THE UNIVERSITY
of EDINBURGH

A Parallel Reconfigurable Single Photon Avalanche Diode Array for Optical Communications

Edward M. D. Fisher

Doctor of Philosophy

THE UNIVERSITY OF EDINBURGH

2nd August 2013

Pleasure to me is wonder— the unexplored, the unexpected, the thing that is hidden and the changeless thing that lurks behind superficial mutability. To trace the remote in the immediate; the eternal in the ephemeral; the past in the present; the infinite in the finite; these are to me the springs of delight and beauty.

'In Defence of Dagon' [1921]

H.P. Lovecraft
(1890-1937)

To my best friends and loving cats,

Spock

1996 – 2012

and

Jim

1996 – 2015

Lay Abstract

There is a pressing need to develop alternative communications links due to the limitations of current connections. A number of physical phenomena, limit the bandwidth and energy efficiency of wire-based systems and economic factors such as cost, material-supply reliability and environmental costs, are beginning to restrict our use of some technologies. The modern internet has moved towards optical connections in an effort to reduce costs, energy use and to supply the high data rates needed.

A primary concern is that the current optical-detection device requires very high optical power to achieve fast data rates with high signal quality. The energy required therefore, quickly becomes a problem. In this thesis, advances in single-photon sensitive devices are utilised to reduce the amount of light needed and to reduce the overall energy required.

Current high performance receivers often use exotic materials, many of which have severe environmental impact and have cost, supply and political restrictions on their use. These present a problem when it comes to connecting the detector to the electronics needed to process their outputs. To achieve high energy efficiency, we must consolidate the separate parts (optical, electrical and processing) into a small unit. Exotic materials hinder this, forcing both expensive technologies and limits to receiver size and efficiency. In this thesis, silicon technology similar to computer processors is investigated, as all components can be integrated onto a small, mass-producible, low power silicon chip.

The thesis concludes that data can be captured well with a single-photon sensitive receiver. More electrical energy is needed at the receiver due to its fundamental operation but overall, optical power can be reduced, allowing significant savings in either transmitter power or the transmission length, along with the advantages of an integrated digital chip.

Abstract

There is a pressing need to develop alternative communications links due to a number of physical phenomena, limiting the bandwidth and energy efficiency of wire-based systems or economic factors such as cost, material-supply reliability and environmental costs. Networks have moved to optical connections to reduce costs, energy use and to supply high data rates. A primary concern is that current optical-detection devices require high optical power to achieve fast data rates with high signal quality. The energy required therefore, quickly becomes a problem.

In this thesis, advances in single-photon avalanche diodes (SPADs) are utilised to reduce the amount of light needed and to reduce the overall energy budget. Current high performance receivers often use exotic materials, many of which have severe environmental impact and have cost, supply and political restrictions. These present a problem when it comes to integration; hence silicon technology is used, allowing small, mass-producible, low power receivers.

A reconfigurable SPAD-based integrating receiver in standard 130nm imaging CMOS is presented for links with a readout bandwidth of 100MHz. A maximum count rate of 58G photon/s is observed, with a dynamic range of $\approx$ 79dB, a sensitivity of ≈ -31.7dBm at 100MHz and a BER of $\approx 1\text{x}10^{-9}$. We investigate the properties of the receiver for optical communications in the visible spectrum, using its added functionality and reconfigurability to experimentally explore non-ideal influences. The all-digital 32x32 SPAD array, achieves a minimum dead time of 5.9ns, and a median dark count rate (DCR) of 2.5kHz per SPAD. High noise devices can be weighted or removed to optimise the SNR. The power requirements, transient response and received data are explored and limiting factors similar to those of photodiode receivers are observed.

The thesis concludes that data can be captured well with such a device but more electrical energy is needed at the receiver due to its fundamental operation. Overall, optical power can be reduced, allowing significant savings in either transmitter power or the transmission length, along with the advantages of an integrated digital chip.

Acknowledgements

ST Microelectronics Imaging Division were helpful in the use of the co-owned SPAD device, the use of their internal IMG175 130nm CMOS imaging process and for *Andrew Holmes'* help during the latter stages of the chip tape out. They also provided a number of pieces of equipment for the optical bench used in the characterisation of this receiver and were helpful when it was found that a number of the wafer runs resulted in very high dark count rates. A number of devices were provided from the low dark count rate sections of the wafers, giving better results than the first batch received.

I would like to acknowledge Dr Henderson, the group, the research institute, the university and the engineering and physical sciences research council, for the opportunity to work towards this PhD, along with the support of colleagues within the CMOS Sensors and Systems group. I'd like to personally thank Professor Ian Underwood who as my second supervisor has given me some invaluable advice. Both the research group and Ian have helped me look deeper into my subject, with colleagues knowledge of CMOS design and Ian's optics background both being tremendously helpful in the understanding of opto-electronic sensors and communications.

I am indebted to Dr David Renshaw and Professor Alan Murray, both of whom helped hugely in my interpretation of microelectronics, throughout my undergraduate, masters of engineering and my Ph.D. Professor Alan Murray especially has helped a great deal in my university career, through various snippets of advice, and some memorable and thought provoking lectures on circuit theory, electro-magnetics and transistors. Within the department, special gratitude goes to Dr Bruce Rae (now at ST Microelectronics) and Dr Katherine Cameron, who were indispensable at the start of my PhD, as well as Dr Justin Richardson, Dr Richard Walker, Dr Eric Webster and many others from the research group, who helped iron out a number of design, mathematical and system ideas. I'd also like to express my gratitude to Camilla Thomson from the Institute of Energy Systems, and Wafi Zuhdi, Krzysztof Nguyen, Neale Dutton, Andrew Cogman, Colin Chung and Dr Nikola Krstajic from the Institute of Micro and Nano Systems. A special mention within the department goes to Keith Muir for a great number of interesting technical discussions and supportive trips to the pub, he has been a good friend within the department during my PhD, and has shown me the trickiness of starting a company from PhD research and results.

A note to Victoria Stone (now at Bristol University Philosophy Department) who prompted me to think about wider philosophical, ethical and socio-economic implications. I have developed a deep interest in philosophy through the works of Hacking, Kuhn, Kant and many others, along with the philosophies of science, mathematics and engineering.

A big part of any work, is the role of family and here I'd like to thank both of my parents, and my brother Matthew, who's support, love and patience helped me though a number of tough

months. Their interesting questions and ideas, helped create a broader context for the work. Discussions with them prompted me to think about why the system may be important and novel, as well as the role of the nature of the particle being counted. My father has an interest in electronics, prompted by his specialisation in Neurophysiology particularly neural EEG and EMG signals. As such, he has prompted a number of interesting conversations regarding the neural 'spike-like' nature of SPAD output pulses. I would also like to thank my grandfather and a number of my uncles, who, being electrical engineers by trade, have inspired me into this discipline from a very early age. A special dedication goes to my long term companion and friend, Spock the cat who died in July 2012. He was a continuous source of undivided affection and comfort.

I am indebted to a number of close friends who have always been a source of reassurance, support and many happy times. In particular I'd like to thank Heather Minto, Steven Garrard, Gary Dorken, Lucy Montgomery, Laura Wake, Matthew Burke and Eleanor Glendenning. Alasdair Clarke, now at the School of Informatics (Edinburgh University), helped me look at communications and photonics from a more mathematical and statistical view. Further gratitude also goes to Joanne Dunwell, Harry Kirkham, Rachel Kelly, Andrew Moore, Louisa Miller and my loving and supportive partner Rebecca Price for their encouragement and many relaxing pub and concert excursions. Rebecca's love has been a source of encouragement and I'm indebted to her for manuscript review, feedback and the support of a strong home-life.

This work was funded by the Engineering and Physical Sciences Research Council (U.K), under the *Hypix* consortium. I'd like to personally thank Dr Tony Kelly and Dr Jonathan McKendry for recent help with modulation studies and optical testing of the manufactured device, using the fabricated μLEDs. Further thanks go to a few other members of the *Hypix* consortium. Professor Peter Blood, mentoring the project, has given me a few pointers throughout the later stages of the project. Likewise Katie Blaney of EPSRC gave me some good writing advice for this thesis, asking how the technical content can be related to real world engineering and society.

Throughout the course of this PhD there have been a number of tough months, however I have taken particular solace in my deep interest in the subject. The ability to spend time reading around a topic including philosophy of science, engineering, the quantum theory of light, or exploring the many circuits used historically or architectures and circuit techniques in the field, has increased my sense of awe and admiration for the subject. This has also been a particular draw for my study of a PhD. I have particularly enjoyed the interaction between the sciences, and the ability to quickly explore contemporary physics literature as easily as the engineering literature. I find it inspiring to read papers regarding photon precursors travelling at the speed of light regardless of a materials refractive index or books regarding the anthropological and socio-economic implications of ubiquitous computing and communications.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

- Supervisors:
  - *Dr Robert Henderson:* CMOS Sensors and Systems Group Leader
  - *Prof Ian Underwood:* Head of The Institute for Integrated Micro and Nano Systems

- Examiners:
  - *Prof John Thompson:* Institute of Digital Communications (IDCOM)
  - *Dr Steve Collins :* Department of Engineering Science, University of Oxford

Edward Michael Dennis Fisher

Date: .....

Contents

List of Figures

List of Tables

List of Mathematical Symbols

α_c	=	Specific Material Absorption Coefficient
α_{mod}	=	Amplitude of Modulation (in Counts)
A_g	=	Generic Amplifier Gain
AP_0	=	Input After-Pulse Counts (pre count competition)
AP_{act}	=	Actual Output After-Pulse Counts (post count competition)
$AQARSTOP$	=	Stop Signal for AQAR (switches to PQPR)
β	=	Average Optical Power (in Counts)
B_r	=	Bit Rate
BW_{avail}	=	Available Bandwidth
BW_{th}	=	Maximum Theoretical Digital Data Bandwidth (Data Rate)
bx^a	=	Power Law Fit, b = multiplier, a = exponent, x = base
c	=	The Speed of Light
C	=	Capacitance, C (Generic) or Count Rate (*similar to* C_{max})
C_{ac}	=	SPAD Diode Anode-Cathode Parasitic Capacitance
C_{cs}	=	SPAD Diode Cathode-Substrate Parasitic Capacitance
C_{as}	=	SPAD Diode Anode-Substrate Parasitic Capacitance
C_D	=	SPAD (Lumped Model) Diode Capacitance
C_L	=	TDM State Machine Counts Per Loop
C_{max}	=	Absolute Maximum SPAD Counting Rate (Hz)
C_{out}	=	Integrated Number of SPAD Counts over a Symbol
C_S	=	SPAD (Lumped Model) Parasitic Capacitance
$C_{SPAD}(t)$	=	Number of SPAD Detections occurring at time, t
D_{BER}	=	Number of Photo Detections Required for BER
DC_0	=	Input Dark Counts (pre count competition)
DC_{act}	=	Actual Output Dark Counts (post count competition)
DCR_{pc}	=	DCR Per Channel Per Symbol Period
DCR_{px}	=	DCR of a Pixel Grouping (All TDM Modes)
DCR_{total}	=	Total DCR (Sum of all enabled SPAD DCRs)
Δf	=	Difference in Frequency (Bandwidth)
ΔT	=	Measurement/Simulation Time Period
D_g	=	SPAD Guard Ring Width
Dia	=	Diameter (units = meters, m)
D_r	=	SPAD Diode Radius
D_s	=	SPAD Guard Ring to Guard Ring Separation

List of Mathematical Symbols

e	=	Euler's Number() = 2.718... (Constant)
$ePDE$	=	Effective Photon Detection Efficiency
E	=	SPAD Event (Generic)
E_{DC}	=	SPAD Event (Dark Count Initiated)
$E[]$	=	Expected Value (= *mean*)
$E[\Psi_{px}]$	=	Expected Photon Rate incident on a Pixel
$E(t)$	=	Expected Inter-Arrival Time
E_{max}	=	Maximum Electric Field Strength (over p-n Junction)
EO_{rec}	=	Received Incident Eye Opening
EO_{ref}	=	Reference Transmitted Eye Opening
η	=	Detector Efficiency (*Cova Model*)
η_{det}	=	Efficiency of a TDM Detector Unit
η_e	=	Effective Pixel Detection Efficiency
f	=	Frequency, f (Generic)
f_{mod}	=	Modulation Frequency
$f(x)$	=	Generic Function, f() of a variable 'x'
F_{max}	=	Modelled Maximum Count Rate over TDM Mode
F_{Ny}'	=	The Nyquist Frequency
FF_{max}	=	Maximum Theoretical Fill Factor
FF_{sq}	=	Fill Factor of Rectilinear Array (*Circular SPADs*)
$\Gamma()$	=	The Gamma Function ($\Gamma(n) = (n-1)!$)
h	=	The Planck Constant
Hz	=	Hertz (unit) [cycles per second]
i	=	Generic Iterator
$i^2_{n,in}$	=	Input Noise Current Per Unit Bandwidth
i_{ph}	=	Photocurrent
I	=	Current, I (Generic)
$I_{avalanche}$	=	SPAD Avalanche Current
I_s	=	SPAD Leakage Current
j	=	Generic Iterator
k	=	Generic Number (Use: Number of Photons)
K	=	Boltzmann Constant
λ	=	Optical Wavelength
λ_r	=	Poisson Rate Parameter
μ	=	Mean Value (Statistics) *or* 'MICRO' = 1x10^{-6}

List of Mathematical Symbols

n	=	Generic Number of Elements/Units
n_{ap}	=	Variation Noise of After-Pulses (*Cova Model*)
n_d	=	Variation Noise of Dark Count Rate (*Cova Model*)
n_p	=	Variation Noise of Photon Rate (*Cova Model*)
$\overline{n_p}$	=	Mean Number of Photons
N_0	=	OOK *Zero*-bit distribution mean
N_1	=	OOK *One*-bit distribution mean
N_a	=	Number of SPADs currently active in TDM Group
N_A	=	CMOS Acceptor Doping Concentration
N_{counts}	=	Number of Counts
N_D	=	CMOS Donor Doping Concentration
n_i^2	=	Silicon Intrinsic Carrier Concentration
$N_{photons}$	=	Number of Incident Photons
$N_{detections}$	=	Number of SPAD Detections
N_s	=	Number of Gain Stages in TIA
$N_{samples}$	=	Number of Samples Per Symbol
N_{SPAD}	=	Number of SPADs within a (Generic) Group
N_{TDM}	=	Number of SPADs within a TDM Group
N_{th}	=	OOK Threshold
p	=	Precision Value (*for asymptotic approaches*)
P	=	Power, P (Generic)
P^+	=	P-type (Acceptor) Silicon Doping
π	=	PI() = 3.141... (Constant)
$P(0\|1)$	=	Probability of a zero given a transmitted one
$P(1\|0)$	=	Probability of a one given a transmitted zero
$P(AP)$	=	After-Pulsing Probability
$P(AP)_{px}$	=	After-Pulsing Probability of a Pixel Grouping (All TDM Modes)
$P(E\|DCR_j)$	=	Probability of an Event given the DCR of the j^{th} SPAD
P_{event}	=	Total Probability of a (*Generic*) Event
Φ	=	Optical Flux Incident on a Surface
$\prod$	=	Product Series
ψ_0	=	Input Photon Counts (pre count competition)
ψ_{act}	=	Actual Output Photon Counts (post count competition)
$\psi(n)$	=	Number of Photons Hitting a Pixel at a Time $t = n$
$\Psi(\tau_b)$	=	Number of Photons Within Symbol Period

List of Mathematical Symbols

$P(Roll-Over)$	=	The Probability a Digital Counter Rolls-Over (to Zero)
$P(SC)$	=	Probability of Spatial Coincidence
$P(TC)$	=	Probability of Temporal Coincidence
$P(TC)_{px}$	=	Temporal Coincidence at Pixel NOR Gate
$P(TC)_{na}$	=	Temporal Coincidence at Pixel NAND Gate
PP_{DCR}	=	Power Penalty: Dark COunt Rate
PP_{EOP}	=	Power Penalty: Eye Opening Penalty
PP_{FF}	=	Power Penalty: Optical Fill Factor
PP_{ISI}	=	Power Penalty: Inter-Symbol Interference
PP_{jitter}	=	Power Penalty: Time Jitter
PP_{PDE}	=	Power Penalty: Photon Detection Efficiency
PP_{RIN}	=	Power Penalty: Relative Intensity Noise
PP_{tot}	=	Power Penalty: Total (*Summation*)
$P_R(\sigma)$	=	Receiver Sensitivity at σ Deviation from Sampling Point
q	=	Electronic Charge
Q	=	Personick BER Q (*Quality*) Factor
R_D	=	SPAD Diode Internal (Series) Resistance
R_L	=	Passive Quenching Load Resistance
R_f	=	Trans-Impedance Feedback Resistance
R_{int}	=	Intensity Fluctuations (Relative Intensity Noise)
R_{on}	=	Transistor *ON* Resistance
R_s	=	Symbol Rate
R_{TIA}	=	Trans-Impedance Series Resistance
σ	=	Standard Deviation
σ_0	=	Standard Deviation of the OOK N_0 Distribution
σ_1	=	Standard Deviation of the OOK N_1 Distribution
$\sigma_{\Delta t}$	=	Standard Deviation of Time Jitter
$S_{in}(t)$	=	Sinusoidal Input Signal as Function of Time
t	=	Time (units, seconds, s)
T	=	Temperature (units, Kelvins)
T_{enb}	=	Enable Period of a SPAD
τ_b	=	Symbol or Bit Period
τ_d	=	SPAD Dead Time (Inactive Period)
τ_p	=	Edge Detection Pulse Period
τ_q	=	SPAD Quenching Time Constant (*Passive Mode*)
τ_r	=	SPAD Recharging Time Constant (*Passive Mode*)

List of Mathematical Symbols

$v_{n,rf}$	=	Feedback Resistance RMS Johnson Thermal Noise
V	=	Voltage, V (Generic)
$Var()$	=	Variance
V_{bi}	=	p-n Junction Built-In Potential
ε_0	=	Vacuum Electric Permittivity
ε_s	=	Relative Silicon Electric Permittivity
VBD	=	SPAD Breakdown Voltage
VEX/Vex	=	SPAD Excess Bias Voltage
VOP	=	SPAD Total Reverse Operating Bias Voltage
V_t	=	Transistor Threshold Voltage
$\frac{W}{L}$	=	Transistor Width and Length (Aspect Ratio)
W_J	=	Width Of p-n Junction
x_d	=	Depth into Material
ξ	=	PTC Non-Linear Multiplier

List of Abbreviations

AP	=	After-Pulsing
APD	=	Avalanche Photodiode
AQ	=	Active Quench
AQAR	=	Active Quench Active Reset
ASIC	=	Application Specific Integrated Circuit
ASK	=	Amplitude Shift Keying
A/W	=	Amps Per Watt (*Photodiode Sensitivity*)
BER	=	Bit Error Rate
BERT	=	Bit Error Rate Test
Bi-CMOS	=	Bipolar CMOS Technology
bps	=	(*Digital*) Bits Per Second
BW	=	Bandwidth
CDF	=	Cumulative Distribution Function
CDR	=	Clock and Data Recovery
CLK	=	Clock *Digital*
CML	=	Current Mode Logic
CMOS	=	Complementary Metal Oxide Semiconductor
CW	=	Continuous Wave
DAC	=	Digital to Analogue Converter
dB	=	Decibel (logarithmic units)
DC	=	Direct Current
DCM	=	Digital Clock Manager
DCR	=	Dark Count Rate
DSP	=	Digital Signal Processing
DT	=	Dead Time
DTCTRL	=	Active/Passive Dead Time Control Voltage ($DTctrl$)
DTI	=	Deep Trench Isolation
EO	=	(Data) Eye Opening
EOP	=	Eye Opening Penalty
EMI	=	Electro-Magnetic Interference
FECC	=	Forward Error Correcting Codes
FF	=	Fill Factor
FIFO	=	First In First Out
FOM	=	Figure of Merit
FPGA	=	Field Programmable Gate Array
FSO	=	Free Space Optical (*link*)

List of Abbreviations

FSM	=	Finite State Machine
FWHM	=	Full Width Half Maximum
GaN	=	Gallium Nitride
Gbits/s	=	Giga ($1x10^{+9}$) bits per second
GBW	=	Gain Bandwdith Product
GM-APD	=	Geiger-Mode Avalanche Photodiode
GUI	=	Graphical User Interface
HDL	=	Hardware Description Language
HS	=	High Speed *CMOS Logic*
IEEE	=	Institute of Electrical and Electronics Engineers (USA)
I/O	=	Input / Output
ISI	=	Inter-Symbol Interference
IP	=	Intellectual Property
LAN	=	Local Area Network
LPS	=	Low Pass Filter
LSB	=	Least Significant Bit
LVDS	=	Low Voltage Differential Signalling
Mb/s	=	Mega ($1x10^{6}$) Bits Per Second (Data Rate)
MIMO	=	Many In Many Out
MOSFET	=	Metal Oxide Semiconductor Field Effect Transistor
MPAM	=	Multi-Level Pulse Amplitude Modulation
MSB	=	Most Significant Bit
MUX	=	Multiplexer
NRZ	=	Non-Return to Zero
OECF	=	Opto-Electrical Conversion Function
OFDM	=	Orthogonal Frequency Division Multiplexing
OK	=	Opal Kelly
OOK	=	On Off Keying
PAM	=	Pulse Amplitude Modulation
PC	=	Personal Computer
PCB	=	Printed Circuit Board
PD	=	Photo Diode
PDE	=	Photon Detection Efficiency
PLL	=	Phase Locked Loop
PMT	=	Photo Multiplier Tube
PNR	=	Photon Number Resolution

List of Abbreviations

POF	=	Polymer Optical Fibre
PPM	=	Pulse Position Modulation
PQ	=	Passive Quench
PQPR	=	Passive Quench Passive Reset
PRNU	=	Photo-Response Non-Uniformity
PTC	=	Photon Transfer Curve
PWM	=	Pulse Width Modulation
QL	=	Quantum Limit
RC	=	Resistor-Capacitor (*Circuit*)
RIN	=	Relative Intensity Noise
RTN	=	Random Telegraph Noise
RZ	=	Return to Zero (*Modulation*)
SC	=	Spatial Coincidence
SDM	=	Space Division Multiplexing
SI	=	Serial Interface
SiGe	=	Silicon Germanium
Si-PM	=	Silicon Photo Multiplier
SMA	=	Sub-Miniature version A Connector
SML	=	Spatially Modulated Light (*Photodiode*)
SNR	=	Signal to Noise Ratio
SOI	=	Silicon-On-Insulator
SPAD	=	Single-Photon Avalanche Diode
STI	=	Shallow Trench Isolation
TC	=	Temporal Coincidence
TCSPC	=	Time-Correlated Single Photon Counting
TDM	=	Time Division Multiplexing
TIA	=	Trans-Impedance Amplifier
USB	=	Universal Serial Bus
VBD	=	Break Down Voltage
VCO	=	Voltage Controlled Oscillator
Vex	=	Excess Bias Voltage
VLC	=	Visible Light Communications
Wi-Fi	=	Wireless-Fidelity
μLED	=	Micro Light Emitting Diode

Chapter 1

Introduction

1.1 Optical Communications and its Application

Optical technologies are now routinely used in inter-continental communications, where both data rates and transmission lengths are high, or in fast networking applications such as data centres or large offices [1, 2, 3, 4]. At the opposite end of the transmission distance scale, data links in microelectronics are moving towards photonic communications [5, 6, 7, 8, 9]. This is due to the bandwidth restrictions and area necessary for traditional metals on silicon chips [10].

Between these extremes, local area networks (LANs) need to expand available bandwidth due to media streaming technologies and the rapidly increasing number of internet connected devices [1, 11, 12]. Optical networking in both home and office environments has also been introduced to promote increased energy efficiency [13, 14]. However the literature indicates that further work is needed to achieve the optimum balance between power efficiency, data transmission performance and manufacturing costs [1].

1.1.1 Current Wi-Fi and Future Optical LAN Networks

Radio-based *wireless* communications, often used in home or office LANs, are restricted by the use of a small slice of the heavily regulated electromagnetic spectrum [15, 16, 17, 18, 19, 20]. While carrier frequencies and link bandwidths can increase, other frequency band users or the design of the transmission systems become limiting factors [1].

Optical *wireless* networks, called visible light communications (VLC) or free-space optical (FSO) links [21], have attracted interest as the spectrum is not licensed. As such, they may obtain a greater aggregate bandwidth and have advantages due to signals being confined to the local target area, ensuring better security than radio frequency signals [19, 20, 22, 23, 24, 25].

VLC signals can be communicated using modulation of the light from solid-state, energy efficient lighting [16, 17, 22]. VLC techniques may however require receiver optical filters to remove un-modulated light. Furthermore, as fast modulation is required, often using lower bandwidth transmitters [17], the modulated signal may be of a low amplitude. A particular issue with home and office VLC, is that for illumination and wide signal spread, the optical sources are often divergent. For a given optical power at the source, the optical power received

on a small area receiver chip is quickly reduced as the optical path expands [17, 18, 20]. VLC (Figure 1.1) therefore requires extremely sensitive receivers able to correctly sense the transmitted signals [18].

Transmitting multiple colours of light in VLC applications, a technique known as wavelength division multiplexing [2, 3], increases the total bandwidth. For illumination [22], white light is produced by mixing many different colours: using red, green and blue transmitters is one such technique. However the receivers need to have narrow colour sensitivity, allowing separation of the transmissions. This requirement reduces the optical intensity at the receiver due to filtering, or, for a given room lighting intensity and number of transmission wavelengths, reduces the optical power per transmission.

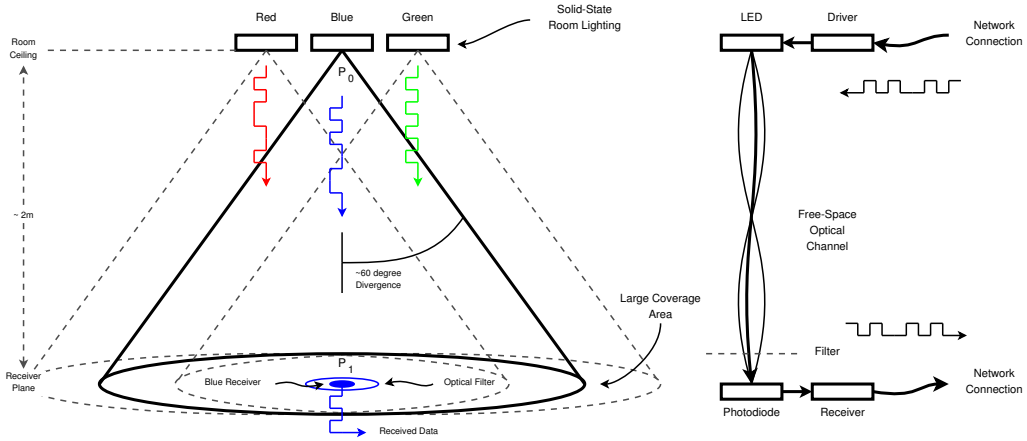


Figure 1.1: A simple representation of free-space optical (FSO) visible light communication (VLC). The divergent beam from a set of LED transmitters is incident on a large receiver plane. The power at a small receiver device is much lower due to filtering and a small capture area.

1.1.2 Receiving Optical Signals

The low-level devices and methods used to receive optical transmissions have been proven to be both reliable and practical and are used as a start for this work [4, 26, 27, 28]. In particular photodiodes (PDs) and avalanche photodiodes (APDs) have both been explored in a variety of complementary metal-oxide-semiconductor (CMOS) receivers [26, 29, 30] for various applications [13, 31, 32, 33, 34]. As with the financial, technical and efficiency costs of achieving reliable high speed communications, the PD and APD technologies currently used also incur costs. These may be increased planar material area, diode bandwidth restrictions, the use of separate chips or depending on the requirements the costs of non-CMOS technologies such as Bi-polar processes. A cost-benefit analysis approach must be taken during design to ensure economic-effectiveness is maintained.

One of the low-level trade-offs that directly affects the energy efficiency, is the parasitic capacitance of a photodiode [2, 31, 35]. In order to obtain a high bandwidth, and hence a high data rate, the photodiode should be small [2, 14, 26, 36]. However a small photodiode gives a correspondingly small photo-current, requiring amplification and the signal to noise ratio of the system may be too low to achieve the target reliability [1, 2, 3, 35].

To ensure both the target data transmission and bit error rates are met, the optical power incident on the photodiode can be increased [31, 37, 38, 39, 40]. This however gives a high overall communications link budget [1]. The link budget is the total sum of the energy needed for encoding, transmitting, overcoming losses within a transmission medium, receiving and decoding the data. Consequently, research is continuing to increase the diode bandwidth and sensitivity [4, 27, 41, 42, 43, 44, 45, 46]. Circuit techniques such as pre-emphasis [35] and equalisation [13, 47, 48] can also be used but increase the complexity, silicon area and power requirements of a receiver.

Avalanche photodiodes [4, 26] alleviate some of the amplification requirements by using the avalanche multiplication effect. This effect moves part of the amplification process into the diode [26] and could reduce the external gain required and therefore the electrical power needed. This increase in photodiode gain can also be used to decrease the required level of incident light, reducing the electrical power needed at the transmitting light source. The use of avalanche multiplication is being increasingly used in receiver design to offset electrical amplification needs. In this work, no radical departure from this trend is proposed, rather the work aims to investigate the use of increased avalanche multiplication. Not only might this help in reducing the external gain needed, as per the primary advantage of current APD systems, it may also add new advantages in the form of photon counting capabilities [49] and the efficient transfer of the electrical signal into the digital domain [49, 50].

1.1.3 Single-Photon Sensitivity for Optical VLC Communications

In this work, advances in CMOS-based Single-Photon Avalanche Diodes (SPADs) are utilized, further exploiting the avalanche gain mechanism used in current APD based receiver systems [49, 50, 51, 52, 53]. While this may require increased electrical power at the receiver, single-photon sensitivity may be suitable in applications where the electrical power needed by a conventional receiver is greater than that required for SPAD circuits. The advantage may be that moving from APDs to SPADs through a higher bias voltage and a higher pn junction electric field, has a higher electrical power efficiency for a given gain, than a similar increase in an external amplifier's gain. Visible wavelength SPADs are particularly attractive as they can be implemented in standard CMOS manufacturing processes [54]. The purpose of this is to increase integration of receiver functions [32], reducing the overall manufacturing costs by implementing single-chip solutions [52, 53] that can use mass production. The noise rates of the diodes are reducing with maturity [53, 55], and the density of SPAD arrays has increased

[56, 57, 58, 59]. The bandwidth of SPADs is limited in comparison to APDs, but as the inactive period following a detection becomes shorter, single-photon count rates have been increasing [49, 60].

This work aims to assess the feasibility, and design factors involved in the use of single-photon receivers which use the already successful avalanche gain mechanism. This will provide exploration into optical power reduction techniques for next generation optical local area network communication receivers. The characteristics and performance of a fabricated device are investigated and the receiver is compared to prior art.

1.2 The Motivation for this Work

The research presented was performed to continue the trend of using avalanche multiplication within communications receivers [41, 42]. This technique which has been successfully used in APD receiver systems, increases the total amplification gain, and provides a higher receiver sensitivity than photodiodes [4]. This project therefore aims to prove the concept and assess the feasibility of free-space VLC using multiple single-photon sensitive detectors. This study focuses on establishing whether single-photon receivers can be used to lower the transmitter optical power, thus saving energy in the overall link budget, while maintaining robust detection of optically encoded signals. Rather than stepping away from APD systems, the aim here is to progress the use of their gain mechanism to higher levels to investigate what further advantages this could bring.

Single-photon devices have been previously used in the communications context [61, 62, 63]; with recent work using CMOS SPADs to achieve inter-chip communications using a pulse position modulation scheme [64]. The work presented here assesses the performance and any limiting factors that may apply for implementing free-space VLC systems [22, 65, 66] using on-off modulated LEDs [23]. A feasibility study investigating the use of analogue silicon photomultipliers, which use large arrays of single-photon detectors, has shown merit in their use in communications [67], while the basic effect they operate on, that of avalanche multiplication, has been progressed successfully yielding significant advantages over traditional photodiode receivers . In this work an all-digital approach is taken, utilising an integrated CMOS process [68, 69] and fully exploiting the *direct-to-digital* nature of CMOS SPADs [64].

Conventional receivers such as photodiode and avalanche photodiode receivers, have been studied allowing full optimisation of performance prior to fabrication. Numerical and analytical models are available to estimate the bandwidth and noise of a photodiode and trans-impedance amplifier (TIA), or the diode capacitive impulse response [13, 31, 35, 26]. These can be used during the prototype design phase, complementing the existing communications theory [1] and link specifications, to enable a prototype that is close to the required performance.

This thesis assesses the performance of single-photon receivers using an experimental receiver, which has been fabricated in an advanced process [30], in order to study real world constraints. Using the experimental receiver, a particular driving force for this work is to explore optimisation of the new receiver, assessing if it is possible to predict the performance during design. To do this, all the factors must be investigated including those initially thought to be unimportant. The end aim of this, with continued future work, is to obtain numerical and analytical models similar to those of conventional systems. This would again allow designs to be as close as possible to their required performance, without too many costly design revisions.

1.3 The Research Objectives

A number of research objectives, listed below, are designed to complement micro-LED transmitter research conducted during the Hypix project [23, 24, 65, 66]. The main objectives of this work are:

Objective 1: To develop an integrated circuit pooling the signals of planar single-photon sensitive photodiodes, in order to tend towards a receiver capable of achieving the specified data and bit error rates, whilst achieving a sensitivity greater than comparable conventional receivers.

Objective 2: Develop printed circuit boards (PCBs), firmware and software to control the circuit, decode the output, adapt it for optimum performance and allow experimental investigation.

Objective 3: Establish the performance of the receiver for data reception, demonstrating link performance with the Hypix project micro-LED (μLED) transmitters.

Objective 4: Model the device to investigate its theoretical function, and experimentally observed performance.

Objective 5: Establish important parameters, linking measurements and models to indicate improvements for a second generation.

1.4 Contributions to Knowledge

A number of contributions were made during the progress of this work, however some work performed has been excluded for clarity and brevity reasons.

1. Firstly, the SPAD-based receiver fabricated during this project is the first use of an all digital CMOS photomultiplier [70, 71] in optical communications [68, 72]. This has been used to explore the potential routes towards increased receiver sensitivity and lowering optical transmission power consumption, that are offered by the increased use of the avalanche gain mechanism.
2. A second contribution is to investigate device behaviour limiting the bit error rate (BER) of the receiver, with a brief assessment of the suitability of mitigation techniques. The step and impulse responses, along with other receiver sensitivity power penalties [2, 3, 14, 36], reduce the photon count separation between received digital bits, giving increased errors.
3. The final contribution of this work, is a number of presented models, such as models of a) receiver responsivity curves, b) step responses and settling points, and c) the action of the circuit upon the low level SPAD performance metrics. These investigate unexpected receiver characteristics and give insights into improving the design of future receivers, optimising designs for specific applications or data and bit error rates.

1.5 Outline of Thesis Chapters

Chapter 2: The second chapter introduces communications [1] and specifically the transition to optical communications [2, 3, 13, 35]. A literature review discusses important affects contributing to the performance of receivers. The different receiver architectures are then discussed with an integrating method [35, 73] focused upon to increase sensitivity. To further elevate the sensitivity, avalanche multiplication gain [26, 74] that exceeds that of conventional APD receivers, has been suggested. Therefore the literature and significant affects of single-photon avalanche diodes [49, 50, 53, 75, 76] are introduced and a possible advantageous multi-diode approach [21, 41, 43, 68, 72, 77, 78, 79] is proposed with respect to SPAD noise sources and non-ideal behaviour.

Chapter 3: Chapter 3 presents models from the literature, which are explored to obtain an initial specification for the receiver [14, 36, 80]. While these are approximate and must be expanded upon with experimental work, they complement the specifications and match the development of the micro-LED transmitters. The chapter culminates, in a set of specifications that can be taken as the design parameters for the receiver prototype fabricated in this work.

Chapter 4: In Chapter 4 a receiver using a SPAD array is fabricated in a standard 130nm CMOS process. A *design for test* approach [69] is used to allow adaptation of circuits within the design. This allows flexibility for characterisation of the array and communication link experiments with the micro-LED transmitters [23, 24]. The chip is designed to operate as the front-end of the receiver only, where clock and data recovery, robust clock generation using phase-locked-loops and communications protocols were envisaged in the follow on project after the Hypix program [66].

Chapter 5: Low level characterisation is investigated in Chapter 5. Here the electrical power consumption, the SPAD dead times and the noise rates are measured allowing their impact to be evaluated. The receiver responsivity [81] is also measured to explore any non-linear effects. Complementing this, a model of the responsivity investigates competition between the different avalanche triggers. This suggests the noise count rate may change with incident optical power. These low level characteristics can then be used when assessing overall performance and experimental communication results.

Chapter 6: Results for the transient responses, important for communications, are presented in Chapter 6. These show behaviour that limit the achievable BER for the target data rate. The step response has been modelled and contributes to the final conclusions. The chapter also presents communication between micro-LED transmitters and the receiver, with analysis of the BER extended to analyse multi-modal effects measured in a real optical system. Finally, figures of merit compare the receiver with prior art, giving favourable outcomes. The results, models and figures of merit then suggest modifications for a second generation receiver design.

Chapter 7: The final chapter, summarises the thesis and concludes with further work that could improve the performance of the receiver designed here. Fundamentally, the breakdown and excess bias voltages of SPADs lead to a high receiver electrical power per bit, but enable single-photon sensitivity that could alleviate optical power requirements. SPAD receivers may be useful for niche applications, but the adoption of the presented receiver is restricted due to electrical power, silicon area and error rate limitations. Single photon sensitivity would however be suitable when the electrical power needed is lower than that required for the necessary gain in a conventional receiver. The overall outcome of this work, is that while the presented receiver works to a point, the observed issues must be investigated and corrected through continued research, prior to any high performance design or commercialisation can be considered.

Chapter 2

Literature Review And Design Specifications

2.1 Chapter Introduction

This chapter discusses optical communication [1, 2, 3, 14, 36] and its application to *wireless* local area networks for homes and offices, along with introducing the use of optical networks for medium distance communication [16, 17, 20]. This work focusses on visible light communications (VLC), where transmitters may be modulated at rates faster than their bandwidth, or where optical filters are needed. In this bandwidth limited or optically filtered situation, the signal amplitude is small and receivers are required to be highly sensitive (Section 1.1.1). The literature review within this chapter, focuses on increasing the receiver sensitivity, firstly through different architectures [35] and secondly through the use of avalanche multiplication [26]. One of the architectures, the integrating receiver [35], is chosen for this work. This gives good sensitivity and is suitable for digital implementation.

Avalanche multiplication [26] enhances the sensitivity, as shown within prior art. Single-Photon Avalanche Diodes [49] are introduced, which exploit this gain mechanism. The benefits and limitations of SPADs are introduced, informing the specification of a receiver (Section 2.9). While SPADs have been used [62, 64, 67], the combination of a specific receiver architecture, on-chip SPADs and a digital implementation, is an area of research that has not yet been fully explored. This work examines a receiver architecture that:

- is suitable for 32x32 μLED transmitter arrays [66] and
- includes scope for circuit testability and experimentation within a communications environment.

2.2 Optical Communication: An Introduction

The demand for communication resources continues to grow in the form of higher data rates and lower error rates [1, 2, 3, 12, 14] (Figure 2.1 [82]). The demand is driven by the increase in the number of connected devices, growth of the accessing population, and the trends towards both real time media streaming and access to cloud-based computing [12, 16].

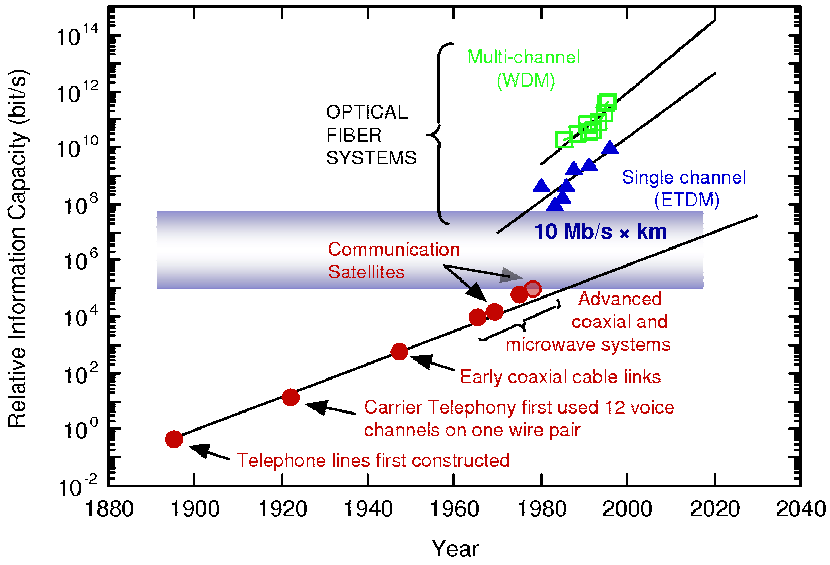


Figure 2.1: An overview and evolution of communications, including Electrical Time Division Multiplexing (ETDM) from MIT Microphotonics Center Industry Consortium 2005.

There are four types of communication network. These are a) electrical communications though wires (electrical wire-line), b) electrical communications through radio-based links (electrical wireless), c) optical communications through fibres (optical wire-line) and d) optical communications through free-space (optical wireless) [1, 3, 21, 32, 35, 83].

1 Wired electrical links have become a mature technology allowing high-speed, reliable communication and have the advantages of privacy and non-line-of-sight communication [1, 35, 84]. Whilst mature, these systems have fundamental limitations due to phenomena such as the skin effect and interference [1, 69, 85, 86], which limit bandwidth and give rise to errors respectively. Material and system costs along with complexity [1, 47, 69, 84], also become issues when operating at multi-Gigabit/s rates, as does energy use.

2 Radio communication is routinely used locally for lower rate communications, often at multi-subscriber distribution points [1]. Radio and Wi-Fi links allow systems to be wireless, reducing physical routing complexity and decreasing material costs for multi-channel, short to medium distance links. However, radio links are limited due to the increasing number of subscribers [11, 12] and the heavily regulated radio spectrum [1, 15, 16].

3 There has been a transition to optical fibre communications for long haul (10–100km) links, with this strategy becoming a mature technology [3, 4, 13, 87, 84]. These systems achieve low electrical energy per bit [2] transmitted with 100's of Gb/s data rates and error rates low enough ($1x10^{-12}$) to assume error free transmission [1, 3, 14, 36]. Optics has been adopted due to the bandwidth, complexity and efficiency issues of electrical links; namely the data, error and development rates that are required by end applications [1, 3, 13, 14, 35].

4 Historically free-space optical communications has been used for millennia in the form of fire beacons, smoke signalling or semaphore [1, 13]. Line of sight free-space optical links have also been previously used as an extension of point to point radio links [21]. High speed laser-based links are now also commercially available [88, 89, 90], offering bandwidth and noise advantages over radio implementations, but suffer from the issues of all free-space optical techniques [1].

Visible light communication (Figure 1.1) has become an option for continued bandwidth scaling [16, 17, 20, 22]. This is required with the increased use of internet connected devices, restrictions on the radio spectrum and the increase in high data rate, real-time streaming of media content [1, 11, 12, 15, 91].

VLC is not intended to supersede radio networks, it is rather designed to complement them, providing the download bandwidths required within the home/office and public environments [16, 17]. When energy reduction is a specific driver, receivers must maintain targeted data and error rates while using lower transmitter powers [4, 13, 31, 92]. This work focuses on high-sensitivity receivers for VLC, complementing micro-LED (μLED) research [23, 24, 93] that follows modulated solid-state lighting trends [16, 17, 20, 22, 94].

VLC, recently ratified by the IEEE (IEEE 802.15.7) [22], has a number of requirements. While speed and reliability are called for, privacy of communication similar to that offered by electrical constrained links is needed. Likewise, *wireless* communication is advantageous for signal routing and lowering material costs. Optical links offer an alternative way to achieve these aims, with few spectrum restrictions, larger bandwidth and no electro-magnetic interference [16, 84]. Increasing data volume and subscriber numbers cause an increase in overall energy usage [11, 12], therefore a critical requirement is a reduction in energy per transmitted bit [1, 13, 31]. The combination of VLC transmission with efficient, long lifetime, solid-state lighting gives a possible way to achieve this reduction in energy use, but requires further research before the energy per bit and data robustness can be comparable to state-of-the-art long haul links [16, 17, 18, 20, 94, 95]. This thesis therefore focuses upon the use of efficient, solid-state visible lighting sources [17, 22], which can be modulated at rates approaching 100Mb/s per LED, and the use of highly sensitivity SPAD-based receivers.

2.3 Key Communication Performance Metrics

A number of metrics characterise communication systems [1, 2, 3, 14, 35, 36]. These are:

2.3.1 The Data Bit Rate

The bit rate, B_r, is the absolute transfer speed [1, 3], measured in binary bits per second (bps). This is a function of the transmitting clock frequency and the modulation technique [1, 2, 3]. For the on-off key (OOK) method [1] used with the μLEDs [23, 24], the symbol rate R_s, which is the number of symbols per second, is equal to the bit rate [1].

2.3.2 Link Bandwidth

To obtain the data rates specified by an application, the bandwidth of a system must be high enough to allow robust data transfer [1]. The bandwidth is a measure of the range of frequencies that can be transmitted. Digital transmission uses square pulses containing a range of frequencies [1, 2, 3]. Frequencies higher than the data rate are included producing clean fast transitions. If these components are removed, the transition points become increasingly smeared over multiple symbol time slots [1, 2, 3, 35, 36]. Despite needing a bandwidth appropriate for the frequency spectrum of the transmission, a bandwidth lower than the data rate may be advisable to reduce the noise frequencies that can corrupt a signal [1]. Depending on the transmission and the receiver implementation, the overall bandwidth may require increased receiver complexity, in the form of equalisation techniques [1, 2, 13, 14, 35].

2.3.3 On-Off Key Modulation

On-Off Key modulation is a simple modulation coding scheme where a binary one is indicated by the presence of a pulse during the symbol period τ_b [1, 2, 3, 14]. A binary zero is likewise indicated by the absence of that pulse [1]. The scheme is a specific form of amplitude shift keying (ASK) [1, 3], where a transmitter shifts between two signal amplitudes to convey binary data. Coding using the presence or absence of a pulse represents a 100% extinction ratio, measuring the ratio between the two amplitudes [1, 2, 14, 36]. For the work presented here, a 100% extinction ratio or modulation depth is assumed for theoretical work as it represents an ideal OOK transmission and has been used in the design and subsequent experimental work with μLED transmitters [23, 68]. In realistic systems a source may not be able to fully turn off [23, 24], either due to the *extrinsic* bandwidth or non-ideal high-speed *intrinsic* bandwidth [1, 2, 14, 36, 44]. The effect of not turning fully off can be accounted for using the eye opening power penalty [14, 36].

A single-bit digital signal can be mapped directly to an optical symbol via a simple driver circuit, enabling dense arrays of μLED transmitters [23]. The design of these circuits [23, 96] aims to achieve full (100% extinction ratio) on-off modulation at high rates for communication and biological fluorescence lifetime applications [23, 24, 66, 96]. With such a transmitter being designed during the project [65, 66], and indeed achieving pulse *turn-off* edges suitable for SPAD-based biological auto-fluorescence imaging [96], the optical signal incident on the receiver is assumed to be ideal for the processes of a) establishing specifications, b) modelling and c) receiver design.

For the μLED transmitters used [23, 93, 97, 98], the prospective use of finite counting rate single-photon diodes [49] and the attenuation and distortion properties of either free-space [17, 20, 22] or large core diameter polymer optical fibres [99, 100, 101], the non-return-to-zero (NRZ) variant of the OOK scheme is desirable [1]. Further, as the project specifications point to OOK modulation [23, 65], the design of a VLC receiver will use this modulation format as a primary design parameter. The project's choice of using OOK modulation not only refers to the driver circuity previously developed for the μLEDs, but also to the likely output performance of the Hypix project organic polymer laser materials [65, 66]. As such, the receiver designed here will not be directly suitable for other modulation types. Despite this restricted design remit, in comparison to a more general receiver, other schemes can be used with the concept [1], if appropriate design specifications are followed throughout the receiver design.

2.3.4 The Bit Error Rate

The bit error rate (BER) is a property of the transmission, medium and receiver system, measuring or specifying the probability a bit is incorrectly received [1, 2, 3, 14, 36]. In a long transmission, erroneous bits can be detected and corrected [1], however low noise and low interference links minimise the native error rate [3]. This ensures forward error correcting codes [1] or data link protocols [1, 102] do not need to be used as they increase power and complexity requirements.

A variety of BER performance levels are conventionally used [1, 14]. These are influenced by the application, speed and scope for error correction. Inter-continental links require error rates of just one erroneous bit in a transmission of one trillion, a BER of $1\text{x}10^{-12}$ [6, 9, 84]. VLC applications however operate at lower data rates and may tolerate error rates of $1\text{x}10^{-6}$ [17, 20, 22]. The overall system can achieve near error free transmission when multiple error correction techniques are used together. While the native error rates in VLC systems may be higher than those of inter-continental links, lower speeds allow correction circuits to operate without complex computation [22]. Thus as long as the native BER is better than the $1\text{x}10^{-3}$ error correction limit [1, 14, 36], error free transmission can be achieved at the system level if these multiple methods are combined [102].

In this work, the target BER is between $1x10^{-6}$ and $1x10^{-9}$ (Table 2.1 on page 33). This is, in part, a product of the Hypix project specifications [66], but also that a transmission link that is dominated by the transmission medium is hoped for, rather than one that is dominated by the receiver. With the performance of a new receiver being as-yet-unknown, performance dominated by the receiver may occur, however the target BER can be a useful guide in the design stage of a new receiver. By designing the transmission and receiver electronics for a low BER, the practical performance of the system would be dictated by effects such as multiple paths in the optical transmission [18] or non-uniform atmospheric conditions [41]. Despite operating at much lower data rates than $1x10^{-12}$ BER inter-continental links, and despite error correcting circuitry being suitable for the clock speeds used by VLC systems, the Hypix project's scope includes only the front-end transmitting μLEDs, simple front-end driver circuits and the front-end receiver circuitry [66]. This is because its focus is on lasing organic polymers with CMOS controlled μLED optical pumping. More advanced driver circuits, transmitter error correction logic, transmission protocols and receiver back-end functions such as clock and data recovery [2], are envisaged in the follow on program to the Hypix project (UP-VLC).

2.3.5 Electrical and Optical Noise and Interference

There are various unwanted signals that contribute to a transmission [1, 2, 3]. These are generally additive and have the effect of limiting the error rate that a communication system can achieve [1, 13, 14, 35, 85, 103]. Electrical noise impacts currents and voltages within transmitters, receivers and, if used, within wire lines [1, 35, 85]. It may be caused by the natural thermal noise (Johnson Noise) of resistances within the circuit [14, 31], non-ideal leakage or noise currents [35, 69] within active devices such as MOSFETs, feedback loops within a system or coupling of a signal from one part of a circuit to another [85], for example clock feed-through. While strictly interference rather noise, broad emission electromagnetic interference (EMI) from nearby equipment [85] can also impact the error rate. In this case, frequencies such as 50Hz mains electrical hum, may be capacitively or inductively coupled into the circuit, thereby changing the position of the received signal amplitude relative to the OOK decision threshold. Electrical noise is present in all systems, however the degree that it impacts the error rate is dependent on the system architecture, and the signal-to-noise ratio (SNR) [1, 14, 36]. Electrical noise and electrical interference can be reduced by a number of techniques [1, 85], however each has disadvantages [35].

Noise is also present within the optical domain [3, 41, 87]. This has a fundamental limitation given by the natural Poission variation of light [80, 104], the so called Quantum Limit (QL) [1, 2, 3, 14, 36]. The Poission variation of light refers to the fluctuation in the number of photons received within a specific time period [80]. The variation represents the lowest noise possible in a system, with sub-Poission sources being particularly difficult to create, operate and modulate [80, 105]. Ambient light, multi-path propagation and atmospheric turbulence

give further optical noise in VLC applications [18, 20, 22]. These often dominate Poission variation, but can be reduced by filtering, design or computational techniques [20].

In the ideal case, communication of distinct transmission levels gives static output amplitudes. In a more realistic case, all encoded symbols are received with an amplitude that varies around some mean level [1], with a particular symbol giving a slightly different amplitude each time it is received. Random noise and interference sources, whether they are electrical or optical, broaden these distributions thereby increasing the rate of erroneous bits [2, 103].

2.3.6 Inter-Symbol Interference

A transmitted symbol $x(N)$, may interfere with previous, $x(N-1)$ and subsequent, $x(N+1)$ symbols [1, 14]. This is called inter-symbol interference (ISI) and adds a further departure from the ideal Poisson-limits of OOK modulation [1, 2, 3, 14, 35, 36]. ISI may be caused by operating at a speed greater than the bandwidth of the system, where an ideal rectangular pulse becomes smoothed over multiple symbols [14, 94]. It can also be caused by delays within a channel, where the receiver may observe the summation of both the current symbol, $x(N)$ and the previous, $x(N-1)$ symbol. If the level of interference is severe, it may add other peaks into the received amplitude histogram, quickly increasing the error rate [3].

ISI can be caused through non-ideal transmitter, non-ideal channel and non-ideal receiver properties [14, 36]. Channel properties such as multiple path propagation in VLC are often dominant and difficult to characterise or remove [18, 20, 21, 41, 43, 106]. Electrical transmitter or receiver properties that also give non-ideal temporal characteristics, such as the step response [1] or voltage variation in the electrical power supply [14, 35], can be made to be negligible through multiple design iterations [69]. However these effects must be characterised and their cause understood before a second prototype can attempt to remove their influence. The forward-in-time interference to a second symbol equates to the amplitude of the primary pulse at the sampling point of the second symbol [1, 14]. ISI can be corrected using pre-transmission schemes such as raised cosine filtering [14]. These work well but require prior knowledge of the system response, including bandwidths and impulse and step responses of all paths in the signal chain [14].

2.3.7 The Ideal Receiver and The Quantum Limit

An ideal receiver is an abstraction for design and analysis [14]. It has no negative traits, such as noise, and receives an encoded signal with an efficiency of 100% [1]. Such an ideal system would never present an erroneous bit, would do so with a minimum of input optical power and with the minimum of electrical power. While the optimum solution in terms of energy use, efficiency and reliability, the concept of the ideal system has proved a powerful drive for previous communication systems [13, 31].

The quantum limit characterises the theoretical minimum power that is required for a particular error rate at a specified transmission rate with a specified modulation scheme [1, 2, 3, 14, 36]. The limit assumes a perfect system with no added noise. The quantum limit is therefore useful when comparing the performance of a realistic receiver, giving indications of parameters that should be optimised.

2.3.8 Receiver Optical Sensitivity

The performance of a receiver could be characterised by the bit error rate experimentally obtained for a particular optical power and data rate. The optical sensitivity views performance from the other side, instead measuring the optimum input signal power that is needed to achieve the specified bit error rate at the target operating data speed [2, 14, 36]. If the data rate increases, the energy per symbol must be kept constant in order to retain the error rate that has been targeted [2, 3]. This forces higher received signal powers to be used [1, 3], and thus a higher communications link budget. The received signal power can be reduced giving a low power communications link [1, 3], if an increase in errors can be tolerated [2]. For long distance, low rate (10–100Mb/s) communications, where received signal strengths are low due to the distances involved, a combined electrical and transmission medium BER of $1x10^{-3}$ may be acceptable, but only if forward error correction codes (FECC) or data link protocols are used to achieve a $1x10^{-6}$ to $1x10^{-9}$ or '*error-free*' overall system performance [1, 22].

While there is a link between the optical sensitivity and the error rate [2], there is also a link between the sensitivity and the bandwidth of the receiver [14, 36]. On one hand a bandwidth greater than the input signal frequency spectrum is needed to ensure low distortion and fast data transitions, however a large bandwidth also allows noise contributions from a large frequency range to mix with the desired signal, thus increasing the error rate. On the other hand, one can imagine a very fine frequency bandwidth, which while it may have low noise by cutting many of the noise contributions in the frequency domain, would increase the error rate through elevated inter-symbol interference [14]. There is therefore an optimum bandwidth for a certain data rate [14]. This is approximately 60% to 70% of the data rate, therefore for a example data rate of 100Mb/s, a bandwidth of 60 to 70MHz is needed.

When investigating receivers for low input signal powers, there is a trade off between the required sensitivity, to achieve communications at those light levels, and the error rate [2]. The inverse relationship between the error rate and the signal power [2, 3, 36], presents a challenge for designs, such as the front-end transmitters and receivers to be developed in Chapter 4 and the wider Hypix project, which do not directly include error correction circuitry. To satisfy the project specifications, a sensitivity close to the ideal quantum limit is needed, and the error rate of the link must be low without needing higher-level correction methods. This prompts continued investigation. For this work, the Hypix project aims to obtain low transmitter-only and receiver-only error rates, with advanced driver circuitry, error correction techniques and

protocols being developed in the follow on projects after Hypix. For Hypix this may require slightly higher optical signal powers to achieve the specifications required, but prompts the project to investigate all factors that may affect the envisaged transmitters and receivers [66]. Complementing this investigation of factors, a bandwidth suitable for the Hypix target data rate is need. This bandwidth presents an issue, as the behaviour of a receiver using SPADs (Section 1.1.3) may not be known in enough detail to be able to ensure the 60 to 70% of the data rate criterion. As such, this work aims to investigate the receiver properties, such as the impulse and step response, that may effect the receiver bandwidth.

2.4 Optical to Electrical Conversion: Photodiodes

CMOS photodiodes [4, 26] use two structures, the p-n and p-i-n junctions. The p-n structure uses adjoining n- (donor) and p- (acceptor) doped silicon [26, 31, 44, 46, 107]. This creates a thin depletion region [4, 31, 55, 75]. A thin interfacing region of intrinsic, un-doped silicon can be used forming a p-i-n junction that expands the depletion region controllably. This allows tailoring of sensitivity and bandwidth [4, 101, 108]. The diode operates in reverse bias [13, 14, 26, 31], widening the depletion region and increasing the electric field [26]. The photodiode operates when photons are absorbed at a depth dictated by the incident optical wavelength [26].

The photon energy [80], if greater than the energy band-gap [26], promotes an electron into the conduction band. This creates an electron-hole pair for each photon absorbed [26]. The pair are then separated by the electric field, forcing a photo-current to flow through external circuitry [2, 14, 13, 26]. There are two fundamental points important for understanding photodiodes. These are the sensitivity and the bandwidth [4, 14, 31, 33, 36, 44].

1. The sensitivity is limited by the quantum efficiency [26]. This is the number of electron-hole pairs created for each absorbed photon. As a photodiode is only sensitive to photons that are absorbed near the depletion region, photons absorbed above or below the depletion region do not contribute to the photo-current [26]. Sensitivity is maximised if the depletion region depth matches the depth of absorption [4, 26] or if the region covers a large proportion of the absorption curve [4, 26, 75].
2. The modulation bandwidth is limited by four properties of silicon technology. The limit in most cases is the extrinsic bandwidth formed by the diode capacitance and resistance [33]. The intrinsic bandwidth [33] is caused by carrier diffusion from regions below the photodiode, implying use of silicon-on-insulator processes for speed [27, 31, 46, 109]. Despite these limits, theoretical bandwidths are limited by the time it takes carriers to recombine [26, 76]. This is called the minority carrier lifetime [26]. A final limitation, is the transit time of a carrier over the depletion region [26]. This forces thin regions and small areas in high-speed designs [31, 37, 39, 110, 111].

A trade off is formed between sensitivity and bandwidth [31, 14, 35, 36]. As a wide depletion region is necessary to capture a large proportion of photons [26], the bandwidth becomes transit time limited [26, 31]. Likewise large areas, creating a large photo-current, become capacitance limited [26, 31].

2.5 Optical Communication Receivers

Three classes of receivers are used for optical communications [2, 35]. In this section each class is described highlighting the continuum between, advantages and disadvantages of single-sample, multiple-sample and integration receivers [2, 13, 31, 35, 73, 83]. The single-sample approach is the largest class due to simplicity and historical separation of photodiodes, signal amplifiers and decision circuits [13, 31, 32, 40]. Both multiple-sample [35] and integration receivers [2, 35, 73], attempt to reduce the noise in a measurement of a modulated input, and can provide advantages for overall receiver sensitivity [2, 35]. Each class is discussed, however the integrating architecture [73], is chosen for further design [2, 35, 73].

2.5.1 Single-Sample (Direct Detection) Receivers

The simplest receiver consists of a photo-detection element and an amplifier chain [2, 13, 31, 32, 35]. These are designed to provide the correct voltage whilst optimising trade-offs between gain, bandwidth, noise and power efficiency [13, 31]. Once a voltage appropriate for the down stream circuitry has been reached [13, 31], the analogue data must be converted back to a digital stream. This conversion uses a clock and data recovery (CDR) circuit in combination with an on-chip phase-locked-loop (PLL) [2, 35]. This regenerates a clock synchronous with the input and reproduces a digital logic level, well synchronised bit stream that can then be de-multiplexed [35], error checked or corrected [1]. In the subsections below, the analogue front-end circuits are described, along with major performance trade-offs [13, 31, 33]. A simple single-sample CDR circuit is then discussed [35].

Front-End Amplification Circuits

Prior to fully integrated receivers [13, 31], the photodiode, initial photo-current to photo-voltage conversion, voltage amplification, and equalisation were performed separately, often in optimised technologies [32]. In modern integrated receivers, a photo-current, i_{ph}, is produced by a photodiode within the CMOS technology [13, 31, 33, 46, 100]. This current is small due to finite quantum efficiencies [26], small active areas and finite optical powers incident on the silicon [31, 108].

There are two main issues in these systems. For a high signal amplitude, and thus lower amplifier gains and areas, a large photodiode is required due to the sensitivity, which is measured in amps per watt of optical power (A/W) [2, 13, 26, 31]. However a large photodiode has a large capacitance, reducing the bandwidth and requiring equalisation techniques [13, 31, 33]. A trans-impedance amplifier (TIA) [2, 13, 14, 31] converts the modulated current from the photodiode into a voltage [35]. The TIA also provides suitable gain to achieve the output voltage, a bandwidth greater than the PD and low noise [1, 14, 31, 36, 100].

Single-sample receivers are the largest class of receiver, with many amplifiers proposed in the literature [1, 2, 3, 33, 35, 38, 40, 79, 110], along with techniques such as differential photodiodes or complex equalisation to increase performance [13, 28, 31, 39, 44, 48, 100, 111]. The signal amplitude and bandwidth trade-off [26] leads to a compromise using small, high bandwidth photodiodes, large amplification circuits and high incident light levels [31]. This trade-off is unsuitable for the specifications of this project, where the trade-off would quickly limit the sensitivity that could be achieved. This suggests that design methods to minimise this trade off should be used, including use of alternative gain mechanisms [4, 26, 43].

Conversion of a photo-current into a photo-voltage, can be achieved using a simple series resistor, R_{TI}, acting as a trans-impedance resistor. Via Ohm's Law the series resistor gives a voltage with amplitude, $V(t) = i_{ph}R_{TI}$. However the bandwidth can be pushed higher when an amplified feedback circuit [2, 112] is used. Equation (2.1) describes this amplified feedback bandwidth, with amplifier gain A_g, feedback resistance R_f, photodiode capacitance C_D and parasitic capacitance C_S (Figure 2.2).

$$BW_{TIA} = \frac{A_g}{2\pi R_f (C_D + C_S)} \tag{2.1}$$

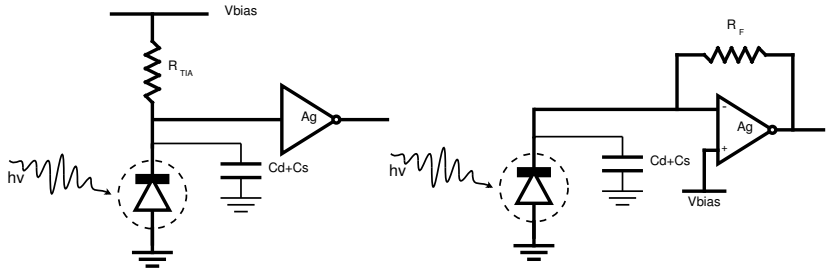


Figure 2.2: Simple Trans-impedance amplifier concepts, a) a resistive TIA using R_{TIA} for current to voltage conversion, and b) an active feedback TIA using R_f for current to voltage conversion with a closed loop gain of A_g.

Simple Clock and Data Recovery Circuits

It is useful to consider the single-sample CDR circuit [2, 35]. This digitises the amplified TIA output and is the corner stone of the single-sample receiver architecture (Figure 2.3). The operation of CDR is to regenerate the clock from the received data [1, 2, 3, 35]. This ensures that the clock does not need to be transmitted separately [14, 36, 83]. Once the clock has been generated, a feedback loop with an accurate voltage controlled oscillator (VCO) and PLL is used to provide a low jitter clock, synchronised with the clock of the transmitter [69].

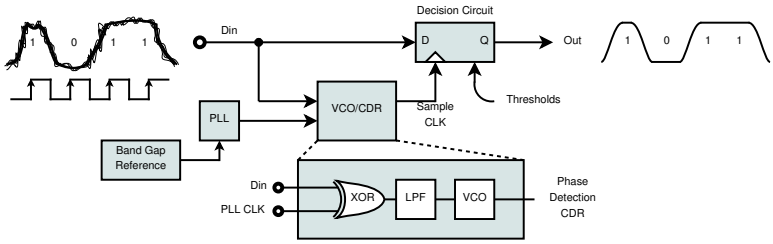


Figure 2.3: A phase detection CDR circuit. A flip-flop takes a decision between a transmitted *zero* or *one*. This is sampled upon the clock edge produced by the phase detecting CDR circuit. This compares the input data stream with a local clock source. The error between the two phases is used to compensate a voltage controlled oscillator. As such, the sampling clock and output data, track changes in the input data rate.

In Figure 2.3, a phase detection CDR [35], modifies the receiver clock frequency and phase. Conceptually, sampling of the data can be achieved using a simple flip-flop, clocked with the inverse of the re-generated clock. By doing so, the circuit can ensure high SNR, low-error data that is re-timed prior to output [2, 3]. However natural voltage noise along with sampling clock jitter, act to degrade the BER performance that can be achieved [14, 36]. With single sample receivers, all of the signal within a symbol is ignored other than during the single short sampling interval [2, 35]. This gives a sensitivity limited by the noise and relative amplitudes of the inputs denoting each symbol. Inter-symbol interference [2, 3], may erroneously produce a sampled voltage greater than the threshold, increasing the error rate further [14, 36]. The single sample must therefore be taken when noise, transients and symbol interference are at a minimum and when the difference in symbol amplitude is maximised [35, 103].

2.5.2 Multiple-Sample Receivers

Multiple sample receivers produce a BER inversely proportional to the number of samples, $N_{samples}$, per symbol. This decreases the BER though oversampling [2, 35]. For a three-sample CDR (Figure 2.4), samples A and C may be used to sample transitions of the input data. Correctly identifying these transition edges, including data jitter, allows sample B to be taken when the amplifier output has settled [35].

Multi-sample receivers encounter a number of problems [35], including elevation of the input capacitance [73, 113, 114]. This impacts bandwidths suggesting low numbers of samples or dedicated low-capacitance circuitry. Accurate sample and hold times and low mismatch between samplers are also needed to ensure all samples are equal in terms of sample duration, noise and output amplitude. Finally an accurate clock is needed with a jitter $\approx N_{samples}$ lower than for a single-sample CDR [2, 35, 73, 113, 114]. Despite implementation issues, multi-sample receivers can be used to provide low-error data. Further, once sampled, low-noise, programmable, small digital circuitry can be used for phase picking or phase tracking [35].

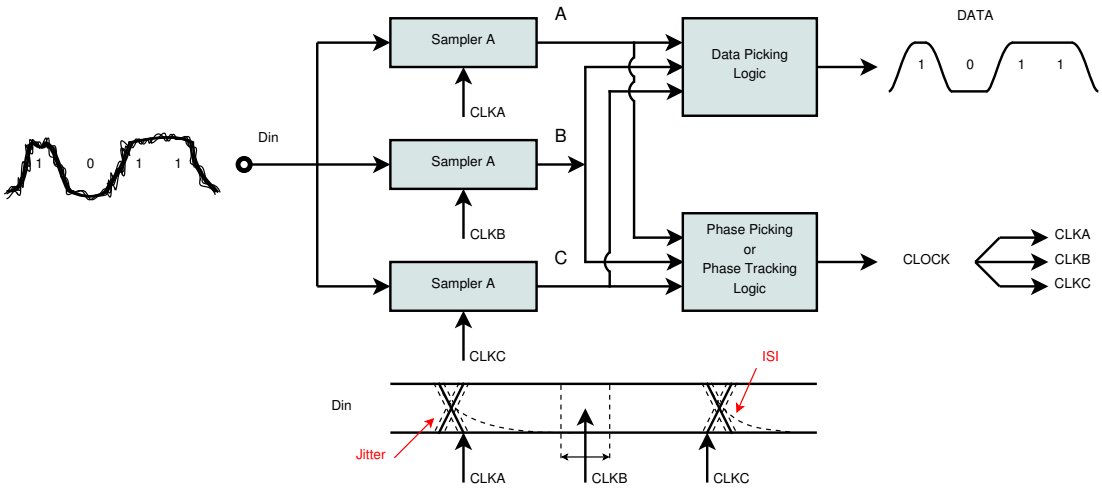


Figure 2.4: A Multi-Sample CDR Circuit. Noisy input data is used to reconstruct a set of sampling clock phases. Re-timed synchronous data streams are output using phase picking between multiple Re-sampled data values from each of the samplers. Phase tracking ensures ISI is minimised.

2.5.3 Integrating Receivers

Taking the multiple sample approach further, the analogue integrating receiver architecture can be used [2, 35, 83]. This class of receiver takes the time integral of the input, measuring the area under the waveform rather than the peak amplitude [2, 35, 73, 113]. If many samples are taken, the continuous time integral can be approximated by taking the integral of all samples that occur in discrete time during the symbol [115, 116] (Figure 2.5).

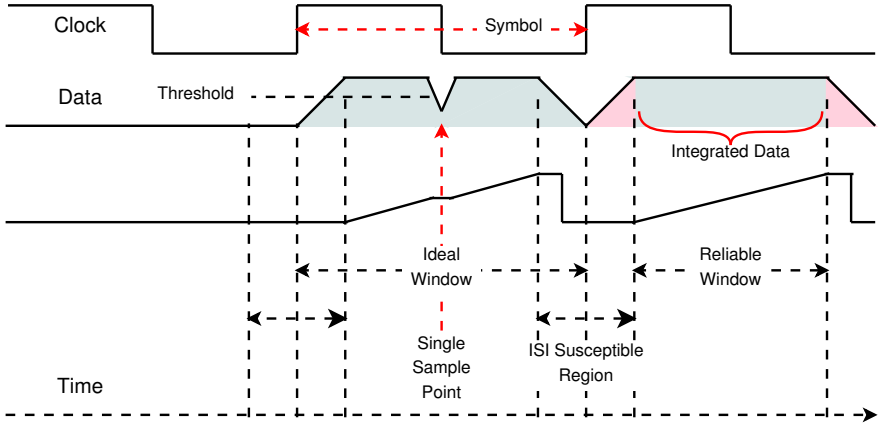


Figure 2.5: The concept of an integrating receiver. During a symbol, an integration window is used to obtain the time integral of the input data signal. This provides some inherent reduction in the influence of in-symbol glitches and noise. At the end of the symbol, the integrated result is sampled and the integrator reset prior to the next symbol. Blanking periods can be used to remove signals from the ISI susceptible regions between symbols.

The integration process provides a number of attributes useful for a receiver design with a high-sensitivity and a low BER [2]. Firstly from Equation (2.2), integration over a time period produces a larger receiver output amplitude than a single, short measurement (Figure 2.5). If the integration period increases as a proportion of the symbol, the output amplitude is elevated correspondingly [2, 35].

$$Vout = \sum_{i=1}^{n} Vin_i = Vin_1 + Vin_2 + \cdots + Vin_n \tag{2.2}$$

$$\approx \int_0^t Vin(t)\ dt \tag{2.3}$$

Secondly, an increased SNR [83] and therefore a lower BER results from an averaging of noise at frequencies higher than the symbol rate. The SNR is further increased through a level of robustness against sampling clock jitter that is provided by the technique [2].

Receivers combining signal integration with conventional amplification techniques have been implemented [35]. Stevens [73] splits the integration into three multiplexed sections forming the integrate and dump methodology. This used separate signal integration, readout and reset phases thereby reducing the clock frequency and high frequency coupling [73]. In [117], integration was included within a differential TIA. This combined signal integration, amplification and bandwidth enhancement through active-inductor peaking [35]. The concept was refined within a multi-phase receiver [113] which obtained a reduction in inter-symbol interference, through time-multiplexing integrations [73] and equalisation. The optimisation of ISI rejection, was however a result of circuit adaptability, modifying the symbol and transition integration periods [113].

The integrating architecture has benefited from development, including a source synchronous clocking method [114]. This matched the integration start and stop edges with the input data, allowing an optimum integration window. Development has continued with Georgas [46] amalgamating the integrating receiver with SOI processes for on-chip photonic links. Their proposed system moves the integration directly into a regenerative latch, similar to [117]. Throughout the development of integrating receivers, authors have noted that it can achieve a greater sensitivity than single-sample TIA circuits [2, 35, 46, 118], and that electrical power can be reduced for the required receiver gain [46, 113, 117]. As a consequence of the low-noise, high-sensitivity and low electrical power characteristics, the integrating receiver and its discrete sample approximation are ideal for the receiver to be developed in this work.

2.6 Avalanche Multiplication: Increasing Senstivity

This project aims to elevate sensitivity further while retaining bandwidth and small circuit areas. One strategy is to combine integration with amplification. This would mitigate bandwidth to sensitivity trade-offs. The challenge however is to obtain the required amplification [26]. Avalanche multiplication used in conventional avalanche photodiodes, has been previously investigated for communications [4, 26, 41, 42, 43, 63, 119]. This moves amplification into the front-end diode providing initial gain without external circuitry.

This process uses an increased diode reverse bias voltage [26]. The resultant increase in the electric field accelerates a free carrier, labelled 1 in Figure 2.6, to a kinetic energy sufficient to overcome the ionisation energy [26]. Upon a collision [26], the accelerated carrier ionises another carrier. An electron-hole pair, labelled 2 and 2′, is generated with the carriers then accelerated by the electric field, causing further ionisation [26, 75, 120, 121, 122]. This process continues exponentially creating a small avalanche of carriers within the depletion region.

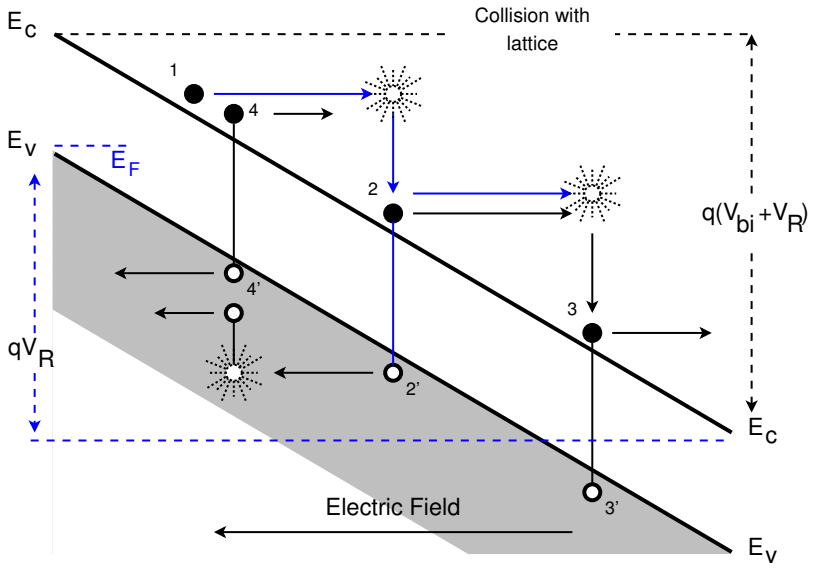


Figure 2.6: Band diagram showing reverse bias *Geiger* avalanche multiplication. When the electric field across a p-n junction is elevated above the ionisation energy level, an accelerated carrier imparts a significant kinetic energy to a bound electron upon collision with the lattice.

Conventional APDs bias the junction such that multiplication and avalanche of carriers achieves linear gain [123, 124] but not run-away avalanche [76]. The diode produces a current dependant on the incident photon flux [125] with the gain [123, 124] and depletion region width being dependant on the bias voltage [26, 76]. Some photodiode structures can be biased further into reverse bias, giving larger gains and greater sensitivities. In exceptional cases the run-away avalanche process [76] can be used to create specialised diodes called Single-Photon Avalanche Diodes (SPADs) [49, 50, 53, 75, 76, 126].

2.7 Single-Photon Avalanche Diodes (SPADs)

As high receiver sensitivity is needed, avalanche gain exceeding that used in APD receivers could be combined with the integrating receiver concept. Therefore, Single-Photon Avalanche Diodes are proposed. These can be pushed further into reverse bias, giving multiplication gains of over $10,000$ [49, 127]. In Figure 2.7, the SPAD replaces the photodiode, TIA and feedback network. The small digital inverter, necessary to prevent loading, replaces the post-amplifier [128, 129]. The gain required by the receiver is moved fully into the SPAD, removing the need for amplification circuitry. The SPAD provides a voltage pulse which can be processed by digital logic. As such, the SPAD itself provides *direct-to-digital* functionality through an efficient transfer between the optical and digital domains [128].

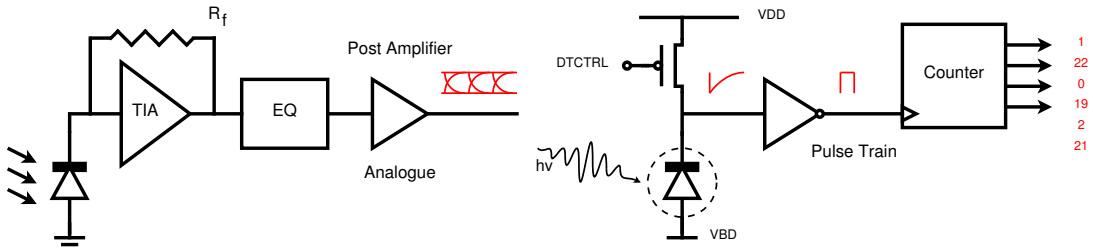


Figure 2.7: Continuous time reception on the left uses a trans-impedance amplifier, equalisation of the photodiode capacitance and post-amplifiers to obtain the correct output voltage swing. Discrete time reception with a SPAD on the right, uses a small inverter to prevent loading along with a digital pulse counter.

2.7.1 Theory of Operation

The Geiger region lies beyond the linear avalanche gain region but prior to breakdown of a guard ring that surrounds this low-level photo detection device (Figure 2.8) [26, 49, 74, 123, 127, 130]. Initial research into avalanche behaviour, centered on Microplasmas [74, 123, 130]. These are small breakdown regions, corresponding to silicon defects [74, 123, 127, 131, 132, 133]. The historical study of microplasmas, resulted in artificial microplasmas [74, 130, 132, 134] with guard rings to force a known breakdown region (Figure 2.9). These artificial structures later became known as Geiger-mode Avalanche Photodiodes (GM-APDs) or Single-Photon Avalanche Diodes (SPADs) [49].

The multiplication mechanism is similar to linear avalanche, however the increased electric field due to the higher reverse bias voltage, results in a large elevation in carrier ionisation rate [26, 127]. The Geiger region provides long periods of quiescent operation [76, 123, 127] at a bias voltage of $VOP = VBD + VEX$, where VEX is the excess bias applied above the breakdown voltage, VBD [49, 75]. Upon the arrival of a carrier, generated by thermal excitation [26], photon-induced or the release of a trapped carrier within the depletion region, multiplication creates a fast avalanche current [76, 135]. The device transitions from the quiescent point (Figure 2.8), to the I-V curve, causing a voltage change over the device [76, 120, 125, 135].

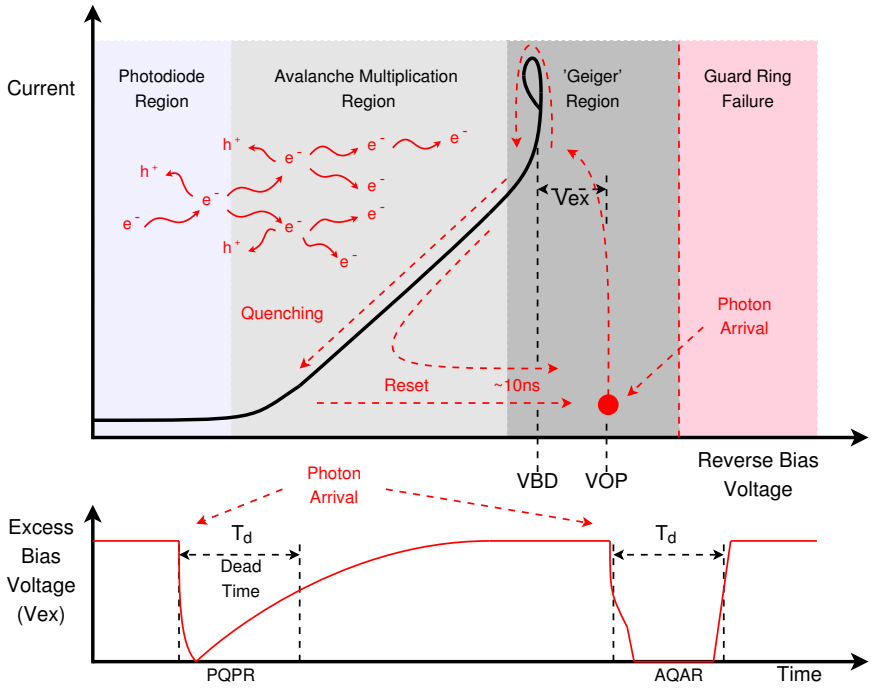


Figure 2.8: Mode of operation of a SPAD showing its bias points and output waveforms. At the left (low voltage) the output current is independent of the bias voltage, being modulated only by temperature and incident light. In the centre the avalanche gain region provides a voltage dependant gain rising as the kinetic energy of carriers increases the probability of carrier ionisation and liberation into the conduction band. As the voltage increases towards the breakdown voltage, the current becomes large producing current pulses indicative of *Geiger* operation. At the bottom, Geiger mode voltage-time pulses are shown for both passive and active quenching cases, denoting the dead time following the avalanche.

A quenching operation decreases the current, eventually dropping down the I-V curve [127]. An external resistance (Figure 2.10) forces a voltage drop across the device [49, 136], bringing it just below VBD, halting multiplication either passively [137] or actively [134]. The sharp voltage discharge is sensed and processed yielding a detection event [49, 138]. The device is then recharged via a passive or active circuit [49]. The total voltage swing over the diode during this quenching and reset operation is VEX, meaning that if the designer arranges that VEX is above the threshold voltage of a CMOS digital circuit, the signal can be processed digitally without further analogue circuitry. In Figure 2.10, two passive quenching topologies are shown. These are suitable for CMOS integration giving either positive or negative going voltage pulses.

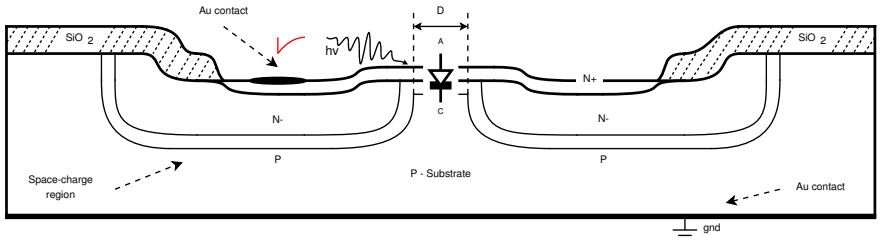


Figure 2.9: The Oldham artificial microplasma test diode. This used a small circular N^+ region abutted to a p-type substrate. A guard ring and anode contact were formed using an n-type diffusion.

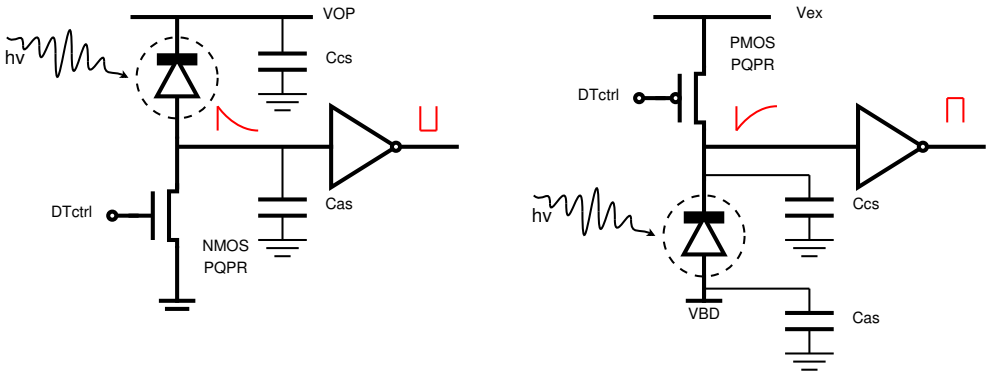


Figure 2.10: SPAD voltage-mode circuit topologies with positive or negative going pulses. The traditional passive quenching topologies have been adapted to use NMOS and PMOS transistors, aiding in CMOS integration and small area quenching circuits. The gate voltage, $DTctrl$ modifies the MOS channel resistance allowing adaptation of the quenching resistance, R_L rather than the fixed resistance values used in earlier circuits.

2.7.2 SPADs in CMOS Technologies

SPADs have been implemented in many materials and processes [50, 126, 134, 137, 139, 140]. However their introduction into CMOS is quite recent (approximately 2003) [52]. CMOS SPADs use the maturity and cleanliness of CMOS to implement structures similar to the Goetzberger [74], Haitz [123, 127], Oldham [132] and Batdorf [130] structures (Figure 2.9). In Rochas et al [52] a SPAD utilized diffusions used to fabricate PMOS transistors. Structural developments allowed investigation of other junctions within standard CMOS [53, 55, 141, 142], along with development of retrograde guard rings for reduced noise [143].

