Distortion Losses of High Speed Single-Photon Avalanche Diode Receivers Approaching Quantum Sensitivity

John Kosman

Doctor of Philosophy

THE UNIVERSITY OF EDINBURGH

2021

Lay Abstract

Data traffic is continuously growing around the world and the available radio frequency spectrum is under pressure to meet these demands. This draws attention towards communication with alternatives such as visible light, which offers significant cost and energy savings since existing lighting infrastructure can be used.

Present receivers need very high optical power to detect a data signal. In this work, single-photon detectors are used to enhance the receiver sensitivity and solve this problem. A simulation model is built to describe the physical drawbacks of the system, such as the speed of the detectors. The model is validated with two real single-photon devices integrated in standard digital silicon technology – enabling small form-factor, low cost, low power, and manufacturability.

The integrated single-photon receiver demonstrates a high data rate with very low visible light levels and is up to 100× more energy efficient than conventional optical receivers.

Abstract

Data traffic is growing exponentially, and the radio frequency (RF) spectrum is under pressure to meet these demands. Visible light communication (VLC) has hundreds of terahertz of unused and unregulated bandwidth and the widespread use of solid-state lighting makes it viable to supplement RF networks. Present optical receivers (RXs) use positive-intrinsic-negative (PIN) diodes or avalanche photodiodes (APDs) and amplification circuitry that impairs RX sensitivity. In this work, the extremely high gain of single-photon avalanche diodes (SPADs) is utilised to remove the need for an amplifier. This offers significantly improved sensitivity and allows the quantum limit of detection to be approached.

A SPAD array integrated in 40 nm CMOS is used to determine the transient response of SPADs and investigate the effect of dead time after a photon is detected. A 130 nm CMOS SPAD array RX in this work achieves 500 Mb/s four-level pulse amplitude modulation and 350 Mb/s OFDM in a 450 nm laser diode-based VLC link within 15.2 dB of the quantum limit. However, SPAD dead time induces around 5.7 dB of transient distortion which restricts error performance and data rate an order of magnitude below that of APDs. This thesis builds a model of a discrete photon counting system which exhibits this nonlinear behaviour and compares it to practical measurements with the RX. A unipolar intensity-modulated optical signal is considered, as opposed to bipolar electric fields in conventional RF systems. Intermodulation is analysed, and the resulting degradation of signal-to-noise-and-distortion ratio and bit error rate is evaluated. The model is a tool for understanding distortion to ultimately allow rectification through RX architecture, modulation scheme, coding, and equalisation techniques. The thesis concludes that the SPAD RX is effective with very low optical power, allowing considerable improvements of two orders of magnitude in transmitter energy efficiency or one order of magnitude in link distance compared to present VLC systems – useful for underwater applications. This work proves that the high electrical power consumption disadvantage due to the SPAD bias can be alleviated by operating the RX in an optimum region determined in the model. Further savings and integration advantages are gained by using CMOS. This SPAD RX demonstrates the lowest power consumption and highest sensitivity to date. The need for narrow bandpass spectral filtering in bright ambient light conditions remains a limitation of the SPAD RX.

Acknowledgements

I would like to thank my supervisors Prof Robert Henderson, Kevin Moore, and Prof Harald Haas for the opportunity to work towards this PhD. I thank STMicroelectronics for sponsoring my work and for use of their 40 nm and 130 nm CMOS imaging processes. I greatly appreciate the support from my supervisors throughout and for giving helpful feedback on my papers.

Thanks goes to Sarrah Patanwala for designing the 40 nm chip along with its PCB. I thank Tarek Al Abbas, Neale Dutton and Richard Walker for the design and layout of the 130 nm chip; Oscar Almer for the RX firmware; Elham Sarbazi for discussion on theory and Stefan Videv for lab equipment. I thank Shenjie Huang for inventing the photon time information-based detection algorithm.

The 130 nm CMOS SPAD IC was funded by the Engineering and Physical Sciences Research Council through the Ultra-Parallel Visible Light Communications Project under Grant EP/K00042X/1. I am grateful to STMicroelectronics for chip fabrication.

I would also like to thank Salvatore Gnecchi for the initial algorithm idea that I developed for my transient simulation code. A special thanks is extended to all my colleagues in the CMOS Sensors and Systems Group and the LiFi R&D Centre for their inspiring enthusiasm and helpful discussions, and to pureLiFi for putting up with my eccentric way of working!

Thank you to my parents and sisters for the support. Most of all, por siempre, Laura mi amor.

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:

Prof Robert Henderson, The Institute for Integrated Micro and Nano Systems, The University of Edinburgh

Kevin Moore, STMicroelectronics

Prof Harald Haas, LiFi Research and Development Centre, The University of Strathclyde

John Kosman

Date: …02/08/2021……

Contents

List of Publications

Journals

J. Kosman, K. Moore, H. Haas, and R. K. Henderson. "Distortion losses of high-speed single-photon avalanche diode optical receivers approaching quantum sensitivity," *The Royal Society Philosophical Transactions A*, March 2, 2020.

Co-author: S. Huang, S. M. Patanwala, J. Kosman, R. K. Henderson, and M. Safari, "Optimal Photon Counting Receiver for Sub-Dead-Time Signal Transmission," in *Journal of Lightwave Technology*, vol. 38, no. 18, pp. 5225-5235, Sept. 15, 2020.

Conferences

J. Kosman, O. Almer, T. Al Abbas, N. Dutton, R. Walker, S. Videv, K. Moore, H. Haas, and R. K. Henderson, "A 500 Mb/s -46.1 dBm CMOS SPAD Receiver for Laser Diode Visible-Light Communications," *IEEE International Solid-State Circuits Conference (ISSCC)*, San Francisco, CA, USA, pp. 468-470, February 17, 2019 (Speaker).

Co-author: J. F. C. Carreira, G. N. Arvanitakis, A. D. Griffiths, J. D. McKendry, E. Xie, J. Kosman, R. K. Henderson, E. Gu, and M. D. Dawson, "Underwater Wireless Optical Communications at 100 Mb/s using Integrated Dual-Color Micro-LEDs," *2019 IEEE Photonics Conference (IPC)*, San Antonio, TX, USA, pp. 1-2, 2019.

J. Kosman, O. Almer, A. V. N. Jalajakumari, S. Videv, H. Haas and R. K. Henderson, "60 Mb/s, 2 meters visible light communications in 1 klx ambient using an unlensed CMOS SPAD receiver," *2016 IEEE Photonics Society Summer Topical Meeting Series (SUM)*, Newport Beach, CA, USA, pp. 171-172, 2016.

List of Figures

List of Tables

List of Mathematical Symbols

$a_1, a_2, a_3, \ldots$	Nonlinear system polynomial coefficients
$a(p)$	Information symbol sequence
A	Amplitude of a modulated signal
A_e	Effective collection area of the detector
B	Half of peak-to-peak amplitude of a modulated signal
BER_{QL}	Quantum limit BER
C_{FF}	Fill factor
e	Euler's number
f	Frequency
F_s	Sampling rate
G	Gain of the nonlinear SPAD system
G_{PQ}	Detection gain of a passively quenched SPAD
h	Planck's constant
i	An integer
I_h	Logic one current level
I_l	Logic zero current level
I_r	Irradiance of incident light
I_{th}	Threshold current
k	Number of photon arrivals
K_n	Mean number of noise photons detected within a symbol
K_S	Mean number of signal photons detected within a symbol
λ	Average photo rate
λ_n	Mean background noise photon rate
λ_S	Mean signal photon rate
μ_0	Mean photocount distribution
μ_x	Mean of an aggregate distribution

m_T	Threshold for photocounts
n	Integer multiples of the reciprocal of τ
p	Symbol index
P_h	Signal power representing an OOK high level
P_l	Signal power representing an OOK low level
$p_K(k)$	PMF of a PQ SPAD
$p_0(k)$	Poisson probability distribution
$p_x(x)$	Aggregate count distribution of a SPAD array
P_{op}	Average optical power
PP_{DCR}	Power penalties caused by DCR
PP_{ISI}	Power penalties caused by ISI
PP_{FF}	Power penalties caused by fill factor
PP_{PDE}	Power penalties caused by PDE
P_r	Average received optical power
$P_{r\ OFDM}$	Average received optical power for OFDM
P_s	Received optical power
π	Pi
R	Symbol rate (reciprocal of period, T)
s(t)	Pulse shape
σ_0	Standard deviation
σ^2	Variance
σ_0^2	Variance of an aggregate distribution
t	Time (discrete)
T	Time interval, symbol period
T_B	Block time
T_L	Time instant when the last photon arrives before a symbol
T_F	Time when a SPAD output pulse falling edge occurs

τ	Dead time
τ_c	Carrier lifetime
t_{on}	Turn-on delay
T_{st}	Beginning time of the counting period of a symbol
ν	Frequency of a wave in Hz
ω	Frequency of a modulated signal in rads^{-1}
$w(t)$	Baseband signal
$x(t)$	Input signal to the nonlinear system
$X(f)$	Fourier-transform of x(t)
$y(x,t)$	Output signal of the nonlinear system

List of Abbreviations

3D	Three-dimensional
ADC	Analogue-to-digital converter
ACO-OFDM	Asymmetrically-clipped optical-OFDM
APD	Avalanche photodiode
ASiPM	Analogue silicon photomultiplier
ASK	Amplitude Shift Keying
AQ	Active quenching
AWG	Arbitrary waveform generator
AWGN	Additive white Gaussian noise
BCH	Bose-Chaudhuri-Hocquenghem
BER	Bit error rate
Bi-CMOS	Bipolar CMOS
BPSK	Binary phase shift keying
BW	Bandwidth
CAGR	Compound annual growth rate
CDR	Clock-and-data recovery
CIR	Channel impulse response
CP	Cyclic prefix
CPC	Compound parabolic concentrator
CMOS	Complementary Metal Oxide Semiconductor
CSOC	Convolutional self-orthogonal code
CW	Continuous wave
DAC	Digital-to-analogue converter
DC	Direct current
DCO-OFDM	DC-biased O-OFDM
DCR	Dark count rate
DD	Direct detection
DDR	Double data rate

DFE	Decision feedback equalizer
DNWELL	Deep NWELL
DSiPM	Digital Si photomultiplier
DSP	Digital signal processing
EMI	Electromagnetic Interference
EQ	Equalization
ER	Extinction ratio
FEC	Forward Error Correction
FF	Fill Factor
FFT	Fast Fourier transform
FOV	Field of view
FPGA	Field-programmable gate array
FSK	Frequency Shift Keying
FWF	Fibre-wireless-fibre
FWHM	Full width half maximum
HD_n	Harmonic distortion of the n-th order
HV	High voltage
IEEE	Institute of Electrical and Electronics Engineers
IFFT	Inverse fast Fourier transform
IM	Intensity modulation
IM3	Third order intermodulation
IoT	Internet of Things
IR	Infrared
ISI	Intersymbol interference
LAN	Local area network
LD	Laser diode
LED	Light-emitting diode
LiDAR	Light detection and ranging
LiFi	Light fidelity
LOS	Line of sight
M2M	Machine-to-machine

µLEDs	Micro-LEDs
NRZ	Non-return-to-zero
OFDM	Orthogonal frequency division multiplexing
OFDMA	Orthogonal frequency division multiple access
O-OFDM	Optical OFDM
OOK	On-off keying
OR	Digital logical disjunction
OWC	Optical wireless communication
PAM	Pulse amplitude modulation
PAPR	Peak-to-average power ratio
PC	Personal computer
PCI	Peripheral Component Interconnect
PD	Photodiode
PDE	Photon detection efficiency
PDP	Photon detection probability
PF	Perfluorine
PIN	Positive-Intrinsic-Negative
PLL	Phase-locked loop
PMMA	Poly(methyl methacrylate)
PHY	Physical layer
Pk	Peak
PMF	Probability mass function
POF	Plastic optical fibre
PP	Power Penalty
PPM	Pulse position modulation
PQ	Passive quenching
PS	Pulse-shortener
PSK	Phase shift keying
PTID	Photon time information-based detection
QAM	Quadrature amplitude modulation
QL	Quantum Limit

RF	Radio frequency
RGB	Red Green Blue
RIN	Noise density specification of the emitter
RLS	Recursive least squares
RS	Reed-Solomon (coding)
RX	Receiver
RZ	Return-to-zero
SNDR	Signal-to-noise-and-distortion ratio
SNR	Signal-to-noise ratio
SPAD	Single photon avalanche diode
SST	Synchronous Summation Technique
TCSPC	Time-correlated single photon counting
TIA	Transimpedance amplifier
ToF	Time of flight
TX	Transmitter
µLED	Micro-LED
USB	Universal serial bus
UV	Ultraviolet
UWOC	Underwater wireless optical communications
VCSEL	Vertical-cavity surface-emitting laser
V2V	Vehicle-to-vehicle
V2I	Vehicle-to-infrastructure
VLC	Visible light communication
WiFi	Wireless network protocols for internet access
XOR	Exclusive OR logic

1. Introduction

1.1 Visible Light Communication

Radio-based wireless communications such as WiFi are ubiquitous in home and office local area networks (LANs) but use a confined portion of regulated RF spectrum [1]. It is forecast that there will be 5.3 billion internet users (around two-thirds of the global population) by 2023 (Fig. 1.1) [2]. Ever increasing demand, multiple users and neighbouring frequency bands put high demands on this limited RF bandwidth [3].

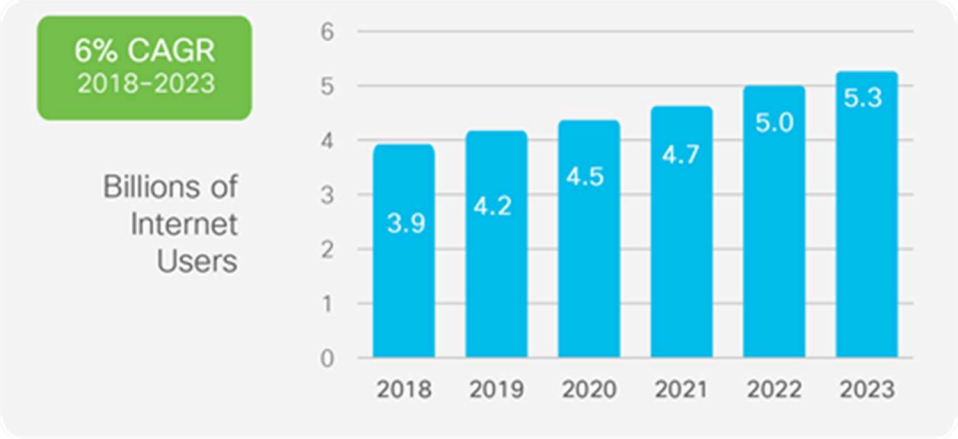

Figure 1.1: Global Internet user growth. 6% estimated compound annual growth rate (CAGR) [2].

Visible light communication (VLC) provides hundreds of terahertz of unused and unregulated bandwidth with the optical wavelengths of the electromagnetic spectrum [4]. The widespread growth of solid-state lighting makes LiFi and VLC viable to support RF networks. VLC offers sustainable, cost-effective and energy efficient systems as LED lighting infrastructure can be used mutually for illumination and communication purposes [5]. VLC can provide an additional layer of small cells to relieve RF channels from increasing demands on bandwidth [6]. Unlike radio waves, light does not propagate through walls, so VLC offers higher security than conventional RF communications since eavesdropping is very difficult. Another advantage of VLC is its immunity to electromagnetic interference [4]. VLC access point coverage aligns with the 5G and beyond direction towards greater cell density to provide higher throughput over more concentrated areas [7, 8]. This work focuses on VLC for these reasons.

1.2 Motivation

VLC has been proven advantageous in a number of applications including lighting, underwater communication, indoor localisation, and transmission in EMI-sensitive environments [3]. All of these use cases require high sensitivity at the RX because information is conveyed in the intensity of the visible light signal. There is ongoing research to close the gap between bandwidth limited VLC implementations and the 100 Mb/s to 100 Gb/s standard with present-day RF and optical fibre technologies [9, 10]. Recent reports attain high data rates by optimising the emitter at the TX to maximise the data rate [11-17]. GaN micro-LEDs (μLEDs) offer a −6 dB electrical bandwidth of over 100 MHz due to their small junction volume and high operating current density [17]. It has been demonstrated that VLC systems with μLED arrays can attain data rates of over 10 Gb/s across 5 m [17]. Off-the-shelf LEDs have been used with OFDM to maximise data rate to 15.73 Gb/s [18] at the forward error correction (FEC) requirement of 3.5×10^{-3} bit error rate (BER) [19]. However, μLEDs have low radiant flux and LEDs without optics have wide half-power semiangles (i.e. 65° in [20]), so link distance remains limited (0.3-5 m) [17-19]. These problems at the TX can be solved by launching an optical fibre laser wirelessly with fibre-wireless-fibre (FWF) tracking [21]. Mirror-based beam-steering was used in [21] to attain 1 Tb/s with ten IR (1546.12 nm to 1553.33 nm) wavelength division multiplexed 4-PAM signals – approaching the channel capacity of optical communication. Modern RF operates in a similar way, where narrow beams are steered to form high-rate line-of-sight links between terminals [22]. [21] also demonstrates bidirectional connection between fibre transceivers with a wide indoor coverage of 2.54 m^2 and [23] raises the FWF data rate to 100 Tb/s, but only with a fixed 1 m point-to-point link. All of these studies highlight that receiver technology lags significantly behind the capability of the TX.

VLC systems make use of PD receivers which require a transimpedance amplifier (TIA) in the receiver chain [10-21, 23]. This introduces thermal noise dominant over the intrinsic shot noise of photons and raises the noise floor, which impairs the sensitivity of the receiver and limits the signal-to-noise ratio (SNR). Although PDs are suitable for fibre communication, they are not ideal for low optical power and long-distance wireless transmission due to this limited sensitivity [24-26]. Received signal power is proportional to the effective light collection area of the detector, which implies that a large area must be used to attain high data rates [26]. However, this is limited because a large collection area increases the device capacitance, which reduces the bandwidth of the receiver. This makes the detector at the receiver front-end the main bottleneck on the performance of the system [27].

SPADs are biased above breakdown and, through rapid carrier impact ionisation, provide a fast-rising pulse when a photon is detected [28, 29]. Therefore, SPADs offer a means to enhance sensitivity via a high internal gain whilst maintaining a small active area. This would benefit portable devices which require high receiver sensitivity to sustain signal levels and low power consumption due to battery limitations [26]. A SPAD RX could operate in assistance with existing APD or PD RXs with the potential to extend range or maintain connectivity in extreme conditions such as dark or diffuse environments [30]. However, a SPAD must be deactivated after a photon is detected to avoid damage from the avalanche process. During this dead time (ranging from 3.5 ns [28] to 50 ns [31]), the SPAD is unable to detect subsequent photons. Therefore, dead time reduces the SNR and bandwidth of a SPAD RX. This work studies the design, simulation, analysis, and experimentation of SPAD RXs for high-sensitivity optical communication. Recent papers report commercial silicon SPAD photomultipliers [31] can support bit rates of up to 3.45 Gb/s with a BER of 10^{-3} [32, 33]. However, these devices combine SPAD pulses into a single analogue signal and require an external electrical amplifier at the output before being captured on an oscilloscope and decoded. Opportunities remain to improve the sensitivity, data rate and native BER and to implement a digital integrated SPAD RX which fully exploits discrete photon counts that can be conveyed directly to DSP and decoded to recover data. Since the information-carrying signal is analogue, methods to sample high speed optical OFDM signals have been simulated in relation to discrete SPAD Geiger counts in [34] and [35]. This work aims to implement a practical SPAD-based OFDM system (Chapter 5, Section 5.4.3).

The probability mass functions (PMFs) of photon detections for SPADs with dead time have been obtained [36], but these models assume that the SPAD is always ready at the start of each symbol to simplify the derivation. This means that intersymbol interference (ISI) from detections in preceding symbols is not considered in these models. SPAD RXs are usually operated under the condition that the symbol period is longer than or approximately equal to the dead time [28, 29, 37, 38]. For high-speed transmission, where the symbol period is shorter than the dead time of individual SPAD detectors, decision-feedback (DFE) equalizers [32, 33, 39, 40], initially designed for linear channel electrical systems [41], have been used. SPAD-based channels are inherently nonlinear and detection techniques have been developed to alleviate ISI caused by dead time [42, 43]. However, prior to the work in this thesis, the resulting power penalty due to dead time and other SPAD constraints has not been quantified.

In this work, the physical parameters of SPADs are translated into a model of a link with a SPAD detector array and data-carrying optical signal. This includes a precise model of all noise sources and system nonidealities such as dead time. Using the model, it should be possible to predict sensitivity, maximum data rate and quantitatively analyse the impact of dead time. The outcome will be a configurable model which can be used to optimise the performance of SPAD RXs for specific applications. Such a model would serve as a quick, effective, and inexpensive method to design RXs in future work.

1.3 Objectives

The aim of this study is to advance understanding of the performance limits of a SPAD RX, to ultimately reach the quantum sensitivity limit. This thesis introduces a SPAD model to mobile VLC systems which are subject to random blockages and rapidly changing SNR conditions with the aim to ensure connectivity when line-of-sight links are interrupted. A detailed analysis of the main characteristics of a SPAD-based RX is presented. Theoretical equations are developed to gain understanding of the effect of dead time. The model assumes a laser diode (LD) source with >1 GHz bandwidth to ensure that only the SPAD RX response is the limiting factor on bandwidth and error performance. Furthermore, the VLC link is simplified to line-of-sight with precise alignment as mobility is not important for SPADs at this stage of research. A single SPAD is considered at first and then expanded to a 64 × 64 array with the objective to accurately predict SNDR. Analytical modelling and simulations are compared to experimental measurements with an integrated device to evaluate the performance of a SPAD RX in a VLC link.

1.4 Contributions to Knowledge

Nonlinear theoretical and numerical models are created in this study to describe the physical properties of SPADs. This enables analysis of the influence of factors such as dead time on the performance of a SPAD-based system. A transient simulation is developed in Matlab to generate and decode an OOK, PAM or OFDM frame propagated through a SPAD-based system model. The model describes the various stages in the signal path including the TX emitter response, a wireless optical channel with ambient light and a SPAD RX, before the frame is recovered and decoded. The transient response to a data-carrying optical signal is investigated and used to estimate the steady-state behaviour of SPADs in an array. It is shown that SPAD transient behaviour narrows the separation between symbol count distributions, resulting in more erroneous bits. Power penalties

due to source extinction ratio, path loss, photon detection losses, fill factor, dead time, SPAD afterpulsing and array crosstalk parasitics and photon count summation are estimated.

The 130 nm CMOS SPAD RX in this work achieves 500 Mb/s in a VLC link within 15.2 dB of the quantum limit and demonstrates the fastest and most sensitive optical RX to-date [44]. This proves higher order four-level pulse amplitude modulation (4-PAM) and adaptive OFDM up to 32-quadrature amplitude modulation (32-QAM) is possible with photon counting RXs. A SPAD RX compliant with high speed OFDM used in IEEE 802.11 standards [45] is an area of research that had not been published until this work. This SPAD RX is the first to demonstrate present-day physical layer (PHY) transmission rates around 100 Mb/s to 500 Mb/s with a minimum incident power comparable to the equivalent radiation sensitivity of modern WiFi RXs [46]. The SPAD RX is around 10× (10 dB) more sensitive than existing APD RXs and 100× (20 dB) more sensitive than conventional PIN RXs. Experiments were conducted in a laboratory bench setup emulating a VLC indoor environment. The system is optimised for sensitivity and throughput by precise optical link alignment as well as extensive optimisation of emitter DC bias and AC amplitude, modulation format, aperture size, bias voltage and RX dynamic range to ensure the received photon counts are minimised. These results were presented at the International Solid-State Circuits Conference in 2019 [44]. The 40 nm CMOS SPAD array chip [47] used in this study allows for testability and access to individual SPADs. As an RX, this IC achieves linear proportionality between the number of SPAD detectors and the data rate. Data was acquired and analysed from the SPAD RXs, which were designed by Sarrah Patanwala and Prof Robert Henderson and members of the CMOS Sensors and Systems Group in the Institute for Integrated Micro and Nano Systems (IMNS) mentioned in the acknowledgements. SNDR and BER is estimated and close matching with measurements validates the effectiveness of the simulation algorithm and the analytical model.

1.5 Outline of Chapters

Chapter 2: A literature review which discusses optical communication and the performance of detectors. The choice of detector and receiver operation is reviewed with the aim to improve sensitivity and system throughput. Avalanche multiplication gain is introduced and APDs and SPADs from prior literature are studied in detail. The advantages and disadvantages of a SPAD receiver front-end are discussed with respect to noise, photosensitivity and nonlinearity constraints.

Chapter 3: Development of theoretical, analytical and simulation models with references to the literature. This mathematical framework improves understanding of the performance limits of a SPAD RX and enables specification of optimal design parameters. The model is expanded and used to delineate post-processing methods to improve performance in a communication link.

Chapter 4: Describes the design of a SPAD IC used to characterise the dead time and other physical parameters of a single SPAD detector and an integrated SPAD array RX.

Chapter 5: Describes measurements and the performance of each IC introduced in Chapter 4 as a SPAD RX. The 130 nm SPAD-based RX demonstrates high speed signal reception at up to 500 Mb/s with photon counting sensitivity. The results are compared to the model.

Chapter 6: Summarises the results and gives insight into further work that could improve the performance of a SPAD RX. Specific applications for SPAD-based receivers and continued research to correct power, area and error rate limitations are suggested.

2. Literature Review

2.1 Introduction

There is growing demand from mobile device users for broadband wireless services with high-speed internet access and high-quality multimedia [2]. User communication needs are addressed by technologies such as copper-based wireline, coaxial and optical fibre cables; fibre-to-home and broadband RF wireless technologies [48-55]. However, radio spectrum is heavily congested [1]. This chapter reviews optical communication as a solution to this limitation [3-5] and its application to wireless LANs for homes and offices.

Wireless channels require wide coverage and device mobility for users. This means that the link is constantly changing due to reflections and attenuation from nearby objects in the surrounding environment. Low transmit power, long distances, and adverse channel conditions can all reduce the received signal amplitude, so the sensitivity of the RX should be high. This review covers detectors in recent literature with the objective of improving sensitivity and two architectures are considered for this work: a single-sample RX and an integrating RX.

Single-photon avalanche diodes (SPADs) are introduced to advance the detection process towards that of an ideal receiver. In addition, SPADs are appropriate for digital integration in CMOS and the state-of-the-art is reviewed [56-59]. The physical behaviour of SPADs is discussed to form the basis of a model to predict performance (Chapter 3). This study focuses on a receiver solution optimised for detecting the fast 450 nm blue component emitted by white LEDs and other solid-state sources such as laser diodes (LDs) [3-5]. VLC is chosen in this work as the most commercially viable path of OWC due to the widespread usage of LED lighting. In addition, SPADs are most sensitive in the visible wavelength range [56]. Nevertheless, analysis is kept broad because SPAD RXs have potential in several different optical wireless and fibre applications.

2.2 Emerging VLC Technology

III-nitride laser diodes (LDs) are promising sources for light fidelity (LiFi) networks (Fig. 2.1), OWC, underwater wireless optical communications (UWOC), and plastic optical fibre (POF) communications [60] due to their high modulation bandwidths (>1 GHz) compared to LEDs (<20 MHz). Their narrow linewidths also enable robust

free-space Gb/s LiFi links in the presence of high intensities of sunlight through selective spectral filtering. Therefore, the main bottleneck on the throughput of the system is placed at the receiver front end. Fully integrated CMOS VLC receivers have recently been developed to enable miniaturised and low-cost Gb/s LD-based links [24]. However, the sensitivity of these devices is constrained by electrical noise sources such as thermal or excess noise related to the employment of PIN PDs or linear APDs and their amplification circuits [11-17, 24, 25]. This inhibits high speed PINs and APDs from being integrated into portable consumer devices which require low power consumption to preserve battery life. The share of machine-to-machine (M2M) connections is projected to grow to 50% of global connected devices by 2023 (14.7 billion connections) [2].

