays for Medical Imaging Applications*. 2000.
78. Llopert, X., et al. *Medipix2, a 64k pixel read out chip with 55 μm square elements working in single photon counting mode*. in *Nuclear Science Symposium Conference Record, 2001 IEEE*. 2001.
79. Ballabriga, R., et al., *The Medipix3 Prototype, a Pixel Readout Chip Working in Single Photon Counting Mode With Improved Spectrometric Performance*. *Nuclear Science, IEEE Transactions on*, 2007. **54**(5): p. 1824-1829.
80. Jackson, J.C., et al., *Toward integrated single-photon-counting microarrays*. *Optical Engineering*, 2003. **42**(1): p. 112-118.
81. SensL, *DigitalAPD: Photon Imagers with High Resolution Timing*. 2009(Draft 1.0).
82. Niclass, C., et al. *Single-photon synchronous detection*. in *Solid-State Circuits Conference, 2008. ESSCIRC 2008. 34th European*. 2008.
83. Gersbach, M., et al. *A parallel 32x32 time-to-digital converter array fabricated in a 130 nm imaging CMOS technology*. in *ESSCIRC, 2009. ESSCIRC '09. Proceedings of*. 2009.
84. Stoppa, D., et al. *A 32x32-pixel array with in-pixel photon counting and arrival time measurement in the analog domain*. in *ESSCIRC, 2009. ESSCIRC '09. Proceedings of*. 2009.
85. Gersbach, M., et al., *A low-noise single-photon detector implemented in a 130 nm CMOS imaging process*. *Solid-State Electronics*, 2009. **53**(7): p. 803-808.
86. Cova, S., et al. *A view on progress of silicon single-photon avalanche diodes and quenching circuits*. in *Advanced Photon Counting Techniques*. 2006. Boston, MA, USA: SPIE.
87. McGuire, G., *Semiconductor Technology Handbook*. 1993 ed, ed. T. Associates. 1993: William Andrew.
88. Sciacca, E., et al., *Silicon planar technology for single-photon optical detectors*. *Electron Devices, IEEE Transactions on*, 2003. **50**(4): p. 918-925.
89. Mosconi, D., et al. *CMOS Single-Photon Avalanche Diode Array for Time-Resolved Fluorescence Detection*. 2006.
90. Gronholm, M., J. Poikonen, and M. Laiho. *A ring-oscillator-based active quenching and active recharge circuit for single photon avalanche diodes*. in *Circuit Theory and Design, 2009. ECCTD 2009. European Conference on*. 2009.
91. Beynon, J.D.E. and D.R. Lamb, *Charge-coupled Devices and their Applications*. 1980: McGraw-Hill.
92. Zanchi, A., et al. *Probe detectors for mapping manufacturing defects*. in *Devices, Circuits and Systems, 2000. Proceedings of the 2000 Third IEEE International Caracas Conference on*. 2000.

93. Megaframe, C., *Megaframe Deliverable D2.1: From Concepts to Requirements*. 2006(1.1): p. 51.
94. Herzel, F. and B. Razavi, *A study of oscillator jitter due to supply and substrate noise*. Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on [see also Circuits and Systems II: Express Briefs, IEEE Transactions on], 1999. **46**(1): p. 56-62.