Despite restrictions from CMOS foundries, a number of developments have tailored the process and doping levels [144, 145, 146], yielding enhanced performance. These structural developments are critical at the architecture level, as cross talk [57, 147] or noise specifications [55, 148] may impact the choice of CMOS process, the layout or impose restrictions on the on-chip logic [149, 150].

2.7.3 Important SPAD Performance Metrics

Investigation of SPADs has highlighted departures from ideal single-photon counting behaviour [49, 50, 75]. These have improved greatly since early developments [74, 123, 127, 130, 132], however the limitations, such as the finite efficiency of detection [151] and the erroneous detections occurring in the dark [123] are discussed in the subsections below. Of course the ideal case would be to detect 100% of incoming photons, with zero erroneous detections over a wide wavelength and optical power range.

The Photon Detection Efficiency

The photon detection efficiency (PDE), is linked to three processes. These are the wavelength specific absorption [26], quantum efficiency [26] and avalanche *turn-on* probability [74, 75, 127, 151, 152]. The absorption of photons and the depth reached at absorption, is wavelength and material dependent. It is dependent on the material band-gap, with absorption yielding an electron-hole pair if the photon energy is greater than the material band gap [26]. The second effect, the quantum efficiency (QE), is limited by the band-gap, the incident photon wavelength and the junction depth and width [26]. This is the number of electron-hole pairs that are created per photon absorbed in the optically receptive volume. Finally, the avalanche turn-on probability is defined as the probability that an electron-hole pair will cause an avalanche [75]. This is primarily dependent on the bias voltage through the carrier kinetic energy and the material ionisation energy within the multiplication region [152], but is also dependent on carrier generation and trapping [74, 127].

The PDE can be increased though a number of techniques. These include doping and structure developments [50, 56, 75, 132, 141, 144], elevating the operating bias voltage leading to increased electric fields within the multiplication region [49, 52, 53, 120, 138, 141, 143] and through wider multiplication regions capturing more of the absorbed photons [26, 49, 75].

The Optical Fill Factor

The fill factor (FF) is the ratio of the optically receptive area to the total device area [56, 81] and is a key parameter in obtaining the goal of a high sensitivity receiver. The fill factor significantly alters the receiver sensitivity as photons incident on inactive regions are not counted but nonthe-less included in the total incident photon flux. Some areas of the silicon are intrinsically optically insensitive due to the metal layers above them, the lack of biased p-n junctions within that circuit or a circuit function, such as a digital circuit, that precludes the detection of signals below the switching threshold. The fill factor, can be difficult to assess for optical receivers, as focusing optics may be used to constrain the flux only to the photodiode [13, 31, 40, 100, 111], or a lens may be used to capture light from a large area and focus it down to a sharp point.

In previous SPAD arrays, the fill factor was often low at 1-2% [70, 148, 153, 154, 155], giving a fixed offset from the 100% efficient quantum limit [14]. For fill factors as low as 1-2%, the optical signal power incident on the receiver would need to be increased by 5 to 10 times that expected to ensure robust detection of the signal can be maintained. The fill factor is often limited by the diode geometries available for a given noise or capacitance specification [53], the complexity integrated onto the device [70, 155] and the CMOS process design rules [53, 55, 56, 70, 148]. It is however advantageous to design a high fill factor receiver [156], to ensure a high receiver sensitivity. While a high fill factor ensures photons have a lower probability of hitting non-optically active circuits and are thus detected, it allows the optical power needed for robust signal detection to be reduced. For a 50% fill factor for example, the optical signal power incident on the receiver would only need to be increased by a factor of 2, far better than the previous factor of 10.

Despite restrictions [53], the fill factor has been increasing through a number of techniques. N-well sharing [141, 157], out-of-array electronics [57], reduced quenching circuit areas [53], NMOS-only logic [56], increased diode radii with retained low-noise [144], use of clustered mini arrays [158, 159], non-circular geometries [53] and micro-lenses [156], have all lead to fill factors reaching recent maximums of 67% [58] and 70% [160]. With a fill factor such as 70%, the optical power would need to be increased by a factor of 1.43, again much less than the factor of 10 for 1% fill factor arrays.

The SPAD Dead Time and Finite Counting

Once a SPAD fires, the external quenching circuit recharges the device [49] (Figure 2.10 on 26). While the diode is below its quiescent voltage it is unresponsive to further incident photons [49, 127, 135, 161, 162]. This period of insensitivity, (the dead time), represents the maximum counting performance that can be achieved [60, 163, 164]. The photon transfer curve (PTC) between input photons and output detections characterises the change in responsivity of the SPAD over incident optical power [163, 165], showing the maximum, along with detection losses and non-linearities [49, 75, 137, 166].

It is advantageous to have a low dead time, τ_d, as this gives high counting rates, low losses and linearity over a large dynamic range [165, 167]. However as it limits the detector, many diodes in a summation architecture may be required in order to achieve the desired specifications [43, 168]. When multiple SPADs are used together, the total maximum count rate of the group can be viewed as a finite counting resource. Clearly a higher number of diodes or a lower dead time, act to increase the total counting resource.

Unfortunately an arbitrarily short dead time cannot be used as other second order effects begin to degrade performance [49, 75, 76]. Despite prior modifications reducing SPAD and circuit dead time contributions [163, 162, 169], research aiming to mitigate second order phenomena

is ongoing [55, 76]. These techniques suggest use of small active areas with low-capacitances [53, 157]. Active quenching, isolation from the substrate and structural design techniques may also reduce the minimum dead time [56, 76]. The photon transfer curve may be affected by variation in the dead time [163], suggesting increased disparity between received symbols if variation is not controlled [2].

Timing Response (Jitter)

A statistical variation in photon arrival [80, 170] is complemented by a timing variation from the SPAD [26, 75, 76, 120, 121, 125, 171]. This is the variation between an avalanche and the electrical pulse with two main contributions.

- The first is the transit time of carriers over the multiplication region [76, 132] which is exacerbated by lateral spreading of an avalanche from a single point [55, 120]. Photons absorbed directly in the multiplication region and the statistics of avalanche build-up contribute to the statistical width of the jitter [75, 120, 121, 125].
- The second source of jitter, adds a decaying tail to the arrival time [49, 75, 171]. This is caused by diffusion of photon-induced carriers [49, 75, 120, 171] or carriers injected into the substrate [56, 149, 150].

Small area devices have reduced the jitter in timing critical applications [53]. However research has also continued with SPADs now isolated from the substrate, mitigating slow carrier diffusion issues [172]. Process modifications have also been utilised [144].

The Dark Count Rate

A SPAD device, exhibits a long quiescent period between detections [127]. However despite the absence of light, it maintains a background count rate [123, 127]. The Erroneous detection rate is called the dark count rate (DCR) [49, 50, 171, 173]. It is quantified as the mean number of dark detections within one second [26, 123, 127] and can be explained by two routes.

- The first is thermal generation due to the random transition of electrons from the valence to conduction band [26]. Traps [49, 174] formed within the band gap elevate thermal generation, with the DCR being volume [55, 75] and bias voltage dependent [26, 123, 152]. Carrier generation via the thermal generation mechanism is important in receiver design as room temperature operation, the CMOS process choice and its cleanliness, both impact the minimum zero bit amplitude. This is important if 100% extinction ratios can be produced by the co-designed μLED transmitters, and their digital CMOS drivers [96].
- The second contributor is band-to-band tunnelling [26, 52, 76, 123]. The energy band gap normally presents a barrier to carriers below this energy, however a carrier has a finite probability of tunnelling [26]. This is dependent on the energy of the carrier and any energy levels (traps) within the band gap.

The DCR, which follows Poission statistics [49, 75], exhibits a number of phenomena. These include random telegraph noise (RTN) which characterises a long term modulation in the mean-DCR of a particular diode [175]. As noted by Spinelli [120], the DCR also acts competitively with photons over the finite period the diode is fully biased. To mitigate the DCR [176] and dead time impacts on photon counting [60], time-multiplexed and time-gated techniques have been proposed [62, 177, 178, 179, 180, 181]. The per-SPAD dark count rate, can be mitigated without system level techniques such as gating, by using a number of methods.

1. As noted by Webster [182], modifying dopant levels and increasing VBD can improve DCR through reduction of tunnelling and trap-assisted thermal carrier generation.
2. Thermal generation can be decreased by cooling the detector using thermo-electric Peltier cooling [53, 63, 143, 177], however these coolers are inefficient and may elevate the overall power consumption.
3. Small volume diodes mitigate the mean-DCR through reduced numbers of traps within each diode [53, 56, 157]. A marked improvement in yield is also observed [53]. This yield is measured as the proportion of diodes that show a DCR below the median [53, 55].
4. The array can be improved by using a binary on-off weighting, turning off a small proportion of above median-DCR devices [53, 55, 56, 57, 148, 158]. Removal of high-DCR SPADs can be effective, but reduces the array fill factor and of course decreases the overall photon counting resources available.
5. CMOS process and structural modifications have also shown reductions in mean DCR [53, 144, 172, 182]. An important one used in this work, is the use of a retrograde guard ring, and the reduced use of shallow trench isolation [53].

The After-Pulsing Probability

After-pulses are a correlated temporal component of the DCR [74, 123, 127]. Upon an avalanche, traps are charged by the large influx of carriers [26, 49]. The traps, begin to release carriers with a temperature dependent lifetime [26, 174, 183]. If a carrier is released when the device is in a high bias state, then a secondary avalanche will result a short time after the initial detection [49, 120, 123]. The after-pulsing probability, $P(ap)$, is the probability of a secondary avalanche occurring after the SPAD dead time. As with the DCR, after-pulsing is bias voltage, temperature, trap and structure dependent [76]. It is also dependent on the number of traps within the SPAD volume, however the number of charge carriers produced in an avalanche will directly influence the charging of those traps [123].

As noted by Zappa [75], the diode is insensitive to after-pulses that occur during the unbiased state. As such, a dead time longer than the carrier and trap lifetimes allows reduction of $P(ap)$ to negligible levels. This is possible as all traps have been depopulated during the low-bias state [50]. After-pulsing can be mitigated without using dead times that are too long for the communications application by using a) small avalanche currents, through the use of low

capacitance, small area diodes, b) lower trap concentrations through clean well-annealed, low-doping concentration CMOS processes [75] and c) dedicated SPAD planar structures [53, 55, 143, 172, 173]. Despite these techniques, a minority carrier issue may fundamentally limit how short the dead time can be, and thus may limit the high photon counting rates needed for robust communications at high data rates [76].

2.8 A Multi-Diode Approach

SPADs have limitations such as the dead time, dark count rate, fill factor and after-pulsing that are discussed in the previous section. These reduce the effectiveness of SPADs [7, 49, 50, 53, 62, 64, 67, 75], but have shown improvement with reductions in device size and have shown dramatic progress with recent industrialisation [53, 55, 59, 182]. In particular, the dead time is lower in small area SPADs due to smaller diode and parasitic capacitances [49, 169]. This short dead time leads to a high diode photon detection rate as the diode can discharge and recharge with a smaller RC time constant [60, 163].

The temporal response is also improved in smaller SPADs due to a reduced lateral area in which avalanche must be initiated [53, 120, 122, 125, 184]. This is advantageous, however other effects benefit more from device scaling [53]. Two such effects that have marked improvements, are the mean DCR and the yield of low-noise devices [53]. This is due to zero or low numbers of traps within each diode, although the total number of traps per unit volume has also been improving through cleaner CMOS processing [55, 59, 76]. A lower volume of silicon within each diode is also adantageous [26, 49, 50], leading to lower numbers of thermal or tunneling generated carriers [26]. The after-pulsing rates, which contribute a temporal noise component, also show improvement with scaling due to fewer devices containing multiple traps [53, 55, 76]. The reduction in diode capacitance contributes to this improvement, with a reduction in current flow during an avalanche event [49, 50, 53, 76].

In order to counter-act some of the issues with SPADs, namely the dark count rate and RC time constant, a multi-diode approach can be used, whereby many smaller diodes are connected in parallel rather than using one diode of the same total active area. This approach reduces the dark count and after-pulsing rates [53] while providing diodes that recharge far more quickly than a single large SPAD with large parasitic capacitances [49, 53]. The scheme allows large sensitive areas integrated with the receiver, however the approach is a trade-off as other limitations, like the fill factor, become important. Using multiple separate diodes is not a new technique as it has previously been sucessful in solid-state SiPMs [77, 145] and has been used in both photodiode and avalanche photodiode receiver [21].

While a multi-diode structure is advantageous with respect to the limitations discussed above [53], it negatively impacts the optical fill factor [53, 56, 185], giving a fixed offset between the quantum limit and the actual sensitivity. This prevents receivers from approaching the quantum limit of an ideal 100% sensitive area, 100% detection efficiency receiver. Despite this reduction in sensitivity, the main SPAD limits are improving for larger area, higher optical fill factor SPAD devices [59], and a number of array layout methods can be used to minimise the impact of the CMOS design rules that reduce the fill factor [141].

2.9 Specifications For Receiver Design

Specifications are derived from the Hypix project [66], which aims to develop receivers working with 32x32 = 1024 μLED transmitters [24, 93, 98] and their CMOS drivers [23, 96]. Taking a *design for test* approach [69] maximises the scope for characterisation and experimentation. This approach is necessary for investigation of performance and future optimisation, but reduces the performance of the initial prototype receiver designed here.

Although fixing a specification is difficult at this stage, the sensitivity should be maximised. Initially a sensitivity of -14dBm was obtained when using μLEDs at 155Mb/s and a BER of $1\text{x}10^{-9}$ [24]. The sensitivity must exceed this with a design margin (e.g. $\approx$ 10dB), to allow for channel losses, optical filtering and differing μLED performance [24]. Using this, a sensitivity of -24dBm is suggested. VLC using LEDs for illumination and communication, can also be used to estimate the sensitivity [18]. For a 100Mb/s transmission at a BER of $1\text{x}10^{-6}$, and scaling the optical power to a single LED, O'Brien et al [18] gives a sensitivity of -21.5dBm.

The important specifications are the target data and error rates [66]. These are established by the project, while others such as wavelengths are flexible [23, 24, 66]. The key specifications for implementation are the multi-diode integrating architecture and the number of diodes to be integrated onto a chip. Although the receiver designed here aims for these targets, its primary function is to find, measure and assess the impact of as-yet-unknown factors. As such, it is expected that the receiver will not achieve the exact performance targeted, but will highlight factors that reduce performance and show routes that can be followed in the future [66].

Table 2.1: Chapter 2 specifications table for the receiver and front-end circuits.

Receiver Specifications:	
Transmitter:	Hypix project μLED structures
Modulation Schemes:	On-Off Keying (*ONLY*)
Number of Channels:	Single and Multiple Channels
Receiver Format:	2D Array
Target Data Rates:	100-500Mb/s per channel
Target Bit Error Rate:	$1x10^{-6}$ to $1x10^{-9}$
Target Sensitivity @ 100Mb/s:	Better than -24dBm
Target Sensitivity @ 500Mb/s:	Better than -17dBm
Target Wavelength:	400nm to 750nm (Visible)
uLED Extinction Ratio:	$\approx 100\%$ [96]
Technology:	all-Digital CMOS
Technology Node:	130 to 180nm
Total Receptive Elements:	32x32 = 1024 Concurrent design with μLED driver
Receiver Size:	1 to 2 mm^2
Front-End Specifications:	
Receiver Architecture:	Integrating Receiver Photon-Counting Mode
Diode Operating Mode:	Avalanche to Geiger-Mode
Avalanche Gain:	$\gg 1000$
Front-End Device:	Single-Photon Avalanche Diode (SPAD)
Front-End Circuit:	Multiple Diode Summation Active (option for passive quenching)

2.9.1 Towards a Receiver Design

In Figure 2.11 both a photodiode-based [35] and a SPAD-based receiver are graphically summarised. Both receivers use the same overall concept, that of the integrating receiver architecture [35, 73] and the summation of multiple optically active front-end diodes [21]. The figure highlights the similar operation between the two, although there are differences principally in a) the ordering of functional blocks and b) the use of analogue or digital signal processing. In order to retain the *design for test* approach, the SPAD-based receiver may be split between the silicon chip that is to be fabricated (Chapter 4) and the field-programmable gate array (FPGA) that is used for testing. As shown, both receivers output a single-bit data stream at 3.3V, although for the SPAD-based receiver the n-bit digital output may be used to facilitate investigation of the concept and limiting factors.

For testing of the device in later chapters, the proposed structure of the SPAD-based receiver prompts two considerations:

- If the bit width of the SPAD-based receiver chip-output is sufficient [69], there will be little further quantisation noise above the natural Poisson variation of photon counting [80].
- For testing, an n-bit digital output is analogous to an n-bit analogue-to-digital converter (ADC) within a bench digital oscilloscope. As such, both the SPAD-based chip-output and the photodiode receiver voltage prior to thresholding (point A in Figure 2.11) should represent the same information content.

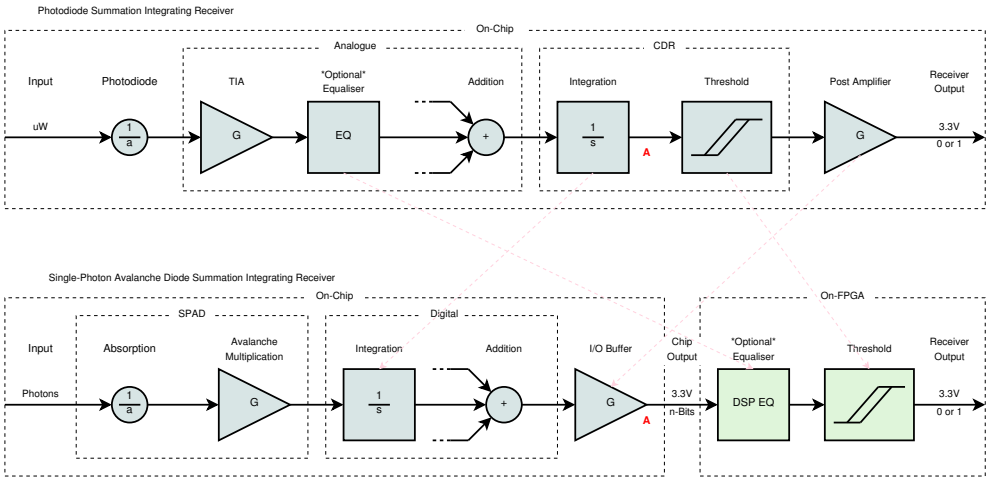


Figure 2.11: Top) A conventional photodiode-based integrating receiver taking the summation of photodiode signals, and Bottom) A SPAD-based, fully-digital integrating receiver again taking the addition of multiple diodes. Both receivers take an optical input converting it to a 3.3V single-bit digital stream. The $\frac{1}{a}$ initial conversion weighting represents photon absorption, s is the Laplace transform variable and G represents a voltage or trans-impedance gain. The hopeful increase in sensitivity is represented by the use of μW for the photodiode receiver input and single-photons for the SPAD receiver input. For the SPAD-based receiver, functionality is a) re-ordered in comparison to the PD receiver, and b) split between the ASIC and FPGA. This promotes logical CMOS implementation and easier testing.

2.10 Conclusions

In this chapter, optical communication systems have first been introduced. A primary aim for transmitter and receiver design, is to increase the speed and error-rate performance while reducing energy use and costs [11]. The key application here is visible light communications (Section 1.1.1), which aims to bolster the throughput of wireless LANs in areas where subscriber numbers and data bandwidth demands either exceed the available radio spectrum [15] or scale faster than the data throughput increases of available Wi-Fi networks.

Concentrating on optical receivers, the current state-of-the-art obtains exceptional speed performance, low error rates, high electrical efficiency and fully integrated solutions in mass-producible CMOS processes. However conventional photodiode and avalanche photodiode receivers require amplification and equalisation circuitry to increase their optical sensitivity. This is required if the receiver area is small, uses optical filtering or is a long way from the transmitter (Figure 1.1).

This work aims to co-design a receiver for the Hypix project's CMOS controlled μLED transmitters, with the wider project investigating both free-space and polymer fibre communications. To address the optical sensitivity, increased avalanche gain is proposed. This leads to the broad definition of receiver specifications shown in Table 2.1. To effectively discuss the details behind this proposal, this chapter has discussed key performance metrics including bandwidth and data rates. The error rate has also been introduced, leading to the sensitivity for a given data and error rate.

From a review of receiver architectures, integrating receivers allow increased received signal amplitudes. This provides a basis for high sensitivity receiver designs. By suggesting avalanche multiplication exceeding that used in APDs, amplification is moved into the diode. This aids sensitivity, prompting introduction of the single-photon avalanche diode that is consequently used.

To summarise the specifications (Table 2.1), in VLC applications the receiver is required to be highly sensitive and capable of both high data rates and low error rates. The transmitter will use OOK modulated μLEDs, however the design of the receiver will combine multiple techniques to increase the sensitivity. In the next chapter, models from the literature are used to further the specifications for a prototype receiver.

Chapter 3

Modelling And Initial Specifications

3.1 Single-Photon Avalanche Diodes for Optical Communication Systems

Given the combination of novel and traditional techniques for receiver design (Chapter 2), this chapter introduces modelling necessary to further the specifications. These are needed for the design of an OOK SPAD-based optical receiver, allowing translation of high level specifications into circuit specifications. Details such as the SPAD count rate, dead time, approximate light level for data reception and the minimum optical efficiency required, can each be estimated.

Firstly, an appraisal of the theory behind the bit error rate is presented to highlight performance values that would be a suitable focus for this work [2, 14, 36, 103]. This BER appraisal prompts increased sensitivity to aid in VLC applications, with either small receiver active areas, or the use of optical filters.

Secondly, models from the literature, including the photon energy [80, 104], the quantum limit [2] and the power penalties of non-ideal effects [1, 14, 36], are explored to refine initial design specifications (Table 2.1). The range of SPAD dead times, complements the design for test strategy [69], giving increased scope for experimental work. The outcome of this section is an initial design flow to obtain values for the relevant design specifications. This design flow ignores as-yet-unknown second order effects but gives specifications suitable for an initial prototype device. Testing and modelling of the prototype can then be used to increase the robustness of this specifications flow.

Modelling informs design undertaken in Chapter 4, creating a best fit to combined project (Chapter 1) and literature review (Chapter 2) specifications. It also helps approach the final performance from a systems level, guiding explanation of inconsistencies that may arise between theoretical and experimental analysis.

3.2 Exploration of a Simple BER Model

In this section, a simple model of the bit error rate is used to identify a direction for the specifications of the prototype receiver. This BER model is used throughout the literature [2, 3, 14, 36, 103], aiming to provide understanding behind the error rate concept. While it is ideal and makes some assumptions, it can still be used to support the design process.

The BER, is a key design parameter dictated by the application. As the BER ultimately defines the number of photons required and the receiver implementation, it can be estimated using a first order model [1, 2, 3, 14, 36, 103]. This estimation assumes that bit amplitudes are Gaussian [3], which while not consistent with the ideal statistics of light [80], can be used to explore key receiver parameters that should be focused upon [13, 31]. This estimation can be done only if the number of photons allows the Poisson distribution to approach the Normal distribution [14, 36, 103, 116]. This process occurs, giving reasonable approximation, with ≈ 10 counts, hence for the target BER, the Gaussian assumption can be used to first order [103, 116].

Within this model, the logic-low (*zero*) distribution is characterised by the mean detections within a symbol, N_0 and the standard deviation, σ_0 [2, 103]. Likewise the logic-high (*one*) bit distribution is characterised by N_1 and σ_1 (Figure 3.1).

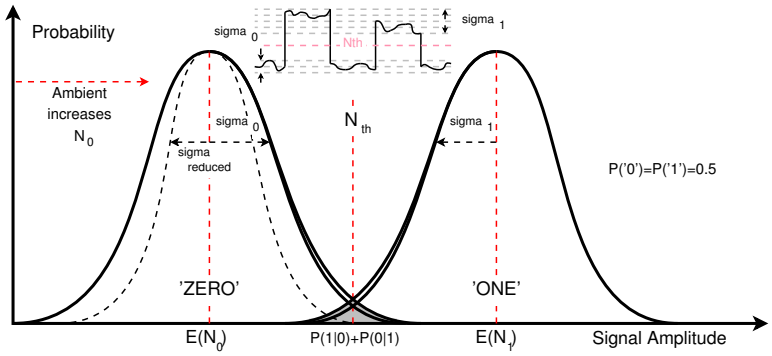


Figure 3.1: Theoretical bit error analysis using the Gaussian histogram method. Here the received *zero* and *one* bits, each form distributions of signal amplitude around some mean level, N_0 and N_1, where the operator $E()$ is the expected value or mean. The standard deviations, σ_0 and σ_1 of the two are caused by random variation in the transmitted, propagated and received signal levels.

A decision made upon reception as to whether a bit is high or low, uses a threshold, N_{th} [2]. The threshold level and the OOK BER are estimated by Equation (3.1) [1, 2, 3, 103].

$$N_{th} = \frac{\sigma_0 N_1 + \sigma_1 N_0}{\sigma_0 + \sigma_1} \qquad BER = \frac{1}{2}\text{erfc}\left(\frac{(N_1 - N_{th})}{\sigma_1\sqrt{2}}\right) \tag{3.1}$$

In Figure 3.1, errors occur between distributions. To achieve low BERs, the probability, $P(1|0) + P(0|1)$ must be reduced [2, 3]. These are the probability of obtaining a one given that a zero was sent and the probability of a zero given a one [1, 115, 116].

Reductions of the error rate can be accomplished in two ways.

- Firstly, the N_1 to N_0 distance can be increased by elevating the receiver SNR [14], transmitter extinction ratio, reducing ambient signals or by increasing modulated signal powers [1, 14, 36].
- Secondly, the standard deviations can be reduced. This however becomes constrained by practical processes or the fundamental statistics of light [2, 14, 36, 80].

A sensitivity analysis can be performed on Equation (3.1), with all variables set to static initial conditions (I_{param}) and varied over the same range. This sensitivity analysis allows the change in BER with each parameter to be assessed, giving the sensitivity to that parameter. In Analysis 1, each parameter is set to $I_{param} = 1$, and then varied separately between $I_{param} \times 0.1$ and $I_{param} \times 10$ (a total parameter change of $10^{2.0}$). This shows that modification of N_1 is preferable, changing the BER by a factor of $10^{1.7}$ (Table 3.1). Altering N_0 also changes the BER, however this is by a lesser extent of $10^{0.6}$. Variations in σ_0 or σ_1 have little effect, indicating these are less important phenomena, although the receiver literature emphasises these two variables as significant issues for modern communications.

Performing a second analysis, N_0, σ_0, N_1 and σ_1 are calculated based upon 100Mb/s, 1024 SPADs formed into a single channel (Table 2.1), 10ns dead time (discussed later in Section 3.3.2) and 2 dark counts per symbol ($\approx$ 200KHz per SPAD) (also discussed later, in Section 3.3.5). This second analysis sets new but fixed initial conditions for each parameter. As in Analysis 1, each parameter is again varied over the range $I_{param} \times 0.1$ to $I_{param} \times 10$, where a 1:1 correspondence between a parameter and the change in the output would give a sensitivity of $10^{2.0}$. For the second analysis, N_1 remains the most sensitive parameter with a change of $10^{17.6}$. Modification of σ_1 produces a change of $10^{5.6}$. The BER is least sensitive to variation in σ_0, although a change in N_0 gives a BER modification of $10^{2.48}$ (Table 3.1). This is in agreement with receiver designs within the literature [13, 31, 110].

Table 3.1: Approximate sensitivity analyses performed on the simple OOK BER model.

Parameter Changed	Analysis 1	Analysis 2
	BER Change	BER Change
N_0	$10^{0.6}$	$10^{2.5}$
σ_0	$10^{0.0}$	$10^{1.7}$
N_1	$10^{1.7}$	$10^{17.6}$
σ_1	$10^{0.0}$	$10^{5.6}$

For project specifications, design should increase the received N_1 mean while lowering light levels. This suggests a prototype remit to increase optical efficiency and sensitivity [13]. The design should aim to reduce the noise, σ_1, of the logic-high bits, along with minimisation of the noise floor, N_0. This analysis is carried forward in subsequent sections, as each, indicates how and in what combinations factors can obtain high-sensitivity [2].

3.3 Deriving Specifications from Literature Models

The project requirements (Section 1.2, Section 1.3 and Section 2.9) can be extended towards implementation specifications needed for design of a VLC receiver [17, 22, 23]. In this section the statistics of coherent light [80] lead onto quantum limits for an ideal receiver [2, 3] (Section 3.3.1). This eventually leads onto a suite of specifications.

The first specification that is derived, is the SPAD dead time, τ_d [49]. This is a function of data rate, targeted error rate and number of channels that can be accommodated on the device. Despite estimation of the sensitivity (Section 2.9), the power penalties due to receiver efficiency, eye opening or source intensity noise, can account for disparities between theoretical and experimental results [1, 2, 3, 14]. These are useful as they bound implementation and performance values such as the fill factor [81] or the photon detection efficiency [49, 50]. Each high-level specification (Table 2.1) is taken towards a circuit-level specification. These are then summarised in Table 3.8 and Table 3.9.

3.3.1 Estimating the Quantum Limit using Poisson Statistics

An electro-magnetic wave has properties including the wavelength, energy, polarisation, coherence and spatial uniformity with many showing random variation [2, 3, 14, 36, 80, 87, 104, 116]. The Poisson distribution models the number of photons within a period for coherent light [2, 3, 80, 104, 116, 186]. The probability of obtaining k photons in a pulse of mean, $\overline{n_p}$, defined using the rate parameter, $\lambda_r = \overline{n_p}$, is given by Equation (3.2) [80, 116].

$$f(k,\lambda_r) = P(\overline{n_p} = k) = \left(\frac{(\lambda_r t)^k}{k!}\right) e^{-\lambda_r t} \tag{3.2}$$

The inter-arrival time is the period between the arrival of one photon and a second [170]. For circuit timing, dead time and competition between photons, dark counts and after-pulses, the Exponential distribution can be used [170]. This is given by Equation (3.3) [116].

$$f(t,\lambda_r,T_{min}) = \begin{cases} 0, & t < T_{min} \\ \lambda_r e^{-\lambda_r t} & t \geq T_{min} \end{cases} \tag{3.3}$$

Before discussing the quantum limit [2, 3, 14], the integrating receiver can be modelled as the continuous time integration of SPAD detections over a symbol [14, 36, 73]. For this receiver, an integration equal to the symbol period of τ_b is used. The number of counts, C_{out}, is given by Equation (3.4), where N_{SPAD} is the number of SPADs enabled, and $C_{SPAD}(t)$ are SPAD detection pulses occurring at times, t.

$$C_{out} = \int_t^{t+\tau_b} N_{SPAD} C_{SPAD}(t)\ dt \tag{3.4}$$

The quantum limit (QL) [1, 2, 3, 14, 36] is an idealisation of Equation (3.1). It assumes no lower noise distribution, $N_0 = 0$, via the assumption that $P(1|0) = 0$. It also assumes a perfectly Poisson N_1 distribution. In Equation (3.5), $\Psi(\tau_b)$ represents the photons received in a symbol of length τ_b, upon an ideal noiseless photon-counting receiver [2].

$$BER_{QL} = \frac{e^{-\Psi(\tau_b)}}{2} \tag{3.5}$$

In Table 3.2, the quantum limited BER [2] is calculated for various numbers of detected photons. With the BER specification defined as $1x10^{-9}$, the minimum number of photons required is ≈ 20.

Table 3.2: The required number of photons necessary to obtain a specific quantum limited bit error rate. Highlighted in red, a BER of $\approx 1x10^{-9}$ can be obtained with 20 photons.

Number of Photons $\Psi(\tau_b)$	Quantum Limit BER BER_{QL}
1	$1.84x10^{-1}$
6	$1.24x10^{-3}$
13	$1.13x10^{-6}$
15	$1.53x10^{-7}$
20	$1.03x10^{-9}$
25	$6.94x10^{-12}$
30	$4.68x10^{-14}$
33	$2.33x10^{-15}$

As the receiver uses SPADs as the optical detector, the DCR, ambient light and transmitter offsets must be accounted for when estimating the number of detections per second. The total number of detections per second will be used to estimate the SPAD dead time, τ_d and other specifications. Despite decreasing DCRs [53, 59, 144, 157, 172, 175, 187], the collection of multiple SPAD signals into a receiver presents a significant number of dark counts. Likewise after-pulsing and inter-symbol interference each elevate the detections per symbol. Assuming a design margin on the total excess counts of $\approx 10\%$ of the quantum limit, yields 22 detections per symbol. In Section 3.3.5, the nature of this design margin is discussed in more detail. Overall, the 10% margin equates to 2 noise counts per symbol. For 100Mb/s operation this equates to a designed built-in tolerance for $2x10^8$ noise counts per second, important when ambient light conditions and dark count rates in the target CMOS process are unknown. Assuming a 1024 diode structure mirroring the μLED transmitters, gives a designed tolerance for worst case total, (DCR + ambient + extinction ratio + after-pulsing), noise rates of 195kcps per SPAD.

3.3.2 Establishing SPAD Quenching Mode and Dead Time Boundaries

The 22 total detections suggested above can be used to establish the minimum detections per second, for both 100Mb/s and 500Mb/s with a target BER of $1\text{x}10^{-9}$. This can be estimated using Equation (3.6), where R_s is the symbol rate and $D_{BER} = 22$ is the detections required for robust reception [1].

$$\lfloor N_{photons} \rfloor = R_s D_{BER} \tag{3.6}$$

This estimates minimum levels of 2.2Gdetections/s and 11Gdetections/s for 100Mb/s and 500Mb/s respectively. If these are treated as detection events per second, which assumes the receiver is ideal for the moment [1], an upper bound on the SPAD dead time, τ_d, can be estimated. Taking the data rate, the maximum SPAD dead time can be calculated for different numbers of channels, with the SPADs either actively or passively quenched [49]. When using active quenching, the dead time can be calculated using Equation (3.7) [60, 163, 164].

$$\tau_d = \frac{N_{SPADs}}{N_{Channels} \lfloor N_{photons} \rfloor} \quad \text{with} \quad C_{max} = \frac{N_{SPADs}}{N_{Channels} \tau_d} \tag{3.7}$$

Passive quenching suffers from paralysis effects [60, 163, 164] that extend the dead time if avalanches occur prior to full diode recharge. This modifies Equation (3.7) by a factor of e^{-1}, giving Equation (3.8).

$$\tau_d = \frac{N_{SPADs}}{N_{Channels} \lfloor N_{photons} \rfloor e^1} \quad \text{with} \quad C_{max} = \frac{N_{SPADs}}{N_{Channels} \tau_d e^1} \tag{3.8}$$

In Table 3.3, the maximum dead time is calculated for each data rate specification, for different numbers of channels per chip and with active or passive quenching. This shows that a lower dead time is necessary when using passive quenching. However, as discussed, reduction of τ_d may not be achievable with after-pulsing, diode capacitance or quenching resistance constraints [49, 135] (Section 2.7.3).

A range of SPAD dead times are produced by the specifications. Therefore an adaptable SPAD quenching circuit is implemented allowing switching between the active and passive modes and providing a range of dead times for experiments [162]. The range, which for this work will be 5–150ns, complements the reconfigurable number of channels available with the CMOS controlled micro-LED driver arrays [23, 24, 66]. For subsequent estimation of the diode area however, a dead time of $(\text{DataRate})^{-1} = 10\text{ns}$ will be assumed.

Table 3.3: Derived maximum SPAD dead times, calculated for both 100 and 500Mb/s operation, over both active (AQ) and passive (PQ) operating modes. The calculated values assume the receiver is quantum limited.

# Channels		1	2	4	8	16
# SPADs	Per Channel	1024	512	256	128	64
Assuming AQ						
Max τ_d	100Mb/s	465.5ns	232.7ns	116.4ns	58.2ns	29.1ns
Max τ_d	500Mb/s	93.1ns	46.5ns	23.3ns	11.6ns	5.8ns
Assuming PQ						
Max τ_d	100Mb/s	171.2ns	85.6ns	42.8ns	21.4ns	10.7ns
Max τ_d	500Mb/s	34.2ns	17.0ns	18.6ns	4.2ns	2.1ns

3.3.3 Estimating the Quantum Limit Light Level

Given an ideal receiver and estimates for required photons per second, the photon energy (Equation (3.9)) [80] can be used to calculate the quantum limited receiver sensitivity (Equation (3.10) and Equation (3.11)) [2, 67, 103]. This provides a lower bound on light levels required for data transmission [2, 14, 36], along with targets for future designs, where h is the Planck constant, c is the speed of light and λ is the wavelength [80].

$$E_{photon} = \frac{hc}{\lambda} \tag{3.9}$$

$$P_{QLlinear} = N_{photons} E_{photon} \tag{3.10}$$

$$P_{Slog} = 10\log_{10}\left(\frac{P_{QLlinear}}{1x10^{-3}}\right) \tag{3.11}$$

The range given for some specifications, such as the data rate or wavelength, give corresponding ranges for circuit level specifications. This gives quantum limited sensitivities in the range of -53dBm to -62.4dBm (Table 3.4).

Table 3.4: The optical power specification in linear and logarithmic units for the quantum limit at 100Mb/s and 500Mb/s at both 450nm and 750nm wavelengths.

Wavelength (nm)	Data Rate (Mb/s)	Photons/s	Power (nW)	Log10 Power (dBm)
450	100	$2.18x10^9$	0.96	-60.2
750	100	$2.19x10^9$	0.58	-62.4
450	500	$1.10x10^{10}$	4.84	-53.2
750	500	$1.10x10^{10}$	2.91	-55.4

These values define the light level needed if the receiver were 100% efficient and satisfy quantum limited noise requirements [2]. A realistic receiver is unlikely to achieve such performance however, especially when using combinations of technologies that have not been fully explored

for communications. As such, knowing the theoretical minimum light levels, the sensitivity requirement estimated previously (Section 2.9) can be used to calculate maximum allowable values for the power penalties that will limit a practical receiver [2, 14, 36].

3.3.4 Estimating First Order Sensitivity Light Level Requirement

The sensitivity estimate (Section 2.9) establishes an upper bound on the light needed for 450 and 750nm operation. As this is based on previous VLC implementations [18, 20, 23], it assumes a realistic receiver. This gives the specifications below (Table 3.5), where the previous sensitivity of -24dBm, for a data rate of 100Mb/s and a BER of $1\text{x}10^{-9}$ (Section 2.9), implies over $9\text{x}10^{12}$ incident photons per second. This increases to $4.5\text{x}10^{13}$ photons per second, for 500Mb/s operation at 450nm.

Table 3.5: Upper bounds for the optical power and number of incident photons for 100 and 500Mb/s operation, with both 450 and 750ns light.

Data Rates	100Mb/s	500Mb/s
Sensitivity (Log)	-24dBm	-17dBm
Sensitivity (Lin)	3.98μW	19.9μW
450nm Photons/s	$9.0\text{x}10^{12}$	$4.5\text{x}10^{13}$
750nm Photons/s	$1.5\text{x}10^{13}$	$7.5\text{x}10^{13}$

The flux is now in the form of lower and upper bounds. The lower bound is a quantum sensitivity that calculates the photons that must be detected by an ideal receiver [2]. In contrast the upper bound assumes realistic behaviour but uses incident rather than detected photons [18]. The difference between the bounds, can therefore be used to establish two constraints, albeit difficult to separate. These are the total losses due to finite efficiencies between the incident photons and detections, and the receiver effects giving realistic rather than quantum limited performance. The mechanisms behind these efficiency losses or non-ideal characteristics are called receiver sensitivity power penalties [1, 14, 36].

3.3.5 Power Penalties and Circuit Specification Estimation

From the above sensitivity constraints, the total power penalty can be estimated [14, 36]. Using the ratio between the $2.2\text{x}10^9$ detections needed assuming a 100% efficient, quantum limited receiver at 450nm and 100Mb/s (Table 3.4), and the $9\text{x}10^{12}$ detections needed for the estimated realistic sensitivity of -24dBm, also at 450nm and 100Mb/s (Table 3.5), the total power penalty, PP_{tot} is 36dB (Equation (3.12)). This increases to 38dB for 750nm light at 100Mb/s. The change between 36dB and 38dB is principally due to the change in photon energy (Equation (3.9)), which changes the number of photons incident for the same optical power. Rather than keeping the optical power static, the number of photons needed for the

quantum limit remains static at 2.19x10^9, while the literature-based sensitivity is used in reverse, via Equation (3.9) to Equation (3.11), to estimate the number of photons within a beam of that optical power at the new wavelength.

$$PP_{tot} = -10\log_{10}\left(\frac{2.19x10^9}{9x10^{12}}\right) = 36dB \tag{3.12}$$

This, will be formed by the detector optical fill factor, the photon detection efficiency, the eye opening penalty (EOP), any threshold offsets, the optical source intensity noise and the jitter of the clock that defines the integration period [3, 14, 36]. These are below with the most significant power penalties discussed first.

Finite Fill Factor Induced Power Penalty

The fill factor for SPAD arrays using in-pixel circuitry, or those constrained by manufacturing design rules, are often limited to 1–5% [70, 77, 128, 155]. Dedicated high fill factor structures can be built, which with combinations of techniques such as N-Well sharing [141], large diameter active areas [144], small radius guard rings [53] and non-circular geometries [53], can reach optical fill factors as high at 70% [56, 57, 145, 158, 180, 188, 189].

To achieve high fill factor, circuitry should be removed from the sensitive area [56, 57, 157, 158, 185]. In this work, the CMOS process [30] and SPADs [143], along with design rules and the use of tested structures, will limit the fill factor. The critical design rule, the N-Well separation, forces inactive areas within the array. These should be used for circuitry to ensure silicon area efficiency, and a scalable routing structure [29, 69].

Presuming a fill factor similar to previous arrays, that did not use the high fill factor techniques [70, 155], a fill factor of $\approx 2\%$ can be assumed. The power penalty due to the fill factor, PP_{FF}, can be estimated to be 17dB, through Equation (3.13).

$$PP_{FF} = -10\log_{10}\left(\frac{FF}{100}\right) = 17dB \tag{3.13}$$

Finite PDE Induced Power Penalty

Many CMOS SPADs have photon detection efficiencies within the 15–30% range [51, 52, 55, 59, 141, 157, 172, 190, 191], although higher values ($\geq 40\%$) have been reported [53, 56, 144, 182, 187]. Assuming the SPADs within the CMOS process follow previously demonstrated performance [53, 143, 173], a PDE of 20% can be used. This gives a power penalty, PP_{PDE} of $\approx$ 7dB (Equation (3.14)).

$$PP_{PDE} = -10\log_{10}\left(\frac{PDE}{100}\right) = 7dB \tag{3.14}$$

Combined, the PDE and fill factor, give a power penalty of $\approx$ 24dB, away from the quantum limit. The remaining difference, must be due to effects such as the eye opening penalty [14]. These are difficult to estimate, prior to experimentation, due to the unique combination of SPADs and integrating receivers. For the specifications, the remaining total penalty, along with estimates of various effects, can be used to establish a power penalty budget [1, 14, 36].

Source Intensity Noise Induced Power Penalty

Intensity noise describes fluctuation in transmitter optical intensity due to power and source variations [14, 36]. Assuming that the power and temperature of the source are well controlled, the main contributors to intensity noise are the *wave fluctuations* [80] in the light itself. The relative intensity noise (RIN) adds a further power penalty, PP_{RIN}, given by Equation (3.15) [14, 36]. Where $Q = 6$, is the Q-factor for the target BER of 1x10^{-9}, and R_{int} parameterises the intensity fluctuations [36].

$$PP_{RIN} = -10\log_{10}\left[1 - R_{int}^2 Q^2\right] = 7.2dB \tag{3.15}$$

For coherent laser sources, RIN may be better than -110dB/Hz and in exceptional cases ≤ -160dB/Hz [14, 36]. However, the application here uses multiple spontaneous emission LED devices, with less stringent power supply and temperature controls [23, 24, 96, 98]. Further, the application includes path reflections [16, 17, 20, 22]. These suggest a RIN and subsequent intensity noise far higher than communication lasers.

There is a critical value for R_{int}, above which there is a steep increase in PP_{RIN} [36]. For the target BER this is approximately 0.15 [14, 36], equating to a PP_{RIN} of 7.2dB. Equation (3.16), where BW is the receiver target bandwidth of $\approx$ 100MHz, allows the approximate calculation of a maximum source relative intensity noise [36].

$$max(RIN_{\mu LED}) = 10\log_{10}\left(\frac{R_{int}^2}{2BW}\right) \tag{3.16}$$

For power penalty budgeting, a value of 5dB can be used [36]. This gives a suitable margin from the critical value, without suggesting a temperature controlled, coherent source with well conditioned power supplies. This equates to an R_{int} of 0.14, and $RIN = -100$dB/Hz. Others have measured the RIN of LED sources [192, 193, 194, 195], however the value for arrays of μLEDs using CMOS controlled drivers has not been assessed.

The Eye Opening Power Penalty

The eye opening penalty (EOP) is a measure of the closure of the ideal transmitted data eye diagram due to channel distortion effects and the transmitter extinction ratio [3, 106]. While unknown at this point, assuming a received amplitude 75% of that transmitted, the power penalty due to the eye opening, PP_{EOP}, can be calculated by Equation (3.17) [14, 36]. Where EO_{ref} is the opening of the reference transmitted data, and EO_{rec} is the eye opening of the data incident on the receiver. The extinction ratio may dominate the receiver DCR noise floor, however this work aims to assess the receiver only performance.

$$PP_{EOP} = 20\log_{10}\left(\frac{EO_{ref}}{EO_{rec}}\right) = 2.5dB \tag{3.17}$$

Inter-Symbol Interference Induced Power Penalty

Inter-symbol interference will also close the received eye diagram [14, 36]. For example if a proportion of a transmitted *logic-high* leaks into a subsequent *logic-low* bit, the signal mean and variance, N_0 and σ_0 will be elevated, increasing the erroneous bit probability. The power penalty due to ISI, PP_{ISI}, can be estimated using Equation (3.18) [14]. Which, if it is assumed to be reduced to $EO_{isi} = 0.8EO_{rec}$, gives a penalty of 2dB.

$$PP_{ISI} = 20\log_{10}\left(\frac{EO_{rec}}{EO_{isi}}\right) = 1.94dB \tag{3.18}$$

Finite DCR Induced Power Penalty

The detection of dark count, ambient light and un-modulated light events adds a threshold offset power penalty, PP_{DCR} [14, 36]. The offset power penalty is given by Equation (3.19), where DCR_{pc} is the total DCR per receiver channel and is assumed to be a worst case of 10% per channel, per symbol, dark count noise design margin, relative to the 20 photons required for the target BER (Table 3.2). In this work, the above design margin is chosen to provide some coverage for three as-yet-unknown factors, a) the average per-SPAD DCR, b) the finite but designed to be 100% extinction ratio of the μLED and CMOS drivers and c) the narrow, filtered 450nm ambient light contribution. For the dark count rate, previous work in an AMS 0.35μm CMOS process pointed to high, $\approx 1-2$kcps, average per-SPAD DCRs, and a significant number of high-noise devices, $\approx 1-2$Mcps. Some data was available from ST Microelectronics CMOS SPADs, however recent waver runs had shown significant average DCR changes of 100cps to $100-200$kcps, due to an N-well doping issue. In Section 5.2.2 the DCR of the receiver front-end developed in this work is shown experimentally, suggesting that a value of the order of 10% is suitable for preliminary design built-in tolerance to noise. This 10% margin gives a penalty of 0.8dB. Assuming 100Mb/s operation and 1024 SPADs in a

single channel, this gives a worst case SPAD total noise of 195kcps. A target such as this, should be achievable in terms of DCR given modern low noise SPAD designs [53], with the excess margin once DCR is accounted for, providing receiver tolerance for ambient light conditions or temperature [143] or fabrication variations. It should be noted, through Equation (3.19), that as we are interested only in the number of 'signal' photon detections, counts arising from within the SPAD, if probable, will mean that more signal photon counts are needed to separate an 'one' from a 'zero', i.e. the power penalty below.

$$PP_{DCR} = 10\log_{10}\left(1 + \frac{2DCR_{pc}}{100}\right) = 0.8dB \tag{3.19}$$

Other Circuit-Implementation-Based Power Penalties

The clock generating the integration period [73], or used in single sample receivers to sample the waveform [2, 35], contributes a further power penalty. For this jitter power penalty, Cvijetic [36] suggests a minimum of ≈ 0.5dB should be assumed to account for this effect.

Other circuit behaviours such as processing errors, signal quantisation or cross talk, may add further power penalties, PP_{spare}. However, the effect of these on the sensitivity is difficult to establish prior to experimentation. With correct digital design, such as pipe-lining or path equalisation [29, 69], the digital signal path should add little noise. Likewise with a suitable digital dynamic range, little quantisation other than natural Poisson variation should result.

Detector cross talk, which is often at the 1–2% level in SPAD arrays using out of array circuitry [56, 57, 147, 180], represents cross talk induced by secondary photons or hot electrons in the substrate [76, 147]. This is SPAD and geometry dependant, however in-array circuitry has also been recognised as a contributor [56, 76, 149, 150]. For design therefore, while area within the array should be used to maximise the functionality to area cost ratio, high-speed circuitry, clock buffers and digital supplies should be placed outside the optically active areas.

Receiver Sensitivity Power Penalty Budget

The separation of the quantum limit and a realistic upper sensitivity bound produced from prior art or required specifications, can be used to budget the receiver power penalties. In Table 3.6, each of the penalties are removed from the total. This leaves $PP_{spare} = 1.25$dB for other as yet unknown power penalties, such as the cross talk.

This penalty budget, while a first order estimation of the effects and their relative impacts, can be used to ensure circuit implementation has known constraints. The design of the receiver, can then aim for better performance, decreasing the gap between the ideal theoretical quantum limit and that realisable with a novel combination of receiver technologies.

Table 3.6: A power penalty budget used to obtain approximate design constraints. This is ordered in descending dB value, with the most significant power penalties towards the top.

Penalty	Amount (dB)	Remaining (dB)
PP_{FF}	17	19.0
PP_{PDE}	7.0	12.0
PP_{RIN}	5.0	7.0
PP_{EOP}	2.5	4.5
PP_{ISI}	1.95	2.5
PP_{DCR}	0.8	1.7
PP_{jitter}	0.5	1.25
PP_{spare}	1.25	0.0
Total Penalty	36	

3.3.6 SPAD Specifications for a Multi-Diode Receiver Array

In this section, parameters for the low level design are discussed. This covers the structure and electrical specifications of the single-photon avalanche photodiode along with the specifications of an array of such diodes and the theoretical properties of receivers using multiple diodes. For example, the diameter of a SPAD or the theoretical maximum fill factor are both estimated, as is the approximate theoretical bandwidth of the system. The maximum fill factor of a small array is presented, based upon the diode diameter and CMOS process used within this work. The outcome of this chapter is a further set of specifications for a prototype SPAD-based OOK receiver for VLC applications. The prototype is used to test the concept, while providing an experimental platform to investigate both as yet unknown phenomena and future routes to design specifications.

Estimating the p-n Junction Depth

Optical energy penetrates, to a wavelength and material dependent depth [26]. Equation (3.20) describes an exponential decrease in the flux, Φ, with depth, x_d (m), into a material with an absorption coefficient of α_c (m^{-1}).

$$\Phi(x_d) = \Phi_0 e^{-\alpha_c x_d} \tag{3.20}$$

For silicon, the absorption coefficient changes between 2.55x10^6 and 1.30x10^5 m^{-1} for 450nm to 750nm light [26]. This equates to depths of 0.4μm and 7.7μm respectively. These are characterised by an e^{-1} reduction in the flux and specify ideal depths for a photodiode junction or SPAD multiplication region [26]. Photons are absorbed above and below the penetration depth [26], however a junction matching the wavelength and a large junction width will maximise PDEs [26, 182]. Given the foundry specified junctions available in CMOS [30], a depth of 0.5μm and an approximate asymmetric 1μm junction width can be assumed [143]. A CMOS SPAD is available [143] with low dark count and after-pulsing rates. This has a PDE of $\approx 25\%$.

Estimating the Diode Capacitance and Internal Resistance

The size of SPADs directly affects performance [53]. Knowing the SPAD dead times that are needed (Table 3.3), the capacitance and therefore the area can be estimated [26, 173]. Equation (3.21) [49] gives an estimate of the passive quench passive reset (PQPR) dead time τ_d, where τ_q and τ_r are the quenching and recharging time constants. R_D and R_L are the diode and load resistances, C_D and C_S are the diode and parasitic capacitances and p is how close the voltage exponentially approaches the asymptote (i.e. 99.9%) [49].

$$\begin{aligned}\tau_d &= -\tau_q \ln(1-p) - \tau_r \ln(1-p) \\ &= -R_D(C_D+C_S)\ln(1-p) - R_L(C_D+C_S)\ln(1-p)\end{aligned} \tag{3.21}$$

The SPAD discharge occurs within 10–100ps [49, 125, 127], due to the speed and magnitude of avalanche currents. Assuming a 10ns dead time (Section 3.3.2), this equates to ≈ 0.1–1% of the total. The capacitance can be estimated by setting the parasitic capacitances to the CMOS substrate, to $\approx 1\%$ of the diode capacitance, C_D, the precision value, p, to 99%, and assuming both a worst case 100ps rise time and a diode resistance, R_D of $1k\Omega$ [49]. This gives a diode capacitance of $C_D \approx 21.5$fF, matching [143].

Estimating the Load Resistance in Passive Quenching Mode

Moving to the larger component of the passive recharging time of the SPAD [49, 60, 163, 196], the quenching load resistance can be approximated. This is often in the 100-500$k\Omega$ region [49], however care must be taken. A large R_L, quenches the diode but presents a significant time constant. A small R_L, yields reduced recharge constants and shorter dead times, however it runs a risk of only partial quenching [49, 127, 138, 152]. This gives an estimated passive quenching resistance, for the target 10ns dead time, of $\approx 100k\Omega$.

Estimating the SPAD Breakdown Voltage, Diameter and Junction Width

The diameter of the SPAD can be estimated from the junction capacitance (Equation (3.24)) [26]. This uses the estimated built-in voltage, calculated using Equation (3.22), and the approximate breakdown voltage, estimated from Equation (3.23) [26]. The built-in potential, V_{bi}, defined under thermal equilibrium and zero applied bias, is formed by the diffusion of carriers over the depletion region [26]. This leaves fixed ions on either side, with V_{bi} then the difference of potential over the region [26]. V_{bi} is calculated via Equation (3.22), where K is the Boltzmann constant, T is the junction temperature in Kelvins, q is the electronic charge, n_i^2 is the intrinsic substrate carrier density and N_A and N_D are the proprietary acceptor and donor concentrations respectively [26]. For the CMOS process used [30], V_{bi} is 0.825V [173].

$$V_{bi} = \frac{KT}{q} \ln\left(\frac{N_A N_D}{n_i^2}\right) \tag{3.22}$$

The break down voltage, VBD can also be estimated, giving an indication of voltage requirements for Geiger-Mode operation [26, 127, 173]. This is calculated using Equation (3.23), where E_{max} is the maximum electric field strength and ε_s and ε_0 are the relative silicon and vacuum permittivities [26]. This estimates junction breakdown at -14.35V [173].

$$V_{BD} = \frac{\varepsilon_s \varepsilon_0 (N_A + N_D)}{2qN_A N_D} (E_{max})^2 \tag{3.23}$$

The diameter, $Dia(m)$, measured in metres, is then estimated by Equation (3.24) [26]. Assuming the previously estimated capacitance of $C_D = 21.5$fF, the SPAD must have a diameter no larger than $Dia(m) = 8.6\mu$m, matching [143].

$$Dia(m) = \frac{2}{1x10^2} \sqrt{\frac{C_D}{\pi} \sqrt{\frac{\varepsilon_s \varepsilon_0 q N_A N_D}{2(N_A + N_D)}} \sqrt{\frac{1}{\sqrt{V_{bi} + V_{BD}}}}} \tag{3.24}$$

Estimating the Theoretical Maximum Optical Fill Factor

As the fill factor is the ratio of the optically active area to the total area of the silicon receiver, the fill factor induced power penalty reduces the achievable receiver sensitivity by approximately 17dB (Equation (3.13)). In order for future designs to achieve much higher receiver sensitivities, both the achievable fill factor and the quantum efficiency are of interest [156]. Fundamentally, if a photon happens to be incident on a portion of the receiver that is covered by the metallisation used in CMOS processes, or hits a non-optically active circuit, it cannot be detected. To address this issue, precise spatial organisation of the SPAD array, the introduction of advanced SPAD array topologies or micro-lenses that focus the incident photons at the centre of each SPAD, may be advantageous [156]. In a receiver constructed of multiple diodes, the fill factor will be reduced [68, 70, 155], from what may be theoretically possible. This is due to manufacturing design rules [30], SPAD geometries [53, 143] and near-SPAD circuitry [56, 162]. An estimate of the theoretical maximum fill factor, allows assessment of how practical it would be to increase the fill factor, may impact the layout or circuit specifications of the SPAD array and gives an estimate of the receiver sensitivity that could be achieved in the CMOS process [156].

The fill factor, with a circular diode of active radius D_r, separate N-Wells and no electrical circuits within the array, can be calculated for a square SPAD array using Equation (3.25). Where D_g is the guard ring width and D_s is the guard ring separation [56].

$$FF_{sq} = \frac{ActArea}{ArrArea_{sq}} = \frac{\pi D_r^2 N_{SPAD}}{[2D_r\sqrt{N_{SPAD}} + (2D_g + D_s)(\sqrt{N_{SPAD}} - 1)]^2} \tag{3.25}$$

Circular SPADs are not area efficient [53], therefore square arrays suggest use of square SPADs with rounded corners (hyper-ellipse) [53]. These can tend towards higher fill factors than circular devices, but experience elevated DCR [53]. Equation (3.25) can be used to estimate the theoretical maximum array fill factor. With the estimated diameter of 8.6μm and $N_{SPAD} = 1024$, the total optically active area will be approximately $0.06mm^2$. Using specific values for D_g and D_s [53, 143], a maximum fill factor, $FF_{max} = FF_{sq}$ of 18.8% is obtained (Figure 3.2). With circuitry and increased diode separations due to manufacturing rules, the obtainable fill factor will decrease rapidly.

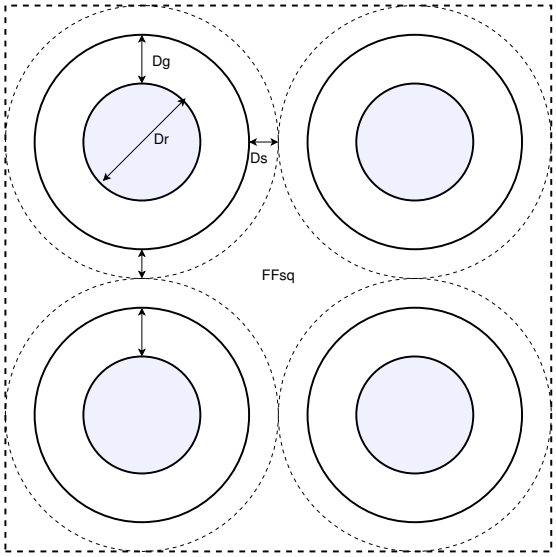


Figure 3.2: Theoretical fill factor of a square array of circular SPADs. Each SPAD has an active area diameter of D_r, a guard ring width of D_g and a required guard ring to guard ring separation of D_s. Circles are not an efficient use of area in this topology indicating the use of hexagonal arrangements, or hyper-elliptical SPADs.

The fill factor, and its large effect on the receiver sensitivity, is key in obtaining a highly sensitive receiver (Equation (3.13)). The obtainable fill factor in this work is limited to approximately 2 to 3%. This means that more signal power is needed to obtain the target BER. As fill factors improve, or as optical solutions such as micro-lenses are used, the overall sensitivity of SPAD-based receivers will increase [156], in a similar fashion to low loss cables developed for long distance wired communications [1].

Estimating the Theoretical Architecture-Independent Data Rate

The analysis above proceeded from specifications in Chapter 1. However the maximum data bandwidth (Mb/s) that could be achieved is also of interest, particularly for future receiver designs [17, 67]. The digital theoretical maximum data rate, BW_{th}, based upon [2], of a noiseless integrating SPAD receiver can be calculated independently from the architecture (Equation (3.26)). This is the native data rate of SPADs as development pushes towards an ideal receiver, independent of clock frequencies, output of the signal, non-ideal effects such as the impulse response and the exact receiver implementation.

$$BW_{th} = \frac{N_{SPAD}}{-\ln(2BER)\tau_d} \tag{3.26}$$

The theoretical architecture-independent data bandwidth is given using different numbers of SPADs at the target BER, in Table 3.7. This assumes a) common dead times of 10ns, b) that all SPADs are turned on and operating at the same count rate, alongside c) a noiseless integration-mode receiver.

Table 3.7: The theoretical architecture-independant data rate achievable for a BER of $1\text{x}10^{-9}$, for different numbers of SPADs per channel.

#SPADs	1024	512	256	128	64
BW_{th}	5.12Gb/s	2.56Gb/s	1.28Gb/s	640Mb/s	640Mb/s

Estimating the Theoretical Nyquist Bandwidth

The Nyquist frequency, F_{Ny}, for analogue modulation standards can also be investigated. This would be important where the sampling rate is a key issue. Two such modulation standards would be orthogonal frequency division multiplexing [17] and traditional amplitude modulation [1, 14, 35]. The Nyquist frequency is estimated, assuming non-uniform sampling due to photon arrival statistics, by Equation (3.27) [2, 14, 80, 116]. Where the operator $E()$ is the expected value or mean [116], and C_{max} is the maximum count rate for SPADs in the receiver (Equation (3.7)).

$$F_{Ny} = \frac{N_{SPAD}E(C_{max})}{2} = \frac{N_{SPAD}}{2\tau_d} \tag{3.27}$$

For 1024 SPADs and a dead time of 10ns (Section 3.3.2), this gives $F_{Ny} \geq 50$GHz. This gives a second estimation as to the theoretical bandwidth that could be achieved with development. Despite this, the troughs of an analogue signal are sampled at a lower count rate, C, due to the lower light intensity and decreased detection rate [80, 163, 165]. An ideal receiver would need to arrange for the maximum modulation bandwidth, BW_{Ny}, to satisfy the Nyquist condition for the average count rate of the minimum signal amplitude [1].

Let β be an average optical power, about which a signal of amplitude α_{mod} is modulated. For an input of the form, $S_{in}(t) = \alpha_{mod}\sin(2\pi f_{mod}t + \phi) + \beta$, with a frequency of f_{mod}, counting rates for a given intensity written as $E[C(\beta)]$ and an absolute minimum count rate of $E[C(\beta - \alpha_{mod})]$, the theoretical bandwidth, BW_{avail}, can be calculated using Equation (3.28).

$$BW_{avail} = \frac{E[C(\beta - \alpha_{mod})]}{2} = \frac{E[C(\beta)] - E[C(\alpha_{mod})]}{2} \tag{3.28}$$

Letting the operating point, $C(\beta)$ be a decade below C_{max} to ensure no non-linearities or clipping, gives $C(\beta) = 10.24$Gphotons/s, with 1024 SPADs and $\tau_d = 10$ns. Presuming the input signal gives a modulation of 10% of $C(\beta)$, $\alpha_{mod} = 1$Gphotons/s. The bandwidth would then be 4.61GHz, slightly below BW_{th}. This scales linearly to 2.3GHz, 1.15GHz, 575MHz and 287.5MHz for 512, 256, 128 and 64 SPAD receivers.

3.3.7 A Multi-SPAD Receiver Array: Conclusions and Specifications

This section has presented a suite of low-level specifications suitable for the implementation, in a CMOS technology, of a SPAD-based highly-sensitive receiver for VLC applications. Throughout, the already present project (Chapter 1) and literature-review based (Chapter 2) requirements, have been used to obtain specifications such as the approximate range of the dead time, the required number of photons per second or even the approximate diameter of the SPAD to be fabricated in the planar CMOS structure.

In the next section, the issue of dead time scalability prompts a method of mitigating non-linear SPAD behaviour. The specifications obtained in this and subsequent sections is then revisited and summarised in Section 3.5, prior to the full discussion, design and implementation of the SPAD-based receiver in Chapter 4.

3.4 Extending Linearity Through Time Division Multiplexing

In the previous section, the specifications allowed the estimation of the dead time (Table 3.3). This assumes that the SPADs, while obtaining the ≈ 20 photons needed for the target BER (Table 3.2), are operating at the maximum counting frequency, C_{max} (Equation (3.7)) [163]. In realistic SPAD systems, the count rate saturates [163] as the optical power comes close to C_{max}, creating a non-linear transfer curve between input photons and output counts [49, 137, 163]. Passively quenched systems also exhibit count paralysis, whereby the number of counts reduces after saturation [60, 163, 164].

In this section, the approach to estimating the dead time specification is retained, but methods of reducing non-linear behaviour are explored. This is used primarily because first order *ideal* dead time estimation may indicate a low dead time is required, although the diode capacitance, structure, quenching circuits or after-pulsing may prevent scaling of the dead time to this low value. As a numerical example, Table 3.3 indicated that a dead time of 2.1ns may be required. If the dead time is fixed at 10ns by a combination of diode diameter and quenching resistance specifications, some technique will be needed to retain correct functionality.

Linearity is less important for binary schemes such as OOK modulation in comparison to analogue modulation schemes, however, depending on the number of SPADs, data rate and BER, the dead time that is estimated may be unobtainable, difficult to scale or present issues for OOK reception (Section 2.7.3). With 16 channels, 64 SPADs per channel and a data rate of 500Mb/s, a dead time of 2.1ns was previously calculated (Table 3.3). For CMOS SPADs, dead times as low as this have not been shown due to the increase in after-pulsing events. Complementing this, use of a dead time shorter than required represents a) a significant design effort if diode and parasitic capacitances are large [49, 134, 162, 169], b) an under-utilisation of available counting resources, and c) an increase in active quenching circuit power though increased diode recharging currents [29, 162].

The impracticality of scaling the dead time [76], implies that a form of linearisation is needed to ensure OOK signals remain error free. This would allow a more practical dead time, for example 10ns rather than 2.1ns to be used without count saturation or paralysis affecting the performance [163]. Count rate paralysis, where the count rate begins to decrease despite higher optical powers, represents the worst case due to the one-to-many mapping that can result [60, 163]. This would severely corrupt the N_1 and N_0 distributions, increasing the BER, however overall gain linearity may also be important.

In [1] and Section 3.2, the data is decoded from an OOK transmission using a single threshold. However the BER is dependent on the mean signal levels, N_0 and N_1 relative to this threshold, and the standard deviations of the two distributions. The closer a distribution to the threshold, N_{th} or the wider the distribution, σ_0 or σ_1, the more the distribution's tail will cross over the threshold [1, 2]. To understand why gain linearity may be important, a numerical example

can be used. If the transmitted N_0 distribution is 0.5 photons/symbol with a receiver gain of, $G = 1$, the received *zeros* distribution will be 0.5 detection per symbol. If the transmitted N_1 is for example 24 photons/symbol, a linear receiver with a consistent gain of 1 would output 24 detections per symbol. Using Equation (3.1) and assuming Poisson limited distributions, this would give a certain BER, in this case a BER of $1.4\text{x}10^{-5}$. However, if the receiver contained a non-linearity of gain, such as the SPAD saturation effect due to the dead time [163], the N_0 gain may well be $G = 1$, but the N_1 gain may instead be 0.75. In this case, the transmitted N_1 of 24 photons per symbol may be artificially reduced, by this lower gain, to 18 detections/symbol. The smaller gap between N_1 and N_{th}, would yield a poorer BER of $2\text{x}10^{-4}$.

3.4.1 Time Division Multiplexing (TDM)

Time division multiplexing (TDM) has been used in photon counting systems [105, 119, 178, 197], directly countering non-linearities. Likewise space division multiplexing (SDM) has been used, partially removing the influence of dead times [168, 177, 198]. In Figure 3.3, TDM is graphically demonstrated. In this mode, called $N_a = 1$, a single SPAD from a group of N_{TDM} SPADs, is enabled at a time. A state machine [69] cycles round the SPADs, holding a fired SPAD off until the state machine re-enables it. In Figure 3.3, $SPAD_1$ is initially enabled. Upon a photon arrival, $SPAD_2$ is enabled after a short unresponsive switching period. The arrival of subsequent photons, iterates the circuit to subsequent SPADs before re-enabling $SPAD_1$ [178].

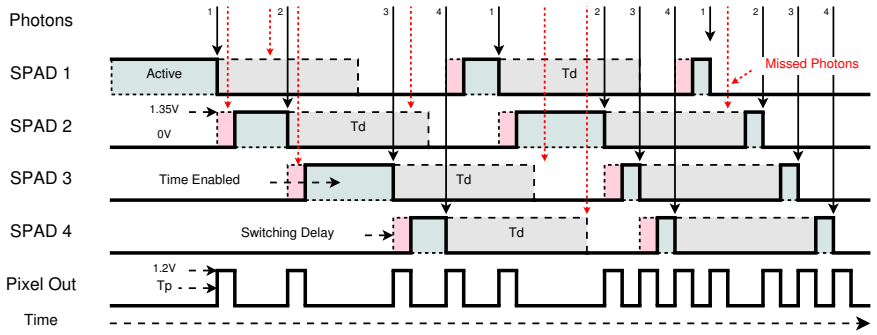


Figure 3.3: Overview of SPAD switching, received and missed photons, and the relative times of SPAD availability as flux increases to the right hand side. Light blue represents SPAD enabled periods, light grey the SPAD dead time, and light red the fully unresponsive periods while TDM switches between diodes.

The dead times are highlighted in grey, however a diode may remain in a hold-off period (Section 2.7.3). During the active period of $SPAD_2$, all photons are missed on other SPADs as $SPAD_1$ is within its dead time, and SPADs 3 and 4 are unbiased. At the right (Figure 3.3), the flux has increased, demonstrating two points.

- The enable time, and thus the time available for a noise count has been reduced [178]
- τ_d leads to groups of pulses. C_{max} is $\frac{N_{TDM}}{\tau_d}$, however the minimum separation and resolution of short inter-arrival times, is a product of switching delays [178].

The switching of SPADs is crucial to the concept, fortunately it acts to reduce after-pulsing and dark count rates of constituent SPADs [62, 178, 179, 180, 181]. This is achieved as the release of trapped carriers, or carriers left in the multiplication region [76], cannot trigger a second avalanche. Enabling a single SPAD and cycling round the group, is one possible TDM mode. A second, denoted as $N_a = 2$, iterates enabling two SPADs at a time. This modifies the hold-off times, reducing the TDM affect on after-pulsing and dark count rates, but increases aggregate sensitivities. Further modes are feasible, however, modes with non-integer values for the ratio, $\frac{N_{TDM}}{N_a}$, where N_{TDM} is the number of SPADs within a group, are not used as they represent a time varying effective photon detection efficiency, $ePDE$. The final mode, given as $N_a = N_{TDM}$, does not represent time division but spatial multiplexing [168], as each SPAD remains enabled. In this mode, the grouping allows each SPAD to fire, quench and be held-off, without impacting other diodes.

3.4.2 TDM SPAD Group Weightings and Applications

The TDM modes, give a mode-dependent change in the aggregate sensitivity by changing the losses involved in a particular TDM operation and the change in the loss behaviour over different optical powers. For a pixel formed from $N_{TDM} = 4$ SPADs, the modes give optical fill factor weightings of $\frac{N_a}{N_{TDM}} = \{0.25, 0.5, 0.75, 1\}$, which globally act upon the sensitivity of the group at all light levels. The state dependent sensitivity, $ePDE$, of the $N_a = 3$ mode is problematic with $N_{TDM} = 4$ SPADs, suggesting this should be used as a way of disabling all SPADs within a group. The weightings then become, $\{0, 0.25, 0.5, 1\}$.

The $N_a = 4$ mode, gives a weighting of $\frac{N_a}{N_{TDM}} = 1$, as all SPADs within the group have the same probability of being struck by an incident photon and the same probability of avalanche [168]. Conversely the $N_a = 1$ TDM mode, will give a weighting of $\frac{1}{4} = 0.25$ [178] as only the SPAD that is currently enabled is able to produce an avalanche [181].

The TDM sensitivity can be explained by referring to the $N_a = 2$ mode. Here, at any instant, two of $N_{TDM} = 4$ SPADs, are enabled giving a $\frac{2}{4} = 0.5$ fill factor weighting. As such, an incident photon hitting either creates an avalanche and a detection pulse on the aggregate output. The increase in sensitivity can also be explained probabilistically through the logical OR of the two active SPADs, where the total output event probability, $P_{event}(Output)$ is given by Equation (3.29) [116, 199].

$$P_{event}(Output) = P_{event}(SPAD_1) + P_{event}(SPAD_2) \tag{3.29}$$

The firing of $SPAD_1$ and $SPAD_2$ are not mutually exclusive [116] as they can occur at the same time. However as output pulses still arise from events in $SPAD_1$ AND $SPAD_2$, the non-mutually exclusive probability is increased by $P_{event}(SPAD_1)P_{event}(SPAD_2)$.

This conceptual view of the weightings may hold near maximum counting rates (Equation (3.7)), where a SPAD will fire once it is enabled, forcing each of the SPADs to have the same probability of an event [178]. The loss behaviour and overall sensitivity will also change over different optical powers (Figure 3.5) as the time gating operation of the circuit changes speed (Section 3.4.1). For photo-detections, the responsivity or photons-in to counts-out gain may also change, but may do so around the fixed weightings discussed above. The TDM concept must be modelled however to understand its affects when the flux is lower than $\frac{1}{\tau_d}$, or when after-pulsing or dark count rates are considered at the group level [178]. This is discussed in the next sub-section.

3.4.3 Modelling The Time Division Multiplexing of SPADs

Castelletto et al [178] have modelled a TDM Geiger-mode APD system, which is used here to investigate affects on linearity. The concept has been further modelled and investigated in [119, 177, 197, 200, 201]. In Figure 3.4, the photon flux has been swept with a SPAD dead time of $\tau_d = 10$ns (Section 3.3.2). The solid coloured lines represent the actively quenched SPAD response [60, 163, 164, 166, 202]. With a single detector, the maximum count rate is equal to $\frac{1}{\tau_d}$ [163], over the entire simulated range (1x10^4 to 1x10^{11} photons/s). This increases linearly at the extreme right of the graph as N_{TDM} increases. The solid black lines represent the total output maximum frequency, with the group of N_{TDM} SPADs in the $N_a = 1$ mode.

Adding more SPADs to the group, increasing N_{TDM}, but retaining the $N_a = 1$ mode, the maximum frequency increases non-linearly. As expected, the maximum scales with N_{TDM}, as shown by the saturation towards the asymptotes at $\frac{N_{TDM}}{\tau_d}$. At low photon counts however, the maximums for any group of N_{TDM} SPADs returns to an effective $N_{TDM} = 1$ maximum count rate. This hints at two implications of TDM. Firstly that the N_0 distribution for OOK has a lower gain, keeping it below the N_1 distribution, and secondly that infrequent after-pulsing and dark counts will be reduced (Section 5.2.2).

In Figure 3.4, the dashed coloured lines represent the combination of the SPAD transfer curve [60, 163, 164, 166, 202] and the TDM transfer curve [119, 178]. This is the actual count rate with photon flux observed at the SPAD group output. An offset is observed between the SPAD only and combined transfer functions, indicative of the shift in $ePDE$.

Plotting Figure 3.5 in the log-log domain and including the 2 dark detections per symbol (with data rate of 100Mb/s and 1024 SPADs per channel), reveals a departure from the single-SPAD or N_{TDM}-SPAD response (Figure 3.4). When operated in the $N_a = 1$ mode, TDM acts to increase the linear region slightly. This is shown at the top right of Figure 3.5, where despite the $ePDE$ offset, the dashed coloured lines become non-linear at a higher flux for the same dead time.

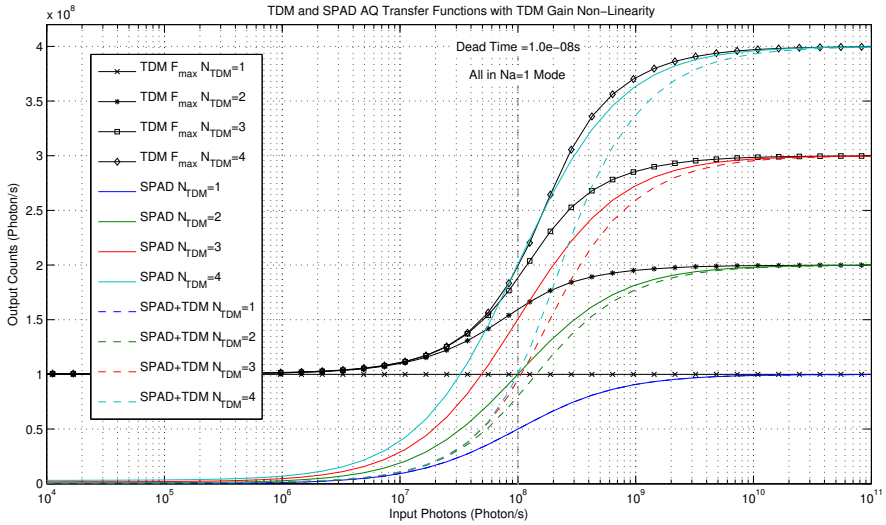


Figure 3.4: Modelling of the photon transfer or receiver gain response for both *SPAD-Only* and combined *SPAD and TDM*. The black lines represent the absolute maximum counting rate achievable for N_{TDM} SPADs, while the solid coloured lines represent the *SPAD-Only* transfer curves. The coloured dashed lines present the combination of the SPAD and TDM circuit responses over a simulated optical input dynamic range.

The previous hints towards after-pulsing and DCR rejection and the lower count level for the OOK N_0 distribution, is seen towards the lower left of Figure 3.5. For $N_{TDM} = 4$, the $N_a = 1$ mode has increased the dynamic range, as indicated by the arrow. Discounting the fill factor weightings caused by TDM for the moment, as the SPAD+TDM traces change gradient over the optical range, further modelling is needed to understand the causes of the change in photons-in to counts-out gain. This change in gain can be seen in Figure 3.5 by comparing the cyan $N_{TDM} = 4$ SPAD-only and SPAD+TDM traces. At a flux of 10^7 photons/s the ratio is clearly $\frac{1x10^7}{4x10^7} = 0.25$, matching the $\frac{1}{4}$ fill factor weighting previously discussed. However one decade higher at 10^8 photons/s the ratio has increased to $\frac{1x10^8}{2x10^8} = 0.5$, i.e. a lower loss. Moving higher to 10^9 photons/s the ratio is almost unity (1). If conversely we move to lower flux levels such as 10^5, the ratio remains the same as the overall TDM fill factor weighting, namely $\frac{8x10^5}{3.1x10^6} = 0.258$. This demonstrates that the overall losses, sensitivity or gain of the receiver changes over the optical range if TDM is involved.

Time division multiplexing of SPADs within small groupings, is an interesting method of obtaining increased linearity [178] or reducing the impact of the SPAD dead time. The implementation of this in a silicon receiver is discussed in Chapter 4, however the implications of this are:

- TDM gives approximate weightings of $\{0, 0.25, 0.5, 1\}$ to a group of N_{TDM} SPADs operating in different N_a modes. The weighting is dependant on the flux giving an element of gain modification.

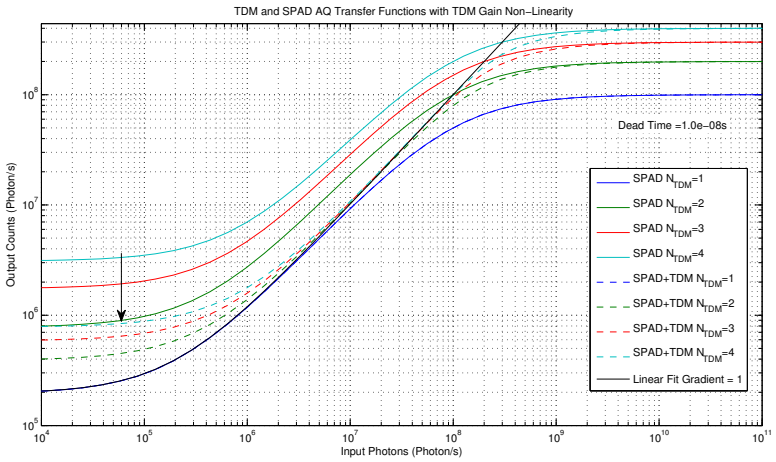


Figure 3.5: A log-log plot of the *SPAD-Only* and *SPAD and TDM* response curves. A trade off exists between the reduction in the noise floor, the increase in linearity and the reduced absolute detection rate. The arrow at the left highlights the reduction in noise floor for a $N_{TDM} = 4$ SPAD pixel, from the SPAD only value of $3.1\text{x}10^6$ detections/s to the SPAD and TDM combined response noise floor of $8\text{x}10^5$ detections/s.

- TDM may increase the linear region when the SPADs are operating near saturation, as they would be if the dead time is specified to obtain the correct number of detections per SPAD per data symbol.
- TDM increases the losses and overall sensitivity at lower photon fluxes, mitigating DCR, after-pulsing and the received amplitude of the N_0 distribution. This is modelled in Section 4.5.9 and Section 4.5.10 and later shown experimentally.
- TDM however creates a sensitivity offset, with the effective group detection efficiency, decreasing with N_{TDM} (assuming the $N_a = 1$ mode)
- Small groupings of SPADs in a TDM scheme, which is suggested by the *ePDE* losses, imply that with multiple diodes within the same receiver, different groupings could be in different TDM modes. This per-pixel mode, could then be used to obtain both low noise and high sensitivity, especially when combined with knowledge of the positions of high-noise devices within the array.

3.5 Extended Specifications for Receiver Design

To complement the high level specifications (Table 2.1), the important receiver and circuit implementation specifications derived in this chapter are presented in Table 3.8 and Table 3.9. The primary considerations are for co-design with CMOS controlled μLED transmitters [23, 24, 65, 66, 93, 98], along with the findings of a literature review looking into contemporary receiver design techniques (Section 2.5) and the factors affecting CMOS SPADs (Section 2.7).

The design specifications (Table 2.1), directly generate estimates for other circuit specifications (Table 3.9). The implementation of a SPAD-based, highly-sensitive receiver, based upon these key values, is therefore discussed in detail in Chapter 4.

Table 3.8: Receiver specifications derived within Chapter 3.

Receiver Specifications:		
Photon Quantum Limit:	20 photons per τ_b	
Dark Counts per Symbol:	≤ 2	
Min Detection Rates:	2.2x10^9 detections/s 11x10^9 detections/s	At 100Mb/s At 500Mb/s
Number of Channels:	1 to 64	Configurable
Integration Methodology:	Digital Pulse Counting	
QL Sensitivity:	0.58 to 2.91nW -53.2 to -62.4dBm	100–500Mb/s (and) 450–750nm
Optical Penetration Depths:	0.39μm 7.69μm	At 450nm At 750nm
Sensitivity:	3.98μW = -24dBm 19.9μW = -17dBm	At 100Mb/s At 500Mb/s
# Photons (450nm):	9x10^{12} 4.5x10^{13}	At 100Mb/s At 500Mb/s
# Photons (750nm):	1.5x10^{13} 7.5x10^{13}	At 100Mb/s At 500Mb/s
Total Power Penalties:	36.16–38.35dB 36.11–38.34dB	At 100Mb/s At 500Mb/s
Theoretical Max Fill Factor:	18.8%	Rectilinear Array
Total Receiver DCR:	$\leq$ 2Mcps	$\approx$ 195kcps per SPAD
Architecture-Independent BW:	$BW_{th} = 5.12$Gb/s	1024 SPADs, $\tau_d = 10$ns
Nyquist Bandwidth:	$F_{Ny} \geq 50$GHz	1024 SPADs, $\tau_d = 10$ns
Estimated Analogue BW:	$BW_{avail} = 4.61$GHz	

Table 3.9: Circuit specifications derived within Chapter 3.

Circuit Specifications:		
Dark Count Rate:	$\leq 0.1QL$	≤ 2 Detections per τ_b
Minimum detections:	22 events	Detections per τ_b
τ_d Passive:	11–170ns 2–34ns	At 100Mb/s At 500Mb/s
τ_d Active:	30–450ns 6–93ns	At 100Mb/s At 500Mb/s
Overall τ_d Range:	5–150	Configurable τ_d Configurable AQ/PQ
Diode Capacitance:	$C_D = 21.5$fF	
Passive Quench Resistance:	$\approx 100k\Omega$	
SPAD Diameter:	$\leq 8.6\mu$m	Assuming Circular
Junction Width:	0.57μm	
Break Down Voltage:	14.35V	
Receiver Clock Jitter:	≤ 250ps ≤ 50ps	At 100Mb/s At 500Mb/s
SPAD Array Architecture:	Clustered	Hierarchical Addition
SPAD Summation Option:	Time Division Multiplex	Mode Configurable $N_a = 3$ mode used for disable
Suggested TDM Grouping:	$N_{TDM} = 4$	SPADs per Pixel

3.6 Conclusions

In this chapter, the specifications for an OOK-only SPAD-based receiver for VLC have been discussed. The receiver aims to decrease the optical energy needed. Therefore the receiver must be efficient, maximally converting photons into usable signals. Models from the literature (Section 3.3) translate combinations of specifications into approximate circuit implementation specifications (Table 3.9). The photon energy, quantum limit and models of SPAD operation, each allow the receiver to be co-designed with an array of 1024 CMOS-controlled μLED OOK transmitters [23, 24]. With bounds for the sensitivity, practical receiver non-idealities can be used to create a power penalty budget (Table 3.6), further generating constraints on specifications.

Numerical modelling of a time division multiplexing scheme, has been performed TDM is used to reduce non-linearities caused by the dead time. Modelling of TDM, points to increased linearity and gain modification. The technique decreases sensitivity however, suggesting limited use and a small number of SPADs within a group. In the next chapter, the design of a prototype receiver is discussed, including on-chip circuitry designed to help assess the concept.