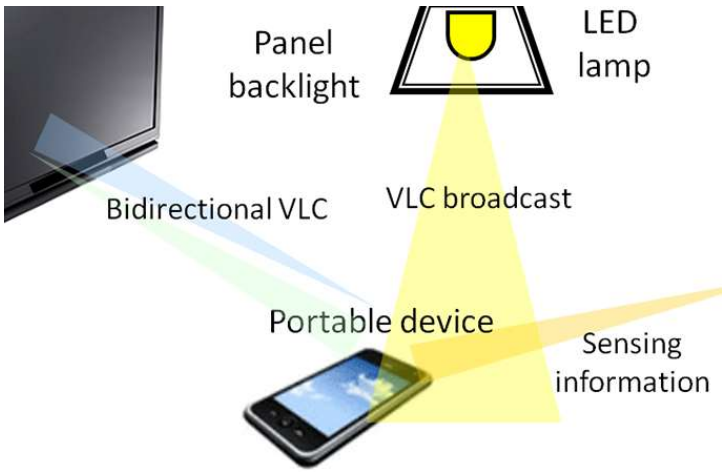

Figure 2.1: VLC and LiFi: LED lighting has dual purpose as internet access points. M2M communication and low power IoT [61].

LEDs are already common in traffic lights and are increasingly being used for automotive lighting, presenting viable prospects for vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication (Fig. 2.2) [62]. Current VLC systems with PDs have ranges of less than 20 metres [11-17, 24, 25], so are inadequate for applications with high path loss [63, 64]. The performance of vehicular VLC systems is severely degraded by adverse weather conditions such as fog [66, 67]. It has been demonstrated that SPAD RXs could solve this challenge and significantly extend the link distance [68]. The RX could be combined with recent LiDAR technology, which already uses arrays of SPADs to construct a 3D image of the surrounding environment [47, 69-73], to enhance vehicle and roadside interaction and transparency. However, it would be essential for a V2V VLC RX to be operational whilst exposed to different outdoor conditions – at night through to broad daylight (100

klx ambient illuminance [74]). This means that the performance of SPADs in these environments is limited without narrow optical filtering at the RX [75].

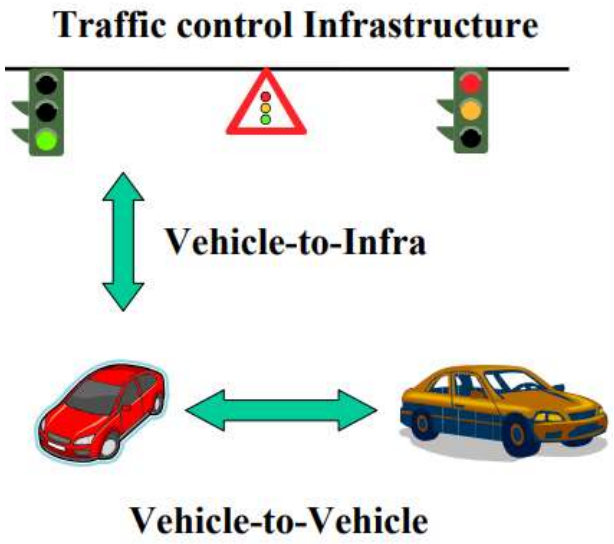

Figure 2.2: V2V and V2I wireless communication transferring velocity and position information for safety [65].

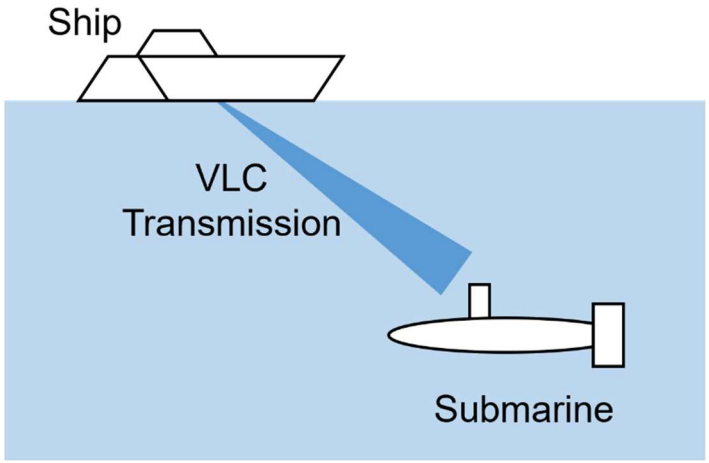

Figure 2.3: UWOC has superior latency to RF underwater but is subject to high scattering and attenuation.

RF is impractical for underwater communication because there is high signal attenuation due to seawater's conductivity [76] and acoustic waves can disturb marine life [77]. VLC has superior data rates and latency underwater because of the faster propagation of light [76] and it has been shown that SPAD RXs allow much longer ranges in these circumstances [77]. Ambient light is usually abundant in underwater environments, but it has been demonstrated that even a small number of transmitted photons reaching a SPAD array detector can ensure reliable communication [77, 78]. An UWOC system could also be used for position reckoning between

divers or submersible vehicles, for example (Fig. 2.3). This would be beneficial for exploration, environmental conservation, commercial offshore diving, and industrial and scientific work [76].

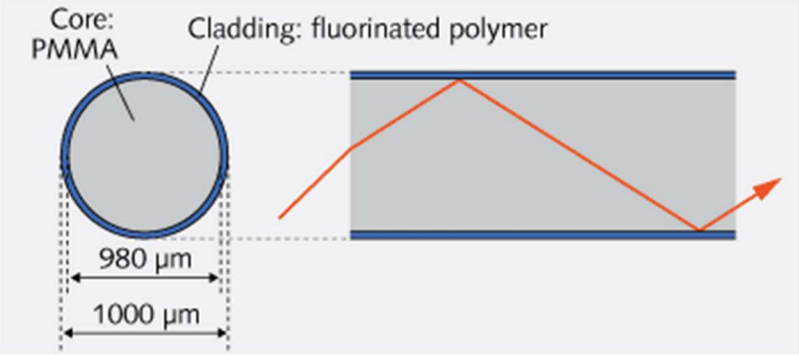

Figure 2.4: POF offers flexible and cost-effective last-mile wireline networking links and can use visible wavelengths [79].

Solid-state drives, multi-core processors, and new standards such as PCI Express 6.0 (64 Gb/s) [80] and USB 3.2 Gen 2×2 SuperSpeed+ (20 Gb/s) [81] raise rates in PCs, tablets, and smartphones to the multi-Gb/s range. Most electrical wireline technologies used in homes and offices are copper-based coaxial cables and twisted-pair cables [48-55, 82] which have a limited bandwidth due to the skin effect and dielectric losses [53]. In addition, twisted-pair cables are susceptible to electromagnetic interference (EMI) that increases bit errors, and coaxial cables are heavy, rigid, and costly to install [83]. Glass optical fibres, such as single-mode and multi-mode silica fibres, deliver data rates of >100 Gb/s over inter-continental links (>100 km between repeaters) [48] with very low attenuation (0.2 to 0.5 dB/km) and gigabit communication in networking applications such as data centres [85]. Glass fibres are immune to EMI and attain an efficient energy per bit [82] and low error rates (10^{-12} BER) [83]. However, glass fibres have small core diameters of around 10 µm [86], which means that accurate alignment is required for installation at the system level.

POF uses materials such as polymethylmethacrylate (PMMA) and low-dispersive perfluorinated (PF) polymers which are more robust under bending and stretching than silica glass fibre [86, 87]. Fig. 2.4 shows a cross-section of a POF. The wide 1 mm core diameter of POF allow inexpensive injection-moulding technologies to be used for connectors and transceiver packaging that are easy to install [84]. Furthermore, POFs are suitable for continuous mass production because they can be fabricated by the coextrusion process [85]. These advantages make POF a favourable solution for low-cost short-haul home and office wiring media to meet the

increasing demand in high-definition video streaming, cloud computing and file sharing [87]. 120 Gb/s was demonstrated across a 50 m cable with twelve POFs in [87] – suitable to support uncompressed 4K/8K video transfer – and 40 Gb/s was demonstrated across 100 m [88, 89]. PF polymers have lower material dispersion than that of silica due to their lower wavelength dependence of refractive indices in visible to near-infrared wavelengths [81]. Therefore, a PF POF can attain a higher bandwidth than a multimode silica glass fibre in this region. Furthermore, POF has potential applications in industrial control [88]. High-sensitivity RXs with large PINs have been used to counteract the high attenuation of POF (0.14 dB/m at 650 nm) and attain a link distance up to 50 m [90-93]. Whilst glass fibre operates from 810 nm to 1610 nm, POF uses visible light – POF has lowest attenuation in the green wavelengths and this aligns with the sensitivity peak of a typical shallow junction Si SPAD [57]. This makes SPADs promising to extend the reach of POF beyond 100 m, and there is ongoing work in applying SPADs to POF systems [94].

2.3 Detectors

2.3.1 Photodiodes

PDs are based on either p-n or p-i-n (PIN) junctions [95, 96]. The p-n structure uses adjoining n (donor) and p (acceptor) doped silicon to create a thin depletion region (Fig. 2.5). The depletion region is expanded in a PIN PD with a thin interfacing region of intrinsic silicon to form a p-i-n junction (Fig. 2.6). This allows the sensitivity and bandwidth to be adjusted [96]. A PD operates in reverse bias, which widens the depletion region and increases the electric field strength across the junction [95]. Photons are absorbed in the junction at a depth dependent on the wavelength of the incident light. If the photon energy is greater than the bandgap, an electron is elevated into the conduction band. This generates an electron-hole pair which is then separated and moved by the electric field, inducing a photocurrent.

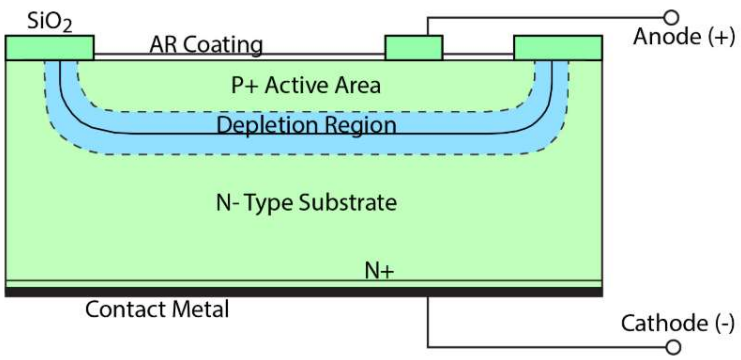

Figure 2.5: PN junction PD cross section [97].

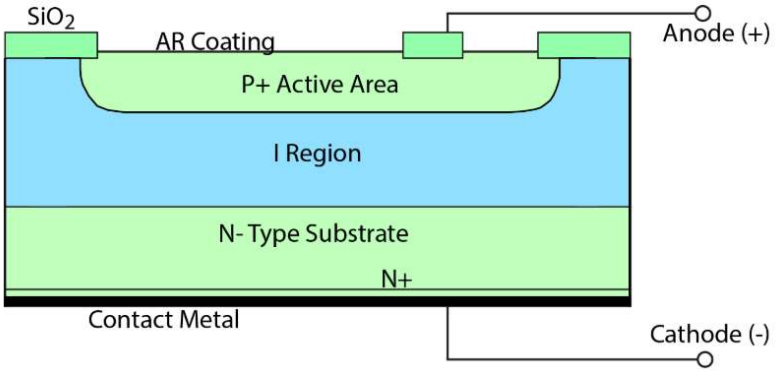

Figure 2.6: PIN photodiode cross section. A PIN has a faster response speed than a PN PD from its increased depletion region [97].

The sensitivity of the PD is limited by the quantum efficiency – the number of electron-hole pairs created for each absorbed photon. A PD is only sensitive to photons that are absorbed near the depletion region – photons absorbed above or below the region do not contribute to the photocurrent. Sensitivity is maximised if the depletion region depth matches the absorption depth or if the region covers a large proportion of the absorption curve [98].

The modulation bandwidth of silicon PDs is mainly limited by the extrinsic bandwidth formed by the diode capacitance and resistance [99]. An intrinsic bandwidth limitation is caused by carrier diffusion from regions below the PD, which can be alleviated with silicon-on-insulator designs [100]. The bandwidth can also be limited by the minority carrier lifetime (the time it takes for carriers to recombine) and the transit time of a carrier over the depletion region. This motivates thin regions and small areas in high speed devices [99]. There is a trade-off between sensitivity and speed because a wide depletion region is necessary to capture a large number of photons, but the bandwidth becomes limited by the transit time. Similarly, large collection areas are capacitance limited.

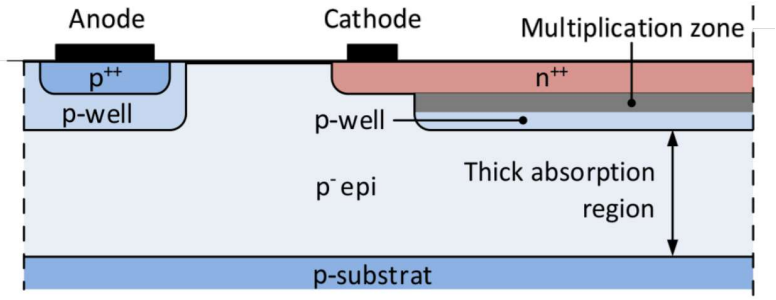

Figure 2.7: Cross section of an APD in pin-PD (Bi)CMOS technology (isolation and passivation stack not shown) [106].

2.3.2 APDs

Recent literature [101-103] and commercial products [104, 105] have developed detectors to solve this sensitivity problem through avalanche multiplication, whilst maintaining high bandwidth [101]. An APD shifts amplification to optical conversion at the diode, which provides an extra gain (of around 100) without additional electrical circuitry. This is achieved by applying a high reverse bias (typically 100–200 V in Si) to increase the electric field in the junction (Fig. 2.7) [106]. The field accelerates drifting charge carriers to a high kinetic energy that makes them ionise lattice atoms they collide with by knocking electrons out of bonds. This effect is known as impact ionisation and creates more electron-hole pairs which are also accelerated by the field, causing further ionisation with exponential growth. APDs therefore produce a large photocurrent which is proportional to the intensity of the incident light. However, the APD current still requires further amplification and conversion into a voltage via a TIA, which introduces thermal noise to the output signal [101-103].

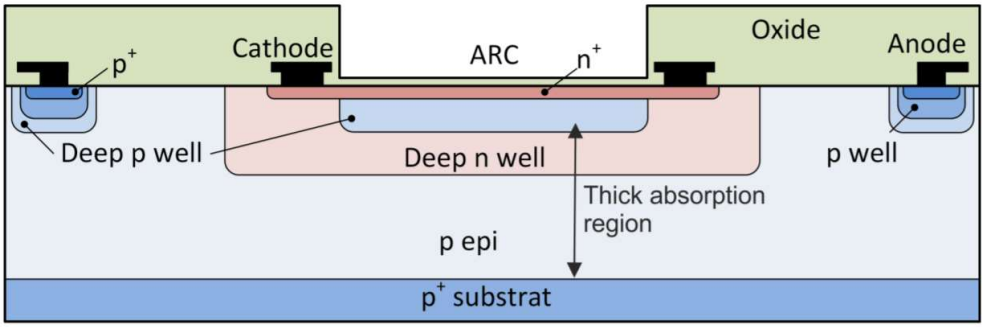

Figure 2.8: Cross section of an APD in high voltage (HV) CMOS technology [106].

Some Si APDs employ alternative doping techniques that allow higher voltages to be applied (> 1500 V), before breakdown is reached, to increase gain to >1000 [106]. APDs achieve higher gain with a higher reverse bias (Fig. 2.9) [106] but stay below the breakdown voltage to maintain linear gain and avoid thermal damage. Fast APD RXs have improved sensitivities of –38 dBm at 280 Mb/s (850 nm) [103], –31.8 dBm at 1 Gb/s (675 nm) [102] and –34.6 dBm at 1 Gb/s (675 nm) [101] but remain two orders of magnitude above the standard quantum limit (QL) determined by photon shot noise [38]. This limit defines the minimum number of photon detections required to ensure a given bit error ratio (BER). BER is a measure of performance and is given by the ratio of erroneous bits at the receiver to the total number of bits sent. An ideal receiver assumes that there is no noise apart from the intrinsic uncertainty of quantised photon arrivals. Such a system could detect and decode a modulated signal within an error probability set by Poisson statistics. The QL characterises the theoretical

minimum optical power required for a given BER and is set by the data rate, modulation scheme and the wavelength of the incident light [37]. The quantum limit is useful for comparing the performance of optical RXs.

2.4 Single-Photon Avalanche Diodes (SPADs)

2.4.1 Operation

SPADs are biased above breakdown in the Geiger region of operation (Fig. 2.9). Due to the very strong electric field in the junction, the absorption of a photon triggers billions of electron-hole pair generations through impact ionisation [107]. This leads to an avalanche multiplication of current with a high internal gain of over 10^6, which produces a digital pulse for an individual photon detection. Additive thermal noise in the RX chain is therefore eliminated since no TIA is needed. This allows quantum sensitivity limits to be approached [28, 37].

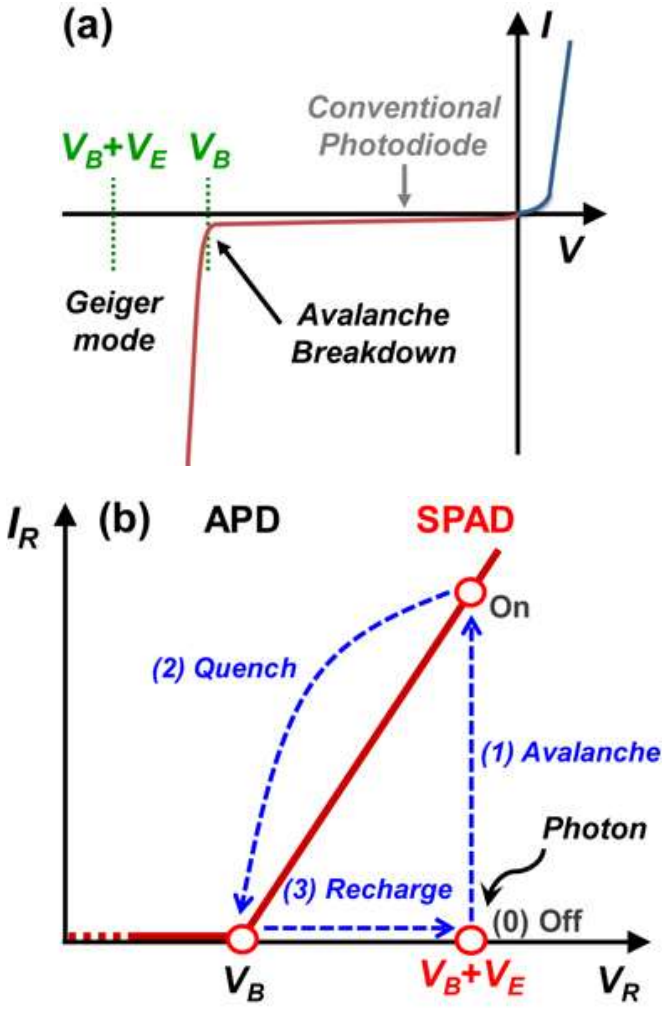

Figure 2.9: (a) Current-voltage characteristics of a p–n junction diode and the three modes of operation left to right: Geiger mode with excess bias V_E above the breakdown voltage V_B, proportional APD mode around V_B,

and conventional unity-gain PD mode at low values of reverse voltage. (b) SPAD operation in reverse I–V (I_R and V_R) characteristics [56].

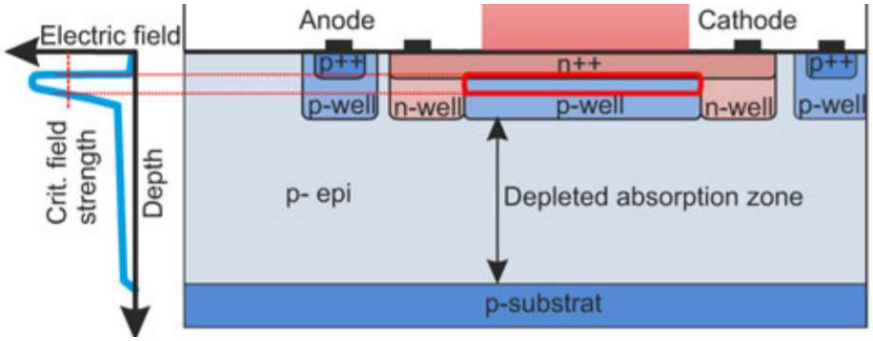

Figure 2.10: Cross section of the structure of a (red-sensitive) SPAD and electric field distribution [28].

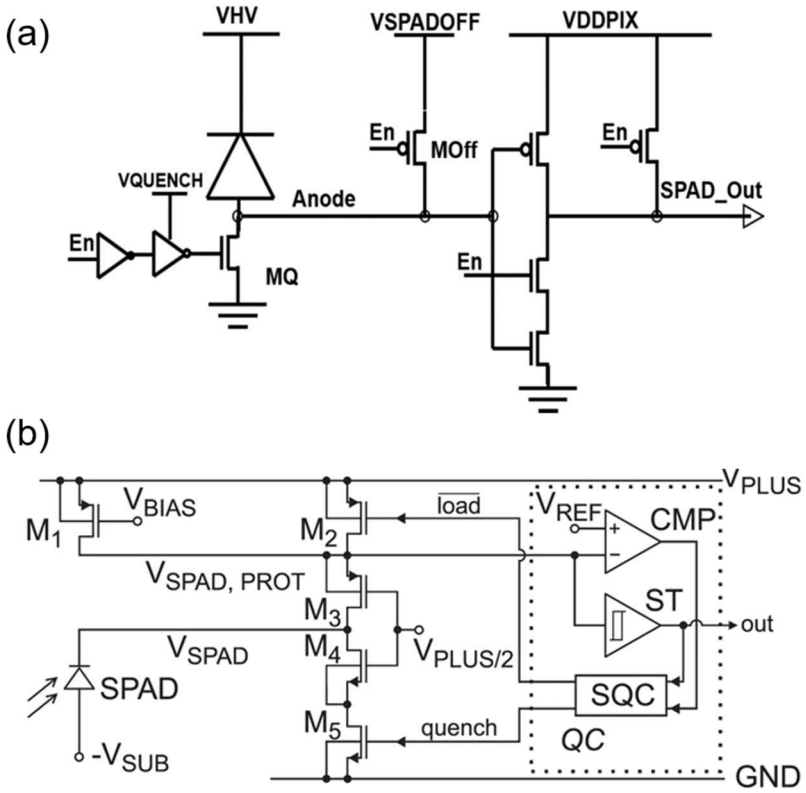

Figure 2.11: Schematic of (a) a PQ and disabling circuit [109] and (b) an AQ circuit [37].

2.4.2 Dead Time

The avalanche process is self-sustaining and must be quenched in order to prevent the SPAD from being thermally damaged. A passive quenching (PQ) or active quenching (AQ) circuit is used to stop the avalanche by temporarily dropping the bias below breakdown V_{br} [108]. The SPAD is then recharged back to above the breakdown voltage ($V_{eb} = V_{br} + V_E$) to be ready for the next photon arrival. V_E is the overvoltage above

breakdown. Fig. 2.11 shows diagrams of a PQ circuit and an AQ circuit. The detection, quenching and recharge process is shown below in Fig. 2.12.

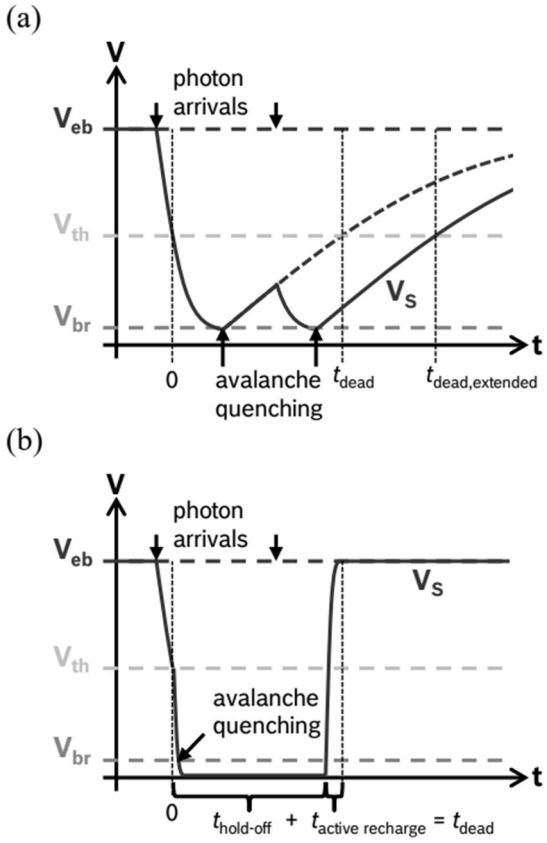

Figure 2.12: SPAD output voltage V_S and dead time t_{dead} (a) PQ: another photon arrival before the SPAD is fully recharged over the threshold voltage V_{th} can extend t_{dead} to $t_{dead,extended}$. (b) AQ: the SPAD is held below the breakdown voltage V_{br} for hold-off time $t_{hold-off}$ and the fast recharge $t_{active\ recharge}$ back to the excess bias V_{eb} ensures t_{dead} is fixed, at the expense of larger circuit area [108].

AQ requires more circuitry and has higher power consumption than PQ (over 4 times higher [28]); this goes against the principal advantage of an ideal photon counting RX. Therefore, PQ is selected in this work to maximise the overall system power efficiency – both at the emitter and the electrical consumption at the RX. Higher V_E gives higher sensitivity, but the high voltage required to bias a SPAD above breakdown increases power consumption. With CMOS processes, V_{br} can be reduced to 12-15 V [57], so power consumption can be minimised. In addition, a SPAD is direct-to-digital, so power consumption can be reduced by offloading RX complexity away from the analogue front-end and into scalable DSP instead. This also enables time gating to be synthesised in digital, where SPADs can be disabled to avoid unnecessary dead time losses [56]. This thesis

focuses on determining the optimum signal power region where PQ SPADs should operate to enhance performance. The success of PQ SPADs relies on their simplicity.

2.4.3 Photon Detection Probability (PDP)

PDP is the likelihood of an avalanche being initiated when a photon reaches a SPAD and is dependent on the probability of a carrier being generated (the quantum efficiency). PDP is an important parameter for characterising device sensitivity as it is a function of the wavelength of the incident light [110]. This means that the centre wavelength of an emitter should be chosen to match the peak of the SPAD spectral response to minimise detection losses [29]. CMOS SPADs have a reduced PDP in IR and near ultraviolet (UV), so are mainly effective in the visible light spectrum (Fig. 2.13) [57]. Fig. 2.14 shows the PDE response of a commercial SPAD.

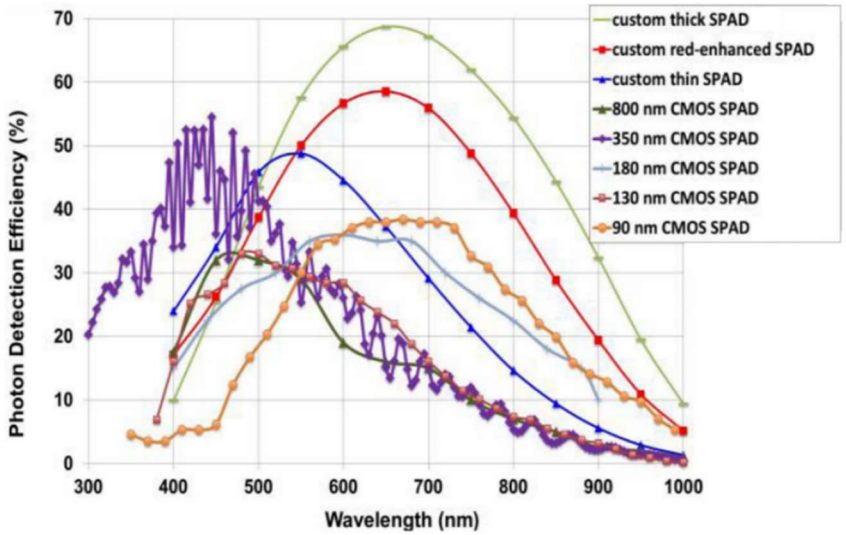

Figure 2.13: PDE against wavelength for various SPAD devices [57].