Chapter 4

An Experimental SPAD Receiver Chip

4.1 Introduction to the Receiver Chip

In this chapter the specifications in Section 2.9 and Section 3.5, are used to implement a SPAD-based integrating receiver [68] with high sensitivity, fabricated in a 130nm CMOS process [14, 35, 49, 69, 73]. The 32x32 SPAD array is designed concurrently with a 32x32 μLED OOK transmitter array [23]. Due to the design of the μLED driver circuits, On-Off Key modulation is targeted only. Other modulation schemes require suitable adjustment of the underlying receiver concept, receiver specification design flow and the μLED driver circuitry.

The receiver operates in two ways; the first, takes the summation of 32x32 = 1024 SPADs into a single output, with each pixel incorporating time division multiplexing [178] (Section 3.4). This is limited by the data bandwidth of the I/O pads. The second mode, gives 16 separate channels, each of 64 SPADs (Table 3.3), that could be used for optical MIMO techniques [16, 20, 95, 203]. This achieves a higher aggregate output due to several, multiplexed outputs.

This chapter introduces the hierarchical design of an integrating receiver, discusses circuit implementations, and presents models used to derive further specifications. The first generation receiver is intended to assess the feasibility of the technique and define further work, while satisfying the discussed specifications and exploring limiting factors.

4.2 SPAD and CMOS Process Availability

The CMOS process choice is restricted in this work. As such, the impacts on the specifications must be indicated. SPADs and optical receivers have been demonstrated in numerous processes, however each has unique trade-offs. Combined bipolar and CMOS (Bi-CMOS) [100, 101, 204] and silicon on insulator (SOI) [27, 109] processes have been previously used for receivers, however these were not available despite advantages for high speed operation [14, 26, 27, 42, 45, 46, 69, 109, 205].

4.2.1 ST Microelectronics: IMG 175

Through collaboration with ST Microelectronics, the device was manufactured on the IMG 175, 130nm imaging CMOS process [30]. This has been modified, with increased quantum efficiency [81] (Section 2.7.3) and a reduced metal stack. The metallisation used is from a 90nm process [30]. SPADs have been developed in this process [53, 143] indicating a route towards high-density integration of SPADs with digital circuits [70, 155, 162, 163, 206, 207].

The process choice has presented a number of problems, limiting this device. A critical limitation is that the available digital input/output pads have a limit of 108Mb/s. Low voltage differential signalling (LVDS) [35, 69], which can be used for faster rates of 600Mb/s, were not available. This severely restricts the rates at which data can be sent to an FPGA-based testing PCB (Figure 4.32).

4.3 Architecture Overview

In Figure 4.1, the receiver hierarchy is presented. Starting at the top, the silicon SPAD is shown. The SPAD with quenching and hold off circuitry is held in a *Sub-Pixel*. It has been denoted as such, because $N_{TDM} = 4$ of these are grouped into *Pixels* implementing TDM and initial summation logic.

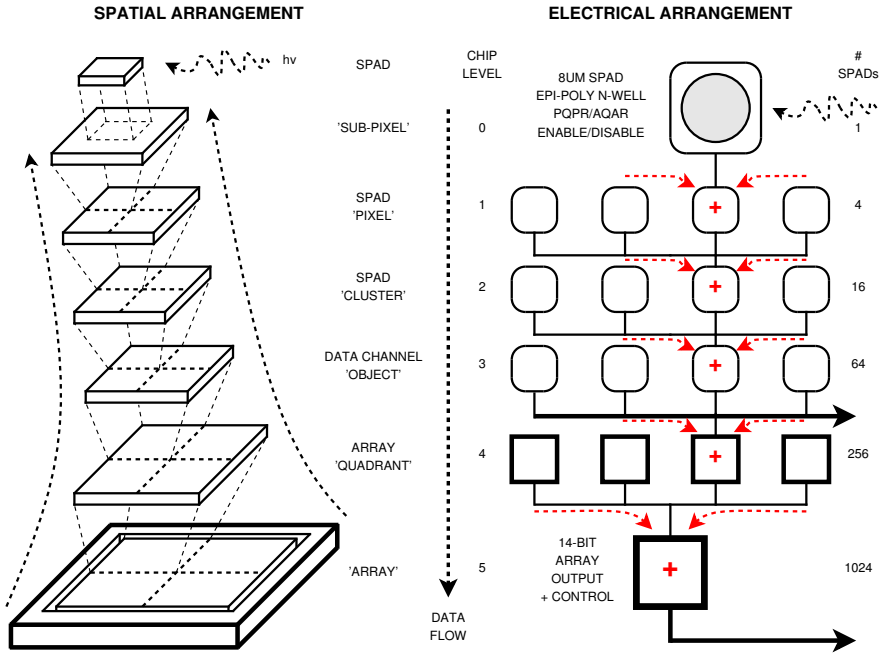


Figure 4.1: The spatial (left hand side) and electrical (right hand side) architecture of the array, showing circuit hierarchy. Each level takes the parallel summation of the lower blocks, giving multiple readout options for the 1024 SPAD, 16 channel receiver array.

A *Cluster* is a group of four Pixels (16 SPADs). It is at this level that the digital signal chain is first encountered. The integration of SPAD pulses within a symbol, assuming OOK operation of one bit per symbol (Section 2.3.3), is implemented here. The Cluster provides serial interfaces [29, 69], setting TDM modes on a *per-pixel* basis along with direct counter readout. The Clusters are grouped into a unit of four (64 SPADs), called *Data Channel Objects*. This represents a single channel in the 16 channel mode (Table 2.1), with each able to be readout as an identical independent receiver front-end.

Moving up the hierarchy, four Objects are grouped together into *Quadrants*. This implements functionality for the final summation and multiplexed channel outputs. Lastly, the *Array* level performs the final summation for the single channel mode. A top level clock tree [29, 69], output multiplexers, control logic and three configuration and readout serial interfaces are also provided.

4.4 The Single-Photon Avalanche Diode and Sub-Pixel

In this section, the low-level implementation of SPAD circuitry is discussed. This includes details of the SPAD in the chosen CMOS process, along with the dual mode quenching and dead time control circuitry. This is important for communications using such a device, as it is at this level where critical performance values will be decided. For example, choice of an incorrect circuit may prevent adequate dead times for the end application, or a silicon design trade-off may reduce the capacity to meet prior design specifications (Table 2.1).

The SPAD used in this design is an 8μm diameter, circular device with the p-n junction, formed from a layer of P^+ doped silicon, a region of P-Well, and a Deep N-Well [143]. The diode and surrounding circuitry is shown as a cross section through the CMOS technology [30] in Figure 4.2. The N-Well provides the SPAD cathode, and contacts the deep N-Well region [173]. The p-n avalanche multiplication region [26] is protected using a retrograde guard ring [143]. Shallow trench isolation (STI) is used near the guard ring to increase electrical and optical isolation between devices, as well as reducing the leakage current [143, 173].

The junction capacitance (C_{ac}) is formed by the depleted space charge regions [26]. This has been estimated [26] independently of Section 3.3.6 to be 18.5fF [173]. The parasitic capacitances were extracted from the device layout [143], giving $\approx$ 14.5fF for the cathode to substrate parasitic (C_{cs}), and $\approx$ 1.49fF for the anode to substrate parasitic capacitance (C_{as}).

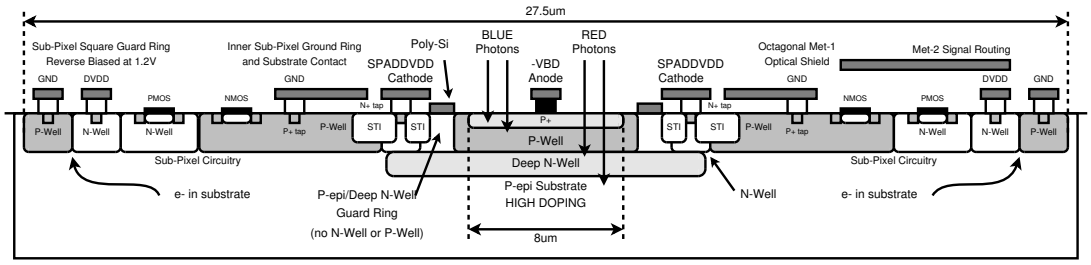


Figure 4.2: A cross section of a Sub-Pixel and SPAD. The multiplication region is positioned at the interface of the P-Well and Deep N-Well regions. The Sub-Pixel circuitry is surrounded by a reverse biased *p-n* junction, shown at the sides of the cross section, which acts as a passive guard ring.

4.4.1 The Optical Guard Ring

The SPAD, in the centre of the 27.5μm cross section (Figure 4.2), is surrounded by the cathode contact [49, 132, 143]. This ring is the moving node of the circuit [143]. In order to retain the optical shield preventing photons entering the substrate around the diode, a secondary wide track is placed outside the cathode. This provides an inner ground ring with substrate contacts, ensuring NMOS devices have a low noise bulk connection [29, 150]. Extensive metal coverage near the SPADs was not possible due to design rules [30], however metallisation over the

circuitry is maximised. This gives the added advantage of wide tracks for power distribution [29, 69, 85].

The metallisation functioning as an optical guard is similar to other SPAD arrays in the chosen CMOS process [56, 57, 70, 153, 154, 163, 208]. Due to the short dead times of 2.1ns and 10.7ns respectively for 500Mb/s and 100Mb/s 16-channel operation (Table 3.3), the parasitic capacitance of the moving node has been reduced by splitting the cathode. This is at the expense of optical shielding but is worthwhile if positive drive [57], low parasitic moving node configurations are unavailable.

4.4.2 The Chosen SPAD Circuit Configuration

The SPAD is used in negative drive configuration using a PMOS quenching transistor [49, 143, 162, 209] (Figure 2.10). While the N-Well [26, 69] becomes the moving node, increasing the capacitance to ground (C_{cs}), positive drive configurations with much lower parasitic capacitances (C_{as}) had not yet been demonstrated in the chosen process [57, 158, 206]. Developmental risk should be minimised, which has lead to a discrete rather than shared N-Well structure [141, 157] and the use of both active and passive quenching methods [49].

The circuit biases the SPAD with the Deep N-well cycling between 0V and *Vex*, and not with the Deep N-well junction to the p-type substrate reversed biased as per a number of other implementations using positive drive quenching [57, 154]. Under certain transient conditions [76] there will be no isolation between the SPAD and the bulk silicon, as the built-in potential is yet to form [26]. The lack of isolation may be problematic [53, 59, 76, 121, 184, 210], however positive drive circuits using this device have now been proven [57, 154].

The Chosen SPAD Geometry

The circular shape was chosen due to demonstrated performance [53, 143, 163]. Circular SPADs achieve lower DCR and a uniform electric field due to the rounded edges [53], however circles decrease array fill factors as they have a low packing density compared to hexagonal, octagonal and rectilinear devices. This continues when the SPAD is surrounded by circuitry, as the native SPAD fill factor remains low due to the guard ring. The fill factor also remains low for circular devices with surrounding circuitry due to the minimum spacings rules and the rectangular nature of digital circuits from foundry libraries [30, 56].

Rectangular photodiodes have been used in CMOS image sensors [81], however the electric field strength is increased significantly when rectilinear shapes are biased into the Geiger region [76, 185]. This increases the DCR and after-pulsing probability [53]. Recently, another class of geometry has been explored [53]. The *Fermat* shape, more aptly named the *Lamé* curve, uses a super-ellipse (square with rounded corners). This can obtain high fill factor without the increased electric field strength of 90° angles [53].

With a Fermat device of 8μm, minimum corner curvature and guard ring dimensions from [173], the SPAD-only fill factor would be 20.6%, much less than the 38.8% of a 16μm Fermat shaped device. Both are significantly higher than circular 8μm and 16μm devices, which have fill factors of 16.4% and 30.9% respectively [173]. The SPAD used here has a native (SPAD-only) fill factor of 20.9% [143].

A Verilog-A SPAD Electrical Model

In order to simulate the SPAD behaviour, an electrical model is employed. Using Verilog-A, a continuous-time simulation can be performed including behaviour due to the transistor, process and chosen circuit topology. A number of models suitable for design have been demonstrated [136, 211], however a reduced Mita et al [211] model is used. This assumes a simple capacitive network shown in Figure 4.3, modelling the junction capacitance (C_{ac}), and the two main parasitic capacitances C_{as} and C_{cs} [26]. The model discharges using a voltage controlled current source, with a forced turn-off based on latching and turn-off currents [49, 123, 136, 138, 211].

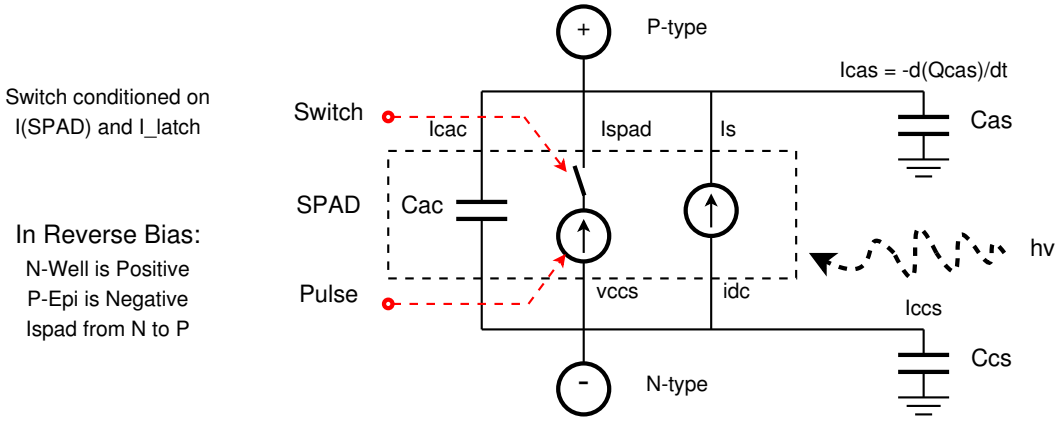


Figure 4.3: Electrical schematic of the Verilog-A model. The voltage controlled current source dynamically discharges the charge held on the junction capacitance, C_{ac} and the two parasitic capacitances, C_{as} and C_{cs}.

The model assumes a fixed quenching resistance [49, 75, 136, 137, 167, 212], however for active circuits, the quenching network may change throughout the cycle [161, 162, 213]. This leads to a departure from the assumptions of the model [211]. The model uses previously defined values for the parasitic and junction capacitances, along with measured values for VBD, the forward voltage and leakage currents (I_s) [143].

4.4.3 The Sub-Pixel

The Sub-Pixel circuitry (Figure 4.4) surrounds the SPAD and is designed to allow the following functionality.

- External SPAD bias control with separate SPAD and digital supplies
- Selectable passive or active quenching, making testing easy
- SPAD enable and disable to force the SPAD into its unbiased state
- Active and passive dead times externally controlled using DTCTRL
- A passive guard ring to minimise cross talk [35]
- Local decoupling on SPAD and digital supplies
- Circuit suitable for free running or fast gated operation
- Compact layout respecting process design rules
- Spurious pulse masking to prevent unnecessary counts
- A gridded power strategy for robust supplies with minimal array droop

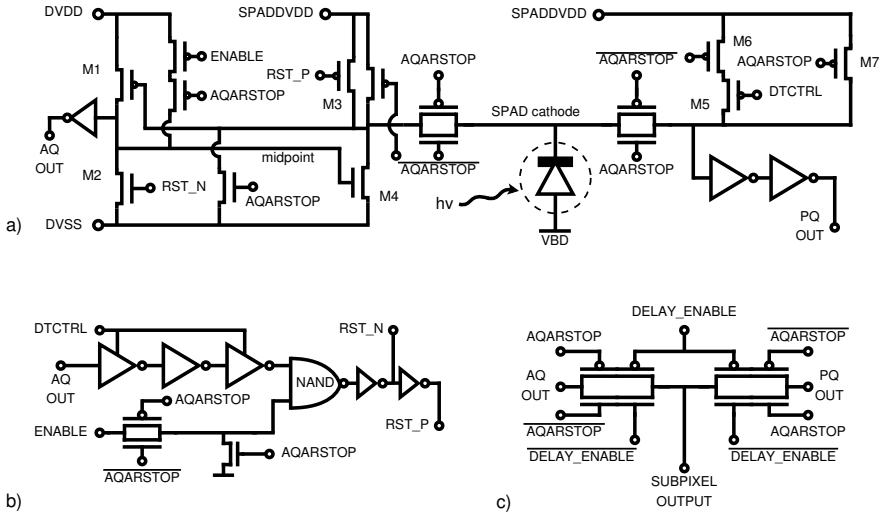


Figure 4.4: Electrical schematic of a Sub-Pixel. a) The AQAR circuit based on the Richardson et al dynamic quenching circuit, b) A current starved inverter chain controlling the AQAR dead time and c) The transmission gate based output multiplexer.

In Figure 4.4.a the selectable active quench active reset (AQAR) or passive quench passive reset (PQPR) circuit, SPAD, quenching selection transmission gates and the enable logic are shown. Figure 4.4.b shows the current starved inverter chain used for the AQAR delay line and Figure 4.4.c details the output multiplexer using transmission gates and delayed enable signals to prevent spurious pulses reaching subsequent logic [29, 69].

The Passive Quench Passive Reset (PQPR) Circuit

Passive quenching is provided through transistors M5 and M6 (Figure 4.4.a), both providing a high series quenching resistance [49, 137]. For M5, $\frac{W}{L} = 0.075$ and M6, $\frac{W}{L} = 0.3$. With $AQARSTOP = 1$, M6 is fully on, allowing M5 to dictate τ_d via the gate bias voltage DTCTRL. When $AQARSTOP = 0$, the minimum dimension PMOS transistor M7, forces the right hand side (high impedance to the SPAD cathode) to logic high. The quenching circuit is simulated in Figure 4.5. This uses the Verilog-A model, Sub-Pixel parasitics and a 500 run Monte Carlo analysis [69] with foundry specified definitions for the process corner variation and device mismatch [30].

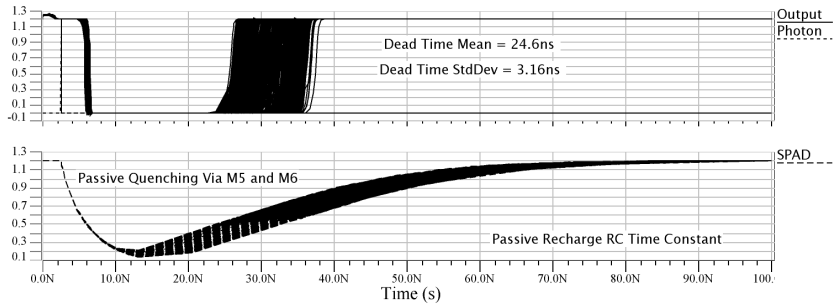


Figure 4.5: Passive quench passive reset circuit simulation with SPADDVDD=1.2V. The lower figure shows the waveform of the SPAD moving node with the slow RC recharge indicative of passive quenching. The top figure shows the variation in the dead time at the output of an inverter, that result from variations in the RC recharging time constant.

The Active Quench Active Reset (AQAR) Circuit

The active quench active reset circuit [49] (Figure 4.4.a) is a re-implementation of a dynamic circuit [162, 214]. This is simulated in Figure 4.6. Assuming the SPAD has not fired, the cathode is high at $Vex = SPADDVDD$, held by leakage currents through transistor M3, $\frac{W}{L} = 39.1$. The midpoint node is held low by leakage through transistor M2, $\frac{W}{L} = 6.9$. These nodes are initially high impedance giving a low power quiescent condition [162, 173].

Upon an event, the SPAD will start to passively quench through M3 [162]. This is shown as the constant slow discharge seen at $\approx$ 3ns in Figure 4.6. This continues until it passes the threshold voltage, V_t, of M1, $\frac{W}{L} = 3$, which starts to pull the midpoint node high. This prompts transistor, M4, $\frac{W}{L} = 1.5$, to turn on, quickly discharging the SPAD [162, 173].

The circuit has been proven [163], however if SPADDVDD is lower than DVDD, the leakage paths become unbalanced creating a discharge/recharge event [162]. Process, mismatch and temperature variations affect leakage paths [29, 162], possibly giving lower array uniformity in comparison to previous arrayable circuits [70, 155, 213]. The circuit was chosen as its prior

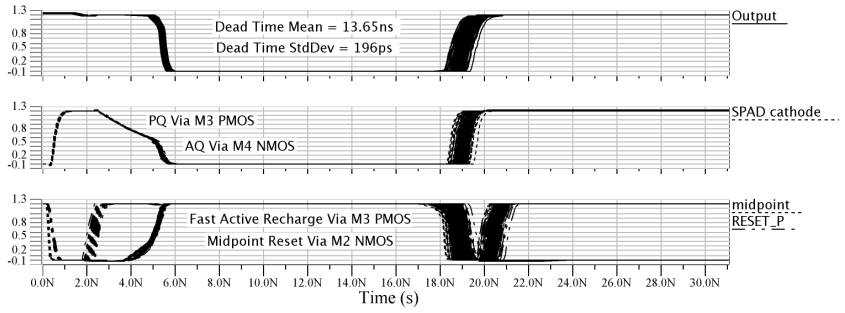


Figure 4.6: Active quench active reset circuit simulation with parasitics and the Verilog-A SPAD model. At $\approx$ 3ns, the SPAD begins to passively quench, until $\approx$ 5ns where the active quenching NMOS is activated.

performance and availability in the chosen process presented a lower developmental risk than other untested circuits.

Simulation with capacitances and Verilog-A model, showed tunability (Figure 4.4.b) of the dead time from 11.53ns at $DTCTRL = 0$mV to 76.7ns at 700mV and an estimated standard deviation of 190ps at $\tau_d = 12.4$ns. This compares well with the required specifications (Section 3.3.2), but may need voltages greater than 700mV to provide longer dead times.

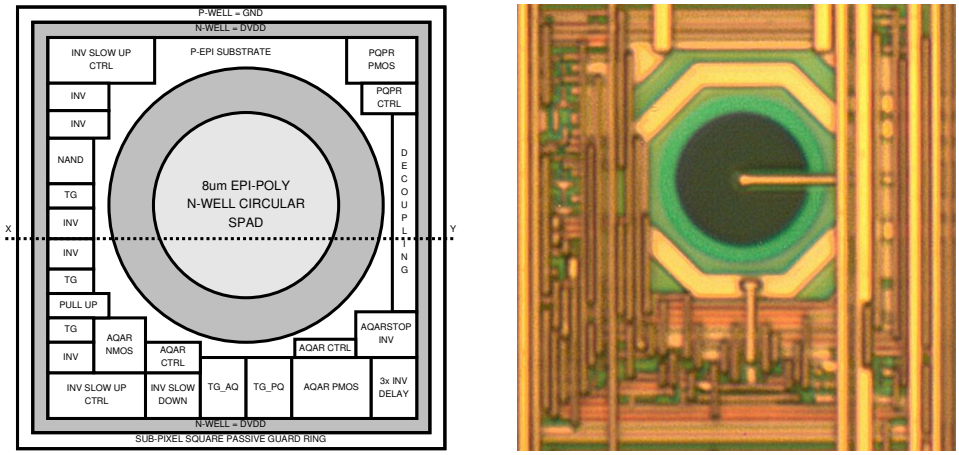
From typical case simulation, the circuit will consume $\approx$ 191fA (average) per pulse, and at its theoretical maximum frequency of 100Mcps will consume 22.9μW from 1.2V. For the 1024 diodes in the array, this gives a worst case SPAD-only power consumption of 23.4mW (all 1024 SPADs firing at maximum rate, with typical $\tau_d = 10$ns specification). This is approximately equal to full array current consumption at saturation, as measured in Section 5.2.1.

Circuit Enable/Disable and SPAD Gating

A number of other transistors are included, such as the *Enable* transistors in Figure 4.4.a and Figure 4.4.b. These allow enabling and disabling in the fastest possible manner [162]. A low leakage PMOS path is created between DVDD and the AQAR midpoint node. Rather than preventing pulses to propagate, if ENABLE is low this quickly forces midpoint high holding the SPAD at the $DVSS = 0V$ supply, reducing DCR and $P(ap)$ [162]. Gating in this manner can cause spurious pulses, these are screened from the output by transmission gates controlled by the enable signal (Figure 4.4.c). However subsequent photon pulses are able to propagate correctly.

The Sub-Pixel Layout

Figure 4.7, shows the block diagram and micrograph of the Sub-Pixel. The square Sub-Pixel guard ring seen in the X-Y cross section (Figure 4.2) surrounds the circuit, providing electrical isolation and a grid-like power supply structure [29, 35, 69, 85]. There is a trade off between compactness, process design rules and functionality. Circuit area is increased by the inclusion of both active and passive circuits, however a) this is used to increase testability and b) the SPAD pitch is dominated by N-Well spacing rules [30].



(a) Block diagram of a single Sub-Pixel.

(b) A micrograph of a single Sub-Pixel.

Figure 4.7: a) Block diagram, and b) micrograph of a single Sub-Pixel. The layout highlights that all area, required by the N-well spacing rules is utilised for functionality within the block.

A distributed logic style around the SPAD has been used. The fill factor is 6.65%, limited by the SPAD native fill factor of $\approx 21\%$ and the spacing rules [30, 143]. Fill factors can be increased by using square or *Lamé* shaped SPADs and smaller guard ring dimensions [53, 173]. These techniques were not implemented however, due to developmental risks. In particular moving away from circular devices increases DCR [53], while reduction of the guard ring width significantly increases the risk of premature edge breakdown [76, 185].

4.4.4 Conclusions from SPAD and Sub-Pixel Implementation

The above section has over-viewed the development of the SPAD and Sub-Pixel levels of a hierarchical single-photon sensitive receiver. For communications, it is development at this level that may dictate a number of higher level receiver performance values. As an example, the guard ring of a SPAD, separations from active MOS devices and the SPAD dead time dictated by the quenching circuitry, can each induce dark and after-pulsing counts [53]. Likewise implementation at this level directly impacts the fill factor and therefore the power penalties of the receiver (Figure 3.2). In the next section, the grouping of SPADs into Pixels each incorporating time division multiplexing functionality (Section 3.4) is presented.

4.5 The Pixel and TDM Circuits

In this section the Pixel and TDM circuits, shown in Figure 4.8, are discussed. This takes the addition of a group of $N_{TDM} = 4$ SPADs, but does so in a set of three different time division multiplexing methods (Section 3.4.1). This is important for device development as it a) allows increased testing of the fabricated device, aiding in experimental work, and b) from previous modelling (Section 3.4.1) may provide a number of advantages for noise reduction when the receiver is used for communications. The Pixel uses a four mode finite state machine (FSM) [69] to implement in-Pixel feedback. This is used in combination with variable next-state enable logic [69, 215] to implement modes. A clock multiplexer selects either a self clocked asynchronous feedback or a synchronous open loop mechanism using an external clock.

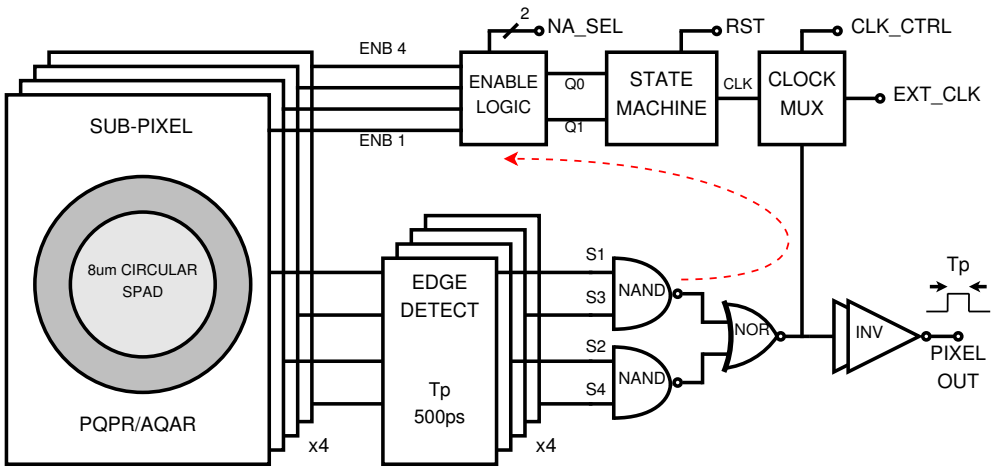


Figure 4.8: Electrical block diagram and logic schematic for the Pixel, using four SPADs, four edge detection circuits and a NAND/NOR signal summation tree. The enable logic is externally controlled via the 2-bit NA_SEL signal. This is configured at the proceeding *Cluster* level of the hierarchy.

4.5.1 Enable Control Logic

To allow reconfigurability, the N_a selection logic generates the truth table shown in Table 4.1, only allowing the chosen generated signals through the enable line multiplexers (Figure 4.9). The majority of the complexity, occurs in the $N_a = 1$ mode. The logic for the $N_a = 2$ mode does not require additional circuitry as it directly creates the enable values using the FSM. The external clock of the FSM can be used to gate all SPADs, while keeping the TDM and non-TDM functionality.

In Figure 4.9, each N_a mode only allows one of the signals to progress to the enable input of the Sub-Pixel. As the number of SPADs within the Pixel increases, the logical overhead required will reduce the fill factor. The design was kept in this implementation, in order to experiment with the various N_a modes. Testing of these and the overall reconfigurability, may reveal insights into mode affects on the noise floor and transient response.

Table 4.1: Truth table for Na mode selection logic

NaSEL0	NaSEL1	Na1	Na2	Na4
0	0	1	0	0
0	1	0	1	0
1	0	0	0	0
1	1	0	0	1

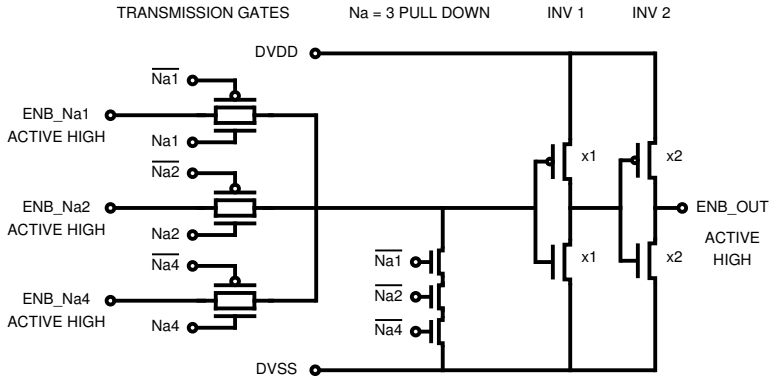


Figure 4.9: Schematic for a non-inverting transmission-gate-based SPAD enable multiplexer for each Sub-Pixel in the Pixel.

4.5.2 The Edge Detection Circuit

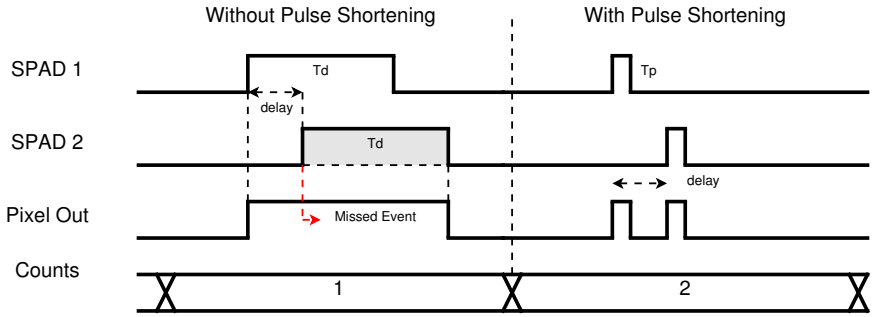


Figure 4.10: Signal propagation with and without edge detection pulse shortening. Two full width pulses merge to a single longer pulse at the pixel output. With pulse shortening, both SPAD pulses are captured, reducing the count losses at light levels giving photon inter-arrival times comparable to the dead time, τ_d.

The Pixel implements an edge detection operation. The dead time pulse generated by the SPAD is shortened, preventing saturation of addition or counting logic. As shown in Figure 4.10, if another SPAD fires without this circuit, pulse overlap may occur in a probabilistic and light level dependant manner. This gives a single count rather than two. The circuit (Figure 4.11) uses a delay path implemented with resistive inverter circuits. A NAND gate takes the addition of the incident and delayed edges, creating an output equal to the designed delay. There is an advantage to shorter τ_p duration pulses (Section 4.5.8), which is complemented by electrical

and area advantages. The width of τ_p, as long as it satisfies $\tau_p \leq \frac{\tau_d}{4}$, is not critical provided that a) it triggers subsequent logic under all process, mismatch and power variations and b) at high flux, τ_d becomes the limiting factor.

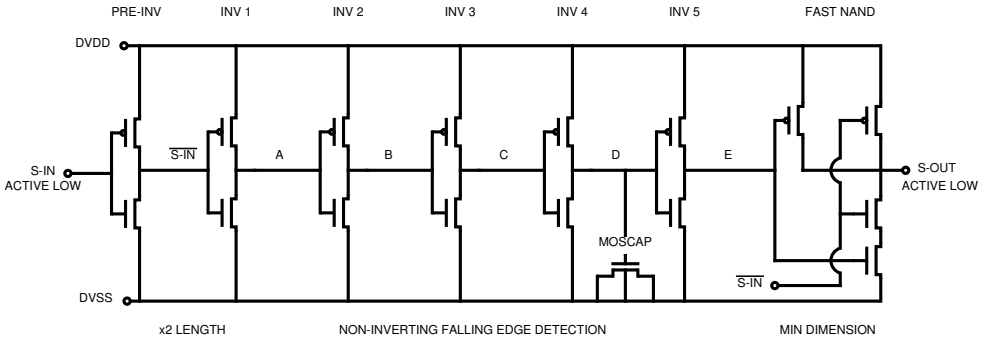


Figure 4.11: Electrical schematic of a non-inverting edge detection circuit with a τ_p pulse width. Inverters A to E implement a slow delay path of appropriate length, with a fast NAND gate providing the pulse to subsequent logic.

4.5.3 SPAD Pulse NAND/NOR Addition

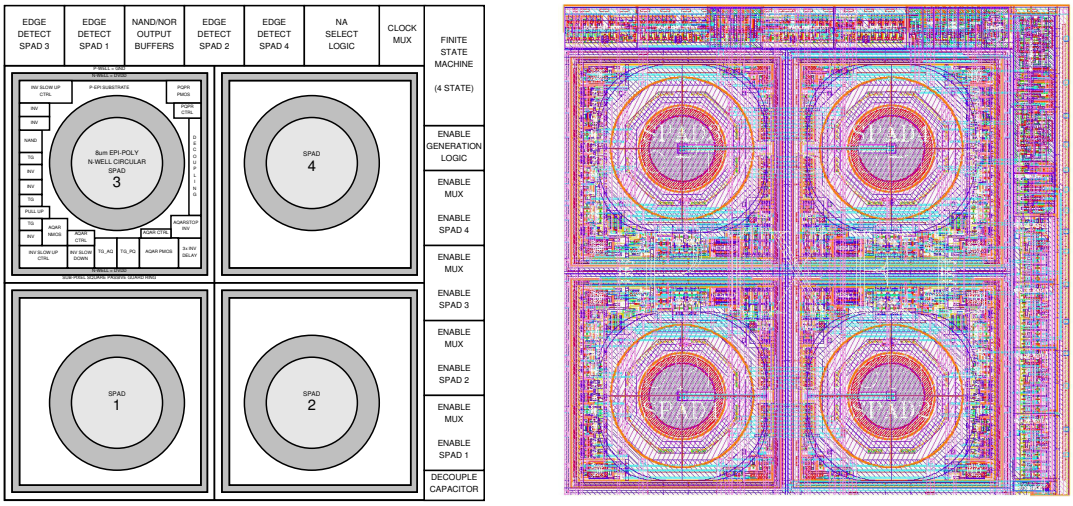
In order to multiplex SPADs into a single output, a NAND/NOR tree is used after edge detection. Other arrays implementing summation have used XOR gates [188], however in this application the NAND/NOR tree allows a smaller distributed structure. The scheme retains the pulse shapes generated by edge detection, rather than requiring a dual edge triggered counter at the Cluster level (Section 4.6).

4.5.4 Synchronous and Asynchronous Operation

A clock multiplexer enables synchronous or asynchronous modes. This is useful as, a) it increases the scope for experimental work, allowing investigation of TDM operation on the DCR and after-pulsing, b) non-TDM gating of SPADs can also reduce these contributions [49, 62] and c) manual iteration of the state machine can be used for array characterisation. For low ISI, low BER communications, the integration period and the time at which the receiver is optically active, could be used to screen transmitter transitions or optimally place the integration period within the symbol [35]. The Stevens et al [73] and Zerbe et al [113] integrating receivers, position a wide integration period at a point where ISI is minimal. This is achieved using reconfigurable phase shifts between the input data and the sampling system, allowing the received bit distributions to be independent of the transmitter and receiver step and impulse responses, thus improving the BER [35, 73, 113].

4.5.5 Pixel Layout and Physical Considerations

The Pixel layout is presented in Figure 4.12. The Sub-Pixels are rotated to create a symmetric central region for power and signal routing. The logic is placed at the top and right edges, creating a symmetric grid at the Cluster level. The Pixel measures 60μm with a maximum $N_a = 4$ fill factor of 5.4%. The splitting of the optically active area into multiple parts has allowed lower parasitic capacitances, dark count rates and jitter, in comparison to a single SPAD device of 45 to 50μm diameter [53].



(a) Block diagram of a Pixel.

(b) Full Pixel Layout (All Layers).

Figure 4.12: Physical block diagram and layout of a 60μm Pixel. Each Sub-Pixel is rotated to ensure a symmetric structure with similar routing strategies for the Sub-Pixel output to the edge detection circuits implemented at the top of the Pixel.

4.5.6 Na = 4 Spatially Multiplexed Summation Mode

In the $N_a = 4$ mode, the Pixel enables all four of the Sub-Pixels at the same time. Any event in any Sub-Pixel is able to progress to the output. This gives the simple sum of SPAD pulses. In this mode, the problem of pulse temporal coincidence impacts the count rate when the incident photon rate approaches $\frac{4}{\tau_d}$.

With the pulse inter-arrival time tending towards zero, the addition of pulses results in a single output pulse, hence the NAND/NOR summation no longer holds. The time dependency, given the stochastic processes involved, should therefore be noted. In this implementation, the dead time and the required maximum count rate necessitate the use of the edge detection circuits (Figure 4.10).

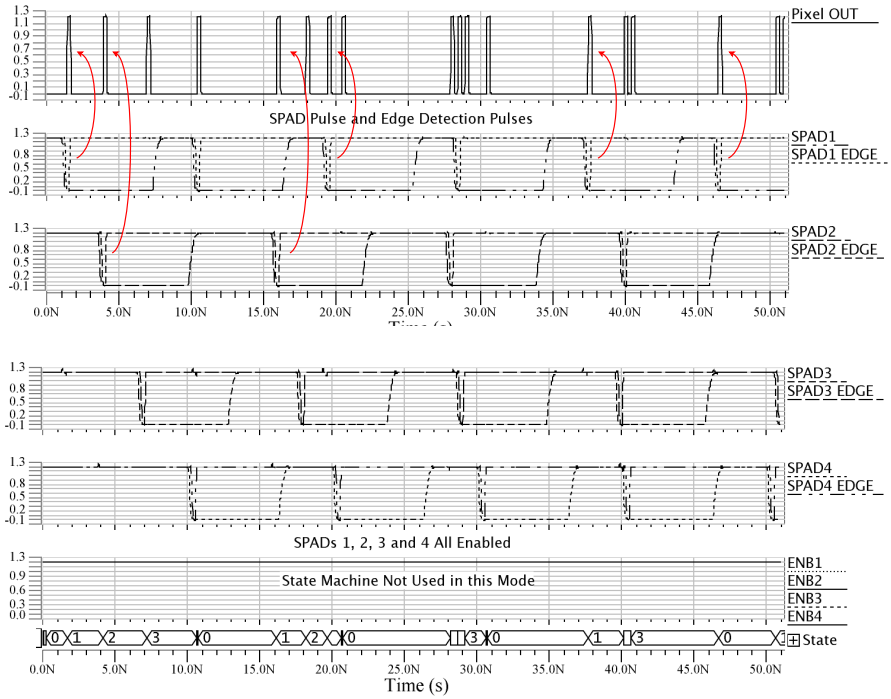


Figure 4.13: Enable line iteration in the $N_a = 4$ TDM mode. All enable lines are held high by the $N_a = 4$ enable generation logic, with any Sub-Pixel pulse able to progress to the output. The mean time between photon events is ≈ 2.75ns.

Figure 4.13 shows that each of the Sub-Pixels is continuously enabled when the Pixel is in this mode. At the top, the output pulses are shown given a random input stream onto the four SPADs. Looking at the SPAD output pulses, *SPAD*1, *SPAD*2 etc (Figure 4.13), it is clear that the full width SPAD pulses would merge into a single pulse, given the NAND/NOR addition method. For example between 5ns and 10ns, all of the SPADs register logic-low, forcing the output to be saturated.

The simulation shows that all shortened Sub-Pixel pulses of width τ_p, are able to progress to the output, however at 39.5ns (Figure 4.13) the temporal coincidence problem is demonstrated. The *SPAD*2*EDGE* and *SPAD*3*EDGE* signals overlap, leading to a single pulse at the output. There are two cases within the simulation, within the first 15ns the light level is low such that output pulses are spaced well apart, allowing counters to correctly iterate. Within the period between 25ns and 35ns however, the simulation shows a pulse separation of just 240ps. This necessitates a dedicated counter design, when the SPADs are operated at event inter-arrival times comparable or shorter than the dead time.

4.5.7 Modelling The Spatial Coincidence Probability

With a grouping of $N_{TDM} = 4$ SPADs, there is a probability that all photons strike the same SPAD. This reduces the counting rate as photons occurring within the dead time are unable to cause a detectable avalanche. This is denoted as the spatial coincidence probability, $P(SC)$. If the number of photons hitting a pixel at time, $t = n$, is denoted as $\psi(n)$, then the probability of spatial coincidence, in each TDM mode, is given by Equation (4.1).

$$P(SC) = \left(\frac{N_a}{N_{TDM}}\right)^{\psi(n)} \tag{4.1}$$

This is derived though the successive logical *AND* operation. For example for four SPADs, A, B, C and D, $P(SC)$ is the probability that an initial photon hits $SPAD_A$, a second photon hits $SPAD_A$, and so on. To minimise this, the number of SPADs within a group should be large in comparison to the number of photons within a dead time, while N_a should be high.

At photon inter-arrival periods longer than the dead time, there is only a small probability of two or more photons within the SPAD inactive period. The bounds for the dead time (Section 3.3.2), are based on a minimum number of detections per symbol. As such, the incident photon inter-arrival time is comparable to the dead time, increasing spatial coincidence. As mentioned in Section 3.4, a shorter dead time than needed could be used, however increased data rates, SPAD capacitances or after-pulsing may prevent short dead times.

4.5.8 Modelling Temporal Coincidence Probability

To register detections within the dead time, pulses from multiple diodes must not prevent the counting of other pulses [168, 188]. This is called the temporal coincidence probability, $P(TC)$. Pulse shortening to widths of τ_p allows the probability of ≥ 2 simultaneous events, giving a single count, to be given by Equation (4.2). This assumes an expected value, $E[]$, for the photon flux incident on a pixel, of E[Ψ_{px}], in units of photons/s.

$$E[P(TC)] = \left(1 - \frac{1}{N_a}\right)\left(1 - \mathrm{e}^{-E[\Psi_{px}]E[\tau_p]}\right) \tag{4.2}$$

Equation (4.2) is based on two circuit properties. Firstly that even with all pulses coincident, a single count is produced. This is modelled by $(1 - \frac{1}{N_a})$. Secondly it uses the cumulative photon inter-arrival distribution, $CDF(exponential) = 1 - \mathrm{e}^{-\lambda_r t}$ [116]. This models the probability that pulses of width τ_p overlap as the flux increases.

In Figure 4.14, the counting rate of a pixel has been swept. The probability of pulse overlap increases if no pulse shortening is used. However reducing the pulse width to $\tau_p = 250$ps, allows approximately two decades of improvement in the incident photon flux for the same

temporal coincidence probability. From Figure 4.14, to maintain low losses while using dead times derived from required detections per symbol (Section 3.3.1), some form of pulse shortening is needed. This gives indications as to pulse shortening requirements in future designs.

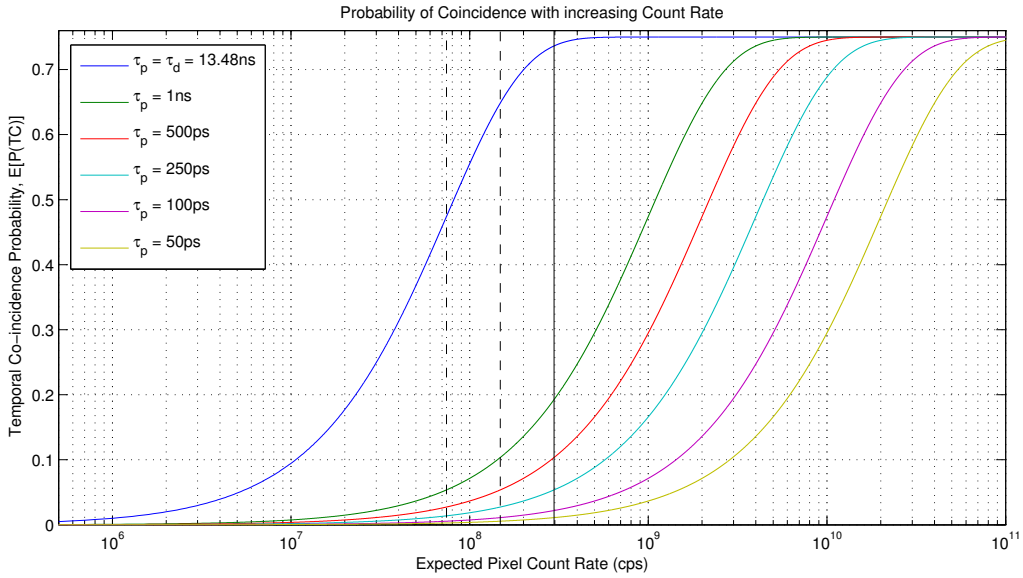


Figure 4.14: Probability of co-incidence with photon flux and pulse width, τ_p. As the flux increases, the probability increases quickly to a maximum where only a single detection can be made. Reducing the pulse width allows a higher flux prior to this saturation. The solid black vertical line at $\frac{4}{\tau_d} = 4\text{x}10^8$ represents the maximum count rate of the pixel.

4.5.9 Modelling Pixel DCR with Simple 'OR' or TDM Circuits

The TDM concept will have an affect on the DCR of a pixel, DCR_{px}, along with the effective photon detection efficiency, $ePDE$ (Section 3.4.2) [178]. Here, the implications of the TDM concept on two different Pixel constructions are explored. Case one explores DCR_{px} when SPADs (with shortened pulses) are multiplexed by an OR-gate. A second case, investigates DCR_{px} for a state machine controlled TDM pixel. Each SPAD is assumed to have a different DCR. DCR_{px} is the native DCR of the group, in perfect darkness under zero after-pulsing conditions.

Case 1: Simple 'OR' SPAD Summation

For the OR-gate case, the output will be the simple addition of the signals of N_a currently active SPADs [69]. The dark count rate of the i^{th} SPAD in the group is denoted by DCR_i. For a group of $N_{TDM} = 4$ SPADs, the combined DCR is given by Equation (4.3).

$$DCR_{px} = \left(\sum_{i=1}^{N_a} (DCR_i) \right) [1 - P(TC)] \tag{4.3}$$

Assuming three SPADs with $DCR = 100$Hz and a SPAD with $DCR = 1000$Hz, the DCR_{px} would be 1300Hz. In this case, the output DCR will be dominated by a high DCR SPAD. With the skew of SPAD DCR distributions [53, 56, 154, 181, 216, 217], the simple 'OR' therefore becomes ineffective.

Case 2: State Machine Controlled (TDM) Summation

In comparison, TDM summation cycles round the SPADs, enabling and disabling each sequentially (Section 3.4.1). Taking $N_a = 1$, a single SPAD is enabled with all others un-biased. While there will be thermal carriers within all SPADs, only the currently active SPAD is able to create an avalanche. If the mode is in its asynchronous state, the SPAD grouping waits for a dark count before the state machine can iterate. The counts per FSM loop can be defined as C_L (Equation (4.4)). This depends on the number of SPADs within the group, N_{TDM}. Here it includes the *gamma* function, $\Gamma = (N-1)!$ [115, 116]. This allows any N_{TDM} SPADs, with any N_a currently active SPADs, under the condition that $N_a \leq N_{TDM}$.

$$C_L = \frac{\Gamma\left(\frac{N}{N_a}+1\right)}{\Gamma\left(\frac{N}{N_a}\right)} \tag{4.4}$$

The combined dark count rate, DCR_{px}, for all N_a cases, is given by Equation (4.5). This has been verified experimentally in Section 5.2.2.

$$DCR_{px} = \frac{C_L[1-P(TC)]}{\sum_{i=1}^{C_L}\left(\frac{1}{N_a}\sum_{j=1}^{N_a}\frac{P(E_{DC}|DCR_j)}{DCR_j}\right)} \tag{4.5}$$

The relative probability of obtaining a dark count event, E_{DC}, given the DCR of a SPAD in a grouping of N_a active SPADs, is given by Equation (4.6).

$$P(E_{DC}|DCR_j) = \frac{DCR_j}{\sum_{k=1}^{N_a} DCR_k} \tag{4.6}$$

The probability $P(E_{DC}|DCR_j)$ is included in Equation (4.5), for the $N_a = 2$ and $N_a = 4$ modes. In $N_a = 1$, the probability $P(E_{DC}|DCR_j)$ falls to unity as there is only a single SPAD turned on and this must fire for the system to iterate. Equation (4.5) uses a weighted average of DCR inter-arrival times. Examining $N_a = 1$, the FSM waits for an average period of $\frac{1}{DCR_j}$ before $SPAD_j$ gives a dark count that iterates to the $SPAD_{j+1}$ state. The time it takes the state machine to complete a single loop, is given by the sum to C_L of these dark count inter-arrival periods [115, 116, 199].

The DCR (cps) can be obtained by dividing the counts per loop, by the average time it takes to iterate round that group. To model N_a modes other than 1 and $N_a = N_{TDM}$, the relative probability of an event, given a SPAD's DCR is needed. This can be seen by examining the $N_a = 2$ case, where either SPAD of two, is able to fire. Clearly, within the smaller N_a grouping, a SPAD with a DCR of 1000Hz will trigger the state machine more often than a SPAD with a DCR of 100Hz.

4.5.10 TDM Pixel SPAD Grouping After-Pulsing

Trapped carriers, which cause after-pulsing, exhibit an exponential decay [26, 49, 76, 127, 174, 176, 218]. In free running SPADs, carrier release at any time after the dead time, can cause an avalanche. If a SPAD fires when a pixel is in a TDM mode, an after-pulse cannot be created until the pixel rearms that SPAD. This gates the time in which a SPAD is enabled as the state machine must iterate through all C_L states before SPAD recharge. The splitting of enable times has a direct affect on the after-pulsing probability [62] of the pixel grouping, denoted as $P(ap)_{px}$.

To derive an approximation (Equation (4.8)), the $N_a = 4$ mode, is assumed first. In this case $P(ap)_{px}$ would be the summation of the after-pulsing rates of all constituent SPADs. Taking the simple 'OR'ed Pixel, any SPAD can create an after-pulse. The combined probability is seen in Equation (4.7), as the summation of separate SPAD after-pulsing probabilities. This is complemented by a product series term, (denoted using $\prod$), due to non-mutual exclusivity [116].

$$\begin{aligned} P(ap)_{px} &= P(AP_1 \text{ or } AP_2 \text{ or } \cdots \text{ or } AP_{N_{SPAD}}) \\ &= [P(AP_1) + P(AP_2) + \cdots + P(AP_{N_{SPAD}})] - P(AP_1 \text{ and } AP_2 \ \cdots \text{ and } AP_{N_{SPAD}}) \\ &= \sum_{i=1}^{N_{SPAD}} [P(ap)_i] - \prod_{i=1}^{N_{SPAD}} [P(ap)_i] \end{aligned} \tag{4.7}$$

While the combination of signals is not mutually exclusive in $N_a = 4$, the product series is negligible. Thus the product term is neglected. To model the time between an event in a particular SPAD and the time at which that SPAD is next able to avalanche, a set of weightings are used, based upon the relative DCRs of the SPADs (Equation (4.6)). For example, if there are two SPADs, with DCRs of 200, and 800, the summation is weighted to give $P(ap)_1\left(\frac{200}{1000}\right)$ and $P(ap)_2\left(\frac{800}{1000}\right)$. Weighting of the average by Equation (4.6) is therefore performed, similar to the model for the TDM group DCR, as a high DCR SPAD has an increased frequency at which it populates charge trapping centres. When viewed from the controlling state machine, a higher DCR SPAD also has a shorter loop time, in that a high DCR SPAD iterates the state to the next SPAD in a shorter time than a low DCR SPAD ($\approx \frac{1}{DCR_j}$). The relative DCR weightings are first performed between SPADs within an N_a group and then by the N_a groups within the Pixel [178].

Finally $P(ap)_{px}$ (Equation (4.8)), must be corrected for the probability of temporal coincidence (Equation (4.2)) at two levels. The quantity $P(TC)_{px}$, measures this probability at the final Pixel 'NOR' and is dominant in $N_a = 4$. $P(TC)_{na}$, measures the same probability but at the combination of SPADs within a N_a group (implemented by NAND gates, Figure 4.8). This has been verified experimentally in Section 5.2.4.

$$P(ap)_{px} = \frac{1}{C_L(1-P(TC)_{px})} \sum_{i=1}^{C_L} \left[\sum_{j=1}^{N_a} \left(\frac{(DCR_j)^2 \, P(ap)_j}{(1-P(TC)_{na}) \left(\sum_{k=1}^{C_L} (DCR_k) \right) \left(\sum_{n=1}^{N_{SPAD}} (DCR_n) \right)} \right) \right] \tag{4.8}$$

TDM can be used to reduce the after-pulsing of a group of SPADs. It does this at the expense of fill factor, creating a trade off. If used on a per-pixel basis, targeting high DCR or high after-pulsing SPADs [53, 55, 59, 76, 144] (Section 2.7.3), it could be used to offset noise contributions without excessive reduction in the receiver sensitivity. An optimum may be found that maximises the SNR, highlighting that a DCR map must also be known [71].

4.5.11 Conclusions from Pixel and TDM Circuit Implementation

In the above section, the Pixel level of the receiver hierarchy has been presented highlighting a number of circuits and issues for implementation. In particular, edge detection circuitry is required to ensure low counting losses. Modelling this allows assessment prior to design in future generations. Likewise models for dark-count and after-pulsing rates of the Pixel grouping have been proposed. These indicated advantages to multiplexing, although it is clear from previous models (Section 3.4.3) and circuit simulation, that each will also impact the absolute sensitivity of a Pixel within the receiver.

The functionality included here, although being useful for characterisation, also provides methods of adapting the SPAD array gain for communication applications. The Pixel also provides the ability to remove high noise SPADs from the signal addition chain, and, subsequent to experimental work, may allow tailoring of the SPAD array post manufacture. In the next section, the Cluster level is introduced. This being the level that implements timing and control, aiming to do so synchronously with the incoming OOK modulated data.

4.6 The Cluster with Integration and Summation

Four Pixels are grouped into a Cluster (Figure 4.15). This provides the first level of counters, latches and adders used in the hierarchy. The Cluster integrates the detections over the majority of the symbol (Equation (3.4)), directly providing functionality necessary for the communication level integrating receiver architecture. However, to read out the data, a short blanking period at the end of the symbol is required. The Cluster implements the summation of four Pixel counters, using a parallel pipelined adder tree [69], presenting a single 8-bit word upon the positive edge of each clock cycle. This represents the number of photo-detections received within the symbol. As the Cluster operates from the system level clock, the timing of the integration and blanking periods are synchronous to this system clock. When used within a communications system, clock and data recovery circuits [2, 35, 219] and on-chip PLLs [83, 205, 219], can be used to synchronise the transmitter and receiver remotely. As this work focuses upon the front-end transmitter and receiver circuits, the PLLs, clocking resources and fabric logic of a testing field programmable gate array (FPGA) could be used to perform synchronisation. Clock and data recovery along with easier prototyping of thresholding schemes in firmware rather than silicon can then be investigated once the operation of the silicon SPAD array front-end circuits have been assessed. Interestingly, once the system is synchronised with CDR techniques, investigation of a low-ISI system could be performed if the blanking period is positioned over transitions of the optical source, in a similar way to some other integrating receivers [73, 113, 114].

The Cluster output is complemented by serial data transfer, formed by a daisy chain of latch-shift registers. This allows readout of the counters before summation. As with previous devices, if a SPAD has a particularly high DCR, it can be turned off [56, 57, 148, 155, 157, 208] using the $N_a = 3$ mode or given a different TDM mode [178]. This functionality is set using a local 8-bit memory (labelled as *8bit N_a Select Serial Interface* in Figure 4.15), which sets TDM N_a values on a per-Pixel basis.

4.6.1 Counter Roll-Over Probability and Dynamic Range

With the data rate dictating the integration period (Equation (3.4)), there is a finite probability a counter will roll-over [69, 215]. This is dependent on the photon flux and the bit depth of the counter. Equation (4.9) estimates the probability that the number of photons in the period (τ_b) is greater than $2^{bitdepth} - 1$.

$$P(RollOver) = 1 - \mathrm{e}^{-\eta_e \xi \tau_b E[\Psi_{px}]} \left(\sum_{i=0}^{2^{BitDepth}-1} \left[\frac{(\eta_e \xi \tau_b E[\Psi_{px}])^i}{i!} \right] \right) \tag{4.9}$$

Equation (4.9) uses the complementary cumulative probability function (cdf) of the Poisson distribution [116, 199]. This measures the probability that the number of arrivals in a specific

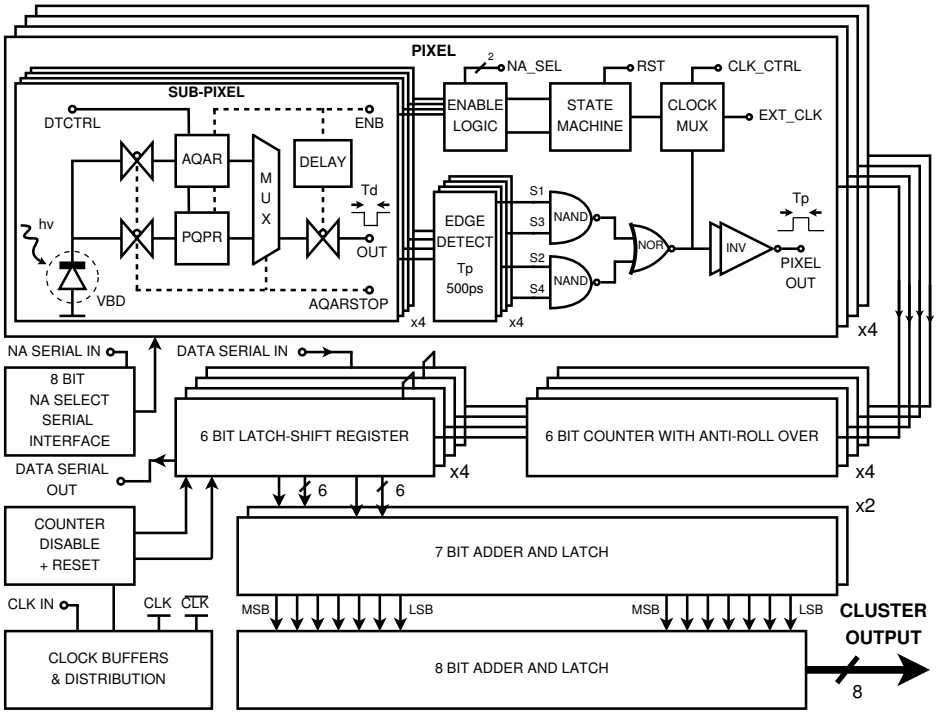


Figure 4.15: A block diagram of a Cluster of four Pixels, each of four Sub-Pixels. A set of four 6-bit digital counters are implemented prior to a duel-clock edge triggered parallel pipelined adder tree. A local 8 bit flip-flip register is provided to set N_a values for each Pixel and four daisy chained latch-shift registers allow pre-summation counter readout.

period, X is greater than a threshold, x, where $P(X > x) = 1 - P(X \leq x)$. Digital counters, if not stopped, iterate to zero once they reach a value of $2^{BitDepth} - 1$ [29, 69]. In Equation (4.9), η_e is the effective Pixel detection efficiency, $E[\Psi_{px}]$ is the expected photon rate incident on the Pixel and ξ is a multiplier that describes the photon transfer curve. Table 4.2 and Equation (4.9), show that as the bit depth increases, the probability the counter will roll-over within a symbol is reduced[1]. Despite this, an anti-roll over circuit is included for characterisation purposes.

Examining both single and 16 channel cases, for a BER of $1\text{x}10^{-9}$, approximately 22 detections are needed (Section 3.3.1). This gives mean photon flux rates incident on each Pixel of 0.086 and 1.375 respectively per symbol. The probability of roll-over is based upon the mean photon flux. The Poisson distribution however has a lengthy positive tail [116]. This suggests that to prevent counter roll-over, the bit depth should be chosen based upon a higher incident flux such as the mean plus two standard deviations ($\mu + 2\sigma$). This gives pixel flux rates per symbol of 0.6725 and 3.72 respectively.

Table 4.2: Counter roll-over probabilities at various count rates

Depth	Max Counts	P(Roll-Over)			
bits	$2^{bitdepth}-1$	$\tau_b E[\Psi_{px}]=0.086$	$=1.375$	$=0.673$	$=3.720$
1	1	0.0035	0.40	0.1463	0.886
2	3	$2.13\text{x}10^{-6}$	0.05	$5.00\text{x}10^{-3}$	0.51
3	7	$6.86\text{x}10^{-14}$	$9.43\text{x}10^{-5}$	$5.72\text{x}10^{-14}$	$3.62\text{x}10^{-2}$
4	15	0	$2.15\text{x}10^{-12}$	$3.3\text{x}10^{-16}$	$1.99\text{x}10^{-6}$
5	31	0	0	0	$6.6\text{x}10^{-16}$
6	63	0	0	0	0

The table assumes the efficiency, $\eta_e=1$ and the response function multiplier, $\xi=1$.

Pixel Counter Bit Depth

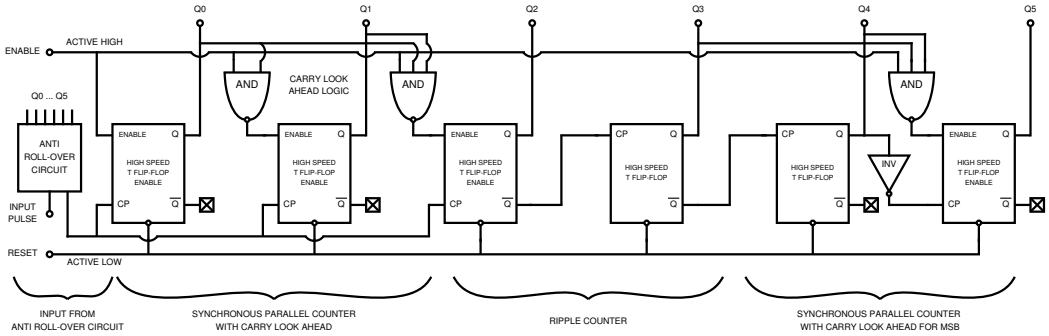
The bit depth, of all counters in this design, was chosen to be 6-bits [29]. This was selected for array characterisation purposes. A larger bit depth can also be used in order to extend dynamic range headroom. This is useful if ambient photon arrivals are included or if the mean DCR changes between manufactured die.

4.6.2 The Pixel Photon Detection Counter

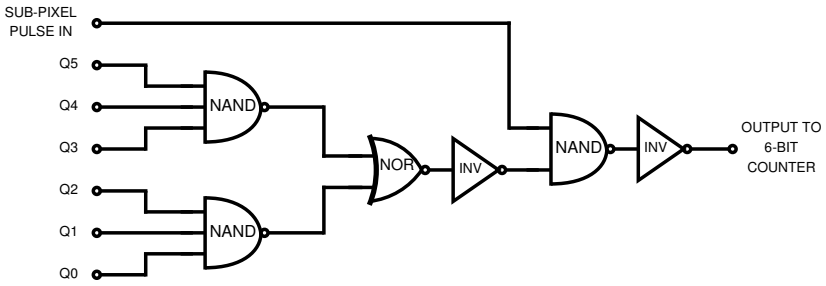
Figure 4.16, shows the design of a 6-bit synchronous parallel counter [29, 69, 215]. With a ripple counter, upon iteration, there is a delay before all bits are valid [29]. The parallel carry look-ahead architecture allows a short final iteration to readout period [29], so that the counter can iterate until the end of the integration period. The counter has access to a buffered high speed *ENABLE* signal. This allows fast gating of the counter for accurate integration times and the disable-latch-reset action that is used within the Cluster.

In Figure 4.16.a, the counter is presented with split carry look-ahead logic [29]. The first three bits use a synchronous parallel architecture [215], ensuring all bits change at the same time. After the first three bits, the look-ahead logic stops and a standard ripple counter is used. This minimises area as look-ahead logic beyond this requires an increase in AND gate inputs [69, 215]. The carry logic is utilized when the separation of pulses becomes short. The architecture reduces the area required, while ensuring the counter copes with short inter-arrival times, of the order of $\frac{N_{TDM}}{\tau_d}$.

The anti-roll over circuit (Figure 4.16.b) prevents the counter from receiving a pulse if it is at the 2^6-1 maximum. This is required to ensure that upon latching, it represents the total number of photons detected. If a counter were to roll over, a one to many mapping would result, severely impacting the BER.



(a) A 6-bit hybrid ripple, synchronous parallel counter with enable and carry look-ahead logic.



(b) An anti-roll over circuit designed to prevent Pixel pulses iterating the counter once it reaches $2^6 - 1 = 63$, a code of '111111'. Its output is connected directly to the synchronous input of the counter in (a).

Figure 4.16: Schematics of a) an individual Pixel 6-bit counter and b) an anti-roll over circuit. Carry look ahead logic is implemented with a split counter design. This ensures fast counter iteration with a small carry look ahead logic area overhead. Fast buffered enable signals allow gating of the counter for use at the end of received data symbols or when implementing an accurate measurement period.

Figure 4.17 simulates the incident pulses, iteration and final latching of data from a counter. An example integration time of 22ns includes a short counter blanking period during which the counter is disabled, latched and reset. At $\approx$ 22ns, a pulse is missed as it is incident during this period. Further, at $\approx$ 50ns, two incident pulses are still counted despite a negligible pulse separation of $\approx 100ps$, demonstrating the use of the fast carry look-ahead scheme. The counter design provides robust counting of photon events and zero roll over probability. This ensures the receiver is limited by SPAD dead times and efficiencies rather than digital circuitry.

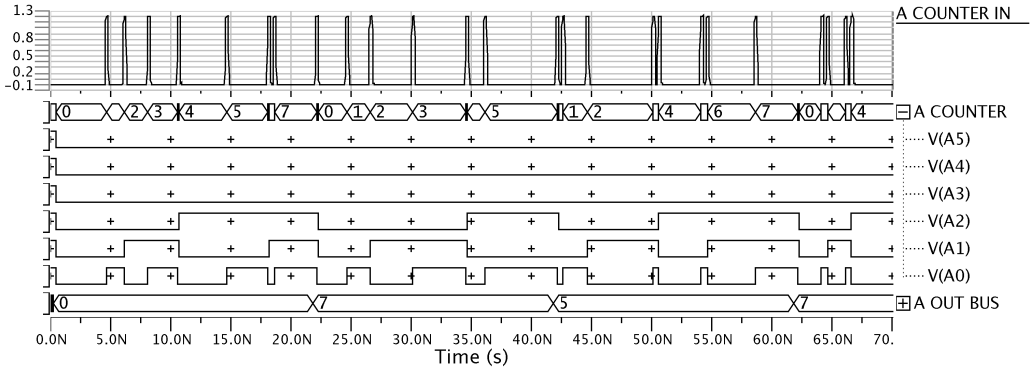
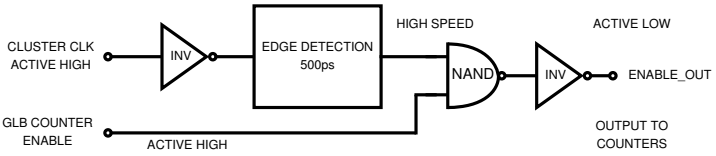
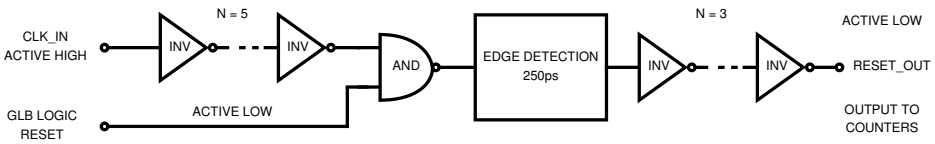


Figure 4.17: Digital counting of Sub-Pixel pulses and the latched 6-bit data. The counter integration time is simulated at 22ns, and shows that events with short inter-arrival times can be captured.

4.6.3 Counter Data Latching and Reset Timing



(a) Counter enable and disable logic.



(b) Counter reset logic using a modified edge detection circuit.

Figure 4.18: Counter enable and resetting pulse generation logic. Crucially both circuits allow their respective global signals to pass, ensuring correct high level operation.

The counters are controlled by three clock edges. A negative going pulse generated by the circuit in Figure 4.18.a, disables the counters for 500ps. This prevents corrupted data further along the data chain [69]. The positive edge of the clock arrives during this time to latch the data into the 6-bit latch-shift registers. A delay on the latching clock pushes it towards the centre of the disable pulse. After a further delay, a 250ps reset pulse (Figure 4.18.b), forces all counters to zero.

In Figure 4.19 the disable, latch and reset action has been simulated. The process requires $\approx$ 1ns, hence in implementations with clock speeds greater than 200MHz, a dual counter, multiplexed method should be used. At 100MHz the integration period is shortened to 9ns, however this is advantageous as a phase shift, Φ_{offset}, can position the integration window within the centre of a symbol, allowing $\approx$ 500ps either side for μLED transitions. Ignoring

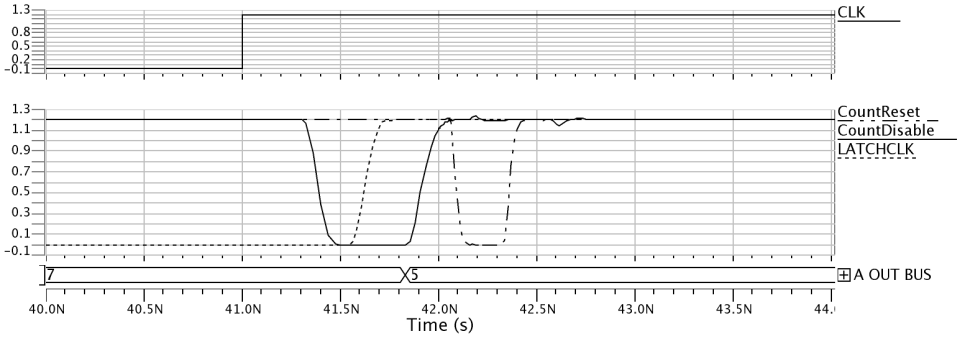


Figure 4.19: Counter disabling, latching and resetting at the end of each clock period. The delay of the latching clock is clearly seen, forcing a a transition near the middle of the counter disable pulse. Likewise the counter reset transitions after correct data latching.

counts near the transitions may help to reduce the BER [1, 14, 35, 46, 73, 83, 113, 117, 220], by decreasing the standard deviation of the distributions (Section 3.2). For experimental work, if $\Phi_{offset} = 0$, the integration begins at the start of a symbol, which may cause issues with impulse or step response induced BER elevation. This will be measured and discussed in Section 6.2.

4.6.4 The Latch-Shift Register

Upon latching, the data is stored in four reconfigurable latch or shift registers (Figure 4.20) [69]. If *LATCH-SHIFT* is low, the circuit behaves as a pipeline register. Data is latched on the clock positive edge and is presented and held on a buffered output [69, 215]. If however LATCH-SHIFT is high, then data is moved from the LSB towards the MSB, creating a shift register. A carry input daisy chains these to form the serial data path used for Pixel readout [69, 215].

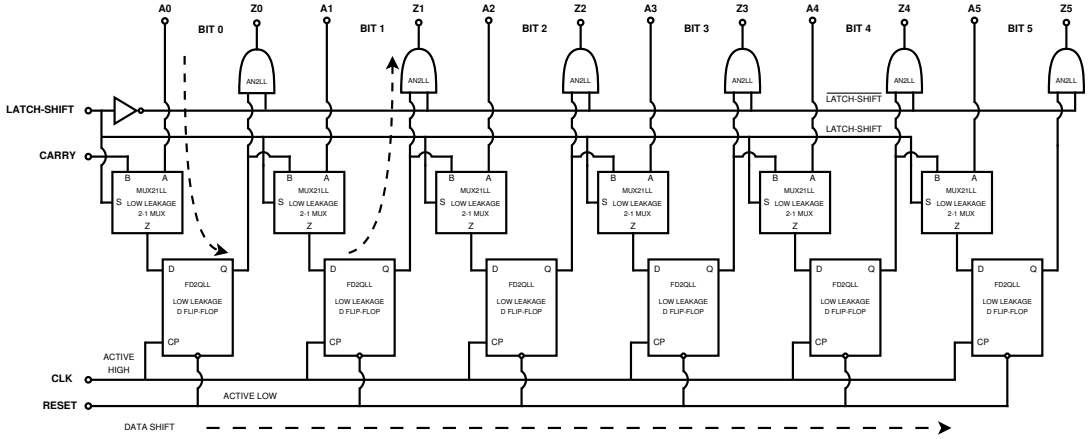


Figure 4.20: A reconfigurable Latch-Shift register.

4.6.5 Na Configuration Serial Interface

A second serial interface (Figure 4.21) is used to configure the N_a values of the constituent Pixels. This is also chained throughout the Array (Section 3.4 and Section 4.5).

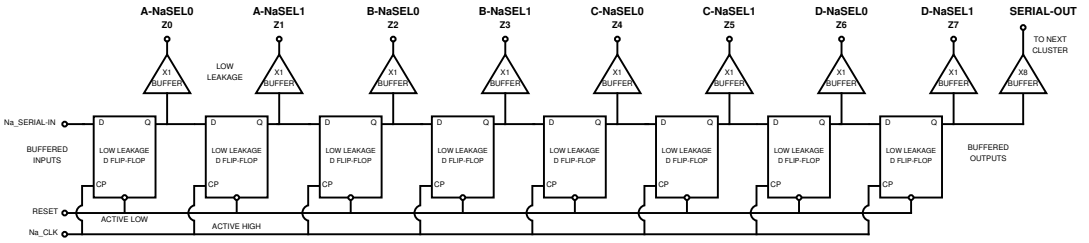


Figure 4.21: An 8-Bit serial interface used within the Cluster. This provides the 2-bit TDM N_a mode selection signal that is used for each of the four Pixels within this group.

4.6.6 The Parallel Pipelined Adder Tree

Figure 4.22 shows the functionality of the parallel pipelined adder tree [69]. As discussed, upon the clock positive edge, the integrated SPAD counts from the preceding cycle are latched, as the values on the *A OUT BUS* show. Two 7-bit ripple adders [29, 69, 215], compute $A+B$ and $C+D$ with small delays. These values, shown in *A+B OUT BUS* and *C+D OUT BUS* are latched on the clock negative edge. A dual edge clocking scheme is used to reduce the latency between data input and the summation output [69], however it doubles the core operating frequency (nominally 200MHz). As such, the adders must settle within half of the clock period (5ns) [69]. These two buses are then added by an 8-bit adder and latched on the next positive edge [29, 69, 215]. The advantage of this scheme for the silicon device, is that the summation computation is not limited to the pad maximum data rate of 108MHz.

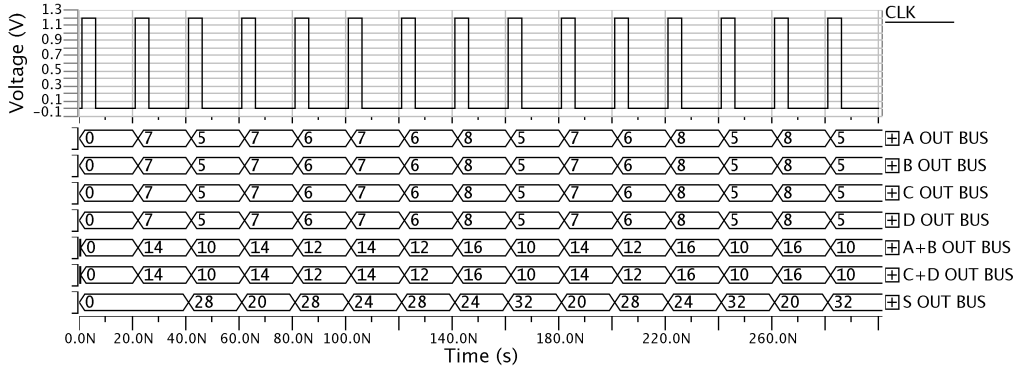


Figure 4.22: Digital summation of four Pixel counters to a single 8-bit output bus showing positive edge summation outputs.

4.6.7 Conclusions from The Cluster Implementation

The Cluster of sixteen SPADs described in this section directly provides functionality to create a discrete time integration for the integrating receiver architecture. The structure and bit depth of *per Pixel* photon counters has been discussed, aiming to reduce issues such as count losses or non-monotonic behaviour at the receiver level. The two serial interfaces provide functionality for device testing but also the idea that each Pixel within the receiver can be individually addressed and independently assigned an operation mode. In the next sections, digital summation is continued, eventually forming functionality for both single-channel and multiple-channel receiver operation.

4.7 A Single 64-SPAD Receiver Channel

Within the hierarchy, a single channel is created from a set of four Clusters (Figure 4.23). This represents one of 16 receiver channels on the silicon device (Table 3.8). The chip could be used for Optical Many-In-Many-Out (OMIMO) techniques [20, 95], as these channels can be read out independently. Along with clock, signal and data buffering, this level continues the parallel addition creating a 10-bit output. The block has a fill factor of 3.2% in the $N_a = 4$ mode and continues the N_a configuration and Pixel read serial interfaces.

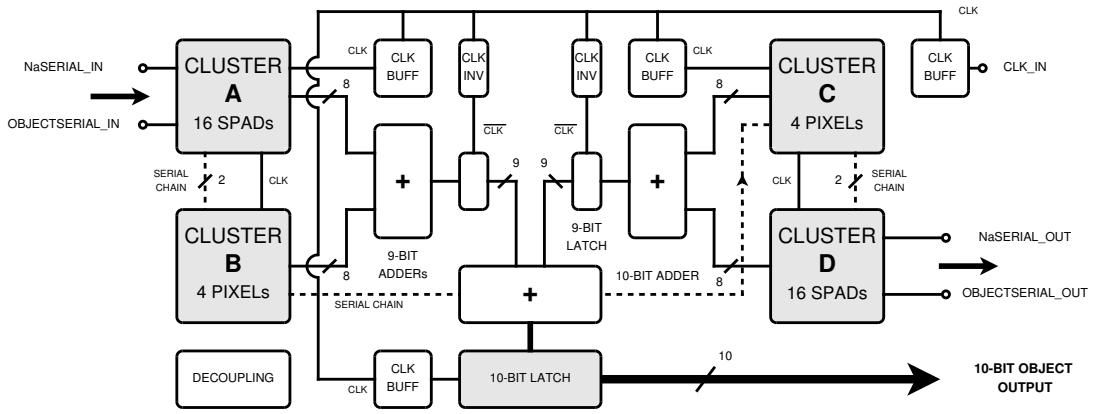


Figure 4.23: Block diagram and clocking strategy of a single data channel Object. The 10-bit output bus is multiplexed at the pad ring allowing operation of 16 channels on the receiver die.

Due to the physical separation of the Cluster clock inputs, these clocks are split into the AB_ClusterCLK and the CD_ClusterCLK. As shown below, each Cluster outputs a new value upon the arrival of the positive clock edge. The addition is performed in the proceeding half clock period, being latched by the AB_LatchCLK or CD_LatchCLK. The 'S OUT BUS' shows the combined summation of all constituent Clusters. Simulating this block (Figure 4.24) shows that the DVDD current draw is predominantly dynamic power [69] due to logic switching and SPAD events. This current increases with both clock frequency and SPAD count rate [69].

This level of the silicon chip hierarchy, together with the levels discussed in the next section, directly implement the single or multi-channel functionality required by the project specifications (Section 2.9). The motivation for this, is an underlying communications effort using multiple μLEDs. The CMOS driver circuits for these LEDs allow both a) reconfigurable grouping of LEDs into channels and b) independent data streams to be sent to separate channels [96].

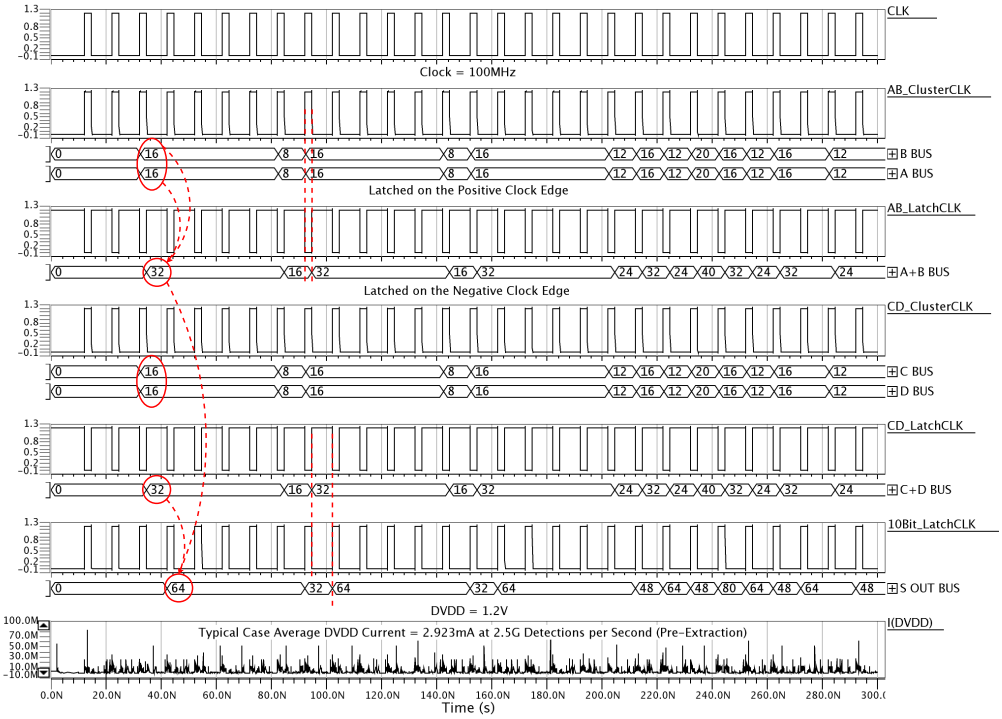


Figure 4.24: Digital simulation with cluster and latch clocking for a data Object. The red lines on the left highlight the timing of the signal summation, with the dual edge clocking and low latency shown between 90 and 100ns.

4.8 The Array and Top Level Device Architecture

In this section, two levels of the hierarchy are discussed. These are the Quadrant and Array levels that were defined in Section 4.3. The main functions or device characteristics discussed here include:

- SPAD photon-count summation using parallel addition circuitry,
- Output provision for both single-channel or multiple-channel operation,
- Synchronous, low skew, symmetric clocking of the chip hierarchy, to ensure equal communications performance over each sub-level,
- An overview of the device optical fill factor and top level silicon implementation, and
- A summary of the testing system printed circuit board, Verilog firmware and Java software architectures.

Implementation at each of the levels, from the 16 individual 64-SPAD channels to the control software, is critical for device characterisation and communication experiments with the μLED transmitter array. As a direct example discussed below, inadequacies in clock distribution implementation will directly impact the receiver inter-symbol interference. Likewise problems in the firmware domain may act to restrict control of the silicon chip or prevent adequate data acquisition for both on-line and off-line communication performance analysis.

4.8.1 Array Quadrants: Collections of Channels

The quadrant level, shown in Figure 4.25, is formed from a group of four separate channels. This provides buffering of the channels to the pad ring multiplexers used for the multi-channel mode, along with circuitry to continue the adder tree. Time multiplexing of the four separate channels within this block, is used to ensure a core-limited silicon design [69].

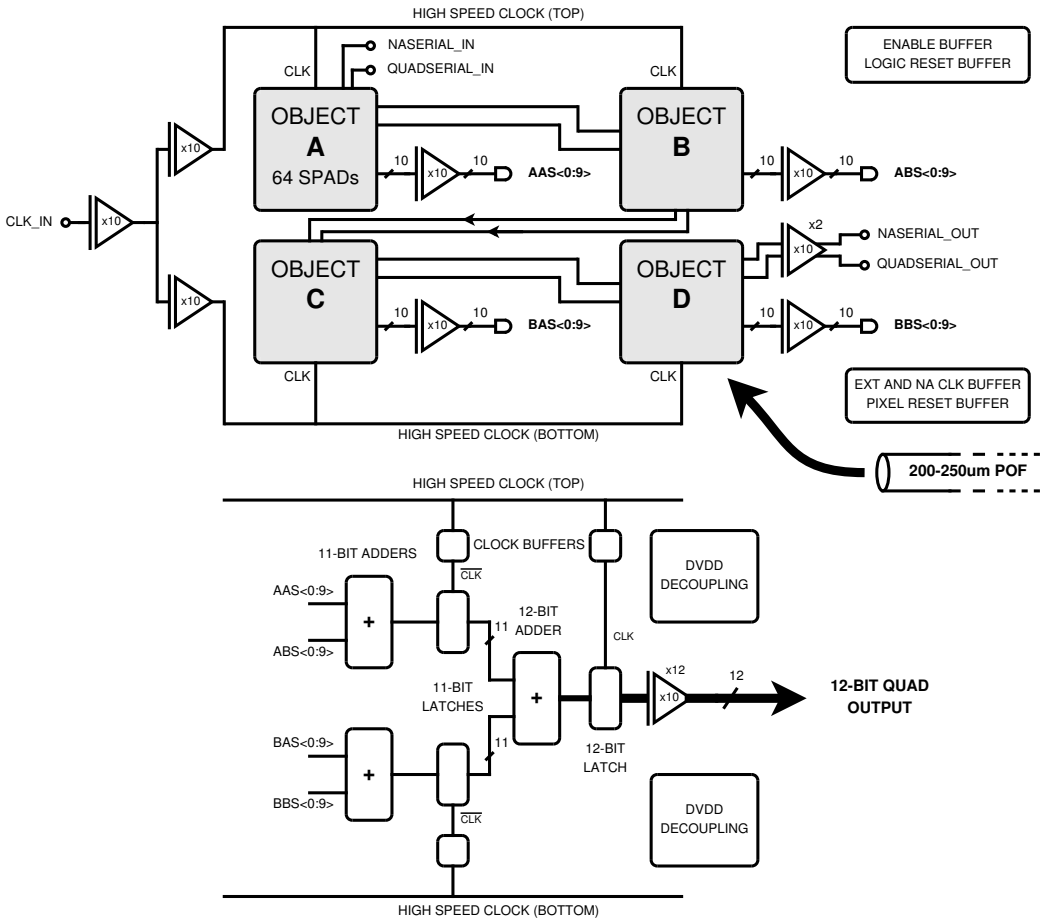


Figure 4.25: Schematic diagram of an Array Quadrant. A weak link exists between the top and bottom high speed clock lines, helping to reduce clock skew. This level of the hierarchy performs buffering of the 16 separate channels to the pad ring, along with continuing the pipelined addition scheme.

While not shown, a weak link exists between the branches of the high speed clock tree. This creates a clock mesh [29, 69, 215], reducing the skew that would have been problematic at the top level. The digital buses labelled as AAS, ABS, BAS and BBS, relate to the four 10-bit outputs from the Quadrant, as shown in the functional schematic (Figure 4.26).

4.8.2 Array Digital Outputs and Maximum Clocking Constraints

The summation of SPAD pulses continues to 14-bits, which can be output as a single channel (Figure 4.27). AOUT_BUS to DOUT_BUS provide access to the 16 separate channels. Each Quadrant, uses a 4–1 10-bit multiplexer to cycle through the channel outputs. This is simulated in Figure 4.28.

The four multiplexers, use a second clock, labelled $CLK2$ (Figure 4.26). In the multi-channel mode, this runs at four times the $CLK1$ frequency, prompting each channel to be presented sequentially within a symbol period. This necessitates a lower clock frequency for the main Array ($CLK1$). In this case, a frequency of 100MHz ($CLK2$) allows all channels to be read within the period of a 25Mb/s per channel transmission ($CLK1 = 25\text{MHz}$). While this limits the clock, a second generation could use on-chip thresholding and dedicated high-speed LVDS I/O for each channel [35, 69], reducing the number of pads and eliminating channel multiplexing.

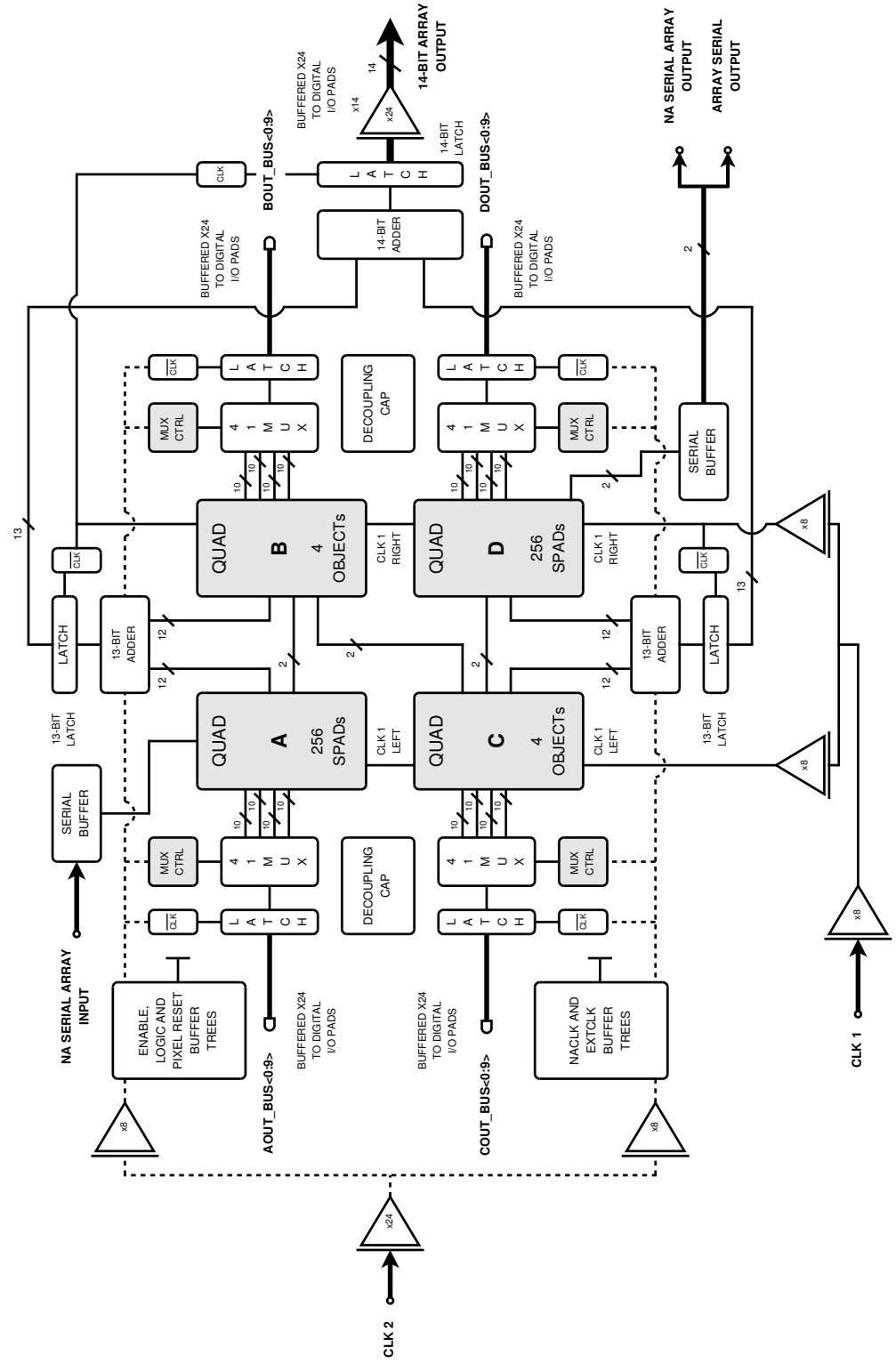


Figure 4.26: Functional schematic of the Array top level and clock trees. This level distributes both $CLK1$ and $CLK2$ clocks, generates the 4–1 multiplexed multiple channel output clocking scheme and continues the full summation pipeline.

4.8.3 The Clock Tree Structures

The distribution of *CLK*1 and *CLK*2 provides stable global clocks with buffering at each stage [69]. The clock H-tree [29, 69, 215] structure is symmetric due to the nested replication down to the lowest levels. This gives approximately uniform delay and jitter values [29, 69]. Jitter varies the start, stop and total integration time of the digital counters and along with voltage and circuit variability, will act to spread the distribution of a measured bit towards a super-Poisson distribution. Hence it will lead to a circuit based increase in the BER (Section 3.2).

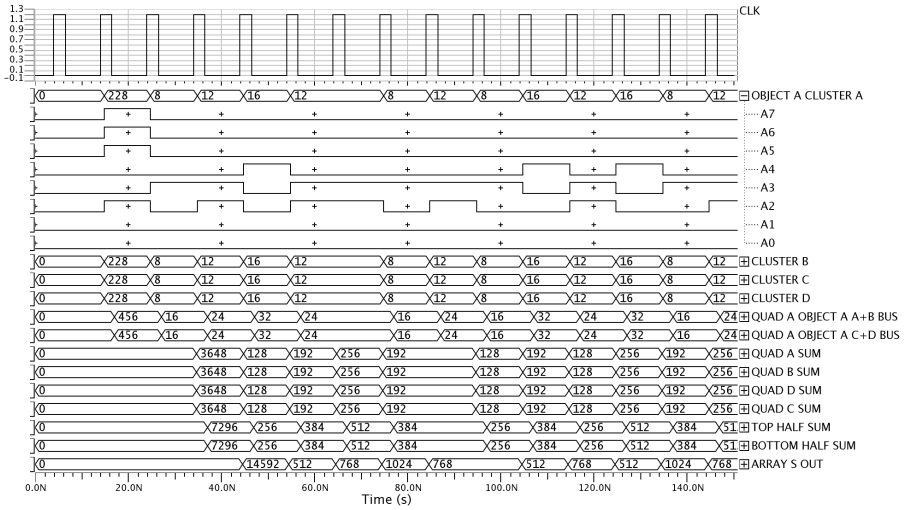


Figure 4.27: Simulated digital summation from the Cluster level to the 14-Bit Array output. The short latency is exemplified by the dual clocking approach, presenting summation data from a Quadrant partial summation (*QUAD A OBJECT A A+B BUS*) on the negative edge of the clock.

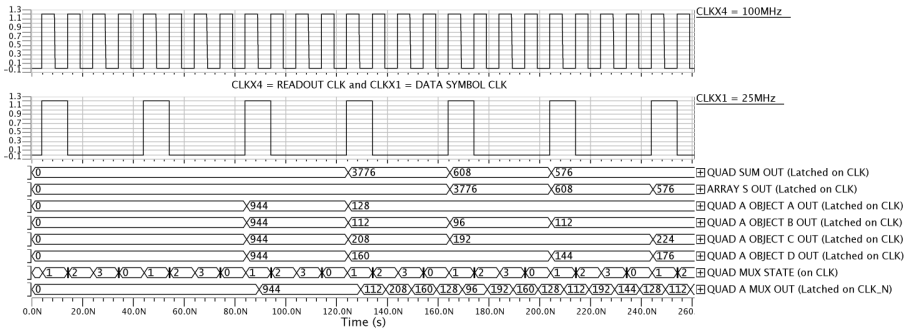


Figure 4.28: Simulation of the Quadrant output multiplexing, showing parallel symbol reception by each data Object, albeit at $\frac{1}{4}$ of the maximum clock frequency.

4.8.4 Receiver Synchronisation with Transmitters

As this receiver uses a clock to perform summation, integration, readout and resetting operations, synchronisation between the μLED transmitters and the receiver is critical [1, 2, 3, 35]. The Hypix project aims to investigate the front-end circuitry needed for transmitter drivers and receiver operation [66], however in this short section, the manner in which the system is to be synchronised is briefly discussed. In Section 4.6, the use of on-chip phase-lock loops and clock and data recovery circuits was suggested for this operation [219]. The aim of these circuits, is to obtain a reliable, low-jitter clock directly from the data received [35, 219] by locally mixing the incoming data with a local oscillator. While outside of the scope and time-scales of the Hypix project, the resources of a testing FPGA can be used for this purpose, or future designs can incorporate the concept monolithically with the main receiver [46, 114, 205, 220, 219].

The system synchronisation would first be based on a short transmitted training sequence. This could simply be a serial digital stream such as, 0101010101010101 etc [35], as commonly used between transmitted frames within data link protocols [22, 102]. If this is mixed with a local digital clock, perhaps using an XOR gate [35], an output proportional to the difference in phase is produced [35, 219]. A phase-lock criterion can be made from this output, depending on the specifics of the architecture and the relative phases of the transmitter and receiver clock. The phase difference proportional output can then be used to modify the phase and frequency of a voltage-controlled oscillator (VCO) or digital delay-locked-loop (DLL) either on-chip [46] or on the system FPGA (Section 4.8.6). For the purely digital, discrete time case, where the FPGA DLL clocking resources are to be used, the frequency and phase can be adjusted using digital control words and a control state machine. Synchronisation would be achieved by iterating the FPGA clock phase (increments of 1.4° degrees for the Xilinx Spartan 3 Digital Clock Manager (Xilinx UG331)), until a suitable output condition is reached. Finer phase control can be achieved by using a multiplied rate clock for the DCM, then using the DCM's clocking resources to divide or process the output to create the final SPAD-array system clock. A periodic check, perhaps at the end of each data frame, could check the phase condition, and if it has departed from the chosen value, can increment or decrement the phase control word until a phase lock is achieved.

In Figure 4.29, an example idealised optical input is shown with $N_0 = 1$ photon/symbol and $N_1 = 20$ photons/symbol. The figure also shows four clock phases, 0°, 90°, 180° and 270°, each relative to the μLED transmitter clock. The 0° phase clock represents perfect synchronisation between the transmitter and receiver. The 180° phase example is interesting as, in this simple example, the cycle to cycle SPAD-array output remains constant. The intermediate 90° and 270° phase examples indicate a route forward in forming a phase locking criterion to be used with the system VCO or DLL. For each of the clock phases, the output represents the SPAD-array summation and integration over the clock period. For example, for the 0° phase, the array will detect and integrate all 20 of the photons within symbol S_N, while for the 90° phase,

the array will detect $\frac{3}{4}$ of the 20 photons within symbol S_N and $\frac{1}{4}$ of the average 1 photons within symbol S_{N+1}. Below this output value, the difference between consecutive symbols $(S_N - S_{N-1})$ is shown. It is clear from Figure 4.29, that the magnitude of this time-varying, phase-proportional output, can be used for transmitter and receiver clock synchronisation.

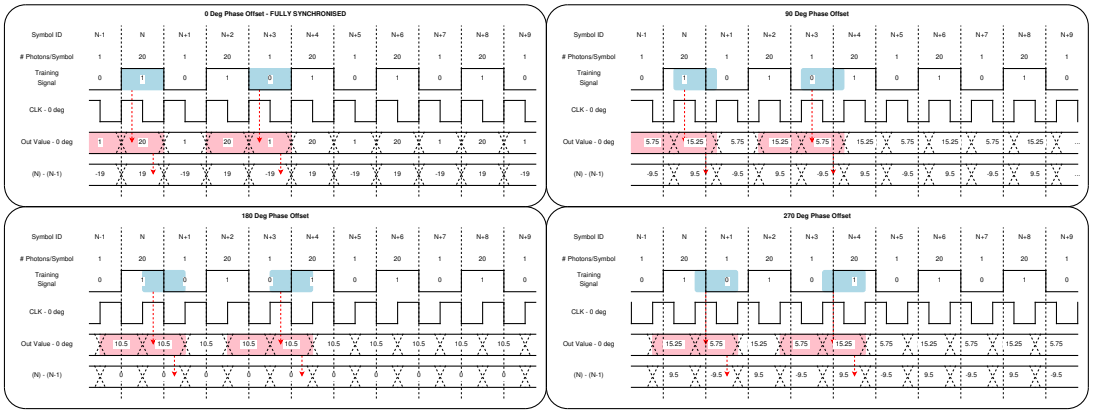


Figure 4.29: Transmitter and receiver clock synchronisation scheme timing diagrams for 0°, 90°, 180° and 270° clock phases, relative to the μLED transmitter clock. The $N_0 = 1$ and $N_1 = 20$ photon/symbol levels are assumed to be consistent for simplicity. The SPAD-array integration is shown in blue, while a $S_N - S_{N-1}$ operation is shown in red. For a phase of 180°, the $S_N - S_{N-1}$ operation yields a constant, near-zero output, while phases of 90° and 270°, show a time-varying output. At 0° clock phase, the magnitude of this $S_N - S_{N-1}$ signal becomes maximised, giving a symbol to symbol magnitude of $19 - (-19) = 39$ counts.

In Figure 4.30, the above clock phase, symbol $S_N - S_{N-1}$ and variation magnitude concept have been modelled. Just as the FPGA digital clock manager would sweep the phase, attempting to lock to the same phase as the transmitter clock, Figure 4.30 sweeps the phase from 0° to 360°. At 180°, the difference between received N_1 and N_0 bits becomes zero, indicating maximal bit errors ($BER = 0.5$). As the phase moves away from 180°, towards either 90° or 270°, the magnitude of the difference between received bits increases and thus the BER improves. One possible criterion then for transmitter and receiver synchronisation, can be defined by Equation (4.10).

$$\text{LOCK} = \text{TRUE}, \text{ if } \max_{0 \leq \phi \leq 360} |S_N - 2S_{N-1} + S_{N-2}| \tag{4.10}$$

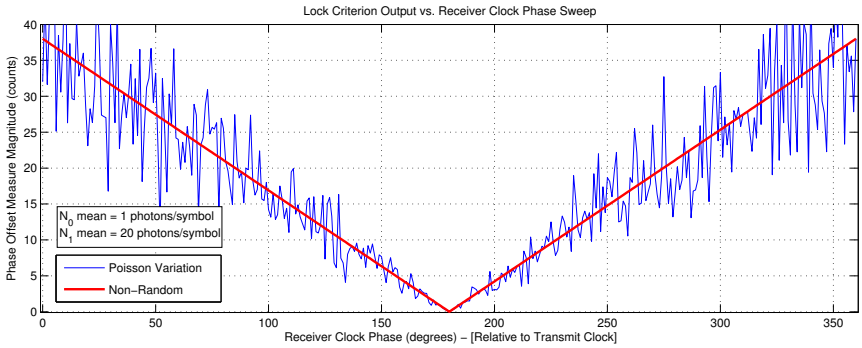


Figure 4.30: A sweep of the SPAD receiver front-end system clock phase with respect to the μLED transmitter clock. At 0°, the receiver is fully synchronous with the transmitter and the magnitude of the difference between received bits reaches a maximum. With a maximum difference between N_0 and N_1 bits, the BER reaches its optimum point. In contrast, at a relative phase of 180°, the N_0 and N_1 bits become approximately equal, yielding a difference magnitude of zero and the worst case system BER of 0.5. The model uses a mean N_1 value of 20 photons/symbol, a mean N_0 value of 1 photon/symbol, and is shown in blue incorporating Poisson statistics on both 'zero' and 'one' bits from the 0101010101010101 synchronisation training sequence.

This synchronisation scheme requires more digital implementation, however in this project, the FPGA and PCB resources are used to allow future work, design adaptability and easy testing of the receiver front-end circuitry. Once synchronisation has been achieved, small, low-speed variations in the transmitter clock frequency or phase can be tracked by the receiver CDR system, to maintain a stable lock [35, 219].

4.8.5 Array Fill Factor

The fill factor is lower than required in single-photon counting applications [158, 221]. The device (Figure 4.31), discounting the pads, achieves a fill factor of 2.42% in the $N_a = 4$ configuration. This is comparable to previous arrays in this process that also do not use enhancement techniques [70, 154, 155]. While this is low in comparison to recent photon counters [56, 57, 71, 157, 206, 221], the receiver compares well to current high speed receivers. For example [31] achieves a 0.16% fill factor, [48] achieves 4.8%, [44] achieves 2.0% and [110] achieves 1.12% (Table 6.3).

The device fill factor should be addressed in future generations to increase sensitivity as illustrated by [57, 58, 141, 157, 159, 160, 208, 221]. However in this work it presents a fixed power penalty (Section 3.3.5), that is expected to scale with advancement.

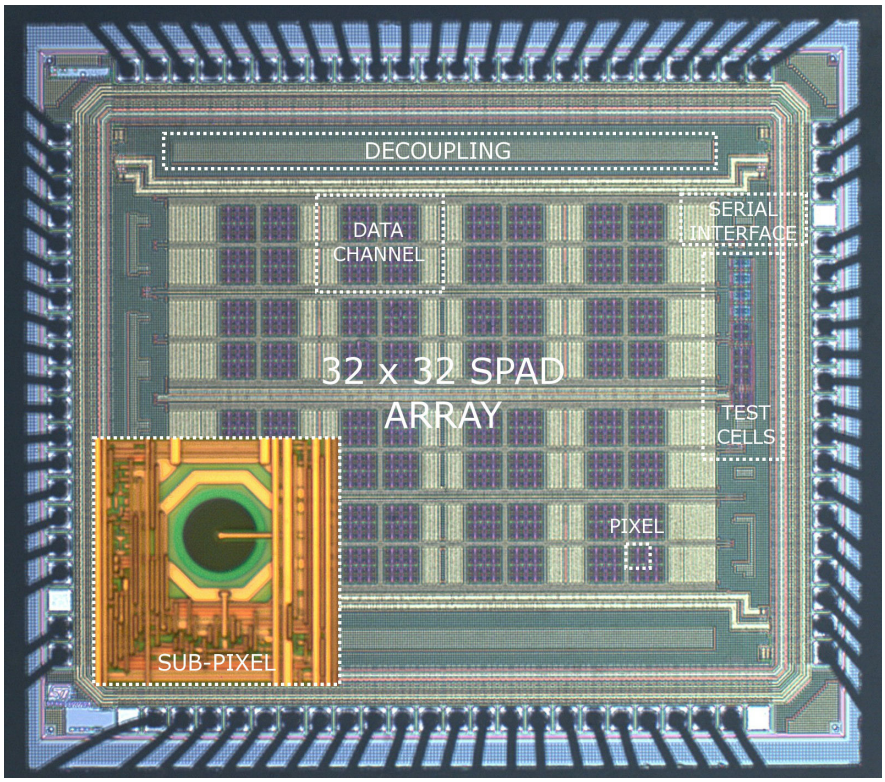


Figure 4.31: Receiver micrograph showing the structure of the final device, along with an inset zoom of a Sub-Pixel. The main 32x32 array, on-chip decoupling capacitors, test cell column and 11-bit serial interface are each highlighted, as is a single receiver channel out of the 16 channels forming the main array. The full array contains 1024 SPADs.

4.8.6 The Final System, PCB, Firmware and Software Architecture

Figure 4.32, shows the overall system. This includes the application specific integrated circuit (ASIC) developed here, but also includes functionality developed for testing purposes [69, 85, 222] (Section 1.3). A PCB with firmware [223] and software [224] is used to allow automated testing. The design for testability methodology is used throughout these subsidiary blocks, giving the maximum scope for experimentation [69]. The system has a number of limitations, with the most significant being use of a parallel data stream. Many devices within the literature use single-bit output streams connected directly to an oscilloscope [31], however in those circuits the threshold applied to the data reduces the ease of testing a prototype device.

USB2.0 communication to a PC (Figure 4.32) is used for data acquisition and configuration. This streams data from the firmware implemented on a Xilinx field programmable gate array (FPGA). However despite this, the USB link remains a bottle neck as the ASIC output rate is higher than the 48MHz USB byte clock.

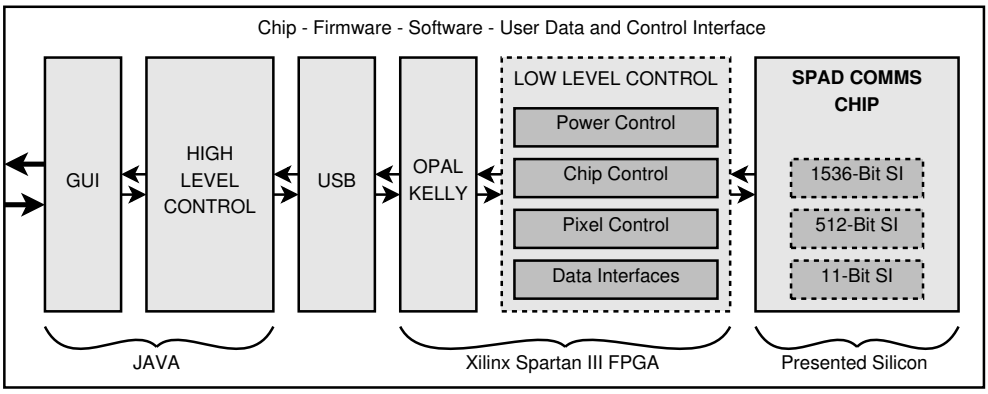


Figure 4.32: Overview System Block Diagram. The overall control and testing system hierarchy is shown driving the CMOS device from a Java software experimental platform. Modular design, and design for testability are used throughout the system, complementing the functionality implemented within the receiver chip.

4.8.7 Conclusions from Silicon and Testing System Implementation

In the above section, the silicon design of the receiver top level, including the input/output pad ring and considerations for synchronously clocking the device, have been presented. Further the testing system, which includes a PCB, firmware implemented on a Xilinx FPGA and software running on a host computer, are each described. This allows testing of the receiver concept using the *design for test* methodology to facilitate experimentation.

4.9 Discussion and Implementation Conclusions

This chapter has presented the implementation and fabrication, in an 130nm CMOS technology, of a 32x32 SPAD-based receiver. This is based on specifications derived in Section 2.9 and Section 3.5. The receiver is designed for concurrent use with μLED transmitters that have been developed using on-off key modulation and visible wavelengths.

The current device is analogous to a digitized version of receiver analogue front-end circuits right up to the data thresholding circuit, yielding full CMOS logic amplitude signals suitable for signal processing. The single or 16-channel receiver offers multiple output modes along with on-chip reconfigurability, for testing purposes. The device obtains a fill factor of 2.42%, which is broadly comparable to other SPAD arrays in the literature that do not use fill factor enhancing techniques, but requires experimentation, modelling and future work to take advantage of the sensitivity of SPADs.

The implication of the developed receiver, is that a fully digital architecture, using CMOS integrated SPADs, can be tested for the first time. In the next two chapters, the low-level performance of the device is first tested, followed by communications tests using the μLED transmitters.

Chapter 5

Low Level Characterisation

5.1 Low Level Characterisation Introduction

This chapter will discuss initial characterisation of the SPAD array that forms the optically active area for the OOK-only, SPAD-based receiver (Chapter 4). This investigates the silicon implementation, centering on low-level results. Doing so indicates the performance of the SPADs and allows initial assessment as to their impact. The results and subsequent models allow insights into noise sources, power penalties and practical limitations (Chapter 3). These insights include the presence and magnitude of as-yet-unknown affects on data reception performance, prior to communication centered experiments presented in Chapter 6.

Section one first deals with the receiver power draw, separating the power for photon counting and that required by the clock distribution network. This helps to evaluate the electrical efficiency and indicates areas in which power can be saved in efficient future designs.

The dead time is then discussed, measured and shown to restrict the number of detections within a short symbol. The section investigates issues that may impact the performance of a SPAD-based receiver, such as the detection to detection variation of the dead time. This impacts methods of specifying SPAD-based designs.

The dark count and after-pulsing rates of SPADs are also measured in this first section. For communications, these results help evaluate some of the receiver power penalties, but also indicate paths to receiver adaptability, post-manufacture, if high noise SPADs are removed.

The change in receiver responsivity with light level reproduces previous measurements [60, 163]. This is also presented in the first section, aiming to indicate any departures from models established in Chapter 3. In particular, the modes in which each pixel can operate are explored, allowing determination of their use in such a receiver.

The chapter concludes, in the second section, with a model exploring the change in SPAD sensitivity. The model looks at competition between photon arrivals, dark count events and secondary avalanches. It seeks to determine if modelling can be extended for the specification and design stages of SPAD-based receivers (Section 3.3.2) and if the native noise of the diodes change when operated at different light levels.

5.2 Low-Level Measurement Results

In this section, low level measurements build up an understanding of the receiver developed in Chapter 4. These test the device against the designed performance, ensuring functionality such as the clock distribution, the SPAD signal counters and the pipelined adder tree operate correctly. The preliminary tests in this section, investigate the performance of the device as a photon counting array, prior to communication tests performed in subsequent chapters. This provides the ground work for communication metrics, or more complex tests such as the impulse response.

When using the device as a receiver, non-ideal behaviour, such as an increase in inter-symbol interference or the BER, may be observed. Therefore, testing the device may give indications as to the roots of particular performance limitations. Further these low-level experiments can be used to fix or refine estimates for the power penalties. These will be used within Chapter 6 to estimate the BER and receiver sensitivity.

The experiments discussed here, are primarily ordered first with factors that are reasonably static in time or use long term average illumination levels. The order then moves onto temporal affects, matching the ordering of experiments in Chapter 6. The experiments are also ordered by experimental complexity. This ensures that measurements with a simple voltage meter or using on-silicon circuits such as counters, are presented before those that require complexities in laser timing, data acquisition or analysis.

5.2.1 Receiver Power Issues

In this section, the power consumption of the receiver is tested. This aims to separate the contributions from the silicon implementation of a distributed digital clock tree and the contribution from the counting of individual photons. There are design techniques available for the optimisation of clock distribution networks, or the reduction in CMOS leakage power, however testing is required to ascertain the exact relationship between incident optical power and the power required for the counting of photons.

This section therefore first discusses the power consumption with increasing clock frequency. The section then goes onto testing the power required by photon counting. In both cases, functionality included on-chip such as global enable/disable, are used to separate the contributions.

The Receiver Power Consumption With Clock Frequency

Without SPAD activity and with the 6-bit counters disabled, the clock power consumption [29, 69, 215] can be measured. In Figure 5.1, as the clock frequency decreases below 1MHz, the current and power consumption saturate at 0.5mW from DVDD.

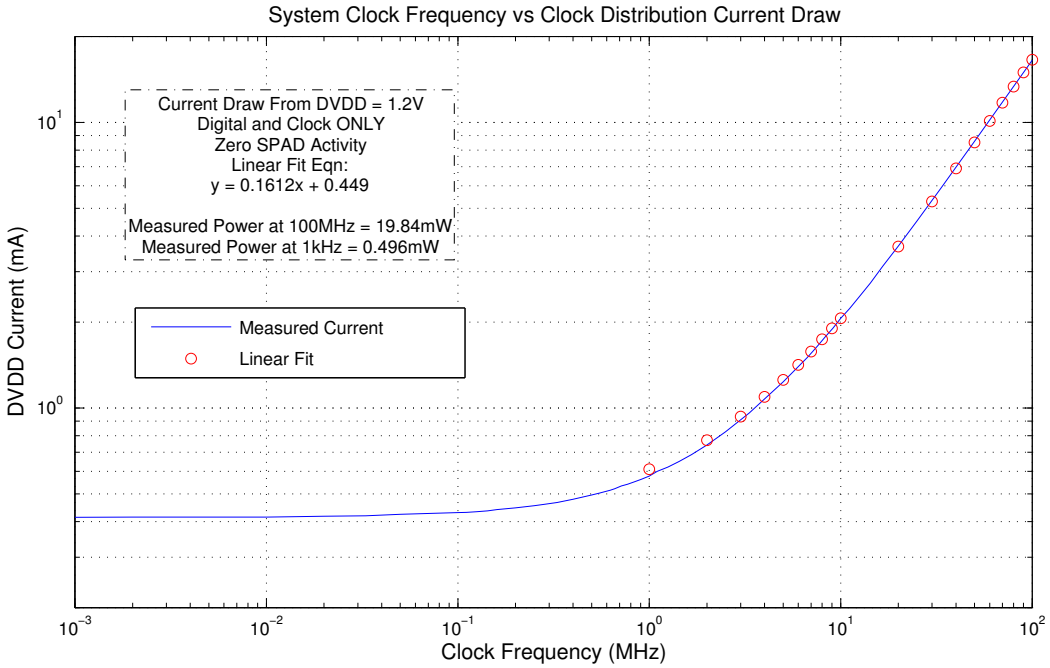


Figure 5.1: DVDD current draw with system clock frequency. The static power due to transistor leakage gives a continuous ≈ 0.5mW power draw. As frequency increases, the dynamic power of switching digital circuits linearly increases to a maximum of ≈ 19.8mW at 100MHz.

This is the combined leakage [26, 69] through all transistors within the device, particularly in high-speed digital circuits. The currents used within the AQAR to maintain quiescent operating points also contribute to this [162].

As the frequency increases, so does the power, until it takes 19.8mW at 100MHz. Over a range of 6MHz to 100MHz, the current draw is linear, as exemplified by a linear equation fitted to the measured current. Figure 5.1 matches the $P = fCV^2$ relationship that is expected for the dynamic power of digital circuits [29, 69]. The power of the system clock distribution can be reduced using low-power circuit techniques [69] or low voltage signalling.

Simulating a 64-SPAD channel (Section 4.7) with parasitic capacitances and the current draw of clocking transistors, yielded a current of 1.17mA at 100MHz. This gives a full device current of 18.7mA, close to the measured 16.5mA (19.8mW). The over estimation may be a product of distribution voltage drops, partial logic swings, or frequency variability [69].

The Receiver Power Consumption With Photon Flux

With the counters and clock disabled, the power consumption of the SPAD, Sub-Pixel and Pixel circuitry under optical intensity can be measured. Generally, as optical power increases, the increased count rate leads to higher dynamic power [69], similar to the clock network. In Figure 5.2, DVDD and SPADDVDD currents are plotted against optical power for active and passive quenching. At 10nW, the receiver draws a leakage current of 456μW from DVDD, for both quenching modes.

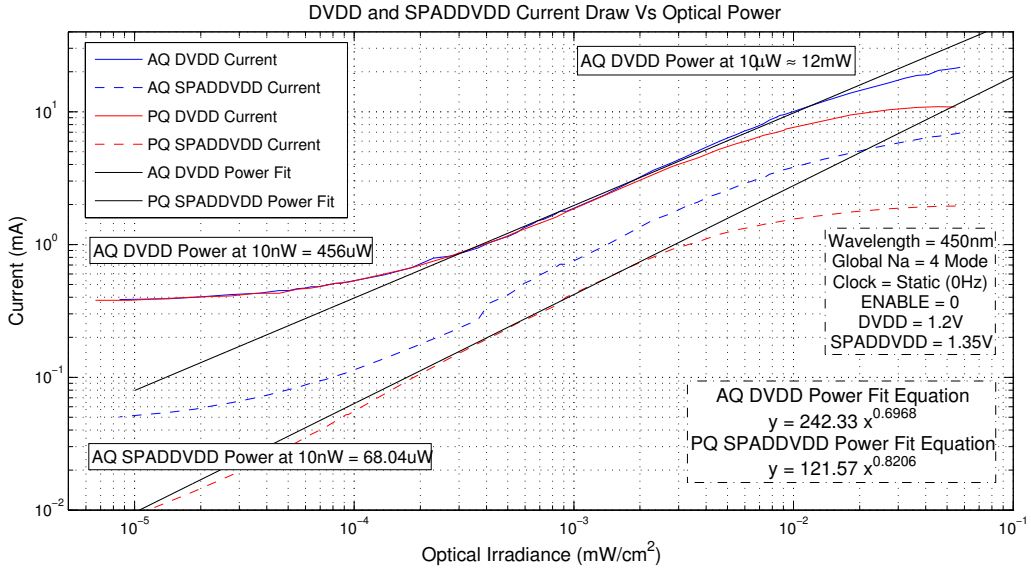


Figure 5.2: Full array SPAD and DVDD current consumption with optical input power. In both active and passive quenching modes the power is expected to increase linearly, however sub-linear current consumption is observed. At saturation ($\approx 10\mu$W) the array, in active quenching mode, draws 12mW from the DVDD supply.

The active quenching SPADDVDD power saturates at 68μW for low flux, suggesting limitation by leakage currents through transistor M3 (Section 4.4.3) [162]. The passive quenching SPADDVDD power draw continues towards the lower experimental limit. The minimum consumption of both supplies will scale with DCR, transistor dimensions and the number of SPADs. At high optical power ($\geq 10\mu$W) the electrical power draw begins to saturate for both DVDD and SPADDVDD. These are restricted by saturation of the achievable count rate, C_{max} [60, 163]. Passive quenching, deviates from the fitting curve, however this is due to paralysis and a lower maximum counting rate (Section 3.4) [60, 163, 164]. Between 0.1μW and 10μW, both supplies demonstrate monotonically increasing current consumption. This is expected to be linear following a $I = fCV$ relationship [69]. However by fitting power law curves of the form, $f(x) = bx^a$, deviation from linear assumptions can be assessed.

The gradient of DVDD (Figure 5.2), shows that the circuit draws less current than would be expected. In simulation, a Sub-Pixel with resistive supplies, also showed sub-linear currents as the pulse inter-arrival time decreased. This suggests that the active quenching circuit causes this behaviour. A parametric simulation over $DVDD = 1.1$V to 1.3V, at the maximum counting rate, showed a linear decrease in DVDD current with voltage, while the SPADDVDD current remained static. As the light level and switching rate increase, a voltage drop may be induced locally [29, 215], thereby reducing the current non-linearly.

The passive quenching circuits also show sub-linear behaviour. This could be due to variation in SPADDVDD or VBD with increased switching. It however has a fitting coefficient a, of 0.82, rather than the 0.67 value for active quenching.

For a communications receiver, this implies careful design of power supply tracks on the chip to reduce any local voltage variation [29, 69]. It also indicates that while a $P = fCV^2$ estimate for power consumption can be used for synchronous digital circuits [69], the relationship over estimates power draw of optically active SPAD circuits. This impacts the receiver power efficiency, giving a better than expected efficiency for a given count rate, but modifies the power efficiency non-linearly for different received power levels.

Discussion of Power Supply Testing Results

In this section, testing of the receiver power supplies has revealed two main conclusions.

- Firstly, clock distribution, while fitting conventional theory, is a large contributor to the power consumption. This is principally caused by physical distribution of clocked logic over the device. Power savings could be achieved using a number of conventional strategies such as lower voltages, differential clocks or low-capacitance routing, however three principal strategies may be appropriate:
    1. Moving synchronous digital processing outside of the optically active array would reduce the area involved, thereby reducing the distances that signals must traverse. This also increases the optical fill factor.
    2. By distributing a slower clock, power can be decreased linearly on long, capacitive routes. Clock multiplication at the point of use can then reduce the length and capacitance of routes operating at the higher clock rate.
    3. The current device uses pipelining and dual-edge clocking. This doubles the operational clock, as operations occur twice per period. Depending on the settling time of the logic, four operations per period could be achieved by distributing two lower frequency clocks with a 90 degree phase shift.

- Secondly, the breakdown of a SPAD is also a large contributor. This is a product of the photon events per second, the SPAD capacitance and the total bias voltage. Despite some deviation, the principals of the $P = fCV^2$ relationship can still be applied to the receiver. For a second generation, further investigation of the non-linear power draw is necessary, although the following three techniques could be used.
  1. Reduction of the total applied bias, although this requires process modification to retain the photon detection efficiency.
  2. Decreasing the SPAD capacitance through area and parasitic reduction. This is also known to benefit the after-pulsing probability.
  3. Minimising the number of detections required per symbol i.e. pushing the receiver closer to the quantum limit, or increasing the BER specification.

In the next section, the static behaviour of SPADs in the array are measured. This includes the dead time and dark count rate, both of which, depending on their magnitude, could affect later communication experiments.

5.2.2 Static SPAD Performance

In this section, the static performance parameters of the SPADs within the receiver front-end are assessed. This refers to parameters that should be independent of time or optical power. The section therefore measures the SPAD dead time using a set of test cells and the dark count rate using all SPADs within the array. While some variation is expected, and for the DCR in particular the literature points to both a time dependent effect and a wide range in mean count rate, here the aim is to measure long term average performance.

- For the dead time, this allows assessment of the differences between active and passive quenching, attempting to determine which would ideally be used for communications.
- For the dark count rate, this allows an indication of the noise floor for the ideally 100% on-off pulse from the μLED transmitters, along with if any SPADs should be disabled to improve the bit error rate.

The SPAD Dead Time and Dead Time Distribution

Dead times were obtained using a test block outside the main array (Figure 4.31). In Figure 5.3, the mean dead time, τ_d, is plotted against the control voltage DTCTRL. Both circuits exhibit an exponential increase as the PMOS devices in the current starved inverter chain or the quenching transistor become more resistive [26, 215]. Agreement with simulation indicates no departure from established theory [29, 49, 211] and little impact on the design of a SPAD-based receiver front-end. There is however a positive shift between simulation and measurement. This is due to additional parasitic capacitances, a test cell multiplexer and an unmatched ohmic connection to the oscilloscope [85].

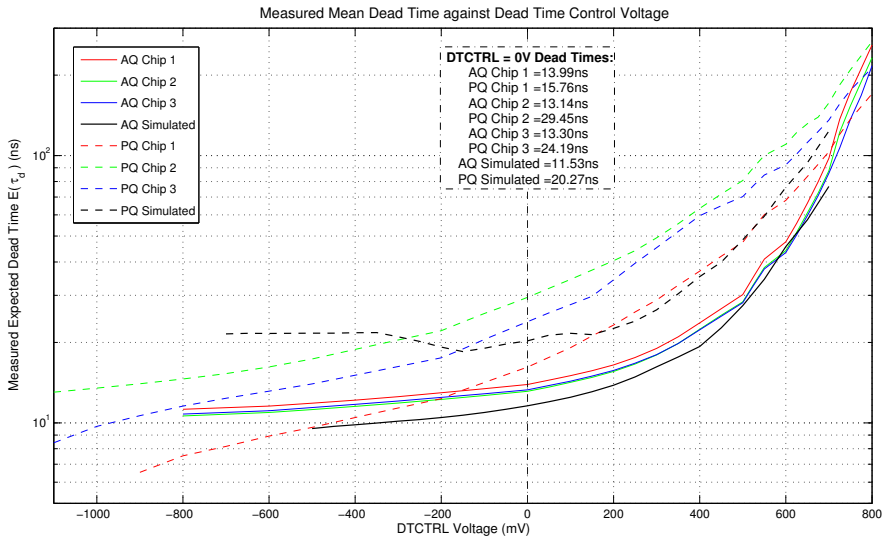


Figure 5.3: Mean active and passive SPAD dead times for different values of the control voltage, DTCTRL. Active quenching closely matches circuit simulation, however passive quenching departs from the simulation.

Investigating the dead time standard deviation, plotted for a number of dead time mean values, shows that active quenching obtains tighter control over the detection to detection dead time width (Figure 5.4). Further, Section 4.4.3 shows that active quenching obtains better circuit to circuit variation in the presence of process, power and matching variations. Passive quenching gives an order of magnitude larger standard deviation, despite a similar range and trend to the active circuit (Figure 5.3). There are two effects contributing here:

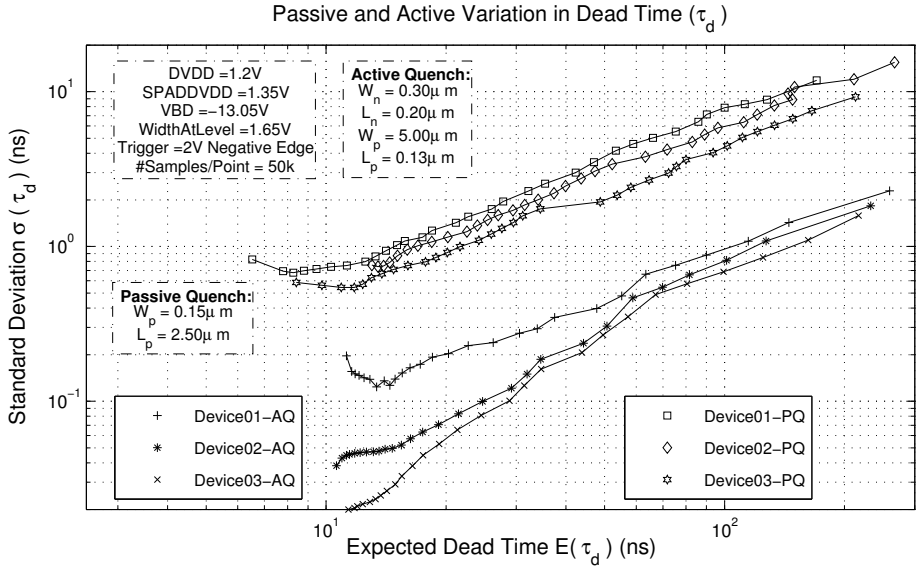


Figure 5.4: Active and passive SPAD dead time standard deviation, $\sigma(\tau_d)$, over different mean dead times. Active quenching obtains tight dead time distributions.

- Marked differences between active and passive quenching are observed (Figure 5.3 and Figure 5.4). This is due to the different circuit implementations and recharging methods. Passive circuits are able to obtain a low dead time with a small area circuit, depending on the resistance (R_{on}) of the quenching PMOS [49]. In contrast, an active circuit may be better for communication receivers, due to a known, but similarly low, designed hold-off period and tighter variability. This is important to ensure performance is the same from one receiver to another. In applications such as VLC using multiple diodes, variation in the dead time may prevent a receiver from obtaining the correct number of detections per symbol (Equation (3.7) and Equation (3.8)) (for example receiving 18 photons rather than 20, Section 3.3.1) and therefore may increase the BER.
- Variability in the dead time may increase uncertainty in the photons received, particularly within short symbols. For example, for a 10ns symbol, an exact dead time of 2ns gives a maximum of 5 photon counts per SPAD per symbol (Figure 5.5). If however the dead time varies from one detection to another, with a 2ns mean and a 1ns standard deviation, the mean number of counts may change only slightly. The standard deviation of the received counts will however change significantly, artificially broadening the received distributions, thus increasing the BER (Section 3.2).

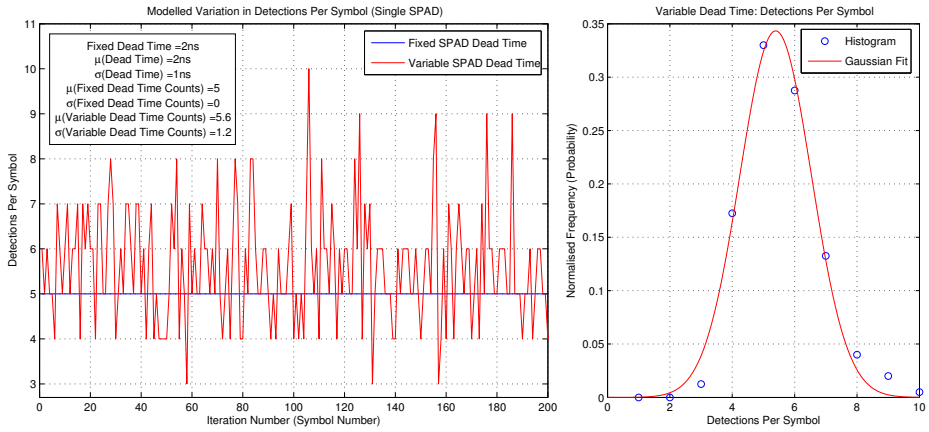


Figure 5.5: Modelling results investigating increased detection variation if the dead time variation is a significant proportion of the symbol. Here a 10ns symbol, with a SPAD dead time of 2ns mean and 1ns standard deviation, as compared to a zero variability model.

The SPAD Dark Count Rates

The DCR of the SPADs within the main array have been measured to give an indication of the expected noise floor when using an assumed perfect OOK modulated signal (i.e. 100% extinction ratio) [23, 24] (Section 2.3.3). The analysis of the receiver will assume this perfect modulation, (a product of the CMOS digital driver circuitry [96]), however the eye opening penalty (Equation (3.17)) can be used to account for non-ideal μLED modulation [2, 3, 14, 36].

In Figure 5.6, the sorted DCR distributions of five chips are presented. Each trace includes 1024 SPADs, with each SPAD measured 500 times with integrations of 1ms and 0.1ms. The total DCR has been verified using the global array summation output. The DCR of a single SPAD ideally follows Poisson statistics [26, 49], however the mean DCR of SPADs over an array, often shows skewed distributions similar to Figure 5.6 [53, 55, 56]. The array mean DCR was 7.27kHz, however as shown by the histogram (Figure 5.6.Inset) for *Chip*01 (plotted on a logarithmic x-axis), the array median of 2.5kHz is more appropriate, given the distribution skew [55, 56, 76, 225]. While the presented results use the active quenching circuit, the passive case showed negligible departure when operating at the same dead time.

The small number of high DCR devices, is caused by the presence of silicon traps [26, 49, 76]. This leads the high DCR devices to exhibit the combined effect of thermal generation, increased generation from mid band gap energy levels (trap assisted tunnelling) and native band to band tunnelling [26, 126, 128, 143, 144, 172].

Marked on Figure 5.6, the mean (μ), mean plus 1σ (one standard deviation) and $\mu + 2\sigma$ are indicated. The mean is used here, despite the skew, to ensure consistency when using the analytically calculable central moments, μ and σ [116] and discussing the spread of a distribution. In comparison, while the median is preferable to discuss distribution centre location

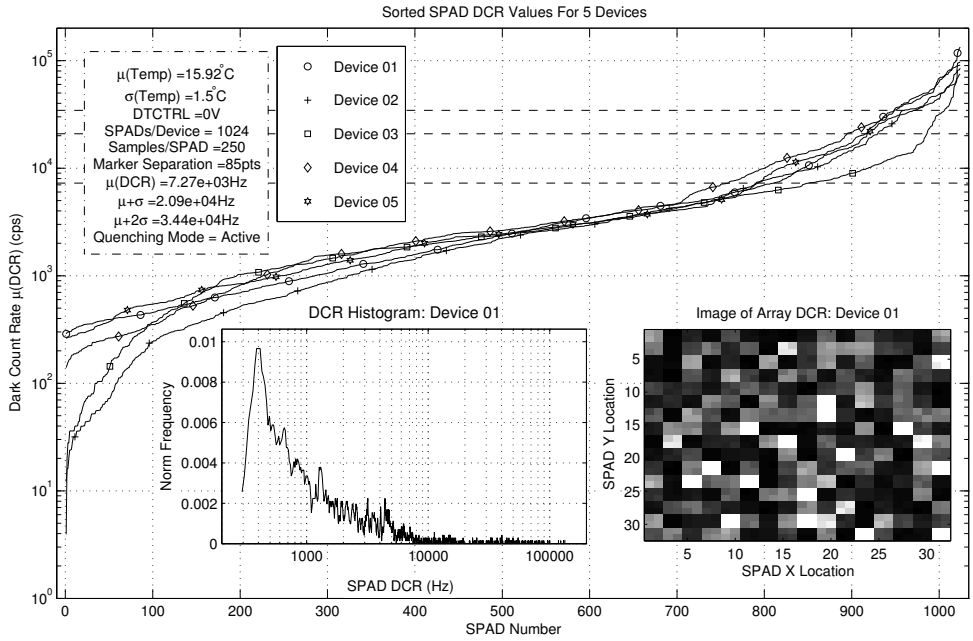


Figure 5.6: Sorted SPAD mean dark count rates for a selection of devices (at 16°C). Shown inset are the log-scale normalised histogram and the 2D spatial mapping for Device 01.

in the presence of outliers, it is neither a moment, central moment or directly related to the variance and standard deviation of a distribution. The mean and median are used simply to help provide a comparison to the bulk of the SPADs within the array, and to SPAD arrays in the literature, however these values can also define which SPADs will be turned off, and which will be assigned a TDM mode (Section 5.2.3). For example, SPADs with a DCR higher than 34kHz can be turned off using the $N_a = 3$ mode (Section 4.5) [70, 77, 208]. SPADs with a DCR between 20 and 34kHz can be assigned the $N_a = 1$ TDM mode which reduces the impact of the high DCR SPAD (Section 4.5). SPADs just above the mean can be assigned the $N_a = 2$ TDM mode, reducing the impact of the high DCR device, but allowing it to contribute slightly more than with the $N_a = 1$ mode. SPADs with a DCR around or below the mean can be given a full weighting in the signal addition scheme, improving the SNR.

If the skew of the DCR distribution is too severe, many SPADs will need to be removed from the summation architecture to prevent the optical power from being increased to ensure accurate distinction between a transmitted 'one' and 'zero' Section 3.3.5. However turning off SPADs, as per the $N_a = 3$ mode, will reduce the overall sensitivity of the SPAD-based receiver front-end. The presence of skew within the DCR distributions of many SPAD arrays [53, 55, 56], indicates that the yield of 'low-noise' SPADs, and indeed the continued reduction of average array DCRs, is needed to prevent undue impact on communications.

Dark Count Rates Comparison with Prior Art

The SPAD used was demonstrated and characterised [143, 163], with a version of the same active quenching circuit [162]. In comparison to the 7.3kHz mean DCR, Richardson et al [143] showed a mean DCR of 20 to 30Hz at 16°C and 1.4V excess bias ($20log_{10}(\frac{25}{7300}) \approx -50$dB noise reduction). The similarity between array DCRs when operated in passive and active quenching, indicates that the immediate circuitry is not a factor that elevates the DCR above that reported. Despite this, high speed logic may increase the DCR if substrate currents are present [56, 149, 150]. Prior results [143], point to a receiver dynamic range [163] that can be increased above that presented here. Cooling to -10°C, [143], showed a mean DCR of 1.4Hz, giving a low total noise floor for devices that do not contain traps.

The cooling used for many commercial optical receivers [13, 14, 36, 62, 63, 226], prompts expected DCR values lower than 0.5Hz [53, 143]. As SPAD noise is discretized, a low DCR such as this, demonstrates the possibility of an ideal receiver (Section 2.3.7), with almost zero continuous time noise, single photon sensitivity, and recently, the combination of low room-temperature DCR with large areas [144].

When using ideal, 100% extinction ratio OOK modulation (Section 2.3.3), the DCR and after-pulsing noise rates constitute the *zero* distribution, N_0 (Section 3.2). This can be obtained with slow ($\leq$ 10MHz) modulation with the Hypix μLED transmitters [23, 24]. Any closure of the transmitted eye diagram, via the effective reduction in modulation depth at higher μLED modulation frequencies ($\geq$ 10MHz), can be quantified using the data eye opening receiver sensitivity power penalty, PP_{EOP} (Section 3.3.5) [1, 14, 36].

Pixel TDM Mode Dark Count Rate Mitigation

The time division multiplexing circuit reduces the overall Pixel group DCR, denoted as DCR_{px} (Equation (4.5)). If all SPADs have equal rates, then the output DCR in $N_a = 1$ will be $\frac{1}{4}$ of the total DCR. Here, DCR_{px} (Section 4.5.9) is experimentally proven using a Pixel within the test column. This has all of its internal and output nodes buffered to a test cell multiplexer (Figure 4.31). It is possible to measure the DCR of each SPAD and to observe signals such as $SPAD1ENB$ or the state machine outputs, $Q0$ and $Q1$. Table 5.1 gives the mean DCR values along with the expected inter-arrival time, $E(t)$, between successive dark counts. The simple addition total DCR that would be expected is also given.

SPAD	DCR(Hz)	#Samples	$E(t)$ (s)
1	629.00	15.5k	$1.58\text{x}10^{-3}$
2	1870.5	45.1k	$0.53\text{x}10^{-3}$
3	206.20	22.6k	$4.85\text{x}10^{-3}$
4	187.20	23.3k	$5.34\text{x}10^{-3}$
Total	2955.9	-	$3.38\text{x}10^{-4}$

Table 5.1: Test Pixel Internal SPAD Dark Count Rates

N_a	DCR_{px}(Hz)	Samples	Theory DCR_{px} (Hz)	ENB1 Width (s)	ENB4 Width
4	2463.5	224.4k	2955.9	-	-
2	703.0	102.1k	698.90	367.7μ	2.49m
1	342.4	33.38k	346.50	1.59m	5.11m

Table 5.2: Test Pixel DCR_{px} for different N_a modes. Note the departure between 2463.5 and 2955.9 ($\Delta = 492.4$), this is due to combined measurement errors and the assumption that $P(TC) = 0$.

The theoretical model developed previously (Equation (4.5)), can estimate the Pixel DCR given the TDM mode (Table 5.2). This is seen by the close agreement between the values of 2463 and 2996, 703 and 699, and 342 and 346, for $N_a = 4$, $N_a = 2$ and $N_a = 1$ respectively. The theoretical DCR_{px}, assumes a zero probability of temporal coincidence (Section 4.5.8) and for the $N_a = 4$ theoretical value, the multiple SPAD measurements each have corresponding measurement errors. Both contribute to the error between the measured and calculated values.

The equation can be simplified to Equation (4.3) for $N_a = 4$, however the two multiplexed modes show a reduction in group DCR, dependent on the constituent SPAD noise rates. If these are equal within the Pixel, then DCR_{px} will reduce to the simple case of $\frac{DCR_{total}}{C_L}$. Equation (4.5) can be used for any number of SPADs within a group, and by using the gamma function $\Gamma(x) = (N-1)!$ [115, 116], can be used to estimate Pixel DCR for the problematic to implement TDM modes such as $N_a = 3$.

Conclusions from Static SPAD Performance Experiments

In the above section, the SPAD dead time and the SPAD dark count rates have been explored. The dead time is critical to communications, as it fundamentally limits the number of photons that can be detected, and thus the achievable bit error rate. Likewise, the dead time may be difficult to scale, (to sub-1ns levels), if higher data rates and lower BERs are required.

The DCR is higher than previous implementations of the SPAD, but below the worst case used within the specifications flow. As the μLED transmitters use digital *on/off* CMOS drivers, 100% extinction ratio modulation is assumed. The sum of SPAD DCRs within a receiver therefore defines the noise floor or N_0 distribution, although finite extinction ratio transmitters and the eye opening power penalty are likely to dominate practical receiver performance.

Both the dead time and SPAD array DCR, can be optimised in future generations, however:

- The dead time is constrained by the diode capacitance, quenching resistance and the after-pulsing probability (Section 2.7.3, Section 3.3.6 and Section 4.4.3). A fundamental limit may also exist [76], preventing reduction necessary for continued data rate scaling. Despite this, future silicon implementations can optimise the system for the minimum dead time that satisfies all specifications. This may require changes in the circuit but may also require modification in the area and number of SPADs used within a receiver.
- Some scope exists for reduction, through design, of the dark count rate. The foundry model of planar silicon technologies however restricts the option of silicon dopant modification. Low DCR structures can be used, however this requires assessment of the trade-offs at a receiver system level, assessment of DCR variability between manufacturing runs and optimisation of SPAD DCR yield. To achieve this, further experimental and theoretical work centering on the causes of the DCR is needed.

5.2.3 Static Array Performance with Reconfigurability

In this section, the photon transfer curve of the SPAD array is measured. This uses a slow sweep of the optical intensity, measuring long-term photon count distributions for each point. By doing so the sensitivity of the array can be measured without time dependent effects. For communications, this allows assessment of any non-linear behaviour such as SPAD paralysis, but also checks for agreement with previous theory, modelling and experimentally observed behaviour.

Knowing the spatial distribution and yield of noisy SPADs, the TDM and disable functionality is then used to remove these devices from the signal chain. Removal of a SPAD reduces the optical fill factor, thus increasing the power penalty, PP_{FF}. However, again assuming a 100% extinction ratio due to the μLED CMOS drivers, this may be a useful strategy. In particular, this may optimise the SNR when a) the overall DCR is very high, as per initial device fabrication runs, or b) the incident optical power, and thus the distance between N_0 and N_1, is low.

The Photon Transfer Curve Over Optical Irradiance and Pixel Modes

With a receiver designed to operate over a range of optical powers [14], the photon transfer curve (PTC) of the SPAD array must be measured. Presented in Figure 5.7, this is the transfer function between input optical power and single-LSB changes (single photon counts) at the array output [81]. This plots the PTC over both active and passive quenching, and over the three pixel modes.

PTC: Maximum Counting Rates

As with the SPAD literature [49, 163] and numerical modelling, not presented here, the $\frac{1}{e}$ factor that limits the maximum count rate, C_{max}, in passively quenched SPADs, is clearly seen at $\approx 10^{-4}$W/cm^2 [60]. Above 5x10^{-4}W/cm^2 count paralysis is observed [60, 163], whereas for active quenching, the count rate correctly saturates at the theoretical value of $\frac{1024}{\tau_d}$ [163].

PTC: Central Region

While active quenching shows a gradient of 1 throughout the central region [163], Figure 5.7 shows that passive quenching follows a slight sub-linear relationship with increasing flux. This is indicative of detector paralysis [163], where a photon incident while the SPAD is recharging, is able to extend τ_d by creating a smaller, lower charge avalanche.

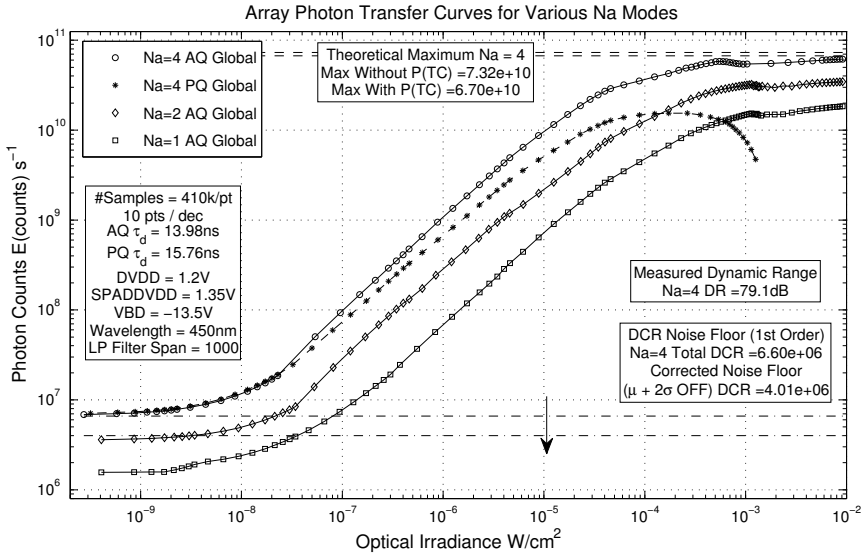


Figure 5.7: The photon transfer curves for globally set N_a values of 1, 2 and 4. Saturation points at the DCR noise floor and the active and passive maximum count levels fits theoretical values, while passive quenching effects such as paralysis are also observed. For the DCR correction, 57 pixels, that had noise levels greater than $\mu + 2\sigma$ were turned off, with a DCR noise reduction of 4dB.

The sub-linear curve and paralysis, may introduce distortion into the received signal [14]. This may not impact OOK modulated signals if:

- the *zero* and *one* distribution distance is large, as with a low target bit error rate such as $1\text{x}10^{-9}$ [1, 14, 36], or,
- if the SPAD array is designed for a dead time lower than that calculated previously.

However in multi-level pulse amplitude modulation, continuous analogue modulation or when the dead time cannot be scaled lower than the dead time calculated during specification, it may constitute an increase in the BER [1].

PTC: Array Noise Correction

Using the DCR map, devices with a noise greater than the mean plus two standard deviations ($\geq \mu + 2\sigma$) were turned off. This was 57 pixels, highlighting that SPADs must be turned off separately rather than the per pixel method that is implemented (Section 4.5), although some high noise rate SPAD grouping was observed. This has reduced the lower saturation level to 4MHz giving a saving of 4.3dB and will be of lower importance as the yield of SPAD devices increases and the noise rates decrease towards sub-100Hz levels [53, 144]. This 4MHz count level equates, using a linear fit to the central region of the $N_a = 4$ PTC, to 3.8nW/cm^2($\pm \approx$ 10%) (or -54.2dBm) of optical power, showing that the noise floor is above the quantum limit (Table 3.4). In [221], turning off 5% of the devices with the highest noise gave a 3dB reduction in the noise floor.

PTC: Dynamic Range

The dark count rate has reduced the dynamic range from previous implementations with this p-n junction and SPAD geometry [163]. DCR values in the 20–100Hz range were previously reported, giving a total dynamic range greater than 170dB at room temperature [143, 163, 172]. The dynamic range of the SPAD array in the $N_a = 4$ mode is 148.32dB (Equation (5.6)) [163]. This may reduce the performance of receiver.

PTC: Action of TDM

The three tested TDM modes, $N_a = 4$, $N_a = 2$ and $N_a = 1$ [178], each affect the photon transfer curve (Figure 5.7). Conceptually, the modes should give weightings of $1, 0.5, 0.25$ due to their respective affects on the fill factor (Section 3.4.2) [178]. These weightings hold at the two saturation levels. At $0.5\text{x}10^{-9}\text{W/cm}^2$, the ratio between the $N_a = 4$ and $N_a = 1$ mode is 4.22, while at $1\text{x}10^{-3}\text{W/cm}^2$, this ratio is 3.63.

The ratio between the TDM modes departs from these conceptual weightings due to the combined action of the SPAD transfer curve and the respective TDM mode response curves (Section 3.4.3) [178]. In Figure 5.7, the ratio between the $N_a = 4$ and $N_a = 1$ modes changes, increasing above a factor of 10 at the centre of the PTC. This action can be investigated using modelling of the different TDM modes in Section 3.4.3.

The fundamental number of devices within a pixel grouping gives a factor of $N_{TDM} = 4$. The ratio between the modelled maximum counting rates, F_{max}, of Figure 3.4 at the centre of the modelled PTC, is 2.2. This gives a combined ratio of 8.8, matched by the 8.8 ratio between the modelled $N_a = 4$ and $N_a = 1$ *SPAD+TDM* responses (Figure 3.4). The ratio between experimental values for the $N_a = 4$ and $N_a = 1$, slightly exceeds that of the numerical model. However the action of the temporal coincidence probability (Equation (4.2) and Figure 4.14), creates a further non-linearity over the optical range. For the $N_a = 1$ mode, the unresponsive switching time between SPADs also creates a slight sensitivity loss, moving the modelled ratio of 8.8 closer to the measured ratio of 10.

PTC: Affect on Communications

For communications, the photon transfer curve indicates trade-offs between the photon counting sensitivity and the noise benefit of time division multiplexing (Figure 5.7). For optimal response and low noise, only pixels that contain high DCR SPADs should be assigned the lossy TDM modes, with the highest noise devices turned off. Combining different pixel modes depending on the noise rates may alleviate the receiver DCR noise floor and temporal noise caused by after-pulses, when using OOK with 100% extinction ratios (Section 2.3.3). This is explored in the next experiment, using the *design for test* functionality included on the chip.

Array Signal-to-Noise Ratio (SNR)

Turning to the achievable SPAD-array signal-to-noise ratio [2, 3] (not yet including ISI etc), the level of noise in the system directly impacts the receiver bit error rate [1, 14, 36, 62]. The SNR defines the ratio of the signal to the noise amplitudes at the sampling instant [1, 2, 35, 36]. Here however the receiver uses an integration (Equation (3.4)) [35, 73, 113]. As the BER looks at the probability that a transmitted zero is incorrectly received as a one [1], the integral of the noise must be below the OOK threshold, N_{th} [2, 3, 103].

The SNR can be improved by removing known high noise devices from the summation tree [70, 77, 208]. Here the SNR is explored as the ratio of the peak achievable count rate to the variation of the DCR, after-pulsing and ambient light noise rates [176]. The maximum count rate includes the effective fill factor reduction, as it is the maximum count rate of the *currently enabled* rather than total SPADs. The absolute DCR represents an unwanted signal, indistinguishable from photon counts, which acts to reduce the available counting time [52, 59, 75, 141, 165, 176]. However the *noise*, being the variation of that level, is also important [1, 49, 163, 176]. The SNR is defined here using Equation (5.1) [176].

$$SNR = \frac{Signal}{Noise} = \frac{Signal}{\sqrt{\sigma^2(DCR) + \sigma^2(AfterPulse) + \sigma^2(Ambient)}} \tag{5.1}$$

For the readout circuitry used, the after-pulsing of SPADs within the array could not be measured. In Figure 5.8, the sorted DCR distribution of Chip01 (Figure 5.6) is used. However it is assumed that a) after-pulsing is much lower than the DCR, b) that no ambient light is present, and c) the receiver-only performance is being assessed, thereby assuming a 100% transmitter extinction ratio [67]. This may be an ideal case, however it allows investigation of the SPAD turn-off functionality [70, 77, 208] using measured DCRs without requiring measurement of the after-pulsing of each SPAD.

The SNR is assessed as contributions from the ordered set of Chip01 are sequentially added (Equation (5.2)). SNR_i represents the noise of the sum of SPADs currently turned on, to the total theoretical maximum, C_{max}, for that set. This is expressed by the iterative SNR, where $i \in (1, 2, \cdots, 1024)$ is the number connected to the output. Calculating the SNR in this manner for different numbers of *turned on* SPADs, allows which and how many SPADs from the ordered set should be used. A similar method could be used with regard to photo response non-uniformity [81], selectively removing low-PDE SPADs to maximise SNR.

With the dead time based on the number of detections required and the data rate (Section 3.3.2), a SPAD is assumed to be fully utilised when obtaining symbol detections. Further, as mentioned (Section 2.7.3), it may not be possible to scale the dead time with data rate to allow operation within the linear PTC region. As such, the maximum signal, C_{max}, is used. Figure 5.8 plots

the SNR (*blue trace*) against the theoretical total output count rate. For completeness, the population turned on is also given. Equation (5.2) uses $20\log_{10}$ as both the maximum number of counts per second and the number of dark counts per second, are amplitudes rather than signal powers.

$$SNR_i = 20\log_{10}\left[\left(\sum_{j=1}^{i}\left(\frac{1}{\tau_d}\right)_j\right)\left(\sum_{j=1}^{i} DCR_j\right)^{-\frac{1}{2}}\right] \tag{5.2}$$

The SNR increases as low noise SPADs are turned on, due to the low dead time, (assumed to be equal for all SPADs), increasing the theoretical maximum counting rate more rapidly than the noise of the sum of low-DCR SPADs. The SNR decreases as the high DCR devices are sequentially turned on. However turning off the high noise devices still allows a maximum count rate in excess of $2\text{x}10^{10}$cps. The maximum SNR of $157.1dB$ is reached at a count rate of $4\text{x}10^{10}$Hz. This turns off an unacceptable 618 of 1024 SPADs, severely reducing the fill factor. With SPAD noise yield increasing past 95% [53, 56, 181, 216], this unacceptable number of turned off devices must be reduced to maximise the fill factor and therefore the counting rates. Other SPAD arrays have shown poorer low noise yields between 60–80% [154, 217], increasing the number of turned off SPADs needed.

For optimisation, the proportion of devices that must be turned off, to maximise both SNR and C_{max} needs to be estimated. To prevent a low fill factor, the higher count rate intersection with an SNR specification should be chosen. This specification can be derived from the required error and symbol rates [1, 14, 36]. Assuming an example required $SNR \geq 153dB$, the maximum SNR value could be taken. However this example specification can also be met at the higher count rate of $8.8\text{x}10^{10}$, operating at the extreme right of Figure 5.8, with SNR of $154.7dB$ and only 140 SPADs turned off. This analysis, has only investigated the effect of turning SPADs off. Using the N_a modes to selectively weight Pixels, may offer other advantages. This however must be investigated with respect to fill factor, SNR, count rate, after-pulsing and the impulse and step responses.

Conclusions from Static Array Performance and Reconfiguration Experiments

In this section, the photon transfer curve has been measured and the array signal to noise ratio has been investigated. The PTC directly impacts the sensitivity of a communications receiver, and depending on how the receiver is used, may present issues for the recovery of accurate data steams. The saturation or paralysis behaviour, indicates that the dead time should be designed to be lower than that estimated in initial modelling, although as mentioned this may be difficult to scale for higher data rates.

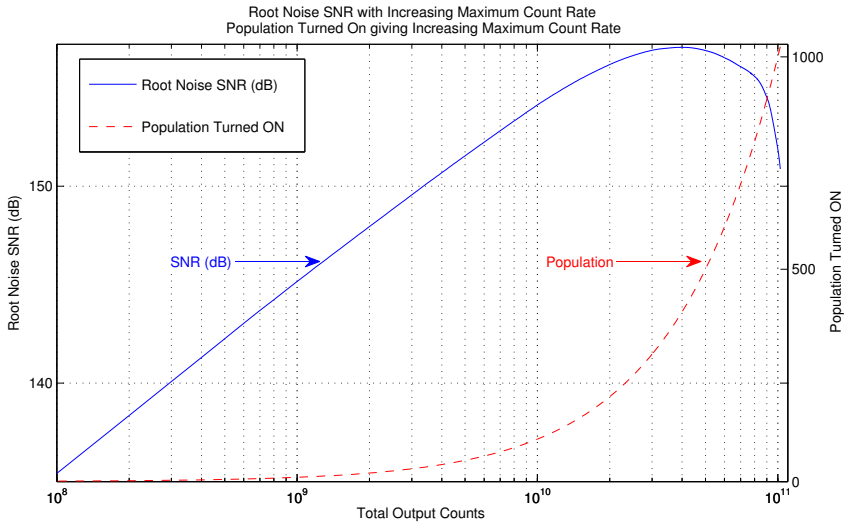


Figure 5.8: The SNR (dB) and the turned on population of Chip 01 with respect to the theoretical maximum count rate. A constant input optical power is assumed, with each SPAD operating at its maximum counting rate.

The *design for test* functionality included on chip allows the signal-to-noise ratio to be optimised. This is particularly useful when the likely SPAD DCR is unknown prior to device fabrication, as was the case with this design. Specifically, such functionality may be critical if the mean DCR is particularly high and if the yield is particularly poor (as demonstrated by initial fabrication wafers). The outcome of this for communications, acts to highlight the advantage of SPAD integration with CMOS digital logic. Future work could, if found to be advantageous, automatically optimise a receiver chip to optimise performance in the short term. Meanwhile, SPAD and process research on lowering the mean DCR and increasing the yield can be performed in the longer term.

5.2.4 Time Dependent and Other SPAD Performance

So far, experiments have concentrated on the time-invariant performance characteristics. In this section, temporal characterisation is begun prior to further time-based and communication-based experiments in Chapter 6. The section first discusses the after-pulsing probability, as this may contribute to elevated receiver bit error rates.

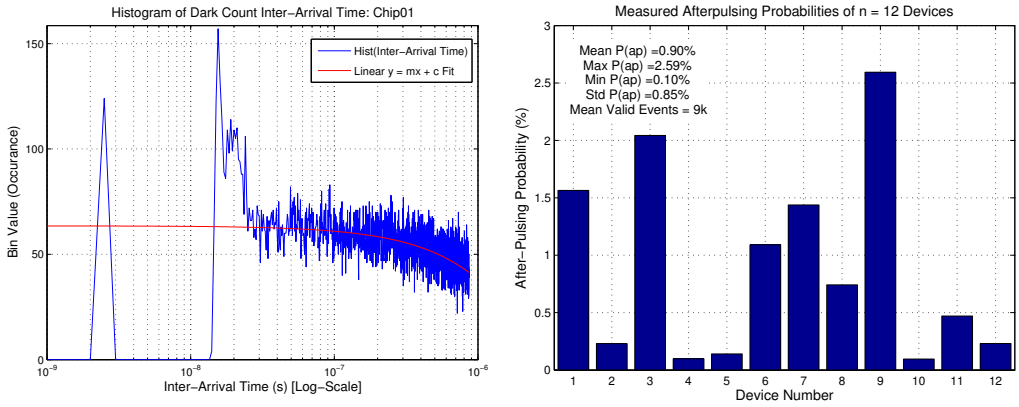
While not strictly time dependent, the later portion of this section briefly discusses other measurements that were performed. This includes the photo-response non-uniformity which acts to spatially change the photon detection efficiency over the optically active area, thus modifying the receiver sensitivity. SPAD to SPAD cross talk is also discussed, however the role this has on communication performance is difficult to ascertain.

Measuring The After-Pulsing Probability

An understanding of after-pulsing is required as it could act to create a form of inter-symbol interference. The after-pulsing was measured in a low ambient temperature, dark environment with two methods used over a selection of devices. In the first method, the histogram of the dark count inter-arrival time was used to observe after-pulses over a 1μs window. After pulsing, is observed as an increase above normal exponential inter-arrival time, and is only visible, for one particular SPAD, in the first 10–20ns after the SPAD has recharged, as shown between $1.5\text{x}10^8$ and $3\text{x}10^8$ in Figure 5.9.a. This, along with the autocorrelation method [75, 128, 172], is useful when assessing the influence of the dead time [75, 163].

In Figure 5.9.b, after-pulsing probabilities are presented for $n = 12$ SPADs using a second method. This uses a correlated counting technique, counting secondary events that again occur within a 1μs window after an initial count. This gives an estimated mean after-pulsing rate of 0.9%, and a standard deviation of 0.85%. A minimum of 0.1% was observed for devices four and ten, while device nine showed the maximum after-pulsing probability of 2.6%.

Due to the construction of the main array, only after-pulsing measurements from a limited number of SPADs could be performed. These used SPADs positioned within the test cells towards the edge of the array (Figure 4.31), although care was taken to ensure the same circuitry was used for quenching the SPAD. The small number of tested SPADs limits the conclusions that can be drawn numerically from the overall mean and standard deviation [116]. However qualitatively, a low after-pulsing distribution may be caused by SPADs without silicon traps [76], while the few above average after-pulsing SPADs may indicate a separate distribution that includes one or more traps.



(a) Inter-arrival histogram for chip01 **(b)** Measured after-pulsing for 12 devices

Figure 5.9: Measured inter-arrival histogram of the dark count rate of a SPAD on Chip 01, and the measured after-pulsing probabilities of 12 separate SPADs.

Pixel TDM Mode After-Pulsing Mitigation

Test cells can be used to investigate the affect of N_a mode on after-pulsing. Table 5.3, gives the after-pulsing rates of four SPADs within a Pixel. To avoid temporal co-incidence, $P(TC)$ (Section 4.5.8), the after-pulsing rates were measured directly using buffered versions of their full width pulse. For validity a large number of pulses were captured, however even with the highest after-pulsing SPAD, only $41.6k$ after-pulses were obtained (Table 5.3).

SPAD	P(ap)	Total #Counts	#APs	DCR (Hz)	$P(E\|DCR_n)$
1	2.6%	1.6M	41.6k	630	0.2
2	0.1%	4.3M	4.20k	1871	0.6
3	0.5%	3.1M	14.8k	205	0.1
4	0.2%	4.0M	9.16k	187	0.1

Table 5.3: Test Pixel Internal SPAD After-Pulsing Rates

In Figure 5.10, the normalised natural logarithm of the inter-arrival histogram is presented for each of the SPADs. In this figure, the time scale is such that the straight line of the DCR inter-arrival, is not visible. This figure only shows the after-pulsing straight line and transition to the lower gradient DCR line. This implies that the straight line for $SPAD_3$ is after-pulsing only, until $\approx 100ns$, when it becomes flat. Due to the windowing technique (1μs), the likelihood of a dark count within this time is low ($\approx 2\text{x}10^{-2}\%$), hence the asymptotic drop towards zero probability at higher pulse inter-arrival times.

In Figure 5.10, each SPAD exhibits a linear region before transitioning towards the linear line expected for the dark count inter-arrival time. $SPAD_1$ has a different after-pulse emission time. This may indicate a second trap within the multiplication region. Overall, for most of the SPADs, the after-pulsing contributes a temporal noise component over an approximate 250ns

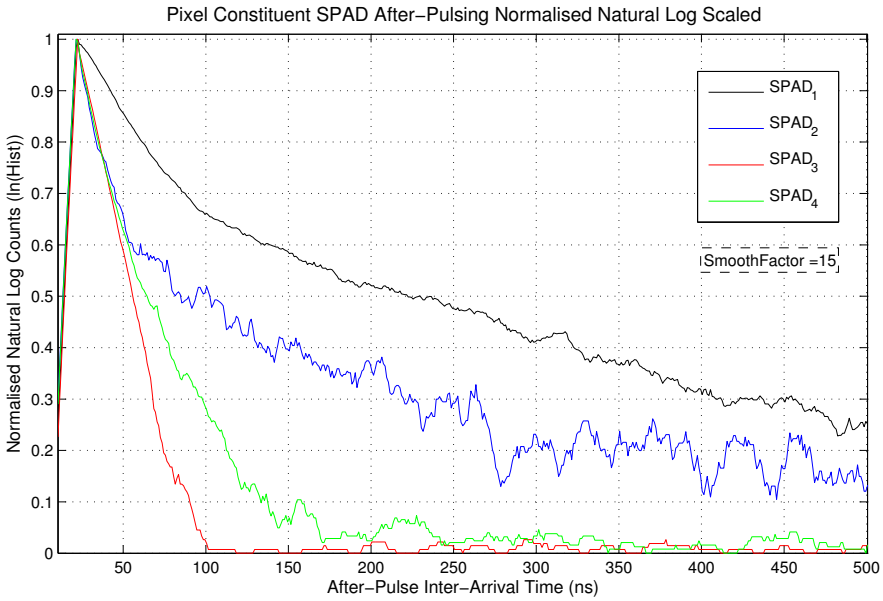


Figure 5.10: After-pulse inter-arrival times for the four SPADs within the test Pixel. The natural log of count inter-arrival time renders 'n' distinct straight lines. The principal line, due to the log of the exponential inter-arrival of dark counts, and a second or third line for the after-pulsing inter-arrival time.

time scale (approximately 10–25bits at 100Mb/s). Increasing the dead time can reduce the influence of after-pulsing [176], however this also reduces the count rate, and therefore the detections per symbol.

Table 5.4, presents the measured and theoretical $P(ap)$ over three modes. The theoretical model, Equation (4.8), estimates the after-pulsing values with reasonable accuracy given measurement error (0.806% vs. 0.792%). The pixel after-pulsing output, $P(ap)_{px}$, of $N_a = 4$, is not the simple summation of the individual SPAD after-pulsing rates, which would be $\approx 3.76\%$. Instead $P(ap)_{px}$ has been dominated by the relative DCR of $SPAD_2$ (Table 5.3).

Likewise $N_a = 1$ and $N_a = 2$, show a decrease in $P(ap)_{px}$, improving inter-symbol interference issues. As $SPAD_2$ shows, (Table 5.3), a high DCR device may have low $P(ap)$, however this is uncommon and will be dependent on if a trap is present. For this SPAD, the DCR is much lower than the mean DCR ($\approx$ 1870 vs. $\approx$ 7250cps), and it has a below average after-pulsing probability ($\approx 0.1\%$ vs. $\approx 0.9\%$). As such, this SPAD may not contain a trap, but create dark and after-pulse counts through other mechanisms [26, 52, 59, 76, 123, 143].

N_a	$P(ap)_{px}$	Total #Counts	#APs	Theory $P(ap)_{px}$
4	0.806%	5.56M	44.8k	0.792%
2	0.15%	5.30M	7904	0.150%
1	0.032%	3.69M	1224	0.040%

Table 5.4: Test Pixel $P(ap)_{px}$ for different N_a modes.

For Table 5.4, the values will include measurement error for the constituent SPAD after-pulsing rates, rounding errors in calculation and the assumption of a finite fixed temporal co-incidence probability. $P(TC)$ is estimated for $N_a = 4$, using the ratio, $P(TC) \approx \frac{Sum\ DCR-Measured\ N_a=4\ DCR}{Sum\ DCR}$. The after-pulsing probability of the Pixel output under the different N_a modes will also include measurement error, as will the DCR values previously obtained (Table 5.3).

In Figure 5.11, the after-pulse inter-arrival histogram is shown for different N_a modes. The normalised natural log $N_a = 4$ histogram, shows a similar shape to those of the constituent SPADs (Figure 5.10). Here, each has been normalised to the same maximum value. Any variation in the straight lines obtained when taking the natural log of the exponential inter-arrival time, indicates modulation in that inter-arrival time.

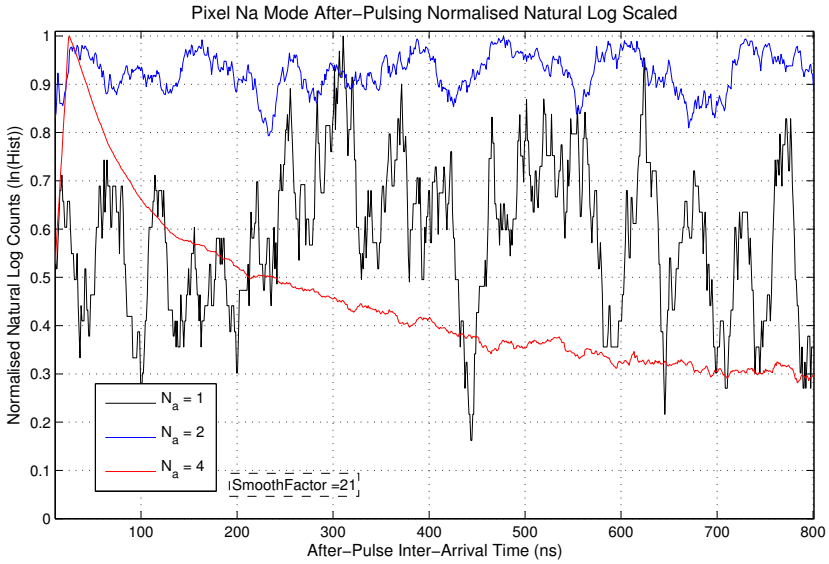


Figure 5.11: After-pulse inter-arrival times for the three Pixel Na TDM modes. The $N_a = 4$ mode shows a reduction in the number of after-pulses similar to measurements of the individual SPADs. The two TDM modes however gate the SPADs at a frequency dictated by the counts per loop and the relative dark count rates.

For $N_a = 1$, the initial peak has been removed, however the inter-arrival has become fragmented, shown by the rapid oscillation in $P(ap)_{px}$. This is expected given the time gating effect caused by the pixel state machine. $N_a = 2$, shows a similar lack of initial peak, however the modulation in inter-arrival probability has been reduced in line with the reduced frequency of state machine iteration. The non-normalised data shows a significant reduction in the overall levels of after-pulsing in each mode. This is in agreement with the numerical values within Table 5.4.

Based upon the analytical model (Equation (4.8)) that has been matched to experimental values in the section above (Table 5.4), as the frequency of state machine iteration increases when the optical flux increases, the overall pixel after-pulsing probability may be suppressed further. Four factors may become involved:

- An increase in the number of avalanches increases the number and periodically refreshes carriers within trapping centres,
- A decrease in the overall enable time, T_{enb} that is available for an after-pulse to create an avalanche will result as photon arrivals remove each SPAD for a period, τ_d,
- An increase in the speed of TDM state machine iteration, will gate the after-pulse lifetime into smaller separate enable periods. As each slice is short, the probability of an event due to an after-pulse will decrease, while the probability an event is caused by a photon increases with the incident optical power.
- Uniform optical flux will give a uniform distribution of event probabilities. Rather than being the relative probability of an event given the DCR, $P(E|DCR_n)$, it will become weighted by the relative probability given the optical flux per SPAD, $P(E|\Psi_n)$. And hence will tend to become the simple summation of gated after-pulsing probabilities.

Other Array Characterisation Experiments

As shown by the BER model (Section 3.2), if the standard deviations of N_0 or N_1 are increased, the BER also increases [2, 3]. As the SPAD array presents the summation of a set of SPADs, as either a single channel ($N_{SPAD} = 1024$) or 16 separate channels ($N_{SPAD} = 64$), variation of the received counts from different SPADs may act to increase the width of the ideally Poisson output distributions [2, 80].