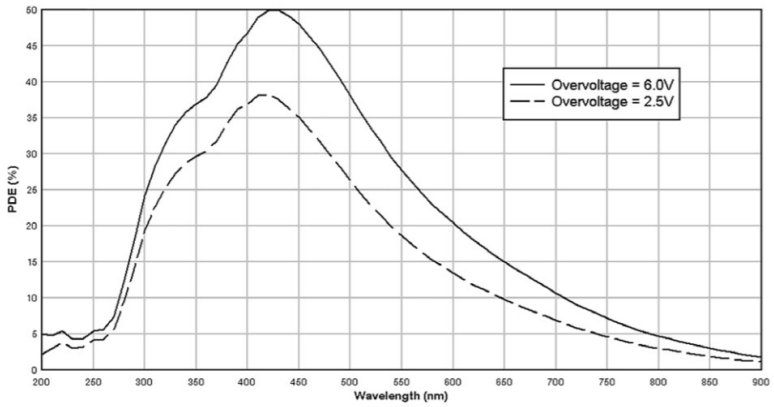

Figure 2.14: PDE response of a commercial SPAD manufactured by ON Semiconductor [31].

PDP is directly proportional to the electric field strength and can be increased if the excess bias is increased further above the breakdown voltage. However, a higher field may in fact impair SPAD performance [38] as it

increases dark counts. The overall probability of detection is the product of the PDP and fill factor of the receiver. The fill factor is the proportion of the light-sensitive surface area of the detector relative to its total area. Some SPAD arrays have low fill factors (around 2.5 %) due to guard rings and circuitry adjoining each pixel [29]. This results in low detection efficiency, as the majority of incident photons hit the inactive surfaces between pixels.

2.4.4 Dark Count Rate

Thermally generated carriers within a SPAD can trigger avalanches even in the absence of illumination [110]. The dark count rate (DCR) is the rate of counts due to these avalanches and is always present as an internal noise characteristic of SPADs. Dark counts are modelled by a Poisson process. False detection events can be subdued by cooling the device or, to some extent, reducing its active area [110]. DCR is also dependent on voltage as increasing the excess bias further above breakdown increases the electric field strength, resulting in a higher number of dark counts due to impact ionisation.

2.4.5 Afterpulsing

Afterpulsing is a second-order effect which is correlated in time with photon arrivals and occurs when carriers become trapped in energy levels between the valence and conduction bands during an avalanche [111]. These carriers are later released and can trigger a spurious avalanche. Release lifetimes are usually in the order of nanoseconds. By ensuring that carriers are released without avalanche, the afterpulsing probability can be reduced if the dead time is made long enough. However, short dead times are necessary to maximise the count rate, so afterpulsing is often unavoidable and can cause inter symbol interference (ISI) due to the dispersion of pulses over time (in a similar fashion to multipath dispersion). Afterpulsing must be taken into consideration if the symbol period is comparable to the dead time. This presents a problem for SPAD-based communications at high transmission speeds. The capability of OFDM in mitigating this effect will be investigated in this project.

2.4.6 Crosstalk

SPAD arrays can be used to improve the throughput of applications [25, 28, 29, 32, 33, 37, 38]. However, these arrays are susceptible to non-uniformities; namely crosstalk between adjacent pixels in the array. Crosstalk occurs when an avalanche in a SPAD triggers a spurious avalanche in another nearby [111]. This introduces additive variabilities in the output photon counts, which can cause errors during demodulation in a SPAD-based communication system. Furthermore, crosstalk can impose a constraint on the dynamic range of the array.

Electrical crosstalk is caused by interference from substrate or supply noise, due to stray capacitive coupling between SPADs. This effect can be practically nullified through the careful layout of supply lines and by taking measures to reject substrate noise [111]. Performance can also be hindered by optical crosstalk. This occurs when an incident photon triggers an avalanche that causes several additional photons to be emitted in the SPAD pixel due to impact ionisation electrons in the avalanche current. These photons can propagate through the bulk and may be absorbed in the active region of a neighbouring pixel, triggering an inadvertent secondary avalanche there. Optical crosstalk is dependent on the pixel pitch and the number of carriers involved in an avalanche.

2.4.7 Combination Techniques

Effective detection rate is dependent on the dead time caused by quenching and recharging a SPAD after it detects a photon. During this time (3.5 ns reported in [37], 12 ns – this work, 10.6 ns [29] and 5 ns [25]), the SPAD is unable to respond to subsequent incident photons. Hence, it is unlikely that more than one photon detection will occur in a symbol period at rates higher than 100 Mb/s, where PD/APD RXs readily operate. Therefore, a single SPAD detector is unable to recover signals above this range with reliable BER – so multiple SPADs are required. Fig. 2.15 shows a representative block diagram of a photon counting RX with this principle compared to a typical PD/APD RX. This presents design challenges to combine, count and sample the outputs of the SPADs with minimal losses and circuit area. [25] accomplishes this with an analogue silicon photomultiplier (ASiPM) of 60 SPAD currents combined via a node connected to a load resistor. Commercial ASiPMs (Fig. 2.16) available on the market comprise of very large arrays of thousands of PQ SPADs (for example, 5,676 elements with 45 ns recharge time each, and 14,410 elements with 15 ns recharge time each) [31] but are currently not optimised for communication applications. [32] proves gigabit OOK communication can be supported with off-the-shelf ASiPMs. However, these analogue summation architectures are inefficient because in a full practical system, an ADC is still required to convert the output voltage into digital data bits. Other RX architectures are fully digital and permit parallel photon event summation to be achieved [112]. These are termed digital SiPMs (DSiPMs).

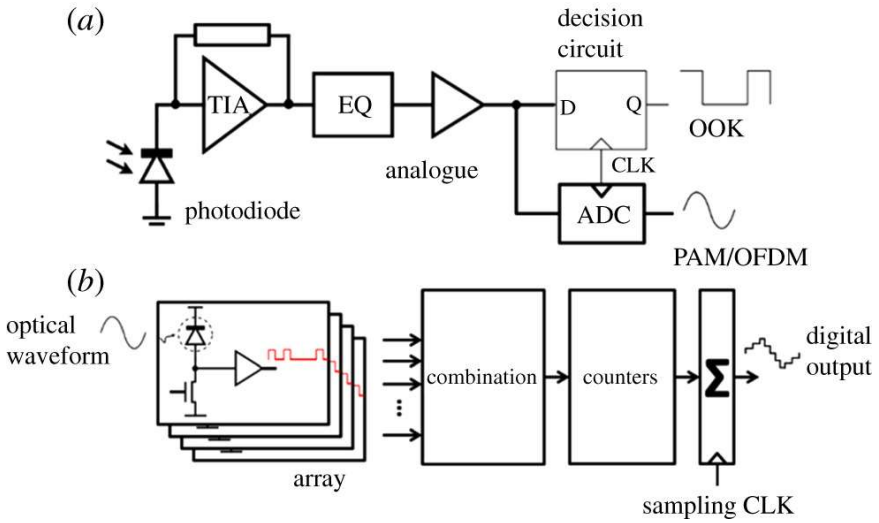

Figure 2.15: Block diagram of a conventional PD RX (top) and a photon counting RX with summed SPAD array (bottom).

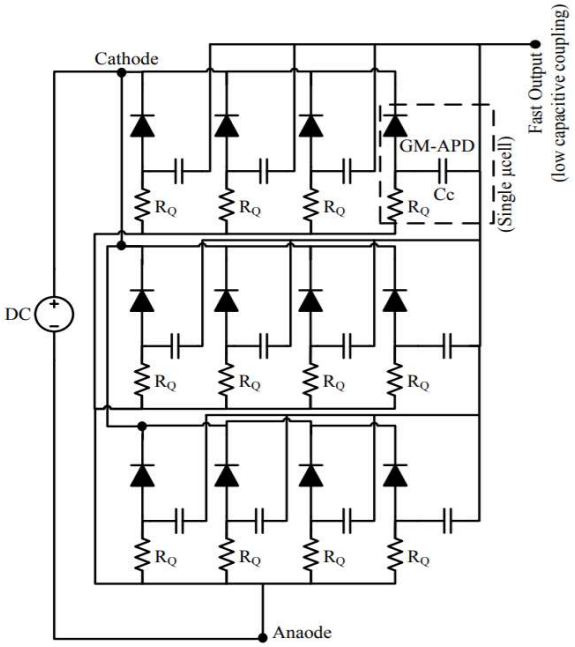

Figure 2.16: Schematic of an ASiPM, with capacitors coupling each microcell to a common output [113].

Below, Fig. 2.17 shows the differences between ASiPMs (a) and DSiPMs (b). Fig. 2.18 shows the combination logic options for DSiPMs. Losses occur in the combining logic due to finite output pulse width. Toggling XOR trees achieve the lowest losses of the three combination techniques [112].

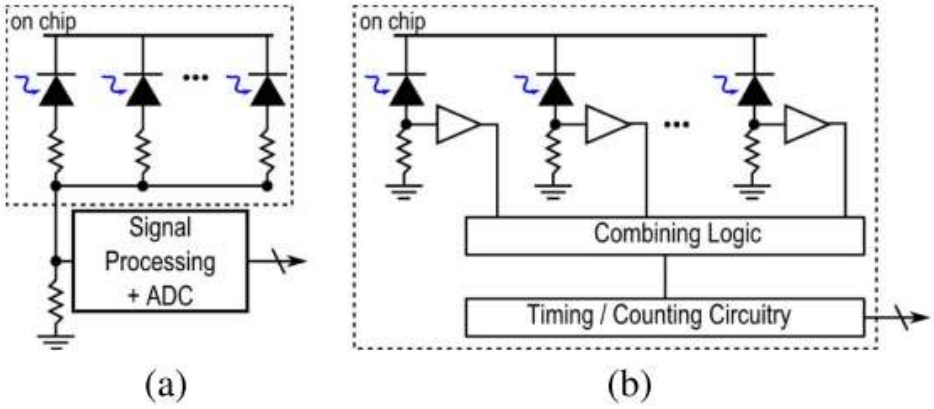

Figure 2.17: The dead time means that SPADs must be combined into arrays to enhance the dynamic range of the sensor. The outputs can be combined in an (a) analogue ASiPM or (b) digital DSiPM [112].

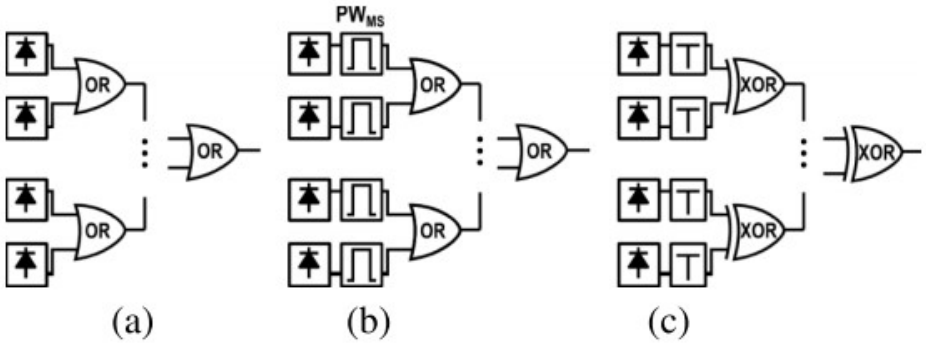

Figure 2.18: Digital combination logic networks—SPAD output pulses can be combined into a single channel through an (a) OR tree, (b) monostable pulse shaper PWMS + OR tree, or (c) toggle + XOR tree [112].

After the photon events are combined, they must be counted and summed. This can be implemented with digital counters and adders in a DSiPM. The total number of photons counted in each clock sample form the instantaneous amplitude of a photon-time signal. This is equivalent to integrating the received photons in each sample. The detection of baseband signals involves translating each received symbol into a single numerical value and comparing this value with a reference (or, for multi-symbol modulation, a set of references) to infer which symbol was transmitted. This is implemented by sampling the aggregate signal at a single point and taking this value to one or more comparators [48]. For equiprobable binary symbols (zero and one) expressed by two signal levels, the mid-point between them is a suitable reference assuming an AWGN channel. This will be different for a photon counting system. There are two ways to reduce a received symbol to a single value: single-sample direct detection or integrating each symbol.

2.5 Receiver Architectures

2.5.1 Direct Detection

The photocurrent produced by a conventional PD is small due to finite quantum efficiencies [96], small active areas and the limited optical power of the incident modulated signal [95]. Therefore, a TIA is used to amplify and convert the photocurrent into a voltage suitable for signal conditioning and processing [95]. The signal amplitude and bandwidth trade-off leads to the use of small, high speed PDs, large amplification circuits and high incident light levels at the expense of sensitivity [97]. Techniques such as differential APDs [102] or analogue equalization (EQ) [98] are implemented at analogue front ends to reduce noise and improve bandwidth.

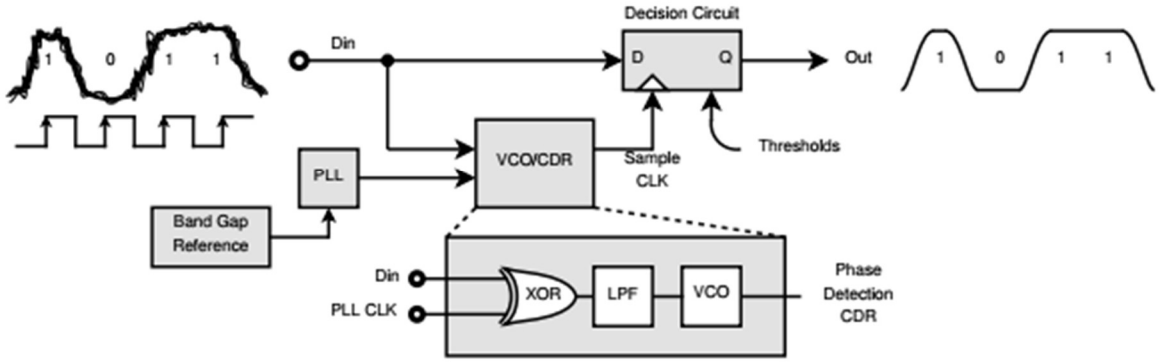

Figure 2.19: CDR circuit. A flip-flop takes a decision between a transmitted 0 or 1 depending on a threshold. This is sampled by the clock (CLK) produced by the phase detecting CDR circuit which compares the input signal with a local clock.

The output of the amplifier chain is an analogue signal that is sampled with a clock and data recovery (CDR) circuit in combination with an on-chip phase-locked-loop (PLL) [48] to regenerate a clock from the received signal. This means that the clock does not need to be transmitted separately. The error between the two phases is determined from a phase detector and used to adjust a voltage-controlled oscillator (VCO). This allows the sampling clock and output data to track changes in the input data rate [29]. A decision circuit converts into a digital bit stream by comparing the sampled signal amplitude to a threshold. A high or low logic level is discerned if the voltage at the sample instant is greater or less than the threshold respectively. In an OOK system, this can be implemented with a flip-flop clocked with the regenerated clock. This reproduces a synchronised bit stream that can then be decoded and error corrected. Noise, sampling clock jitter and ISI [98] may cross the threshold during the sampling interval, causing a bit error and therefore degrading BER. The sample point is optimised

when noise and neighbouring symbols in time are minimised and when the current symbol amplitude is maximised [28]. This usually occurs in the middle of a symbol period.

2.5.2 Integrating Receiver

An integrating receiver architecture takes the integral of multiple samples over each symbol period. The area under the time domain signal is thresholded at the decision circuit rather than the peak amplitude, which produces a larger output amplitude than a single sample capture. This means that the symbol amplitude can be increased with a longer integration time [29]. Furthermore, the receiver SNR is less vulnerable to sampling clock jitter with this architecture. An integration window is used to obtain the integral of the received signal over a portion of a symbol between transition points. At the end of each symbol, the integration summation is sampled, and the integrator is reset prior to the next symbol. The main disadvantage an integrating RX is a that a high oversampling factor is required for resolution to approximate a modulated continuous time signal with discrete samples. This results in added complexity because a PLL and high CLK rate must be implemented on chip.

2.6 Modulation Schemes

2.6.1 On-Off Keying (OOK) & Pulse Amplitude Modulation (PAM)

The physical layer bit rate R_b is the gross data transfer speed measured in bits per second (b/s) [44]. This is a function of the number of signalling events across the transmission medium per second (the symbol rate R in baud (Bd)) and the modulation technique [45]. The gross bit rate includes user data as well as protocol overheads such as coding redundancy [46]. The ‘1’ and ‘0’ symbols are random, so typically have an equal probability of occurrence in a bit stream. The reciprocal of R_b is the bit period, T_b. Bit rate is dependent on the bandwidth (BW) of the communication system. The bandwidth is defined as the range of frequencies with less than 3 dB of attenuation by a system. The capacity of a communication system is defined as the maximum information rate it can transfer and this depends on the bandwidth and the signal to noise ratio (SNR) at the receiver.

On-off keying (OOK) is a modulation scheme where a binary ‘1’ bit is conveyed by the presence of a pulse during a symbol period T and a binary ‘0’ is conveyed by the absence of a pulse [14]. OOK is a binary form of amplitude shift keying (ASK), where different signal amplitudes indicate distinct data symbols. Alternative digital schemes express signalling distinctions in the phase or frequency of a carrier with phase shift keying (PSK) or

frequency shift keying (FSK) respectively. FSK is used for improved energy efficiency at the expense of bandwidth [48]. PSK is often combined with amplitude distinctions to give quadrature amplitude modulation (QAM) in industry [45, 46] and research [10].

R is equal to R_b in OOK because only one bit is conveyed per symbol [46]. Multiple bits can be sent in each symbol with the use of pulse amplitude modulation (PAM). For example, four-level PAM (4-PAM) has an alphabet of 4 symbols, which means $\log_2(4) = 2$ bits are carried in each symbol level. As long as the SNR is sufficient to reliably distinguish between the received symbols, this effectively doubles the bit rate which can be attained with the same symbol rate and bandwidth as OOK. Non-return-to-zero (NRZ) line coding, where is the pulse amplitude is held constant throughout the entire symbol period, is implemented in this work. Return-to-zero (RZ) line coding is also considered, where the pulsed signal drops to a rest state at the zero level for the latter portion of the symbol period. RZ pulse width is set by the duty cycle γ and narrower pulses increase the spacing between symbol transitions at the expense of increased bandwidth. The main benefit of this is to mitigate ISI.

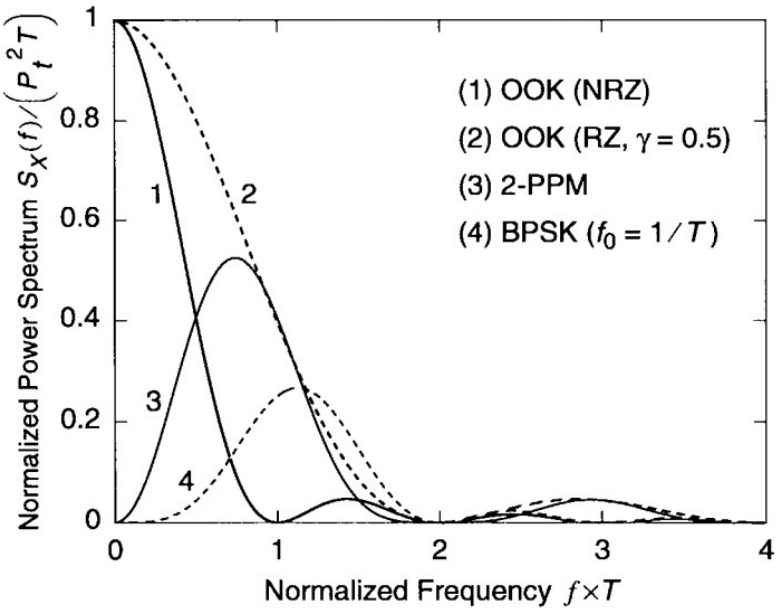

Figure 2.20: NRZ, RZ, 2-PPM and BPSK spectra [26].

Fig. 2.20 shows the power and bandwidth needed for various signalling schemes. The bandwidth required for RZ BW is inversely proportional to the duty cycle, given by the expressions in Table 1 [26]. The expressions in Table 1 are approximations assuming SNR is high. OOK NRZ has power and bandwidth requirements of 0 dB and unity, respectively.

Table 1: Comparison of average power and bandwidth requirements of various modulation schemes normalised to OOK with NRZ on ideal distortion-less channels with AWGN [26].

Modulation Scheme	Normalized Average Power Requirement (Optical dB)	Normalized Bandwidth Requirements
OOK, NRZ	0	1
OOK, RZ, duty cycle γ	$5 \log_{10} \gamma$	$\frac{1}{\gamma}$
N BPSK subcarriers	$1.5 + 5 \log_{10} N$	2
N QPSK subcarriers	$1.5 + 5 \log_{10} N$	1
L-PPM (soft decisions)	$-5 \log_{10}(\frac{L \log_2 L}{2})$	$\frac{L}{\log_2 L}$
L-PPM (hard decisions)	$-5 \log_{10}(\frac{L \log_2 L}{4})$	$\frac{L}{\log_2 L}$
L-DPPM (hard decisions)	$-5 \log_{10}[\frac{L \log_2 L}{8}]$	$\frac{L+1}{2 \log_2 L}$

This means that RZ requires more bandwidth compared to NRZ to attain a given bit rate (Fig. 2.20). For example, the spectrum of an RZ signal with 50 % duty is contained in a frequency band that is twice as wide as the NRZ spectrum. On the other hand, an RZ signal is inherently self-clocking because impulses at multiples of 1/Tb are present in the spectrum. This simplifies the re-timing task of the CDR circuit at the receiver.

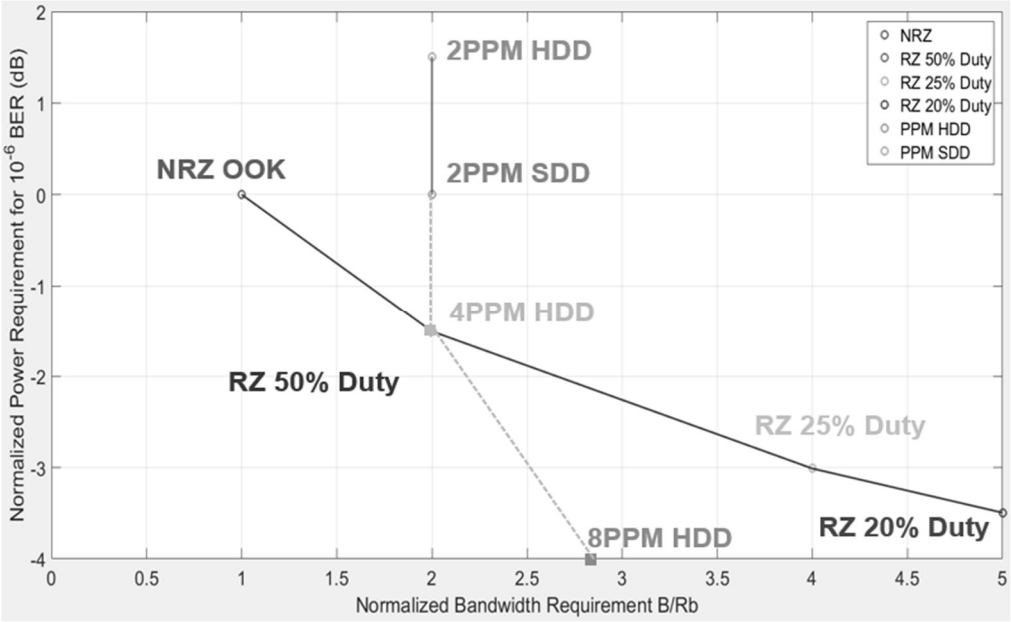

Figure 2.21: Comparison of BW and power requirements for NRZ, RZ and PPM, plotted from Table 1.

Data signals comprise of a range of frequencies (Fig. 2.20), including frequencies higher than R_b, to produce sharp pulses with short rise and fall times. Most of the signal power is contained in the frequencies below f=1/T_b (the bit rate) and there are no spectral components at integer multiples of R_b. Therefore, a signal can be transmitted over a BW lower than the bit rate and the high frequency components and noise can be filtered out with a low pass filter. A bandwidth of 60-70 % of R_b is typically chosen as optimal to receive data, taking into account random jitter and data-dependent jitter caused by additional poles above the –3 dB receiver cut-off frequency [98]. This causes signal transition points to spread out in time over multiple symbols, which may be tolerable to an extent. However, additional complexity with pre-distortion at the TX or equalization at the RX may be required in bandwidth-limited implementations with extensive symbol spreading. Optical communication systems transmit intensity modulated/direct detection (IM/DD) signals by varying the instantaneous optical power of an emitter. This means that the signals are unipolar because the optical power cannot go below zero, and therefore an impulse is present at DC in the spectrum.

2.6.2 Pulse Position Modulation (PPM)

PPM encodes bits in one of L pulse time slots within a symbol period [3, 26]. L-PPM has a lower power requirement than OOK because the low frequency components are absent from its spectrum, but L-PPM requires more BW than OOK (Table 1 and Fig. 2.21) due to the shorter pulses. In addition, PPM is much more complex to implement than OOK because it needs more precise symbol synchronisation at the RX. Hard (HDD) or soft decision decoding (SDD) can be applied for simplicity or power efficiency, respectively [26]. Differential PPM (DPPM) can be used to reduce the power requirement by eliminating the unused time slots in each symbol. PPM is vulnerable to intrasymbol interference when a pulse is shifted in time and misinterpreted as an incorrect symbol, which makes it unideal for VLC channels that are subject to multipath reflections. For these reasons, PPM is not used in this work.

2.6.3 Orthogonal Frequency-Division Multiplexing (OFDM)

Single carrier modulation schemes such as OOK and PAM are simple and inexpensive to implement in VLC but are vulnerable to ISI at high transmission rates due to the low-pass characteristics of optical front ends. Orthogonal frequency division multiplexing (OFDM) is a multicarrier scheme which splits the channel into several narrowband channels [10, 34]. This allows OFDM to be transmitted across non-flat frequency-selective fading channels with single-tap, low complexity equalization in the frequency domain and this aids in mitigating ISI. The

signal is divided into a set of overlapping subcarriers in the frequency domain [35]. The subcarriers are orthogonal to each other, and each is modulated with a scheme such as QAM and then combined to produce the OFDM signal that is transmitted. The spectra of the subcarriers overlap without interfering with each other because the maximum power of each subcarrier corresponds directly with the minimum power of the other subcarriers. This significantly improves spectral efficiency, allowing more information to be conveyed over the available channel bandwidth.

OFDM is used in many modern broadband communication standards such as IEEE 802.11 [45, 46] for multipath dispersive channels. Higher data rates utilise a more robust technique such as OFDM. Furthermore, adaptive bit loading and power allocation can be applied on a substantial number of OFDM subcarriers to improve spectral efficiency and approach channel capacity (the maximum rate information that can be transferred across a given channel). Bits are loaded onto each subcarrier according to the properties of a communication system found from channel estimation made with BPSK pilot signals at the start of an OFDM packet. Power allocation adjusts the SNR on subcarriers with the same QAM depth to a level sufficient to be decoded within the FEC threshold floor. This allows the available spectral resources to be utilised to their full potential and offers an improvement in performance, especially in channels where signal attenuation is significant [34]. OWC systems use IM to encode data in optical intensity due to the incoherent light output of LEDs. This means that only real-valued and positive signals can be transmitted. In contrast, complex-valued and bipolar signals are utilised in conventional RF communication systems. Therefore, standard OFDM must be modified to produce only real-valued symbols for optical OFDM (O-OFDM) in OWC systems. O-OFDM variants include DC-biased optical OFDM (DCO-OFDM) [35], asymmetrically clipped optical OFDM (ACO-OFDM) [10] and unipolar OFDM (U-OFDM) [35]. Another advantage of O-OFDM is that it allows a multiple access scheme to be implemented with ease in an optical attocell network by dividing time and frequency resources among several users. This is similar to orthogonal frequency division multiple access (OFDMA) in RF systems [17].