To assess this possible increase in the BER, other array characterisation experiments were performed including assessment of the photo-response non-uniformity (PRNU) [81] and pixel to pixel cross talk [57]. Despite the possible factors that can influence PRNU, this was measured to be $\approx 2\%$, indicating that it will have a negligible affect on the BER. The overall level of cross talk was also negligible in a measured device, suggesting little affect on the receiver BER and sensitivity power penalties.

Conclusions from Time Dependent and Other SPAD Performance Experiments

The after-pulsing probability from the summation of multiple SPADs may impact the achievable BER through receiver inter-symbol interference. This section has measured the after-pulsing and shown how the TDM modes of the pixels can be used to reduce it. One method to reduce the after-pulsing probability is to increase the dead time, however this solution is not suitable here as it limits the maximum counting rate of the SPADs and therefore the maximum achievable detections per symbol.

Two other array characterisation experiments are also discussed, however these are shown to be negligible, thus reducing their possible affect on receiver performance. The characterisation and in particular time-dependent performance of the SPAD array is continued in the first sections of Chapter 6. However prior to that, the next section aims to model the performance and photon transfer curve of a single SPAD.

5.2.5 Conclusions from Low-Level Receiver Characterisation

The low-level characterisation above has briefly investigated three main themes important for ascertaining the communications performance of the SPAD-based receiver. These included power supply issues, static SPAD and array performance and time-dependent characteristics. The aim of this, is to have knowledge of the presence or magnitude of an effect prior to communications experiments performed in the next chapter. For example, the photon transfer curve directly impacts the receiver sensitivity, while the after-pulsing probability should ideally be known in order to separate effects that contribute to the receiver impulse response.

The results here, indicate this first generation prototype requires further work to become fully optimised for the targeted VLC application. In particular, the clock distribution network consumes a significant proportion of the device power budget, while variability in the passive quenching dead time may contribute to the BER. Likewise, the dark count rate is higher than expected, reducing the prospects for increasing the receiver sensitivity. In the next section, the photon transfer curve and a possible competitive iteration at the SPAD level is modelled.

5.3 Modelling the SPAD Sensitivity

In this section, the responsivity of a single SPAD is modelled. This gives insights into the operation of the SPAD for use as the optically sensitive device in high-sensitivity receivers [67]. The technique, establishing dead time specifications, assumed that all SPAD counting resources were used (Section 3.3.2). These are discrete slots, each the length of the dead time, that are available within a finite counting period. Reducing the dead time further than this requirement allows operation within a linear region, but may present a problem for design depending on after-pulsing trap lifetimes or capacitive recharge time constants [49, 76]. Due to this estimation methodology, operation near the saturation region of the SPAD transfer curve is of interest. As such, the model presented aims to assess the noise level and SPAD SNR once competition between photons, dark counts and after-pulses have been included.

5.3.1 SPAD Photon Transfer, Count Competition and Noise Modelling

A simple model, including input photons, $N_{photons}$, photon detection efficiency, PDE, fill factor, FF, after-pulsing, $P(ap)$ and dark count rates, DCR, can be used (Equation (5.3)). This is appropriate under some circumstances, however it does not account for the finite dead time, which is assumed to be zero. The model also assumes that photons, dark counts and after-pulses can be separated, which while achievable with careful experimentation, cannot be performed during normal operation. As the cause behind an avalanche is a carrier within the multiplication region (Section 2.6), which of the three (photons, dark counts or after-pulses), that causes an avalanche cannot be determined [26]. Further, a photon arrival and after-pulse (or a dark count and after-pulse) are not statistically independent [49]. For a model such as Equation (5.3), it is difficult to investigate the affects of a particular parameter on the count or noise rates, as certain relationships, particularly the dead time, are not adequately included. Separability of affecting phenomena, can lead to increased understanding, as such a second model tries to allow better investigation of the parameters when the dead time is included. In Equation (5.3), $PDE(\lambda)$ represents the effective PDE at a particular wavelength, λ, while ΔT represents the time period over which N_{counts} are collected.

$$N_{counts} = [1 + P(ap)](PDE(\lambda)\, N_{photons}(\lambda)\, FF + DCR)\Delta T \tag{5.3}$$

Equation (5.3) can be expanded to include the dead time, τ_d. This can account for a contention between photon, dark count and after-pulse events when each creates a short unresponsive period [120, 183]. As the dead time decreases, a new model (Equation (5.5)) will tend towards the ideal simple model of Equation (5.3). However at high counting rates, the dead time will prevent the separability, independence, and static contribution assumptions being used in Equation (5.3).

Starting with the arrival of a photon, the detector is unresponsive for a finite period, τ_d. When using an active circuit, τ_d represents a period in which $V_{ex} = 0$V. Fundamentally, no other avalanches can occur [26, 49]. This is assumed to only block further photon detections [49, 127, 133, 163], however the thermal generation of an electron-hole pair, i.e. a dark count, or the release of a trapped carrier, i.e. an after-pulse [26], will also be prevented from causing an avalanche [120, 183].

The arrival of a photon when actively quenched always predicates loss of a photon, dark count or after-pulse if one occurs during the dead time. In the passive case, events are still lost, however this occurs in a time dependent manner depending on the recharging constant of the device [211]. This creates competition at higher count rates, in which photons prevent dark counts or after-pulses, dark counts prevent photon counts or after-pulses and after-pulses prevent photon counts or dark counts. This has been explored numerically using Verilog-A in [183, 211] and a statistical jitter and avalanche build-up model in [120]. This is modelled using an enable time, T_{enb}, as a proportion of ΔT. An avalanche induced dead time will remove a short period, τ_d, from ΔT. The time in which a SPAD is in its *enabled* receptive state, is then given by Equation (5.4).

$$T_{enb} = \Delta T - [(1 + P(ap))(\psi_{act} + DC_{act})(\tau_d)] \tag{5.4}$$

$P(ap)$ is the after-pulsing probability, assumed to be static with optical power and τ_d is the dead time, assuming active quenching. ψ_{act} and DC_{act} are the actual number of photon and dark count detections respectively, along with the number of actual output after-pulses which is modelled explicitly using $AP_{act} = P(ap)(\psi_{act} + DC_{act})$. However these are intermediate output values of Equation (5.4) in a time ΔT, and must be corrected for the action of competition between them. This is rather than the ideal input values that would be calculated using incident optical power, fill factor or carrier generation rate at a specific temperature [26].

A regression can be made back to the input values, ψ_0, DC_0 and AP_0 prior to the action of competitive interaction. The final counts N_{counts}, including competitive interaction and the finite dead time, is then given by Equation (5.5). For example 90 photons may be detected, however due the competition between these and other events, the detection of 10 other incident photons may not have occurred. The input incident photon rate would then be 100 photons.

$$\begin{aligned} N_{counts} = &\left(\frac{\Delta T}{1 + (1 + P(ap))(\psi_0 + DCR_0)(\tau_d)} \right) (\psi_0 + DCR_0) \\ &+ \left(\frac{\Delta T}{1 + (1 + P(ap))(\psi_0 + DCR_0)(\tau_d)} \right)^2 (\psi_0 + DCR_0) P(ap) \end{aligned} \tag{5.5}$$

To check the validity of the model (Equation (5.5)), the simple and competitive models are compared with experimental results for the photon transfer curve in Figure 5.12.

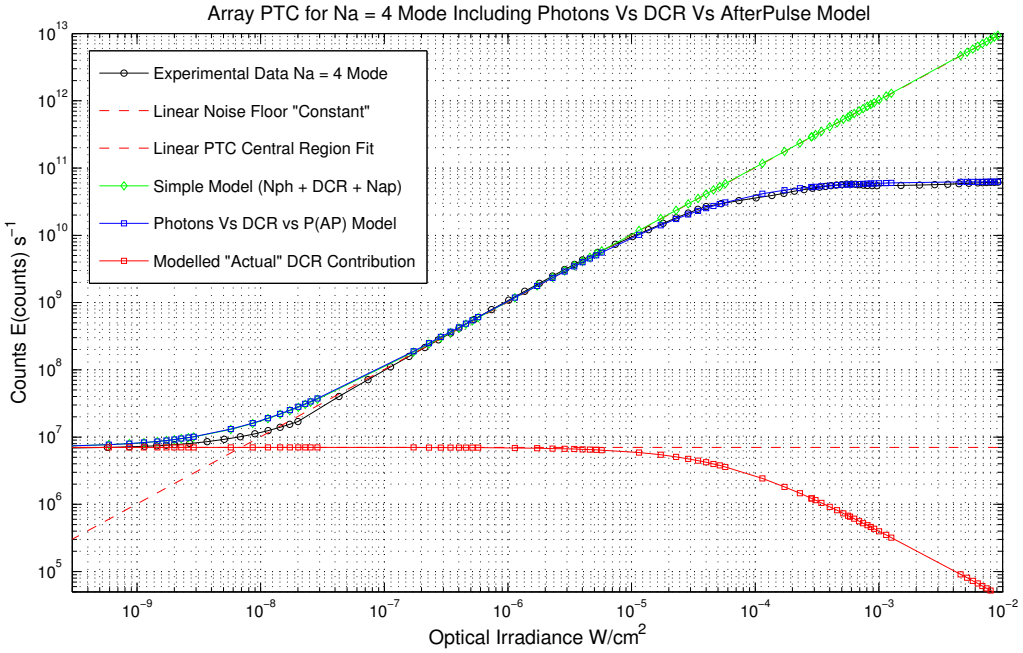


Figure 5.12: Count competitive and simple addition models for the photon transfer curve. The fitting parameters are: $\lambda = 450$nm, $PDE = 23\%$, $FF = 2.4\%$, $P(ap) = 1\%$, $\tau_d = 16$ns, $DCR = E(DCR_{meas})$.

The simple model, fits for low counting rates but departs at higher rates due to the assumption of a zero dead time. The new model matches experimental results, shown in black (Figure 5.12). This allows exploration of quantities unavailable experimentally, such as the proportion of detections caused by dark counts. Figure 5.12 shows the idealised linear fit for the DCR noise floor, and can be viewed as the assumption of a static noise floor over the optical range. The ideal linear fit for the central region is also given. There is a departure of experiment at both ends of the optical range, due to some uncertainty of fitting parameters such as the area of the beam and the precise value of the PDE.

5.3.2 Modelling The Dynamic Range of a Single SPAD

While the dynamic range is the ratio of maximum and minimum observable signals, it can be expanded to be the observable signal at some optical bias. This is important for finite extinction ratios common in optical sources or when dealing with a small signal modulation superimposed on a continuous level [14].

Many SPAD, application independent, models [75, 163, 167] assume that the DCR and noise floor $\sqrt{DCR}$, are constant over the optical range. This is understandable considering thermal carrier generation [26], however this assumes the infinitely short dead time used in Equation (5.3) [163]. Non-communications applications of SPADs such as fluorescence lifetime analysis [208] may use this assumption if the integration time of the count measurement is much larger than the dead time, or if count rates are low in comparison to C_{max}. However for communications with short symbols, competition between photons, dark counts and after-pulses, or operation near the maximum counting rate cannot be ignored. In [163] the dynamic range of a SPAD is defined using Equation (5.6).

$$DR_{dB} = 20log_{10}\left(\frac{C_{max}}{rms(DCR)}\right) = 20log_{10}\left(\frac{C_{max}}{\sqrt{DCR}}\right) \tag{5.6}$$

This assumes, that the noise floor is static with increasing flux, and that at the point C_{max}, where a photon arrives just as the SPAD has recharged, that a dark count can still be detected with the same probability [163]. This seems paradoxical as [163] also defines the maximum count rate, $C_{max} = \frac{1}{\tau_d}$, putting the non-zero dead time into Equation (5.6) with a denominator that assumes $\tau_d = 0$. For a more suitable dynamic range definition, second order effects such as after-pulsing, must be included, particularly in applications such as communications where dead times comparable to symbol widths, are required [7, 61, 64].

In Figure 5.12, the competitive model is used to investigate DCR as a function of optical power. It is not possible to measure this experimentally due to inseparability of counts, however with close fitting to experiment, this shows a reduction in DCR as the flux increases.

In [216] the noise floor, in this case $\sigma = \sqrt{Photons + DarkCounts + AfterPulses}$, is shown decreasing with optical power, just as the count rate starts to saturate and deviate from its ideal linear fit. This is similar to that observed in Figure 5.12. The reduction in the combined photon, dark count and after-pulse output signal noise can be explained, firstly if the DCR has fallen towards zero. And secondly, if saturation at high counting rates forces the summation of the SPAD dead times, i.e. T_{enb}, to be a significant proportion of the measurement. This gives little room for photon number variation, exponential count arrival or counts other than photons. At perfect saturation, the variation in output counts should therefore tend towards zero, and no extra counts above the maximum can be detected, simply as T_{enb} is zero at that point.

For dark counts, the reduction in enable time due to photon arrivals ($\Delta T N_{detections} \tau_d$), prevents a thermal carrier from causing an avalanche when the photon flux is much greater than the dark count rate (Equation (5.7)) [213].

$$N_{counts} = \frac{N_{detections}}{\Delta T - \tau_d N_{detections}} \tag{5.7}$$

At saturation, photons represent the highest probability of creating an avalanche. A thermal carrier [26] is unable to create an avalanche as the detector is likely to be in its dead period with zero bias [49]. Thus, the actual number of dark counts, DC_{act}, when operating at some light level is given by Equation (5.8). For the dynamic range at a particular point, this implies that as the photon flux increases, the significant contribution to the noise floor switches from the variability of the DCR to the photon shot noise, $\sqrt{\psi_{act}}$.

$$\begin{aligned} DC_{act} &= DCR_0 \left(\frac{T_{enb}}{\Delta T} \right) \\ &= \frac{DCR_0}{(1+P(ap))(\psi_0 + DCR_0)(\tau_d)} \end{aligned} \tag{5.8}$$

5.3.3 The Signal-To-Noise Ratio Of A Single SPAD

To define the SNR of a single-SPAD rather than an array of SPADs, Cova et al [75] include the efficiency, η, along with the variation noise of the DCR, n_d and photon rate, n_p, in a period of ΔT (Equation (5.9)). This points to a definition that includes after-pulsing, photon noise, correlation factors and DCR in order to obtain a reliable estimate. The hope is, that this can begin assessment of the signal chain in order to estimate the BER. However as with Equation (5.3), the model assumes a static DCR as the incident flux increases.

$$SNR = \frac{\eta n_p \Delta T}{\sqrt{\eta n_p \Delta T + n_d \Delta T}} \tag{5.9}$$

With inclusion of count competition, a new SPAD SNR can be defined, starting from Equation (5.10). This will be the signal amplitude to the expected noise floor, as in the Cova et al definition [75] (Equation (5.9)), but at that particular signal input level. This is important as a modulated signal may not have a *zero* signal at the noise floor, but may instead be at some level below the *one* signal (Equation (3.17)). The μLED transmitters used within this project are assumed to use an extinction ratio of 100% due to their digital *on/off* drivers and their prospective use in biological fluorescence lifetime analysis applications [96]. As such, the SPAD SNR derivation below, will only be of use when the N_0 level is not at the receiver noise floor, as in the case of a finite extinction ratio.

In Equation (5.10), $\psi_{act}(\psi_0)$ is the actual photon counts detected by a single SPAD, given some input photon level, $n_{ap}(\psi_0)$ is the after-pulse contribution and $DC_{act}(\psi_0)$ is the dark count rate at that same photon level, ψ_0. This assumes that each process, photons, dark counts and after-pulses, is Poisson distributed, however like Staples et al [176], the noise fluctuations can be described using the standard deviations, $\sigma(\psi_{act}(\psi_0))$, $\sigma(DC_{act}(\psi_0))$, and $\sigma(n_{ap}(\psi_0))$.

$$SNR(\psi_0) = \frac{\psi_{act}(\psi_0)}{\sqrt{\psi_{act}(\psi_0) + n_{ap}(\psi_0) + DC_{act}(\psi_0)}} \tag{5.10}$$

Including regression back to pre-competition values (Equation (5.5)), the competitive model SPAD SNR can be defined using Equation (5.11). The relative significance of the input, shifts from dark counts to photon counts as the photon signal, ψ_0, increases. Once the SPAD has a greater probability of firing due to photons rather than dark counts, the relative significance shifts to the after-pulsing and photon shot noise terms. If the after-pulsing probability, $P(ap)$, is low, such as the $\approx 1\%$ measured (Section 5.2.4), or if photon arrivals prevent avalanches due to trapped carriers, then the SPAD SNR will be dominated by the photon shot noise, $\sqrt{\psi_0}$.

$$SNR(\psi_0) = \frac{1}{\sqrt{1 + (1 + P(ap))(\psi_0 + DCR_0)(\tau_d)}} \left(\frac{\psi_0}{\sqrt{\psi_0 + DCR_0 + n_{ap0}}} \right) \tag{5.11}$$

While this is the SPAD SNR, for the 1024-SPAD array as a whole, the combined array SNR, SNR_{array} including temporal noise, could be used to estimate the receiver BER [14]. This however requires further work to estimate. As discussed by Sackinger [14], the conversion of SNR to BER requires knowledge of the ratio between the *zero* and *one* amplitude and noise levels. As such, this estimate needs further specification of the intended VLC application along with estimation of the optical path, losses, optical components and packaging options for the device. This however is too early to specify during design and testing of a first generation prototype receiver.

The noise at a particular level, used in the above SPAD SNR to ascertain the ratio between *zero* and *one* noise levels, becomes useful when combined with noise analysis of the signal processing chain to estimate SNR_{array}. The BER, in the case that the mean noise of a received *one* is much larger than the mean noise of a received *zero*, can roughly be estimated by Equation (5.12) [2, 14, 103].

$$BER = \frac{1}{2}\text{erfc}\left(\frac{\sqrt{2\,SNR_{array}}}{\sqrt{2}} \right), \quad \text{if } noise(1) \geq noise(0) \tag{5.12}$$

The translation of SPAD SNR (Equation (5.11)) to Array SNR (SNR_{array}) requires further modelling to include SPAD grouping effects and temporal noise sources. In particular this must include DCR, PDE and P(ap) variation over the 2D array, SPAD time division multiplexing,

SPAD to SPAD cross talk, the affect of high-noise SPAD removal and the combined effects of receiver inter-symbol interference mechanisms such as the receiver impulse or step response.

Vitally, the array SNR cannot be modelled or represent the likely BER correctly, until temporal effects such as the μLED bandwidth, receiver impulse and receiver step responses have been measured and modelled. As such, estimation of the BER is deferred to Chapter 6 where, a) the impulse and step responses are measured, b) the experimental setup for communications experiments is discussed, and c) any deviation of the μLED away from the assumed 100% extinction ratio, dictated by the digital CMOS driver circuits, is assessed and quantified.

5.3.4 Conclusions from Modelling the SPAD Sensitivity

In this section, modelling has attempted to redefine the SNR of a single SPAD. This has been done as previous SNR estimation models made a number of assumptions that may not have been suitable for communications. Here, the competitive interaction between photons, dark counts and after-pulses is included, as is provision for the SNR as evaluated at a particular optical bias point. This is akin to differing voltage amplifier SNR performance depending on where on the transfer curve the input signal is biased.

With the base competitive model fitting the experimental PTC, the SNR model suggests that the dominant noise changes depending on the optical input. At low optical power, the DCR dominates. As the optical power increases, the dominant noise source switches to the photon shot noise. This suggests further effects that may change the performance of a receiver that uses a SPAD array as the optical element. The outcome of this section is therefore a SNR definition that could, with further modelling, be used during the specification stage of a second generation receiver. This would be particularly useful, along with the models within the next chapter, if clearer specifications of the intended application are also included during the initial design phase.

5.4 Conclusions

This chapter described characterisation of the 32x32 SPAD array (Table 5.5). This allows knowledge of effects that may impact communications performance, before on-off key signals are sent to the CMOS receiver.

In the first section, the power draw with clock frequency and optical flux were measured. This implied that clock distribution consumes a significant amount of power, indicating room for optimisation in future receivers. The power draw due to photon detection is lower than expected, prompting further investigation into possible power supply non-uniformity and ripple issues.

The first section also measured the dead time, which limits detections per symbol and the BER. Active quenching has a tighter distribution which, through modelling, has been shown to be advantageous in reducing some of the variability of the detected photon flux. Thus a modified, accurate dead time circuit may be helpful in reducing the BER towards the quantum limit. The transmitter extinction ratio is likely to be the dominant factor in practical μLED links, however DCR measurements investigated the theoretical noise floor when using ideal OOK modulation (Section 2.3.3).

As a prelude to receiver sensitivity estimates in the next chapter, the array responsivity was investigated using photon transfer curves. These directly impact the communications receiver sensitivity and can include non-ideal effects such as count paralysis. This was later modelled implying that the detected noise rate may change with incident optical power.

In the next chapter other effects such as the transient responses, are explored along with communication experiments using a μLED device, on-off key modulation, and a visible wavelength (Blue at 450nm), suitable for VLC application.

Device Property	Value	Conditions
DVDD Leakage Power	$\approx 500\mu$W	Zero SPAD Activity and $CLK \leq$ 1MHz
DVDD Leakage Power	$\approx 450\mu$W	Minimum SPAD Activity and Zero Clock
SPADDVDD Leakage Power	$\approx 70\mu$W	Minimum SPAD Activity and Zero Clock
Max Clock Distribution Power	$\approx$ 20mW	Zero SPAD Activity and CLK at 100MHz
Max AQ DVDD Power	$\approx$ 12mW	Maximum SPAD Activity and Zero Clock
AQ Dead Time Range	10–250ns	Using $V(DTCTRL)$
PQ Dead Time Range	5–200ns	Using $V(DTCTRL)$
Base AQ Mean Dead Time	13.5ns	Sim = 11.5ns, $V(DTCTRL) = 0V$
Base PQ Mean Dead Time	23.1ns	Sim = 20.3ns, $V(DTCTRL) = 0V$
$V(DTCTRL)$ Tuning Range	-900 to $+800$mV	
DCR (ALL SPADs)	Mean = 7.27kHz	At 16°C
-	Median = 2.5kHz	At 16°C
Photon Transfer Curve	$C_{max} = 6.5\text{x}10^{10}$cps	TDM Mode $N_a = 4$
-	$C_{min} = 6.6\text{x}10^{6}$cps	TDM Mode $N_a = 4$
After-Pulsing, $P(ap)$	mean($P(ap)$) = 0.9%	1μs Window, 12 SPADs

Table 5.5: A short summary table of the main results discussed within this chapter.

Chapter 6

High Level Characterisation And Communication Results

6.1 Higher-Level Characterisation Overview

Higher level characterisation of the SPAD array, along with communications results with a visible light μLED [23, 24], are presented in this chapter. This allows some assessment as to the effectiveness of the techniques used to improve the receiver sensitivity, leading to the conclusion that while data can be received with such a receiver, some restrictions will be placed on the data rates and performance achievable with this first generation prototype. The funding project [65] specified On-Off Key modulation, the receiver designed here is therefore only suitable for this technique. Other modulation types such as multi pulse amplitude modulation (M-PAM) or pulse width modulation (PWM), require only a slight change in the concept [1].

For realistic communication systems, the measured impulse and step responses may limit the receiver, degrading the achievable BER for the target data rate. These will act to compound any effects from the transmitter device (such as the extinction ratio) or the optical channel, and are therefore important effects to assess for technology scalability. Both of these are measured in the first section of this chapter. The aim here is to evaluate the receiver performance only, independent of the transmitter device. Doing so allows the concept, that of three combined sensitivity enhancement techniques, to be evaluated. This is similar to recent feasibility studies of multiple SPAD detectors for communications [67], where the extinction ratio requires an increased optical power, i.e. a power penalty, to retain correct operation.

As discussed (Section 3.4), lowering the dead time can allow operation within the linear region, although this can also cause issues with data rate scaling. A step response model suggests an advantage to lower dead times, that of better settling characteristics. Together, the dead time and the step response indicate future work necessary on dead time scaling and reduced after-pulsing, both of which are necessary before commercialisation of the technology.

In Section 6.4 communication results between a single μLED and the SPAD-based receiver, first investigates an ideal zero ambient light case. This uses a large modulation in the received number of photon counts per symbol and may be more of a proxy for visible communications within polymer optical fibres. Complementing the transient and communications results, the

bit error rate is estimated. With sweeps of unmodulated optical power and the simple BER model [3], the error rate only due to the receiver continuous wave response, can be found. The analysis of the BER is then extended [3] to analyse effects in realistic systems. This takes a multi-modal model of the received bit distributions.

To conclude, important performance issues are discussed with respect to communication link designers, modellers and future designs. To aid in this discussion, figures of merit are used to compare the receiver with prior art. The receiver is shown to perform favourably, although being a first generation, there is room for improved performance and further characterisation.

6.2 Measurement Of The Receiver Transient Response

The response of a system to an impulse or step is important for understanding its behaviour [1, 14, 36]. A SPAD gives a fast response [121, 125, 127], however the extent that one symbol interferes with proceeding symbols is critical [14, 36]. Here the transient responses are measured over optical intensities, dead times, pixel configurations and quenching modes.

As symbol widths decrease, the transient response causes inter-symbol interference for more symbols [14, 36]. At 100Mb/s the transient may only affect a few subsequent bits. However the scalability of a technology to higher data rates, such as 1Gb/s, is important if the technology is to fit with the overall increase in bandwidth demand [11, 12] (Section 1.1).

6.2.1 Transient Experimental Setup

Blocks of memory on the testing FPGA [222] obtain data from the receiver. For this experiment, the memory allows a programmable flag to be raised once a certain amount of data has been obtained. Firmware and software modules save a frame of data while triggering a laser pulse. By repeating this, the average transient response can be calculated.

For the impulse response, a 53ps, $\lambda = 442$nm pulsed laser was used. This was replaced for the step response with a $\lambda = 840$nm continuous wave laser, that allowed triggered activation. While jitter of the trigger was observed, measurements with incorrect acquisition were removed.

The results plot the mean, maximum and minimum of multiple transient experiments. For impulses, a double exponential is fitted, allowing comparison across parameters using the decay constant. This is the time for the fitted curve to reach $\frac{1}{e}$ of the maximum. The clock speed was set to the maximum of 100MHz, which limited the experimental time resolution.

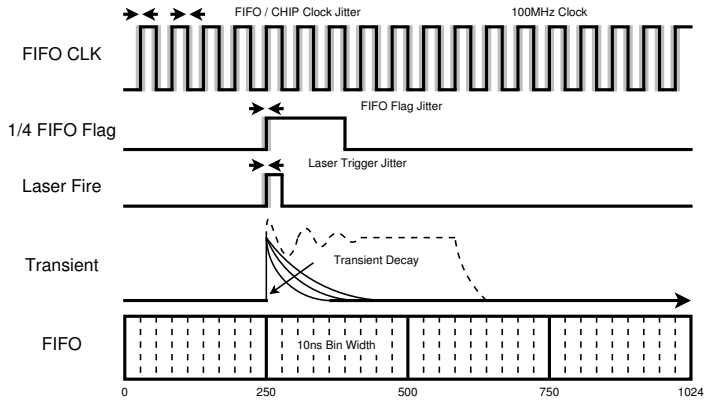


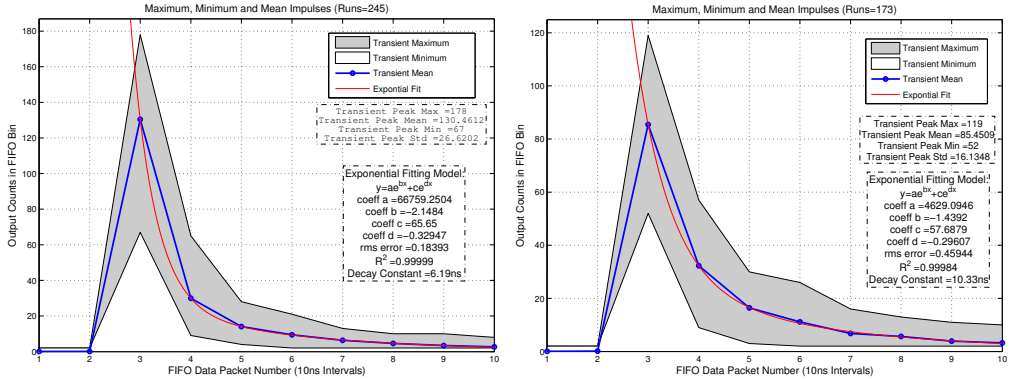
Figure 6.1: Data acquisition and output timing diagram for impulse and step response experiments. The FPGA on the testing PCB controlled the triggering of a laser using a programmable flag. This allowed the impulse or step response to be measured, but also allowed measurement of the output immediately prior to the laser trigger.

6.2.2 Impulse Response under Optical Intensity Variation

To obtain the specified BER (Table 2.1), a quantum limit optical power, optical extinction ratio, ([1] page 207), SNR [14] and N_0 to N_1 separation, assuring the correct number of received photons is needed [2, 3, 14] (Table 3.2 and Equation (3.1)). This is exponentially dependent upon the required BER [14] (Section 3.2), and linearly dependent upon the detection efficiency [3]. To optimise communication efficiency, the receiver sensitivity, along with the properties of the transmitter and channel, are used to assess the amount of light necessary, i.e. a communications power budget [1, 14]. Transmitting optical powers greater than this budget allows a lower BER than specified [2], but does so with increased energy consumption [2, 14].

As inter-symbol interference can determine technology scalability, the impulse response is measured against different optical powers. At low powers (i.e. high BERs such as $1x10^{-3}$ [2]) only a few SPADs will fire, causing low probabilities of after-pulsing in subsequent symbols. At high powers (i.e. low BERs such as $1x10^{-12}$ [2]), almost all SPADs will breakdown. This gives a large number of filled traps and consequently larger after-pulsing probabilities in subsequent bits [49, 76, 123, 174]. Additionally, an increased number of photo-generated electron-hole pairs may occur within the SPAD guard ring, Sub-Pixel circuit and other areas of the substrate [26], due to the optical window needed for the SPADs (Figure 4.2). This may contribute to the impulse response and intrinsic bandwidth, as seen in CMOS photodiode receivers [31, 33, 44].

Figure 6.2.a shows impulses using active quenching and a globally set $N_a = 4$ mode. The quoted optical power is measured separately with the average power measured when the laser is set to a 100MHz repetition rate. The optical energy per pulse is then calculated using the $\approx$ 1KHz repetition rate of the experiment.



(a) Low optical power of 4.78μW **(b)** High optical power of 530μW

Figure 6.2: Transient responses at 4.78μW and 530μW optical power ($\tau_d = 13.5ns$). Both plots show a typical impulse response shape with an increase in the fitted time constant as the optical power of the laser pulse increases. Shown with the mean impulse response of multiple runs in blue, are the maximum and minimum impulse responses recorded, along with the double exponential fit (red curve).

In Table 6.1, an increase in decay constant is observed if the average optical energy is above 20fJ. Below this it decreases towards 1.9ns. With the average $P(ap)$ across twelve devices, found to be 0.9% at the average active quenching (DTCTRL=0V) dead time of 13.5ns (Section 5.2.4), lower decay constants would be expected. For lower powers such as 1.25μW (trace not shown), the exponential decay can be explained by a mean after-pulsing probability of $\approx 1\%$. This arises from values of 0.45 and 0.2 in the proceeding 10ns bins after an average impulse response peak of 37 counts, giving an estimated full array total after-pulsing probability of 1.2%.

For higher powers, the decay becomes longer than can be accounted for by after-pulsing. At 4.78μW, an after-pulsing probability of 23% would be required, this is clearly not the case due to the measured low optical power impulse responses and native after-pulsing probability. This *effective after-pulsing probability* is defined as the probability that would be needed to account for the decay with after-pulsing alone. This increases with optical power to an effective after-pulsing probability of 30% at 261μW and 37% at 530μW. As this cannot be after-pulsing, some other mechanism, linked to incident power, must be contributing.

Table 6.1: Transient response decay constant with measured optical power. Measured values are the average decay constants from multiple runs, with erroneous runs removed prior to analysis. The detection rate of impulses was 100%, with a repetition rate of ≈ 1kHz.

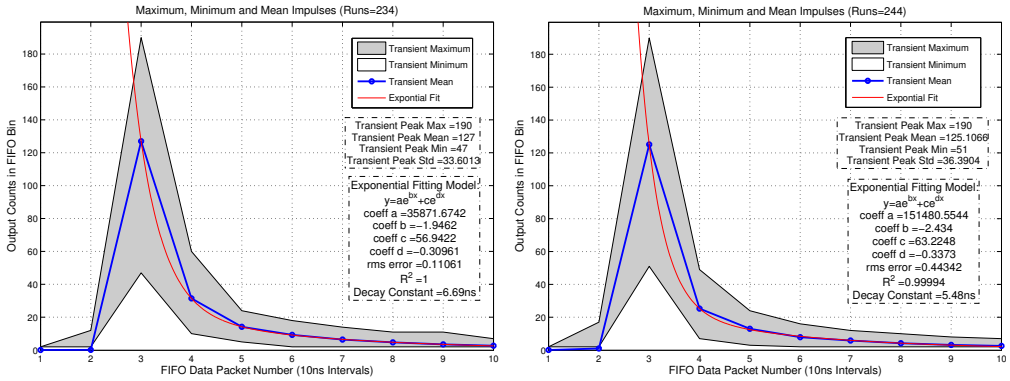
Optical Power (uW) @ 100MHz Rep Rate	Optical Energy (J) Per Pulse	Photons per SPAD Per Pulse	Decay Constant (ns) Time to e^{-1}
1.25	12.5f	27	1.91
1.55	15.5f	34	1.89
4.78	48.7f	106	6.30
5.4	54.0f	118	6.70
110.1	1.10p	2393	7.01
205.1	2.05p	4458	7.88
261.0	2.61p	5673	8.70
271.0	2.71p	5891	8.87
530.0	5.30p	11520	10.41
1093	10.93p	23758	12.34
1417	14.17p	30800	14.79

The increase in decay constant could be explained by photo-generated carriers [31, 33]. After the laser pulse and dead time, minority carriers could diffuse into the depletion regions creating secondary pulses not related to trapped carriers. This is similar to intrinsic bandwidths [33] in photodiode and avalanche photodiode receivers [26, 28, 31, 40, 48], where photo-generated carriers reach the photodiode with a random delay, thus contributing to ISI [31, 33].

Photodiodes within CMOS have used approaches of in-substrate structures, similar to the Sub-Pixel guard ring (Figure 4.2), to increase the intrinsic bandwidth [26, 33]. Future SPAD-based receivers, must assess this contribution prior to device floor planning, with increased guard ring structures [35] as appropriate, given that they can reduce the optical fill factor.

6.2.3 Impulse Response under Active and Passive Quenching

The impulse response is measured under passive and active quenching with a pulse energy of 55fJ. Sweeping the dead time shows that with low values, the decay becomes longer. This can be attributed to increased after-pulsing, or increased time available for secondary avalanches. This is shown, when using dead time control voltages of -150mV, -100mV and -50mV, which give decay constants of 5.67ns, 5.63ns, and 5.48ns respectively.



(a) Active Quenching **(b)** Passive Quenching

Figure 6.3: Impulse response under active and passive quenching modes. Both measurements use the same light level, the same experimental setup and the same dead time of 13.5ns. The fitted curve (red) indicates only a slight change in the impulse response decay constant. The light level used here is 5.4μW, when the laser was set to 100MHz repetition rate. The energy per pulse (experiment run at $\approx$ 1kHz rep rate) was 54fJ.

In Figure 6.3 the response is measured at the same dead time ($\tau_d = 13.5$ns) for both active and passive quenching modes. The active quenching decay constant of 6.69ns reduces to 5.48ns in the passive case. For the same dead time, the active circuit recharges the SPAD fully before the passive quench, which is slowly recharging through its exponential RC time constant. Therefore a carrier arriving shortly after the dead time will have a lower voltage and charge per pulse in the passive mode and a correspondingly reduced probability of a second pulse [49].

6.2.4 Impulse Response under Pixel Na Summation Modes

The impulse response is measured over different N_a modes to understand the temporal affects on ISI. In Figure 6.4, the response is shown for globally set modes of $N_a = 4$ and $N_a = 1$. In $N_a = 4$ (Section 4.5.6), the decay constant is 6.41ns. The $N_a = 1$ mode shows a reduced decay constant of 2.35ns. With after-pulsing and optically generated carriers being possible causes of the response, the influence of the N_a mode, may be caused by two effects:

- Firstly in $N_a = 1$, there is a reduced number of SPADs that are able to fire upon laser triggering. In this case, $N_{SPAD}\left(\frac{N_a}{N_{TDM}}\right) = 1024\left(\frac{1}{4}\right) = 256$. This will lead to a reduced number of traps being filled by avalanche currents, lowering the after-pulse contribution.
- Secondly, $N_a = 1$ gates the SPADs once they have fired, only recharging them once the TDM loop has iterated through all N_{TDM} devices. As such, traps or carriers near a SPAD are unable to create an avalanche. Recombination or release of trapped carriers [26] during this period, gives a smaller probability of secondary avalanche once the SPAD has been re-enabled by the pixel state machine.

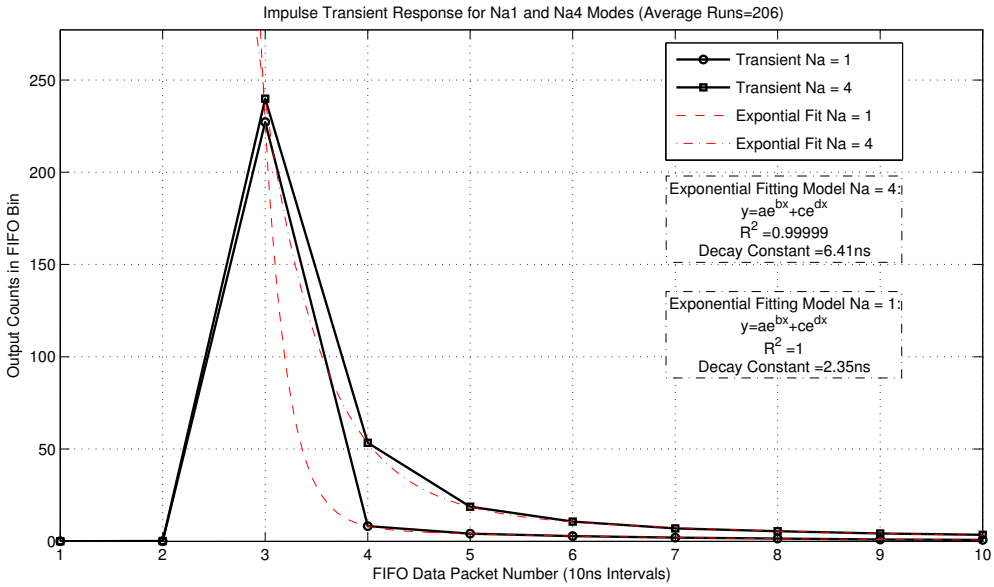


Figure 6.4: Impulse response measurements under different globally set pixel N_a modes. Despite a reduction in the overall detection probability, the $N_a = 1$ mode shows a decreased impulse response decay constant when operating under the same light level and experimental conditions. The light level used here is 0.26mW, when the laser was set to 100MHz repetition rate. The energy per pulse (experiment run at $\approx$ 1kHz rep rate) was 2.6pJ.

6.2.5 Step Response under Optical Intensity Variation

This experiment assesses the transient inter-symbol interference between a transmitted *one* and a second transmitted *one*, due to the receiver step response. If the step response acts to reduce the detection efficiency, the distance between a received amplitude, N_1 and the OOK threshold, N_{th} may be reduced, giving an increased probability of error (Section 3.2). In Figure 6.5, the step response is shown for a number of optical power densities. The most dramatic change is in the steady state value. This decreases in comparison to the initial peak as the optical power increases. For the upper most trace (6.75μW) the peak reaches 240 counts, the steady state however is approximately 190 counts. This trace also shows a severe dip in the received counts in the second time period ($\approx$ 175 counts per symbol), i.e. the second 'one' symbol after a $0 \rightarrow 1$ transition. Overall, the step response appears to improve if the expected count rate is a small proportion of the total array counting resources. For example, for the 0.5μW trace at the bottom of the figure, the steady state is very close to the initial number of counts, and the peak and trough behaviour is minimal. As the number of detections depends on the optical power, the influence of this can be removed by normalising the data to the steady state for each trace. Doing this allows better comparison of how the form of the step response changes, i.e. the relative sizes of the peaks, troughs and the settling time.

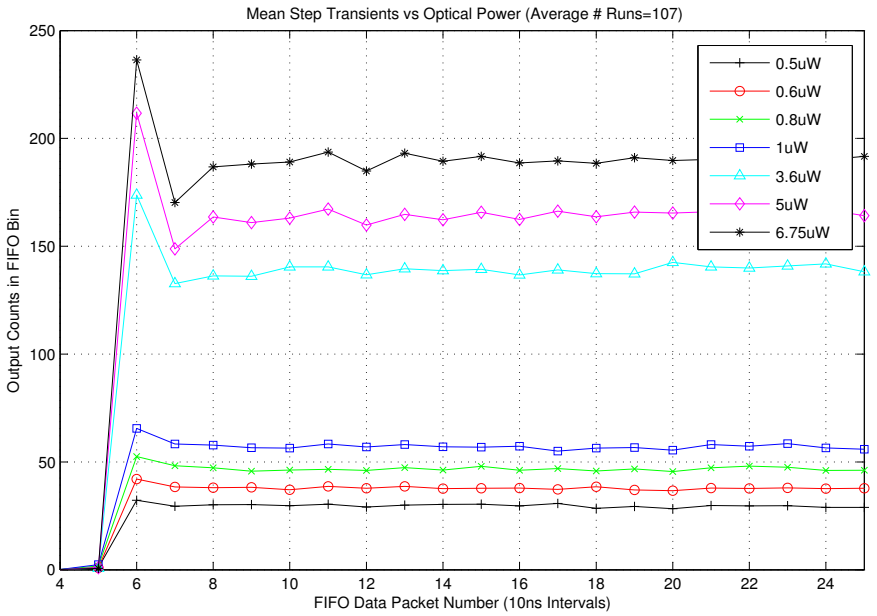


Figure 6.5: Absolute count number (non-normalised) average step responses over optical power. The peak and the subsequent ripple in the number of counts per 10ns symbol, gets worse as the optical power increases. The response also decreases as a proportion of the initial number of counts in the optical step, suggesting that the N_1 distribution of an OOK signal may become split into two, one main peak that represents the steady state for data transitions such as $0 \rightarrow 1 \rightarrow 1$, and a higher peak that represents the initial peak of the step response for data transitions such as $0 \rightarrow 1 \rightarrow 0$. This splitting is explored and shown experimentally in optical communication experiments later in this chapter.

Figure 6.6 shows the normalised step response with increasing optical power. The experiment shows variation from one run to another, possibly due to the characteristics of the laser. The validation process has reduced this, by not only removing runs with obvious jitter but also runs that show unusually low count rates. Each trace has been normalised to its long term steady state, thereby removing the large changes in the steady state. The normalisation approach therefore gives an initial count value of 110, a normalised value of 1.1 for a steady state of 100. Figure 6.6 clearly shows a large change in the step response with optical power. For the 5μW step response, the peak is 29% above the steady state. This will give a separate peak in data communication histograms depending on the data sequence (Figure 6.14). The following data symbol shows a value 10% below the steady state. In comparison, lower levels such as 0.5μW steps, give peaks only 10% above the steady state, followed by a minimal trough. When combined with the observation that the lower light level responses have a steady state closer to the initial number of counts, any splitting of the N_0 and N_1 data histogram will be reduced.

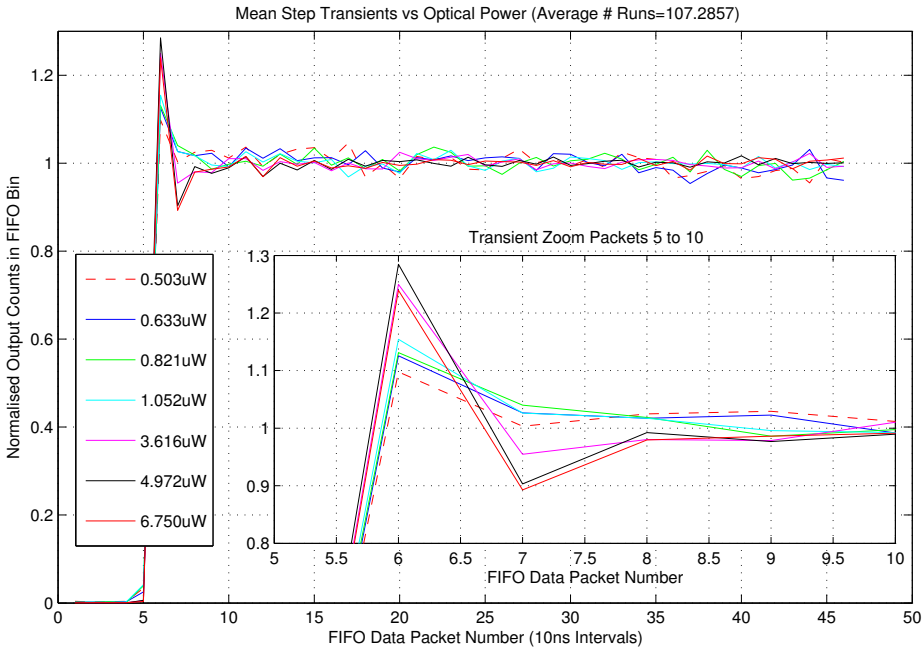


Figure 6.6: Normalised average step response over a sweep of optical power. Multiple valid runs are used to generate each step response trace, with all traces normalised to their long term steady state values. The peak and the subsequent ripple in the number of counts per 10ns bin, is highlighted in the inset zoom (50ns) after the start of the laser step.

At high light levels, the typical step response rise, settling time and over and undershoot are seen. Beam powers were measured at a 12MHz repetition rate and 100ns optical pulse width. This effectively produced a continuous wave source, as 12MHz gives a pulse start separation of 80ns. This was then used when calculating the power per pulse at the slower $\approx$ 1kHz repetition rate of the experiment. A power of 1.05μW relates to an effective energy per 10ns interval of $10.5fJ$.

The incident power is difficult to ascertain experimentally. While all of the beam can be ensured to be incident on the power meter, the beam diameter at the testing plane cannot be observed directly. With 840nm light, a fluorescent viewing card must be used, however this gives too high a variability in beam diameter to accurately calculate beam intensity.

The response at low power per pulse shows a small peak and obtains steady state within one or two cycles. As the power increases (Figure 6.6), the initial peak becomes larger, increasing towards 30% above the steady state (at 4.97μW). The settling time becomes longer, suggesting optical powers must be moderated for low BERs. For the step response, there are three affects that may be contributing to this behaviour.

1. If a readout clock is used such that integration periods are shorter than τ_d, then at the next clock edge, a number of SPADs have broken down and not fully recharged. The available pool of devices that can fire is reduced and the effective total sensitivity decreases. If the dead time and integration time are comparable but includes dead time variability, at the start of the next cycle a varying number of devices will have fired (Equation (3.2)) and recharged (Equation (3.3)) [80, 116].
2. As the optical power increases, the probability that photons arrive closer to the start of the pulse, rather than distributed within the symbol period, increases. This effect is called first-photon bias, which is known to increase as the number of photon arrivals increases [227, 228, 229, 230]. A large, almost instantaneous current spike will be seen by SPADDVDD and VBD if this occurs. The ripple in the step response could be the action of power supply variation, modulating the SPAD *PDE*. This was tackled through on-chip grid-like power supplies with local decoupling used throughout.
3. With the avalanche creating a large number of carriers, a local heating effect may contribute. This has been noted by a number of authors, albeit with large area diodes [140, 165, 176, 231, 232]. Thermal variation and modulation of junction breakdown voltages has however been reported to be negligible [137, 144].

6.2.6 Step Response under Active and Passive Quenching

The step response with different quenching methods and a small sweep of the passive quenching dead time is shown in Figure 6.7. For the same optical power ($\approx 5\mu$W), active quenching shows a smaller initial peak, a steady state that is higher than the passive quenching case and a faster settling time. This difference is key for the assessment of future designs, as active quenching has already been linked to lower dead time variability (Figure 5.4) which through modelling has been shown to be beneficial for the standard deviation of the N_1 distribution, σ_1 (Figure 5.5), and thus a better BER. However active quenching requires more electrical power (Figure 5.2) and often reduces the optical fill factor in comparison to passive quenching, and therefore must only be used if necessary. In the inset zoom of Figure 6.7, all five traces follow the same peak then exponential fall to steady state, but the point at which they get to steady

state changes. The two dead times for active and passive quenching become equal with PQ DTCTRL $\approx -150mV$, allowing direct comparison between the active and passive quenching schemes. Active quenching settles at a steady state of 160 counts, while the passive case for the same dead time settles at ≈ 125 counts per symbol. Further, a slight trough in the active case is not present in passive quenching.

The small sweep in the passive quenching dead time, while only a small change in the steady state, suggests that a lower dead time is required for fast settling and a steady state a higher proportion of the number of input counts. The steady state changes between 113 counts per symbol and 125 counts (a change of 12 counts per symbol) for a dead time change of ≈ 2.5ns. A numerical and analytical model of the step response is presented in the next section to investigate this, allowing general non-numerical estimation of the direction that should be taken in the near future.

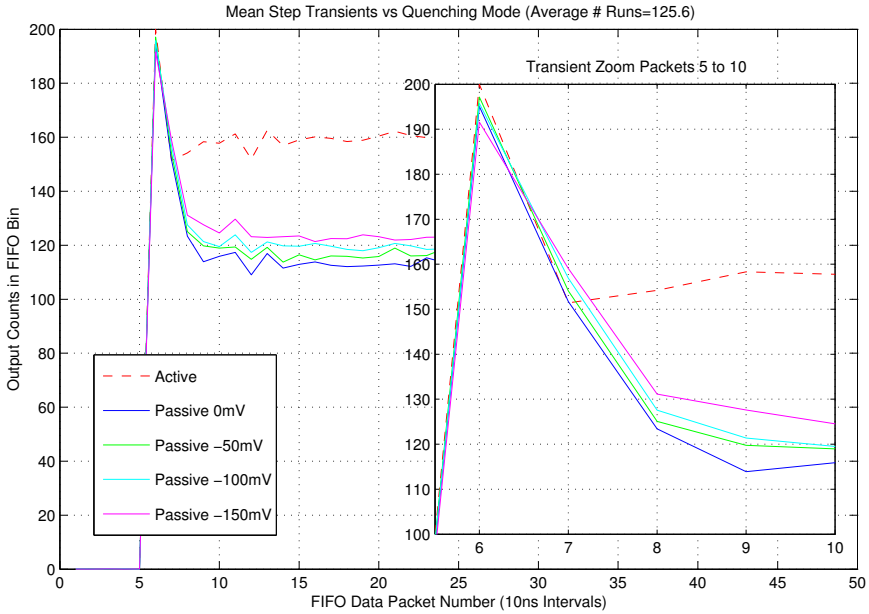


Figure 6.7: Averaged absolute counts per symbol step responses for active and passive quenching. Active quenching achieves a steady state value more quickly and with lower under- and over-shoot in comparison to passive quenching. This complements the earlier advantage of a low dead time variability, suggesting a trade off is needed when the power use and fill factor are taken into account. Active quenching reaches a steady state of 160 counts per symbol, while for the same dead time, passive quenching only achieves a steady state of 125 counts per symbol.

As with the sweep of optical power above, the step responses can be normalised to each of their long term steady state values. Doing so allows better visual comparison of the relative size of the initial peak in comparison to the steady state. In Figure 6.8 the traces have been normalised. Active quenching has a peak 21% higher than its steady state, while the worst case (DTCTRL = 0mV) passive quenching trace has a peak 72% higher than the steady state. The peak in the

passive mode decreases as the SPAD dead time decreases, which is consistent with the overall faster recharging of SPADs after they have broken down. The inset zoom shows the changes in the step response in better detail. While it would be difficult to ascertain experimentally, there is a large difference in the form of the active vs passive responses despite the same dead time. This may be linked to the manner in which the SPAD is recharged. In particular, for the same dead time, the passive case has a larger initial peak. This could be a product of the passive quenching dead time variability, with some SPADs recharging either:

- early and hence breaking down again before the end of the integration, or
- late, giving a lower pool availability, and fewer detections within the integration period, when at steady state.

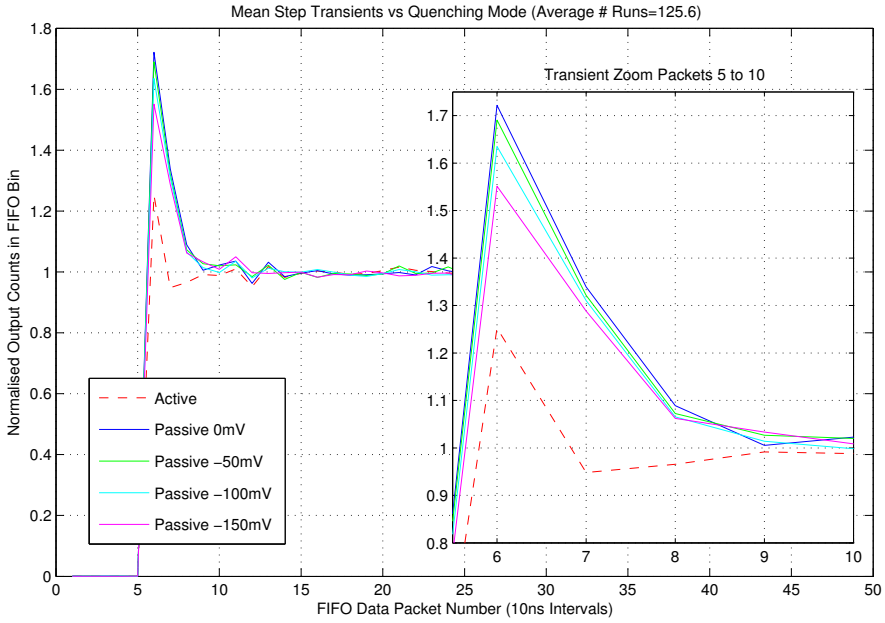


Figure 6.8: Normalised average step response in active and passive quenching. Active quenching achieves a steady state value more quickly and with lower under- and over-shoot in comparison to passive quenching. The slight reduction in peak amplitude and the time to steady state is consistent with faster SPAD recharge as the dead time decreases.

6.2.7 Step Response under Pixel Na Summation Modes

Both $N_a = 4$ and $N_a = 1$ show the typical step transient (Figure 6.9). The latter mode, shows a smaller initial peak and a trough of much lower depth, highlighting a faster settling time. With the $N_a = 1$ TDM method, the pool of available SPADs is reduced in comparison to the $N_a = 4$ mode, to 256 (Section 3.4.1). Once a photon event occurs, the Pixel switches to the next SPAD, meaning that the pool is again 256. The total detection efficiency is lower than $N_a = 4$, however the effective detection probability due to the step response alone, remains closer to 100%, despite the finite dead time.

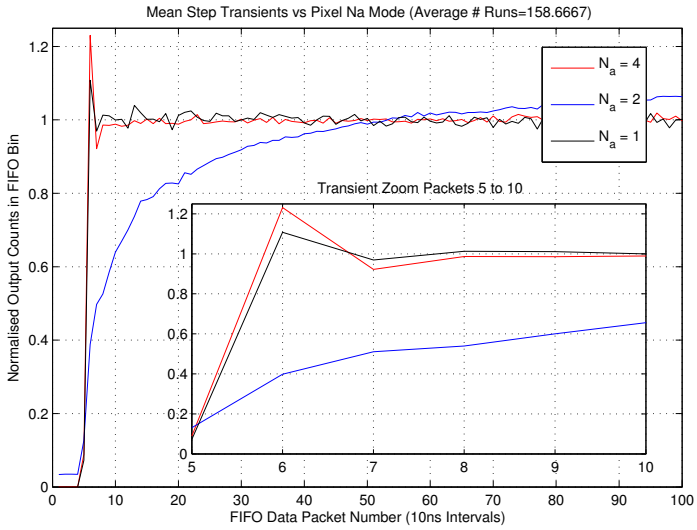


Figure 6.9: Normalised average step response over various Na modes. $N_a = 1$ shows a faster settling time and a smaller initial peak in comparison to $N_a = 4$. The $N_a = 2$ mode however shows a slow over-damped response indicating it is impracticable for communications.

Figure 6.9 also shows an over-damped response for $N_a = 2$. This takes longer to reach steady state, indicating it is not suitable for communications. The mechanism behind this over-damped response is unclear, however it may be related to double iteration of the state machine. To reduce the step response issues of $N_a = 2$, the average incident photon rate could be reduced. This however is not a practical solution due to the increase in BER if the number of photons required is not reached.

Drawing Conclusions from Measured Receiver Transient Responses

Both the impulse and step responses will contribute significantly to the overall performance of the receiver as they operate over time scales similar to the intended 100MHz $\rightarrow$ 10ns time period of a data symbol. This indicates an area that must be addressed through future work before such a system is suitable for the application. In order to begin such an optimisation, the next section discusses a model of the step response. This points methods whereby any step response induced variability in the N_1 signal, and therefore the BER, can be reduced.

6.3 Modelling the Detector Pool Dynamic Response

The detector dynamic response (Figure 6.6 and Figure 6.8) can be modelled to investigate the above measured transient issues, eventually feeding back to help the initial specification modelling of a second generation receiver. Here a model for the step transient is presented allowing further assessment of ideal parameters for the number of SPADs within a receiver and the SPAD dead time. This is based on a pool of detectors. Initially, the receiver has a full complement of detectors, all of which are available. At some time later, a number of diodes have broken down, giving a number at the output. These have been removed from the pool, and hence the effective detection efficiency, *ePDE*, for subsequent photon arrivals reduces.

Together with other models, for example the impulse response, the total ISI of the SPAD receiver could, with further work, be estimated prior to design. This section therefore first aims to investigate a single contributing factor to the receiver ISI. As OOK data can be modelled as a random data sequence, the ISI only due to the step response can be investigated by looking at the $0 \rightarrow 1 \rightarrow 1$ data transition. Conceptually this will be worse if the dead time is greater than the symbol period. The model developed here therefore aims to:

- explain and fit to the transient issues seen in experimental data, and
- assess the impact of the dead time, in the context of the scalability of the technology to faster data rates. This is important as:
  - the dead time may be fundamentally limited to the nano-second range by minority carrier effects [76], or
  - the diode capacitance and quenching resistance time constant may be difficult to overcome [49].

The removal of SPADs from the pool, and the reduction in detection probability, leads to decreased output counts despite continuous input levels. Some time later, the detectors that were initially removed will be recharged back into the available pool. The model makes a number of assumptions, however fitting to measured results and correct estimation of long term steady state values, indicates the assumptions hold for this device. The model assumes:

1. the array is globally set to its most sensitive mode, $N_a = 4$,
2. the input counts prior to the step change are exactly zero,
3. that the clock frequency of the pipelined readout scheme, is initially equal to the detector dead time, allowing the model to use discrete time steps, n and $n+1$ etc (*this is later expanded to integer multiples and eventually continuous time*)... and,
4. that the SPADs break down at the start of the symbol integration.

Each of these assumptions can be tackled with further work, adding complexity and second order effects as required to fit new experimental results. Here, only simple first-order effects were considered, limiting the accuracy of obtained numerical conclusions. However due to close fitting with experiment (Figure 6.10), the model allows general array size and dead time conclusions to be made.

6.3.1 The Detector Pool Model

The received output counts $C_{out}(n)$ at a given discrete time n, is given by Equation (6.1).

$$C_{out}(n) = ePDE(n)C_{in}(n) \tag{6.1}$$

Where $ePDE(n)$ is the effective detection efficiency. PDE_0 used in Equation (6.2), is the initial photon detection efficiency, given by the SPAD PDE, the fill factor and number of devices, N_{array}. The instantaneous number of devices that are available, N_{avail}, will satisfy the condition $N_{avail} = N_{array}$, when $n = 0$.

$$ePDE(n) = PDE_0 \left(\frac{N_{avail}(n)}{N_{array}} \right) \tag{6.2}$$

As this is a recursive model, N_{avail} will be dependent on the number of available detectors in the previous time steps. The step response model can be written as Equation (6.3).

$$N_{avail}(n) = N_{avail}(n-1) - C_{out}(n-1) + C_{out}(n-2) \tag{6.3}$$

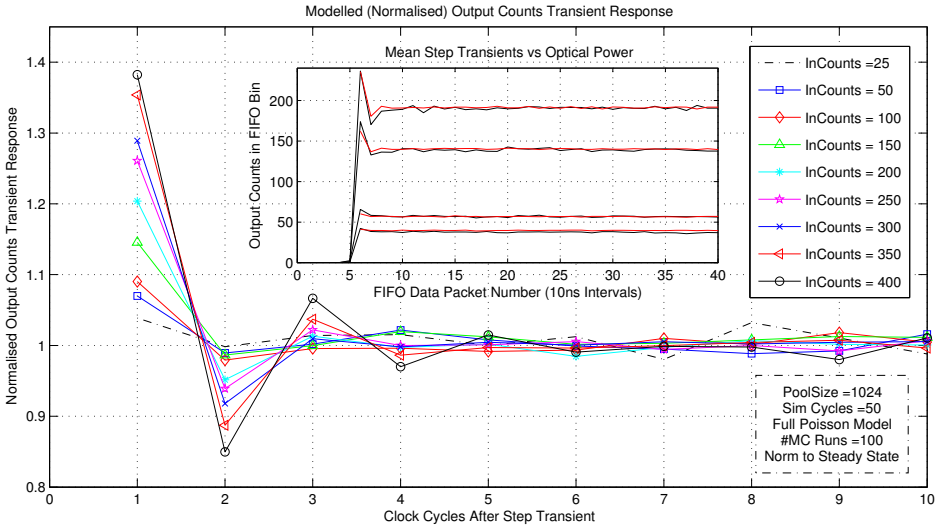


Figure 6.10: A modelled set of step transients with increasing optical flux. Each Monte Carlo run, including Poisson photon variation, is averaged and normalised to the steady state mean value. Inset, the detector pool model is shown, fitting experimentally obtained (non-normalised) step transients.

In Figure 6.10, the discrete time model has been run using a Monte Carlo technique (including Poisson variation of incident photons) and plotted for the first 10 readout cycles after the input step. In the numerical examples using this model along with the sections below, the number of SPADs is 1024 matching the SPAD-array developed in Chapter 4, the dead time is set to 10ns approximating the measured average AQAR dead time, and the readout clock

frequency is set to 100MHz matching the silicon chip's maximum output data rate and input clock frequency. Depending on C_{in}, the settling time and amplitude of count variation increases. Figure 6.10 has been normalised to steady state conditions for each model run, in the same manner as Figure 6.6. However a fitting to experimental data is shown non-normalised for clarity (Figure 6.10.inset). As can be seen by these inset experimental results, this simple discrete time model fits the observed step response in that it replicates the magnitude of the initial peak and the long-term steady state with good accuracy. This is despite the natural Poisson noise included in the experiment. To investigate the closeness of fit between the model and the experimental results, the residual between the two was calculated over time. This gave a residual range of ≈ 18 counts. In comparison, the Poisson noise of the experiment was approximately 15 counts. This shows that for the non-normalised experimental results, the closeness of fit is comparable to the natural optical noise of the experiment.

6.3.2 Estimation of the Steady State Solution

The steady state solution is obtained when N_{avail} remains constant (Equation (6.4)). This can be evaluated by a rule of thumb based on the initial increase of the input signal to $C_{in}(0)$ (denoted as the simple-model) (Equation (6.5)) or estimated accurately using the long term average input counts $\langle C_{in} \rangle$ after the transient (denoted as the α-model) (Equation (6.6)).

$$C_{out}(n) \approx C_{out}(n-1) \tag{6.4}$$

$$C_{out}(n) \approx C_{in}(0)\left[1 - \frac{C_{in}(0)^2}{N_{array}}\right] \tag{6.5}$$

$$C_{out}(n) \approx \langle C_{in} \rangle \left[1 - \frac{\langle C_{in} \rangle}{N_{array}}\left(1 + \alpha \frac{\langle C_{in} \rangle}{N_{array}}\right)\right] \tag{6.6}$$

An estimation of the steady state, and the manner in which the receiver responds, allows understanding of detector affects during design and specification. For example, with estimates for photon numbers required for communications, a further estimate on the number of SPADs per receiver can be used as a design constraint.

The simple-model holds for low input counts as a proportion of the total number of SPADs ($\langle C_{in} \rangle < 200$, where 200 input counts is $\approx 19.53\%$ of the $N_{SPAD} = 1024$ diodes within the receiver front-end), but departs due to the use of initial photon number (Figure 6.11). Likewise, estimates using the long term input count average also suffer from a similar departure. The fitting factor α, can be obtained iteratively, however $\alpha = 1$ still gives a maximum deviation of only 17 counts, at $\langle C_{in} \rangle = 400$, approximately equal to the expected Poisson noise. By minimisation of the residual, $\alpha = 0.725 \approx \sqrt{2}$, yielded a minimum residual of 1.1 counts at $\langle C_{in} \rangle = 250$, as shown in Figure 6.11.

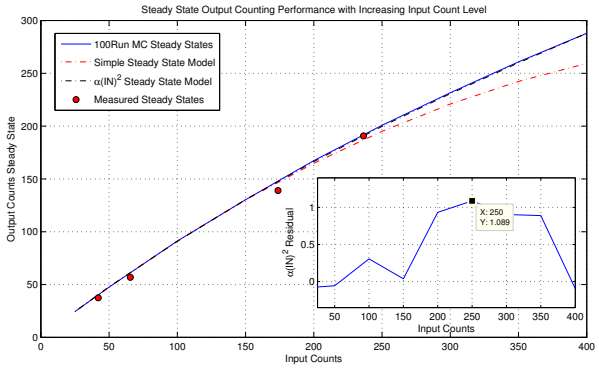


Figure 6.11: The detector pool model steady state value, along with measured results and two steady state value estimation models. Close agreement is seen between the model and the measured results, while the simple estimation model departs at higher incident photon levels. The α model fits accurately as evidenced by the inset residual.

6.3.3 Tending Towards A Continuous Time Detector Pool Model

The advantage of the full model, over the steady state estimation, is that it allows assessment of under- and over-shoot observed during step transients. For a complete understanding, a continuous time model requires inclusion of the dead time distribution (Section 5.2.2), and the arrival of photons throughout the symbol.

The model can be converted into continuous time as discrete samples occur at spacings of Δt. In this case $n = t$ and $n-1 = t-\Delta t$. The sampling period leading to Δt, is a product of the readout frequency, F, hence $\Delta t = 1/F$. The probability a SPAD has recharged, a time t after it has broken down can be defined (Equation (6.7)), using the cumulative density function (CDF), with the assumption that the dead time distributions display Gaussian behaviour (Section 5.2.2).

$$P(T \leq t) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{(2/F) + \mu(\tau_d) + t}{\sigma(\tau_d)\sqrt{2}}\right)\right] \tag{6.7}$$

The continuous model can be computed using Equation (6.8), where the first term models recharging SPADs. The second term represents output counts, reducing the available detectors each time a photon arrives and $C_{out}(t)$ represents the instantaneous output. This is a time sweep, with photons removing detectors from the *available* pool, but which are only returned once the time axis reaches a point where there is a high or unitary probability they have recharged.

$$\frac{\mathrm{d}\left(N_{avail}\right)}{\mathrm{d}(t)} = \int_{t-\frac{1}{F}}^{t} P(T \leq t)\, C_{out}(t)\, dt - \int_{t}^{t+\frac{1}{F}} C_{out}(t)\, dt \tag{6.8}$$

6.3.4 Use of the Detector Pool Model

Despite the assumptions, the discrete time detector pooling model achieves good agreement with measurements in Figure 6.10.inset and Figure 6.11. The step response directly contributes to inter-symbol interference, modifying the receiver total detection efficiency. An assessment of the response during design specification, would allow determination of the number of SPADs to be implemented per receiver. It would also couple to the dead time, impacting both the magnitude of the dead time and constraints on its variation. This model, and the concept of a pool of available detectors, can be used to minimise the step response and the inter-symbol interference caused by it, reducing the overall bit error rate.

6.3.5 Optimum Dead Time and Bit Integration

This analysis suggests implications for SPAD dead times. From the initial design stage, the symbol period is fixed by the data rate. Therefore to prevent inter-symbol interference, caused by the step response, the inequality, Equation (6.9), must be satisfied. To explore this, two cases can be used,

- A dead time longer than the integration period, $\tau_d \geq \tau_b$, or
- An integration period longer than the dead time, $\tau_b \geq \tau_d$

$$\tau_b = \mu(\tau_d) \quad and \quad \sigma(\tau_d) << \mu(\tau_d) \tag{6.9}$$

SPAD Dead Times Longer Than Symbol Periods

Section 6.4 will show communication results between the μLEDs and SPADs, however the impact of the step response can be assessed by thinking of specific transitions within random OOK data. Clearly the step response has little impact on $0 \rightarrow 0 \rightarrow 0$ transitions, and the $1 \rightarrow 0 \rightarrow 0$ transition, as discussed, is influenced by the receiver impulse response. The step response can however be linked to transitions from a low count level to a high count level such as a $0 \rightarrow 1 \rightarrow 1$ transition. Here the step response will give rise to a first 'one' bit that is close to the actual number of photons within that symbol. This is effectively the peak of the step responses shown in both experimental work and the discrete time model (Figure 6.12). The second 'one' however becomes a significant issue for the BER. Here the incident photon flux should represent a 'one' however the step response has decreased the sensitivity of the array (Equation (6.2)), decreasing the actual number of detections significantly. This will have two effects, firstly that the ideal N_1 distribution will be split into 'ones' that follow a 'zero' and 'ones' that follow previous 'ones'. Secondly, as the received signal of some 'ones' is reduced, the gap between the N_0 and N_1 distributions will be reduced. This leads to an increased error rate, as discussed in Equation (3.1) and shown in the sensitivity analysis of Table 3.1.

In order to obtain an optimal step response, Equation (6.9) states that the dead time should be equal to the symbol time, however symbol periods less than τ_d may result in some designs. This would occur especially when large, passively quenched diodes are used, or where designers aim to scale the bit rate faster than the proposed limit [76]. Exploring this, a symbol less than the dead time would lead to worse symbol to symbol variation and therefore increased ISI.

In the discrete model (Equation (6.3)), SPADs return to the available pool after one time step (as dead times equal to the symbol period are assumed for simplicity). For a dead time an integer multiple of the symbol period, $\tau_d = x_{recharge}\tau_b$, the number of devices recharging is $N_{recharged} = C_{out}(n - x_{recharge} - 1)$. In other word, SPADs do recharge back into the pool, but some time later than the $C_{out}(n-2)$, previously used. As an example, a symbol a quarter of τ_d can be used. Upon a step, the *ePDE* would decrease for four cycles before beginning to increase. This increases the depth and number of subsequent bits affected by the step response trough. Assuming 1024 SPADs and a step giving 150 initial counts, the effective efficiency would follow the series: $100\% \rightarrow 85.3\% \rightarrow 72.85\% \rightarrow 62.2\% \rightarrow 67.7\% \rightarrow 70.3\%$.

A step transient, allowing assessment of its contribution to the multi-component ISI, is modelled (Figure 6.12) with 100 photons incident on the array. This figure uses an array of 1024 SPADs and a SPAD dead time of $\tau_d = 10$ns. The time step is chosen with each marker representing a different bit, while the model is intended to investigate only the contribution of the step response to the bit to bit ISI. At 50Mb/s a symbol of 20ns is longer than the dead time, giving only a slight reduction in output counts. Decreasing the data period to 5ns symbols (equivalent to 200Mb/s), the time to steady state and number of bits within the step response ripple increase. The steady state value decreases by $\Delta = 90.23 - 79.99 = 10.24$, i.e. a factor of -11.34%, indicating a split in the N_1 distribution, matching results in Section 6.4. With 1ns symbols (equivalent to a data rate of 1Gb/s) significant variation in the output is observed. This increase in severity is linked to the fixed SPAD dead time. As will be seen below, the dead time must scale with the bit period to prevent this increase in ripple severity.

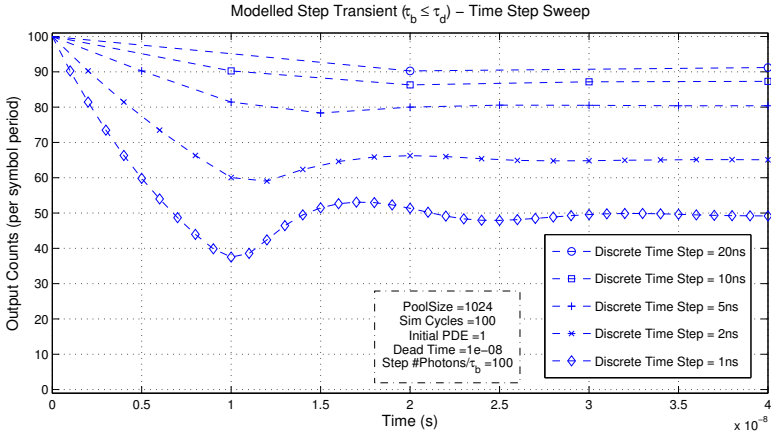


Figure 6.12: The modelled step response under a variety of simulation time steps, which act as proxies for different data rates (assuming OOK). The dead time is fixed at 10ns for all traces (100MHz max count rate per SPAD), with 1024 SPADs. As the time step or clock period decreases, the number of bits (in this step response scenario) affected by ISI increases, as does the amplitude of the under- and over-shoot. Overall, to reduce the impact of the step response, the dead time must scale with the clock period, ensuring that SPADs have recharged by the start of the next integration. While not shown, the trace for a 1ns discrete time step, improves significantly to a steady state of 87.2 counts (up from 49.8 counts) and only 4ns settle time, when the dead time is reduced to 1ns.

Figure 6.13 shows that increasing the number of photons in the step, significantly reduces the steady state value in comparison to the initial input count value and increases the amplitude of the under- and over-shoot. This, as shown below, is linked to the number of SPADs within the array, and the dead time of those SPADs. For a well controlled step response, the incident optical power should be as close as possible to the quantum limit (Table 3.2 and Table 3.4), i.e. zero or few power penalties other than the fill factor and photon detection efficiency power penalties.

Before moving onto modelled sweeps of the number of SPADs and dead time of those SPADs, an assessment can be made as to the effect of the step response on the N_1 distribution. In Figure 6.14, the time traces of step responses similar to Figure 6.13 are converted into histograms. The model parameters are slightly different in that the time step is set to 10ns to match the readout speed of the SPAD-based receiver. As the photon flux increases, the steady state decreases away from the initial peak number of detections. This splits the distribution into two sub-distributions, severely degrading the BER performance through a lower N_1 to N_0 separation, and a significantly increased distribution standard deviation (σ_1). For example at low step flux levels (e.g. 50photons/symbol), the received N_1 distribution is approximately a single Gaussian ($\mu = 47.6$, $Median = 47.3$, $\sigma = 6.9$). In comparison, at high flux levels (250photons/symbol), the distribution has become split into a small peak around 250, corresponding to the step response peak, and a larger lower peak at ≈ 181, corresponding to the steady state value (overall $\mu = 199.3$, $Median = 188.2$, $\sigma = 32.6$). This splitting behaviour is

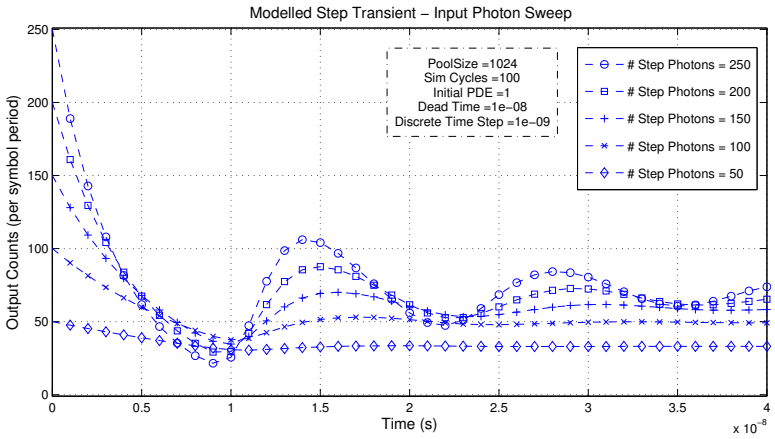


Figure 6.13: Modelled step response with a time step of 1ns and a sweep of incident photons. The dead time is fixed at 10ns. As the number of incident photons increases, the severity of the step response also increases. The number of SPADs and the dead time of those SPADs must scale to mitigate the step response. For example if the 1024 SPADs used here is increased to 2048, the 100 step photon steady state changes from 48 to 67, while the time to steady state remains the same. Likewise if the dead time is reduced to 1ns, the steady state increases to 87 counts, while the time to steady state decreases to 3ns. It would appear then that reducing the dead time is preferable, as it acts to improve both the steady state and the time to steady state, while increasing the number of diodes does little to the time to steady state. This is discussed through the next figures.

shown to occur experimentally in μLED communication experiments (Section 6.5.2).

Moving back to sweeping the model over other parameters, and using the previous time step value of 1ns, in Figure 6.15 and Figure 6.16, the number of SPADs within the array, and the dead time of those SPADs are swept respectively. Holding all other parameters constant, an increase in the number of SPADs, or rather an increased ratio of available SPADs to used SPADs, is beneficial as the ripple becomes less pronounced and the steady state increases back towards the 100 count/period ideal case. The step response model can be used to estimate the size of the SPAD array needed for a future design. If a criterion is applied such that the steady state is within the natural Poisson variation of the N_1 distribution, no splitting of that distribution would be observed, although the distribution may still be wider than the Poisson limited case. Using the step response model, and an input step value of 100 counts/symbol, a steady state of $100-\sqrt{100}=90$ counts/symbol is required. This is achieved with a SPAD-array size greater than 9500 SPADs. The nearest binary value, assuming a square array, is 16384. This suggests future development of very dense, high fill-factor, low dead time arrays above 128x128. To date, arrays such as this have been achieved, such as [160], which incorporated a 340x94 = 32640 SPAD array, with 70% fill-factor and a low mean DCR of 2.68kcps.

In Figure 6.16, all of the parameters remain the same, such as an array of 1024 SPADs. However here the SPAD dead time is swept between 1ns and 20ns. As the dead time decreases, the step response again improves, with the steady state increasing towards the ideal 100counts/period

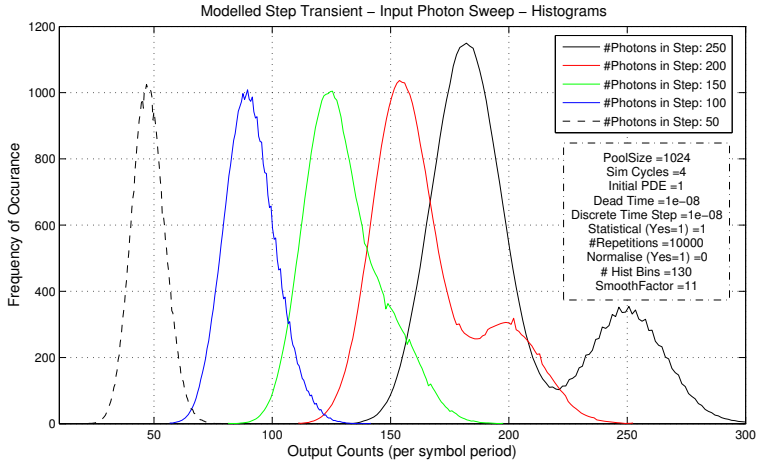


Figure 6.14: For a modelling scenario close to the experimental silicon receiver, the above time model with a sweep of the number of photons/symbol, is run with a statistical model. This allows the form of the N_1 distribution to be assessed. In this figure, the pool size of 1024 SPADs, the dead time of 10ns and the values of the number of input photons/symbol, are kept the same. The figure uses a time step of 10ns, matching the silicon readout clock frequency of 100MHz.

level and the magnitude of the ripple also decreasing. In comparison to the sweep of the number of SPADs however (Figure 6.15), the time at which the first step response minimum occurs is reduced significantly. This points not to an argument of either increasing the number of SPADs or decreasing the dead time, but to the concept of increasing the overall counting resources, measured as $C_{resource} = C_{max} = \frac{N_{SPADs}}{\tau_d}$, in comparison to the input light level. The need to have a large number of SPADs, or a large overall counting resource, was suggested previously in Section 4.5.7 and a short dead time or shortened pulse was suggested in Section 4.5.8.

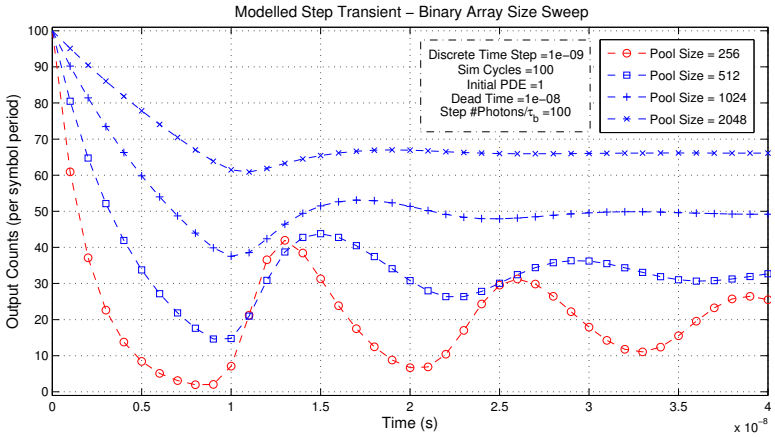


Figure 6.15: Modelled step response with a sweep of the number of SPADs within the array. Here the time step is 1ns, the dead time is fixed at 10ns and the number of photons within the incident optical step is 100 photons/symbol. As with the symbol period sweep, the dead time must scale in complement with the symbol period. Here a dead time longer than the symbol has purposely been chosen to show an initial worst case. To show consistency between the figures, the 1024 SPAD trace in this figure is the same as the 1ns discrete time step trace in the above symbol period sweep figure. As the number of SPADs increases to 2048, the severity of the ripple decreases and the steady state value increases. Conversely, if the number of SPADs is decreased to 512 or 256, the step response becomes very severe and the steady state decreases below 40% of the actual photon input.

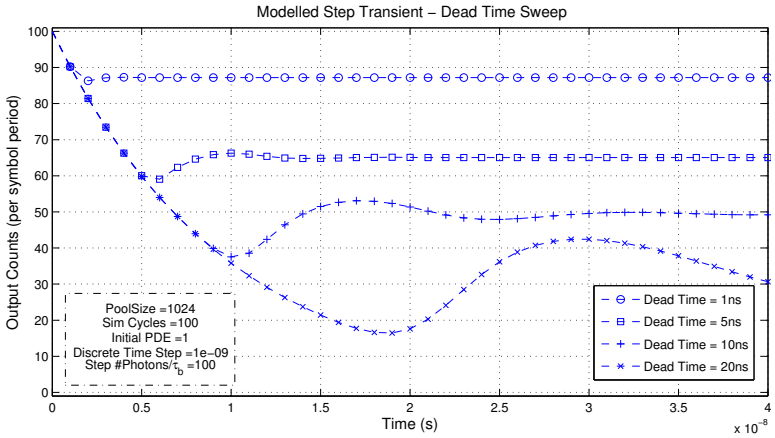


Figure 6.16: Modelled step response with a sweep of the SPAD dead time. Here the discrete time step is 1ns, the SPAD pool size is fixed at 1024diodes and the incident optical step is 100photons. As the dead time decreases to 5ns and 1ns, the step response improves, both in terms of the steady state value reached, and the time to settle. As the dead time increases to 20ns, perhaps for diodes with a larger parasitic capacitance, two interesting phenomena occur. Firstly, the position of the first minimum extends outwards in time, reaching a minimum of 16 counts at 19ns. Secondly, the steady state decreases from the ideal 100photons/symbol to just 33 counts. The best option for a stable step response, that is a small proportion of the data symbol period, is therefore to design an array with both a large number of SPADs and low dead times. Of course to achieve an overall optimum receiver performance, these step response requirements must be balanced against the optical fill factor and the practicalities of reducing the dead time.

Symbol Periods Longer Than SPAD Dead Times

If τ_b is chosen to be larger than τ_d (Equation (6.9)), then a SPAD can detect multiple photons per symbol. The positive effects this would have are:

- an increase in overall receiver sensitivity,
- a reduction in step transient under- and over-shoot, and
- steady states closer to the initial value.

However, symbols longer than the dead time would limit the receiver readout, yielding lower data rates. Further, scalability of the technology to higher data rates is reduced. This would again be due to the fundamental limit on SPAD dead times [76].

An Engineering Optimum Solution

While dead times may scale to lower values, the two cases indicate a trade-off. As receiver design maximises both data rates and sensitivity, a dead time equal to the symbol period should be chosen. This allows the dead time to scale at the same rate as the receiver readout frequency, while giving the best step response for that frequency. As data rate scaling is a high priority [11], the previous case of symbols shorter than or equal to the dead time is preferable. As shown however, this will require careful modelling to compensate poor step responses with correctly chosen numbers of SPADs or SPAD dead times.

This may be a clear result, however it may be useful for designs attempting to push the sensitivity, speed or error rate. If, for example, the dead time were to be fundamentally limited at 1ns [76], the data rate may not need to be limited to 1Gb/s. A data rate of 1–2Gb/s might be supported if modelling and further experimental results showed that an increase in photon counting resources, (N_{SPADs} x $\frac{1}{\tau_d}$), needed to obtain the specifications, could be achieved via an increase in the number of SPADs. In effect this would use a different parameter in the step response model, with the dead time fixed, to compensate the step response caused by that dead time. For example, as shown by Figure 6.13, reducing the other power penalties so that the input optical power is as low as it can be, can reduce the severity of ISI and as shown by Figure 6.15, the same reduction in step response severity could be achieved with an array of more SPADs. This tactic of using modelling to influence a second generation would be a similar approach as the modelling used in SML diodes to estimate the number and width of p/n regions or the modelling of equalisation techniques that aim to remove ISI prior to clock and data recovery.

6.3.6 Design Conclusions Drawn From Detector Response Modelling

The model of the step response can be used in the initial modelling and specification of a second generation receiver. This will use the measured and modelled insights gained to further constrain the values of the number of SPADs in a receiver along with the dead time of SPADs. Likewise, prior to silicon implementation, a modelled second generation could be tested for application suitability, where trade offs between sensitivity and bandwidth could be assessed for the specific situation in which the receiver will be used. Overall, the model points to use of a large number of diodes, each with a dead time equal to or shorter than the symbol period. Likewise it indicates that the maximum number of detections per symbol should be a small proportion of the total number of diodes. This allows any step response steady state to be close to the initial number of detections, decreasing both the N_1 distribution variability and the BER.

Now that transient behaviour has been established, and modelling has pointed to system optimisation, communication experiments can be performed. This is the subject of the next section, exploring the receiver under various conditions. The section directly establishes the effect that the transient responses have on this first generation receiver.

6.4 Received Data Streams Using a GaN Micro-LED

In this section, waveforms and histograms of experimentally received data are presented, aiming to prove the concept behind the presented receiver front-end. Overall, data streams are received, however the μLED and front-end transient responses both contribute to bit errors when operated above $\approx$ 25MHz. The experiments used the SPAD-based receiver, with no colour filter, and with a small 34μm 450nm μLED. No filtering was included, as preliminary tests aimed to ascertain only the performance of the silicon device. In VLC receiver implementations aiming for commercialisation, narrow wavelength optical filters can be used to allow colour separation in WDM systems or to remove ambient light contributions. However, such optical filtering adds expense to a system. In ambient light experiments performed here, the influence of broad-spectrum ambient light needed to be found in order to assess if costly narrow-λ filters were necessary. The μLED was biased at 3mA, producing a narrow-band DC optical power of $20\mu W/cm^2$ at the SPAD receiver. The μLED has a bandwidth of $\approx$ 220MHz [24]. As this is close to the 100MHz modulation rate, an eye opening penalty was incurred due to RC charging and discharging [14, 36] (Section 3.3.5). This has reduced the modulation in photon detections and caused μLED induced ISI, prior to the influence of receiver ISI (Section 3.3.5). The direct mapping between 100% electrical extinction ratio and the μLED output optical extinction ratio is unknown due to the short experimental period available. However to counteract changes in the modulation depth, the SPAD-array recorded eye opening has been investigated and the eye opening power penalty, PP_{EOP} is used in the estimation of the BER and sensitivity (Section 6.5.1). The bandwidths of the testing μLEDs force the error rates to be total link error rates, rather than that for the receiver only. This makes determination of the BER performance of the receiver difficult to resolve, and therefore prevents full assessment of the combined avalanche, integrating, multi-diode receiver. As such, other approaches to estimate the BER are presented.

A bit error rate tester (BERT) applied a non-return-to-zero (NRZ) on-off key (OOK) modulation, with an electrical modulation depth of $2V_{pp}$ (100% electrical extinction ratio), directly onto the μLED. This was similar to the technique used in [23, 24], but removed the CMOS drivers [23] to increase the bandwidth. A pseudo-random bit stream (PRBS) of length 2^7-1 is used for testing and is repeated during the experiment. The light from a single 450nm μLED was focused onto the receiver. This was placed $\approx$ 1m away using absolute dark conditions or a standard desk lamp to simulate low level ambient light.

As the SPAD-array incorporates a number of configurable factors, the receiver was setup in its single-channel 1024 SPAD mode, with an active quenching active reset dead time of 13.5ns ($DTCTRL = 0$mV). The readout clock was set to 100MHz throughout all the experiments in this section and the temperature of the testing PCB was monitored. The SPAD pixels were globally set to the $N_a = 4$ mode, no high DCR devices were turned off and the SPADDVDD and VBD voltages were held static. Each of these parameters could be investigated with respect to communications in future work, however in the initial experiments shown here, each is fixed allowing assessment only of the μLED and SPAD-array responsivity and transient issues.

In order to obtain the communication results within this section, the photon arrivals per clock period were output via a testing FPGA and saved for off-line analysis. The data was also thresholded and re-timed within the FPGA and provided back to the BERT for real-time comparison with transmitted signals. Experimental issues such as clock stability, synchronisation and equipment issues prevented the BERT from obtaining a correct lock to the returned data, prompting the use of estimation techniques for BER determination [3].

6.4.1 Zero Ambient Light and High Count Modulation Depth

A high received photon count modulation depth and zero ambient light, provides a useful proxy for constrained optical links using optical fibres or VLC without ambient light. The optical power modulates the photon detections about the middle of the SPAD receiver photon transfer curve, ensuring minimal intensity distortion due to the receiver DC responsivity (Section 5.2.3).

A voltage eye diagram is unsuitable for this receiver due to the 14-bit digital output, however the digital eye diagram at 10Mb/s is shown in Figure 6.17. Here, the eye is open but a number of samples show inter-symbol interference due to the impulse and step responses, while the μLED bandwidth (rise/fall time of 1.45ns) will also cause ISI. The combination of these is shown by the slow turn off ($1 \rightarrow 0$) transients. The eye diagram (Figure 6.17) also shows a number of incomplete transitions, such as the partial $1 \rightarrow 0 \rightarrow 1$ transition at $t = 20$ns.

In Figure 6.18, a 50Mb/s stream shows the repeatability of the transmitted data and a clear gap between N_0 and N_1 distributions. There is agreement between the transmitted and received data waveforms. The histogram indicates wide distributions showing that it is the large modulation in counts that enables the bits to be received.

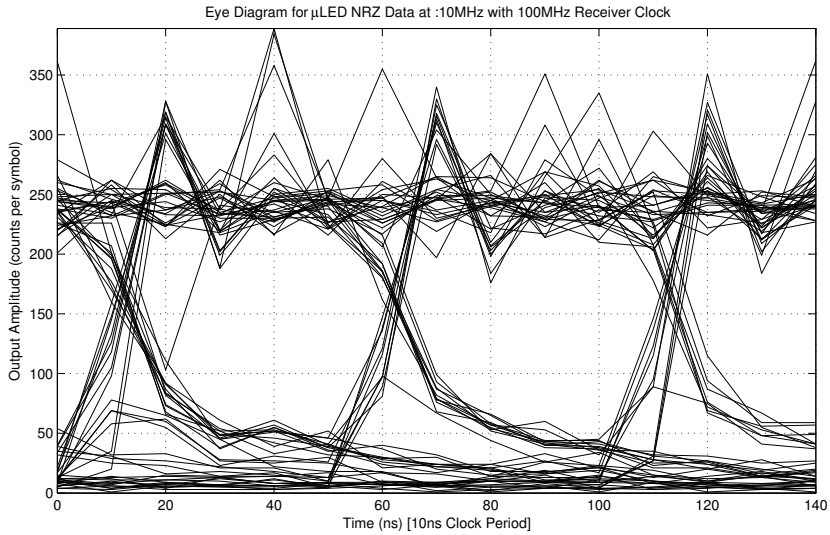


Figure 6.17: Output digital eye diagram for 10Mb/s NRZ OOK PRBS data with a 100MHz receiver clock. The eye remains open despite influence from the μLED bandwidth and the impulse and step response of the SPAD-based receiver. As previously modelled, the step response is responsible for increasing the ≈ 250 count/sample mean value's variation (Poisson variation of $\approx \sqrt{250} = \pm 15.8 = 31.6$), towards a variation of $390 - 190 = 200$ counts.

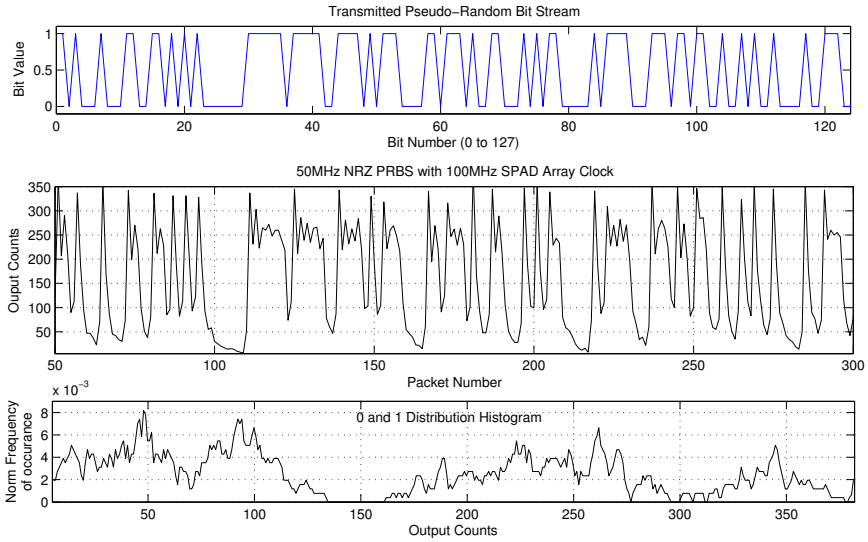


Figure 6.18: Output data using 50Mb/s NRZ OOK PRBS data and a 100MHz system clock. The top two plots show a 1:1 correspondence between transmitted and received data. The lower plot shows a normalised histogram of the counts, indicating the multi-modal nature. The SPAD receiver TDM pixels were each set to the $N_a = 4$ mode.

For N_0, an increase in the width of the distribution could be caused by both μLED bandwidth and SPAD receiver impulse responses. For the N_1 distribution, the under- and over-shoot of the step response causes a separation of the distribution into two. The higher distribution ($\mu(counts) = 340$) may be due to the initial peak in the step response. While the lower distribution ($\mu(counts) = 250$), may be due to the steady state reached just before the data transitions. The 50Mb/s eye diagram (Figure 6.19), shows a closing of the eye, in comparison to Figure 6.17, along with worsening of the transient mid level and $0 \rightarrow 1$ distributions.

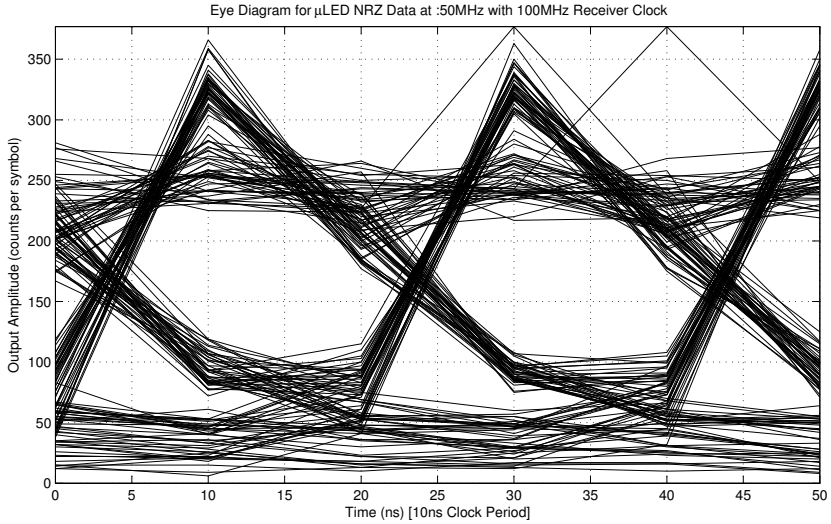


Figure 6.19: Output eye diagram for 50Mb/s data with a 100MHz receiver clock. The eye has closed due to influence from μLED bandwidth and receiver transient response.

At 100Mb/s (Figure 6.20), an increase in errors is observed. An example of this can be seen at $x = 72$ in Figure 6.20. This is not because of N_0 and N_1 overlap, but due to three effects. The first and second are that the impulse and step responses have rendered incomplete transitions. The third however, is that the peakiness of the $0 \rightarrow 1$ transition has increased, along with a reduction of the step response steady state (shown between packets 45 and 52). Looking at the histogram, the upper peaks have reduced in amplitude, bringing the peak down from ≈ 350 to ≈ 300. The peaks caused by the step steady state and the incorrectly transitioning 1 bits have combined, giving the main N_1 peak at 175 counts.

Figure 6.21, shows the eye diagram at 100MHz transmission clock. Clearly data thresholding into accurate data will be difficult due to the mid level distributions and the increase in N_0 mean (compared to the N_0 mean at 10Mb/s) due to the tails of the μLED response. In Figure 6.21, the middle of the eye diagram is shown at $t = 10$ns, however as the symbol rate is equal to the sample rate, the preceding and proceeding bits are at $t = 0$ns and $t = 20$ns respectively.

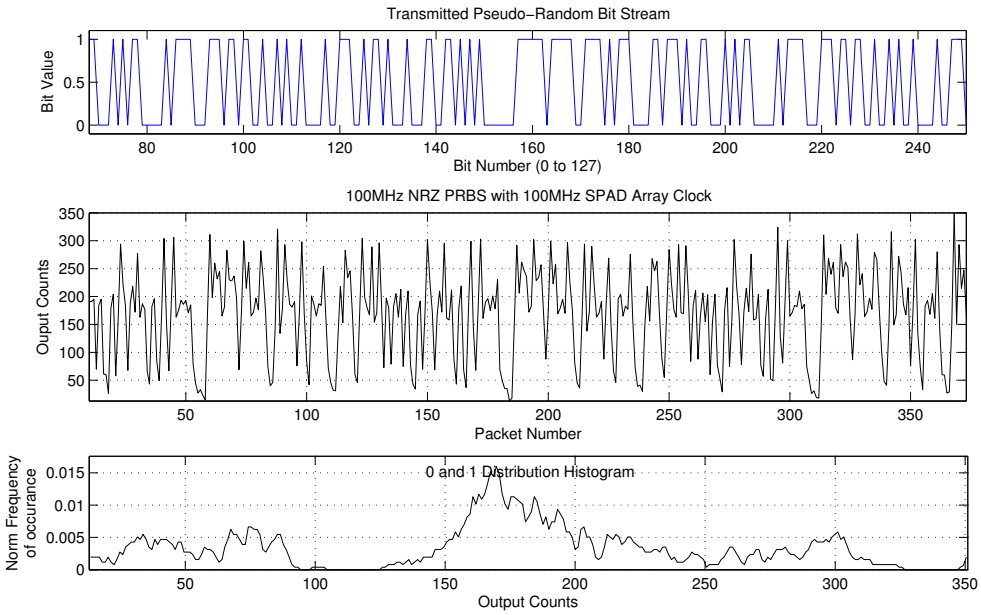


Figure 6.20: Output data using 100Mb/s data and a 100MHz system clock. Some correspondence remains between the transmitted and received data streams, however more bit errors exist. From the histogram, incomplete transitions and significant step response ripple, gives a large distribution near the threshold value.

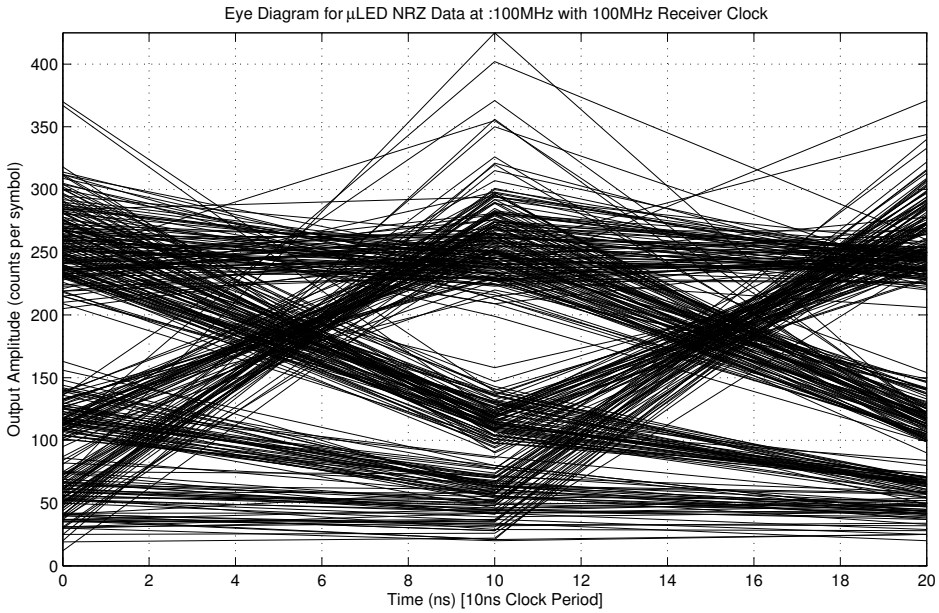


Figure 6.21: Output eye diagram for 100Mb/s data with a 100MHz receiver clock. The eye has closed indicating significant bit errors.

Transient Analysis from Received Data Streams

Figure 6.22 shows a subsection of bits received at 50Mb/s. Transient behaviour already linked to the impulse and step responses, is clearly shown at bit transitions. With the receiver clock at 100MHz, the bin-width is the same as the transient analysis experiments, allowing identification of features in the received waveform and their attribution to measured effects.

Initially, a $0 \rightarrow 1$ transition is observed. This shows the cycle to cycle step response (Section 6.2.5). The noise in counts at the top of this first bit is a product of the step response, Poisson photon arrival noise and noise contributions from the μLED. Each of the $0 \rightarrow 1$ transitions, show the high initial peak and lower subsequent counts attributed to the removal and subsequent return of SPADs into the detector pool (Section 6.3.1).

Shown at $1 \rightarrow 0$ transitions, the decay after a high power is visible and is over the same time scale as the impulse decay constant. Due to some oversampling used during the communication demonstration experiment, samples at $x = 124$ and $x = 166$ may be taken during the transitions of the μLED source. These show a small number of counts received prior to the full height peak, but could also be caused by synchronisation issues between the transmitter and SPAD-based receiver.

The initial peaks, at the start of the step response, fall at similar values throughout Figure 6.18 and Figure 6.22. This can be seen in the histogram for the 50Mb/s data, where these have created a separate peak at 345 counts. This indicates that the reduction in counts during the step transient is creating the majority of the N_1 distribution variability between 160 and 275 counts in the histogram of Figure 6.18 and that this is severely increasing the BER. This can be reduced in further generations through application of the detector pooling model (Section 6.3) during design specification (Section 3.3). Reduction of the step response can be achieved using the $N_a = 1$ mode (Section 6.2.7), however this must be used sparingly due to the reduction in sensitivity (Section 3.4).

Despite a design targeting 100Mb/s operation at $BER = 1\text{x}10^{-9}$, the unexpected severity of the impulse, step and μLED responses clearly limit this receiver. As such, further work is needed to address these issues. Data transmission should use multiple μLED and receiver channels, each operating at speeds below this bandwidth limit [16, 17, 18, 20, 23, 24, 233, 234].

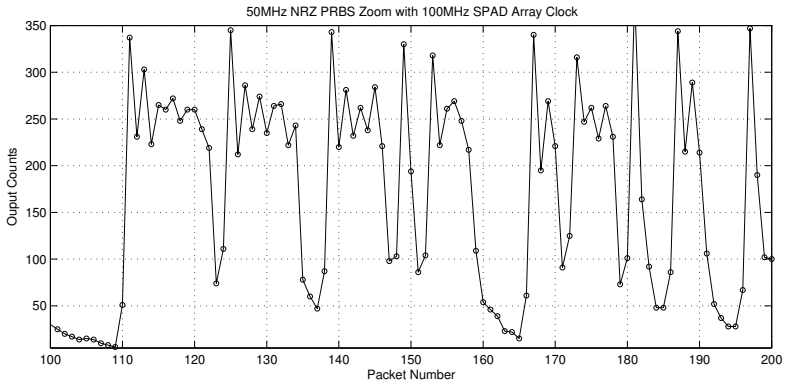


Figure 6.22: A zoom of output data using 50Mb/s NRZ OOK data from a single μLED and a 100MHz system clock to the SPAD-based receiver. The receiver pixels are all set to the $N_a = 4$ TDM mode, maximising sensitivity but demonstrating step and impulse response issues at the transitions between bits. The step response in particular has increased the N_1 distribution in excess of Poisson photon arrival variation.

6.4.2 Low Ambient Light Conditions with Low Count Modulation Depth

In the following experiments, a larger 450nm μLED of 84μm diameter has been used. This transmitter had a native bandwidth of $\approx$ 150MHz due to its increased capacitance. During initial experiments, a low modulation depth was used with a high optical power. This was used as these gave the maximum modulation speeds observed during previous μLED experiments [23]. Low power ambient light from a 40W incandescent desk lamp (0.5–1mW/cm^2) was also present to model broad-spectrum ambient lighting in free-space VLC environments. As no optical filter was used on the SPAD array, to ascertain if a narrow-wavelength filter was indeed needed, the ambient light as compared to the mono-chromatic 450nm modulated source must be multiplied by a) the predominantly red and near-IR emission spectra of incandescent lighting and b) the predominantly blue and near-UV photon detection spectra of the SPADs. With further work this can be measured experimentally, however here the background SPAD array count level is used as a proxy for this measurement. Unfiltered broad-spectrum ambient light represents the worst case for VLC systems, but is used primarily because the receiver must cope with other optical sources in the home/office lighting environment. As shown below, it is clear that some form of narrow-wavelength filter is needed.

Operation at 10Mb/s

The received data and histograms shown, can give an insight into the operation of the receiver in this worst case situation. In the first of the figures, data was received using a 100MHz receiver clock and an optical OOK data rate of 10Mb/s (Figure 6.23).

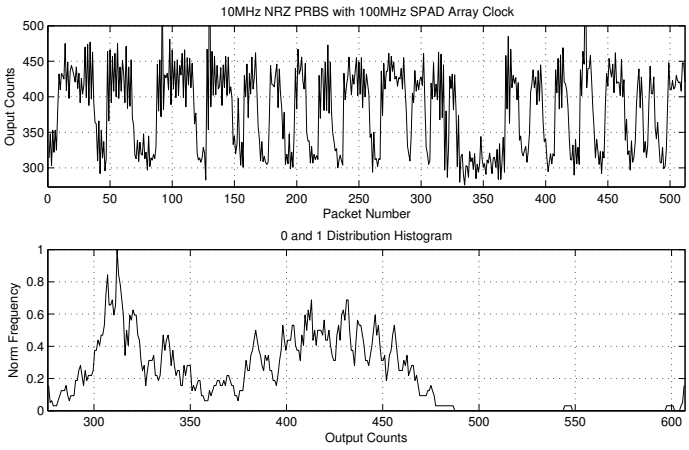


Figure 6.23: Output data using a 10Mb/s OOK data transmission and a 100MHz SPAD receiver system clock. In comparison to the zero ambient light case above, the N_0 distribution is much higher at the 300 counts per 10ns region, while the modulation in counts is now ≤ 100 counts.

Figure 6.23 shows the typical NRZ data stream that was transmitted, however there is a high degree of cycle to cycle noise within both zero (N_0) and one (N_1) bits. The histogram shows that there are two distinct peaks. However the level of zero bits is much higher than in the zero ambient light experiments (Figure 6.17). In the zero ambient light case, slow modulation (10MHz) gave zero levels at the 10 to 20 counts per sample level (Figure 6.17). In comparison the lower count level in the ambient light experiments at the same frequency, (Figure 6.23), is approximately 300 counts per sample.

Three effects have come into this experiment, forcing the distributions to be numerically close. Firstly, there is the high continuous power of the μLED, which is a combination of the high LED bias, the bandwidth of the LED and the non-zero mean transmission power of OOK modulation [1]. Secondly, there is the ambient light that has elevated the previously μLED extinction ratio limited lower count level. This contribution has added to both the N_0 and N_1 distributions and reduced the available SPAD-count modulation dynamic range (Section 5.2.3). The ambient light used here may be modest in comparison to home/office lighting requirements (≈ 500 lux), however it demonstrates the dynamic range problem associated with unmodulated incident light.

A third effect is that the high μLED and ambient optical powers, have pushed the modulation further up the SPAD array photon transfer curve and may have forced the receiver to operate near the saturation region (Section 5.2.3). The modulation depth of the transmitted signal

will be reduced by the gain compression [14] within this region, creating a further receiver sensitivity power penalty [14, 36]. Overall, the three effects are likely to have degraded the BER performance. However, as ambient light contributions are a direct product of the VLC application, this worst case, unfiltered experiment shows that in future receivers narrow-wavelength filters should be used to reduce SPAD counts from unmodulated light.

Operation at 50Mb/s

At 50Mb/s (Figure 6.24), the data becomes highly distorted. Some bits can be observed, as can pattern repeatability. The distribution of zeros and ones have overlapped forcing significant errors. The histogram, still shows two peaks, although the errors, optical source transitions and close mean values have disrupted the ideal statistics.

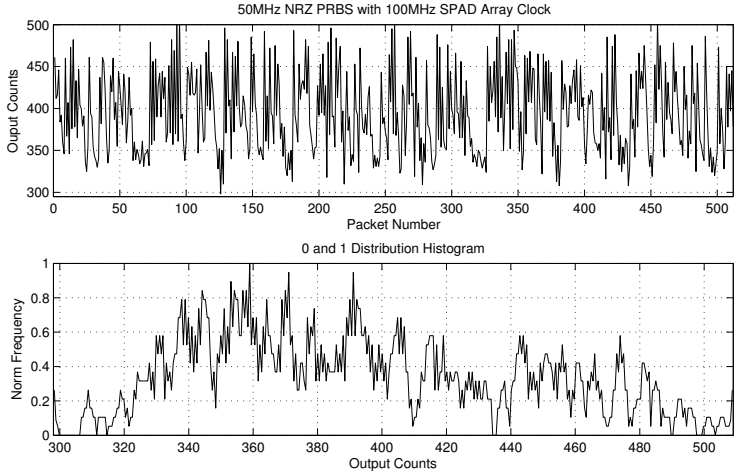


Figure 6.24: Receiver output data using a 50Mb/s OOK NRZ data transmission and a 100MHz SPAD-based receiver system clock. Bit to bit transitions are still visible, however the variability in the N_0 and N_1 distributions is now extreme enough to merge the two distributions. The simple thresholding scheme of OOK modulation would yield a significant number of bit errors under these circumstances, reducing the applicability of the receiver to ambient light, high speed applications.

Operation at 100Mb/s

At 100Mb/s (Figure 6.25), the two distributions have overlapped completely, giving a single broad distribution. The data trace shows some vestiges of the transmitted signal, however taking a threshold to reconstruct the transmitted signal yields the worst case BER of 0.5.

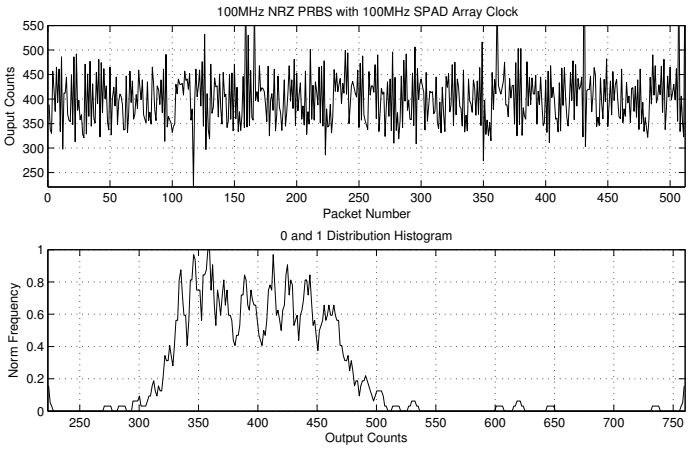


Figure 6.25: A received data stream using 100Mb/s OOK NRZ modulation of a blue (450nm) μLED and a receiver clock frequency of 100MHz. Unlike the 10Mb/s and 50Mb/s data streams, peaks and troughs reaching between 250 and 760 counts are evident.

6.4.3 Conclusions from Received Data Streams

In this section, small blue (450nm) μLEDs have been used to transmit OOK modulated signals to the SPAD-based receiver. Testing a variety of conditions including modulation frequency and the presence of ambient light, has shown distinct issues with this first generation receiver. These correspond to the measured impulse and step responses indicating a possible route forward when considering a second generation receiver for free-space VLC applications.

As per the zero ambient light experiments, despite design for 100Mb/s, the receiver response requires operation at slower rates. With knowledge from transient response experiments and appropriate models, further prototypes may tend towards the target performance, particularly if all crucial factors can be found, modelled and taken into account during prototype design. Further characterisation of the receiver, using μLEDs and bit error rate testing equipment, is needed to investigate operation near the bandwidth limit of the receiver.

In the above section, the BER is difficult to ascertain empirically [1]. This is due to experimental issues along with the severe transient issues demonstrated in the proceeding sections. As such, the next section investigates methods of estimating the BER of the receiver, allowing estimation of power penalties along with clarification of steps for future designs.

[1] Measured BER results in this work are within the estimated range of 10^{-5} to 10^{-6} as the project targeted a BER of 10^{-6} to 10^{-9}. Experimental issues prevented robust BER measurement, forcing estimation based upon the data sets obtained within this section.

6.5 The Bit Error Rate

For data transmission experiments, closure of the eye diagram due to the μLED bandwidth cannot be separated from the received data distributions. Likewise parameters such as the μLED relative intensity noise were unavailable. This reduces the prospect of reversing the power penalty budgeting (Section 3.3.5) concept to investigate the causes of receiver behaviour.

In this section, two methods estimate the receiver BER, aiming to take into account the μLED behaviour. While comparison of bit streams derives the BER empirically by comparing each received bit with the transmitted data, this is not always feasible [3]. For BERs such as $\leq 1\text{x}10^{-9}$, when the receiver is tested over multiple parameters such as frequencies, power supplies, light levels and modulation depths, the time to test becomes impractically long [3].

An initial estimate of the BER can be obtained using the Gaussian simple model discussed in Section 3.2 [2, 3, 14, 36, 103]. This is used below to assess the receiver only BER as a result of the photon transfer curve and continuous light sensitivity. The outcome of this, inclusive of estimated power penalties, indicates a high sensitivity. Despite this, the presence of transient issues within the received data, impacts the conclusions that can be drawn from this simple model. This therefore leads onto testing using a more accurate estimation, allowing multi-modal distributions with complex overlaps [3]. This can include transient issues and attempts again to obtain a BER estimate from a smaller data set. This model gives an insight into the cause of errors and is used to try to ascertain the receiver performance, independent from the single transmission device that was available.

6.5.1 Receiver BER Estimation Using A Simple Gaussian Model

For estimation, experimentally measured values for N_1, N_0, σ_1 and σ_0 are used with a sweep over optical power and system clock frequency. While the Gaussian model [2, 3, 103] does not separately include DCR or after-pulsing contributions, both are included in measurements. A 100% modulation depth is assumed, however the receiver power penalties such as the fill factor, PP_{FF} (Section 3.3.5), can be used to obtain a clearer receiver performance estimate.

To generate Figure 6.26, a similar setup to the PTC experiment (Section 5.2.3) has been used. This uses an unmodulated source which is filtered, creating 450nm light. The setup was modified to include a diffuser, iris and a highly defocused beam to ensure beam uniformity. The received distributions were generated from $\approx 400k$ samples each, for clock frequencies of 25, 50 and 100MHz. The light level was swept creating a distribution for each point on Figure 6.26.

This experiment is a simplification, as it removes the influence of extinction ratio and transmitter and receiver transient behaviour. The technique is not as robust as bit comparison experiments, however it offers a method of measuring native detector-only performance, without transmitter problems and the synchronisation issues experienced in the above μLED experiments. This approach, when combined with the power penalty budget (Section 3.3.5), the array

characterisation experiments (Chapter 5) and multi-modal BER estimation [3] (Section 6.5.2), allows investigation of how particular factors influence the BER.

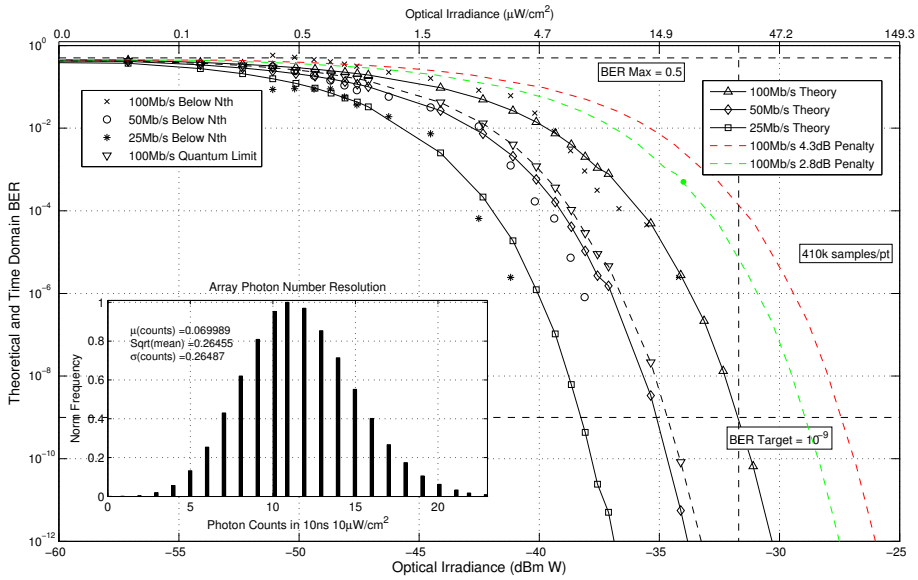


Figure 6.26: The Gaussian model estimate for the receiver-only BER over incident optical power and different system clock frequencies. The target BER of $1\text{x}10^{-9}$ from the project specifications, is estimated to be achieved at an optical power of -31.7dBm ($\approx 0.67\mu$W). The Gaussian estimation model is plotted along with the number of samples that dipped below the optimum threshold and the quantum limit (dashed). Although the model does not include receiver non-ideal behaviour other than the PDE, DCR and FF power penalties, the estimated receiver sensitivity can be used with receiver power penalty budgeting. Taking into account μLED and receiver temporal power penalties, the sensitivity curve is pushed towards higher optical powers (red) curve, while a 100Mb/s μLED experiment, gives the green dot, on the green curve. This represents a measured BER corrected for the μLED extinction ratio but not for the ISI components.

A BER of $1\text{x}10^{-9}$ is estimated to be achievable at a sensitivity of -31.7dBm for an on-off key modulated signal at 100Mb/s (Figure 6.26). This is just under a micro-Watt of optical power ($\approx 0.67\mu$W) [2]and is primarily due to the single-photon sensitivity of the SPADs.

The quantum limit at 100Mb/s, corrected for the fill factor and PDE, is presented as a dotted line on Figure 6.26. This is based on the counts per symbol in the experimental data and reaches a quantum limited BER of $1\text{x}10^{-9}$ at a sensitivity of -35.9dBm. The observed gap between the estimated BER and the corrected quantum limit could be due to source intensity noise, DCR, after-pulsing, and photon transfer curve issues.

[2]The optical power, -31.7dBm or $\approx 0.67\mu$W, sits mid way between the grid lines marked on the top x-axis as $14.9\mu W/cm^2$ and $42.2\mu W/cm^2$. This top x-axis is an area normalised version of the $\approx 0.67\mu$W unit linear optical power units, but is scaled according to the area of the array and the area of the testing 450nm optical beam.

Including Approximate Power Penalties

Using power penalty budgeting, approximate penalties for the transmitter extinction ratio or inter-symbol interference can be included in this estimate. Initially using the values of the power penalties defined in Section 3.3.5, the above estimated sensitivity of -31.7dBm becomes -26.5dBm. This includes the eye opening penalty (≈ 2.5dB), inter-symbol interference penalty (≈ 2dB) and the DCR penalty (0.8dB). The jitter power penalty is already included, as each measurement for Figure 6.26, will include jitter on the integration start and stop edges. This is within specification (Table 2.1) but does not utilise the low-level characterisation results from Chapter 5.

These low-level measurement results can be used to update the initial values for power penalties (Section 3.3.5). This gives a first order estimate for the receiver sensitivity once the transmitter and receiver properties are included. In descending order, the power penalties are:

- The total ISI induced power penalty is ≈ 2.57dB. This can be split into three contributions, ISI due to the transmitter, $PP_{ISI}(\mu LED)$, ISI due to the receiver impulse response, $PP_{ISI}(Impulse)$, and ISI due to the receiver step response, $PP_{ISI}(Step)$.
  - The impulse response at $\approx 0.67\mu$W, is close to the minimum impulse decay constant of 1.9ns (measured at 1.25μW, see Table 6.1). Assuming a single exponential decay and 10ns symbols, this gives an ISI power penalty of $PP_{ISI}(Impulse) = 1.82$dB.
  - The ≈ 220MHz bandwidth of the μLED, gives a rise/fall time of $\frac{1}{\pi BW} \approx 1.45$ns [235]. Assuming a linear discharge, $\approx 7.25\%$ of a previous *logic-high* bit will leak into a subsequent bit. This gives an ISI power penalty of $PP_{ISI}(\mu LED) = 0.65$dB.
  - At the optical power ($\approx 0.67\mu$W) estimated to give the target BER (Figure 6.26), the variation of counts due to the step response is comparable to or less than natural Poisson noise (Figure 6.6). As such, a power penalty of 0.1dB is assumed.
- The extinction ratio of the μLEDs has been estimated (Section 6.5.2) to be approximately 82% with 100Mb/s modulation and 97% with 10Mb/s modulation. This gives an eye opening power penalty of $1.7dB$ at 100Mb/s.
- The measured total array DCR is much lower than the worst case (2 counts/τ_b = 200Mcps) assumed during design and seen in initial wafer runs. Here, the whole array DCR has been measured to be 6.6Mcps or 4Mcps if high noise devices are turned off (Section 5.2.3). This gives an estimated power penalty of 0.05dB for the DCR-induced offset from the theoretical OOK threshold.

Taking these power penalties, estimated from low-level receiver characterisation measurements, the total power penalty is ≈ 4.3dB. Adding this to the initial single-Gaussian, DC response only sensitivity estimate of -31.7dBm, gives an estimated full receiver sensitivity of -27.4dBm. This is shown by the red dashed curve in Figure 6.26.

6.5.2 Bit Error Rate: Further Statistical Estimation

Binh [3] discusses techniques used to estimate the BER, using multiple fitted distributions and maximum likelihood estimation. These techniques are useful when full bit comparison testing is impracticable [3] and when the simple model is inadequate. In order to obtain fast estimates of the BER, some BERT test-sets use the simple Gaussian model and the statistical techniques below prior to obtaining full bit comparison based measurements ([3] page 103). The length of a test is dependent on:

- the BER, for example the lower the BER the larger the number of bits that must be transmitted to receive a single error [2, 3],
- the number of erroneous detections required to get an accurate BER [3, 116],
- the data rate being tested and hence the time it takes to transmit the total number of bits required, and
- the number of parameters to be tested against, for example, multiple data rates, multiple optical powers, different modes or sweeps of power supplies and bias voltages [2, 3, 13, 14, 31, 36].

The Gaussian model [2, 3, 14] can be used to estimate the BER (Section 3.2 and Section 6.5.1), given photon shot noise, in-symbol noise, source intensity noise and the number of photons required for the target BER [3]. However it is only valid if enough samples and photons are obtained [2, 3, 14, 36, 103, 116] and if no transient issues force deviation away from two simple distributions.

The simple fit does not directly include behaviour such as the receiver transient responses (Section 6.2), or the non-linear behaviour of free-space links [3, 14, 16, 17, 20, 21, 41, 106]. In this section, the multiple distribution method [3] is discussed and used to explore the receiver-only BER. Doing so aims to assess the feasibility of SPAD-based receivers in the free-space communication [22] applications focussed upon here (Section 1.1.1).

Multiple-Gaussian Distributions and Expectation Maximisation

The received distributions N_0 or N_1 may not be Gaussian [3] (Section 6.4.1). This is because effects other than the signal amplitude and shot noise, will render peaks within the data. For example in Figure 6.17, Figure 6.18 and Figure 6.20, the receiver step response has created one distribution for the initial peak and a second for the steady state. Likewise the impulse and μLED responses have increased the variability of the *logic-low* distribution by creating a level that is dependent on the previous bit. If the last transmitted bit was a one, a *logic-low* bit will contain some of the impulse response from that previous bit (Figure 6.22).

The received distribution, $p(N)$, can be modelled as the summation of multiple distributions [3], where the probability density function (pdf) of each is weighted. If weightings, w_i, are assumed with $\sum_{i=1}^{M} w_i = 1$, and the normalised sub-distribution pdfs are denoted as $p(N|f(\mu_i, \sigma_i^2))$, i.e. the pdf is characterised by its own mean, μ_i and variance, σ_i^2, the received distribution is given by Equation (6.10).

$$p(N|w_1, w_2, \cdots, w_M;\ \sigma_1^2, \sigma_2^2, \cdots, \sigma_M^2) = \sum_{i=1}^{M} w_i\ p(N|f(\mu_i, \sigma_i^2)) \tag{6.10}$$

The 1^{st} and 2^{nd} derivatives estimate the number of Gaussians, along with their weightings, means and variances using expectation-maximisation [3]. This finds fitting parameters, maximising the probability that the results came from those ideal distributions [3, 116]. This concept is used later to assess the number of distributions and their parameters using the data sets obtained in 100Mb/s transmission experiments (Section 6.4.1) . As the receiver operates using integration (Section 2.5.3), at 100Mb/s with a 100MHz clock, the histogram represents N_0 and N_1 levels presented directly to clock and data recovery circuitry (Section 2.5.1).

In the simple Gaussian case, an integral is performed on the probability density function [1, 3, 103, 116], however the process could be the integration over any density function, which here will be denoted as $pdf(N_1)$ [3]. For the theoretical behaviour, this is integrated analytically, as discussed in [1, 2, 3, 14, 36, 103], over ideal distribution pdfs [67, 116]. However experimentally, the BER is either, a) measured using exhaustive bit comparison testing, or b) estimated by obtaining the N_0, σ_0, N_1 and σ_1 values and assuming only two distributions that are both Gaussian. Assuming equal '0' and '1' bit probabilities, the theoretical BER can be estimated using Equation (6.11) [1, 3, 103].

$$BER = 0.5 \left[\int_{-\infty}^{N_{th}} pdf(N_1) + \int_{N_{th}}^{+\infty} pdf(N_0) \right] \tag{6.11}$$

Taking the multi-Gaussian assumption and summation rules over integration [115, 116], the BER can be expressed as the summation of multiple integrated distributions [3]. For example, the integral of p_i, may represent the total probability that an inter-symbol interference modified output, is above N_{th} [3]. The BER is rewritten in Equation (6.12), where M_0 and M_1, are the number of sub-distributions within N_0 and N_1 respectively [3]. Clearly if $M_0 = M_1 = 1$, Equation (6.12) regains the form of the simple two distribution model (Equation (6.11)).

$$BER = 0.5 \left[\sum_{i=1}^{M_1} w_i \int_{-\infty}^{N_{th}} p_i(N|f(\mu_i, \sigma_i^2))\, dN + \sum_{j=1}^{M_0} w_j \int_{N_{th}}^{+\infty} p_j(N|f(\mu_j, \sigma_j^2))\, dN \right] \tag{6.12}$$

Bit Error Rate Estimation with Dynamic Behaviour

The BER can be investigated further using the μLED experiments (Section 6.4.1). Section 6.5.1 made the assumptions that received data could be approximated by two Gaussian distributions and that time dependent effects [1, 3, 67, 103] could be accounted for by approximate power penalties. The μLED transmission experiments suggest the multi-Gaussian fit [3], should be used. This aims to explore the receiver BER [3], while investigating the causes behind the received statistics.

While the μLED experiments did not yield BER results directly, the obtained μLED to SPAD receiver results, would have been the BER of the overall link. This would include the data generator, μLED, SPAD receiver, FPGA thresholding, and the single-bit return path to the BERT. Below, offline analysis of the saved, 14-bit per point, received data streams, aims to identify the contributions of the μLED and receiver. This allows investigation of the receiver only performance, thereby assessing the feasibility of using multi-SPAD integrating receivers for high sensitivity OOK optical communications.

The first contributions that can be checked are the affect of the photon transfer curve (Section 5.2.3) and any photo response non-uniformity (Section 5.2.4) on the BER estimate. Using the 1^{st} and 2^{nd} derivatives [3], the distributions for each data point in Figure 6.26 showed single-Gaussian behaviour. This suggests the receiver is not creating estimation issues, i.e. multiple distributions, from a continuous-wave (un-modulated) viewpoint. Certainly these effects may modify the mean and variance of the distributions, however N_0, σ_0, N_1 and σ_1 can still be used for this simple estimation (Section 3.2 and Section 6.5.1) [67, 103].

The second class of contributions that can be investigated are transient issues at low data rates. Using the derivatives to estimate the number of distributions [3] in the 10Mb/s experiments (Figure 6.17), the *logic-high* N_1 distribution showed single-Gaussian behaviour. This indicates that despite the receiver step response being observable in Figure 6.17, it is not yet severe enough to create two distributions. In contrast, the *logic-low* N_0 distribution gave two peaks when using the derivatives [3]. The secondary N_0 peak has a smaller probability and a large variance. This can be attributed to:

- the μLED discharge tail [23, 24, 96, 98], convolved with
- the impulse ($1 \rightarrow 0$) response of the SPAD-based receiver.

The peaks within the 50Mb/s and 100Mb/s data streams may be a product of increasing receiver transient issues as the modulation frequency and average optical power increase, alongside the non-idealities of the μLED and clock synchronisation.

Thirdly, to assess the contributions of transient issues at higher data rates, Figure 6.27 shows a selection of multiple-Gaussian curves, fitted to received data using 100Mb/s transmission from a μLED and a 100MHz SPAD receiver clock. Curve fitting with $N_g = 5-$, $6-$ and $7-$ Gaussians, each fit the data well [3]. This does not suggest there are exactly N_g Gaussian distributions, more that the technique can be used to better approximate the complex received statistics and give indications as to underlying processes [3].

In Figure 6.27, there appears to be four main distributions, with a varying number of smaller distributions depending on the fitting N_g value. The four main distributions are similar in normalised peak height (Table 6.2), suggesting they may correspond to the $0 \rightarrow 0$, $1 \rightarrow 0$, $1 \rightarrow 1$ and $0 \rightarrow 1$ data transitions respectively.

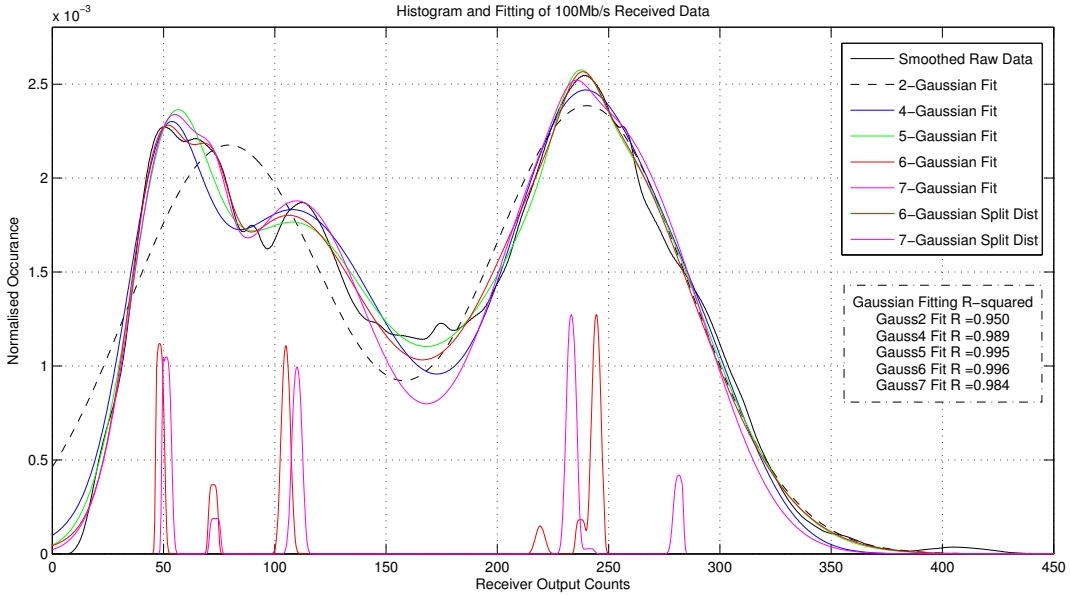


Figure 6.27: Histogram fitting for 100Mb/s NRZ μLED data, with a range of multi-Gaussian fitted curves. The R-squared fitting parameter (inset) indicates 6-Gaussian curve fitting approximates the received data set well. A 2-Gaussian fit shows departure from the data for the N_0 distribution but fits the N_1 distribution more accurately. At the bottom, the curve fitting parameters for the 6 and 7-Gaussian fits have been used to plot the individual sub-distributions. In order to observe them rather than the combined histogram, the standard deviation from the fitting has been reduced. This is used for illustration purposes only.

Assigning the peaks letters, A to G from left to right (Figure 6.27 and Table 6.2), it is possible to start suggesting the ordering of peaks and attributing effects to a particular peak by taking each in turn.

The assignment of peaks to various transitions, has been verified by selectively filtering the 100Mb/s data stream. If data points $n-1$, n and $n+1$ are taken using a 3-point window that is swept over the data set (similar to a 1x3 FIR filter), a particular transition can be captured (Figure 6.28). For example the $1 \rightarrow 0 \rightarrow 1$ transition matches peak C, with a strong peak at ≈ 110 counts. This transition also contains the B distribution, prompting a second 4-point window looking for the $1 \rightarrow 0 \rightarrow 0 \rightarrow 1$ transition. This yields the B distribution only, with a strong peak at a mean value of $\mu \approx 65$ counts.

- As a string of zero bits would give the absolute minimum signal, with any transient issues having had time to decay, the $0 \rightarrow 0 \rightarrow 0$ data transition can be attributed to peak A at a mean value of $\mu = 51$ counts. As such, this indicates that from 10Mb/s to 100Mb/s operation, the extinction ratio of the μLED has reduced, increasing the BER for the given light level and requiring an eye opening power penalty, PP_{EOP}, to account for the reduced N_0 to N_1 distance [14].
- Likewise the B and C distributions may be caused by the $1 \rightarrow 0$ transition, where both the receiver impulse response and μLED turn-off tail elevate the *zero* proceeding a high bit [1, 14].
  - For the C distribution, the $1 \rightarrow 0 \rightarrow 1$ transition, would include the receiver impulse response and μLED discharge tail. Crucially however, as the following bit is a *one* the $1 \rightarrow 0$ transition would be stopped before discharging down to the minimum level of distribution A.
  - For the B distribution, the $1 \rightarrow 0 \rightarrow 0 \rightarrow 1$, would again result in an impulse decay that is stopped prior to full discharge, but due to the second *zero* bit, it falls to a lower value than the C distribution.
- The F peak for the 6-Gaussian fit, or D peak for the 7-Gaussian fit, is consistent with the continuous $1 \rightarrow 1 \rightarrow 1$ transition. This is also the steady state count value reached after a $0 \rightarrow 1 \rightarrow 1$ step transient.
- In the 7-Gaussian fit, a further but smaller peak is seen at 280 counts. This is likely to be the initial peak observed in the receiver step response. From 10Mb/s to 100Mb/s the rippling period becomes a large proportion of the symbol period, indicating issues for technology scaling to higher data rates.

Figure 6.28 validates the multi-distribution model by splitting the receiver 100Mb/s data into eight sub-distributions. The $0 \rightarrow 0 \rightarrow 0$ distribution, as split from the measured data, peaks at 50 counts/bit ($\mu = 58.1$, $\sigma = 25.7$, $median = 54$), matching the 7-Gaussian 'A' peak at 51.2 counts per bit ($\sigma = 23$). Similarly, the $1 \rightarrow 1 \rightarrow 1$ distribution, which peaks at 232 counts/bit ($\mu = 5234.2$, $\sigma = 35.9$, $median = 234$), matches the 'D' peak at 233.2 counts/bit ($\sigma = 49$). By ensuring that the summation of the sub-distributions is equal to the total number of bits in the data set, the contributions of each to the whole data stream can be validated. Figure 6.28 shows a few interesting properties that link back to the impulse and step responses, tested (Section 6.2.2 and Section 6.2.5) and modelled earlier (Section 6.3).

Table 6.2: Coefficients for 6 and 7-Gaussian fitting curves for the received 100Mb/s μLED to SPAD-receiver OOK NRZ experiment.

Coefficient	6-Gaussian Fit	7-Gaussian Fit
Peak A	$1.7\text{x}10^{-3}$	$1.9\text{x}10^{-3}$
Mean A	48.2	51.2
Sdev A	20	23
Peak B	$5.0\text{x}10^{-4}$	$3.0\text{x}10^{-4}$
Mean B	72.3	73.2
Sdev B	11	9
Peak C	$1.8\text{x}10^{-3}$	$1.9\text{x}10^{-3}$
Mean C	104.9	109.9
Sdev C	54	47
Peak D	$2.0\text{x}10^{-4}$	$2.2\text{x}10^{-3}$
Mean D	219.2	233.2
Sdev D	66	49
Peak E	$3.0\text{x}10^{-4}$	$2.0\text{x}10^{-4}$
Mean E	237.3	233.9
Sdev E	12	10
Peak F	$2.1\text{x}10^{-3}$	$5.0\text{x}10^{-5}$
Mean F	244.5	241.3
Sdev F	62	17
Peak G	-	$8.0\text{x}10^{-4}$
Mean G	-	281.5
Sdev G	-	38

- Firstly, the 000 and 111 transitions each show approximately Gaussian distributions with lower variation than other transitions such as 011. This is linked to the settled nature of the response at this point, in that the 'zeros' are free from the impulse response and the 'ones' are free from the initial peak and ripple of the step response.
- Moving from 000 to 100 to 101, the mean value of the middle 'zero' increases as more of the impulse response is included. This matches peaks 'B' and 'C' where the bit is a zero because it is lower than the threshold, but due to its proximity to a previous 'one' includes a large proportion of leaked impulse response. Therefore the bit's N_0 value is close to the threshold, N_{th}, thereby increasing the BER significantly.
- The 010 data transition equates to the very large peak of the step response, but in this case only for transitions that have a following 'zero'. It is difficult to separate why the peak is slightly lower than the 011 transition that also includes the step response initial peak. However the issue is seen in the time trace of the 100Mb/s data (Figure 6.20).
- The 011 transition has a much wider variance than the other step response transition (010), this is because it includes the ripple and decay to steady state of the step response (Figure 6.5). This variability, and the reduction of counts/sample for the steady state can only be minimised by addressing the step response, as modelled in Section 6.3.5.

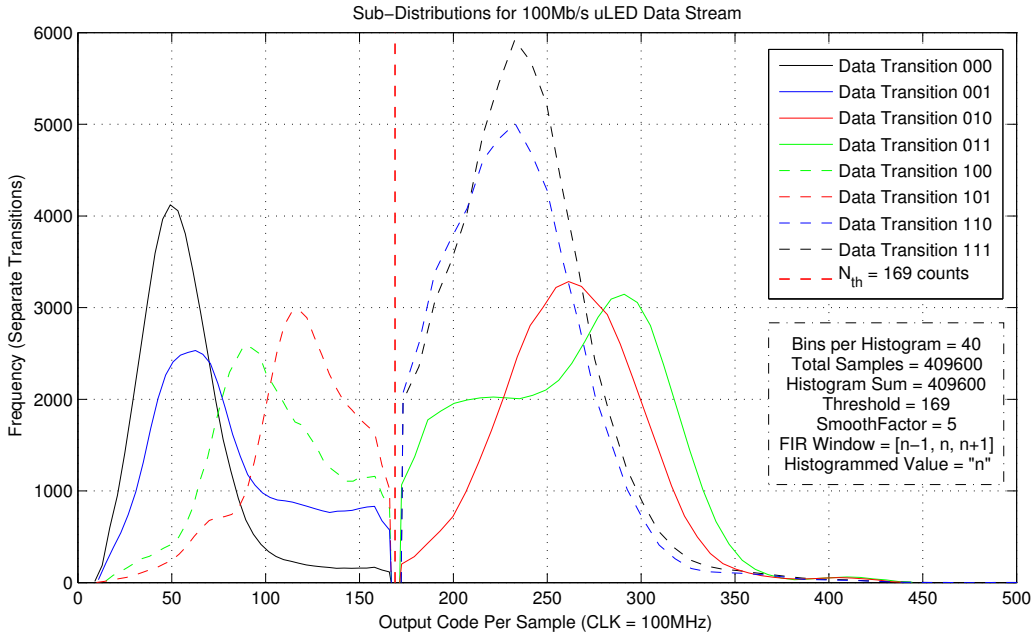


Figure 6.28: A manual split of the 100Mb/s received data. Each transition, 000 or 101 etc is split from the data set using a 3-tap filter. All histograms use 40 bins with a smooth factor of 5. The OOK threshold is set to 169 counts/symbol. As with the μLED experiment, the data rate is 100Mb/s with a SPAD receiver clock of 100MHz. To ensure all samples are included and no bits are counted twice, the summation of histogram bin values for all transitions, is equal to the total number of bits in the data set (409600 i.e. 400 FIFO frames of 1024 samples each). The FIR technique, with samples $n-1, n, n+1$ puts the central bit into the sub-distribution histogram. Therefore for the $1 \rightarrow 0 \rightarrow 1$ transition, the sub-distribution corresponds to 'zero' bits that occur after a 'one', therefore including the impulse response, but before another 'one' which stops the impulse response prematurely.

The BER of the above μLED experiment can be calculated using the multi-modal BER estimation above (Equation (6.10)) [3] and the fitting parameters for the 7-Gaussian fitting curve. The 7-Gaussian fitting has been chosen as it includes the ≈ 280 count histogram peak that represents the high initial peaks in Figure 6.20. In comparison, the 6-Gaussian fitting obtains a higher R fitting coefficient, but ignores this separated N_1 distribution. Doing this yields a rather poor BER of $2.8\text{x}10^{-2}$. However as mentioned above, the extinction ratio of the μLED has reduced, impacting the ability to only measure the BER of the receiver. The poor BER is also directly due to experimental clock synchronisation and equipment specification issues during the tests. From the fitted maximum and minimum count distributions of measured experimental data, the μLED extinction ratio at 100Mb/s is 82%. This is much lower than the $\approx 97\%$ extinction ratio observed at 10Mb/s.

The BER of the experiment, without the reduced μLED extinction ratio, can be estimated by scaling down the *A, B* and *C* distributions, but keeping their relative positions. Here the minimum count level is 51.2 counts. In the 10Mb/s experiment the minimum count level is 7 counts. A scaling factor of 7.2 can therefore be used. The *B* and *C* distributions are both comprised of background μLED induced counts and the two ISI contributions from the receiver impulse response and μLED turn-off tail. Scaling these distributions down in mean value removes the *LED off* counts but retains the SPAD detections induced by both ISI mechanisms. Doing this yields a BER of $5\text{x}10^{-4}$, just below the forward error correction limit [1].

The BERT experiments used an average optical power of $\approx 20\mu W/cm^2 \approx -34$dBm and hence will naturally deviate from the BER of $1\text{x}10^{-9}$ at an approximate sensitivity of $-31.7dBm$. This is clear from Figure 6.26, where reducing the optical power, quickly increases the expected BER. This is also clear from the quantum limit perspective (Equation (3.5) and Table 3.2), where a reduced optical power reduces the number of photons per symbol, leading to an increased BER [2, 3, 14].

In order to compare the μLED experiments with the initial BER estimate (Figure 6.26), either a) the light level must be appropriately scaled up to the -31.7dBm that was estimated, or b) the BER must be appropriately scaled to this new optical power. Given a measured point (in green in Figure 6.26) that includes both transmitter and receiver temporal characteristics, scaling repositions the 100Mb/s BER curve to the green dashed curve in Figure 6.26. This is similar to the power penalty scaling of the sensitivity shown in red (Section 6.5.1). This will not remove the influence of ISI, however the general curve of the relationship can be be used to estimate the BER at the slightly higher light level. Doing this gives either:

- An estimated BER of $1\text{x}10^{-5}$ at the previous -31.7dBm optical power, or
- An estimated sensitivity of -28.5dBm for a BER of $1\text{x}10^{-9}$.
- An estimated sensitivity of -30.5dBm for a BER of $1\text{x}10^{-6}$.

The 2.8dB power penalty, in comparison to the previous estimate of 4.3dB, suggests at least one of the penalties, if not all of them, are lower than originally thought in the power penalty budgeting sections. Further the above analysis scaled the μLED extinction ratio back to the $\approx 97\%$ for the 10Mb/s experiments. It would however be possible to scale the extinction ratio back towards a 99.99% extinction ratio, thus estimating the BER of the receiver only.

6.5.3 Conclusions from BER Estimation

In this section, two methods have been used to estimate the sensitivity and BER of the tested receiver front-end. Initially, a simple Gaussian model investigated the sensitivity due to the receiver photon transfer curve and the overall single-photon counting functionality. This uses estimated power penalties for factors such as the μLED bandwidth or receiver step response to estimate the sensitivity of the receiver.

As both the Gaussian estimation model and estimation of power penalties make simplifying assumptions, a second multi-distribution model is fitted to experimentally received OOK data. This helps identify the main problematic data transitions, such as the $1 \rightarrow 0 \rightarrow 1$ or $0 \rightarrow 1 \rightarrow 1$ transitions. To assess the performance of the receiver, independent of the transmitter, the results are scaled back towards a 100% extinction ratio. Doing this and fitting to measurement allows some reassessment of the power penalties.

Now that the performance of the receiver has been established, through communication results and estimation of the bit error rate, how the device performs against a selection of prior art can be discussed. Presented in the next section, this will use a figure of merit methodology to estimate a device ranking. This enables some insight into factors that should be optimised in the future.

6.6 Benchmarking with Prior Art

Comparing the combined measured and estimated performance of the receiver against similar receivers developed in CMOS technologies, can illuminate suitable paths towards higher performance. In this section, different receivers are numerically compared using three figures of merit. As a more qualitative outcome however, the comparison indicates that the optical fill factor and the slow data throughput (a product of OOK modulation and 100MHz system clock specifications), must be improved in future receiver generations.

6.6.1 A Receiver Figure of Merit (FOM) from Literature

In Table 6.3, the SPAD receiver is compared with other receivers using the figure of merit (FOM) below (Equation (6.14)). This is adapted from the approximate scale presented by Steyaert et al in 2009 [31], aiming to put all fully integrated CMOS receivers onto a scale as fairly as possible. Prior to the discussion of the overall figure of merit, the sensitivity, S, must be normalised to allow a fair comparison, independent of wavelength and data rate. This is denoted as S_{norm}, and uses the same bit error rate (BER_{target}), wavelength (λ_{target}) and bit rate ($BitRate_{target}$) for all devices (Equation (6.13)). The normalisation assumes testing used a) a fine line-width source, b) that the relationship between input photons and BER can be used (Equation (3.5)) to perform the scaling, and c) that receiver efficiency remains constant over the visible band. The sensitivity is normalised so that a receiver tested at 450nm can be approximately compared to a receiver tested at 650nm, although, as the wavelength approaches the UV band (400nm) or the near-IR band (900nm), the suitability of the approach will be reduced. This approximate scaling is useful in order to make general comparisons, knowing that variations in how and if values are quoted in the literature, are the largest contributors to FOM error.

$$S_{norm} = 10log_{10}\left[\left(\frac{BER_{target}}{BER}\right)\left(\frac{BitRate_{target}}{BitRate}\right)\left(\frac{10^{\frac{S}{10}}}{1000}\right)\left(\frac{\lambda}{\lambda_{target}}\right)\right] + 30 \qquad (6.13)$$

In order to create a suitable figure of merit, the below high-level performance characteristics were included by Steyaert et al [31]:

- Firstly, all receivers must integrate detectors, current to voltage amplifiers and conditioning circuits on the same chip, i.e. fully-integrated. The receivers must also use standard planar processes, where the minimum dimension gives a weighting allowing fair comparison of a receiver front-end that is larger than a competitor, not because of added gain or functionality, but because the manufacturing process is physically larger.
- If two receivers are compared with similar performance parameters, the bit rate is included to ensure that the faster receiver obtains a higher FOM.

- The bit error rate is included together with the sensitivity as a high optical power yields a low BER, however a receiver that obtains the same BER with less light should be ranked with a higher FOM.
- The area of the photo detector is included as a large area is easier to handle from a system packaging viewpoint [31], but also represents a larger parasitic capacitance. If two receivers both obtain excellent performance, but one must overcome a significantly larger capacitance to obtain the required bandwidth, data rate and BER, the large diode receiver should have a higher FOM.
- The power needed by the receiver is also included, as modern requirements are pushing for low energies per bit to allow continued data rate and throughput scaling [11]. The power is on the denominator to ensure high power consumption receivers have a lower FOM than receivers obtaining the same performance at a lower power draw.
- In the Steyaert et al FOM [31], the overall trans-impedance gain is included to allow for some receiver implementations that do not include post-amplifiers and do not give outputs suitable for direct connection to subsequent signal blocks such as CDR circuits or data multiplexers. Here the output voltage is instead included to achieve the same aims as trans-impedance gain is not defined for SPAD-based receiver front-ends. If a receiver outputs a larger voltage swing that a competitor with the same overall performance, the voltage swing that is suitable for connection to subsequent blocks is given a higher rating. Further to this, if both receivers drive similar output capacitances, including the voltage in the FOM allows the $P = fCV^2$ relationship to weight the overall power and energy per bit performance of the receiver.

With the above performance parameters, Steyaert et al [31] attempted to make a fair ranking scale. However in this work, single-photon sensitivity, which only SPAD-based receivers achieve, is excluded from the FOM (Equation (6.14)). Further, while the Steyaert et al FOM should include receiver silicon area, quotation carries errors, and may require assumptions to be made. It should be noted that due to differences in how or if a parameter is quoted in the literature, a figure of merit will always be inexact. As an example, for the receivers presented in Table 6.3, the inclusion or exclusion of pad ring power supplies in the overall receiver power draw is not always stated.

$$FOM = \frac{BitRate(bit/s)\ S_{norm}(dBm)\ log_{10}(BER_{target})\ V_{out}^2(V^2)\ PD_{area}(\mu m^2)}{Technology^2(nm^2)\ Power(W)} \tag{6.14}$$

6.6.2 An Optical Fill Factor Normalised Figure of Merit

With the FOM (Equation (6.14)) not including area or fill factor, the SPAD-based receiver must be compared equally to others despite its design for test and separated diode structure. Some designs [31, 110, 111] achieve 100% optical fill factor by abutting fibres to the active regions. However here, the fill factor of such devices is defined in the same way as a multi-diode receiver [79] such as the one presented [68]. This is taken to be the ratio of optically sensitive area to the total functional area.

In [31], the fill factor, with fibre abutment is $\approx 100\%$. However the fill factor, if defined as above, is only $\approx 0.16\%$. The SPAD receiver obtains a fill factor of $\approx 2.43\%$. Adding this to Equation (6.14), ensures that a device achieving its performance with a high photodiode to circuit area ratio, obtains a high FOM. The fill factor normalised figure of merit is denoted as FOM_{FF}.

6.6.3 A Circuit Area Normalised Figure of Merit

The FOM can be modified further to incorporate circuit area, FOM_{FFA}. Thus a physically large device scores a worse FOM_{FFA}, especially if the fill factor indicates significant non-optically active area. This metric embodies the performance to cost ratio, as area is directly proportional to cost. However unlike monetary cost, the metric can be used for small-run prototypes or devices that used academic discounts. The metric is useful to add to the receiver FOM, as a receiver obtaining a specific performance whilst using half the silicon area of a different implementation, clearly obtains that performance at a lower cost per unit.

6.6.4 Performance Comparison with Recent Prior Art

In Table 6.3, the SPAD-based receiver is compared to receiver prior art. This presents device parameters such as the bit error rate, data rate, optical wavelength and power per bit, along with the three figures of merit, FOM [31], FOM_{FF} and FOM_{FFA}.

Table 6.3: Performance Comparison with Recent Prior Art

Ref	Technology	BER	Bit Rate	λ	Area	Power/Bit	S_{norm}	FOM	FOM_{FF}	FOM_{FFA}
		(measured)	Gb/s	nm	μm^2	W Hz^{-1}	dBm			
[31] (1)	130nm CMOS	$1x10^{-12}$	4.5	850	$1.8x10^{6}$	$3.0x10^{-11}$	-18.8	11.92	11.12	4.88
[31] (2)	130nm CMOS	$1x10^{-12}$	4.5	850	$1.8x10^{6}$	$1.7x10^{-11}$	-18.4	12.18	11.38	5.14
[110]	0.6μm CMOS	$3x10^{-10}$	0.5	850	$5.7x10^{5}$	$3.4x10^{-11}$	-12.5	11.68	11.73	5.98
[40]	180nm CMOS	$1x10^{-12}$	3.125	850	$7.0x10^{5}$	$4.6x10^{-11}$	-17.6	10.90	10.67	4.83
[44]	0.6μm CMOS	$1x10^{-9}$	0.25	860	$1.3x10^{4}$	$1.6x10^{-11}$	-19.1	10.68	11.98	7.88
[38]	180nm CMOS	$1x10^{-12}$	2.5	850	$5.3x10^{5}$	$4.3x10^{-11}$	-17.0	10.68	10.35	4.63
[99]	0.35μm CMOS	$1x10^{-9}$	0.6	680	$4.5x10^{4}$‡	$5.0x10^{-11}$	-16.0	10.37	12.07	7.41
[236]	0.6μm BiCMOS	$1x10^{-9}$	2.5	660	$7.7x10^{5}$	$6.8x10^{-11}$	-32.4	10.72	11.68	5.79
[48]	130nm CMOS	$1x10^{-12}$	8.5	850	$1.0x10^{5}$	$5.5x10^{-12}$	-21.0	11.68	12.36	7.35
[111]	130nm CMOS	$1.5x10^{-12}$	2.4	850	$3.0x10^{6}$	$9.4x10^{-11}$	-15.2	11.55	10.52	4.05
[39]	130nm CMOS	$1x10^{-12}$	4.5	850	$1.8x10^{6}$	$1.6x10^{-10}$	-18.4	11.19	10.39	4.15
[33]	130nm CMOS	$1x10^{-11}$	3.0	850	$2.1x10^{5}$	$1.7x10^{-11}$	-31.9	10.28	10.36	5.04
[100]	0.6μm BiCMOS	$1x10^{-9}$	1.8	655	$1.3x10^{6}$	$5.6x10^{-11}$	-30.4	11.95	12.93	6.81
This Work	130nm CMOS	$1x10^{-9}$††	0.1	450	$2.1x10^{6}$	$8.0x10^{-10}$†	-31.7	12.20	12.58	6.26

Table 6.3. † measured at maximum counting rate, expected to be lower for quoted BER. ‡ estimated from device micrograph. †† BER estimated using simple-Gaussian and multi-distribution fitting. Pertinent FOM_{FFA} Statistics: $Mean = 5.7$, $Mean + 1\sigma = 6.98$, $Max = 7.9$, $Min = 4$, $Std = 1.25$ and $Median = 5.5$.

6.6.5 A Discussion of Compared Receiver Technologies

The SPAD-based receiver has been used to investigate the concept and explore affecting parameters. As such, the receiver is not optimised. Due to the combination of an integrating receiver architecture, multiple diodes and high avalanche gains, optimisation routes were not available. In this section, the above table is discussed.

Receiver Power per Bit Performance

The SPAD-based receiver power of $8x10^{-10}$ WHz^{-1}, is measured at the maximum count rate, C_{max}, for the base dead time (13.5ns). This inflates power consumption above other receivers that detect just enough light for communications but do not provide single-photon sensitivity. The high power is also due to non-optimised design, and the increased output voltage amplitude compared to other receivers.

Due to differences in how power draw is quoted in the literature, namely if pad ring power is included or if photodiode bias voltages are included, the voltages and currents used in the power draw estimation are for the receiver core only. For a photodiode receiver, this may need to exclude the power supply of the final buffers that drive the output pads if the current for that supply is not quoted. For the SPAD-based receiver, the power discounts the current drawn from the VBD supply. This is a limitation, as current is drawn by SPADs but is lower than SPAD quenching, digital logic and clock distribution currents. As the VBD voltage is negative, conventional current flow, if applicable, would suggest current flows from the SPADDVDD, DVDD and ground nets, through the SPAD and out of the device via the VBD net. For a SPAD, the current flows are complex, due to the nature of the SPAD planar structure (Figure 4.2), and the capacitive network that surrounds the p-n junction (Figure 4.3), with extra carriers being liberated from the lattice during a breakdown, and carriers drifting from elsewhere in the silicon CMOS substrate. There is ongoing research in this area using TCAD simulators [76, 185], however if the convention holds VBD should be discounted to ensure SPAD current draw is not taken into account twice. The power per bit of the receiver, with its large clock tree, non-optimised quenching and digital circuits and limited pad output rate, is open for optimisation (Section 7.4).

The SPAD-based receiver is able to achieve a photon detection energy efficiency of $\approx$ 3.2pJ/photon. This is far in excess of other receivers which would draw significant power if their gain was scaled to achieve single-photon detection. The energy per bit, is $\approx$ 370pJ/bit. This was calculated using the measured power draw when the receiver was operating at its maximum counting rate, linearly scaling to the photon count rate giving the target BER and then scaling to a single bit. SPAD receivers will draw increased electrical power due to the bias voltages required. However, including the efficiency of the transmitter, increased sensitivity will lead to a reduced link power budget once SPAD-based receivers are optimised. The system

proceeding receivers that use LVDS outputs, will require electrical power for amplification if processing by CMOS logic is to be used. As such, the FOM (Equation (6.14)) [31] of low voltage output implementations will be decreased from that shown in Table 6.3. The presented receiver is suitable for direct connection of signal processing blocks.

Normalised Sensitivity Performance

The sensitivity of the SPAD-based receiver has been discussed in the preceding sections. While this was based on estimation, the receiver should remain in a high position relative to the others due to the single-photon sensitivity of the SPADs. Due to the assumptions and nature of using quoted values from the literature, each value in Table 6.3 contains some error.

Comparison Using The Steyaert et al FOM

Steyaert et al [31], implemented a low power, high-speed receiver with a small photodiode and 950mV output. This obtained a comparable FOM (12.18 vs 12.2), but requires a high input optical power. In comparison, this work presents a pre-optimised design with an active area $\approx$ 18.2 times larger, a normalised sensitivity (linear units) $\approx$ 21 times higher, 1.2V CMOS logic-compatible output and includes a number of test circuits used during prototyping. Comparing [236] and [99] with the SPAD receiver, despite both operating in the visible, they have reduced FOMs due to their lower output voltages ($\approx$ 225mV).

It is clear from Table 6.3, that for BER, speed and area, the presented SPAD-based receiver does not compare favourably. The overall FOM of 12.20 has been reduced by the 0.1GHz data rate. However despite limitations of 108Mb/s per pad, schemes such as MPAM could be used to increase the bit rate for the same receiver clock frequency [1, 3]. The closest bit rate neighbour [44] obtains a rate of 250MHz. Its poor FOM is a product of mid range sensitivity and use of a small photodiode. It however obtains its performance, even at the 0.6μm node, within a very small area budget.

Comparison Using The Fill Factor Normalised FOM

The fill factor inclusive FOM (FOM_{FF}) for [111], decreased from 11.55 to 10.52. This is due to a fill factor of $\approx 0.094\%$. The greatest improvement from FOM to FOM_{FF}, was observed with [99] as it achieves a fill factor of $\approx 50\%$ (estimated from micrograph). This was achieved by using photodiode structural equalisation rather than circuit techniques. In fact the TIA was astoundingly small at $\approx 50\mu$x50μm. As this device was implemented with structural equalisation, the sensitivity has been reduced and hence the FOM_{FF} is lower than other designs using combinations of the two techniques.

Under the fill factor normalised FOM, the SPAD-based receiver retains its high position, however the order of the receivers has changed. Under the initial figure of merit (Equation (6.14)), a weak-inversion receiver [31] obtained 12.18. The fill factor modified FOM of this receiver is far lower at 11.38. The FOM_{FF} of the SPAD-receiver improves slightly, however [100] obtains far higher values of 12.70 and 12.93 at 1.25GHz and 1.8GHz respectively. These are achieved using a BiCMOS process, large integrated photodiode (400μm dia), circuit based equalisation, an output voltage of 720mV and fill factor of $\approx 9.6\%$.

Comparison Using The Fill Factor and Area Normalised FOM

Once the area is included, the positions of the receivers change. The SPAD-based receiver no longer obtains scores near the top. This is because the receiver obtains poor area utilisation due to the *design for test* structure and the design rules regarding SPAD guard ring widths and N-well separations. The receiver is an initial prototype. As such, further work (Section 7.4) will give improved performance. Despite this, the receiver obtains a fill factor, circuit area and output voltage normalised figure of merit towards the middle of the range. The concept therefore has potential to scale if further work tackles initial limitations.

6.6.6 Conclusions from Comparison Figure of Merits

The FOMs presented, indicate improvements necessary for SPAD-based receivers. If SPAD receivers could be equalised to improve the impulse and step response or if topological techniques are combined to yield high fill factor and high detection probabilities, high-speed, low-BER, fibre-coupled, high-sensitivity receivers could be possible.

Likewise, cheap CMOS, mass-producible receivers with small but highly sensitive active areas, may promote further growth in LED-based data transfer and solid-state illumination applications, such as VLC. From the presented figures of merit, methods to increase the data rate should be explored in the future to achieve higher FOMs. One method would be to increase the operating clock towards the GHz levels used in some of the other receivers within the comparison table. Another method, would be to explore modulation schemes such as MPAM that can be combined with silicon implementation optimisation. These data throughput techniques, and reduction of ISI must be explored before SPAD-based receivers can lead in mobile platform applications of visible light communications. In the next section, system level factors such as the typical hierarchical data protocols used for communication, or the possible use of such a high sensitivity receiver are considered.

6.7 Important Issues for Communication Designers

The communications community covers optical source and detector design [2, 14, 35], transmitter and receiver design [2, 13, 14, 31, 35, 36], modulation and coding scheme design [1, 3], right through to link and protocol modelling, design and standardisation [1, 15, 22]. Comparison using figures of merit allows investigation of how to optimise a receiver. In this section however, the system level of communications is also considered. This points towards protocol levels or further integration of functions to aid in electrical and economic cost efficiency.

The receiver exists only at the physical layer of the International Organisation for Standardization's Open Systems Interconnection (OSI) model [102]. This structure and the Internet Protocol model [1], ensures separability between functional layers. As such, once a receiver operates within specification [1], its impact on the higher levels is minimised. Throughout, the receiver design has progressed from the project specifications (Table 2.1), and has inferred specifications from literature (Section 2.5 and Section 2.7.3) and initial modelling (Section 3.3 and Section 3.3.5). The use of integrating architectures [35, 73, 113] (Section 2.5.3), high avalanche gain [26, 49] (Section 2.6) and multiple diodes [41, 43] (Section 2.8) is feasible, as shown by experiments with a CMOS implementation (Section 6.4), and recent feasibility studies [67]. While there is a need for further characterisation, design and experimental communication work (Section 7.4), there are a number of conclusions that can be drawn.

- Optical signals can be received using the presented combination of technologies,
- The impulse and step responses require further work to mitigate inter-symbol interference induced BER contributions,
- Small, cheap, CMOS-only, high-sensitivity receivers could be used for mobile applications, and
- CMOS integration allows tasks such as clock and data recovery, threshold calculation, equalisation or parity checking, to be computed on-chip.

Along with these, the work has presented a specification flow for a SPAD-based integrating receiver (Section 3.3). While this requires understanding of such receivers and inclusion of further modelling, future designs could use a similar flow to create high-sensitivity receivers. The project aimed to explore optimisation routes for SPAD-based receivers. Once added to and adequately modelled, this would allow the best engineering solution for a given specification. Here, some experimental effects could not be separated due to other silicon requirements. Likewise, modelling the link, inclusive of observed effects, was complex due to practical issues and inherent random variation of some effects.

Two implementations that could be conceivable within the current receiver's performance and through the suggested specification flow, are:

1. A small area, high-sensitivity receiver for 25–50Mb/s VLC, that is suitable for mobile applications. This could use native BERs of $1x10^{-6}$ with on-chip correction to achieve error-free reception. White light, suitable for illumination, could be used, or optical filters could be incorporated if fill factor power penalties can be reduced. The IEEE standard [22] currently contains some functionality required for the OSI or IP data link layers. However once handshaking between transmitters and receivers and on-error retransmit are included, robust communications will be possible.
2. The design of a physical layer optical receiver for the 100Mb/s, 850nm, 100BASE-SX Ethernet industry standard [1]. This could include the advantages of the presented receiver to extend the current 550m operation distance, or the advantages of the Ethernet IP protocol for high reliability communications.

To conclude, future work on the specific silicon implementation of SPAD-based receivers may be able to meet the specifications of conventional protocol implementations within the OSI and IP communication hierarchies. Once this occurs, SPAD-based designs may tend towards the ideal receiver, that of highly sensitive, inexpensive receivers that leverage cost reductions from increased integration. Future work, addressing the temporal response, may enable designs operating close to the quantum limit, with direct applications where sensitivity is the main priority over absolute speed.

6.8 Higher-Level Characterisation Conclusions

In this chapter, experimental work evaluating the performance of the receiver (Table 6.4), including the impulse (Section 6.2.2) and step (Section 6.2.5) responses are discussed. These are limiting factors, degrading the achievable BER for the target specifications (Table 2.1). With this SPAD unprotected from substrate carriers (Figure 4.2), photon-induced carrier diffusion is proposed (Section 6.2.2) as a mechanism for impulse issues requiring characterisation before mitigation can be implemented. Complementing this, the step response shows significant contributions to BER (Section 6.2.5). In particular, a peak is followed by ripples tending towards a smaller settling value. This broadens the *logic-high* distribution (Section 6.4.1), reducing the quality of received data. Experimentation and modelling suggest this is linked to the number of SPADs, the number of SPADs removed and the number of SPADs recharged, while pointing towards methods that may reduce it.

The BER of the receiver has been estimated, firstly using a simple Gaussian model (Section 6.5.1), and secondly a multi-modal distribution technique using μLED data (Section 6.5.2). These aimed to assess only the BER contribution of the receiver, finding that while data can be received, there are a number of power penalties that must be reduced in future generations. The step and impulse responses both impede robust communication, while the optical fill factor and photon detection efficiency contribute significantly to a departure from the ideal quantum limit.

To compare the receiver with prior art, figures of merit are used (Section 6.6). The Steyaert et al FOM is modified to obtain an equal footing. The SPAD-based receiver performs well, however it requires further work before it reaches the speeds of prior art. The FOMs are also useful from a specifications perspective, suggesting modifications that should be prioritised.

Finally, this chapter discusses implications for communications. The broad conclusion is that the concept is feasible and that further work is required. Such devices include routes towards CMOS only, high-sensitivity receivers, that could incorporate error correction on the die. In the next chapter, the concept, design and performance values of the SPAD-based receiver are summarised, and a number of topics for future work are suggested. Chapter 7 concludes the work, briefly commenting on applications of such a receiver and its suitability with regard to ongoing trends.

Table 6.4: A short summary table of the main results discussed within this chapter.

Parameter:	Value:	Units:	Notes:
Impulse Time Constant	1.9-14.8	ns	27-30,000 photons/pulse
			12.5fJ to 14.2pJ per pulse
	6.4	ns	$N_a = 4$ TDM Mode
	2.35	ns	$N_a = 1$ TDM Mode
	6.7	ns	AQAR at $\tau_d = 13.5$ns
	5.5	ns	PQPR at $\tau_d = 13.5$ns
Received Data	10	Mb/s	Eye-open but slow $1 \rightarrow 0$
Ambient=None	50	Mb/s	Eye-closure worse $1 \rightarrow 0$
Modulation=High	100	Mb/s	Eye-closed, high BER
			Combined µLED and
			Receiver Transient Issues
			Note: Reduced Extinction Ratio
			Maintained 1:1 Correspondence
Received Data	10	Mb/s	Data-visible but high noise
Ambient=Low	50	Mb/s	Data-corrupted
Modulation=Low			Partial N_0 and N_1 overlap
	100	Mb/s	Full Data corruption
			Full N_0 and N_1 overlap
			Combined µLED and
			Receiver Transient Issues
			With high N_0 signal amplitude
			Note: Reduced Extinction Ratio
			1:1 Correspondence Lost
Estimated DC Sensitivity	-31.7	dBm	At $BER = 1\text{x}10^{-9}$
			Gaussian Assumption [1, 3, 103]
			Continuous-Wave
			$\approx 0.67\mu$W at 450nm
Estimated DC QL	-35.9	dBm	*As above*
			Corrected for PDE and FF
Gauss-Model	4.3	dB	From Measurements
Power Penalties			(*impulse, step, µLED, etc*)
Gauss-Model DC Sensitivity	-27.4	dBm	Inc PP_{ISI}, PP_{EOP} etc
Multi-Gaissian Fitting	-31.7	dBm	At $BER = 1\text{x}10^{-5}$ *or...*
Corrected	-28.5	dBm	At $BER = 1\text{x}10^{-9}$
Energy Per Bit	≈ 370	pJ/Hz	Estimated from C_{max}
Photon Detection	≈ 3.2	pJ/photon	Estimated from C_{max}
Energy Efficiency			
FOM_{FFA} SPAD-receiver	6.26	-	Within $\mu + 1\sigma = 7$
FOM_{FFA} Mean	5.73		$Max = 7.88$, $Min = 4.05$,
			$Median = 5.46$, $\sigma = 1.25$

Chapter 7

Consequences and Conclusions Of Presented Research

7.1 Thesis Motivation and Overview

Visible Light Communication (VLC) is able to increase data transmission capacity. This is required where the number of subscribers restricts the *per subscriber* bandwidth, or where download intensive tasks are required despite regulation of the electro-magnetic spectrum. VLC is an attractive solution to the problem, due to the introduction of solid-state LED lighting as a replacement for incandescent lighting in home and office environments. A variety of data modulation schemes can be implemented using light-emitting-diodes. For example, multiple colours not only provide white light but also allow wavelength division multiplexing, thereby increasing the utilised bandwidth.

As optical efficiency decreases with inclusion of filters or non-ideal transmission medium effects, multiple techniques have been used to increase sensitivity. It is hoped that these could be utilised to enable small, CMOS-only receivers with sufficient sensitivities to address three key issues.

- Firstly, optical filters could allow multiple wavelengths to be used for modulation, while still providing white light illumination required by IEEE VLC standards,
- Secondly, reduction in the modulated optical transmission power could aid in decreasing link power budgets, allowing energy efficient designs,
- Thirdly, optical powers on the receiver can be reduced, allowing packaging flexibility or use of smaller active areas allowing multiple receiver channels per device.

The feasibility of multiple SPADs within integrating receivers has been investigated. This work presents initial modelling to produce specifications, the implementation of a device, and demonstrates the concept through characterisation and transmission experiments (Chapter 6). This is complemented by modelling to explore observed issues. The positive outcome, is that data can be received using such a device. This implies feasibility if future work tackles some of the non-ideal behaviour. It also indicates a route to reduce the increasing energy requirements of communication links, while supporting VLC concepts.