An OFDM time domain signal has a high crest factor, which results in a high peak-to-average power ratio (PAPR) [34]. This is considered as a disadvantage in conventional RF systems, as it demands a TX with a large dynamic range, and therefore reduces power efficiency (due to power amplifier nonlinearities) [10]. This inefficiency is because the amplifier operates at only a fraction of its full capability for most of the time. In a VLC OFDM system, the high peak-to-average characteristic can be utilised productively since the signal is used to modulate the intensity of an LED TX. The signal variations can be centred on an operating point appropriate for the LED in use, such that it functions in the linear region of its current against intensity curve. SPADs are highly nonlinear

and have low dynamic range due to dead time, so the high PAPR of OFDM could cause errors in a SPAD-based system due to clipping of the high amplitudes of the OFDM signal. Nonlinear distortion of SPADs at high irradiances will be quantified in this work in Chapter 5 (Section 5.4.3). Fig. 2.22 summarises modulation schemes suitable for optical communication – VLC in particular.

2.7 Transmitters

An optical source can be either directly or externally modulated. In a direct modulation system, the optical power of the source is modulated via the current from a driver with a time-varying voltage signal. For OOK, the source is turned on and off directly for a logical one and zero, respectively [26]. When external modulation is applied, the source emits in continuous wave (CW) mode at full power. The modulation is performed by an external optical shutter that is open for a logical one symbol and closed for a zero symbol [28]. Direct modulation is simpler, cheaper, and more compact than external modulation, so it is used throughout this work.

2.8 Summary of SPAD Receivers

OWC channels suffer from high attenuation (with a typical DC gain around -50 dB across a 1 m link [103]) and low link margin compared to RF and wireline technologies due to the inverse square law. This is because SNR of an OWC RX is proportional to the square of the received optical power. OWC links often employ high transmit power and operate over a limited range as a result. Non-imaging optics such as CPC concentrators [103] can be employed to increase the received power and improve the SNR, but this is limited to an optical gain of only up to around 10 due to the Etendue law (around the detector areas used in OWC) and these can be bulky and therefore unsuitable for consumer applications such as smartphones. In addition, concentrators also add complexity to the assembly process and manufacturability issues because the optics have to be attached with bonding such as epoxy on the surface of the detector package or window. Therefore, a large detector area is required to achieve the same SNR for error-free demodulation at the RX. This is made worse in high-speed systems, because a small diode capacitance (and therefore small active area) is required to provide the extrinsic bandwidth needed to achieve high symbol rates. An array of PIN or APD detectors can be used [96] but for the ranges for VLC links (2-3 m up to 10 m) [10-17] and high speeds (>1 GHz), the array size becomes unmanageable because the voltage signals after each TIA must be combined via a summing amplifier and this is challenging to implement. It becomes preferable to combine hundreds, if not thousands, of detectors into a single node and an ASiPM is more efficient.

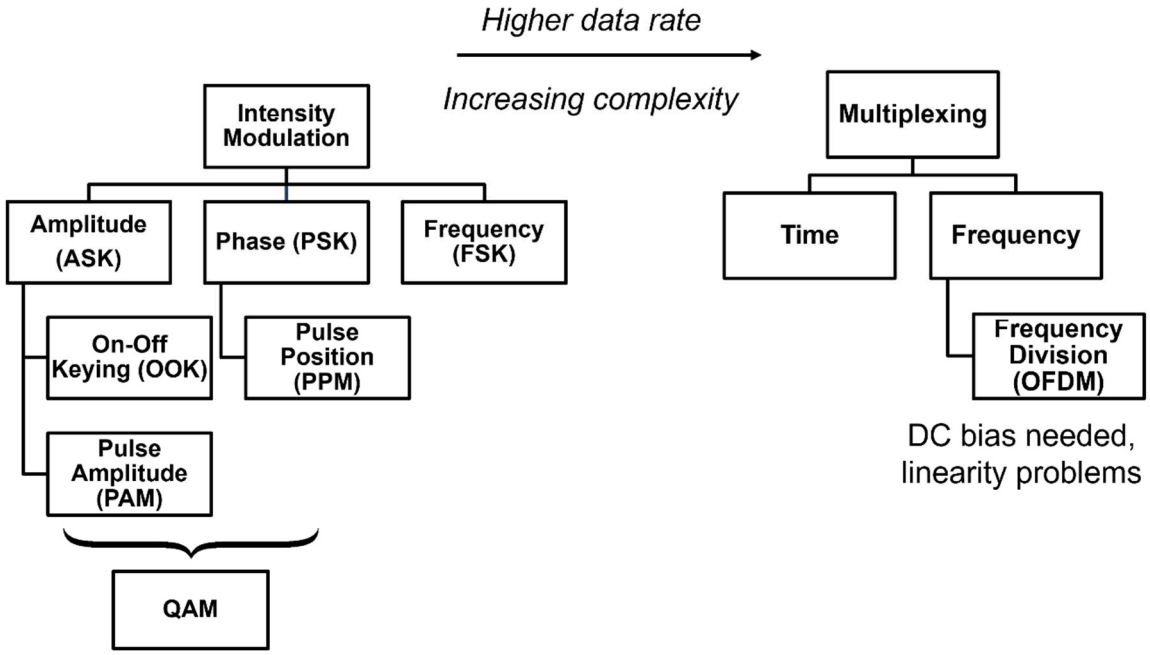

Figure 2.22: Comparison of modulation schemes.

3.45 Gb/s OOK with 10^{-3} BER has been attained with commercial ASiPMs and laser diodes [32], but the detected counts need to be calculated from the analogue voltage signal captured on an oscilloscope. A DSiPM would avoid this problem. It has been demonstrated SPADs can enhance the sensitivity of optical receivers in communication systems, offering to increase RX SNR and therefore potentially channel capacity [28, 32]. Furthermore, 4-PAM can be supported with the increased SNR made available using an array of SPADs and [114] establishes a technique to determine the optimum signal count levels used to represent 4-PAM symbols with a commercial ASiPM module. Two methods of counting detected photons are: peak counting, using the output voltage peaks to count photons, and summation, where for each pulse all the sampled outputs within a time interval after the pulse are summed together and then divided by a specific value and rounded to the nearest integer. The total number of counts within the integration time is equal to the sum of these integers. Digital summation is chosen for this work.

Low noise operation is desired to ensure high sensitivity, and low power consumption is needed to allow integration into a small form factor module. Sensitivity and power consumption have a trade-off: reducing the power consumption of the TIA at the expense of sensitivity requires a corresponding increase in the TX output power, increasing the TX power consumption and reducing the emitter lifetime. In addition, the optical gain of an RX is limited by the required field of view. This highlights the necessity for a fast RX with high sensitivity. It is critical to move the received signal to digital as soon as possible [82] and SPADs exploit the discrete nature of photons and act as an optical ADC. Improvements in computing power and FPGA technology enable complex

DSP and equalization to be applied in the digital domain with ease. In this work, a SPAD array and combination circuitry replaces the TIA converting the photocurrent into a sufficiently large voltage swing, such that the subsequent clock-and-data recovery (CDR) circuit can accurately recover the sent bit stream.

RF systems use complex modulation schemes such as OFDM and spatial multiplexing to obtain high bit rates [11]. These techniques are viable due to the high sensitivity of RF RXs (compared to optical RXs) and the ability to embed complexity in low-cost silicon ICs. A large array DSiPM has potential to bridge this gap, so a digital architecture forms the scope of this thesis.

3. Simulation of a SPAD VLC System

3.1 Introduction

Given the background literature reviewed in Chapter 2, this chapter introduces a complete model to quantify the quantum limit of detection and estimate error rate performance. This is used to assess the effectiveness of optical RXs. The model assumes an LD source with >1 GHz bandwidth; precise alignment; dark ambient light conditions (0 lx illuminance); and a narrow bandpass optical filter at the detector matched to the LD centre wavelength to ensure that the SPAD RX is the only limiting factor on bandwidth and error performance. A CMOS SPAD with passive quenching is also assumed to minimise power consumption in a real RX implementation.

State-of-art theory can explain power penalties due to optical losses, TX extinction ratio and second-order effects, however, the dead time in practical SPAD RXs remains as an opaque extra penalty from the QL. This penalty cannot be explained with current SPAD models, so a novel method of describing nonlinearity due to dead time is established. In this section, a model is developed to bring together all the channel impairments introduced by the physical parameters of SPADs and determine the resulting effect on SNR, data rate and BER.

3.2 Single Photon Model of Light

The detection process of an ideal photon counting receiver is modelled using Poisson statistics which describe the shot noise effect [9]. Photon shot noise is an uncertainty related to the quantised nature of light and statistical fluctuations in the time between photon arrivals. This sets the standard quantum limit of detection. The number of photon arrivals k in a time interval T is a realisation of a random variable that follows a Poisson probability distribution $p_0(k)$ with a constant average photon rate λ (in s^{-1}) [115]:

$$p_0(k) = \frac{(\lambda T)^k e^{-\lambda T}}{k!} \tag{1}$$

The mean of the photocount distribution, μ_0, is equal to the variance σ_0^2:

$$\mu_0 = \sigma_0^2 = \lambda T \tag{2}$$

The inter-arrival times between photons can be determined from a realised set of values of k. Fig. 3.1 shows the probability mass function (PMF) found from (1) for different mean incident photon rates and T =1 μs. At

higher intensities, the Poisson distribution approaches a Gaussian distribution about its mean, in accordance with the Central Limit Theorem [116], making the shot noise akin to Gaussian noise. A Poisson process with a mean of around 10 photons and above can be approximated by a Gaussian distribution [116].

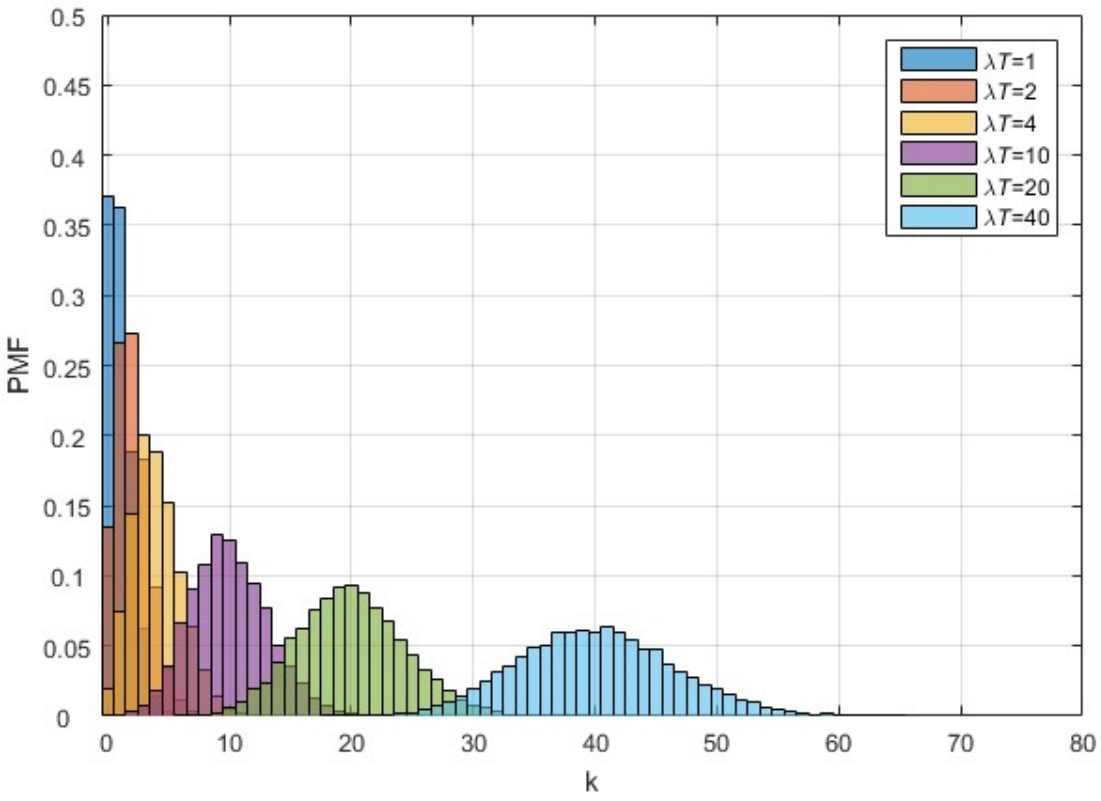

Figure 3.1: Poisson PMF of k photon arrivals with average incident photons/symbol λT.

BER is a key performance metric used to evaluate the integrity of a communication link. A first order BER model for AWGN channels is recognised in prior literature [48]. This model estimates the probability of error P_e by determining the overlap between the received symbol distributions. BER= P_e for OOK. Fig. 3.2 shows an example of an OOK signal with Poisson statistics. The received high and low amplitude distributions are interpreted by the RX as '1' and '0' symbols affected by noise and distortion. A similar model for BER is applied to a photon counting system with Poisson statistics [38]:

$$BER = \frac{1}{2}\sum_{k=0}^{m_T}\frac{(K_s+K_n)^k}{k!}e^{-(K_s+K_n)} + \frac{1}{2}\sum_{k=m_T+1}^{\infty}\frac{K_n^k}{k!}e^{-K_n} \quad (3)$$

where $K_n = \lambda_n T$ is the mean number of photons detected within a '0' symbol of length T (due to background counts arising from dark counts, afterpulsing, ambient light, and finite emitter ER) and λ_n is the mean background

noise photon rate. $K_s = \lambda_s T$ is the mean number of signal photons detected within a '1' symbol of length T for a mean signal photon rate λ_s. This model assumes the data sent is a random bit sequence and the 0 and 1 bits are equiprobable.

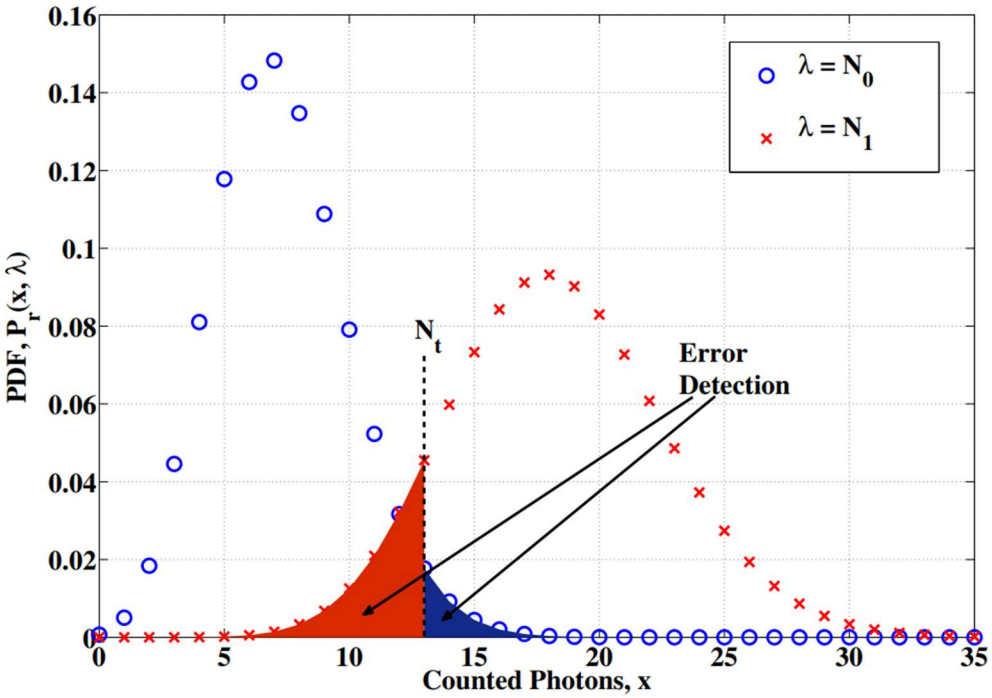

Figure 3.2: Probability functions for OOK detection with threshold N_t [117]. Each x symbol count distribution has a mean, N_1 and N_0. The standard deviations of each distribution (σ_1 and σ_0) are caused by noise. It is not possible to reduce σ_1 and σ_0 below $\sqrt{N_1}$ and $\sqrt{N_0}$, respectively, due to the Poisson statistics of light.

The RX discerns the symbol distributions by applying a threshold m_T to the photocounts. A decoding error occurs if $k \leq m_T$ when a '1' symbol is sent, or if $k > m_T$ when a '0' symbol is sent. $m_T + 1$ is used for the lower bound of the logic zero probability expression because k can only have integer values. BER depends on the choice of m_T, which can be selected to minimise the probability of making an error. This occurs at the value of m_T where $\frac{dBER}{dm_T} = 0$. Differentiating Eqn. (3), rearranging, and taking the natural logarithm on both sides gives the optimum threshold:

$$m_T = \frac{K_s}{\ln\left(1 + \frac{K_s}{K_n}\right)} \tag{4}$$

BER can be improved if K_s and K_n are spaced further apart, by increasing the SNR or TX ER, reducing ambient light, or increasing the transmit power. However, emitter output power is often limited by eye safety restrictions in OWC systems [103] and even with optical filtering, background light is almost always present in any consumer application [74]. This implies that the RX should have high detection efficiency to maximise K_s.

In most electronic and RF systems, AWGN is additive and affects each symbol equally. However, in a system with only shot noise, the noise standard deviation of each symbol is proportional to the amplitude of the symbol, meaning that higher signal levels have more noise than lower signal levels.

The quantum limit (QL), expressed in Eqn. (5), is a reduction of Eqn. (3) that assumes there are no photons on the '0' symbol ($K_n = 0$) and therefore it is impossible for a received '0' symbol to be mistaken for a '1' [49]:

$$BER_{QL} = \frac{e^{-K_s}}{2} \tag{5}$$

K_s is the mean number of photons detected in a symbol period by an ideal photon counting RX. (5) also assumes K_s has an ideal Poisson process and the only noise source present in the system is the unavoidable shot noise induced by the signal itself. Rearranging Eqn. (5) gives

$$K_s = -\ln(2BER_{QL}) \tag{6}$$

which can be used to determine the mean number of detected photons per symbol required for a specific BER, assuming an ideal RX at the QL.

Table 2 shows the solution of (6) for different BER specifications. BER_{QL} in the second column is the minimum native BER required for FEC to produce an output BER of 10^{-9}, which is considered low enough to assume error-free communication [48]. It should be noted that redundant bits must be sent with the data payload for error correction purposes, so the net data rate (the effective information rate) is lower than the gross line bit rate. Redundancy ratio is the ratio between the number of information bits covered by the FEC and the number of bits after encoding, comprising the information bits plus the added redundant bits. This ratio depends on the FEC method used. For example, there is 24.48% redundancy with concatenated Reed-Solomon (RS) outer and convolutional self-orthogonal code (CSOC) inner code super FEC [19] and there is no redundancy if on-board hardware supports 10^{-9} BER or lower because no FEC is required.

Table 2: Mean number of photons per symbol required to obtain a quantum limited BER and BER after FEC. 3.5×10^{-3} BER can be obtained with a mean of 4.96 photons per symbol (in bold).

Photons/symbol K_s	BER BER_{QL}	Redundancy for 10^{-9} output BER (%)	Sensitivity at 100 Mb/s, 450 nm (dBm)
1	1.84×10^{-1}	–	–76.56
4.34	6.5×10^{-3}	24.48	–70.18
4.96	**3.5×10^{-3}**	**6.69**	**–69.61**
5.52	2×10^{-3}	6.69	–69.14
20.03	10^{-9}	0 (uncoded)	–64.08

The minimum BER specification for this project is chosen to be 3.5×10^{-3} because this value gives a balance of low bit redundancy (6.69%) and realistic attainable BER after research into practical implementations in prior literature (mostly reporting 10^{-3} [32, 33, 38, 39] to 6.5×10^{-3} [28]). Therefore, the absolute minimum mean at the detector is 4.96 photons/symbol for a '1' bit. This value will be higher in reality because more photons will be required to account for losses due to PDP, FF, background counts, DCR, dead time, afterpulsing, and crosstalk at the RX, as well as ER and relative intensity noise (RIN) at the TX.

The photon energy for a given emitter wavelength can be used to quantify the QL receiver sensitivity. K_s is related to the average received optical power P_r by [118]

$$P_r = \frac{K_s}{2T}h\nu = \frac{K_s R}{2}h\nu \tag{7}$$

where h is Planck's constant, ν is the frequency of the light and R is the symbol rate ($R = \frac{1}{T}$). Assuming $K_n = 0$, an average of $\frac{K_s}{2}$ photons are required per bit (zero or one). P_r is proportional to R, so more power is required to attain higher data rates. For OFDM, the received time domain signal is discretised and consists of photons per sample, so P_r is proportional to the sampling rate of the RX, rather than the symbol rate as with single carrier modulation schemes, and therefore the average received optical power is given by

$$P_{r\,OFDM} = \frac{K_s F_s}{2}h\nu \tag{8}$$

where F_s is the sampling rate. The received power (in W) is related to the irradiance of the incident light I_r (in Wm^{-2}) and the effective collection area of the detector A_e (m^2) [26]:

$$P_r = \int_{A_e} I_r dA \tag{9}$$

I_r depends on the radiation pattern of the emitter and the distance and angle of the RX in respect to the TX. This expression implies the detector area should be large to capture as many photons as possible. P_r is typically expressed in dBm (the power with reference to 1 mW) to define the sensitivity of a receiver [49]:

$$P_r(\text{dBm}) = 10\log_{10}\left(\frac{P_r}{1\times 10^{-3}}\right) \tag{10}$$

This is the minimum power required at the detector to decode the signal at a given BER. 100 Mb/s QL sensitivities at 450nm are shown in column four of Table 3.

3.3 Power Penalty Budget

For a SPAD RX, the fill factor (due to in-pixel circuitry or manufacturing design rules) reduces the effective detector area. In addition, the SPAD photon detection efficiency (PDE) is finite because there are photon losses. Modifying Eqn. (7), both of these effects increase the received optical power P_s required on the SPAD active area:

$$P_s = \frac{1}{C_{FF}PDP}\frac{K_s R}{2}h\nu \tag{11}$$

where C_{FF} is the SPAD fill factor and PDP is the SPAD photon detection probability at the wavelength of the light. These effects induce power penalties because a higher signal power is required to overcome the losses. Therefore, the effective RX sensitivity is degraded and raised above the QL.

Eqns. (12) and (13) describe the power penalties caused by fill factor (PP_{FF}) and PDP (PP_{PDP}), respectively.

$$PP_{FF} = -10\log_{10}(C_{FF}) \tag{12}$$

$$PP_{PDP} = -10\log_{10}(PDP) \tag{13}$$

where PP_{FF} and PP_{PDP} are in dB.

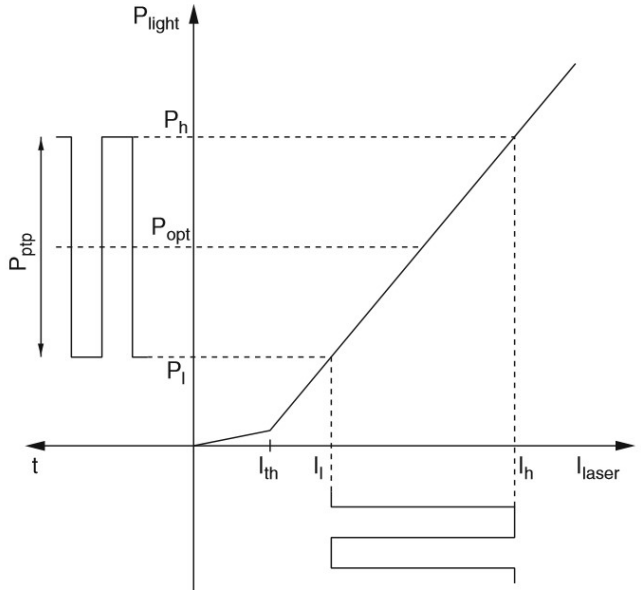

Figure 3.3: Input-output characteristic of a typical semiconductor source [98].

In addition, inefficiencies at the emitter will induce additional losses in the channel and further degrade sensitivity. An ideal source would have infinite extinction ratio (ER) – the ratio between the high and low levels (P_h and P_l) corresponding to logical one and zero for OOK [98]. ER is given by

$$ER = \frac{P_h}{P_l} \tag{14}$$

In reality, light sources are not capable of fully turning off at high speeds due to their extrinsic or intrinsic bandwidths, which means that a threshold current is required [98]. Therefore, a DC offset is imposed on the data-carrying optical signal. The input-output response of a practical source is shown in Fig. 3.3. The current through the emitter should always be larger than the threshold current I_{th} to minimise the turn-on delay t_{on} when changing from a zero to a one [53]:

$$t_{on} = \tau_c \ln\left(\frac{I_h - I_l}{I_h - I_{th}}\right) \tag{15}$$

where τ_c is the carrier lifetime, typically around 3 ns for laser diodes [98]. I_h is the logic one current. t_{on} decreases if the logic zero current I_l is increased, and $t_{on} \to 0$ if $I_l \geq I_{th}$. t_{on} restricts the BW of the TX, so I_l must be increased and therefore a reduced ER is unavoidable. In addition, the emitter should always be on to minimise jitter caused by relaxation oscillation when the source is driven by a step [49]. The effect of non-ideal ER can be accounted for by computing the eye-opening power penalty (PP_{ER}) from the received eye diagram [98]:

$$PP_{ER} = 10 \log_{10}\left(\frac{ER + 1}{ER - 1}\right) \tag{16}$$

PP_{ER} (in dB) relates the average optical power P_{op} to the average signal power P_s:

$$P_s[\text{dBm}] = P_{op}[\text{dBm}] - PP_{ER}[\text{dB}] \tag{17}$$

where $P_s = \frac{P_h - P_l}{2}$ and $P_{op} = \frac{P_h + P_l}{2}$. Now P_{op} sets the RX sensitivity. The relationship between ER and PP_{ER} is shown in Fig. 3.4. Higher ER means the power is used more efficiently because a smaller portion of the total power is used to bias the source. ER can be improved with external modulation [28] but with compromised TX power efficiency at system level because the source is continuously emitting. Finite extinction ratio therefore causes background counts and $K_n \neq 0$, making Eqn. (5) invalid. Instead, Eqn. (3) must be used to estimate BER.

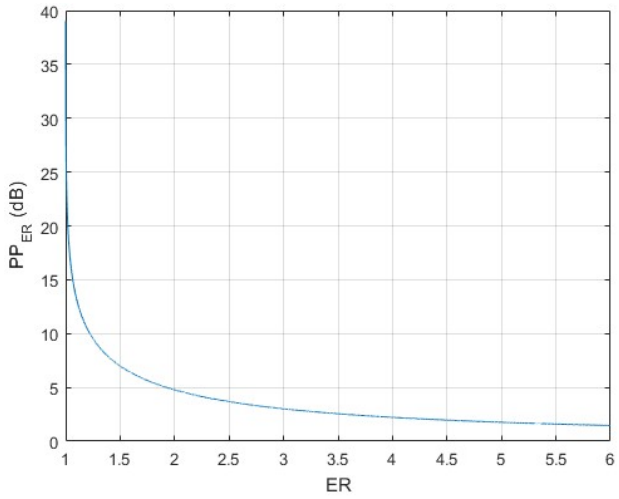

Figure 3.4: Emitter extinction ratio ER and the corresponding power penalty it induces: PP_{ER}.

SPAD DCR will also add background counts to the symbol distributions, causing λ_n to increase. This means a higher λ_s is required to separate a '1' from a '0' and maintain BER – creating a power penalty PP_{DCR}:

$$PP_{DRC} = 10\log_{10}\left(1 + \frac{2DCR}{\lambda_s}\right) \tag{18}$$

At high speeds, the mean number of dark counts per symbol $\frac{DCR}{R}$ will be very low because DCR is typically in the range of 10-10 ks^{-1} per SPAD [57]. This means that it is improbable a dark count will cause a bit error (only a '0' being mistaken for a '1' because the total counts are above the threshold), making it almost negligible as a penalty. Additive counts also arise from parasitic afterpulses and crosstalk, but these effects cannot be estimated without specific SPAD parameters in practical experimentation [119].

Relative intensity noise (RIN) describes fluctuation in emitter optical intensity due to power supply and source variations [49]. Wave fluctuations in the light are the main contributors to RIN because the power and temperature of the emitter can be controlled well with VCSELs [98] and heat sinking [49].