In the proceeding sections, the work within each chapter is briefly summarized into its key conclusions, whether that be specifications or measured results. Future work is then discussed aiming to address issues observed experimentally. Finally, the two key issues from the work are discussed in the context of the overall VLC application. This highlights a) the mobile computing, high-download capacity target application of such research and b) that commercialisation and research is progressing the concepts discussed.

7.2 Summary Of Work Presented

The first chapter introduced optical communications, the motivation for which are electromagnetic spectrum restrictions and its use in solid-state, lighting-based visible-light communication standards such as IEEE 802.15.7. This lead on to the motivation for this work (Section 1.2), that of increased receiver sensitivity. This would allow robust communications despite optical filters or small receiver areas. The chapter concluded indicating the overall drive behind the work presented. This takes the form of overall project specifications, the most important of which are briefly summarised in Table 7.1.

Table 7.1: Selected overall project specifications.

Parameter:	Value:
Technology:	all-Digital CMOS
Number of Channels:	Multiple Channels
Target Data Rates:	100-500Mb/s per channel
Target Bit Error Rate:	$1x10^{-9}$
Target Sensitivity @ 100Mb/s:	Better than -24dBm
Target Wavelength:	400nm to 750nm (Visible)
Compatibility	Co-Design with μLED transmitter

Chapter 2 introduced optical communications in more detail. Key performance metrics such as the bit error rate and inter-symbol interference were discussed prior to a review of receiver literature. The techniques used within this work were then established. In turn these were:

- use of integrating receivers, giving increased received signal amplitudes and some noise immunity (Section 2.5),
- use of avalanche multiplication exceeding that used in avalanche photodiode receivers, thereby moving most, if not all, the necessary gain into the diode (Section 2.6), and
- a multi-diode approach that allows lower noise rates and higher photon count rates to be obtained (Section 2.8).

The reviewed literature suggested the techniques can be used additively, quickly enabling the high-sensitivities that VLC applications require. Chapter 2 therefore concluded with specifications (Table 7.2) that are required throughout the design chapters, simultaneously establishing concurrent design with a multi-channel, CMOS-controlled, visible-light μLED transmitter.

Table 7.2: Selected high-level receiver specifications.

Parameter:	Value:	Notes:
Modulation Schemes:	On-Off Keying	NRZ
Receiver Architecture:	Integrating	Discrete Time
Diode Operating Mode:	Avalanche to Geiger-Mode	-
Front-End Circuit:	Multiple Diode Summation Active (+ passive quenching)	- -
Total Receptive Elements:	1024	-

In Chapter 3 a model of the error rate was first used to assess parameters that should be optimised to create a high-sensitivity receiver. This was then built upon using models from the literature to indicate how silicon circuits and the prototype VLC receiver should be implemented. In a similar way to the preceding chapter, Chapter 3 concluded by generating a list of specifications. These specifications, the crucial points of which are summarised in Table 7.3, lead directly into the CMOS design of the receiver in Chapter 4. A key conclusion from Chapters 2 and 3, which is included within the overall specifications, is that of design for testability. This directly allows testing of the concept and is later shown to be a key requirement in order to test and understand two fundamental limitations of the receiver front-end.

Table 7.3: Selected final specifications for a high-sensitivity SPAD receiver.

Parameter:	Value:	Units:	Notes:
Quantum Limit	20	Photons	$BER = 1\text{x}10^{-9}$
Dark Count Rate:	$\leq 0.1QL$	-	≤ 2 Detections per τ_b
Total detections:	22	-	Detections per τ_b
SPAD Diameter:	≤ 8.6	μm	Assuming Circular
Array Architecture:	Clustered	-	Hierarchical Addition
SPAD Summation: -	Time Division Multiplex -	- -	Mode Configurable $N_a = 3$ for disable

Chapter 4 implemented the receiver within a CMOS technology. As the literature within Chapter 2 pointed to a small proportion of SPADs with dark count rates far higher than the average, functionality was included to turn off noisy groups. This was achieved using added circuits that allow receiver configuration when in use. Furthermore, the excess bias, dead time, quenching mode and summation method were each configurable to aid in assessment of their impacts within subsequent chapters. The chapter conclusions centre around the fabricated device, in that it acts as a testing platform for the concepts. As such, the included functionality and how it can aid experimental work are direct outcomes. Some of the key parameters from the design are outlined in Table 7.4, indicating a) the functionality included, b) design parameters important for subsequent testing and c) factors that act to limit the device.

Table 7.4: Selected details of the fabricated SPAD-based receiver chip.

Parameter:	Value:	Units:	Notes:
CMOS Process	130nm Imaging	-	90nm Metallization
# SPADs	1024	-	Circular
SPAD Diameter	8	μm	Active
Quenching	Active/Passive	-	Adjustable
Peak PDE	20 [143]	%	$Vex = 1.2V$ 450nm
Digital DVDD	1.2 / 3.3	V	Core / Pads
Array Fill Factor	2.42	%	Exc Pads etc
Max Clock	108	MHz	*Pad Limited*
# Channels	1 or 16	-	Configurable
Die Size	2.4x2.1	mm	$\approx 5mm^2$

Characterisation of the manufactured SPAD array has been presented in Chapter 5. SPAD specific values such as the dead time, dark count rate, after-pulsing probability and photon transfer curve were explored and are briefly summarized in Table 7.5. This allowed some knowledge of the magnitude of a particular factor prior to experimental communication work. For example, the power consumption of the device indicates the need of future work on silicon implementation, to optimise the energy per bit performance.

A significant and novel contribution this thesis has made, has been a number of models. One such model has investigated the receiver responsivity due to a suggested competitive interaction between photon, dark and after-pulse counts. It was concluded that this may indicate a reduction in detected thermal carriers as the optical flux increases, further suggesting use of all of a SPAD's available counting resources. The chapter concluded that some SPAD array characteristics will need to be optimised in a second generation. This is required in order to obtain either desired communication performance (explored in Chapter 6) or the scaling of that performance, with higher optical and electrical efficiency.

Table 7.5: Selected initial characterisation results of the SPAD-based receiver.

Parameter:	Value:	Units:	Notes:
Max Clock Distribution	≈ 20	mW	-
Max AQ DVDD	≈ 12	mW	-
AQ Dead Time	10–250	ns	Using $V(DTCTRL)$
PQ Dead Time	5–200	ns	-
DCR (ALL SPADs)	Median = 2.5	kHz	At 16°C
Photon Transfer Curve	$C_{max} = 6.5\text{x}10^{10}$	cps	TDM Mode $N_a = 4$
-	$C_{min} = 6.6\text{x}10^{6}$	cps	-
After-Pulsing	mean$(P(ap))$ = 0.9	%	1μs Window, 12 SPADs

Chapter 6 presented the receiver impulse and step responses, both of which impact the BER. It can be concluded that this receiver includes significant inter-symbol interference, with this being a particular drive for future work. This lead to a novel step response model that can be used to influence the number and dead time of SPADs in subsequent designs (Section 6.3) and is a direct outcome matching the contributions to knowledge in Chapter 1. Experiments using a μLED, demonstrated the positive result that data can be received, although three factors influence the immediate use of the technology (Section 6.4).

- The impulse and step responses are problematic at the 100Mb/s target data rate,
- The transient response of the μLED adds further inter-symbol interference, increasing the error rate, and
- The performance of single-photon sensitive receivers is reduced when unmodulated light is present.

Chapter 6 concluded with a comparison with prior CMOS optical communication receivers. A number of figures of merit were given, attempting to evaluate performance fairly. The SPAD-based receiver performed well. However, further work is needed to remove observed non-ideal behaviour and to push obtainable figure of merits further (Table 7.6).

Table 7.6: Selected transient response and data reception characterisation results.

Parameter:	Value:	Units:	Notes:
Impulse Time Constant	1.9-14.8	ns	27-30,000 photons/pulse
Estimated DC QL	-35.9	dBm	*As above* Corrected for PDE and FF
Gaussian-Model Power Penalties	4.3	dB	From Measurements (*impulse, step, μLED, etc*)
Gaussian-Model DC Sensitivity	-27.4	dBm	Inc PP_{ISI}, PP_{EOP} etc
Multi-Gaussian Fitting *Corrected*	-31.7 -28.5	dBm dBm	At $BER = 1\text{x}10^{-5}$ *or...* At $BER = 1\text{x}10^{-9}$
Energy Per Bit	≈ 370	pJ/Hz	Estimated from C_{max}
Photon Detection Energy Efficiency	≈ 3.2	pJ/photon	Estimated from C_{max}

This work has introduced communications using a multi-SPAD, all-digital integrating receiver. The implication for communications requires work to address power usage and transient response issues. Power penalty budgeting, along with the FOMs of Section 6.6, indicate that with recent advances in SPAD array fill factor and photon detection efficiency, highly sensitive receivers are feasible [67]. Such receivers provide increased flexibility for VLC, allowing optical filters to be used for bandwidth increasing wavelength division multiplexing. Reduction of the power budget is also needed to counteract the escalating energy use of increasing demand for bandwidth. With respect to VLC commercialisation [233], high-sensitivity allows a smaller active receiver area for a given sensitivity. This is crucial if VLC is to be introduced into mobile, low-power, hand-held devices as indicated by the literature and VLC standards.

7.3 Obtained Performance and Comparison to Targets

In this section, the performance and learning outcomes of the prototype receiver front-end are compared to the project specifications. Chapter 1 discusses that overall, the project and the prototype receiver front-end have been developed to assess the feasibility [67] and design factors [68, 156] that are involved in using SPADs for optical communications (Section 1.2). As Chapters 5 and 6 have shown, there were a number of factors that were not known prior to the project, with the investigation of these becoming a paramount point after early experiments showed them to be detrimental to communications. In Section 1.3, a brief summary of some of the implementation objectives was presented. These covered the development of the receiver front-end IC, a testing PCB and related firmware and software. It also discussed objectives such as establishing the performance of the prototype and modelling it in order to make suggestions for future designs. Each of these outcomes has been achieved, with Section 1.4 and Section 7.2 setting out the main points for each chapter.

Chapter 2 presented more specifications from the Hypix project. This included the target communications specifications such as the BER of $1x10^{-6}$ to $1x10^{-9}$, the data rate of 100Mb/s to 500Mb/s, the wavelength range of 400nm to 750ns, the use of OOK modulation and the co-design with the Hypix project's 32x32 μLED transmitter array. The sensitivity of the receiver was also specified, aiming for -24dBm, although the exact figure was accepted to be dependent on as-yet-unknown factors and the feasibility of the concept.

The performance of the SPAD-based receiver front-end can be compared to the target specifications put forward in Chapters 1, 2 and 3. For some specifications the chip directly satisfies the the project aims. For example, the chip uses a 32x32 SPAD structure, mirroring the μLED transmitters and the functionality of reconfigurable grouping of devices into multiple data channels. The chip also obtains wide wavelength sensitivity as demonstrated by previous SPAD characterisation [143]. The receiver front-end also demonstrates some suitability for OOK signalling with chip readout clock speeds up to 108MHz, although for the most part, the chip does not obtain the required communications performance (Section 6.4 and Section 6.5).

Through device characterisation, modelling and communication experiments, it is clear the impulse and step responses of the SPAD-based receiver have drastically impacted the communications performance. This isn't to say that the prototype cannot operate at a clock rate of 100MHz, but that it cannot obtain 100Mb/s while also obtaining the target BER and sensitivity. Prior to design, there was little indication that either factor would be as severe as they have been measured to be. The performance is therefore either an estimated BER of $1x10^{-5}$ with a sensitivity of -31.7dBm, or an estimated sensitivity of -28.5dBm with a BER of $1x10^{-9}$. This approximated performance satisfies the specifications with a margin of -4.5dB, however the finding-of, investigation-of and modelling-of factors impacting communication performance were the primary requirements and aims of the project (Section 1.3).

7.4 Future Work For >1Gb/s OOK VLC SPAD Receivers

While a SPAD-based integrating receiver has been fabricated and demonstrated (Chapter 6), the future work required for VLC receivers obtaining native OOK data rates greater than 1Gb/s is of interest. VLC is intended, alongside conventional Wi-Fi, for download intensive or multi-user situations (Section 1.1.1). With the increase in connected devices, number of subscribers and bandwidth demands per person, VLC receivers will need to scale appropriately to meet these increasing demands.

The presented receiver front-end circuits used standard CMOS processing technology, allowing the scaling of this receiver type to use techniques and advances in CMOS processing, digital circuit scaling and of on-chip signal processing. The use of SPADs implemented in CMOS is becoming better understood and practical solutions have been emerging for low noise diode structures, fill factor enhancement techniques and circuit implementations.

7.4.1 Specific Future Work

Following a bottom up approach, specific ideas for future work are presented below. Each modification attempts to reduce the penalties that prevent a quantum limited receiver.

- To target the inter-symbol interference caused by the impulse response, Silicon-on-Insulator (SOI) processes, despite high costs, often achieve system and photodiode speeds near the upper end of the performance scale [31, 46]. For integrated photodiodes, SOI increases the intrinsic bandwidth. This may be useful for SPAD technologies.
- For SPAD design, PDE performance can be improved using double junction SPADs. This increases the capture of photon absorptions using larger silicon depths. To address impulse response issues, positive drive SPAD configurations have shown isolation from the substrate, reducing detections of slow diffusing carriers. Small diodes ($\approx 2\mu$m diameter) have reduced DCRs, lower capacitances, shorter dead times, lower after-pulsing probabilities and can use micro-lens arrays to increase fill factor.
- The step response induced ISI can be mitigated by feeding modelling back into specifications, with the model outcome suggesting an advantage to arrays of many small SPADs, each with very short dead times, or rather a larger total counting resource available compared to the photon flux.
- To address the principal power penalty (Table 3.6), the fill factor can be increased using shared N-well structures, non-circular geometries, optimal packing, out-of-array logic and NMOS only circuitry.
- To increase data throughput, modulation schemes such as multiple pulse amplitude modulation (M-PAM), wavelength division multiplexing (WDM) and optical many-in-many-out (OMIMO) can be investigated.

7.5 Conclusions

This work discussed the feasibility and issues surrounding multiple SPAD, integrating, all-digital receivers for optical communications. This concentrated on visible light communications as an application. VLC is progressing into commercialisation principally due to the restrictions imposed on the electromagnetic spectrum by the United Kingdom's regulator, Office of Communications (OFCOM) [15], the increase in subscribers and increases in download intensive media streaming.

High optical sensitivity, mass producible and small receivers are required for VLC's target of reaching mobile computers such as smart phones and tablets. A combined approach using high-gain avalanche multiplication, integrating receivers and multiple separate diodes has been explored, with SPAD-based sensors already commercialised into smart phone platforms [237]. The concept and issues have been discussed, with positive communication results showing the scope for further work.

There are two key conclusions from this work. These are:

- Combining the advantages of multiple sensitivity enhancing techniques, while minimizing each of their disadvantages, will prompt high-sensitivity receivers suitable for VLC applications.
- The impulse and step responses of SPAD-based receivers must be addressed to achieve communication bandwidth scaling towards Gigabit per second speeds.

SPAD receivers may be useful for niche applications but their immediate usefulness is restricted due to power, area and error rate limitations. With progression, the advantages of all-digital CMOS can be utilised, incorporating automatic data thresholding and bit error correction onto the device. Single photon sensitivity is suitable when the electrical power to operate the SPADs is lower than the analogue circuits within conventional receivers for the same amplification gain. Likewise, extreme sensitivity may be required to obtain link budgets if small active areas or optical filters are required. It is hoped that the ideas, concepts, implementation and models provided can be useful in continuing progress towards ideal communications receivers.

Appendix A

Author Publications

A Reconfigurable 14-bit 60GPhoton/s Single-Photon Receiver for Visible Light Communications

Edward Fisher, Ian Underwood and Robert Henderson
The Institute for Integrated Micro and Nano Systems, The School of Engineering, The University of Edinburgh
Email: E.Fisher@ed.ac.uk

***Abstract*— A reconfigurable Single-Photon Avalanche Diode integration mode receiver in 130nm CMOS is presented for optical links with an array readout bandwidth of 100MHz. The all-digital 32x32 SPAD array achieves a minimum dead time of 5.9ns, and a median dark count rate of 2.5kHz per SPAD. Pulse shortening increases the dynamic range by preventing pulse overlap. An in-pixel feedback loop allows synchronous or self-clocked asynchronous time division multiplexing. The internal gain of SPADs and spatio-temporal summation removes the need for analogue amplification. A maximum count rate of 58GHz is observed, with SNR of 79dB, a sensitivity of -31.7dBm at 100MHz and a BER of 10^{-9}. The sensor core draws 89mW at maximum count rate.**

I. Introduction

Visible light communications (VLC) [1] is an emerging application for local capacity enhancement in the 100Mb/s to 1Gb/s range utilizing the growth of LED lighting or polymer optical fibers (450-650nm). Recent trends towards integrated CMOS high speed optical receivers have employed photodiodes (PDs) and avalanche photodiodes (APDs) [2]. The low PD gain and thermal noise necessitate the use of intricate high gain transimpedance amplifiers, limiting amplifiers, adaptive equalization or BiCMOS processes.

Single-Photon Avalanche Diodes (SPADs) are p-n diodes biased beyond breakdown into the 'Geiger' region [3]–[5]. Due to the high electric field, avalanche multiplication leads to a large internal gain, creating a density modulated pulse train. Single-photon sensitivity and pico-second time resolution have led to their introduction into a broad range of applications, including quantum key distribution [6]. SPAD arrays in standard CMOS processes have improved with respect to jitter, noise and dead times [3], [4], [7]. SPADs have been used here with the long term aim of power efficient, high sensitivity receivers able to approach the photon shot noise limit. We translate the ability of previous SPAD arrays to suppress noisy diodes [8] in order to improve the noise floor. SPAD communications [6] has so far been limited to the MHz region by dead time, τ_d, and the after-pulsing probability, P(ap). This paper introduces a first generation proof of concept SPAD array (Fig.1) that solves these issues by system level techniques. The array has a sensitivity of -31.7dBm, limited by photon detection efficiency (PDE), fill factor and SPAD noise. The quantum limit of -60.5dBm could be approached in the future, with ≈ 25dB recovered by high fill factor [9], direct fibre abutment and high detection efficiency [3].

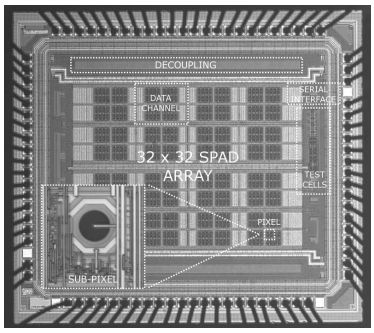


Fig. 1. Sensor micrograph. Die = 2.4 x 2.1mm. Inset: Sub-pixel with SPAD.

II. Array Architecture

The array is designed for on-off keying (OOK), pulse amplitude (PAM) and pulse width modulation (PWM) in conjunction with a CMOS controlled 450nm GaN micro-LED array [1]. It could also be used for multi-level PAM and optical multi-in-multi-out (MIMO) techniques. System configuration, control and data acquisition uses a FPGA and USB link.

A. The SPAD Sub-Pixel

The array is arranged in a hierarchy with the SPADs [4] in $27.5\mu m$ square sub-pixels (Fig.1 inset). This is comprised of configurable active quench active reset (AQAR) or passive quench passive reset (PQPR) circuits (Fig.2). To reduce electrical cross talk, a split power architecture and local decoupling are used. A local (ENABLE) signal can pull SPADs firmly off. The voltage DTCTRL is used to tune the PMOSs in the PQ and AQAR circuits, and directly controls τ_d (Fig.5). The sub-pixel fill factor is limited to 6.65%.

B. The Pixel, Cluster, Data Channel and Full Array

Pixels consist of four sub-pixels (Fig.3) and provide low level spatio-temporal multiplexing [10], [11], which alters the sensitivity of the photon transfer curve (PTC), via the number of currently active SPADs, $N_a = \{1, 2, 4\}$. Pixel area can be split into smaller units which lowers parasitic capacitance, dark count rate (DCR) and jitter [3]. Mono-stables (Fig.2.c) reduce the SPAD τ_d to a shorter pulse $\tau_p = 500ps$. The NAND/NOR tree is used as the output signal, but also iterates

Figure A.1: E.Fisher, I. Underwood and R. Henderson, *A Reconfigurable 14-bit 60GPhoton Per Second Single-Photon Receiver for Visible Light Communications*, The 38th European Solid-State Circuits Conference (ESSCIRC), Bordeaux, France, September 2012

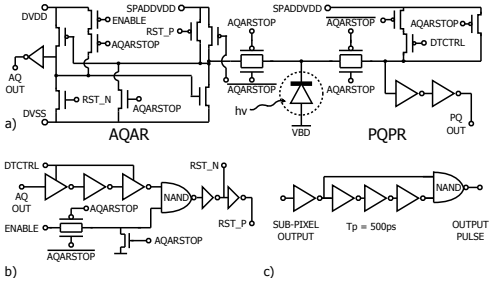


Fig. 2. a) switchable AQAR/PQPR sub-pixel circuit, b) current starved inverter delay adjusting AQAR dead time and c) pixel edge detection circuit.

a state machine, active in $N_a = \{1, 2\}$. $N_a = 1$ enables a single SPAD, iterating sequentially in a time multiplexed manner around the four SPADs, while $N_a = 2$ uses pairs of SPADs, this increases sensitivity. $N_a = 4$ provides maximum sensitivity as a mini silicon photo-multiplier (Si-PM) module. Clusters of $M = 4$ pixels (Fig.3) implement a spatial sum, and the first level of counters, adders and control. In cluster memory, accessed by a serial interface (SI), sets N_a values on a per pixel basis. Upon a clock edge (Fig.4), latch-shift registers capture data from the 6bit counters while a short disable pulse prevents iteration, before resetting. In subsequent cycles, a parallel pipelined adder tree computes the sum of photons within a symbol time (τ_b). A clock multiplexer allows an external pixel clock and a nested decoupling and guard ring structure is used. The 6bit latch can be reconfigured to a chained shift register for DCR maps and imaging.

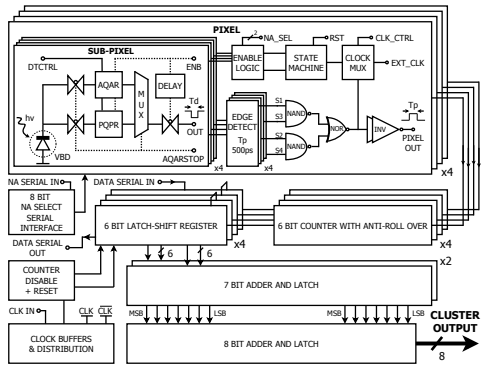


Fig. 3. Cluster block diagram showing a hierarchical addition of pulses. Sub-blocks are rotated to ensure most electronics is at cluster circuit edge.

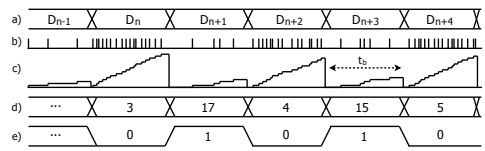


Fig. 4. Timing diagram showing a) input data b) pixel SPAD pulses c) counter iteration during τ_b d) latched 6b data e) data output post threshold.

The hierarchy continues to 16 data channels, each of $R = 4$ clusters and finally to a 100MHz 14-bit array sum with a 4 clock period latency. Global sensor conditions are set by an 11-bit SI. The weighted summation algorithm of SPAD pulses allows an unprecedented theoretical maximum count rate, C_{max}, while maintaining single photon sensitivity.

$$C_{max} \approx \#SPADs/\tau_{dmin} \tag{1}$$

III. Motivation and Theory

A. Temporal Coincidence Reduction

To register detections within τ_d, pulses must not prevent counting of other pulses [8], [11], an effect called temporal co-incidence (TC). With pulse edge detection, the probability of $\geq$ 2 SPADs firing simultaneously, giving a single count with an expected pixel count rate, E[Ψ_{px}], is to first order:

$$E[P(TC)] = \left(1 - \frac{1}{N_a}\right)\left(1 - \mathrm{e}^{-E[\Psi_{px}(\tau_b)]\,E[\tau_p]}\right) \tag{2}$$

If the pixel is at half maximum count rate ($\approx$ 1.5x10^8), a typical AQ $\tau_d = 13.48ns$ and in $N_a = 4$ mode then $P(TC) = 0.054$, reducing τ_p to 250ps gives 0.027. If $\tau_p = \tau_d$, then $P(TC)$ increases to 0.648. N_a should be small for spatially multiplexed groups in order to reduce count losses.

B. Time Division Multiplexing (TDM)

TDM [10], [11] allows a per pixel reconfigurable weighted sum of counts, with weights of $\{0, 0.25, 0.5, 1\}$. The technique also prevents powerful pulses from saturating all SPADs. Forcing SPADs below breakdown while others are operating, also allows reduction of P(ap) and DCR, as carriers can be released from traps without inducing avalanche multiplication.

C. Spatial Oversampling and Photon Counting

For photon counting applications, sampling over a matrix of identical independent SPADs, can reduce the signal variability. Assuming negligible correlated cross talk, uniform noise and a uniformly distributed coherent 2D Poisson process over the data channel (250μm square), let $\Psi(\tau_b)$ be a noiseless photon number $|n\rangle$ and $\eta(\tau_b)$ be additive photon shot noise. The SPAD output is $\Phi(\tau_b) = \mathrm{c}\,(\Psi(\tau_b) + \eta(\tau_b))$ where 'c' is a sigmoid function with the constant value of one over the central region of the PTC. The data channel takes the sum of these signals, $S_{dc}(\tau_b)$ with a division by n_{div} performed in firmware, however this could be implemented on chip. If non-TDM operation is assumed, the mean counts and standard deviation to first order will be:

$$E[S_{dc}(\tau_b)] = \frac{\sum_{k=1}^{R}\sum_{j=1}^{M}\left(1 - E[P(TC)_{j,k}]\right)\sum_{i=1}^{N_a}\left(c_{i,j,k}\,\Psi_{i,j,k}(\tau_b)\right)}{n_{div}} \tag{3}$$

$$\sigma_{S_{dc}(\tau_b)} = \frac{\sqrt{RMN_a}(1 - E[P(TC)])\left(E(c)\,\sigma_{\eta(\tau_b)}\right)}{n_{div}} \tag{4}$$

This maximises measurement accuracy, and increases the probability P($n|n$). Array noise, and non-uniformity, will increase $\sigma_{S_{dc}(\tau_b)}$ to super-Poisson statistics. Note that by $\sigma^2(c.X) = c^2.\sigma^2(X)$ incident photon statistics remain valid.

Figure A.2

D. Theoretical BER Analysis

Over the operating region for a required error rate, the DCR noise distribution (N_0, σ_0) is much lower than the photon distribution received (N_1, σ_1). The decision point, N_{th} for OOK is taken to ensure the minimum BER.

$$N_{th} = \frac{\sigma_0 N_1 + \sigma_1 N_0}{\sigma_0 + \sigma_1} \quad BER = \frac{1}{2}\text{erfc}\left(\frac{(N_1 - N_{th})}{\sigma_1\sqrt{2}}\right) \quad (5)$$

Sensitivity analysis shows that for measurement accuracy $n_{div} \gg 1$ and $n_{div} \gg |n\rangle$, however from a communications viewpoint $\#SPADs \gg n_{div}$. The extinction ratio, relative intensity noise, clock jitter and average optical power will all impact the bit error rate (BER). Correlation between noise sources can be modelled by the covariance, $Cov(X, Y)$. By time integrating n counts per bit, the impact of in-symbol intensity noise is reduced, similar to integration mode receivers. Inter-symbol intensity noise will significantly increase the σ_1 and σ_0 values. The non-uniform Nyquist sample rate is $E[C_{max}]/2$.

IV. Experimental Results

Fig. 5 shows the achieved dead time of a test sub-pixel. The AQAR circuit has a minimum τ_d of $10.6ns$ limited by the designed delay chain. The PQPR circuit is able to achieve a lower τ_d, limited by the R_{on} of the quenching PMOS and SPAD parasitic capacitances but has a significantly worse standard deviation giving additional PTC noise.

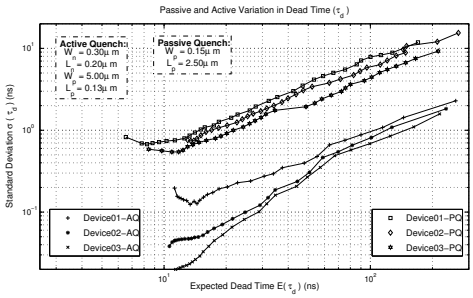


Fig. 5. Dead time standard deviation. AQ consistently has lower $\sigma(\tau_d)$.

The array has a mean DCR of $7.27kHz$ (Fig.6). This is higher than reported in [4] for the same junction, which suggests a non-corrected noise floor better than 100kHz. Noisy devices can be weighted or removed which lowers the noise floor and increases dynamic range. As the symbol time decreases, the probability of a dark count in τ_b reduces.

After-pulsing of $(n = 8)$ sub-pixels was measured at $\tau_d = 13.5ns$ using a $1\mu s$ window, triggered by dark count events. After-pulsing, which would cause inter-symbol interference, varied with $\mu = 0.92\%$, $\sigma = 0.73\%$, a minimum of 0.1%, and a maximum of 2%. The PTC (Fig.7) shows that the device saturates at C_{max} and the various N_a modes show the sensitivity and noise trade offs. The dynamic range can be improved by lower dead times, the small capacitance of CMOS SPADs and high speed of integrated nanometre scale digital circuits [3], [5], however lower DCR noise would

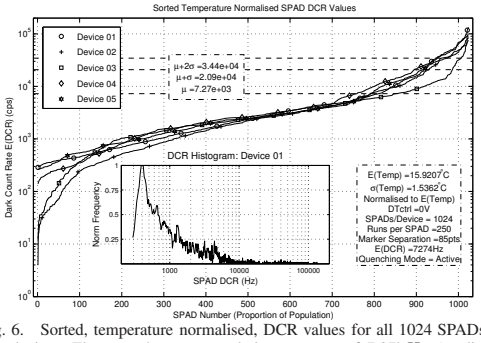


Fig. 6. Sorted, temperature normalised, DCR values for all 1024 SPADs on five devices. The array has a mean dark count rate of $7.27\text{k}Hz$ (median = $2.53\text{k}Hz$, $\sigma(DCR) = 1.36kHz$). For Device01 89.6% are below $\mu + \sigma$.

give a significant improvement. Device01 homogeneity under uniform illumination $(30\mu s, \approx 1.4\mu W/cm^2)$ was on average 2.26% from E(signal), with $\sigma(counts) = 0.0826 * \mu$, and only 1.2% of devices outside $\pm 3\sigma$. This non-uniformity has no spatial component which suggests it is caused by SPAD, and circuit variation rather than power supply variation.

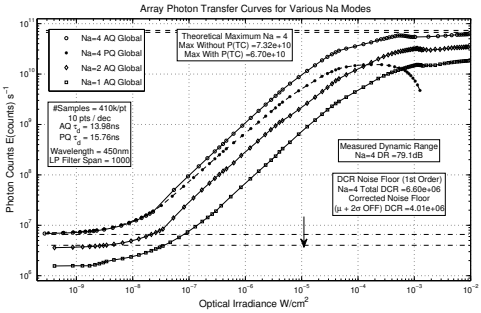


Fig. 7. Photon transfer curve. The $N_a = 4$ PQ curve shows poor gain linearity, the 1/e reduction in counting rate and SPAD paralysis. The AQ also shows paralysis above $10mW/cm^2$. The $Na = 1$ curve will tend to C_{max} at optical powers approaching $100mW/cm^2$ [10]. 57 Pixels with SPAD DCRs $\geq \mu + 2\sigma$ are turned off using $Na = 3$, reducing noise by 4.3dB.

Fig.8 presents BER analysis over a selection of frequencies and optical powers and assumes a 100% extinction ratio. The BER was obtained using a CW light source, experimental values for N_1, N_0, σ_1 and σ_0 and equation (5), this removes the influence of extinction ratio and source bandwidth (Fig. 9). The modulated count range of Fig. 9, suggests the array is suitable for $\{4, 8, 16\}$-PAM. A time domain sweep shows the number of samples below the N_{th} threshold. The quantum limit at 100MHz, corrected for the array fill factor and PDE is presented, based on experimentally derived counts per τ_b. The observed gap could be due to source intensity noise, array noise, non-linearity and the assumptions of equation 5. The array achieves a sensitivity of -31.7dBm for a BER of 10^{-9} at $100MHz$. If a fill factor of 50% and a PDE of 45% are assumed, a similar array could achieve a quantum limit of -52.5dBm. I/O frequency can be increased by conventional low voltage and differential techniques. In comparison [2] achieves -19dBm (30.7pJ/bit) corrected for the same BER and

Figure A.3

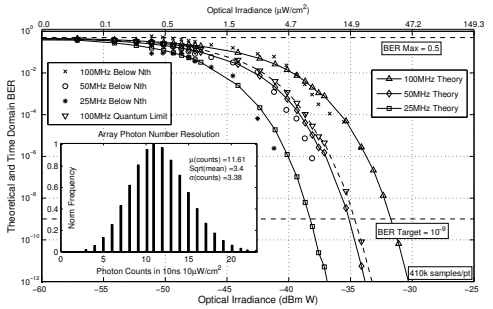


Fig. 8. BER analysis using the theoretical BER and a time domain search of samples below the N_{th} threshold. Analysis assumes a 100% extinction ratio. Target BER of 10^{-9} is reached at -31.7dBm for 100MHz. Photon number resolution is shown inset, $\mu = 11.61$, $\sigma = 3.38$. Array area = $0.0212cm^2$

frequency. The core power consumption of $87.9mW$ scales with photon flux and system clock giving, 372pJ/bit and a detection power efficiency of 3.16pJ/photon.

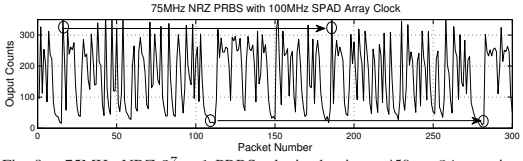


Fig. 9. 75MHz NRZ $2^7 - 1$ PRBS, obtained using a 450nm 34μm micro-LED [1]. Measured with high DCR SPADs turned off and $\approx 40uW/cm^2$. Array hysteresis is observed from $1 \rightarrow 1$ and $0 \rightarrow 0$.

V. SPAD Trends for Optical Communications

SPAD developments are expected to allow the quantum limit to be approached by improvements in fill factor and detection efficiency. A deep junction CMOS SPAD shows spectral sensitivity above the blue, its PDE of 44% at 690nm and 20% at 850nm place it at a prime region for polymer and silica optical fibre attenuation windows [12]. Quenching circuits have concentrated on high density integration [3] and increased count rates [7]. The fill factor of other devices has been increased by N-Well sharing [9] but circuits are now limited by the SPAD itself. Circuits with sine wave and hybrid gating show sampling at 2.23GHz [6], albeit with lower sensitivity. System level techniques [6], [10], [11] are necessary while research continues on noise sources, integration, bandwidth and yield. A $2\mu m$ SPAD has achieved median DCR of 0.5Hz and a single photon jitter of 103ps [13]. The internal front end gain of multiple adaptive detectors, active quenching circuits and pulse shortening can allow lower optical powers achieve standard BERs and longer transmission range.

VI. Conclusion

This work presents a SPAD array designed for optical communications and photon counting, summarised in Table.I. Specifically spatio-temporal summation improves BER, and reconfigurability of individual pixels allows high noise SPADs to have a lower weight in the sum. Further work would include on chip PLLs, clock data recovery, data thresholding and increased fill factor for abutted optical fibres. Further characterisation, modelling and Multi-PAM modulation studies will be carried out to investigate additional design and bandwidth improvements.

TABLE I
General Results and Chip Information

Parameter:	Value:	Units:	Notes:
CMOS Process	130nm Imaging		90nm Metallization
Max Count Rate	58.13x10^9	s^{-1}	@ $\tau_d = 13.85$ns
Sensitivity	-31.7	dBm	450nm 10^{-9} BER
Dynamic Range	79.1	dB	At C_{max}
Peak PDE	20 [4]	%	Vex=1.2V 450nm
Digital DVDD	1.2 / 3.3	V	Core / Pads
SPAD DVDD	1.35	V	Excess Bias
SPAD VBD	-13.05	V	Break Down
Min Dead Time	5.87 ± 2.7	ns	Passive Quench
Min Dead Time	10.61 ± 0.04	ns	Active Quench
Max Count Power	190	mW	Exc Pads, Inc VBD
Array Fill Factor	2.42	%	Exc Pads etc

Acknowledgments

The authors thank ST Microelectronics and both Jonathan McKendry and Anthony Kelly for μLED modulation assistance. The work was funded by the Hypix project, EPSRC UK. www.hypix.photonics.ac.uk.

References

[1] J. McKendry, D. Massoubre, *et al.*, "Visible-Light Communications Using a CMOS-Controlled Micro-Light- Emitting-Diode Array," *Lightwave Technology, Journal of*, vol. 30, no. 1, pp. 61 –67, 2012.
[2] F. Tavernier and M. Steyaert, "High-Speed Optical Receivers With Integrated Photodiode in 130 nm CMOS," *Solid-State Circuits, IEEE Journal of*, vol. 44, no. 10, pp. 2856 –2867, 2009.
[3] J. Richardson, E. Webster, *et al.*, "Scaleable Single-Photon Avalanche Diode Structures in Nanometer CMOS Technology," *Electron Devices, IEEE Transactions on*, vol. 58, no. 7, pp. 2028 –2035, 2011.
[4] J. Richardson, L. Grant, *et al.*, "Low Dark Count Single-Photon Avalanche Diode Structure Compatible With Standard Nanometer Scale CMOS Technology," *Photonics Technology Letters, IEEE*, vol. 21, no. 14, pp. 1020 –1022, 2009.
[5] M. Ghioni, A. Gulinatti, *et al.*, "Progress in Silicon Single-Photon Avalanche Diodes," *Selected Topics in Quantum Electronics, IEEE Journal of*, vol. 13, no. 4, pp. 852 –862, 2007.
[6] J. Zhang, P. Eraerds, *et al.*, "2.23 GHz gating InGaAs/InP single-photon avalanche diode for quantum key distribution," *PROC.SPIE*, vol. 7681, p. 76810Z, 2010.
[7] A. Eisele, R. Henderson, *et al.*, "185 MHz Count Rate, 139 dB Dynamic Range Single-Photon Avalanche Diode with Active Quenching Circuit in 130nm CMOS Technology," in *2011 International Image Sensor Workshop*, IEEE. IEEE Electron Devices Society, 2011, pp. 278–280.
[8] L. Braga, L. Pancheri, *et al.*, "A CMOS mini-SiPM detector with in-pixel data compression for PET applications," in *Nuclear Science Symposium and Medical Imaging Conference*, 2011, pp. 548 –552.
[9] R. Henderson, E. Webster, *et al.*, "A 3x3, 5um pitch, 3-transistor single photon avalanche diode array with integrated 11V bias generation in 90nm CMOS technology," in *Electron Devices Meeting (IEDM), 2010 IEEE International*, 2010, pp. 14.2.1 –14.2.4.
[10] S. Castelletto, I. Degiovanni, *et al.*, "Reduced Deadtime and Higher Rate Photon-Counting Detection using a Multiplexed Detector Array," *Journal of Modern Optics*, vol. 54, no. 2, pp. 337–352, 2007.
[11] W. H. Farr, J. Gin, *et al.*, "Gigahertz bandwidth photon counting," in *Proc. SPIE 7320*. SPIE, 2009, p. 732006.
[12] E. A. G. Webster, J. A. Richardson, *et al.*, "A Single-Photon Avalanche Diode in 90-nm CMOS Imaging Technology With 44% Photon Detection Efficiency at 690 nm," *Electron Device Letters, IEEE*, vol. PP, no. 99, pp. 1 –3, 2012.
[13] E. Webster, J. Richardson, *et al.*, "Single-Photon avalanche Diodes in 90nm CMOS Imaging technology with sub-1Hz Median Dark Count Rate," in *International Image Sensor Workshop*, 2011, pp. 262–265.

Figure A.4

1638 IEEE JOURNAL OF SOLID-STATE CIRCUITS, VOL. 48, NO. 7, JULY 2013

A Reconfigurable Single-Photon-Counting Integrating Receiver for Optical Communications

Edward Fisher, *Student Member, IEEE*, Ian Underwood, and Robert Henderson, *Member, IEEE*

***Abstract*—A reconfigurable Single-Photon Avalanche Diode integrating receiver in standard 130 nm CMOS is presented for optical links with an array readout bandwidth of 100 MHz. A maximum count rate of 58 G photon/s is observed, with a dynamic range of ≈79 dB, a sensitivity of ≈−31.7 dBm at 100 MHz and a BER of $\approx 1 \times 10^{-9}$. The sensor core draws 89 mW at the maximum count rate and obtains a peak SNR of ≈157 dB. We investigate the properties of the receiver for optical communications in the visible spectrum, using its added functionality and reconfigurability to experimentally explore non-ideal influences. The all-digital 32 × 32 SPAD array, achieves a minimum dead time of 5.9 ns, and a median dark count rate of 2.5 kHz/SPAD. The internal gain of SPADs and spatio-temporal summation removes the need for analogue amplification. High noise devices can be weighted or removed to optimize the SNR. The power requirements, transient response and received data are explored and limiting factors similar to those of photodiode receivers are observed.**

***Index Terms*—CMOS, IEEE802.15.7, integrated, integrating receiver, interchip, low-light receiver, OEIC, photon counting, single-photon avalanche diode (SPAD), visible light communications (VLCs).**

I. Introduction

VISIBLE light communications (VLCs) [1], [2] is emerging for local capacity enhancement at 100 Mb/s to 1 Gb/s, using solid-state LED lighting [3] or polymer optical fibers (450–650 nm) [4] and is being standardized (IEEE P802.15.7) [1]. The impetus has been the wide, licence free, visible and near-IR spectrum and immunity to electromagnetic interference [5], [6]. VLC is established at bit error rates (BERs) of 10^{-3} to 10^{-6} [3], [7] with forward error-correcting codes (FECC) already specified [1]. A number of transmitters and receivers [2] are aiming for native BER $\leq 10^{-9}$. Optical communications [6], [8], [9] may be a route for on-chip interconnects, optical PCB [10], chip to chip and rack to rack backplanes [5]. These are high-performance applications; however low-cost, moderate performance implementations may be important for last mile links [7]. We focus on CMOS integrated receivers combinable with on-chip processing, without large buffers or dedicated analogue blocks. Integrated CMOS receivers have employed photodiodes (PDs) and avalanche photodiodes (APDs) [6], [11]–[14]. The low PD sensitivity requires high gain transimpedance amplifiers (TIA), limiting amplifiers (LA) and adaptive equalization. The challenge has been amplification with low thermal noise, while ensuring a high front-end bandwidth. Equalization compensates *extrinsic bandwidth*, being initial RC circuit poles, and *intrinsic bandwidth*, being carrier diffusion limitation [15]. APDs shift some gain into the diode utilising avalanche multiplication of electron-hole pairs into many charge carriers (≈1000). By doing so, TIAs can use lower gain and area. Despite advances, high optical powers are required for low BERs, increasing total link power budgets. In this paper we expand upon earlier work [16]. High reverse bias, moves gain fully into the diode, requiring only a small digital inverter (Fig. 1). The output rate is limited; however we aim to explore factors that may affect SPAD receivers using numerous diodes. Section II introduces the SPAD structure, operation and performance. In Section III the proof-of-principle, experimental receiver is discussed (Fig. 2). The aim of this receiver and the added real-estate for testability, is to discover the factors that affect high level performance, interrelationships and optimization paths, allowing dynamically reduced noise floor, nonlinearities and retained sensitivity. In Section IV motivation is presented. Experimental results are discussed in Section V. We then examine, in Section VI, the trends in SPAD research. Section VII, compares the device with traditional receivers and Section VIII gives conclusions along with further work.

II. Single-Photon Avalanche Diodes

Single-photon avalanche diodes (SPAD), are p–n diodes biased beyond reverse breakdown in the "Geiger" region [17]–[20]. Due to the high electric field, avalanche multiplication of an electron-hole pair, leads to a large internal gain, ≥10,000. SPAD discharge creates a digital level, pulse-density modulated output, proportional to the photon flux, with additive random dark count rate (DCR). Due to high speed multiplication, the edge is accurate to sub-ns levels [17], [21]. Once a SPAD fires, it quenches and resets using a passive or active circuit [17]. During this cycle, it is unresponsive for a SPAD and circuit specific dead time, τ_d. This limits counting rates, however second order effects such as after-pulsing $P(ap)$, prevent arbitrarily short τ_d. SPAD circuits tend to be small digital implementations [22], scaling with process [20], [23] and have been used in a range of applications, including quantum key distribution [24], 3-D imaging [25] and biomedical sensing [26]. Arrays in standard CMOS [25] have improved jitter, noise and dead times [19], [20], [22], [27]. High sensitivity receivers using SPADs, aim to approach the photon shot noise limit. We

Manuscript received November 30, 2012; revised February 16, 2013; accepted March 01, 2013. Date of publication April 08, 2013; date of current version June 21, 2013. This paper was approved by Guest Editors Antonio Liscidini and Doug Smith.

The authors are with the Institute for Integrated Micro and Nano Systems, School of Engineering, University of Edinburgh, Edinburgh EH9 3JL, U.K. (e-mail: E.Fisher@ed.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSSC.2013.2253222

Figure A.5: E.Fisher, I. Underwood and R. Henderson, *A Reconfigurable Single-Photon-Counting Integrating Receiver for Optical Communications*, The IEEE Journal of Solid-State Circuits: ESSCIRC Special Edition 2013, Published 1st July 2013

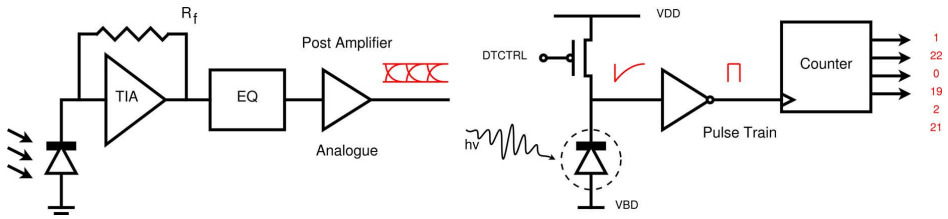


Fig. 1. Left: A traditional photodiode and TIA in continuous time operation. Right: A SPAD and digital counter operating in discrete time.

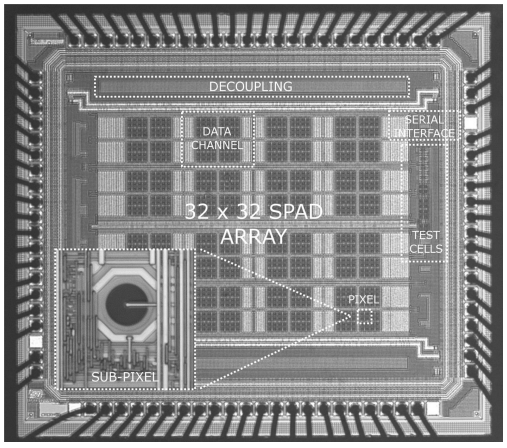


Fig. 2. Sensor micrograph. Die = 2.4 × 2.1 mm. Inset: Subpixel with SPAD.

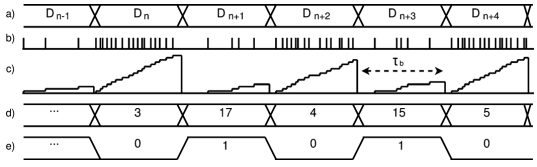


Fig. 3. Timing diagram showing (a) input data, (b) pixel SPAD pulses, (c) counter iteration during τ_b, (d) latched 6b data, and (e) data output post threshold.

translate the ability of previous SPAD arrays to suppress noisy diodes [28], in order to improve the signal-to-noise ratio (SNR), but extend this to a weighting algorithm. SPAD communications [24], [29], [30] has been limited to the megahertz region by τ_d, and $P(ap)$, due to the use of single discrete or III–VI SPADs. ON–OFF Key (OOK) modulation with SPADs has been used previously, as initial proof of principle for on-chip links [30], along with pulse-position modulation (PPM) [29].

III. Array Architecture and Integrating Receivers

Clock and data recovery (CDR) circuits [31], use single or multisampling to threshold and retime the data. As SPADs give a discrete time signal, a classical CDR circuit, such as phase picking, is not suitable. Instead detections must be decoded before a CDR-like operation can be performed. The receiver uses an integrator, taking a summation of samples, giving a level of data robustness, due to the integrated intrasymbol intensity [32] and Poisson noise. The array is designed for OOK, pulse amplitude (PAM) and pulse width modulation (PWM) with a CMOS controlled 450 nm GaN μLED array [2]. For OOK and MPAM, the array provides a digital output proportional to the optical pulse height. For PWM, the number of photons is dependent on the pulse width. These modulation schemes will return binary data once thresholded. The reconfigurable hierarchical receiver, creates a H-tree-like parallel pipelined adder tree, trading off responsivity and noise. The sensor presents 16 identical independent channels, for abutted fibers or multiple-input-multiple-output (MIMO) transmission. Each channel provides a 10-bit number, representing the number of photons received within an integration window. Within a data symbol, we wish to maximize the number of photon-detections Fig. 3. This is problematic, as a fired SPAD saturates proceeding logic. To combat this, two approaches could be taken; to use a counter per SPAD, constituting an area increase, or a small "OR" gate signal summation. The integration is equal to the clock period, although a short blanking interval allows data latching. This increases sensitivity, implying a high integration to symbol period ratio. However this period should be short to ensure Poisson limited counts, rather than widening by photons at transitions. The array is similar [33], where 256 direct-to-digital, reconfigurable receiver-pixels, achieve 500 kb/s/channel at −47 dBm.

A. SPAD Subpixel

SPADs [19] are held in subpixels (Fig. 2 inset), using configurable active quench active reset (AQAR) [34] or passive quench passive reset (PQPR) circuits (Fig. 4), decreasing fill factor (FF), but allowing investigation of their merits. AQAR requires more area and power, but has higher counting rates, no SPAD paralysis [27] and tight τ_d control. PQPR uses transistors "PQ pMOS" with 2.5 μm length while AQAR uses the wide "AQ pMOS," (W/L = 38.5) for active reset, and a small (W/L = 1.5) "AQ nMOS," for active quenching. The sizes are designed for fast quenching but suitable leakage paths for quiescence. The core uses a 1.2 V $DVDD$ supply, and the excess bias voltage, $SPADDVDD$, is separate, with distributed decoupling, reducing electrical crosstalk. A local $ENABLE$ reduces the excess bias to 0 V. $DTCTRL$, directly controls τ_d through gate bias, by modifying the PQ PMOS R_{on}, or the transitions of AQ current-starved inverters. Each subpixel is surrounded by a 1.2 V reverse biased junction reducing cross talk [11], [31], [35] and provides further distributed decoupling. This

Figure A.6

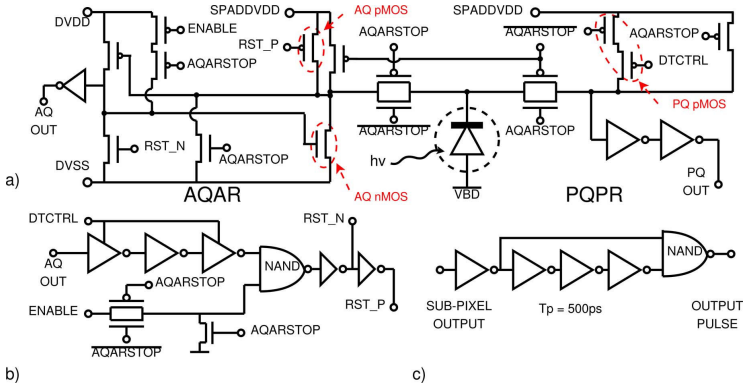


Fig. 4. (a) Switchable AQAR/PQPR subpixel circuit, (b) current starved inverter delay adjusting AQAR dead time, and (c) pixel edge detection circuit.

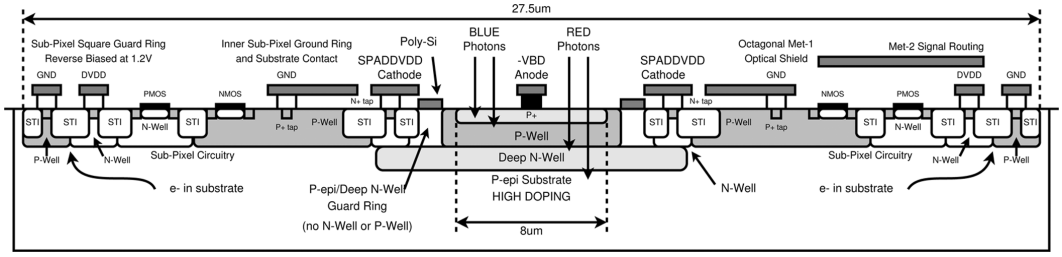


Fig. 5. Cross section of a subpixel and SPAD within a standard 130 nm CMOS imaging process.

should attract and sink carriers to ground, reducing transient decay constants. An electron will diffuse towards the P-well due to concentration gradients, where it will recombine. Likewise, an electron may drift by attraction to the positively biased N-Well. Secondary photons [21], [35] can propagate in direct or indirect paths, causing optical cross talk. Although the outer guard ring will present a barrier to direct path photons [35]. The subpixel fill factor is limited to 6.65%, due to the SPAD guard ring and design rules. The negative drive AQAR is a reimplementation [34], which has already shown high counting rates [27]. The SPAD has also been implemented [19] obtaining a DCR of 25 Hz at room temperature, and a dynamic range of 139 dB [27]. This was achieved using a retrograde guard ring [19], a low capacitance, and a low AQ τ_d [27], [34]. The SPAD in its CMOS environment is presented in Fig. 5. The multiplication region is formed at the P and N wells, exhibiting maximum photon detection efficiency (PDE) and lowest jitter [19] at blue wavelengths, due to the shallow junction.

B. Pixel, Cluster, Data Channel, and Full Array

Pixels (Fig. 6) consist of four subpixels providing spatio-temporal multiplexing. This alters the sensitivity, giving weightings via the number of currently active SPADs, $N_a = \{1, 2, 4\}$. Pixel area is split lowering SPAD capacitance, DCR, and jitter [20]. Using values from [20], yield improves with smaller devices. With four 8 μm diameter SPADs, the median pixel DCR would be 180 Hz. In comparison, a single SPAD of equal area, (16 μm diameter), gives a DCR of over 1 kHz with an increased probability of a SPAD containing a trap. In [36] DCR as a function of area shows increase in both overall DCR and the cumulative distribution (CDF), reducing yield. Jitter improves with smaller area [20], due to faster avalanche spreading [21] and lower capacitance. Mono-stables [Fig. 4(c)] reduce τ_d to a shorter pulse $\tau_p = 500$ ps, reducing pulse overlap [37] at high counting rates [16]. While the mono-stables add jitter, the leading edge is fast, the jitter is dominated by the SPAD itself (≈200 ps) [19] and the integration with a blanking period reduces jitter influence. The NAND/NOR tree is used as the output, iterating a state machine, active in $N_a = \{1, 2\}$. $N_a = 1$ enables a single SPAD, iterating in a time multiplexed manner, while $N_a = 2$ uses pairs of SPADs, increasing sensitivity. $N_a = 4$ provides maximum sensitivity as a mini silicon photo-multiplier (Si-PM) module. Clusters of $M = 4$ pixels implement spatial summation into 8-bits, along with counters, adders, and control. Fig. 6 shows the addition of two 6-bit counters into a 7-bit bus, replicated, operating in parallel, for the other pixels. An 8-bit memory, accessed by a 512-bit serial interface (SI), sets N_a values on a "per pixel" basis. Upon a clock edge (Fig. 3), latches register data from the counters, whilst a disable pulse prevents iteration, before reset and reenable. The 6-bit latches can be reconfigured to a chained shift register (1536-bit) for device characterization. Jitter in the pipeline is due to clock stability and the large clock network, representing statistical fluctuation in integration period. For single-photon counting, Poisson variation and source noise will be dominant. The hierarchy continues to an object (one of 16 channels) (Fig. 7), each of $R = 4$ clusters, where four 8-bit buses are added through two 9-bit adders. This forms a final 10-bit representation of the number of pho-

Figure A.7

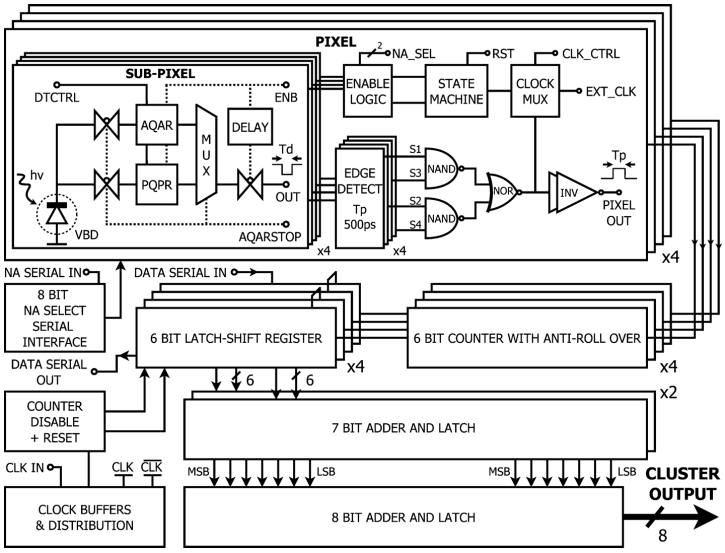


Fig. 6. Cluster block diagram showing a hierarchical addition of pulses. Subblocks are rotated to ensure most electronics is at cluster circuit edge.

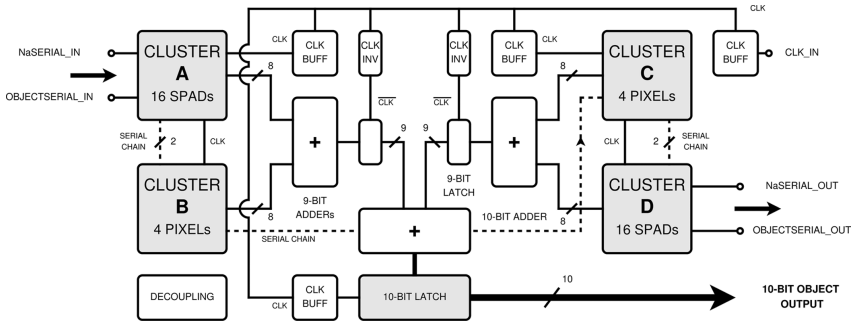


Fig. 7. Data channel block diagram showing a hierarchical addition of four cluster summation buses.

tons received on the channel active area in a symbol integration period. The hierarchy culminates in a 100 MHz 14-bit array sum with a 4 clock period latency. For the 1024 SPADs, and a measured AQ $\tau_d = 13.5$ ns (0 V control bias), the weighted summation allows an unprecedented theoretical maximum count rate, $C_{\max}$, of 75.8×10^9 photons/s, while maintaining single photon sensitivity

$$C_{\max} \approx \#\text{SPADs}/\tau_d. \tag{1}$$

IV. Motivation and Theory

A. Time Division Multiplexing (TDM)

TDM [37], [38] allows a per pixel, reconfigurable, sum of counts, with weights of $\{0, 0.25, 0.5, 1\}$. As SPADs are forced below breakdown while others are operating, $P(ap)$ and DCR should be reduced, due to carrier release without avalanche. In the $N_a = 1$ TDM mode, SPAD_1 is enabled with SPADs 2, 3, 4 below breakdown. The circuit will remain in this state until an event occurs (Fig. 8). Dark counts within other SPADs are ignored, reducing group DCR to that of the lowest noise device. It reduces the effective PDE by 6 dB with reduction in PDE for $N_a = 2$ lessened to 3 dB. Used sparingly with the locations of high DCR SPADs, it could abate the receiver non-idealities without sensitivity degradation.

B. Theoretical First-Order BER Analysis

Over the operating region, the DCR noise (N_0, σ_0) is much lower than the photon distribution (N_1, σ_1). The dc optical power BER can be estimated using these, with the decision point, N_{th}, for OOK modulation, taken to ensure minimum BER.

$$N_{\text{th}} = \frac{\sigma_0 N_1 + \sigma_1 N_0}{\sigma_0 + \sigma_1}, \quad \text{BER} = \frac{1}{2}\,\text{erfc}\left(\frac{(N_1 - N_{\text{th}})}{\sigma_1\sqrt{2}}\right). \tag{2}$$

Figure A.8

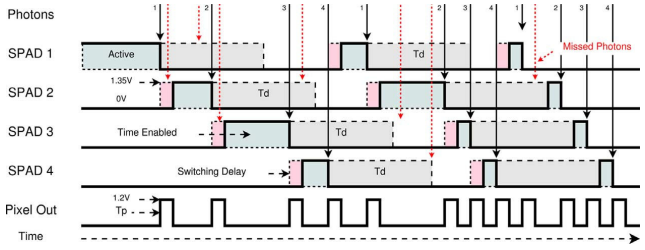


Fig. 8. Timing diagram for the $N_a = 1$ time division multiplexing mode. Red arrows indicate missed photons incident on inactive SPADs or during switching, causing a reduction in effective PDE.

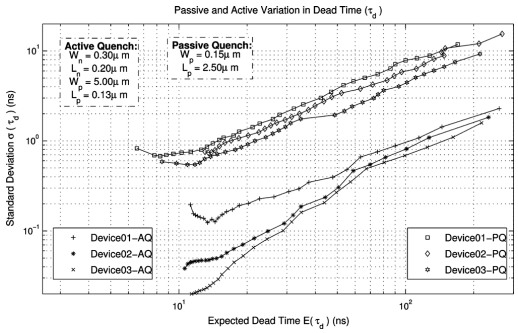


Fig. 9. Expected dead time, controlled via a PMOS gate bias, $DTCTRL$, and the observed standard deviation in dead time. Active quenching consistently has lower $\sigma(\tau_d)$.

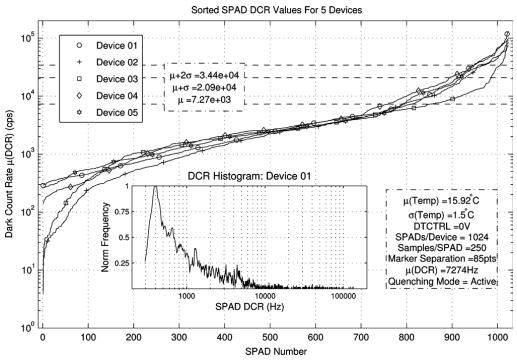


Fig. 10. Sorted DCR values for all 1024 SPADs on five devices. The array has a mean dark count rate of 7.27 kHz (median $= 2.53$ kHz, $\sigma(\mathrm{DCR}) = 1.36$ kHz). For Device01 89.6% are below $\mu+\sigma$. Two integration periods were run with 250 experiments per integration per SPAD. The average temperature was measured using an on PCB sensor at 15.9 °C.

The extinction ratio, intensity noise, and clock jitter will all impact the BER, causing deviation from this first order estimate. Sensitivity analysis shows that the N_0 to N_1 count gap is of high importance [16]. Ideally, N_0 should be close to zero, a case called the quantum limit, QL, limited only by the statistics of light. For design, we require a low N_0, a large $N_1 - N_0$ separation and ideally Poisson limited distributions with no mid-gap counts. Due to the exponential interarrival time, the receiver samples the input in a random manner. This is known as nonuniform sampling and has a Nyquist rate of $E[C_{\max}]/2$.

V. Experimental Results

Configuration, control, and data acquisition, uses a Xilinx Spartan III FPGA and USB 2.0 PC link. The output is stored in a first-in-first-out (FIFO) register prior to streaming. This differs from the optimized single-bit low-voltages of other receivers, but allows characterization, along with signal statistics. This was chosen to maximize the scope of experimental results, including the ability to read individual SPAD noise levels. We characterize the array and 14-bit output, as time-interleaving was used for the 16 channels to prevent pad limitation.

A. SPAD Dead Time (τ_d)

The graph (Fig. 9) plots the mean τ_d against the standard deviation of the τ_d, controlled by $DTCTRL$. The AQAR has a minimum τ_d of 10.6 ns, limited by the designed delay and obtains τ_d distributions an order of magnitude lower variability. Variation in τ_d increases uncertainty in photon number introducing non-ideal secondary behavior. PQ obtains a lower τ_d in this device, limited by the PMOS and the parasitic capacitances, but with severe variability. Active quenching is preferable for low BER and high $C_{\max}$, and can be designed for arbitrary τ_d.

B. SPAD Dark Count Rate

SPADs exhibit a DCR similar to PD dark current. This is caused by thermal carriers and is exacerbated by traps. The DCR from 1024 SPADs are presented from five chips. These are sorted by mean, allowing observation of the pseudo-CDF. The array has a mean DCR of 7.27 kHz (Fig. 10). This is higher than [19], [27] for the same junction, which suggested a preimplementation noise floor $\leq$100 kHz. The DCR change is due to foundry process changes between [19] (2009) and this device. As the distribution has a high positive skew, (shown in the logarithmic x-axis DCR histogram for device 01, shown inset), the median DCR of 2.5 kHz is more representative. With the DCR being predominantly influenced by process, arrays must be viewed in terms of the proportion above an acceptable limit. For device 01, 89.6%, and 92.6% of devices are below the levels of $\mu+\sigma$ and $\mu+2\sigma$, respectively, and the array achieves 50.3%

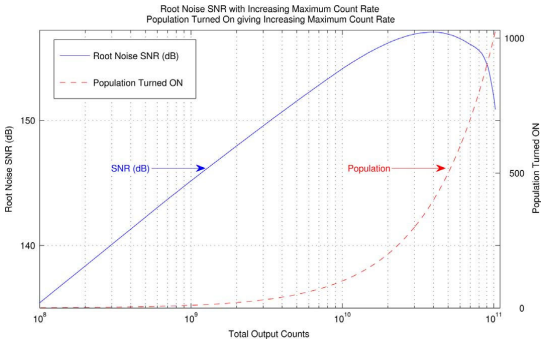


Fig. 11. Maximum signal to DCR noise at $C_{\max}$ with increasing "Enabled" population (dashed line).

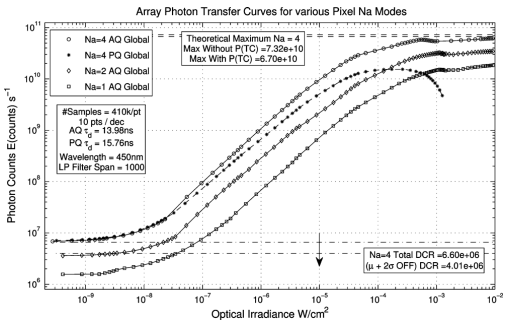


Fig. 12. Photon transfer curve. The $N_a = 4$ PQ curve shows poor gain linearity, the 1/e reduction in counting rate and SPAD paralysis. The AQ also shows paralysis above 10 mW/cm². The $Na = 1$ curve will tend to $C_{\max}$ at optical powers approaching 100 mW/cm² [38]. 57 Pixels with SPAD DCRs $\geq \mu + 2\sigma$ are turned off using $Na = 3$, reducing noise by 4.3 dB.

below the median. The DCR map, obtained using the pixel read SI, shows no spatial dependence, however there is local correlation within pixels.

C. Reconfigurability of Signal to DCR Noise Ratio

In Fig. 11, we assess the SNR, as we sequentially add contributions from the ordered DCR set of device 01. SNR_i represents the noise of the sum of SPADs currently turned on (from 1024), and the total expected maximum signal for that set. This is expressed by the iterative SNR [27], where $i \in 1, 2, \ldots, 1024$ is the number currently connected to the output

$$\text{SNR}_i = 20 \log_{10} \left(\sum_{j=1}^{i} \left(\frac{1}{\tau_d} \right)_j \left(\sum_{j=1}^{i} \text{DCR}_j \right)^{-\frac{1}{2}} \right). \quad (3)$$

The SNR increases due to low τ_d, (assumed equal here), increasing $C_{\max}$ more rapidly than the noise of the sum of low-DCR SPADs. For system optimization, we wish to know what proportion must be turned off to maximize both SNR and $C_{\max}$. Assuming we have a specification SNR of $\geq$150 dB, we would initially pick the peak in Fig. 11. This would turn off over half the array. Instead the higher count rate intersection with specification should be chosen, with the specification derived from the required error and symbol rates.

D. After-Pulsing Probability

After-pulses are correlated to detections by the time dependent release of trapped carriers [18]. This is reduced using long τ_d, ensuring release without avalanche. We require a short τ_d meaning after-pulsing constitutes intersymbol interference (ISI). The probability, $P(ap)$, of $(n = 8)$ subpixels was measured at $\tau_d = 13.5$ ns and varied with $\mu = 0.92\%$, $\sigma = 0.73\%$, a minimum of 0.1%, and 2% maximum. This was measured using an oscilloscope [39] to measure the time between a (low probability) primary dark count and a secondary event within a 1 μs window. Counting primary and secondary pulses yields a good estimate for $P(ap)$ given suitable numbers of secondary pulses. As the frequency increases, ISI due to P(ap) will increase due to lower τ_d requirements and will affect an increased number of bits.

E. Receiver Photon Transfer Curve

The photon transfer curve (PTC) (Fig. 12) between optical power and single-photon counts, saturates at $\approx C_{\max}$, with the sensitivity and noise tradeoffs of N_a modes. This used a variable, broad spectrum CW lamp, a 450 nm (10 nm FWHM) band-pass filter, variable density filters providing $\approx$ 8 decades of optical power, and a calibrated meter. This relates to [1], [33] free-space optical communications, where a broad spectrum source (indoor lighting), or ambient light are present. 410 k samples per point were taken. We observe a gradient of one throughout the central region, however PQ shows a sublinear relationship. SPAD paralysis [27] is observed, impacting high count rates. The PQPR circuit obtains a $C_{\max}$, $(1/e)$ of the AQAR circuit [27]. The absolute dynamic range (DR) of 79.1 dB can be improved by lower τ_d, small capacitance, and high speed of nanometer circuits [18], [20], although lower DCR would give a significant improvement. The DR has been measured to the absolute noise floor, $C_{\max}/\text{DCR}$, rather than variation of noise, as it represents an absolute minimum on a single-shot basis. Low noise devices could give absolute DR greater than 122 dB at 25 °C and 142 dB at 0 °C, based upon [19], 1.35 V excess bias, 1024 SPADs, and $C_{\max}$ of 58.13×10^9. Device 01 homogeneity under uniform illumination, (30 μs, $\approx$1.4 μW/cm², 250 frames), was on average 2.26% from $\mu(\text{signal})$ and only 1.2% of devices were outside $\pm 3\sigma$. This photo-response nonuniformity has no spatial component suggesting SPAD and circuit variation.

F. Receiver Power Consumption With Photon Flux

In Fig. 13, current consumptions are plotted with the clock at 0 V. AQ draws higher current because of its large recharge transistor used to prevent paralysis. Both AQ and PQ follow the logistic curve of the PTC. The central region is expected to follow a $I = fCV$ relationship, but both supplies show sublinear behavior in both quenching modes. This has been linked to local voltage drops, despite the grid power structure and distributed power pads. The efficiency is reduced by the high VBD voltage. Charge pumps can generate VBD [40], [41], however these create further PTC nonlinearity if unable to provide the

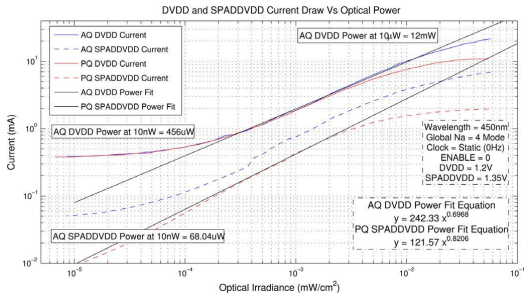


Fig. 13. Receiver SPAD and DVDD power supply current consumption with increasing optical input power.

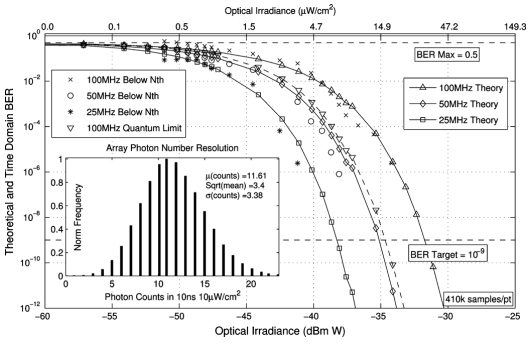


Fig. 14. BER analysis using the theoretical BER and a time domain search of samples below the N_{th} threshold. Analysis assumes a 100% extinction ratio, and uses 410 k samples per point. Target BER of 10^{-9} is reached at -31.7 dBm for 100 MHz. Photon number resolution is shown inset, $\mu = 11.61$, $\sigma = 3.38$. Array area $= 0.0212$ cm^2.

fast switching currents required. While nonlinearity can be compensated, the array draws 7.6 mA from $VBD = -13.05$ V at $C_{\max}$, suggesting significant area and power for an on-chip charge pump.

G. Estimate of Receiver BER

Fig. 14 presents first-order BER analysis over frequency and optical power, assuming a 100% extinction ratio. This used the same setup as the PTC, experimental values for N_1, N_0, σ_1, σ_0, and (2). The influence of extinction ratio, source bandwidth, and receiver transient behavior have been removed. Other experiments are presented for variable separation. With a μLED used as a modulated source, it was necessary to estimate the expected N_0 to N_1 distribution gap, without the influence of the experimental, similar bandwidth transmitter. The μLED slow turn off will artificially raise N_0, forcing the BER to deviate from the estimate and forcing it to be the combined link, rather than the receiver only BER. Each point of Fig. 14 estimates the BER to first order, using the count mean and standard deviation at a particular light level and clock frequency including the dark noise floor. The inset histogram, at 100 MHz and 10 μW/cm^2, shows 410 k samples of ≈ 11.6 photons per sample. Using the first and second derivatives, the distributions for each data point showed single-Gaussian behavior. Low photon points showed skew and discretization due to the binary counters. A multi-Gaussian fit, (including the action of receiver transient behavior), will allow more detailed BER determination than the first order estimate presented here. With increasing flux, any non-ideal distribution shapes will become increasingly negligible as the N_0 to N_1 separation increases. The count range of Fig. 15 suggests suitability for $\{4, 8, 16\}$-PAM. A time domain sweep shows samples below N_{th} and agrees with the theoretical first order result. The experimentally derived QL at 100 MHz, corrected for FF and PDE, is also presented. The gap could be due to source intensity noise, array noise, nonlinearity, and the assumptions of (2). The array has an estimated sensitivity of -31.7 dBm for a BER of 10^{-9} at 100 MHz, primarily due to the sensitivity of the SPADs. For a fill factor of 50% and a PDE of 45%, a similar array could achieve $QL = -52.5$ dBm. Photon number resolution is shown inset in Fig. 14, demonstrating that pure Poisson statistics can be approached. I/O frequency could be increased by low voltage and differential techniques and the overall throughput can be increased by thresholding to single-bit. The core power of 87.9 mW scales with photon flux, Fig. 13 and is linear with the system clock giving, 372 pJ/bit and a detection power efficiency of 3.16 pJ/photon. In comparison [6] achieves a normalized sensitivity of -18.8 dBm (30.7 pJ/bit).

H. Received NRZ-OOK Modulated Data Streams

Non-return-to-zero (NRZ) OOK modulated data is transmitted in free-space with zero ambient light, using a bare 450-nm μLED [2] ($dia = 34$ μm, $\mu(I) = 3$ mA) with ≈ 200 MHz bandwidth [42]. This relates to [1] where solid-state lighting is expected rather than laser sources [2], [3], [42]. A long pseudo-random bit stream (PRBS) was used ($\approx 2V_{p2p}$, $\mu(P) = 20$ μW/cm^2), with clear repeatability in the received data. Poisson, intensity noise, and hysteresis effects from $1 \rightarrow 1$ and $1 \rightarrow 0$ transitions are observed. N_0 and N_1 show wide distributions, although a clear gap is seen at 50 MHz (Fig. 15). A voltage eye diagram is unsuitable due to the 14-bit output, however the digital eye for 50 MHz is shown (Fig. 16). The open eye shows sporadic peaks at $0 \rightarrow 1$ transitions and an intermediate level due to incomplete $1 \rightarrow 0$ transitions. With transient μLED tails due to GaN carrier lifetime [2], an integration $\tau_b \leq 75\%$ of the clock period may be needed. At 75 MHz separation closes, indicating increased overall link BER. In the 100 MHz eye (Fig. 17), a large number of "0" bits are not fully transitioning. The eye remains partially open, however midlevel codes are now problematic for CDR. Derivatives of 10 MHz data under the same conditions as Fig. 16, showed an almost Gaussian N_1 distribution. A small, large variance secondary peak, attributed here to the $1 \rightarrow 0$ impulse response, is present in N_0, (multi-Gaussian fitted), occurring with $\approx 14.3\%$ probability. This is not negligible to the other N_0 distribution peaks. Multiple peaks within the 50 and 100 MHz data may be a product of increasing receiver transient issues from increased modulation frequency and average optical power, alongside the non-idealities of the μLED, clock synchronization and FIFO constraints. At 100 MHz a 7-Gaussian fit, shows a split N_1 distribution due to the average "1" optical power, (similar mean to the 10 MHz N_1 distribution), and a lower probability peak ≈ 50

Figure A.11

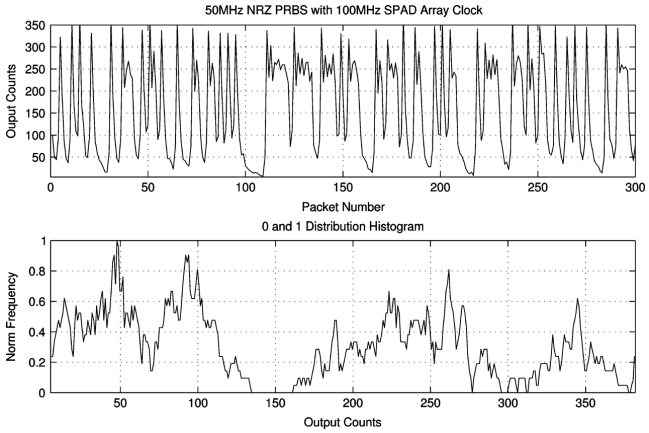


Fig. 15. 50-MHz NRZ $2^7 - 1$ PRBS, obtained using a 450 nm 34 μm micro-LED [2]. Measured with high DCR SPADs turned off and ≈ 20 μW/cm^2. Array hysteresis is observed from $1 \rightarrow 1$ and $1 \rightarrow 0$ causing distinct super-Poisson statistics. Total Packets $= 512$.

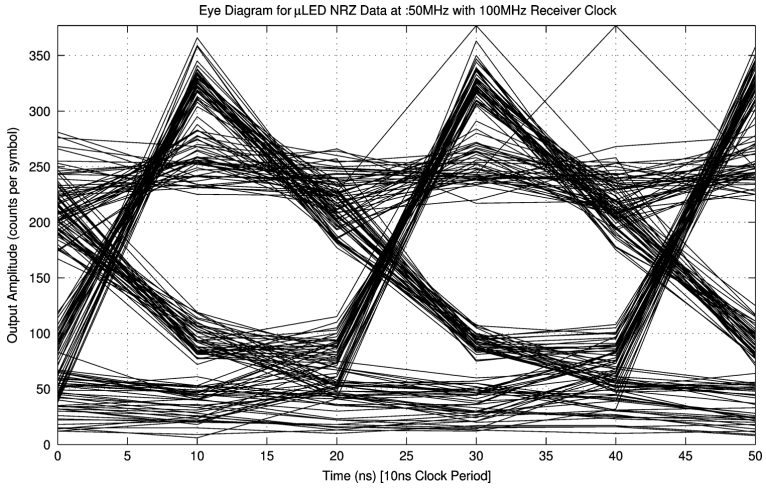


Fig. 16. Eye diagram for a 50 MHz NRZ data stream from a μLED, with a 100 MHz receiver clock. The eye is open but shows sporadic peaks at the $0 \rightarrow 1$ transitions and an intermediate level due to incomplete $1 \rightarrow 0$ transitions.

codes higher, due to the receiver step response (Fig. 18). N_0 is split with means ($\mu = 52, 73, 111$) corresponding to 1) full receiver transition down to the μLED minimum, 2) full μLED transition but partial receiver transition because of receiver impulse response, and 3) partial transitions for both μLED and receiver. With the peak positions and causes known, the μLED response can be removed, giving the BER at ≈ 20 μW/cm^2, as estimated by a fit to receiver only data, as 3.7×10^{-5}. This, however, is not under the same conditions as the first-order BER estimate, as the transmitter extinction ratio is increasing N_0 above the noise floor, elevating BER. We can scale N_0, using the 10 MHz lower limit of $\mu(N_0) = 9$ counts per τ_b. This brings the lower μLED power back towards 100% extinction ratio, removing the transmitter effect. Doing so gives a BER of 1.07×10^{-6}. This agrees with the BER of 1.5×10^{-6}, obtained from the 1st order estimate, at the ≈ 20 μW/cm$^2 = -33.6$ dBm optical power that the μLED experiments were conducted. We conclude, that when the optical power is at ≈ -31.7 dBm and transmitter behavior has been removed, the BER will be close to 10^{-9}.

I. Receiver Step Response Under Varying Optical Power

The step response, measured using a fast switching CW (840 nm) laser, is a proxy for $0 \rightarrow 1 \rightarrow 1$ transitions (packets 105 to 130 in Fig. 15). Laser triggering used a FIFO 1/4 full flag, a 100 MHz clock and global AQAR and $N_a = 4$ modes. Multiple experiments were performed and averaged with only valid runs showing no timing problems considered. Each is

Figure A.12

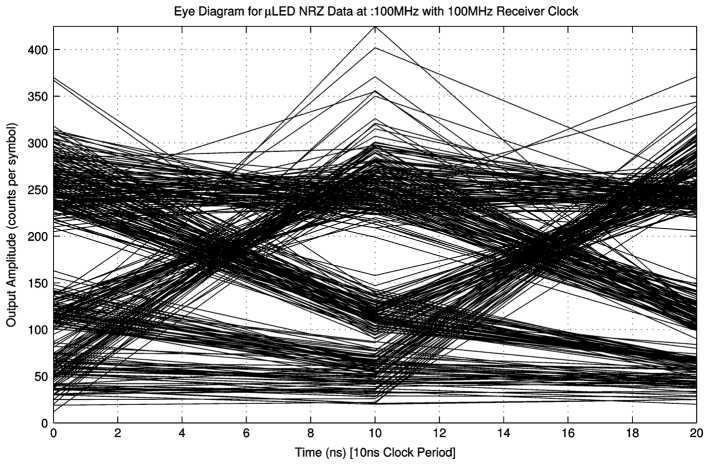


Fig. 17. Eye diagram for a 100 MHz NRZ data stream from a μLED, with a 100 MHz receiver clock. The eye remains partially open, however mid level codes are now problematic for correct thresholding.

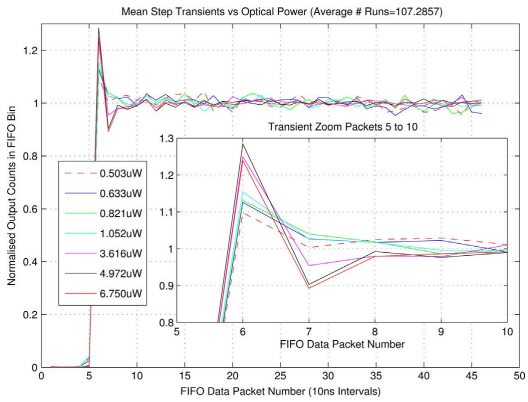


Fig. 18. Normalized average step response optical power sweep.

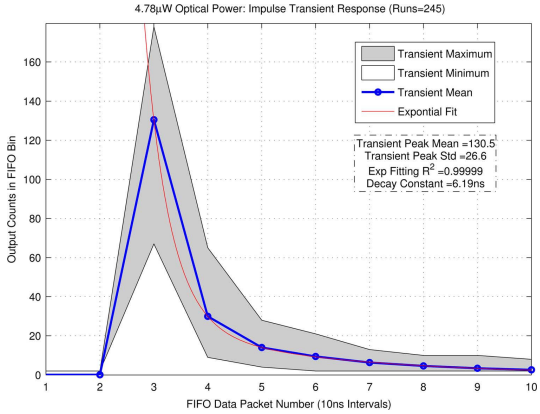


Fig. 19. Mean impulse response with a low average optical power of 4.78 μW.

normalized to its long term steady state for different average optical powers from variable neutral density filters. An initial transient is exacerbated at high powers, increasing settling time, constituting further ISI. A detector pool model where SPADs are removed by events and returned to the available pool after a period (of length τ_d), correctly models this behavior. The equilibrium effective PDE will be closer to unity if the optical power is low, or if $\tau_d \ll \tau_b$. The response is likely to be worsened by finite current handling in the power supplies.

J. Receiver Impulse Response Under Varying Optical Power

The impulse response to a 53 ps laser pulse ($\lambda = 442$ nm) can be used to investigate non-ideal $1 \rightarrow 0 \rightarrow 0$ transitions. The setup is identical, other than the laser, to the step response. A double exponential (Fig. 19) was fitted to calculate the decay constant over different average optical powers. The constants

TABLE I
TRANSIENT RESPONSE DECAY CONSTANT WITH OPTICAL POWER

Optical Power (μW)	Optical Energy (J)	Photons per SPAD	Decay Constant (ns)
@ 100MHz Rep Rate	Per Pulse	Per Pulse	Time to e^{-1}
1.25	12.5f	$\approx$ 30	1.9
4.78	48.7f	$\approx$ 100	6.3
110	1.10p	$\approx$ 2400	7.0
205	2.05p	$\approx$ 4400	7.9
261	2.61p	$\approx$ 5700	8.7
530	5.30p	$\approx$ 11.5k	10.4
1093	10.93p	$\approx$ 23.7k	12.3

c and d were negligible compared to the a and b fitting constants, showing mono-exponential curves. The decay constant (Table I) increases with optical power and could be accounted

Figure A.13

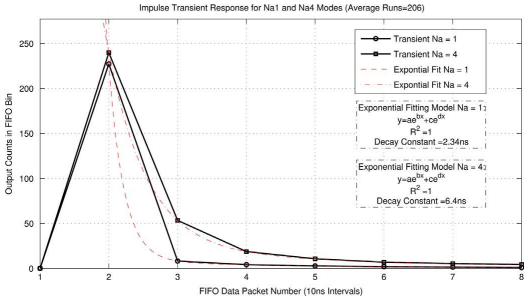


Fig. 20. $N_a = 4$ and $N_a = 1$ pixel modes with a reduction in the impulse response decay constant. Fitting coefficients suggest an approximately mono-exponential decay.

for by an elevation of the after-pulsing, over the dark measurement. This could be due to photon-induced electron-hole generation close to the SPAD. Despite the deep N-Well, [19] shows increased diffusion-tails [18] at 815 nm, suggesting photo-carriers diffuse over this region. This indicates a tradeoff between high optical intensity for a low BER and a low intensity for high bandwidth, preventing SPAD receivers achieving low BERs. In PD and APD receivers [6], [8], [10], slow diffusion of minority carriers has prompted buried oxide layers in SOI processes [43], guard rings to sink carriers [11], [31], [44], spatially modulated light (SML) PD structures [9], [10], and differential circuits [6], [8]. As SML structures are problematic for SPAD design, due to edge breakdown, SPAD circuits will suffer from similar finite physical limitations. Long decay constants are present despite the subpixel guard ring. Although a buried N-well [13] improved bandwidth significantly, here, the deep N-well cycles between 0 V and 1.35 V and is close to the optically active junction. Positive drive quenching [36], [41], with deep N-well SPADs, bias the substrate junction in high reverse bias, with a large depletion region, isolating SPADs [13] and could mitigate the impulse response.

K. Receiver Impulse Response Reduction With Pixel Mode

The decay constant can be reduced from 6.4 to 2.34 ns when using $N_a = 1$ (Fig. 20). We posit this is a product of forced SPAD hold-off. At high flux, numerous photon-induced carriers are present and could inflate the effective $P(ap)$. If SPADs are disabled, close minority carriers will recombine or diffuse prior to recharge. Hence they will not generate a carrier-induced count. This result sets N_a value globally, however this could be used dynamically, minimizing the BER using the tradeoff between transient behavior and sensitivity.

VI. SPAD Trends for Optical Communications

The QL can be approached by FF and PDE improvements. A deep junction SPAD, shows PDE of 44% at 690 nm and 20% at 850 nm, placing it within polymer and silica fiber attenuation windows [39]. Si-Ge SPADs [45] integrated with CMOS, pushed the PDE into IR (800–1150 nm) with reasonable DCR. A PDE of 50% with low DCR and diffusion tails [46] indicates

TABLE II
General Results and Chip Information

Parameter:	Value:	Units:	Notes:
CMOS Process	130nm Imaging		90nm Metallization
Max Count Rate	$58.13\text{x}10^9$	s^{-1}	@ $\tau_d = 13.85$ns
Sensitivity	≈ -31.7	dBm	450nm 10^{-9} BER
Dynamic Range	79.1	dB	Abs(DR) @ C_{max}
Peak PDE	20 [19]	%	$Vex = 1.2V$ 450nm
SPAD DVDD	1.35	V	Excess Bias
SPAD VBD	-13.05	V	Break Down
Min Dead Time	5.87 ± 2.7	ns	Passive Quench
Min Dead Time	10.61 ± 0.04	ns	Active Quench
Max Count Power	190	mW	Exc Pads, Inc VBD
Array Fill Factor	2.42	%	Exc Pads etc

TABLE III
Performance Comparison with Recent Prior Art

Ref	Technology	BER	Bit Rate	λ	Area	Power/Bit	S_{norm}	FOM
			GHz	nm	μm^2	W Hz^{-1}	dBm	
[6] (1)	130nm CMOS	$1\text{x}10^{-12}$	4.5	850	$1.8\text{x}10^6$	$3.0\text{x}10^{-11}$	-18.8	11.92
[6] (2)	130nm CMOS	$1\text{x}10^{-12}$	4.5	850	$1.8\text{x}10^6$	$1.7\text{x}10^{-11}$	-18.4	12.18
[8]	0.6μm CMOS	$3\text{x}10^{-10}$	0.5	850	$5.7\text{x}10^5$	$3.4\text{x}10^{-11}$	-12.5	11.68
[10]	180nm CMOS	$1\text{x}10^{-12}$	3.125	850	$7.0\text{x}10^5$	$4.6\text{x}10^{-11}$	-17.6	10.90
[9]	0.6μm CMOS	$1\text{x}10^{-9}$	0.25	860	$1.3\text{x}10^4$	$1.6\text{x}10^{-11}$	-19.1	10.68
[11]	180nm CMOS	$1\text{x}10^{-12}$	2.5	850	$5.3\text{x}10^5$	$4.3\text{x}10^{-11}$	-17.0	10.68
[48]	0.35μm CMOS	$1\text{x}10^{-9}$	0.6	680	$4.5\text{x}10^4$‡	$5.0\text{x}10^{-11}$	-16.0	10.37
[47]	0.6μm BiCMOS	$1\text{x}10^{-9}$	2.5	660	$7.7\text{x}10^5$	$6.8\text{x}10^{-11}$	-32.4	10.72
[49]	130nm CMOS	$1.5\text{x}10^{-12}$	2.4	850	$3.0\text{x}10^6$	$9.4\text{x}10^{-11}$	-15.2	11.55
This Work	130nm CMOS	$1\text{x}10^{-9}$	0.1	450	$2.1\text{x}10^6$	$8.0\text{x}10^{-10}$†	-31.7	12.20

Table III. † measured at maximum counting rate, expected to be lower for quoted BER. ‡ estimated from device micrograph.

larger area SPADs could be combined with topological techniques, such as noncircular geometries. The FF of other devices has also been increased by N-Well sharing [36], [41]. Quenching circuits have concentrated on high integration and $C_{\max}$ [20], [27], along with a reduction in SPAD τ_d [22]. Sine wave gating shows sampling at 2.23 GHz [24], albeit with lower sensitivity. System level techniques [24], [37], [38] are necessary while research and commercialization continues on integration, bandwidth, and yield. A 2 μm SPAD has achieved median DCR of 0.5 Hz and a single photon jitter of 103 ps [23]. There is a trend towards small area devices and SPADs in smaller geometry processes. Moving electronics to array edges and use of NMOS logic is a route towards higher FF.

VII. Comparison With PD and APD Receivers

In Table III, the SPAD receiver and PD/APD receivers are compared. Bandwidth, sensitivity, and noise analyses can optimize PD receivers. In contrast, this receiver investigates the roles of diode noise, transient behavior, and summation. The power, 8×10^{-10} WHz^{-1}, is measured at the absolute $C_{\max}$ for base τ_d and discounts power from VBD, which will inflate this consumption. The power, with large clock tree, nonoptimized circuit, and limited I/O, is open for significant optimization. The

sensitivity, S, has been normalized, S_{norm}, to the same BER ($\mathrm{BER}_{\mathrm{target}}$), wavelength ($\lambda_{\mathrm{target}}$), and bit rate ($\mathrm{BitRate}_{\mathrm{target}}$) for all devices

$$S_{\mathrm{norm}} = 10\log_{10}\left[\left(\frac{\mathrm{BER}_{\mathrm{target}}}{\mathrm{BER}}\right)\left(\frac{\mathrm{BitRate}_{\mathrm{target}}}{\mathrm{BitRate}}\right) \times \left(\frac{10^{\frac{S}{10}}}{1000}\right)\left(\frac{\lambda}{\lambda_{\mathrm{target}}}\right)\right] + 30. \quad (4)$$

The figure of merit presented (5), shown at the bottom of the page, is an adaptation of the FOM in [6]. We consider receivers with integrated PDs in standard CMOS processes. BiCMOS implementations tend to achieve higher frequency and sensitivity, shown by high S_{norm} [47]. The output voltage must also be considered, here the power per bit (W/Hz) is corrected assuming $P = fCV^2$. Photon-number resolution, which only this presented receiver achieves, is excluded from this FOM. While the FOM should include area, quotation carries errors, and may require assumptions. The receiver achieves poor area use due to a *design for test* structure and high power use due to a large clock tree, increased output amplitude, and measurement at C_{max}. SPAD receivers will draw increased power. However with increased sensitivity, the link power including the source power efficiency, can be reduced once optimized. As a further consideration, the proceeding system, taking in LVDS signals, will require power before processing by CMOS logic. As such, the FOM of low voltage output implementations will be decreased, as the presented receiver is suitable for the direct (no decoder) connection of signal processing blocks, such as FECC, signal processing or dynamic feedback. The nearest neighbors have comparable FOMs, as [6] have implemented a low power, differential, high-speed receiver with a small photodiode and a 950 mV output, but requiring a high optical input power. In comparison this work presents a preoptimized design with a total photodiode area $\approx$ 18.2 times larger, a normalized sensitivity (linear units) $\approx$ 21 times higher, a 1.2 V logic-compatible output voltage, and configuration/test circuits. Both [47] and [48], while operating in the visible, output low voltages, reducing their figures of merit.

VIII. Conclusion

ON–OFF key signals are received using single-photon avalanche diodes integrated into CMOS. The receiver (Table II), uses addition to increase sensitivity. The diodes give high sensitivity with their noise abated using system level techniques. It has been used to explore sensitivity, transient, and power factors of such receivers. High DCR devices can be weighted and the dynamic range improved with lower DCR or moderate cooling. Intersymbol interference is an issue due to after-pulsing and non-ideal transient responses. $P(ap)$ can be reduced, but prompts continued development. A pool of finite dead time detectors causes increasing ISI with optical power. Power consumption is linear with the clock and tracks the photon flux. RC equalization, spatially modulated photodiodes, and analogue differential techniques are not suitable, reducing impulse response mitigation options. To combat this, SPAD equalization such as subtraction of the impulse response, must be developed. Further work requires on chip CDR and increased fill factor. Error correction could also be implemented on chip. In contrast to other receivers, single-photon resolution and a high photon detection power efficiency are demonstrated. Demodulation of SPAD outputs into analogue, may allow schemes such as orthogonal frequency division multiplexing to be used. Optimized SPAD receivers may find a niche where optical power is extremely low, or if reconfigurability is required. Free-space optical communications with atmospheric turbulence, or space-borne transmitters are two such applications [37]. Similarly, consumer "last mile" receivers with POF, low optical-power or long fiber lengths could be developed if link power budgets are suitably low.

Acknowledgment

The authors would like to thank ST Microelectronics along with Dr. J. Richardson, Dr. B. Rae, Dr. R. Walker, and Dr. E. Webster for their insights. Further thanks go to Dr. J. McKendry and Dr. A. Kelly for LED modulation assistance. The work was funded by the Hypix project, EPSRC U.K., www.hypix.photonics.ac.uk.

References

[1] *P802.15.7—IEEE Approved Draft Standard for Short-Range Wireless Optical Communication Using Visible Light*, C/LM-LAN/MAN Standards Committee Std., Rev. P802.15.7, Apr. 2011 [Online]. Available: http://standards.ieee.org/develop/project/802.15.7.html

[2] J. McKendry *et al.*, "Visible-light communications using a CMOS-controlled micro-light-emitting-diode array," *J. Lightw. Technol.*, vol. 30, no. 1, pp. 61–67, 2012.

[3] J. Vucic *et al.*, "White light wireless transmission at 200+ Mb/s net data rate by use of discrete-multitone modulation," *IEEE Photon. Technol. Lett.*, vol. 21, no. 20, pp. 1511–1513, 2009.

[4] M. Atef and R. Swoboda *et al.*, "1.25 Gbit/s over 50 m step-index plastic optical fiber using a fully integrated optical receiver with an integrated equalizer," *J. Lightwave Technol.*, vol. 30, no. 1, pp. 118–122, Jan. 2012.

[5] C. Gunn, "CMOS photonics for high-speed interconnects," *IEEE Micro*, vol. 26, no. 2, pp. 58–66, 2006, IEEE Computer Society.

[6] F. Tavernier and M. Steyaert, "High-speed optical receivers with integrated photodiode in 130 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 44, no. 10, pp. 2856–2867, Oct. 2009.

[7] S. Loquai *et al.*, "10 Gbit/s over 25 m plastic optical fiber as a way for extremely low-cost optical interconnection," in *Optical Fiber Communication Conf.*, 2010, Optical Society of America.

[8] C. Hermans and M. Steyaert, "A high-speed 850 nm optical receiver front-end in 0.18 μm CMOS," *IEEE J. Solid-State Circuits*, vol. 41, no. 7, pp. 1606–1614, Jul. 2006.

[9] C. Rooman *et al.*, "Asynchronous 250-Mb/s optical receivers with integrated detector in standard CMOS technology for optocoupler applications," *IEEE J. Solid-State Circuits*, vol. 35, no. 7, pp. 953–958, Jul. 2000.

[10] W.-Z. Chen *et al.*, "A 3.125 Gbps CMOS fully integrated optical receiver with adaptive analog equalizer," in *Proc. IEEE Asian Solid-State Circuits Conf., ASSCC'07*, 2007, pp. 396–399.

[11] W.-Z. Chen and S.-H. Huang, "A 2.5 Gbps CMOS fully integrated optical receicer with lateral PIN detector." in *Proc. IEEE Custom Integrated Circuits Conf., CICC'07*, Sep. 2007, pp. 293–296.

$$\mathrm{FOM} = \frac{\mathrm{BitRate(bit/s)}S_{\mathrm{norm}}\mathrm{(dBm)}\log_{10}(\mathrm{BER}_{\mathrm{target}})V_{\mathrm{out}}^2(V^2)PD_{\mathrm{area}}(\mu m^2)}{\mathrm{Technology}^2(\mathrm{nm}^2)\mathrm{Power}(W)} \quad (5)$$

Figure A.15

[12] J. Campbell, "Recent advances in telecommunications avalanche photodiodes," *J. Lightw. Technol.*, vol. 25, no. 1, pp. 109–121, 2007.

[13] B. Ciftcioglu *et al.*, "Integrated silicon PIN photodiodes using deep N-well in a standard 0.18-μm CMOS technology," *J. Lightw. Technol.*, vol. 27, no. 15, pp. 3303–3313, Aug. 1, 2009.

[14] S. Csutak *et al.*, "CMOS-compatible high-speed planar silicon photodiodes fabricated on SOI substrates," *IEEE J. Quantum Electron.*, vol. 38, no. 2, pp. 193–196, Feb. 2002.

[15] S. Radovanovic *et al.*, "A 3-Gb/s optical detector in standard CMOS for 850-nm optical communication," *IEEE J. Solid-State Circuits*, vol. 40, no. 8, pp. 1706–1717, Aug. 2005.

[16] E. Fisher *et al.*, "A reconfigurable 14-bit 60 GPhoton/s single-photon receiver for visible light communications," in *Proc. ESSCIRC*, 2012.

[17] S. Cova *et al.*, "Avalanche photodiodes and quenching circuits for single-photon detection," *Appl. Opt.*, vol. 35, no. 12, pp. 1956–1976, 1996.

[18] M. Ghioni *et al.*, "Progress in silicon single-photon avalanche diodes," *IEEE J. Sel. Topics Quantum Electron.*, vol. 13, no. 4, pp. 852–862, 2007.

[19] J. Richardson *et al.*, "Low dark count single-photon avalanche diode structure compatible with standard nanometer scale CMOS technology," *IEEE Photon. Technol. Lett.*, vol. 21, no. 14, pp. 1020–1022, 2009.

[20] J. Richardson *et al.*, "Scaleable single-photon avalanche diode structures in nanometer CMOS technology," *IEEE Trans. Electron Devices*, vol. 58, no. 7, pp. 2028–2035, 2011.

[21] A. Spinelli and A. Lacaita, "Physics and numerical simulation of single-photon avalanche diodes," *IEEE Trans. Electron Devices*, vol. 44, no. 11, pp. 1931–1943, 1997.

[22] C. Niclass and M. Soga, "A miniature actively recharged single-photon detector free of afterpulsing effects with 6 ns dead time in a 0.18 μm CMOS technology," in *Proc. 2010 IEEE Int. Electron Devices Meeting (IEDM)*, 2010, pp. 14.3.1–14.3.4.

[23] E. Webster *et al.*, "Single-photon avalanche diodes in 90 nm CMOS imaging technology with sub-1Hz median dark count rate," in *Proc. Int. Image Sensor Workshop*, 2011, pp. 262–265.

[24] J. Zhang *et al.*, "2.23 GHz gating InGaAs/InP single-photon avalanche diode for quantum key distribution," in *Proc. SPIE 7681*, 2010, p. 76810Z.

[25] C. Niclass *et al.*, "Design and characterization of a CMOS 3-D image sensor based on single photon avalanche diodes," *IEEE J. Solid-State Circuits*, vol. 40, no. 9, pp. 1847–1854, Sep. 2005.

[26] D.-U. Li *et al.*, "FPGA implementation of a video-rate fluorescence lifetime imaging system with a 32 × 32 CMOS single-photon avalanche diode array," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, 2009, pp. 3082–3085.

[27] A. Eisele *et al.*, "185 MHz count rate, 139 dB dynamic range single-photon avalanche diode with active quenching circuit in 130 nm CMOS technology," in *Proc. 2011 Int. Image Sensor Workshop*, 2011, pp. 278–280.

[28] L. Braga *et al.*, "A CMOS mini-SiPM detector with in-pixel data compression for PET applications," in *Proc. Nuclear Sci. Symp. Med. Imaging Conf.*, 2011, pp. 548–552.

[29] C. Favi and E. Charbon, "Techniques for fully integrated intra-/inter-chip optical communication," in *Proc. 45th ACM/IEEE Design Automat. Conf. (DAC)*, Jun. 2008, pp. 343–344.

[30] M. Sergio and E. Charbon, "An intra-chip electro-optical channel based on CMOS single photon detectors," in *IEEE Int. Electron Devices Meet. (IEDM) Tech. Dig.*, 2005, Paper no. 822.

[31] F. Yuan, *CMOS Current-Mode Circuits for Data Communications*, ser. Analog Circuits and Signal Processing. New York, NY, USA: Springer, 2007.

[32] S. Sidiropoulos and M. Horowitz, "A 700-Mb/s/pin CMOS signaling interface using current integrating receivers," *IEEE J. Solid-State Circuits*, vol. 32, no. 5, pp. 681–690, May 1997.

[33] B. Leibowitz *et al.*, "A 256-element CMOS imaging receiver for free-space optical communication," in *Proc. IEEE Custom Integr. Circuits Conf. (CICC)*, 2004, pp. 303–306.

[34] J. Richardson *et al.*, "Dynamic quenching for single photon avalanche diode arrays," in *Proc. Int. Image Sensor Workshop*, 2007, pp. 258–260.

[35] I. Rech *et al.*, "Optical crosstalk in single photon avalanche diode arrays: A new complete model," *Opt. Exp.*, vol. 16, no. 12, pp. 8381–8394, Jun. 2008.

[36] L. Pancheri and D. Stoppa, "Low-noise CMOS single-photon avalanche diodes with 32 ns dead time," in *Proc. 37th Eur. Solid State Device Res. Conf. (ESSDERC)*, Sep. 2007, pp. 362–365.

[37] W. H. Farr *et al.*, "Gigahertz bandwidth photon counting," in *Proc. SPIE 7320*, 2009, pp. 732006–7320068.

[38] S. Castelletto *et al.*, "Reduced deadtime and higher rate photon-counting detection using a multiplexed detector array," *J. Modern Opt.*, vol. 54, no. 2/3, pp. 337–352, Jan/Feb. 2007.

[39] E. A. G. Webster *et al.*, "A single-photon avalanche diode in 90-nm CMOS imaging technology with 44% photon detection efficiency at 690 nm," *IEEE Electron Device Lett.*, vol. 33, no. 5, pp. 694–696, May 2012.

[40] M. Al-Rawhani *et al.*, "Photocurrent dependent response of a SPAD biased by a charge pump," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2011, pp. 789–792.

[41] R. Henderson *et al.*, "A 3 × 3, 5 μm pitch, 3-transistor single photon avalanche diode array with integrated 11 V bias generation in 90 nm CMOS technology," in *Proc. IEEE Int. Electron Devices Meet. (IEDM)*, 2010, pp. 14.2.1–14.2.4.

[42] J. McKendry *et al.*, "High-speed visible light communications using individual pixels in a micro light-emitting diode array," *IEEE Photon. Technol. Lett.*, vol. 22, no. 18, pp. 1346–1348, 2010.

[43] S. Csutak *et al.*, "High-speed monolithically integrated silicon photoreceivers fabricated in 130-nm CMOS technology," *J. Lightw. Technol.*, vol. 20, no. 9, pp. 1724–1729, Sep. 2002.

[44] L. Zhiqun *et al.*, "40-Gb/s, 0.18 um CMOS front-end amplifier for VSR parallel optical receiver," in *Proc. Symp. Photon. Optoelectron. (SOPO)*, 2009, pp. 1–4.

[45] A. Sammak *et al.*, "A CMOS compatible Ge-on-Si APD operating in proportional and Geiger modes at infrared wavelengths," in *Proc. IEEE Int. Electron Devices Meet.*, pp. 8.5.1–8.5.4.

[46] D. Bronzi *et al.*, "Low-noise and large-area CMOS SPADs with timing response free from slow tails," in *Proc. ESSDERC*, 2012, pp. 230–233.

[47] R. Swoboda and H. Zimmermann, "2.5 Gbit/s silicon receiver OEIC with large diameter photodiode," *Electron. Lett.*, vol. 40, no. 8, pp. 505–507, Apr. 2004.

[48] K. Kuwabara and Y. Matsumoto, "Development of CMOS integrated optical receiver for short-range data communication," in *Electron. Commun. Jpn.*, 2004, vol. 87, no. 12, pp. 10–18 [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1002/ecjb.20099/abstract

[49] F. Tavernier and M. Steyaert, "A high-speed fully integrated optical receiver in standard 130 nm CMOS," in *Proc. 15th IEEE Int. Conf. Electron., Circuits, Syst. (ICECS)*, 2008, pp. 806–809.

Edward Fisher (S'08) received the M.Eng. degree in electronic and electrical engineering from The University of Edinburgh, Edinburgh, U.K., in 2009. He is currently working toward the Ph.D. degree at the Institute of Micro- and Nanosystems, researching the use and factors involved with single-photon avalanche diodes for optical communications.

He was funded by the EPSRC "Hypix" project which aimed to use Gallium Nitride μLED structures, smart CMOS drive, control and reception electronics and lasing organic polymer materials for optical communications. His current research interests include device modelling for system level design of SPAD receivers and characterization of second order SPAD array effects.

Ian Underwood received the Ph.D. degree in applied physics from the University of Edinburgh, Edinburgh, U.K., in 1987.

He was appointed Lecturer in Electrical Engineering at the University of Edinburgh in 1989. He is a pioneer of liquid crystal on silicon (LCoS) technology, publishing his first paper on that topic in 1985. In 1991, he was a visiting Fulbright Fellow at The University of Colorado, Boulder, CO, USA, where he codesigned an award winning LCoS spatial light modulator in collaboration with Boulder Nonlinear Systems. He became active in the research and development of LCoS microdisplay devices and systems and was active in the formation of Micropix Technologies. From 1998 until 2008 he pioneered polymer OLED microdisplays as a cofounder of Microemissive Displays. He was promoted in absentia by the University of Edinburgh to Professor of Electronic Displays in 2004. In 2008 he returned to his alma mater where he leads the Institute for Integrated Micro and Nano Systems.

Dr. Underwood was appointed a Fellow of the Royal Society of Edinburgh in 2004 then a Fellow of both the Royal Academy of Engineering and the Institute of Physics in 2008. From 2008 to 2013, he was a member of the Scottish Science Advisory Council. He was the recipient of awards, such as Entrepreneur of the Year and Scotland's top Innovator.

Figure A.16

Bibliography

[1] I. Glover and P. Grant, *Digital Communications, 3rd Edition*, 3rd ed. Pearson, 2010.

[2] B. Razavi, *Design of Integrated Circuits for Optical Communications*, 1st ed. McGraw-Hill, 2003.

[3] L. N. Binh, *Digital Optical Communications*, 1st ed. CRC Press, 2008.

[4] J. Campbell, "Recent Advances in Telecommunications Avalanche Photodiodes," *Lightwave Technology, Journal of*, vol. 25, no. 1, pp. 109 –121, 2007.

[5] D. Miller, "Rationale and Challenges for Optical Interconnects to Electronic Chips," *Proceedings of the IEEE*, vol. 88, no. 6, pp. 728 –749, june 2000.

[6] I. Labs, "The 50G Silicon Photonics Link," Intel Labs," White Paper, July 2010.

[7] C. Favi, "Single-Photon Techniqies for Standard CMOS Digital ICs," PhD, Ecole Polytechnique Federale De Lausanne, 2011.

[8] R. Soref, *Introduction: The Opto-Electronic Integrated Circuit, in Silicon Photonics: The State of the Art*, G. T. Reed, Ed. John Wiley & Sons, Ltd, June 2008.

[9] M. Zuffada, "The Industrialization of the Silicon Photonics: Technology Road Map and Applications," in *ESSCIRC, Proceedings of*, 2012.

[10] D. Miller and H. Ozaktas, "Limit to the Bit-Rate Capacity of Electrical Interconnects from the Aspect Ratio of the System Architecture," *Journal of Parallel and Distributed Computing*, vol. 41, no. 1, pp. 42 – 52, 1997.

[11] C. S. Inc, "The Zettabyte Era - Trends and Analysis," Cisco Systems Inc," White Paper, May 2013. [Online]. Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/VNI_Hyperconnectivity_WP.html

[12] Cisco Systems Inc, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013 to 2018," Cisco Systems Inc, White Paper FLGD 11446, Feb 2014. [Online]. Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.html

[13] F. Aznar, S. Celma, and B. Calvo, *CMOS Receiver Front-ends for Gigabit Short-Range Optical Communications*, ser. Analog Circuits and Signal Processing, M. Ismail and M. sawan, Eds. Springer [New York] Science and Business Media, 2013.

[14] E. Sackinger, *Broadband Circuits for Optical Fiber Communication*. Wiley, 2005.

[15] *The United Kingdom Frequency Allocation Table*, OFCOM UK Std. 17, Feb 2013. [Online]. Available: http://stakeholders.ofcom.org.uk/spectrum/information/uk-fat/

[16] L. Hanzo, H. Haas, S. Imre, D. O'Brien, M. Rupp, and L. Gyongyosi, "Wireless Myths, Realities, and Futures: From 3G/4G to Optical and Quantum Wireless," *Proceedings of the IEEE*, vol. 100, no. Special Centennial Issue, pp. 1853 –1888, 13 2012.

[17] H. Elgala, R. Mesleh, and H. Haas, "Indoor broadcasting via white LEDs and OFDM," *Consumer Electronics, IEEE Transactions on*, vol. 55, no. 3, pp. 1127 –1134, 2009.

[18] D. O'Brien, H. Le Minh, L. Zeng, G. Faulkner, K. Lee, D. Jung, Y. Oh, and E. T. Won, "Indoor Visible Light Communications: Challenges and Prospects," in *Free-Space Laser Communications VIII: Proceedings of the SPIE*, A. K. Majumdar and C. C. Davis, Eds., vol. 7091, 2008, pp. 709 106–709 106–9.

[19] I. Alsolami and D. O'Brien, "Multiresolution PPM for Broadcasting over Asymmetric Photon Counting Channels," *Communications Letters, IEEE*, vol. 15, no. 12, pp. 1268–1270, 2011.

[20] L. Zeng, D. O'Brien, H. Minh, G. Faulkner, K. Lee, D. Jung, Y. Oh, and E. T. Won, "High Data Rate Multiple Input Multiple Output (MIMO) Optical Wireless Communications

using White LED Lighting," *Selected Areas in Communications, IEEE Journal on*, vol. 27, no. 9, pp. 1654 –1662, december 2009.

[21] A. Belmonte and J. Kahn, "Sequential Optimization of Adaptive Arrays in Coherent Laser Communications," *Lightwave Technology, Journal of*, vol. 31, no. 9, pp. 1383–1387, 2013.

[22] W. W. P. A. N. W. W. Group, *P802.15.7 - IEEE Approved Draft Standard for Short-Range Wireless Optical Communication Using Visible Light*, C/LM - LAN/MAN Standards Committee Std., Rev. P802.15.7 - Apr 2011, Apr 2011.

[23] J. McKendry, D. Massoubre, S. Zhang, B. Rae, R. Green, E. Gu, R. Henderson, A. Kelly, and M. Dawson, "Visible-Light Communications Using a CMOS-Controlled Micro-Light- Emitting-Diode Array," *Lightwave Technology, Journal of*, vol. 30, no. 1, pp. 61 –67, 2012.

[24] J. McKendry, R. Green, A. Kelly, Z. Gong, B. Guilhabert, D. Massoubre, E. Gu, and M. Dawson, "High-Speed Visible Light Communications Using Individual Pixels in a Micro Light-Emitting Diode Array," *Photonics Technology Letters, IEEE*, vol. 22, no. 18, pp. 1346 –1348, sept.15, 2010.

[25] K. Wang, A. Nirmalathas, C. Lim, E. Skafidas, and K. Alameh, "High-Speed Reconfigurable Free-Space Card-to-Card Optical Interconnects," *Photonics Journal, IEEE*, vol. 4, no. 5, pp. 1407–1419, 2012.

[26] S. M. Sze, *Semiconductor Devices: Physics and Technology, 2nd Edition*, 2nd ed. John Wiley and Sons, Inc, 2001.

[27] S. Csutak, J. Schaub, W. Wu, R. Shimer, and J. Campbell, "CMOS-Compatible High-Speed Planar Silicon Photodiodes Fabricated on SOI Substrates," *Quantum Electronics, IEEE Journal of*, vol. 38, no. 2, pp. 193 –196, feb 2002.

[28] ——, "High-Speed Monolithically Integrated Silicon Photoreceivers Fabricated in 130-nm CMOS Technology," *Lightwave Technology, Journal of*, vol. 20, no. 9, pp. 1724 –1729, sep 2002.

[29] D. H. N. Weste, *CMOS VLSI Design: A Circuits and Systems Perspective*, 3rd ed. Pearson, 2005.

[30] M. Cohen, F. Roy, D. Herault, Y. Cazaux, A. Gandolfi, J. Reynard, C. Cowache, E. Bruno, T. Girault, J. Vaillant, F. Barbier, Y. Sanchez, N. Hotellier, O. LeBorgne, C. Augier, A. Inard, T. Jagueneau, C. Zinck, J. Michailos, and E. Mazaleyrat, "Fully Optimized Cu Based Process with Dedicated Cavity Etch for 1.75um and 1.45um Pixel Pitch CMOS Image Sensors," in *Electron Devices Meeting, 2006. IEDM '06. International*, dec. 2006, pp. 1 –4.

[31] F. Tavernier and M. Steyaert, "High-Speed Optical Receivers With Integrated Photodiode in 130 nm CMOS," *Solid-State Circuits, IEEE Journal of*, vol. 44, no. 10, pp. 2856 –2867, 2009.

[32] A. Carusone, H. Yasotharan, and T. Kao, "Progress and Trends in Multi-Gbps Optical Receivers with CMOS Integrated Photodetectors," in *Custom Integrated Circuits Conference (CICC), IEEE*, 2010, pp. 1–8.

[33] S. Radovanovic, A.-J. Annema, and B. Nauta, "A 3-Gb/s Optical Detector in Standard CMOS for 850-nm Optical Communication," *Solid-State Circuits, IEEE Journal of*, vol. 40, no. 8, pp. 1706 – 1717, aug. 2005.

[34] A. Sammak, M. Aminian, L. Qi, W. de Boer, E. Charbon, and L. Nanver, "A CMOS Compatible Ge-on-Si APD Operating in Proportional and Geiger Modes at Infrared Wavelengths," in *Electron Devices Meeting (IEDM)*, 2011, pp. 8.5.1 –8.5.4.

[35] F. Yuan, *CMOS Current-Mode Circuits for Data Communications*, ser. Analog Circuits and Signal Processing Series. Springer, 2007.

[36] M. Cvijetic, *Optical Transmission Systems Engineering*. Artech House, 2003.

[37] W.-Z. Chen, Y.-L. Cheng, and D.-S. Lin, "A 1.8-V 10-Gb/s Fully Integrated CMOS Optical Receiver Analog Front-end," *Solid-State Circuits, IEEE Journal of*, vol. 40, no. 6, pp. 1388 – 1396, june 2005.

[38] W.-Z. Chen and S.-H. Huang, "A 2.5 Gbps CMOS Fully Integrated Optical Receicer with Lateral PIN Detector," in *Custom Integrated Circuits Conference, CICC. IEEE*, 2007, pp. 293 –296.

[39] F. Tavernier and M. Steyaert, "Power Efficient 4.5Gbit/s Optical Receiver in 130nm CMOS with Integrated Photodiode," in *Solid-State Circuits Conference, 2008. ESSCIRC 2008. 34th European*, sept. 2008, pp. 162 –165.

[40] W.-Z. Chen, S.-H. Huang, G.-W. Wu, C.-C. Liu, Y.-T. Huang, C.-F. Chin, W.-H. Chang, and Y.-Z. Juang, "A 3.125 Gbps CMOS Fully Integrated Optical Receiver with Adaptive Analog Equalizer," in *Asian Solid-State Circuits Conference, ASSCC*, 2007, pp. 396 – 399.

[41] V. Vilnrotter and M. Srinivasan, "Optical Communications Through Atmospheric Turbulence Using Photodetector Arrays," in *Photonics West Proc SPIE ol 4272 Free-Space Laser Communication Technologies XIII*, 2000, pp. 282–292.

[42] B. Yang, J. Schaub, S. Csutak, and J. Campbell, "10 Gbps All Silicon APD Optical Receiver," in *Lasers and Electro-Optics Society, 2002. LEOS 2002. The 15th Annual Meeting of the IEEE*, vol. 2, nov. 2002, pp. 681 – 682 vol.2.

[43] M. Ferraro, W. Freeman, R. Mahon, J. Murphy, P. Goetz, M. Colbert, W. Rabinovich, H. Burris, C. Moore, W. Clark, and W. Waters, "Large Area Adaptive Avalanche Photodetector Arrays For Free-Space Optical Communications," in *Proc SPIE vol 7091: Free-Space Laser Communications VIII*, 2008.

[44] C. Rooman, D. Coppee, and M. Kuijk, "Asynchronous 250-Mb/s Optical Receivers with Integrated Detector in Sandard CMOS Technology for Optocoupler Applications," *Solid-State Circuits, IEEE Journal of*, vol. 35, no. 7, pp. 953 –958, july 2000.

[45] H. Zimmermann, B. Muller, A. Hammer, K. Herzog, and P. Seegebrecht, "Large-area Lateral P-I-N Photodiode on SOI," *Electron Devices, IEEE Transactions on*, vol. 49, no. 2, pp. 334 –336, feb 2002.

[46] M. Georgas, J. Orcutt, R. Ram, and V. Stojanovic, "A Monolithically-Integrated Optical Receiver in Standard 45-nm SOI," *Solid-State Circuits, IEEE Journal of*, vol. 47, no. 7, pp. 1693 –1702, july 2012.

[47] D. George, R. Bowen, and J. Storey, "An Adaptive Decision Feedback Equalizer," *Communication Technology, IEEE Transactions on*, vol. 19, no. 3, pp. 281–293, 1971.

[48] D. Lee, J. Han, E. Chang, G. Han, and S. M. Park, "An 8.5Gb/s CMOS OEIC with On-Chip Photodiode for Short-Distance Optical Communications," in *Solid-State Circuits Conference Digest (ISSCC)*, 2010, pp. 362 –363.

[49] S. Cova, M. Ghioni, A. Lacaita, C. Samori, and F. Zappa, "Avalanche Photodiodes and Quenching Circuits for Single-Photon Detection," *Appl. Opt.*, vol. 35, no. 12, pp. 1956–1976, 1996.

[50] M. Ghioni, A. Gulinatti, I. Rech, F. Zappa, and S. Cova, "Progress in Silicon Single-Photon Avalanche Diodes," *Selected Topics in Quantum Electronics, IEEE Journal of*, vol. 13, no. 4, pp. 852 –862, 2007.

[51] C. Niclass, M. Sergio, and E. Charbon, "A Single Photon Avalanche Diode Array

Fabricated in Deep-Submicron CMOS Technology," in *Design, Automation and Test in Europe, 2006. DATE '06. Proceedings*, vol. 1, 2006, pp. 1–6.

[52] A. Rochas, M. Gani, B. Furrer, P. A. Besse, R. S. Popovic, G. Ribordy, and N. Gisin, "Single Photon Detector Fabricated in a Complementary Metal-Oxide-Semiconductor High-Voltage Technology," *Review of Scientific Instruments*, vol. 74, no. 7, pp. 3263 –3270, 2003.

[53] J. Richardson, E. Webster, L. Grant, and R. Henderson, "Scaleable Single-Photon Avalanche Diode Structures in Nanometer CMOS Technology," *Electron Devices, IEEE Transactions on*, vol. 58, no. 7, pp. 2028 –2035, 2011.

[54] C. Niclass, A. Rochas, P. Besse, R. Popovic, and E. Charbon, "CMOS Imager Based on Single Photon Avalanche Diodes," in *Solid-State Sensors, Actuators and Microsystems, The 13th International Conference on TRANSDUCERS*, vol. 1, 2005, pp. 1030 – 1034 Vol. 1.

[55] E. Webster, J. Richardson, L. Grant, and R. Henderson, "Single-Photon Avalanche Diodes in 90nm CMOS Imaging technology with sub-1Hz Median Dark Count Rate," in *International Image Sensor Workshop*, 2011, pp. 262–265.

[56] E. Webster, R. Walker, R. Henderson, and L. Grant, "A Silicon Photomultiplier with >30% Detection Efficiency from 450-750nm and 11.6um Pitch NMOS-Only Pixel with 21.6% Fill Factor in 130nm CMOS," in *ESSCIRC, Proceedings of*, 2012.

[57] R. Walker, E. Webster, and R. Henderson, "High Fill Factor Digital Silicon Photomultiplier Structures in 130nm CMOS Imaging Technology," in *IEEE Nuclear Science Symposium and Medical Imaging Conference*, 2012.

[58] E. Vilella, O. Alonso, A. Montiel, A. Vila, and A. Dieguez, "A Low-Noise Time-Gated Single-Photon Detector in a HV-CMOS Technology for Triggered Imaging," *Sensors and Actuators A: Physical*, vol. 201, no. 1, pp. 342 – 351, 2013.

[59] D. Bronzi, F. Villa, S. Bellisai, S. Tisa, G. Ripamonti, and A. Tosi, "Figures of Merit for CMOS SPADs and Arrays," *Proc. SPIE 8773, Photon Counting Applications IV; and Quantum Optics and Quantum Information Transfer and Processing*, vol. 8773, pp. 877 304–877 304–7, May 2013.

[60] L. Neri, S. Tudisco, F. Musumeci, A. Scordino, G. Fallica, M. Mazzillo, and M. Zimbone, "Generalization of DT Equations for Time Dependent Sources," *Sensors*, vol. 10, no. 12, pp. 10 828–10 836, 2010.

[61] M. Sergio and E. Charbon, "An Intra-Chip Electro-Optical Channel Based on CMOS Single Photon Detectors," in *Electron Devices Meeting, IEDM, IEEE International*, 2005, pp. 4 pp. –822.

[62] J. Zhang, P. Eraerds, N. Walenta, C. Barreiro, R. Thew, and H. Zbinden, "2.23 GHz Gating InGaAs/InP Single-Photon Avalanche Diode for Quantum Key Distribution," *PROC.SPIE*, vol. 7681, p. 76810Z, 2010.

[63] S. Takahashi, A. Tajima, and A. Tomita, "High-Efficiency Single Photon Detector Combined with an Ultra-Small APD Module and a Self-Training Discriminator for High-Speed Quantum Cryptosystems," in *In Proc. 13th Microoptics Conference, Japan 2007*, 2007.

[64] C. Favi and E. Charbon, "Techniques for Fully Integrated Intra-/Inter-Chip Optical Communication," in *Design Automation Conference, DAC. ACM/IEEE*, 2008, pp. 343 –344.

[65] Engineering and P. S. R. Council. (2012, Oct) (HYPIX) Hybrid Organic Semiconductor/Gallium Nitride/CMOS Smart Pixel Arrays: Details of Grant.

Website. Engineering and Physical Sciences Research Council. EP/F060076/1 Grant for Edinburgh but also EP/F059922/1, EP/F05999X/1 and EP/F061609/1. [Online]. Available: http://gow.epsrc.ac.uk/NGBOViewGrant.aspx?GrantRef=EP/F060076/1

[66] M. Dawson. (2012, Oct) (HYPIX) Hybrid Organic Semiconductor / Gallium Ntride / CMOS Smart Pixel Arrays. Web. [Online]. Available: http://hypix.photonics.ac.uk/

[67] G. Zhang, C. Yu, C. Zhu, and L. Liu, "Feasibility Study of Multi-Pixel Photon Counter serving as the Detector in Digital Optical Communications," *Optik - International Journal for Light and Electron Optics*, vol. 124, no. 22, pp. 5781 – 5786, 2013.

[68] E. Fisher, I. Underwood, and R. Henderson, "A Reconfigurable Single-Photon-Counting Integrating Receiver for Optical Communications," *Solid-State Circuits, IEEE Journal of*, vol. 48, no. 7, pp. 1638–1650, 2013.

[69] H. Kaeslin, *Digital Integrated Circuit Design: From VLSI Architectures to CMOS Fabrication*. Cambridge University Press, 2008.

[70] R. Walker, J. Richardson, and R. Henderson, "A 128x96 Pixel Event-Driven Phase-Domain ΔΣ-Based Fully Digital 3D Camera in 0.13μm CMOS Imaging Technology," in *Solid-State Circuits Conference (ISSCC), IEEE International*, 2011, pp. 410 –412.

[71] T. Frach, G. Prescher, C. Degenhardt, and B. Zwaans, "The Digital Silicon Photomultiplier; System Architecture and Performance Evaluation," in *Nuclear Science Symposium Conference Record (NSS/MIC)*, 30 2010-nov. 6 2010, pp. 1722 –1727.

[72] E. Fisher, I. Underwood, and R. Henderson, "A Reconfigurable 14-bit 60GPhoton/s Single-Photon Receiver for Visible Light Communications," in *ESSCIRC, 2012. ESSCIRC '12. Proceedings of the*, 2012, pp. 85–88.

[73] A. E. Stevens, "An Integrate-And-Dump Receiver for Fibre Optic Networks," Ph.D. dissertation, Columbia University School of Arts and Sciences, 1995.

[74] A. Goetzberger, B. McDonald, R. Haitz, and R. M. Scarlett, "Avalanche Effects in Silicon p-n Junctions. II. Structurally Perfect Junctions," *Journal of Applied Physics*, vol. 34, no. 6, pp. 1591–1600, 1963.

[75] F. Zappa, S. Tisa, A. Tosi, and S. Cova, "Principles and Features of Single-Photon Avalanche Diode Arrays," *Sensors and Actuators A: Physical*, vol. 140, no. 1, pp. 103 – 112, 2007.

[76] E. A. G. Webster, L. A. Grant, and R. K. Henderson, "Transient Single-Photon Avalanche Diode Operation, Minority Carrier Effects, and Bipolar Latch Up," *Electron Devices, IEEE Transactions on*, vol. 60, no. 3, pp. 1188 –1194, march 2013.

[77] P. Eraerds, M. Legré, A. Rochas, H. Zbinden, and N. Gisin, "SiPM for fast Photon-Counting and Multiphoton Detection," *Opt. Express*, vol. 15, no. 22, pp. 14 539–14 549, Oct 2007.

[78] H. Hinton, T. Cloonan, F. McCormick, A. Lentine, and F. Tooley, "Free-Space Digital Optical Systems," *Proceedings of the IEEE*, vol. 82, no. 11, pp. 1632 –1649, 1994.

[79] B. Leibowitz, B. Boser, and K. Pister, "A 256-Element CMOS Imaging Receiver for Free-Space Optical Communication," in *Custom Integrated Circuits Conference. Proceedings of the IEEE*, 2004, pp. 303 – 306.

[80] R. Loudon, *The Quantum Theory of Light*, 3rd ed. Oxford Science Publications, 2000.

[81] J. Janesick, *Photon Transfer: DN to λ*. SPIE (The Society of Photo-Optical Instrumentation Engineers), 2007.

[82] M. M. C. I. Consortium, "Communications Technology Roadmap: Microphotonics: Hardware for the Information Age," Massachusetts Institute of Technology," White Paper: Executive Overview, 2005.

[83] S. Sidiropoulos and M. Horowitz, "A 700-Mb/s/pin CMOS Signaling Interface using Current Integrating Receivers," *Solid-State Circuits, IEEE Journal of*, vol. 32, no. 5, pp. 681 –690, may 1997.

[84] C. Gunn, "CMOS Photonics for High-Speed Interconnects," *Micro, IEEE*, vol. 26, no. 2, pp. 58 –66, 2006, IEEE Computer Society.

[85] H. Johnson and M. graham, *High-Speed Digital Design: A Handbook of Black Magic*. Prentice Hall, 1993.

[86] J. Gambino, F. Chen, and J. He, "Copper Interconnect Technology for the 32 nm Node and Beyond," in *Custom Integrated Circuits Conference, CICC*, 2009, pp. 141 –148.

[87] K. Kao and G. Hockham, "Dielectric-Fibre Surface Waveguides for Optical Frequencies," *Electrical Engineers, Proceedings of the Institution of*, vol. 113, no. 7, pp. 1151 –1158, july 1966.

[88] R. S. I.-M. enterprise. (2014, Feb) Communication Equipment: Free Space Optics. Online. http://www.grpz.ru/en/products/Communication/models/. [Online]. Available: http://www.grpz.ru/en/products/Communication/models/

[89] M. Vorontsov, G. Carhart, J. Gowens, and J. Ricklin, "Adaptive Correction of Wave-Front Phase Distortions in a Free-Space Laser Communication System and Method," U.S. Patent 7 197 248, 2007. [Online]. Available: http://www.freepatentsonline.com/7197248.html

[90] Fast Links: Wireless Network Solutions. (2014, Feb) Free Space Optics Wireless Wide Area Networks: Free Space Optical & infrared Laser Systems. Online. Fast Links: Wireless Network Solutions. http://www.fastlinks-wireless.com/wireless-network/free-space-optics-networks/. [Online]. Available: http://www.fastlinks-wireless.com/wireless-network/free-space-optics-networks/

[91] P. Dourish and G. Bell, *Divining a Digital Future: Mess and Mythology in Ubiquitous Computing*. MIT Press, May 2011.

[92] K. Birnbaum, W. Farr, J. Gin, B. Moision, K. Quirk, and M. Wright, "Demonstration of a High-Efficiency Free-Space Optical Communications Link," in *Proc. SPIE 7199, Free-Space Laser Communication Technologies XXI, 71990A*, vol. 7199, 2009, pp. 71 990A–71 990A–12. [Online]. Available: http://dx.doi.org/10.1117/12.809720

[93] D. Massoubre, J. McKendry, B. Guilhabert, Z. Gong, I. Watson, E. Gu, and M. Dawson, "Fabrication of Planar GaN-Based Micro-Pixel Light Emitting Diode Arrays," in *LEOS Annual Meeting Conference Proceedings, 2009. LEOS '09. IEEE*, oct. 2009, pp. 84 –85.

[94] J. Vucic, C. Kottke, S. Nerreter, A. Buttner, K.-D. Langer, and J. W. Walewski, "White Light Wireless Transmission at 200 + Mb/s Net Data Rate by Use of Discrete-Multitone Modulation," *Photonics Technology Letters*, vol. 21, no. 20, pp. 1511 –1513, 2009.

[95] T. Q. Wang, Y. A. Sekercioglu, and J. Armstrong, "Analysis of An Optical Wireless Receiver Using a Hemispherical Lens With Application in MIMO Visible Light Communications," *J. Lightwave Technol.*, vol. 31, no. 11, pp. 1744–1754, Jun 2013.

[96] B. Rae, "Micro-Systems for Time-Resolved Fluorescence Analysis using CMOS Single-Photon Avalanche Diodes and Micro-LEDs," PhD Thesis, The School of Engineering, The University of Edinburgh, 2009.

[97] D. Elfstrom, B. Guilhabert, J. McKendry, S. Poland, Z. Gong, D. Massoubre, E. Richardson, B. R. Rae, G. Valentine, G. Blanco-Gomez, E. Gu, J. M. Cooper, R. K. Henderson, and M. D. Dawson, "Mask-less Ultraviolet Photolithography Based on CMOS-Driven Micro-Pixel Light Emitting Diodes," *Opt. Express*, vol. 17, no. 26, pp. 23 522–23 529, Dec 2009.

[98] Z. Gong, S. Jin, Y. Chen, J. McKendry, D. Massoubre, I. M. Watson, E. Gu, and M. D. Dawson, "Size-Dependent Light Output, Spectral Shift, and Self-Heating of 400 nm InGaN Light-Emitting Diodes," *Journal of Applied Physics*, vol. 107, no. 1, p. 013103, 2010.

[99] K. Kuwabara and Y. Matsumoto, "Development of CMOS Integrated Optical Receiver for Short-Range Data Communication," in *Electronics and Communications in Japan*, vol. 87, no. 12, 2004, pp. 10–18.

[100] M. Atef, R. Swoboda, and H. Zimmermann, "1.25 Gbit/s Over 50m Step-Index Plastic Optical Fiber Using a Fully Integrated Optical Receiver With an Integrated Equalizer," *J. Lightwave Technol.*, vol. 30, no. 1, pp. 118–122, Jan 2012.

[101] M. Fortsch and H. Zimmermann, "2.5Gbit/s fiber receiver with integrated PIN photodiode in low-cost 0.6um BiCMOS," *Detectors and Associated Signal Processing II: Proceedings of the SPIE*, vol. 5964, pp. 596 401–596 401–7, 2005.

[102] I. O. for Standardization (ISO), *Information Technology - Open Systems Interconnection - Basic Reference Model ISO/IEC 7498-1*, Online, International Organisation for Standardization (ISO) Std. 7498-1, Rev. 2, Nov 1994. [Online]. Available: http://standards.iso.org/ittf/PubliclyAvailableStandards/s020269_ISO_IEC_7498-1_1994(E).zip

[103] S. D. Personick, "Receiver Design for Digital Fiber Optic Communication Systems, I," *Bell System Technical Journal*, vol. 52, no. 6, pp. 843–874, 1973. [Online]. Available: http://dx.doi.org/10.1002/j.1538-7305.1973.tb01993.x

[104] A. Funk and M. Beck, "Sub-Poissonian Photocurrent Statistics: Theory and Undergraduate Experiment," *American Journal of Physics*, vol. 65, pp. 492–500, 1997.

[105] P. Kok, "Limitations on Building Single-Photon-Resolution Detection Devices," in *Frontiers in Optics*. Optical Society of America, 2003, p. MT5.

[106] P. Babu and B. Srinivasan, "Countering Atmospheric Turbulence in Free Space Optical Links Using Wavelet Based Signal Processing," in *Global Telecommunications Conference. IEEE GLOBECOM. IEEE*, 2008, pp. 1 –4.

[107] F. Chou, C. Wang, G. Chen, and Y. Hsin, "An 8.7 GHz Si Photodiode in Standard 0.18-um CMOS Technology," in *OptoElectronics and Communications Conference (OECC), 2010 15th*, july 2010, pp. 826 –827.

[108] B. Ciftcioglu, L. Zhang, J. Zhang, J. Marciante, J. Zuegel, R. Sobolewski, and H. Wu, "Integrated Silicon PIN Photodiodes Using Deep N-Well in a Standard 0.18- um CMOS Technology," *Lightwave Technology, Journal*, vol. 27, no. 15, pp. 3303 –3313, 2009.

[109] M. Ghioni, F. Zappa, V. Kesan, and J. Warnock, "A VLSI-Compatible High-Speed Silicon Photodetector for Optical Data Link Applications," *Electron Devices, IEEE Transactions on*, vol. 43, no. 7, pp. 1054 –1060, jul 1996.

[110] C. Hermans and M. Steyaert, "A High-Speed 850nm Optical Receiver Front-end in 0.18um CMOS," *Solid-State Circuits, IEEE Journal of*, vol. 41, no. 7, pp. 1606 – 1614, July 2006.

[111] F. Tavernier and M. Steyaert, "A High-Speed Fully Integrated Optical Receiver in Standard 130nm CMOS," in *Electronics, Circuits and Systems. ICECS. 15th IEEE International Conference on*, 31 2008-sept. 3 2008, pp. 806 –809.

[112] F. Aznar, S. Celma, and B. Calvo, "A 0.18-um CMOS 1.25-Gbps Front-End Receiver for Low-Cost Short Reach Optical Communications," in *ESSCIRC, Proceedings of the*, sept. 2010, pp. 554 –557.

[113] Zerbe, Garlepp, Chau, Donnelly, Horowitz, Sidiropoulos, Garrett, and Werner, "Integrating Receiver with Precharge Circuitry," U.S. Patent 12/897c61, 2011.

[114] J. Zerbe, B. Daly, L. Luo, B. Stonecypher, W. Dettloff, J. Eble, T. Stone, J. Ren, B. Leibowitz, M. Bucher, P. Satarzadeh, and Q. Lin, "A 5Gb/s Link with Clock Edge Matching and Embedded Common Mode Clock for Low Power Interfaces," in *VLSI Circuits (VLSIC), IEEE Symposium on*, 2010, pp. 75 –76.

[115] D. W. Jordan and P. Smith, *Mathematical Techniques: An Introduction for the Engineering, Physical, and Mathematical Sciences*, 3rd ed. Oxford University Press, 2002.

[116] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, 2nd ed., ser. McGraw-Hill Series in Electrical Engineering. McGraw-hill, Inc, 1984.

[117] T. Wang and F. Yuan, "A New CMOS Current Integrating Receiver For GByte/s Parallel Links," in *Circuits and Systems, MWSCAS. 49th IEEE International Midwest Symposium on*, vol. 1, aug. 2006, pp. 230 –233.

[118] T. Morf, C. Brys, P. De Dobbelaere, P. Van Daele, P. Demeester, T. Martinson, and W. Bachtold, "Integrating Optical Receiver Transplanted by Epitaxial Lift Off," in *Gallium Arsenide Integrated Circuit (GaAs IC) Symposium, 17th Annual IEEE*, oct-1 nov 1995, pp. 189 –192.

[119] H. Coldenstrodt-Ronge and C. Silberhorn, "Avalanche Photo-Detection For High Data Rate Applications," *Journal of Physics B: Atomic, Molecular and Optical Physics*, vol. 40, no. 19, pp. 3909–3921, 2007.

[120] A. Spinelli and A. Lacaita, "Physics and Numerical Simulation of Single-Photon Avalanche Diodes," *Electron Devices, IEEE Transactions on*, vol. 44, no. 11, pp. 1931 –1943, 1997.

[121] M. Assanelli, A. Ingargiola, I. Rech, A. Gulinatti, and M. Ghioni, "Photon-Timing Jitter Dependence on Injection Position in Single-Photon Avalanche Diodes," *Quantum Electronics, IEEE Journal of*, vol. 47, no. 2, pp. 151–159, 2011.

[122] C. Tan, J. Ng, G. Rees, and J. David, "Statistics of Avalanche Current Buildup Time in Single-Photon Avalanche Diodes," *Selected Topics in Quantum Electronics, IEEE Journal of*, vol. 13, no. 4, pp. 906–910, 2007.

[123] R. H. Haitz, "Mechanisms Contributing to the Noise Pulse Rate of Avalanche Diodes ," *J. Appl. Phys*, vol. 36, no. 10, pp. 3123–3131, 1965.

[124] R. Mcintyre, "The Distribution of Gains in Uniformly Multiplying Avalanche Photodiodes: Theory," *Electron Devices, IEEE Transactions on*, vol. 19, no. 6, pp. 703–713, 1972.

[125] J. Blazej and I. Prochazka, "Avalanche Photodiode Output Pulse Rise-Time Study," in *Proc. SPIE 7355, Photon Counting Applications, Quantum Optics, and Quantum Information Transfer and Processing II, 73550M*, vol. 7355. SPIE, May 2009, pp. 73 550M–73 550M–11.

[126] S. Pellegrini, R. Warburton, L. Tan, J. S. Ng, A. Krysa, K. Groom, J. David, S. Cova, M. Robertson, and G. Buller, "Design and Performance of an InGaAs-InP Single-Photon Avalanche Diode Detector," *Quantum Electronics, IEEE Journal of*, vol. 42, no. 4, pp. 397 – 403, april 2006.

[127] R. H. Haitz, "Model for the Electrical Behavior of a Microplasma," *Journal of Applied Physics*, vol. 35, no. 5, pp. 1370–1376, 1964.

[128] A. Rochas, M. Gosch, A. Serov, P. Besse, R. Popovic, T. Lasser, and R. Rigler, "First

Fully Integrated 2-D Array of Single-Photon Detectors in Standard CMOS Technology," *Photonics Technology Letters, IEEE*, vol. 15, no. 7, pp. 963 –965, july 2003.

[129] C. Niclass, A. Rochas, P.-A. Besse, and E. Charbon, "Design and Characterization of a CMOS 3-D Image Sensor based on Single Photon Avalanche Diodes," *Solid-State Circuits, IEEE Journal of*, vol. 40, no. 9, pp. 1847 – 1854, 2005.

[130] R. L. Batdorf, A. Chynoweth, G. C. Dacey, and P. Foy, "Uniform Silicon pn Junctions. I. Broad Area Breakdown," *Journal of Applied Physics*, vol. 31, no. 7, pp. 1153–1160, 1960.

[131] R. J. McIntyre, "Theory of Microplasma Instability in Silicon," *Journal of Applied Physics*, vol. 32, no. 6, pp. 983–995, 1961.

[132] W. G. Oldham, R. Samuelson, and P. Antognetti, "Triggering Phenomena in Avalanche Diodes," *Electron Devices, IEEE Transactions on*, vol. 19, no. 9, pp. 1056–1060, 1972.

[133] H. Sigmund, "Photoelectrical Properties of Spherical Avalanche Diodes in Silicon," *Infrared Physics*, vol. 8, no. 4, pp. 259 – 264, 1968.

[134] P. Antognetti, S. Cova, and A. Longoni, "A Study of the Operation and Performances of an Avalanche Diode as a Single-Photon Detector," in *Euratom Publication EUR 5370e: Proc. 2nd Ispra Nuclear Electronics Symposium: Stresa*, May 20-23 1975, pp. 453–456.

[135] D. A. Ramirez, M. M. Hayat, G. J. Rees, X. Jiang, and M. A. Itzler, "New Perspective on Passively Quenched Single Photon Avalanche Diodes: Effect of Feedback on Impact Ionization," *Opt. Express*, vol. 20, no. 2, pp. 1512–1529, Jan 2012.

[136] F. Zappa, A. Tosi, A. D. Mora, and S. Tisa, "SPICE Modeling of Single Photon Avalanche Diodes," *Sensors and Actuators A: Physical*, vol. 153, no. 2, pp. 197 – 204, 2009.

[137] R. G. W. Brown, K. D. Ridley, and J. G. Rarity, "Characterization of Silicon Avalanche Photodiodes for Photon Correlation Measurements. 1: Passive Quenching," *Appl. Opt.*, vol. 25, no. 22, pp. 4122–4126, Nov 1986.

[138] G. Giustolisi, R. Mita, and G. Palumbo, "Behavioral Modeling of Statistical Phenomena of Single-Photon Avalanche Diodes," *International Journal of Circuit Theory and Applications*, vol. 40, no. 7, pp. 661–679, 2012.

[139] S. Verghese, K. McIntosh, R. Molnar, L. Mahoney, R. Aggarwal, M. Geis, K. Molvar, E. Duerr, and I. Melngailis, "GaN Avalanche Photodiodes Operating in Linear-Gain Mode and Geiger Mode," *Electron Devices, IEEE Transactions on*, vol. 48, no. 3, pp. 502 –511, mar 2001.

[140] N. S. Nightingale, "A New Silicon Avalanche Photodiode Photon Counting Detector Module for Astronomy," *Experimental Astronomy*, vol. 1, pp. 407–422, 1991.

[141] L. Pancheri and D. Stoppa, "Low-Noise CMOS Single-Photon Avalanche Diodes with 32 ns Dead Time," in *Solid State Device Research Conference, ESSDERC. 37th European*, sept. 2007, pp. 362 –365.

[142] N. Faramarzpour, M. Deen, S. Shirani, and Q. Fang, "Fully Integrated Single Photon Avalanche Diode Detector in Standard CMOS 0.18um Technology," *Electron Devices, IEEE Transactions on*, vol. 55, no. 3, pp. 760–767, 2008.

[143] J. Richardson, L. Grant, and R. Henderson, "Low Dark Count Single-Photon Avalanche Diode Structure Compatible With Standard Nanometer Scale CMOS Technology," *Photonics Technology Letters, IEEE*, vol. 21, no. 14, pp. 1020 –1022, 2009.

[144] D. Bronzi, F. Villa, S. Bellisai, B. Markovic, S. Tisa, A. Tosi, F. Zappa, S. Weyers, D. Durini, W. Brockherde, and U. Paschen, "Low-Noise and Large-Area CMOS SPADs with Timing Response Free from Slow Tails," in *ESSDERC, Proceedings of*, 2012.

[145] T. Frach, W. Ruetten, K. Fiedler, G. Maehlum, T. Solf, and A. Thon, "Assessment of Photodiodes as a Light Detector for PET Scanners," in *Nuclear Science Symposium Conference Record, 2004 IEEE*, vol. 7, oct. 2004, pp. 4177 –4181.

[146] T. Frach, "Photodiodes and Fabrication Thereof," US Patent 0 127 314A1, 2010.

[147] I. Rech, A. Ingargiola, R. Spinelli, I. Labanca, S. Marangoni, M. Ghioni, and S. Cova, "Optical Crosstalk in Single Photon Avalanche Diode Arrays: A New Complete Model," *Opt. Express*, vol. 16, no. 12, pp. 8381–8394, 2008.

[148] B. Rae, D. Renshaw, E. Charbon, and R. Henderson, "A 2-D Time Resolved Photon Counting Imager in 0.35um Technology," in *European Optical Society Conference on Frontiers in Electronic Imaging*, 2007, pp. 50–51.

[149] E. Charbon, P. Miliozzi, L. Carloni, A. Ferrari, and A. Sangiovanni-Vincentelli, "Modeling Digital Substrate Noise Injection in Mixed-Signal IC's," *Computer-Aided Design of Integrated Circuits and Systems*, vol. 18, no. 3, pp. 301 –310, mar 1999.

[150] S. Crespo, S. Ouzounov, and T. Guerrero, "Estimation of Substrate Noise from Digital Circuits in Very Large Sensor Arrays," in *Proceedings of ICT.OPEN 2011 (proRISC)*, 2011.

[151] V. Savuskan, I. Brouk, M. Javitt, and Y. Nemirovsky, "An Estimation of Single Photon Avalanche Diode (SPAD) Photon Detection Efficiency (PDE) Nonuniformity," *Sensors Journal, IEEE*, vol. 13, no. 5, pp. 1637–1640, 2013.

[152] C. Kimura and J. Nishizawa, "Turn-on Mechanism of a Microplasma," *Jpn. J. Appl. Phys*, vol. 7, no. 12, pp. 1453–1463, Dec 1968.

[153] M. Gersbach, Y. Maruyama, E. Labonne, J. Richardson, R. Walker, L. Grant, R. Henderson, F. Borghetti, D. Stoppa, and E. Charbon, "A Parallel 32x32 Time-to-Digital Converter Array Fabricated in a 130nm Imaging CMOS Technology," in *ESSCIRC '09. Proceedings of*, 2009, pp. 196 –199.

[154] M. Gersbach, Y. Maruyama, R. Trimananda, M. Fishburn, D. Stoppa, J. Richardson, R. Walker, R. Henderson, and E. Charbon, "A Time-Resolved, Low-Noise Single-Photon Image Sensor Fabricated in Deep-Submicron CMOS Technology," *Solid-State Circuits, IEEE Journal of*, vol. 47, no. 6, pp. 1394 –1407, june 2012.

[155] M. Gersbach, R. Trimananda, Y. Maruyama, M. Fishburn, D. Stoppa, J. Richardson, R. Walker, R. K. Henderson, and E. Charbon, "High Frame-Rate TCSPC-FLIM Using a Novel SPAD-Based Image Sensor," in *Proc. SPIE 7780, 77801H*, 2010.

[156] D. Chitnis and S. Collins, "A SPAD-Based Photon Detecting System for Optical Communications," *Lightwave Technology, Journal of*, vol. 32, no. 10, pp. 2028–2034, May 2014.

[157] R. Henderson, E. Webster, R. Walker, J. Richardson, and L. Grant, "A 3x3, 5um Pitch, 3-Transistor Single Photon Avalanche Diode Array with Integrated 11V Bias Generation in 90nm CMOS Technology," in *Electron Devices Meeting (IEDM), 2010 IEEE International*, 2010, pp. 14.2.1 –14.2.4.

[158] L. Braga, L. Gasparini, L. Grant, R. Henderson, N. Massari, M. Perenzoni, D. Stoppa, and R. Walker, "An 8x16-Pixel 92kSPAD Time-Resolved Sensor with on-Pixel 64ps 12b TDC and 100MS/s Real-Time Energy Histogramming in 0.13um CIS Technology for PET/MRI Applications," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2013 IEEE International*, 2013, pp. 486–487.

[159] L. Braga, L. Pancheri, L. Gasparini, R. Henderson, and D. Stoppa, "A Mini-SiPM Array for PET Detectors Implemented in 0.35-μm HV CMOS Technology," in *Ph.D. Research*

in Microelectronics and Electronics (PRIME), 2011 7th Conference on, july 2011, pp. 181 –184.

[160] C. Niclass, M. Soga, H. Matsubara, and S. Kato, "A 100m-Range 10-Frame/s 340x96-Pixel Time-of-Flight Depth Sensor in 0.18um CMOS," in *ESSCIRC, Proceedings of*, 2011, pp. 107 –110.

[161] R. Mita and G. Palumbo, "High-Speed and Compact Quenching Circuit for Single-Photon Avalanche Diodes," *Instrumentation and Measurement, IEEE Transactions on*, vol. 57, no. 3, pp. 543–547, 2008.

[162] J. Richardson, R. Henderson, and D. Renshaw, "Dynamic Quenching for Single Photon Avalanche Diode Arrays," in *Proceedings of 2007 International Image Sensor Workshop*, 2007.

[163] A. Eisele, R. Henderson, B. Schmidtke, T. Funk, L. Grant, J. Richardson, and W. Freude, "185 MHz Count Rate, 139 dB Dynamic Range Single-Photon Avalanche Diode with Active Quenching Circuit in 130nm CMOS Technology," in *2011 International Image Sensor Workshop*, IEEE. IEEE Electron Devices Society, 2011, pp. 278–280.

[164] S. Lee and R. Gardner, "A New G-M Counter Dead Time Model," *Applied Radiation and Isotopes*, vol. 53, pp. 731–737, 2000.

[165] M. Stipcevic, "Active Quenching Circuit for Single-Photon Detection with Geiger Mode Avalanche Photodiodes," *Appl. Opt.*, vol. 48, no. 9, pp. 1705–1714, Mar 2009.

[166] L. Neri, S. Tudisco, F. Musumeci, A. Scordino, G. Fallica, M. Mazzillo, and M. Zimbone, "Note: Dead Time Causes and Correction Method for Single Photon Avalanche Diode Devices," *Review of Scientific Instruments*, vol. 81, no. 8, p. 086102, 2010.

[167] C. Hu, M. Liu, X. Zheng, and J. Campbell, "Dynamic Range of Passive Quenching Active Reset Circuit for Single Photon Avalanche Diodes," *Quantum Electronics, IEEE Journal of*, vol. 46, no. 1, pp. 35 –39, jan. 2010.

[168] W. H. Farr, J. Gin, and D. Nguyen, "Gigahertz Bandwidth Photon Counting," in *Proc. SPIE 7320*. SPIE, 2009, p. 732006.

[169] C. Niclass and M. Soga, "A Miniature Actively Recharged Single-Photon Detector Free of Afterpulsing Effects with 6ns Dead Time in a 0.18 um CMOS Technology," in *Electron Devices Meeting (IEDM), 2010 IEEE International*, 2010, pp. 14.3.1 –14.3.4.

[170] B. Picinbono, "Output Dead-Time in Point Processes," *Communications in Statistics - Simulation and Computation*, vol. 38, pp. 2198–2213, 2009.

[171] A. Rochas, "Single Photon Avalanche Diodes in CMOS Technology," PhD Thesis, Ecole Polytechnique Federale De Lausanne, 2003, no: 2814 (2003).

[172] M. Gersbach, J. Richardson, E. Mazaleyrat, S. Hardillier, C. Niclass, R. Henderson, L. Grant, and E. Charbon, "A Low-Noise Single-Photon Detector Implemented in a 130nm CMOS Imaging Process," *Solid-State Electronics*, vol. 53, no. 7, pp. 803 – 808, 2009.

[173] J. A. Richardson, "Time Resolved Single Photon Imaging in Nanometer Scale CMOS Technology," Ph.D Thesis, The School of Engineering, The University of Edinburgh, 2010.

[174] D. V. Lang, "Deep-Level Transient Spectroscopy: A New Method to Characterize Traps in Semiconductors," *Journal of Applied Physics*, vol. 45, no. 7, pp. 3023 –3032, jul 1974.

[175] M. Karami, L. Carrara, C. Niclass, M. Fishburn, and E. Charbon, "RTS Noise Characterization in Single-Photon Avalanche Diodes," *Electron Device Letters, IEEE*, vol. 31, no. 7, pp. 692 –694, july 2010.

[176] C. Stapels, W. Lawrence, F. Augustine, and J. Christian, "Characterization of a CMOS Geiger Photodiode Pixel," *Electron Devices, IEEE Transactions on*, vol. 53, no. 4, pp. 631 – 635, april 2006.

[177] R. Hadfield, "Single-Photon Detectors for Optical Quantum Information Applications," *Nature Photonics*, vol. 3, no. 12, pp. 696–705, 2009.

[178] S. Castelletto, I. Degiovanni, V. Schettini, and A. Migdall, "Reduced Deadtime and Higher Rate Photon-Counting Detection using a Multiplexed Detector Array," *Journal of Modern Optics*, vol. 54, no. 2, pp. 337–352, 2007.

[179] F. Zappa, M. Ghioni, S. Cova, C. Samori, and A. Giudice, "An Integrated Active-Quenching Circuit for Single-Photon Avalanche Diodes," *Instrumentation and Measurement, IEEE Transactions on*, vol. 49, no. 6, pp. 1167–1175, 2000.

[180] A. Vila, E. Vilella, O. Alonso, and A. Dieguez, "Crosstalk-Free Single Photon Avalanche Photodiodes Located in a Shared Well," *Electron Device Letters, IEEE*, vol. 35, no. 1, pp. 99–101, Jan 2014.

[181] Y. Maruyama and E. Charbon, "An All-Digital, Time-Gated 128X128 SPAD Array for On-Chip, Filter-Less Fluorescence Detection," in *Solid-State Sensors, Actuators and Microsystems Conference (TRANSDUCERS)*, june 2011, pp. 1180 –1183.

[182] E. A. G. Webster, J. A. Richardson, L. A. Grant, D. Renshaw, and R. K. Henderson, "A Single-Photon Avalanche Diode in 90-nm CMOS Imaging Technology With 44% Photon Detection Efficiency at 690 nm," *Electron Device Letters, IEEE*, vol. PP, no. 99, pp. 1 –3, 2012.

[183] G. Giustolisi, R. Mita, and G. Palumbo, "Verilog-A Modeling of SPAD Statistical Phenomena," in *Circuits and Systems (ISCAS), 2011 IEEE International Symposium on*, may 2011, pp. 773 –776.

[184] M. Hsu, S. 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*, vol. 30, no. 6, pp. 641 –643, 2009.

[185] E. Webster, "Single-Photon Avalanche Diode Theory, Simulation, and High Performance CMOS Integration," PhD, The School of Engineering, The University of Edinburgh, 2013.

[186] W. E. Lamb, "Anti-Photon," *Applied Physics B: Lasers and Optics*, vol. 60, pp. 77–84, 1995.

[187] T. Leitner, A. Feiningstein, R. Turchetta, R. Coath, S. Chick, G. Visokolov, V. Savuskan, M. Javitt, L. Gal, I. Brouk, S. Bar-Lev, and Y. Nemirovsky, "Measurements and Simulations of Low Dark Count Rate Single Photon Avalanche Diode Device in a Low Voltage 180-nm CMOS Image Sensor Technology," *Electron Devices, IEEE Transactions on*, vol. 60, no. 6, pp. 1982–1988, 2013.

[188] L. Braga, L. Pancheri, L. Gasparini, M. Perenzoni, R. Walker, R. Henderson, and D. Stoppa, "A CMOS Mini-SiPM Detector with In-Pixel Data Compression for PET Applications," in *Nuclear Science Symposium and Medical Imaging Conference*, 2011, pp. 548 –552.

[189] C. Degenhardt, B. Zwaans, T. Frach, and R. de Gruyter, "Arrays of Digital Silicon Photomultipliers; Intrinsic Performance and Application to Scintillator Readout," in *Nuclear Science Symposium Conference Record (NSS/MIC), 2010 IEEE*, 30 2010-nov. 6 2010, pp. 1954 –1956.

[190] L. Pancheri and D. Stoppa, "Low-Noise Single Photon Avalanche Diodes in 0.15um

CMOS Technology," in *Solid-State Device Research Conference (ESSDERC), 2011 Proceedings of the European*, Sept 2011, pp. 179–182.

[191] M. Karami, H. Yoon, and E. Charbon, "Single-Photon Avalanche Diodes in sub-100nm Standard CMOS Technologies," in *Proceedings of International Image Sensor Workshop (ISW)*, 2011.

[192] S. L. Rumyantsev, S. Sawyer, N. Pala, M. S. Shur, Y. Bilenko, J. P. Zhang, X. Hu, A. Lunev, J. Deng, and R. Gaska, "Low Frequency Noise of Light Emitting Diodes (Invited Paper)," in *SPIE Proceedings Vol. 5844: Noise in Devices and Circuits III.* SPIE, 2005, pp. 75–85. [Online]. Available: http://spie.org/x648.xml?product_id=608559

[193] P. Kappe, J. Kaiser, W. Elsasser, R. Wirth, and K. Streubel, "Investigations of the Fundamental Quantum Noise Properties of Resonant-Cavity Light-Emitting Diodes (RCLEDs)," *Annales Des Telecommunications*, vol. 58, no. 9-10, pp. 1424–1431, 2003. [Online]. Available: http://dx.doi.org/10.1007/BF03001738

[194] S. L. Rumyantsev, C. Wetzel, and M. S. Shur, "Wavelength-Resolved Low-Frequency Noise of GaInN/GaN Green Light Emitting Diodes," *Journal of Applied Physics*, vol. 100, no. 8, p. 084506, 2006. [Online]. Available: http://scitation.aip.org/content/aip/journal/jap/100/8/10.1063/1.2358409

[195] J. J. Brophy, "Fluctuations in Luminescent Junctions," *Journal of Applied Physics*, vol. 38, no. 6, pp. 2465–2469, May 1967.

[196] S. Lee, R. Gardner, and M. Jae, "Determination of Dead Times in the Recently Introduced Hybrid G-M Counter Dead Time Model," *Journal of Nuclear Science and Technology*, vol. 4, pp. 156–159, 2004.

[197] J. Peng, Y. Fu, L. Yao, X. Shang, Z. Lu, B. Yang, and L. Yu, "Quick Single-Photon Detector with Many Avalanche Photo Diodes Working on the Time Division," *Chin. Opt. Lett.*, vol. 6, no. 5, pp. 320–322, 2008.

[198] E. A. Dauler, A. J. Kerman, B. S. Robinson, J. K. W. Yang, B. Voronov, G. Goltsman, S. A. Hamilton, and K. K. Berggren, "Photon-Number-Resolution with sub-30-ps Timing using Multi-Element Superconducting Nanowire Single Photon Detectors," *Journal of Modern Optics*, vol. 56, pp. 364–373, 2009.

[199] V. Krishnan, *Probability and Random Processes*, ser. Wiley Survival Guides in Engineering and Science, E. Desurvire, Ed. Wiley, 2006. [Online]. Available: http://books.google.co.uk/books?id=RwLiEqDU5fcC

[200] M. J. Fitch, B. C. Jacobs, T. B. Pittman, and J. D. Franson, "Photon-Number Resolution using Time-Multiplexed Single-Photon Detectors," *Phys. Rev. A*, vol. 68, p. 043814, Oct 2003.

[201] P. Kok and S. L. Braunstein, "Detection Devices in Entanglement-Based Optical State Preparation," *Phys. Rev. A*, vol. 63, p. 033812, Feb 2001.

[202] S. H. Lee, M. Jae, and R. P. Gardner, "Non-Poisson Counting Statistics of a Hybrid G-M Counter Dead Time Model," *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, vol. 263, no. 1, pp. 46 – 49, 2007.

[203] C.-H. Yang and D. Markovic, "A 2.89mW 50GOPS 16x16 16-Core MIMO Sphere Decoder in 90nm CMOS," in *ESSCIRC, Proceedings of*, 2009, pp. 344 –347.

[204] S. Goswami, T. Copani, A. Jain, H. Karaki, B. Vermeire, H. Barnaby, G. Fetzer, R. Vercillo, and S. Kiaei, "A 96Gb/s-Throughput Transceiver for Short-Distance Parallel Optical Links," in *Solid-State Circuits Conference, ISSCC*, 2008, pp. 230 –609.

[205] F. Liu, D. Patil, J. Lexau, P. Amberg, M. Dayringer, J. Gainsley, H. Moghadam, X. Zheng, J. Cunningham, A. Krishnamoorthy, E. Alon, and R. Ho, "10-Gbps, 5.3-mW Optical Transmitter and Receiver Circuits in 40-nm CMOS," *Solid-State Circuits, IEEE Journal of*, vol. 47, no. 9, pp. 2049 –2067, sept. 2012.

[206] SPADnet. (2012) Fully Networked, Digital Components for Photon-starved Biomedical Imaging Systems. Accessed: 26/07/12. [Online]. Available: www.spadnet.eu

[207] J. Richardson, R. Walker, L. Grant, D. Stoppa, F. Borghetti, E. Charbon, M. Gersbach, and R. Henderson, "A 32x32 50ps Resolution 10-Bit Time to Digital Converter Array in 130nm CMOS for Time Correlated Imaging," in *Custom Integrated Circuits Conference. CICC. IEEE*, 2009, pp. 77 –80.

[208] D. Tyndall, B. Rae, D. Li, J. Arlt, A. Johnston, J. Richardson, and R. Henderson, "A High-Throughput Time-Resolved Mini-Silicon Photomultiplier With Embedded Fluorescence Lifetime Estimation in 0.13 μ m CMOS," *Biomedical Circuits and Systems, IEEE Transactions on*, vol. 6, no. 6, pp. 562 –570, dec. 2012.

[209] C. Niclass, A. Rochas, P.-A. Besse, and E. Charbon, "Toward a 3-D Camera Based on Single Photon Avalanche Diodes," *Selected Topics in Quantum Electronics, IEEE Journal of*, vol. 10, no. 4, pp. 796–802, 2004.

[210] M. Gersbach, C. Niclass, E. Charbon, J. Richardson, R. Henderson, and L. Grant, "A Single Photon Detector Implemented in a 130nm CMOS Imaging Process," in *Solid-State Device Research Conference, ESSDERC, European*, 2008, pp. 270 –273.

[211] R. Mita, G. Palumbo, and P. Fallica, "Accurate Model for Single-Photon Avalanche Diodes," *Circuits, Devices Systems, IET*, vol. 2, no. 2, pp. 207 –212, april 2008.

[212] S. Cova, A. Longoni, and A. Andreoni, "Towards Picosecond Resolution with Single-Photon Avalanche Diodes," *Review of Scientific Instruments*, vol. 52, no. 3, pp. 408–412, 1981.

[213] M. Gronholm, 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*, aug. 2009, pp. 5 –8.

[214] G. Kim, M.-K. Kim, B.-S. Chang, and W. Kim, "A Low-Voltage, Low-Power CMOS Delay Element," *Solid-State Circuits, Journal of*, vol. 31, no. 7, pp. 966 –971, 1996.

[215] J. F. Wakerly, *Digital Design: Principals and Practice*, 3rd ed., ser. Prentice Hall Xilinx Design Series. Prentice Hall, 2000.

[216] D. Stoppa, D. Mosconi, L. Pancheri, and L. Gonzo, "Single-Photon Avalanche Diode CMOS Sensor for Time-Resolved Fluorescence Measurements," *Sensors Journal, IEEE*, vol. 9, no. 9, pp. 1084 –1090, sept. 2009.

[217] S. Tisa, F. Guerrieri, A. Tosi, and F. Zappa, "100 kframe/s 8 bit Monolithic Single-Photon Imagers," in *Solid-State Device Research Conference. ESSDERC, European*, sept. 2008, pp. 274 –277.

[218] M. Ghioni, S. Cova, F. Zappa, and C. Samori, "Compact Active Quenching Circuit for Fast Photon Counting with Avalanche Photodiodes," *Review of Scientific Instruments*, vol. 67, no. 10, pp. 3440–3448, 1996.

[219] B. Razavi, *Monolithic Phase-Locked Loops and Clock Recovery Circuits: Theory and Design*, B. R. (Editor), Ed. Wiley - IEEE Press, 1996. [Online]. Available: http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0780311493,miniSiteCd-IEEE2.html

[220] S. Palermo, A. Emami-Neyestanak, and M. Horowitz, "A 90nm CMOS 16Gb/s Transceiver for Optical Interconnects," in *Solid-State Circuits Conference, ISSCC. IEEE International*, 2007, pp. 44 –586.

[221] T. Frach, G. Prescher, C. Degenhardt, R. de Gruyter, A. Schmitz, and R. Ballizany, "The Digital Silicon Photomultiplier; Principle of Operation and Intrinsic Detector Performance," in *Nuclear Science Symposium Conference Record (NSS/MIC), 2009 IEEE*, 24 2009-nov. 1 2009, pp. 1959 –1965.

[222] O. Kelly. http://www.opalkelly.com/. Website. Opal Kelly.

[223] P. M. D. Thomas, *The Verilog Hardware Description Language*, 1st ed., J. Allen, Ed. Kluwer Academic Publishers, 1991.

[224] D. Flanagan, *Java in a Nutshell*, 5th ed. O'Reilly, 2005.

[225] R. Henderson, E. Webster, and L. Grant, "A Dual-Junction Single-Photon Avalanche Diode in 130-nm CMOS Technology," *Electron Device Letters, IEEE*, vol. 34, no. 3, pp. 429–431, 2013.

[226] A. Lacaita, F. Zappa, S. Cova, and P. Lovati, "Single-photon Detection Beyond 1um: Performance of Commercially Available InGaAs/lnP Detectors," *Appl. Opt.*, vol. 35, no. 16, pp. 2986–2996, Jun 1996.

[227] Y. Yang, A. Marshak, S. Palm, T. Varnai, and W. Wiscombe, "Cloud Impact on Surface Altimetry From a Spaceborne 532-nm Micropulse Photon-Counting Lidar: System Modeling for Cloudy and Clear Atmospheres," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, no. 12, pp. 4910–4919, Dec 2011.

[228] T. W. Murphy, J. D. Strasburg, C. W. Stubbs, E. G. Adelberger, L. Tom, A. E. Orin, E. L. Michelsen, J. Battat, C. D. Hoyle, E. Swanson, and E. Williams, "APOLLO: Meeting the Millimeter Goal," in *Proc. 14th International Workshop on Laser Ranging*, 2004.

[229] D. Imaging and R. S. L. R. I. of Technology. (2014, May) Digital Imaging and Remote Sensing Image Generation (DIRSIG) Model: LIDAR Modality Handbook. Online : Accessed 16/05/14. Rochester Institute of Technology. [Online]. Available: http://www.dirsig.org/docs/new/lidar.html

[230] T. Murphy, "APOLLO Documentation: Statistics of Photon Arrival Time," Apollo Project: Department of Physics: University of California San Diego," Online Project Documentation, 2001. [Online]. Available: http://www.physics.ucsd.edu/~tmurphy/apollo/doc/arrival.pdf

[231] R. G. W. Brown, R. Jones, J. G. Rarity, and K. D. Ridley, "Characterization of Silicon Avalanche Photodiodes for Photon Correlation Measurements. 2: Active Quenching," *Appl. Opt.*, vol. 26, no. 12, pp. 2383–2389, Jun 1987.

[232] N. Brunner, "Modeling Internal Heating of Optoelectronic Devices Using COMSOL," in *COMSOL Conference 2012, Proceedings of*, 2012. [Online]. Available: http://www.uk.comsol.com/offers/conference2012papers/papers/presentation/area/optics/id/13085/

[233] P. LiFi, "Pure lifi: The home of li-fi," Online, June 2014, http://purelifi.com/, was PureVLC, name change to PureLiFi. [Online]. Available: http://purelifi.com/

[234] S. Zhang, S. Watson, J. McKendry, D. Massoubre, A. Cogman, E. Gu, R. Henderson, A. Kelly, and M. Dawson, "High-Bandwidth Parallel Data Transmission Using GaN/CMOS Micro-LED Arrays," in *Photonics Conference (IPC), 2012 IEEE*, Sept 2012, pp. 38–39.

[235] H. W. Ott, *Noise Reduction Techniques in Electronic Systems*, 2nd ed. John Wiley & Sons, 1988.

[236] R. Swoboda and H. Zimmermann, "2.5 Gbit/s Silicon Receiver OEIC with Large Diameter Photodiode," *Electronics Letters*, vol. 40, no. 8, pp. 505 – 507, april 2004.

[237] S. M. Press. (2013, Feb) STMicroelectronics Press Release: Proximity Sensor

Solves Smartphone Hang-ups. Online. ST Microelectronics. [Online]. Available: http://www.st.com/web/en/press/p3397