RIN adds a further penalty, PP_{RIN}, given by

$$PP_{RIN} = -10\log_{10}\left(1 - BW\,10^{\frac{RIN}{10}}SNR\right) \tag{19}$$

where RIN (in dB/Hz) is the noise density specification of the emitter, BW is the bandwidth, and SNR is the linear signal-to-noise ratio required for the target BER. RIN is typically around -110 dB/Hz or less [49].

The summation of all of these penalties can be used to establish a link budget for a SPAD-based communication system:

$$P_{op} = P_{r\,QL} + PP_{RIN} + PP_{ER} + PP_{DCR} + PP_{FF} + PP_{PDP} + PP_{ISI} \tag{20}$$

where $P_{r\,QL}$ is the QL sensitivity at the centre wavelength of the emitter and the symbol rate of the system. From Eqn. (19), PP_{RIN} ≈ -0.04 dB with 60 MHz BW and 10 dB SNR [49]; PP_{ER} is less than 2 dB for most laser diode sources (Fig. 3.4); and assuming a single SPAD detector in Eqn. (18), PP_{DRC} ≅ 0.0001 dB with λ_s = 4.96×100 Mb/s = 496 Ms^{-1} (at the QL in Table 2) and a typical DCR = 6 ks^{-1} [57]. Combined PDP+FF values are up to 38 % with state-of-the-art CMOS SPADs [57], so $PP_{FF} + PP_{PDP}$ ≅ 4.2 dB (Eqn. (12) plus Eqn. (13)). Adding all of these losses together gives a combined upper bound 4.2 + 2 + 0.04 + 0.0001 = 6.24 dB gap from the QL at 100 Mb/s. The dominant losses are PDP and FF.

SPAD dead time induces intersymbol interference (ISI), which is included in the link estimation in Eqn. (20) as another power penalty, PP_{ISI}. The effect of dead time on the performance of SPAD RXs has not been explained in literature. Some studies assume that the received counts follow a Poisson-distributed random variable, but ignore dead time, so BER is often considerably overestimated [38]. Other studies approximate the received counts as a Poisson distribution with a mean rate reduced by dead time [7, 8]. This rate is derived from renewal theory that assumes the incident photon rate is constant [9], therefore these approximations are only valid if symbol periods are much longer than dead time; ISI is ignored. For high-speed transmission, when the symbol period is shorter than the dead time, the approximation cannot be used. [36] derives probability mass functions (PMFs) of photocounts detected by PQ and AQ SPADs; however, ISI is also omitted by assuming that the SPAD

is always ready to detect at the start of each symbol. To alleviate dead time ISI, the bit rate of SPAD-based implementations is often selected such that the symbol period is larger than or approximately equal to the dead time [28, 29, 37].

3.4 Completing the Link Budget of a SPAD-based System

ISI reduces the eye opening of the received eye diagram, therefore increasing the error rate [48]. For example, if the bandwidth of the front-end is insufficient, a transmitted one symbol pulse $x(N)$ at symbol index N may spread into a subsequent zero symbol $x(N+1)$ over an extended fall time, causing the zero to be mistaken for a one at the decision instant. Properties such as multipath channel propagation in VLC cause significant ISI and are difficult to characterise and eliminate [42]. ISI can also be caused by delays in the channel and may create peaks to the received distributions, degrading BER. Dead time τ causes ISI in a similar way in a discrete counting system because an avalanche pulse started from a detection in symbol $x(N)$ extends to subsequent symbols $x(N+1)$, $x(N+2)$, $x(N+3)$, …, preventing the SPAD from detecting photons in '1' symbols. This is more likely to happen when $x(N)$ is a '1', and especially likely if a long string of consecutive '1's is sent. The inactive period between the beginning of a symbol T_{st} and the falling edge time (T_F) from the last photon detection is defined as the block time T_B. This effect is illustrated in Fig. 3.5 for a photon arrival (before the current symbol) at time T_L. For a PQ SPAD, where τ is extended at high irradiances, T_B can only be defined from the falling edge T_F of the last pulse in the previous symbol as [42]

$$T_B = \max\{T_F - T_{st}, 0\} \tag{21}$$

A SPAD is active from $T_{st} + T_B$ rather than T_{st}. T_L and T_F are random variables, so the ISI causes a random inactive time which changes the photocount statistics. Raised cosine pulse shaping [48] and pre-emphasis at the TX [96] can help to reduce the impact of ISI. However, these techniques only work effectively if the system response is known, including the overall bandwidth, transfer function, and impulse response of the signal path. Novel methods such as photon time information-based detection (PTID) have been designed to reduce ISI by exploiting the high timing resolution of SPADs [47]. PTID uses both the received counts and photon arrival times to recalculate a threshold based on the instantaneous block time during a symbol. This method is similar to fading channel signal detection where the threshold is adaptively adjusted with respect to the instantaneous fading status to minimise BER [18]. Another scheme that employs SPAD time-correlated single photon counting (TCSPC) encoding demonstrates improved resilience to background light [14]. The main disadvantage of this

method is that the data rate is significantly reduced due to the histograms that need to be made in order to decode.

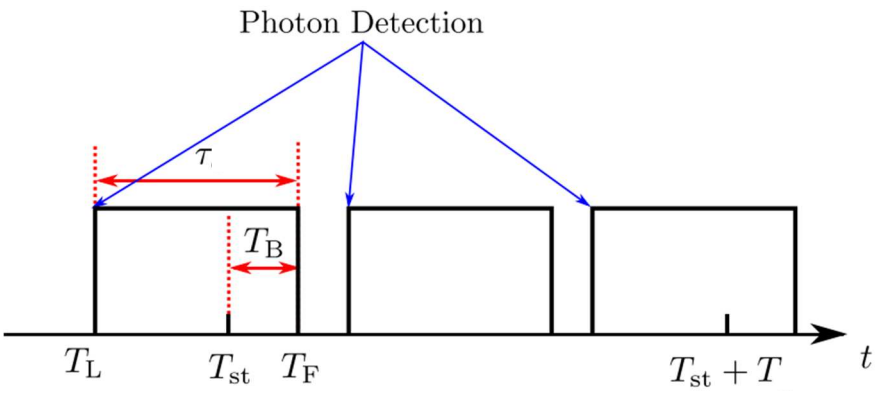

Figure 3.5: Photon detection sequence with block time T_B, where T_L is the last photon arrival time before a given symbol of length T; T_{st} and $T_{st}+T$ are the start and end of the symbol, respectively; and T_F is the end of the SPAD pulse with dead time τ. [42]

Digital equalization such as decision-feedback equalizers (DFEs) [48], originally intended for RF and wireline systems with linear channel expressions, have been employed effectively in SPAD-based OWC systems to mitigate sub-dead-time ISI to some extent [32, 38]. However, SPAD RXs are inherently nonlinear and ISI induced by dead time remains the dominant non-ideal effect in SPAD RXs. Explaining this effect mathematically is the main contribution of this thesis. The resulting power penalty, PP_{ISI}, cannot be estimated in the same way to the previous effects discussed earlier because the mathematical derivation depends on too many variables, especially when considering an array of SPADs. This is because a SPAD photon counting system has memory (due to previous detections), the received discrete signal is both power-dependent and time-dependent, and because the response to dead time depends on the physical steady-state behaviour unique to the SPAD device being used, which could only be determined from practical experimentation [119], prior to this work. Given this challenge, the scope of this work is to model and solve a steady-state response with dead time and then quantify PP_{ISI}.

3.5 A Nonlinear Numerical and Analytical Model

3.5.1 SPAD Response

I first consider a conventional circuit model and then apply a similar principle to an optical system. Amplifiers experience gain compression where the gain decreases for increasing amplitude, since eventually the output

reaches a limit due to supply voltage, for example (Fig. 3.6) [120]. This results in a nonlinear input/output relationship. A SPAD has a similar transfer relationship with greater deviation from linearity as the incident photon rate increases (Fig. 3.7) due to an undesirable dead time caused by resetting the SPAD after it detects a photon. During this time interval (in the order of 10-14 ns [38]) the SPAD is unable to respond to impinging photons.

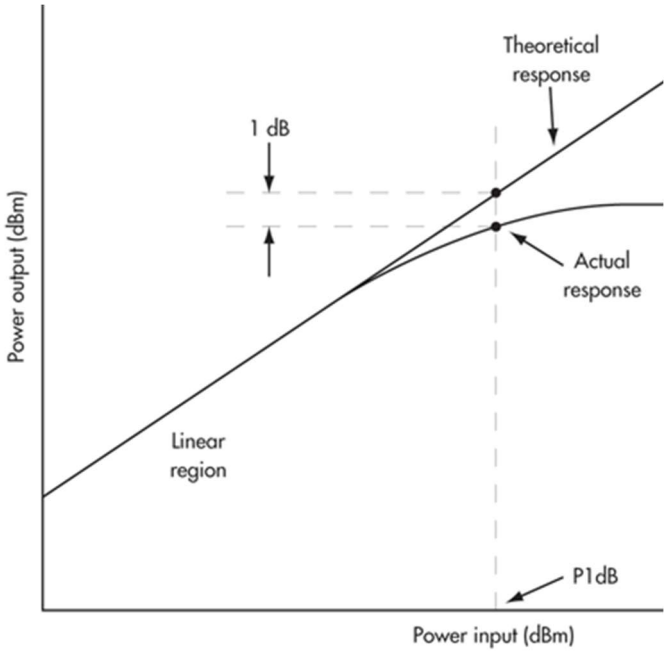

Figure 3.6: Input/output relationship of a nonlinear system. Amplifiers experience gain compression where gain decreases for increasing amplitude since the output reaches a limit due to the supply voltage [122].

A nonlinear system with input x and output y can be represented with a power series [121]

$$y(x,t) = a_1 x + a_2 x^2 + a_3 x^3 + \cdots \tag{22}$$

where a_1 is the linear, small signal gain (unity in this case). I intend to describe the large signal curve of a SPAD (Fig. 3.7) with dead time τ by this polynomial. The objective is to use this to predict the distortion induced on a modulated input signal. Background noise from dark counts and ambient light is omitted at this point to concentrate on the effect of dead time. Photon shot noise is included in the model since it cannot be eliminated from a practical system.

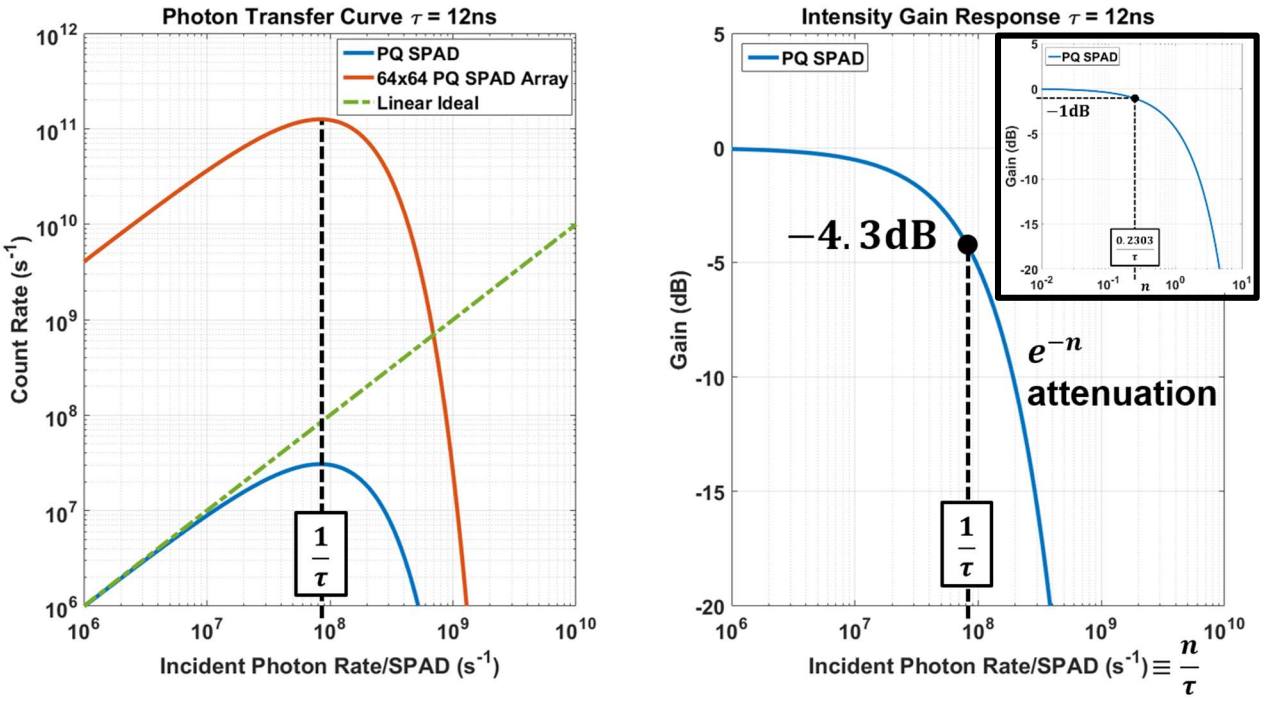

Figure 3.7: Transfer curves of a PQ SPAD and a 64 × 64 array (left). Corresponding gain (right). Inset: gain as a function of multiples n of $\frac{1}{\tau}$. –1 dB compression point (highlighted) at $n = 0.2303$ for a DC arrival rate.

For a PQ SPAD, any photon arriving during the dead time of a previous detection causes τ to be extended. A PQ SPAD output event rate is modelled in [108] as a paralyzable detector [123, 124]:

$$y_{PQ} = Ae^{-A\tau} \tag{23}$$

This model assumes a constant (DC) mean incident photon rate A. τ sets effective count rate and the saturation level is scaled by the array size. An expression for the DC gain G_{PQ} as the input to output ratio is:

$$G_{PQ} = \frac{Ae^{-A\tau}}{A} = e^{-A\tau} = e^{-n} \tag{24}$$

G_{PQ} can be reduced to e^{-n} where $A = \frac{n}{\tau}$ is a normalisation based on multiples of the maximum input event rate $\frac{1}{\tau}$. As the value of factor n increases, the SPAD becomes more nonlinear. Count rate maxima occur at the dead time cut-off point $\frac{1}{\tau}$ marked in Fig. 3.7. As a means of quantifying gain compression, the input level where G_{PQ} has dropped by 1dB is found: $G_{PQ} = e^{-n} = -1$ dB when $n = 0.2303$. Furthermore, the output drops by 4.34 dB when $n = 1$ and attenuation continues to grow exponentially as n increases, with –43.4 dB a decade higher at $n = 10$. DC gain is plotted in Fig. 3.7 and is identical for any number of SPADs in terms of incident rate per

element. Before considering a modulated signal, which is mathematically involved, a single tone sinusoidal input signal with frequency ω_1 is applied,

$$x = A\cos(\omega_1 t) \tag{25}$$

A is the amplitude rate proportional to the incident optical power. The incident photon rate now varies in time for this input. From Eqns. (22) and (25), up to the third term, the output expansion produces

$$y = a_1 A\cos(\omega_1 t) + a_2 A^2\cos^2(\omega_1 t) + a_3 A^3\cos^3(\omega_1 t)$$

$$= \frac{a_2 A^2}{2} + \left(a_1 A + \frac{3a_3 A^3}{4}\right)\cos(\omega_1 t) + \frac{a_2 A^2}{2}\cos(2\omega_1 t) + \frac{a_3 A^3}{4}\cos(3\omega_1 t) \tag{26}$$

Unwanted higher harmonics (second $2\omega_1$ and third $3\omega_1$) are also generated in (26) and second order nonlinearity causes a DC shift of $\frac{a_2 A^2}{2}$. Harmonic distortion is due to self-mixing of the signal. It can be suppressed by low pass filtering the higher order harmonics. The third order generates both third order harmonic distortion and a fundamental component which distorts the linear term. The gain of the system is then deduced:

$$G = \frac{y_{\omega_1}}{x} = \frac{\left(a_1 A + \frac{3a_3 A^3}{4}\right)}{A} = a_1 + \frac{3a_3 A^2}{4} = a_1\left(1 + \frac{3a_3 A^2}{4a_1}\right) \tag{27}$$

If $\frac{a_3}{a_1} < 0$, the gain compresses with increasing amplitude and since $a_1 = 1$, a_3 must be negative. The –1 dB compression point can be determined:

$$10\log_{10}\left(1 + \frac{3a_3 A^2}{4a_1}\right) = -1\text{dB} \tag{28}$$

Note ×10 rather than the conventional voltage ×20 as this is typified by discrete photon detections proportional to the received optical power. If two tones are applied to the system,

$$x = B\cos(\omega_1 t) + B\cos(\omega_2 t) \tag{29}$$

where B is half of A, such that the peak-to-peak amplitude of the waveform ($2B = A$) is equal to the rate in (23):

$$y = a_2 B^2 + \left(a_1 B + \frac{9a_3 B^3}{4}\right)\cos(\omega_{1,2} t) + a_2 B^2\cos((\omega_1 \pm \omega_2)t) + \frac{3a_3 B^3}{4}\cos((2\omega_1 \pm \omega_2)t) + \frac{3a_3 B^3}{4}\cos((2\omega_2 \pm \omega_1)t) + \frac{a_2 B^2}{2}\cos(2\omega_{1,2} t) + \frac{a_3 B^3}{4}\cos(3\omega_{1,2} t) \tag{30}$$

where $\cos(k\omega_{1,2}t) = \cos(k\omega_1 t) + \cos(k\omega_2 t)$ and $k = 1, 2\ or\ 3$. Intermodulation terms arise at $\omega_1 \pm \omega_2$, $2\omega_1 \pm \omega_2$ and $2\omega_2 \pm \omega_1$. This is caused by the two signals cross-mixing. When B is sufficiently small, the higher order nonlinear terms are negligible, and the gain remains approximately a_1. As B increases, the fundamentals

increase proportionally, whereas the third order intermodulation (IM3) products increase in proportion to B^3. Eqn. (30) is modified by considering an IM/DD optical signal which is unipolar (as there is no negative light), in contrary to bipolar radio frequency (RF) with electric fields. A unipolar tone with a DC bias is equivalent to setting $\omega_2 = 0$ in (29):

$$x = B\cos(\omega_1 t) + B_{DC} \tag{31}$$

DC bias B_{DC} is later set to be different to signal amplitude to investigate the effect of a TX with finite ER. Therefore,

$$y = a_1 B + \frac{3a_2 B^2}{2} + \frac{5a_3 B^3}{2} + \left(a_1 B + 2a_2 B^2 + \frac{15a_3 B^3}{4}\right)\cos(\omega_1 t) + \left(\frac{a_2 B^2}{2} + \frac{3a_3 B^3}{2}\right)\cos(2\omega_1 t) + \frac{a_3 B^3}{4}\cos(3\omega_1 t) \tag{32}$$

It is seen that intermodulation between the DC and signal components occurs and the higher order terms in (30) fold down to DC and the fundamental, adding distortion. Now, distorted DC and fundamental gains are obtained:

$$\begin{aligned} G_{DC} &= \frac{y_{DC}}{x_{DC}} = \frac{\left(a_1 B + \frac{3a_2 B^2}{2} + \frac{5a_3 B^3}{2}\right)}{B} = 1 + \frac{3a_2 B}{2} + \frac{5a_3 B^2}{2} \\ G_1 &= \frac{y_1}{x_1} = \frac{\left(a_1 B + 2a_2 B^2 + \frac{15a_3 B^3}{4}\right)}{B} = 1 + 2a_2 B + \frac{15a_3 B^2}{4} \end{aligned} \tag{33}$$

Gain is no longer dependant on just a_3, as in (27), but a_2 as well. This means greater distortion occurs than for a constant rate. Computing the second and third order coefficients which delineate both the DC and fundamental distortion,

$$10\log_{10}\left(1 + \frac{3a_2 B}{2} + \frac{5a_3 B^2}{2}\right) = -1\text{dB} \qquad 10\log_{10}\left(1 + 2a_2 B + \frac{15a_3 B^2}{4}\right) = -1\text{dB}$$

$$a_2 = \frac{\left(\left(10^{-\frac{1}{10}}\right) - 1 - \frac{15a_3 B^2}{4}\right)}{2B} \qquad a_3 = \frac{\left(2\left(10^{-\frac{1}{10}}\right) - 2 - 3a_2 B\right)}{5B^2}$$

Substituting,

$$\Rightarrow a_3 = \frac{0.1645}{B^2} \qquad a_2 = -\frac{0.4113}{B} \tag{34}$$

These values hold for –1dB compression and are generalised by expressing in terms of n,

$$a_3 = \frac{0.1645}{B^2} = \frac{0.7143n}{B^2} \qquad a_2 = -\frac{0.4113}{B} = -\frac{1.7859n}{B} \tag{35}$$

Unlike the conventional polar case (27), a_2 is now negative and causing compression and a_3 is positive. The second and third harmonic distortion HD_2 and HD_3 can also be estimated:

$$HD_2 = \frac{harmonic\ amplitude}{a_1 B} = \frac{\left(\frac{a_2 B^2}{2} + \frac{3a_3 B^3}{2}\right)}{a_1 B} = \frac{a_2 B}{2} + \frac{3a_3 B^2}{2} = 0.0411 = 0.1786n \tag{36}$$

$$HD_3 = \frac{\left(\frac{a_3 B^3}{4}\right)}{a_1 B} = \frac{a_3 B^2}{4} = 0.0411 = 0.1786n$$

Substituting the coefficients (35) into the system polynomial (22),

$$y = x - \frac{1.7859n}{B}x^2 + \frac{0.7143n}{B^2}x^3 \tag{37}$$

With this general equation, the response can be determined for any input x, such as a sinusoid or data carrying OOK signal. The polynomial model with a unipolar signal is compared to the paralyzable DC model in Section 3.6.2, Fig. 3.10.

3.5.2 Detection Statistics

In this section, the model is developed to specify the detection statistics of a SPAD RX. As introduced above, distortion causes a reduction of the received SNDR which cannot be understood with prior constant rate models [108]. Consider a baseband signal $w(t)$ given by

$$w(t) = \sum_p a(p)s(t - pT) \tag{38}$$

where $a(p)$ is the information symbol sequence, p denotes symbol index, $s(t)$ is pulse shape and T is symbol period. $a(p)$ is a stream of randomly generated bits in Matlab. $s(t)$ sets duty cycle and rectangular NRZ signalling is considered, where the pulse amplitude is held constant throughout the symbol period: $s(t) = 1 \text{ for } 0 \leq t \leq T \text{ and } s(t) = 0$ otherwise. T is equal to the bit period since a binary alphabet is used (0 and 1 mapping). With sufficient SNR, higher-order modulation could be realised with M-PAM to increase net bit rate and spectral efficiency, where M is alphabet size. The signal is transmitted with an LD or light emitting diode (LED). Assuming an ideal lossless channel, let peak-to-peak amplitude be A, such that the received signal x is

$$x = Aw(t) \tag{39}$$

This can be separated into a DC component B_{DC} and AC swing B. A TX with finite ER means symbol rate affects AC swing. In contrast to a tone which corresponds to an impulse in the frequency domain, the NRZ spectrum

contains many frequencies enveloped with a sinc function. This makes it difficult mathematically to expand polynomial (15) with input (17), so analysis is shifted to the frequency domain in a simulation to investigate baseband gain and distortion. We develop a simulator to describe a SPAD RX with this instance and find a solution to the equation from the spectrum rather than the time domain. Considering Parseval's identity, which states that the energy of a signal $x(t)$ is conserved in temporal and spectral space [125]:

$$\int_{-\infty}^{\infty}|x(t)|^2dt = \int_{-\infty}^{\infty}|X(f)|^2df \tag{40}$$

where $X(f)$ is the Fourier transform of $x(t)$, the output response can be determined. Photon noise is introduced to the model, but standard Poisson statistics cannot be assumed because the output count distribution is distorted by the dead time. In the absence of dead time, ideal photon detections follow a Poisson process and the probability of counting k photons during a symbol interval $(0,T)$ is given by [115]:

$$p_0(k) = \frac{(\lambda T)^k e^{-\lambda T}}{k!} \tag{41}$$

where λ is instantaneous mean photon rate, hence λT is the average number of incident photons in T. λ is related to received optical power P_r by

$$\lambda = \frac{P_r PDE}{h\nu} \tag{42}$$

where PDE is SPAD photon detection efficiency; h is Planck's constant and ν is the frequency of the light. In the presence of dead time, however, distortion causes the counts to deviate from a Poisson distribution [126]. It is assumed that an event is counted from each SPAD pulse. Therefore, the total number of events in counting interval $(0,T)$ is obtained by the number of pulse transitions and cannot exceed $k_{max} = \left\lfloor \frac{T}{\tau} \right\rfloor + 1$, where $\lfloor z \rfloor$ denotes the largest integer that is smaller than z. The probability mass function (PMF) of a PQ SPAD during $(0,T)$ is expressed as [36]:

$$p_K(k) = \sum_{i=k}^{k_{max}-1} (-1)^{i-k} \binom{i}{k} \frac{\lambda^i (T - i\tau)^i e^{-i\lambda\tau}}{i!} \tag{43}$$

for $k < k_{max}$ and $p_K(k) = 0$ for $k \geq k_{max}$. i is integer index. PDE is defined as the product of PDP and fill factor C_{FF} and treated as a constant 8 dB loss (at 450 nm, $V_E = 1.3$ V and $C_{FF} = 43$ %) in this model. It can be seen that the bandwidth of the RX is also affected by τ, so performance is susceptible to both the intensity and frequency of the modulated signal. An array is employed to increase capacity and improve SNR and the output is given by the superposition of the detector elements. Eqn. (43) is applied for each SPAD and expanded to a

64 × 64 array of independent variables combined into a single process. The aggregate count distribution of the array is approximated by a Gaussian distribution [116]

$$p_X(x) \sim \mathcal{N}(\mu_X, \sigma_X^2) \tag{44}$$

where μ_X and σ_X^2 are the mean and variance of the aggregate distribution. Numerical methods are developed to determine the achievable BER for OOK modulation. Although τ remains the main limiting factor on RX performance, noise causes BER to increase further. Dominant noise sources are background events from dark counts, afterpulsing, ambient light and ER. ER becomes the principal noise factor at high symbol rates and appears as a DC offset. Furthermore, feedforward ISI arises when $\tau \geq T$, since previous counts from a logical '1' may disperse in time into a subsequent '0' and therefore add to the '0' distribution. These overflow events are almost unnoticeable in the '1' distribution, especially if the mean is large, so ISI is most prevalent for 1→0 symbol transitions. Let λ_s and λ_n be the mean rates from signal and background noise, respectively.

The average rates per element are $\frac{\lambda_s}{N_{SPAD}}$ and $\frac{\lambda_n}{N_{SPAD}}$, where N_{SPAD} is the number of SPADs in the array. When a '0' bit is transmitted, the average number of counts per symbol is $\mu_0 = \lambda_n T$, and when a '1' is transmitted, the average is $\mu_1 = (\lambda_s + \lambda_n)T$, where $\lambda_s T$ is mean signal counts per symbol. We obtain $p_0(x)$ and $p_1(x)$, probabilities that x photons are detected in the counting interval T, when '0' or '1' are sent:

$$p_0(x) \sim \mathcal{N}(\mu_0, \sigma_0^2) \qquad\qquad p_1(x) \sim \mathcal{N}(\mu_1, \sigma_1^2) \tag{45}$$

with variances σ_0^2 and σ_1^2. Decoding is implemented by comparing the received counts to a threshold x_T [116],

$$x_T = \frac{\mu_1 \sigma_0 + \mu_0 \sigma_1}{\sigma_1 + \sigma_0} \tag{46}$$

An error occurs if $x \leq x_T$ when a '1' bit is sent and, vice versa, if $x > x_T$ when a '0' is sent. Probability of error is equal to BER for OOK and, assuming equiprobable symbols, is

$$BER \cong Q\left(\frac{\mu_1 - \mu_0}{\sigma_1 + \sigma_0}\right) = Q\left(\sqrt{SNDR}\right) \tag{47}$$

where $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{\alpha^2}{2}} d\alpha$ is the Q-function. Independent statistics are assumed for each transmitted bit and it is assumed that the array elements are identical. Eqn. (47) is used to estimate BER in a nonlinear SPAD system in Chapter 5 Section 5.4.2 (Fig. 5.16).

3.6 Transient Simulation

3.6.1 Methodology

A simulation model is developed to estimate the behaviour of a SPAD RX and I compare this to the mathematical framework in Section 3.5. Matlab is used to simulate the stochastic events that occur in the photon counting system, expanded from imaging sensors [127]. With statistical analysis (Eqn. (43)), it is shown that the counting process does not follow a Poisson distribution.

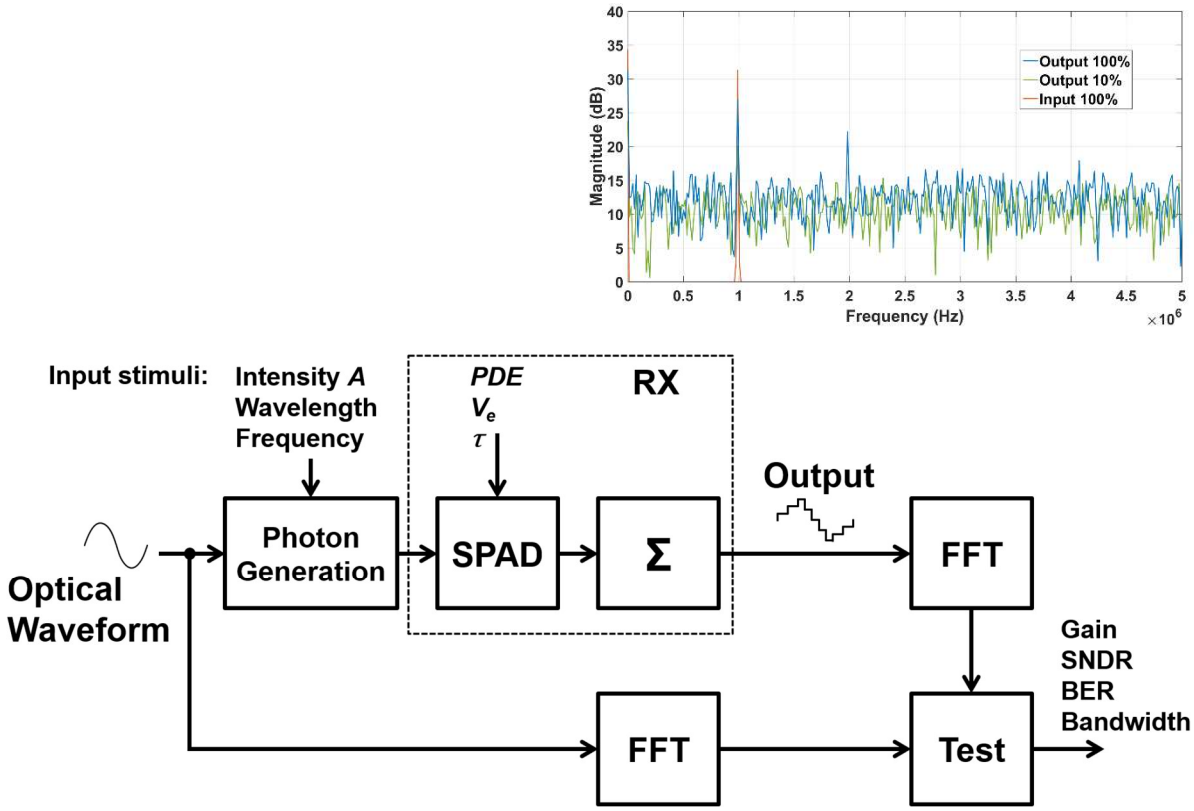

Figure 3.8: Simulation methodology. Input and output spectra for a unipolar 1MHz cosine input at 10 % and 100 % of maximum intensity (top right).

Fig. 3.8 displays a block diagram of the simulation method. As in Eqn. (31), a noiseless sinusoidal input signal is generated with an instantaneous optical power at each sample. For a selected time window, the instantaneous mean photon rate is distributed in time with a Poisson process. The simulator generates a stream of photons, represented by a discrete sequence of ones and zeros corresponding to photons or no photons in a given time step. This is the input to an algorithmic block I designed which describes the statistical detection and integration process of a PQ SPAD RX. The physical parameters of the SPAD including dead time, PDP and V_E are coded in this block.

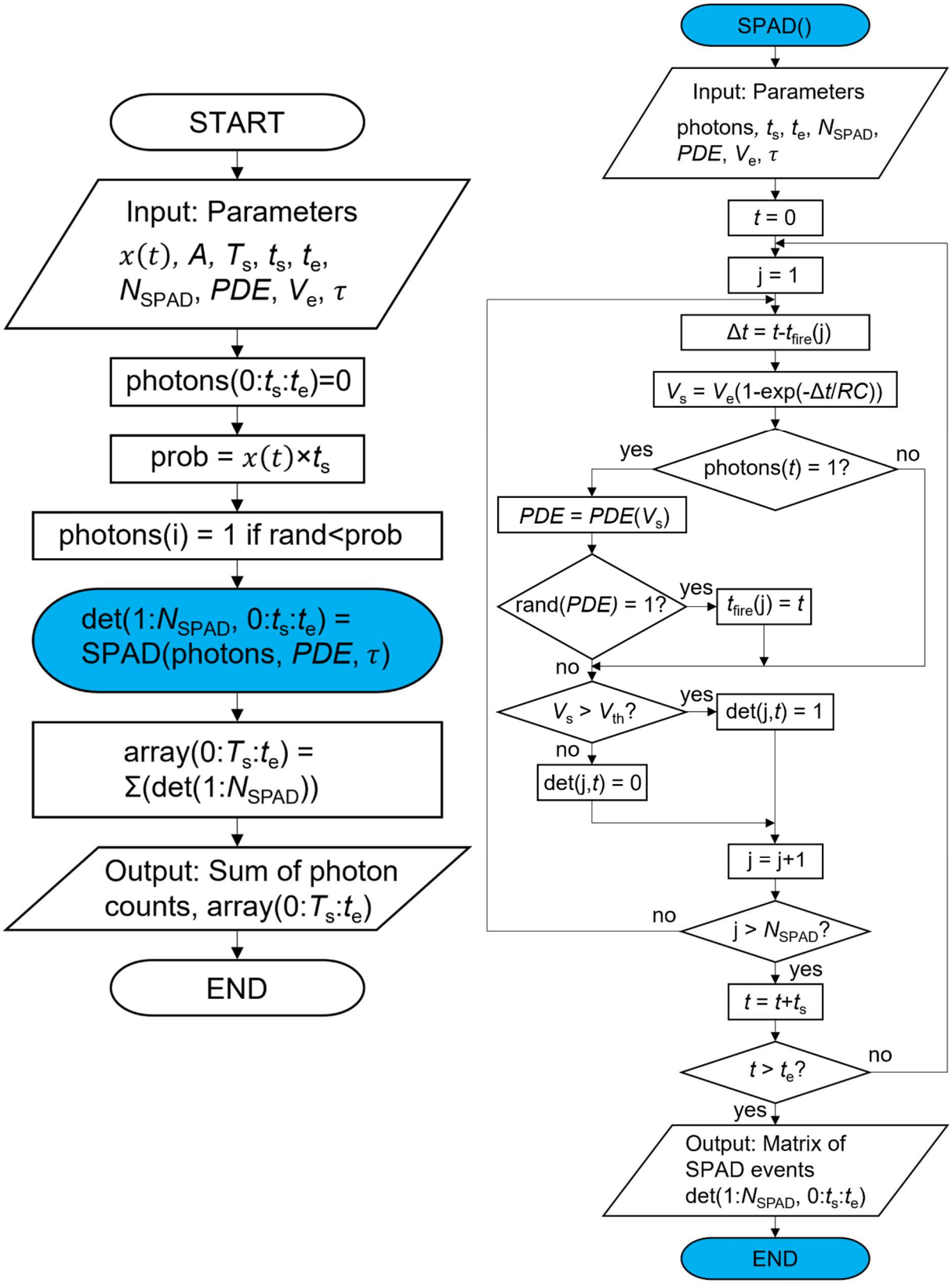

Figure 3.9: Algorithm for an integrating PQ SPAD RX with N_{SPAD} array, T_s sampling period, t_s simulation step time, and t_e end time (length of the data frame). Photon impulses are distributed in time according to a probability (prob) and the instantaneous photon rate A. $RC = -\frac{\tau}{\ln\left(1-\frac{V_{th}}{V_e}\right)}$ is the time constant of a single SPAD.

Fig. 3.9 shows a flowchart of the algorithmic block. Each photon event is simulated individually, and the output is expressed as a digital vector where every rising and falling edge represents a detected photon. This is the output of a SPAD paired with a buffer which can be seen as a 1-bit ADC. The block is initialised in a rest state at zero.

The output of the algorithmic block is the time domain response and vectors are run for each SPAD in the array and combined with an 800 MHz counter implemented in Matlab to recover the modulated signal. Sampling rate is much higher than τ, so it can be assumed that counting losses arising from finite sampling rates are negligible. The signal is then analysed in the frequency domain by taking a fast Fourier transform (FFT) and distortion metrics are obtained. The RX output is directly compared to its input waveform with respect to count rate and frequency and SNDR is calculated from the signal, noise, and harmonic power [128].

3.6.2 Sinusoidal Response

Using the simulator, the response of a PQ SPAD RX is estimated. Fig. 3.8 shows input and output spectra for a unipolar 1 MHz cosine input signal at 10 % and 100 % of the maximum peak-to-peak rate $A = 8.33\times10^7$ s^{-1}. Gain is determined from the difference between the input and output tones. 7 dB output SNDR is computed from the output signal power and noise floor at 10 % intensity, whilst arrival SNR = 9 dB. A second harmonic (2 MHz) appears at maximum input intensity and DC (–3 dB) and signal gains (–5 dB) are found from subtracting the corresponding magnitudes. At this intensity, SNDR = 4.7 dB, arrival SNR = 19 dB. Fig. 3.10 shows the output derived from Eqn. (37) together with simulated DC and signal gains found for different values of n. These estimates are plotted against the standard paralyzable model in Fig. 3.10. The corresponding gain is computed from the quotient of the output response calculated in Eqn. (37) and the input. It is concluded that higher distortion occurs with a unipolar waveform than for a constant incident level, which results in SNDR degradation.

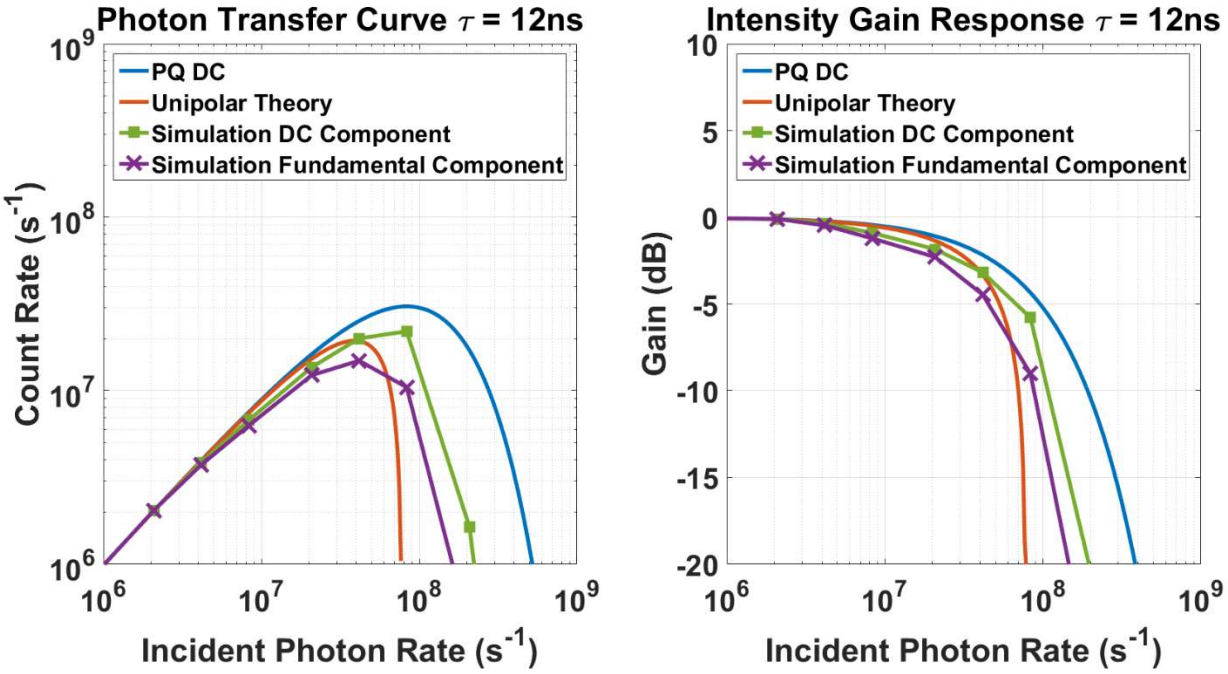

Figure 3.10: Comparison between the paralyzable SPAD model (with DC rate A equal to the peak-to-peak amplitude of the unipolar waveform), analytical prediction and simulation results.

Near matching between unipolar theory (Section 3.5.1, Eqn. (37)) and simulation indicates a reasonably accurate numerical estimate. Simulation accuracy diverges for extremely low event rates due to random noise generation; however, moderate precision is observed around the intensity range (1 % to 100 % of capacity) the RX would nominally operate in for VLC. Earlier inflection occurs at approximately $n = 0.32$ in theory and simulation rather than $n = 1$ as in the DC model.

3.6.3 BER Estimation

The photocount distribution of a SPAD as described in Eqn. (43) is simulated and plotted in Fig. 3.11 with mean incident rate $A = \lambda = 8.33\times10^6$ s^{-1}, a decade below the maximum of 8.33×10^7 s^{-1}. Symbol rate R is increased from 1 MBd ($T = \frac{1}{R} = 1$ µs) to 10 MBd ($T = 100$ ns). $AT = 8.33$ photons/symbol for a logical '1' bit at 1 MBd, which ensures non-zero integer photon arrivals within a symbol period in the simulation – although SNR is insufficient for 3.5×10^{-3} BER reception. The average number of events per symbol is inversely proportional to R. Poisson distribution at 1MBd also plotted (blue) for an ideal RX without dead time. It can be confirmed that the PMF expression (43) simplifies to a Poisson form with mean λT when τ is set to zero. At 1 MBd, the '1' level mean is 7 counts/symbol, which is slightly lower than the arrival mean, and this translates to –0.757 dB average

loss. This continues with means of 3 counts/symbol, 1 count/symbol and 0.8 counts/symbol at 2 MBd, 5 MBd and 10 MBd respectively. Thus, loss is approximately unchanged and there is minor nonlinearity at this incident power. In addition, variance diverges to 2 counts/symbol, 0.9 counts/symbol and 0.5 counts/symbol at 2 MBd, 5 MBd and 10 MBd respectively, whereas arrival variance is 8.33 ($\sqrt{8.33}$ standard deviation). As R increases, the distribution distorts and tends towards a unit Dirac impulse at 0 counts. Hence, the PMF approaches a point where the SPAD is persistently in a recovery state and no following photons are detected.

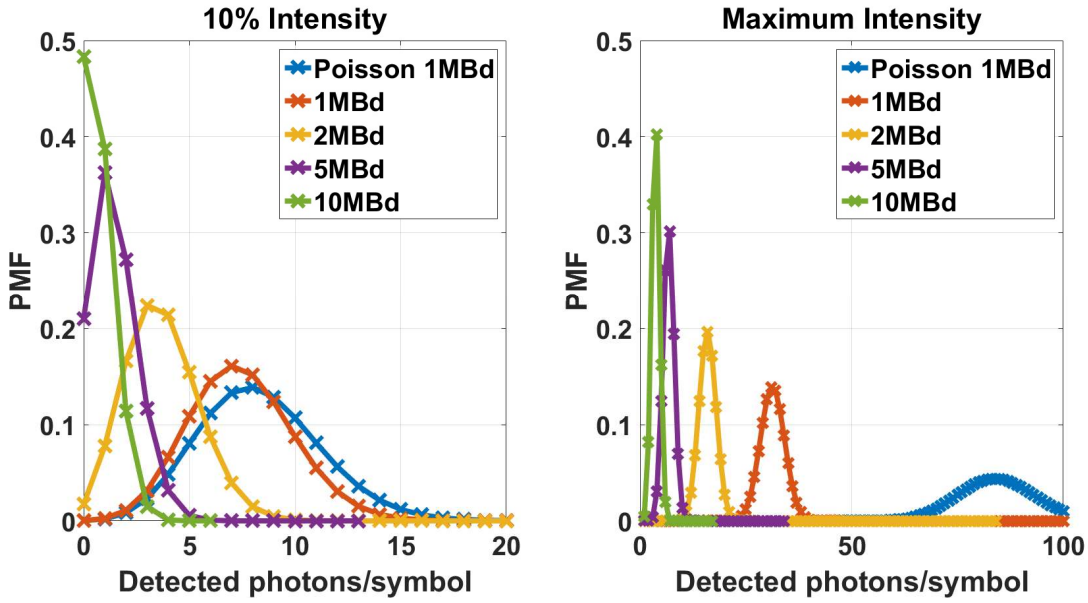

Figure 3.11: Simulated photocount distributions of a SPAD for $\tau = 12$ ns, fixed mean $A = 8.33\times10^6$ s^{-1} (a decade below the maximum incident rate) on the left and $A = 8.33\times10^7$ s^{-1} on the right. $R = 1$ MBd to 10 MBd. 1 MBd Poisson distribution without dead time (blue) for reference.

Similar analysis is carried out at maximum intensity $A = 8.33\times10^7$ s^{-1}. At 1 MBd, '1' mean is 31 counts/symbol, –4.29 dB distortion loss from $AT = 83.3$ photons/symbol arrival mean. The distribution deviates from Poisson form by –4.29 dB, –4.16 dB, –3.77 dB and –3.19 dB with respect to the corresponding photons/symbol at each R. SNDR = 8 dB at 4 Mb/s from Eqn. (47), so it is deduced that this is the maximum OOK bit rate of a single SPAD with these parameters to sustain the 3.5×10^{-3} BER threshold required for forward error correction (FEC) to achieve an output BER of 10^{-9} using concatenated Bose-Chaudhuri-Hocquenghem (BCH) inner and outer codes with 6.69% overhead [19].

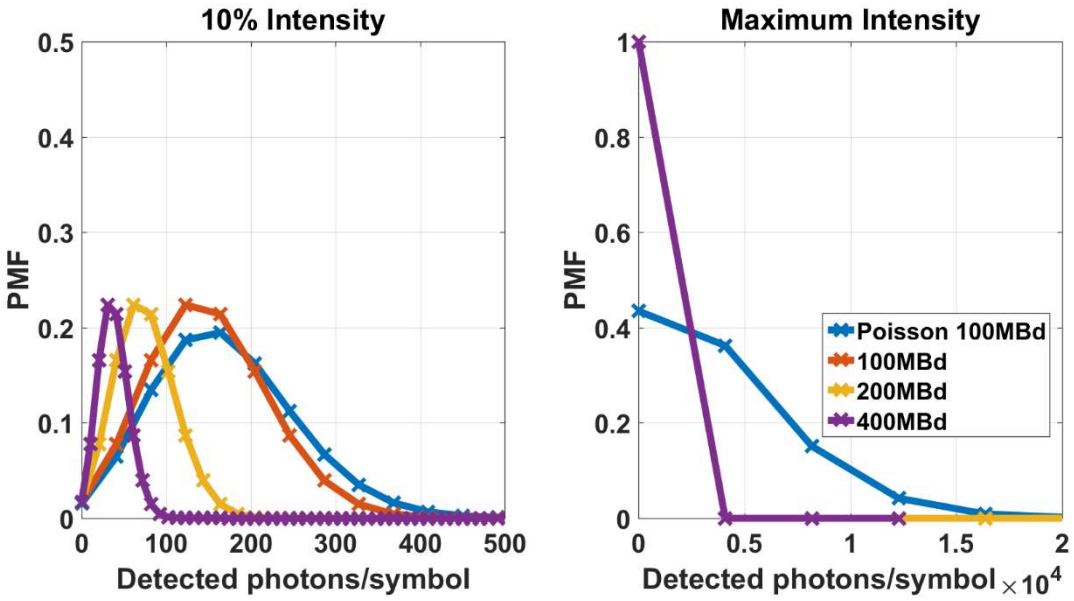

Figure 3.12: Simulated PMF count distributions of a 64 × 64 array for $\tau = 12$ ns, $A = 8.33\times10^6$ s^{-1} (left) and maximum $A = 8.33\times10^7$ s^{-1} (right) at $R = 100$ MBd to 400 MBd.

Fig. 3.12 shows simulated PMF distributions of a 64 × 64 array with $\tau = 12$ ns, $A = \lambda = 8.33\times10^6$ s^{-1}, $C_{FF} = 0.43$ and $R = 100$ MBd to 400 MBd (100 Mb/s to 400 Mb/s NRZ). 400 MBd is likely to be the highest baud rate achievable by the SPAD RX in practice. Since the number of array elements is large, the output count distribution approaches a Gaussian form (45), as according to the Central Limit Theorem [116]. SNDR is calculated from Eqn. (47) with 5.78 ER (1.5 dB penalty) added to the model. PMFs shown (Fig. 3.12 right) with the same parameters, but at maximum intensity $A = 8.33\times10^7$ s^{-1}. 100 MBd and above results in complete saturation of the array, with a certainty of zero counts per given symbol period, without any chance of a SPAD having time to recover to detect the next symbol. At these rates, the output process is approximated by a Bernoulli distribution, as explained in [133], since T is shorter than the τ of each SPAD.

3.7 Summary

A Matlab model describing the behaviour of a SPAD-based OWC system is established to estimate the effect of dead time on the reception of modulated signals. Both a single SPAD detector and an array are considered. The polynomial model developed in Section 3.5.1 shows that the paralyzable DC model significantly overestimates the response of a SPAD RX to a unipolar signal (Section 3.6.2, Fig. 3.10). Near matching between the polynomial model (Eqn. (37)) and the frequency domain simulation results in Fig. 3.10 reinforce the accuracy of the model. However, these two methods should be validated with real SPADs, and Chapters 4 and 5 cover

implementation and experimental results, respectively. From the RX response modelling, it is shown that dead time is the main limiting factor on performance, as other power losses can be accounted for with closed-form expressions and easily known constants. Dead time boundaries are quantified and combined with statistics to estimate SNDR and effective bandwidth along with achievable BER and bit rate performance metrics. It is concluded that the effect of dead time is a key design constraint that can now be predicted and mitigated in SPAD RXs.

4. Design of a Digital SPAD RX

4.1 Objectives

This chapter describes two existing devices used to verify the simulation model developed in Chapter 3: a 40 nm CMOS SPAD IC, with full access to all SPAD characteristics, and a 130 nm CMOS parallel summing SPAD array IC used to demonstrate integration and high-order complex modulation schemes and maximise bit rate. These two ICs establish a foundation for the implementation of a DSiPM that approaches the performance of an ideal photon counting RX bound by the QL. The 1 GHz and 800 MHz sampling rates are the main limitations of the ICs.

4.2 A 40 nm CMOS SPAD Array

4.2.1 Architecture

A 40 nm CMOS SPAD array chip [47], with various SPAD pitch options from 8 µm to 30 µm, is used in this work. Fig. 4.1 shows a TCAD simulation of the doping density and the electric field profile of the SPAD device on the chip. A PWELL and a deep NWELL (DNWELL) implant form the multiplication region optimised for DCR and PDE through junction engineering, maintaining a low breakdown voltage of 15.5 V at 60 °C across wafer production. The guard ring is formed by retrograded epitaxial layer (EPI) and shallow trench isolation.

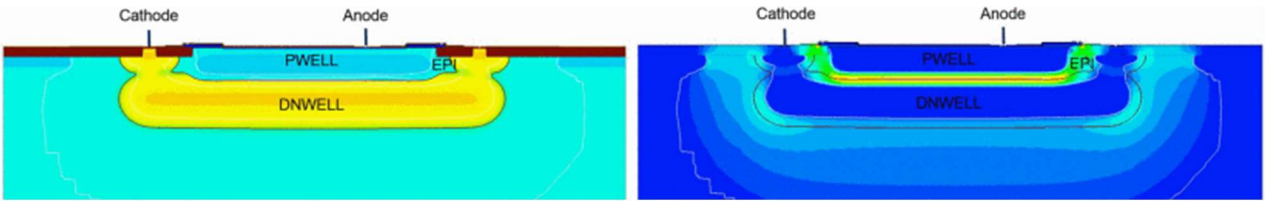

Fig. 4.1: Left: doping density of the SPAD. Right: SPAD electric field profile. PWELL, deep NWELL (DNWELL) and anode and cathode contacts annotated [109].

Dedicated microlens fabrication on top of the SPADs is integrated in the identical technology node, which enhances the fill factor to >70 % at SPAD level and around 40 % at pixel level. The small digital node allows more logic to be integrated inside the pixel, which is portable to a 3D stacked technology, where logic is realised in a fully digital layer [109]. The anode contacts are positioned to ensure track length uniformity, pixel symmetry,

and to minimise optical losses. The SPAD array layout allows crosstalk measurements to be taken in all directions.

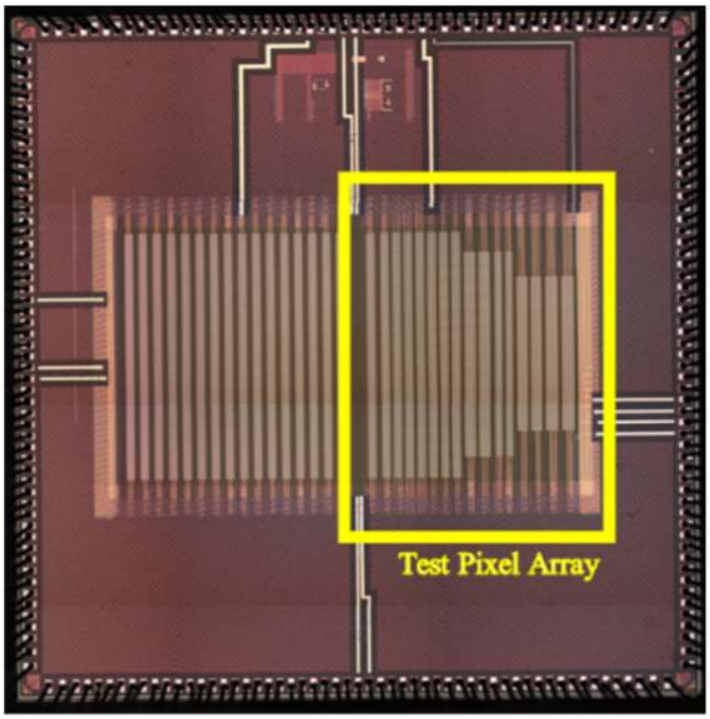

Figure 4.2: 40 nm chip micrograph (test pixel array highlighted) [47].

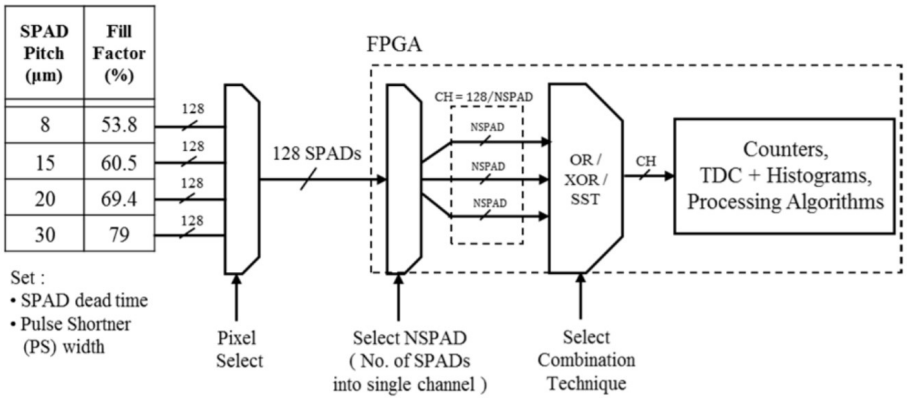

Figure 4.3: Interface and configuration options available with the chip & FPGA setup [47].

128 PQ SPAD outputs sampled at 1 GHz are conveyed off-chip simultaneously to a Xilinx Kintex-7 FPGA [47]. Each output pulse corresponds to an avalanche caused by a photon detection, dark count or parasitic count and has a minimum pulse width equal to the dead time during which the SPAD is blind. In this work, 112 SPADs with 15 µm pitch and a fill factor 60 % are used for characterisation – the remaining 16 SPADs on the die used for these experiments were defective. 15 µm is chosen as a middle ground between active area, fill factor and DCR.

The median DCR at room temperature is 50 s^{-1} per SPAD. Fig. 4.2 shows a micrograph of the chip. The number of SPADs enabled and the logic to combine them into a single channel can be selected on the FPGA (Fig. 4.3). The configuration options include an OR tree, OR tree with pulse-shortener (PS), XOR tree, and a novel Synchronous Summation Technique (SST) intended for automotive LiDAR applications, which require high dynamic range to image moving objects with varying speeds, distances, and reflectivity [47]. SST improves dynamic range by 30× compared to OR, 15× compared to PS OR and 7.5× compared to XOR. A larger array of SPADs will provide more counts at the expense of more readout bits. In addition, a multiple-event histogramming time-to-digital converter can be implemented on the FPGA.

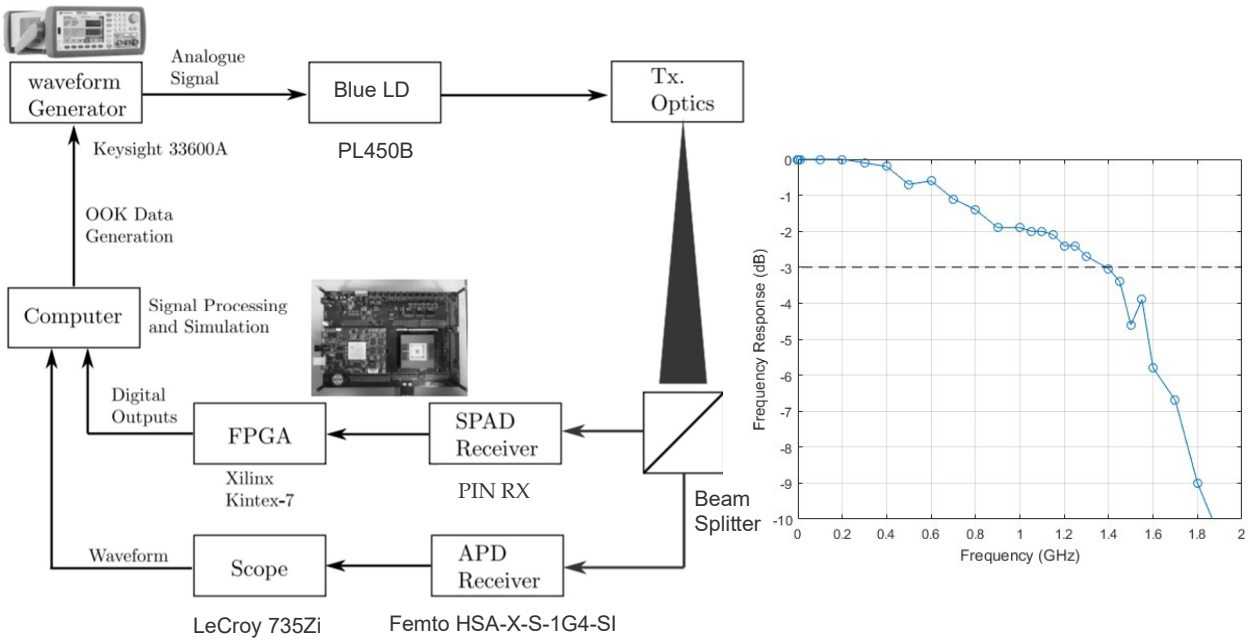

Figure 4.4: Block diagram of the experiment setup (left). Frequency response of the LD+PIN RX (right).

The digital output from the FPGA is sent to a PC for offline processing in Matlab, where the rising edge of each pulse is counted for each SPAD. The counts are combined to form a summation from the array. This is equivalent to an OR tree combination technique with an ideal pulse shortener. There are fewer counting loses with this method compared to practical combining networks such as OR and XOR trees [112]. This allows the dead time of individual SPAD detectors to be characterised in digital and in more detail for communication purposes than before. This readout technique could be implemented with edge detection, counters, and an adder on the FPGA in a similar way to SST proven in [47].

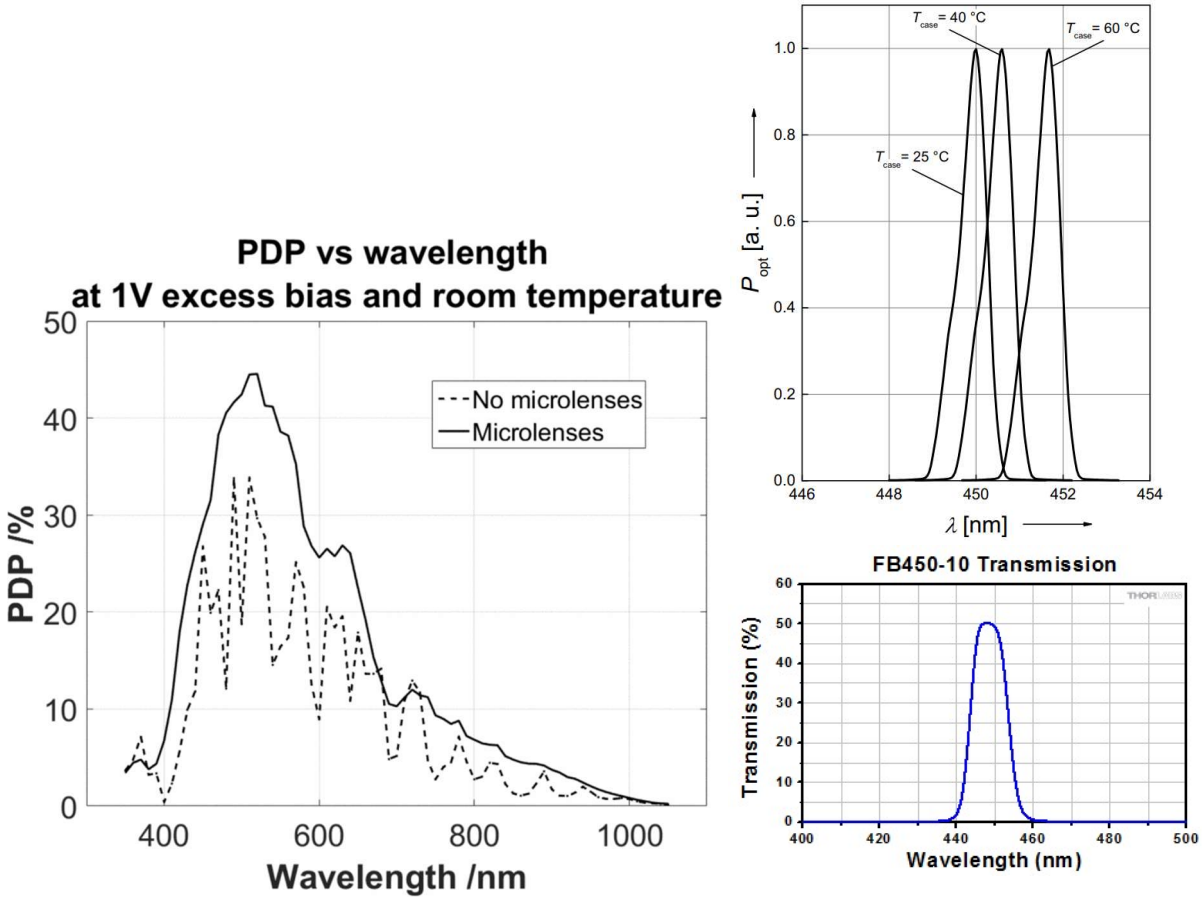

Figure 4.5: PDP spectrum of the 40 nm SPAD – dotted line is without microlenses; solid line is with microlenses. Top right: relative spectral emission of the LD. Bottom right: transmission of the 450 nm optical filter [109].

4.2.2 Methodology

A diagram of the experiment setup is illustrated in Fig. 4.4. A 450 nm blue LD (Osram PL450B) is driven by a Keysight 33600A waveform generator which generates a 50-100 Mb/s unipolar OOK NRZ or RZ signal. An LED could also be used for this experiment because there is a small number of SPADs and therefore the symbol rate is limited. Fig. 4.4 also shows the frequency response of the LD measured with a Si PIN reference RX (Femto HSA-X-S-1G4-SI) and LeCroy 735Zi oscilloscope by sweeping the frequency of a sinusoid from a fast waveform generator (Agilent 81180A). The PIN is specified with bandwidth of 1.4 GHz, and this is the –3 dB point in Fig. 4.4. Taking the –6 dB point in the frequency response to account for the PIN RX, the LD bandwidth is over 1.6 GHz. The LD drive current is 50 mA. Fig. 4.5 shows the spectral emission of the LD (top right) and the PDP spectrum of the SPAD used on the 40 nm chip with and without microlenses (left). The microlenses improve

PDP by smoothing the large oscillations associated to interferences in the optical stack and by concentrating the beam into the centre of the SPAD, allowing more signal to be collected. PDP is calculated relative to the active area of the SPAD. The introduction of microlenses increases the average PDP and lets all metal layers on the chip be retained (allowing for high digital gate density) without affecting optical performance.

The PDP of the SPADs at 450 nm is 20 %. This detection probability value is low and results in a high power penalty of $PP_{PDP} = -10\log_{10}\left(\frac{20}{100}\right) = 6.99$ dB (Chapter 3 Section 3.1) but not a concern in this case because the experiment is focused on investigating the deterministic effect of dead time. Additionally, with mircolenses, the PDP is 31 % and therefore the penalty can be reduced to $PP_{PDP} = -10\log_{10}\left(\frac{31}{100}\right) = 5.09$ dB. This PDP is still low compared to typical values in literature (32 % to 70 % [57]). A 450 nm optical bandpass filter (Thorlabs FB450-10 with 10 nm full width half maximum (FWHM)) is mounted on the active area of the SPAD IC. The filter transmission is also shown in Fig. 4.5. It is assumed that the 55 % transmission loss of the filter at 450 nm is counteracted by the nonideal spectral profile of the LD and filter, allowing a band of wavelengths to be captured with a total of 0 dB loss in the link budget.

After synchronisation, the total number of photon detections in each sample is obtained and integrated to counts per symbol period. BER performance is evaluated with three traditional demodulation schemes: exhaustive search (ES), a linear equalizer, and a DFE equalizer (chosen in line with prior literature which showed DFE to be very effective at overcoming ISI in SPAD RXs [32, 33, 39, 40]). The DFE equalizer is implemented in Matlab with the recursive least squares (RLS) algorithm and feedforward and feedback weights of 41 and 15, respectively. RLS is chosen for its extremely fast convergence compared to other algorithms such as least mean squares (LMS) [48]. However, it should be noted that this benefit comes at the expense of higher DSP complexity.

4.3 A Large Parallel 130nm CMOS SPAD Array RX

4.3.1 Architecture

This work reports a 2.8 mm by 2.6 mm SPAD RX integrated in 130 nm CMOS imaging technology which incorporates 64 × 64 receiver elements at 21 μm pitch. The large array provides high dynamic range and SNR intended for high order modulation. Each element contains a single p-well/deep-n-well SPAD biased at 15.2 V (1.3 V excess bias V_e) with a mean dead time of 12 ns, median dark count rate of 6 kHz at room temperature

and PDP of 37 % at 450 nm. This corresponds to a penalty of $PP_{PDP} = -10\log_{10}\left(\frac{37}{100}\right) = 4.32$ dB. The fill factor is 43 %, which is a $PP_{FF} = -10\log_{10}\left(\frac{43}{100}\right) = 3.67$ dB penalty. Therefore, the total detection loss is the sum of these two penalties: 7.98 dB. 1.34 × 1.34 mm active area is chosen for ease of alignment to POF or VLC optics but can be adjusted electronically by disabling areas of detector elements. Fig. 4.6 is a schematic of the chip architecture.

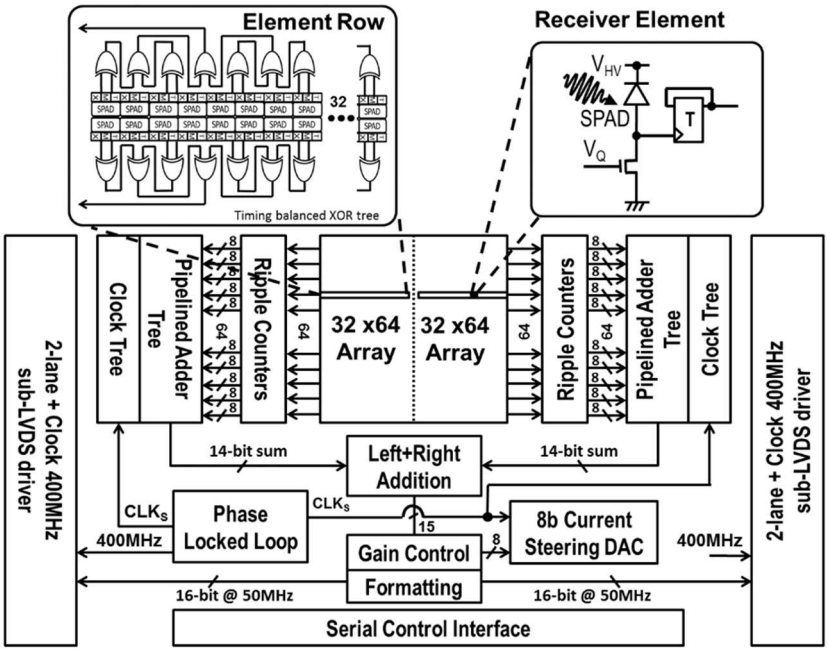

Figure 4.6: Simplified chip block diagram.

A receiver element comprises of a SPAD interfaced to an NMOS passive quench, enable SRAM and toggle-flop. The toggling output encodes photon events on both rising and falling edges. 32 elements are combined with a row XOR tree into an asynchronous double data rate (DDR) sequence at up to 900 MHz (1.8 Gphotons/s) limited only by wiring parasitics (not SPAD dead time).

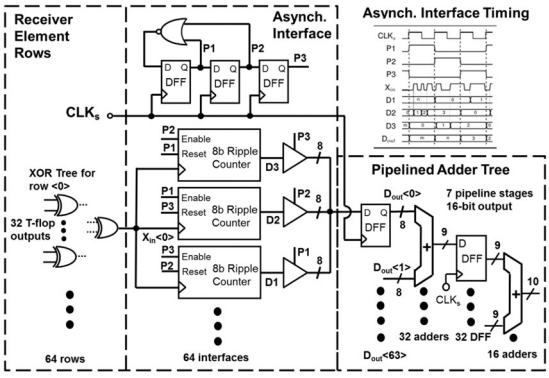

Figure 4.7: Asynchronous interface between (right) SPAD XOR tree and pipelined added tree. Inset: interface timing diagram.

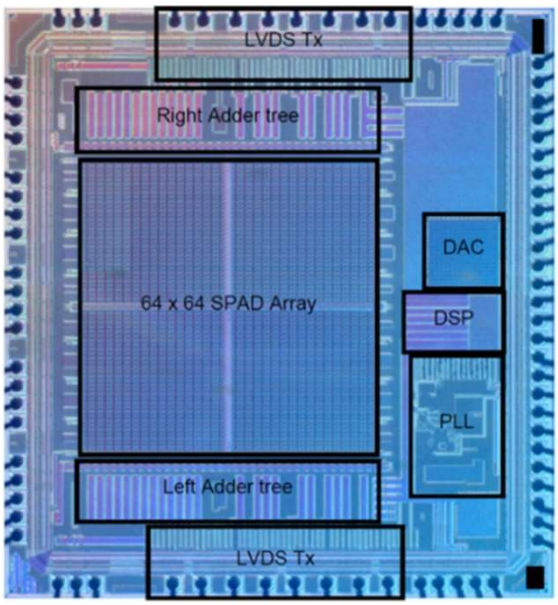

Figure 4.8: Annotated 130 nm chip micrograph.

The detector array divides into two sets of 64-row XOR trees feeding digital readout chains positioned on the flanks of the active area. Fig. 4.7 shows the row parallel interface circuit, consisting of three 8-bit ripple counters, sampling and converting the asynchronous DDR row signal into a synchronous binary count without dead time. The three counters operate in a round robin fashion, so that in every clock cycle one counter is reset, one is being read out and one is counting. A local state machine rotates continuously through the counters. The 128 sets of row counters operate in parallel and their outputs are added through a pipelined 7-stage adder tree to give an overall, 16-bit synchronous sum of SPAD events. This means that the device operates as a photon integrating RX which produces an output signal that can then be sampled or integrated before demodulation.

The entire digital readout operates from a sample clock generated by an on-chip PLL with programmable frequency up to 800 MHz and distributed through clock trees to the pipelined adder. This digital readout replaces the TIA, analogue signal conditioning and ADC chain of conventional PD/APD-based RXs and will scale favourably to advanced nanometre process nodes. It also lends itself to the integration of the DSP required for complex modulation schemes. After digital gain control and formatting, the 16-bit summation is conveyed off-chip using two D-PHY transmission blocks, each dual lane, with line rates of 400 Mb/s, for a total off-chip bandwidth of 1.6 Gb/s. Fig. 4.8 is an annotated micrograph of the manufactured chip. In addition to the D-PHY blocks, an 8-bit 400 MHz bandwidth current-steering digital-to-analogue converter (DAC) generates an analogue output compatible with existing ADC-based RX paths for testability. The total dynamic power consumption of the RX is 115 mW, comprising of 15 mW of SPAD power and 100 mW of DSP.

4.3.2 Methodology

I generated sinusoids, and OOK and 4-PAM waveforms, from a pseudorandom binary sequence (PRBS) via a Mersenne Twister with a pattern length of 2^{19937}–1 and non-return-to-zero (NRZ) line coding. RRC pulse shaping is also applied and DCO-OFDM. These signals are transmitted with an Agilent 81180A Arbitrary Waveform Generator (AWG) driving an LD (Osram PL450B, 450 nm). 450 nm centre wavelength is chosen because it close to the wavelength at which the SPADs have maximum PDP. The AWG sampling rate is 4 GS/s and the LD is collimated onto the SPAD RX at 1 m. A 450 nm bandpass filter (Thorlabs FB450-10, 10 nm FWHM) and two neutral density filters (31 % and 10 % transmission) are placed above the array to attenuate ambient light and LD power, respectively.

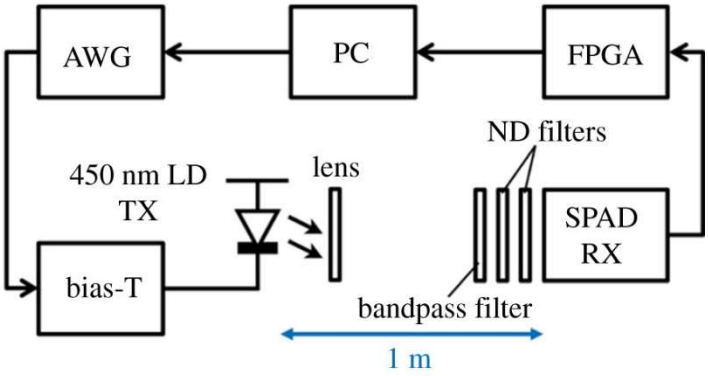

Figure 4.9: Measurement setup.

After frame synchronization, decoding of the received waveform is performed offline. Fig. 4.9 shows a block diagram of the setup and Fig. 4.10 shows an annotated photograph. Experiments are conducted with the source directly in front of the RX to minimise reflections in the channel.

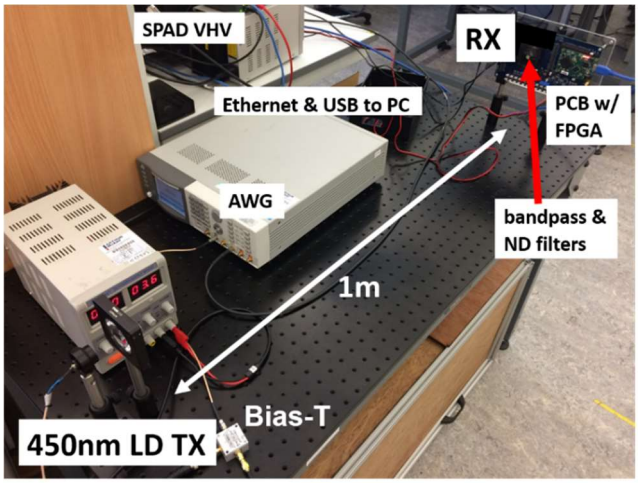

Fig. 4.10: High speed photon counting RX measurement setup.

4.4 Summary

The 40 nm CMOS IC has digital readout via an FPGA, which allows data to be analysed in Matlab. Physical properties such as dead time, PDE, and afterpulsing can be measured with individual SPADs. In Chapter 3, it was shown that these properties degrade the link budget of a SPAD-based system, and the next step is to validate the model estimates with this hardware. In addition, the chip technology enables characterisation of crosstalk between SPADs in the array by measuring photon timing information with output pulses directly from the SPADs. This provides detail and timing precision not possible with an ASiPM. The IC will be used to implement a digital SPAD RX by integrating the photon counts per sample into symbol bins. Since the aggregate counts are combined in software, the performance of the 40 nm chip is expected to be closer to theory relative to practical SPAD RX implementations which combine counts on-chip [28, 37-39] because there is no loss of counts in the combining logic. This is justified because the chip is not intended to be used as a full-scale RX, but instead as a test platform for characterisation. Instead, the 130 nm chip takes this role by providing a fast readout of combined SPAD counts via several parallel DDR toggling XOR trees. The 16-bit aggregate output enables up to 2^{16}-1 = 65535 counts conveyed off-chip in each sampling period. Each row can go up to 1.8 Gphotons/s, giving a total of 230 Gphotons/s with all 128 rows. Sampling at 800 MHz, this means the architecture has a maximum SNR of 23 dB per sample in theory (Eqn. (47)).

5. Experimental Validation of The Model

5.1 Objectives

In this chapter, the physical limitations of SPAD detectors are characterised for RX implementations. BER and bit rate performance metrics are evaluated for the 40 nm and 130 nm SPAD ICs with various modulation schemes, including NRZ, RZ, PAM and DCO-OFDM, and different demodulation techniques: either exhaustive search or digital (linear and nonlinear) equalization in Matlab.

5.2 SPAD Characterisation

5.2.1 Dead Time

Fig. 5.1 shows the photon transfer curve of a single PQ SPAD measured on the 40 nm chip against a linear light level sweep with the LD. These measurements are compared to a simulation with the paralyzable model (Chapter 3 Section 3.2). The maximum DC count rate is 17.5 Ms^{-1}, which corresponds to a dead time of 21 ns.

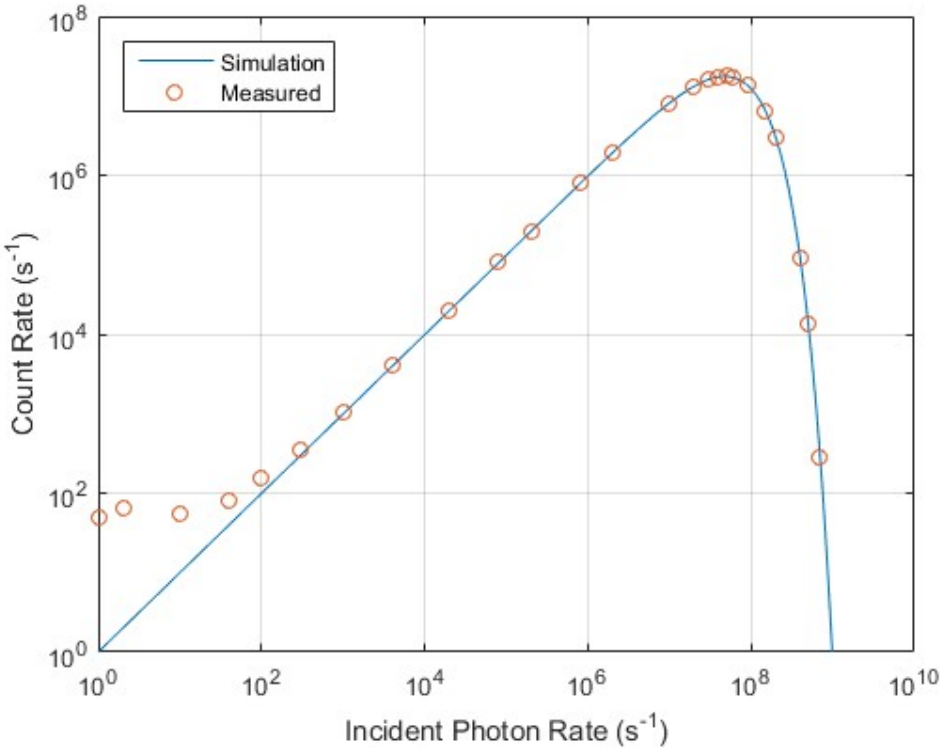

Figure 5.1: Single SPAD photon transfer curve. $\tau = 21$ ns.

It can be seen in Fig. 5.1 that the SPAD has a dynamic range of around 5 orders of magnitude (50 dB) from the DCR floor (50 s^{-1} median) to saturation. The PDFs of the 112 SPAD pulse widths at high and low intensities are shown in Fig. 5.2.

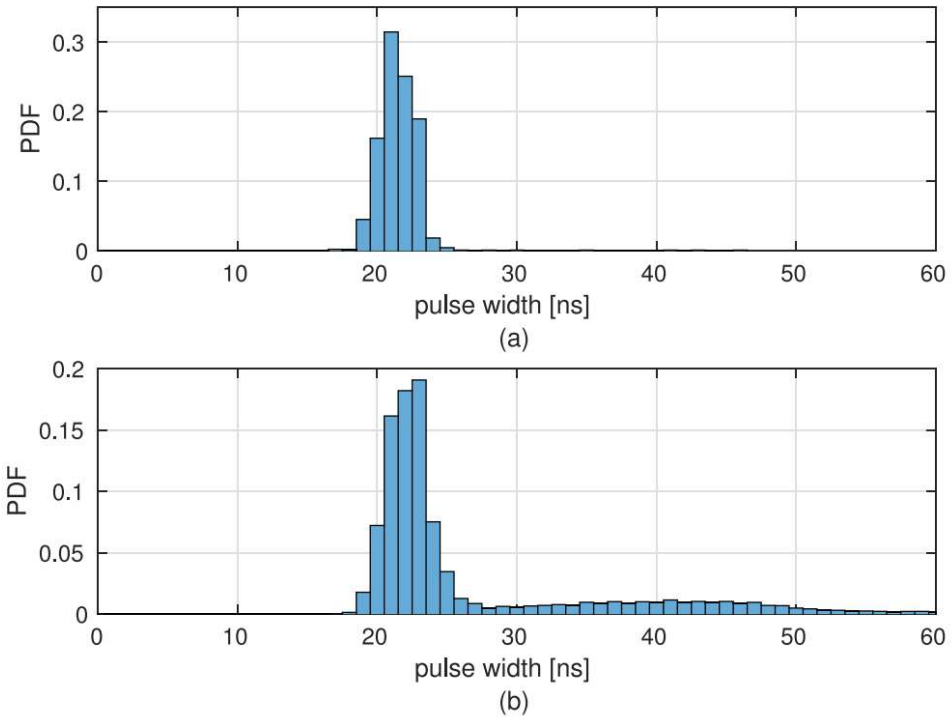

Figure 5.2: PDF of 112 SPAD output pulse widths at (a) 5 % light intensity and (b) 80 % light intensity.

At low light intensity (Fig. 5.2(a)), the mean pulse width is around 21 ns, which is therefore considered to be the mean dead time of the SPAD array. Longer output pulses are seen at higher optical power (Fig. 5.2(b)) because any photon detection during the dead time extends the dead time with a PQ SPAD. The quench voltage can be adjusted to reduce the mean dead time to 10 ns.

5.2.2 Parasitics

Afterpulsing probability is measured for each SPAD in the array by plotting a histogram of the inter-arrival times of SPAD pulse rising edges and integrating the portion above the ideal Poisson curve at short time intervals. A straight line is projected to 0ns and the percentage above the line is determined. Fig. 5.3 shows count inter-arrival time histograms for SPAD No. 1, 2, 14 and 32 localised in the same array column. The average APP per SPAD in the array is found to be around <0.1 %. The maximum number of samples from the chip that can be held in the output buffer is 10^5 in the 1GHz sampling mode, so the measurement acquisitions are iterated multiple times in a loop and added to make a histogram with over 10^6 pulse inter-arrival times. The dead time can be

seen in Fig. 5.3 as a null in the PDF around 10 ns inter-arrival time: the minimum possible time between two pulses.

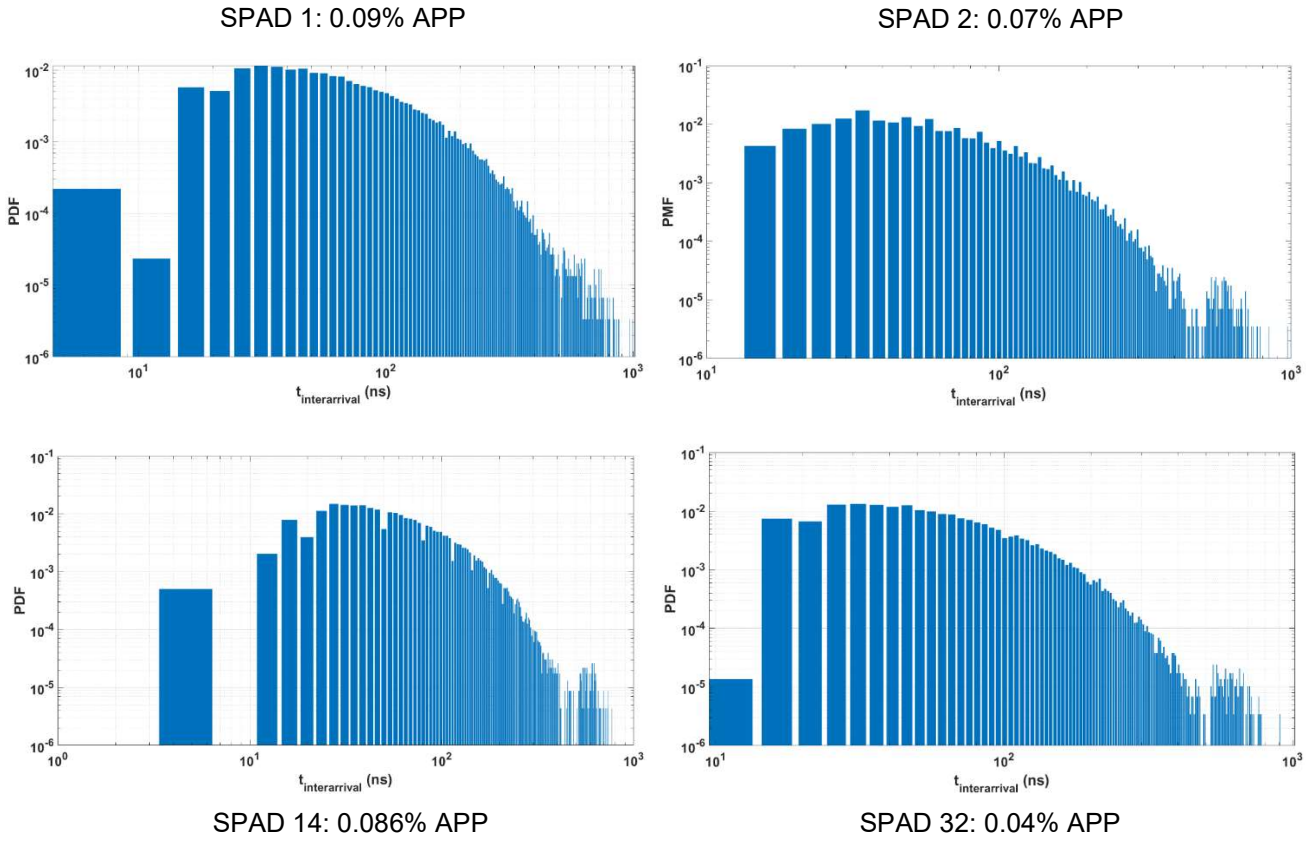

Figure 5.3: Inter-arrival time histograms and the corresponding APP calculated for each SPAD.

Crosstalk (XTLK) between adjacent pairs of SPADs is measured by looking for times when one SPAD fires and another fires at the same time. A histogram of the inter-arrival times between pairs of SPADs is plotted in Fig. 5.4 and the peak which rises above Poisson statistics at 0 time interval is looked for. This is the coincidence when two neighbouring SPADs fire simultaneously. With 112 SPADs, there are a lot of pairs to search through, so the layout geometry is considered to figure out detector neighbours exclusively. There is a very low probability of two SPADs firing at the same time, so rather than increasing the exposure time or number of iterations, the crosstalk measurements are taken in the presence of background light (1 klx lab lighting fixtures) to speed up the acquisition process. It would take a much longer time to gather this data with just DCR.

At high intensities, it will be very likely that two SPADs will fire simultaneously even without crosstalk, but these events are encompassed by the expected Poisson distribution which can be seen in Fig. 5.4. XTLK can therefore

be differentiated from these events with a probability which can be extrapolated from the slope of the Poisson curve – anything above the slope at time 0 is crosstalk.

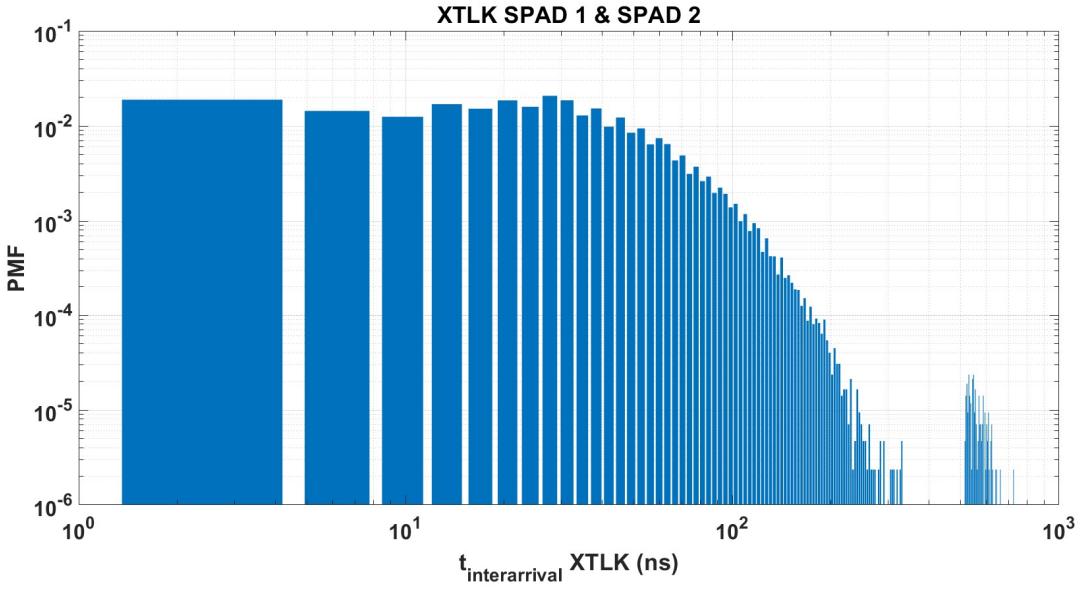

Figure 5.4: Histograms of the inter-arrival times between a pair of adjacent SPADs. The corresponding horizontal XTLK probability is 1.6 %.

5.3 Towards a Linearisation of SPAD Elements with Data Rate

Ideally, symbol rate – and therefore bit rate for a given modulation scheme and alphabet depth – should follow one-to-one mapping with the SNR and hence number of detector elements. This way the gross data rate is directly proportional to the total active area like a conventional RX and should scale linearly along the slope given by the quantum limit (Chapter 3). With this objective in mind, an experiment is carried out to find the maximum data rate attainable with a single SPAD on the 40 nm chip. An integrating RX architecture are chosen over a single-sample architecture to increase the number of counts per symbol and enhance SNR. OOK is selected because of the limited dynamic range of the small array size of SPADs. RZ is chosen as the line coding scheme due to its effectiveness at reducing $x(N+C)$ future ISI in C following symbols by allowing the SPADs to recover at the end of a symbol period $x(N)$, proven in previous literature [28, 37]. Fig. 5.5 shows the received histogram at 1 Mb/s RZ (50 % duty cycle) with a single SPAD detector, 4 Vpp transmit voltage, and no equalization.

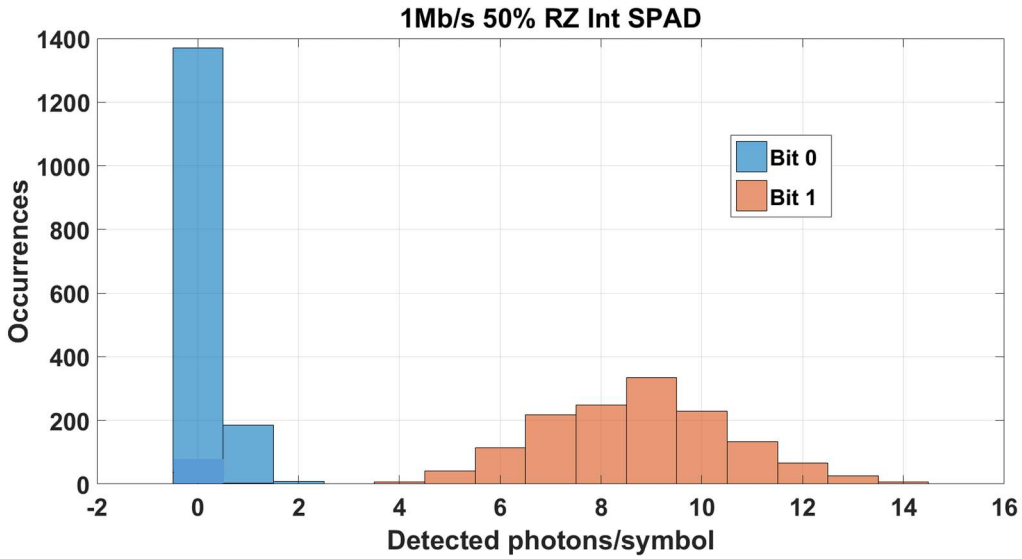

Figure 5.5: Single integrating SPAD RX histogram at 1 Mb/s, 50 % RZ.

A fixed decision threshold m_T is applied to the received signal, determined by exhaustive search (ES) based on the histogram of the output photocounts over all possible symbol sequences, to minimise the BER. This demodulation method is used in literature [38]. However, the main drawback of ES is that a long pilot signal must be sent prior to the data frame to establish a detailed photocount histogram. In a practical real-time system, this would cause significant latency. In this work, the data itself is used to build the histogram. A BER of 3.4×10^{-3} is measured with this demodulation technique. 50 % duty cycle corresponds to 500 ns signal pulse width, which is 23.8 times longer than the 21 ns mean dead time. From Eqn. (23) in Section 3.5.1, this means that a maximum average of $\frac{23.8}{e} = 8.759$ photons can be counted in each symbol after integration.

The mean of the '1' distribution in the histogram is around 9 photons/symbol, therefore it can be concluded that the incident modulated signal has a swing of 100 % intensity ($n = 1$) at the SPAD. This is not optimal, because the model developed in Chapter 3 Section 3.6.2 determines 10 % of maximum intensity ($n = 0.1$) as the point where the linearity of a SPAD is approximately maintained. However, with OOK modulation and RZ pulses, this higher intensity is acceptable because the SPAD is only saturated for the first half of the symbol – it has time to recover during the second half of the symbol and this ensures that it is highly probable the SPAD will be ready to detect the photons making up the pulse in the subsequent symbol. In addition, OOK modulation is more resilient to nonlinearity than schemes such as OFDM. It is concluded that 1 Mb/s is the maximum data rate possible to attain the BER for FEC with a single PQ SPAD and 21 ns dead time.

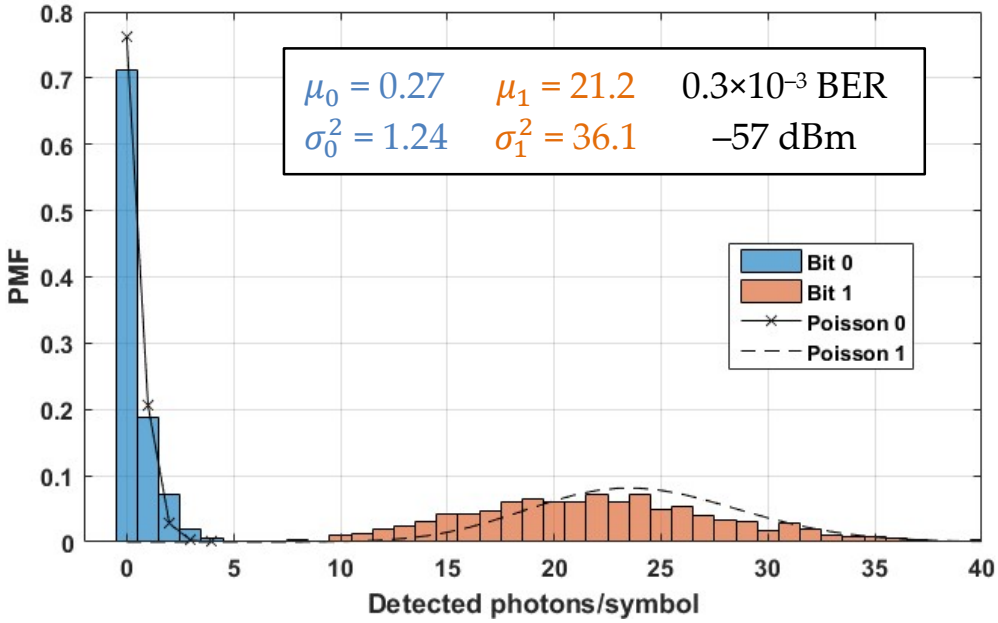

Figure 5.6: Integrating SPAD array RX histogram at 80 Mb/s, 50 % RZ.

Fig. 5.6 shows the received RZ histogram with 112 SPADs and the same transmit power as in the previous experiment (Fig. 5.5). Now the output counts are an aggregation from an array, so the maximum SNR, dynamic range, and RX BW are all increased. 80 Mb/s can be demodulated comfortably, with a BER of 0.3×10^{-3}. As computed in Eqn. (43), the means of the photocount distributions are not equal to their variances – this is not a Poisson process. Instead, the dead time has the distorted the Poisson statistics of the incident photons. The symbol distributions are close to the PMF predicted in Eqn. (43).

The received optical power at the SPAD array is calculated from the mean counts and scaled by the PDP penalty of the SPADs, giving –57 dBm sensitivity. The same is done at 100 Mb/s (Fig. 5.7): –55 dBm. At 450 nm, the 100 Mb/s QL is -69.6 dBm (Table 2, Section 3.2), so the SPAD array RX is operating 14.6 dB away from the QL.

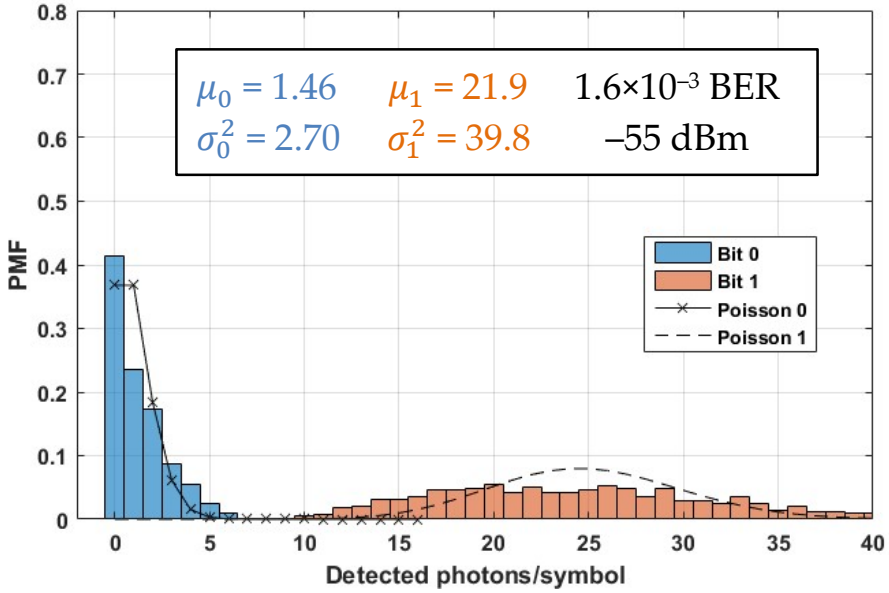

Figure 5.7: Integrating SPAD array RX histogram at 100 Mb/s, 4 Vpp transmit, 50 % RZ.

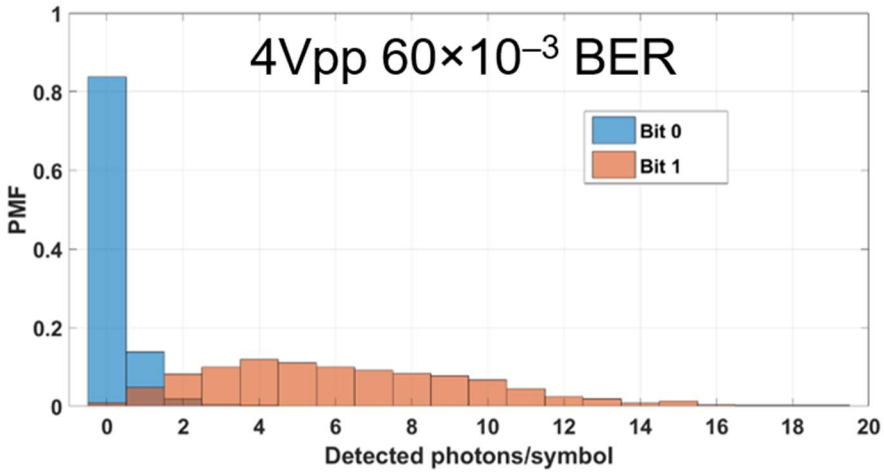

Figure 5.8: Integrating SPAD array RX histogram at 100 Mb/s, 4 Vpp transmit, 25 % RZ.

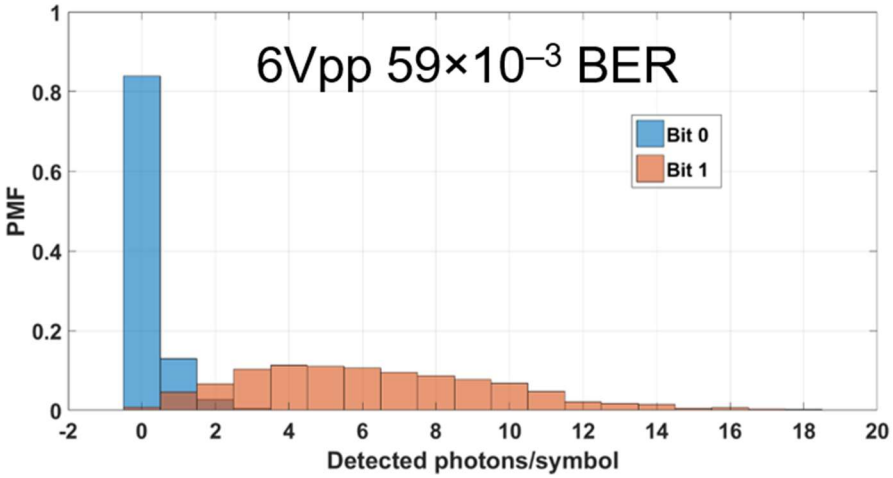

Figure 5.9: Integrating SPAD array RX histogram at 100 Mb/s, 6 Vpp transmit, 25 % RZ.

Figs. 5.7 and 5.8 show the received histograms for 50 % duty cycle and 25 % duty cycle, respectively. The narrowing in pulse width corresponds to doubling the required BW, and the resulting BER degradation from 1.6×10^{-3} to 60×10^{-3} proves this. Fig. 5.9 shows the histogram with a higher transmit power compared to Fig. 5.8. BER does not change between the two cases, which confirms that the system is hitting a BER floor due to a BW limitation rather than an SNR limitation. Therefore, the signal pulse width is too short in comparison to the dead time. This is known because the mean and variance of the symbol distributions does not change. The BER cannot be improved at 25 % RZ without reducing the bit rate. NRZ is also transmitted and Fig. 5.10 shows the eye diagram taken from the array in Matlab at 50 Mb/s – which results in a BER of 3×10^{-3} at best. This proves RZ is the superior choice for a small array of SPADs that can be accessed directly.

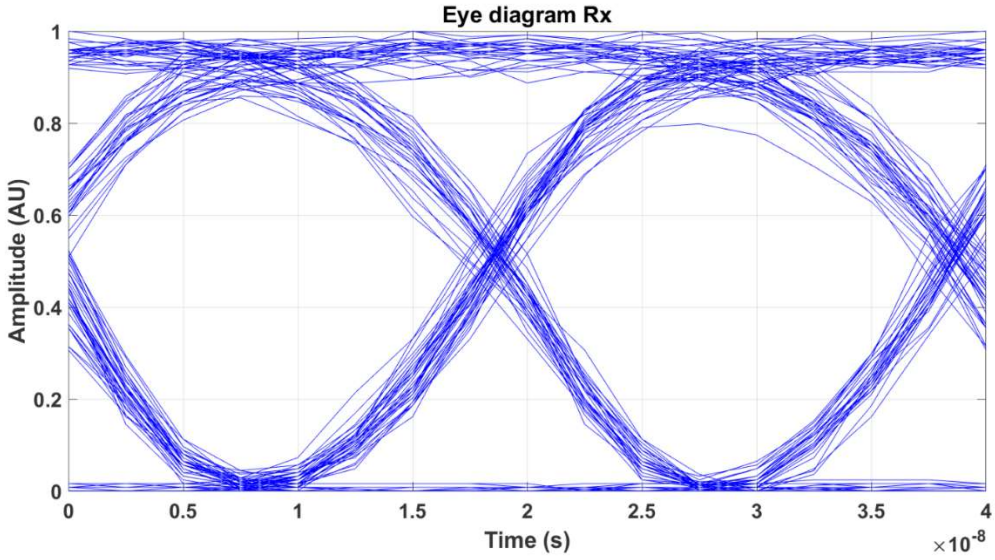

Figure 5.10: Eye diagram of 50 Mb/s NRZ with an integrating SPAD array RX. Samples taken from the RX were overlaid in Matlab to build the eye diagram.

Fig. 5.11. shows the BER measured for a TX intensity sweep at (a) 80 Mb/s and (b) 100 Mb/s. 112 SPADs are enabled on the IC and the mean dead time is 21 ns. 4 Vpp is the optimum transmit signal swing – sufficient to demodulate at the SPAD RX while preventing the emitter from saturating. In addition to ES, a linear adaptive equalizer with an RLS algorithm and a DFE equalizer are implemented. The tap weight coefficients of the equalizer are calculated for this specific SPAD channel in [42]. Mean τ is reduced to 10 ns and 8 ns and the experiment is repeated at 100 Mb/s (Fig. 5.12). These results highlight that τ has a significant influence on the performance of the RX and a shorter τ reduces the ISI. DFE shows best performance, due to its ability to combat

nonlinearity [48], but the linear RLS equalizer is also effective. This leads us to conclude linear EQ is sufficient at these symbol rates.

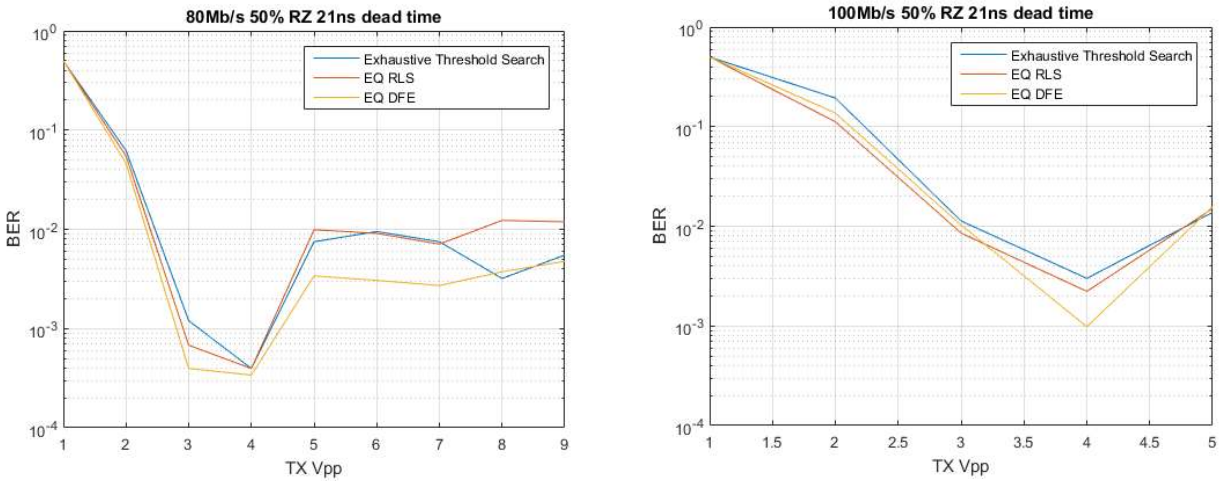

Figure 5.11: 80 Mb/s 50 % RZ, $\tau = 21$ ns (a) and 100 Mb/s 50 % RZ, $\tau = 21$ ns (b). BER increases above 4 Vpp transmit in both cases because the SPADs are saturating. BER floor above 7 Vpp because the LD is saturated.

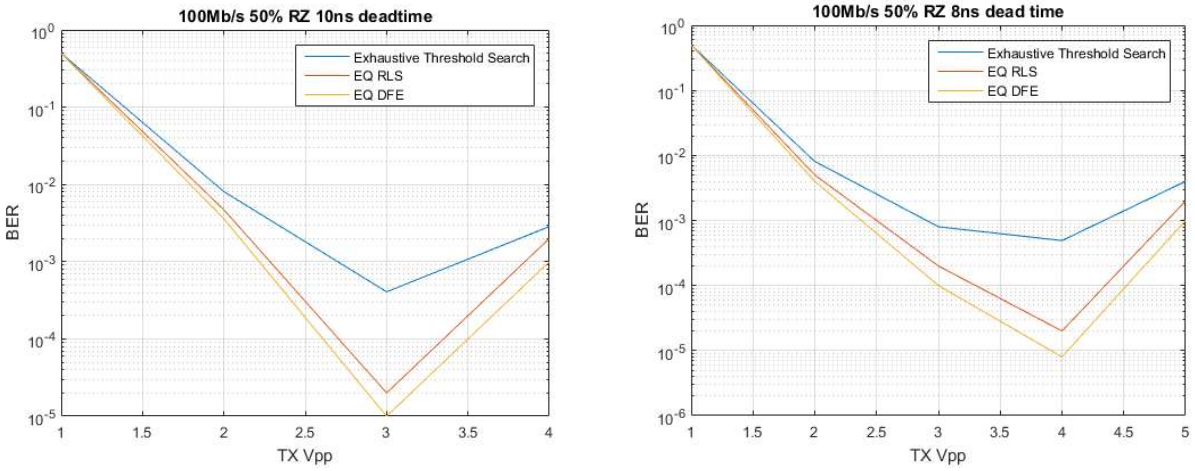

Figure 5.12: 100 Mb/s RZ with (a) $\tau = 10$ ns and (b) $\tau = 8$ ns.

The DFE is consistently the best performance, proving it is effective at responding to nonlinear systems with ISI. RLS also shows good performance compared to ES.

5.4 Large Parallel Photon Counting Results

5.4.1 OOK SNDR

At 400 Mb/s OOK, 1.8×10^{-3} BER is reached at –49.9 dBm average optical power. The detected signal is processed by a recursive least squares (RLS) linear equalizer in Matlab, which is tuned with a training sequence of 200 symbols in the preamble at the start of the data frame.

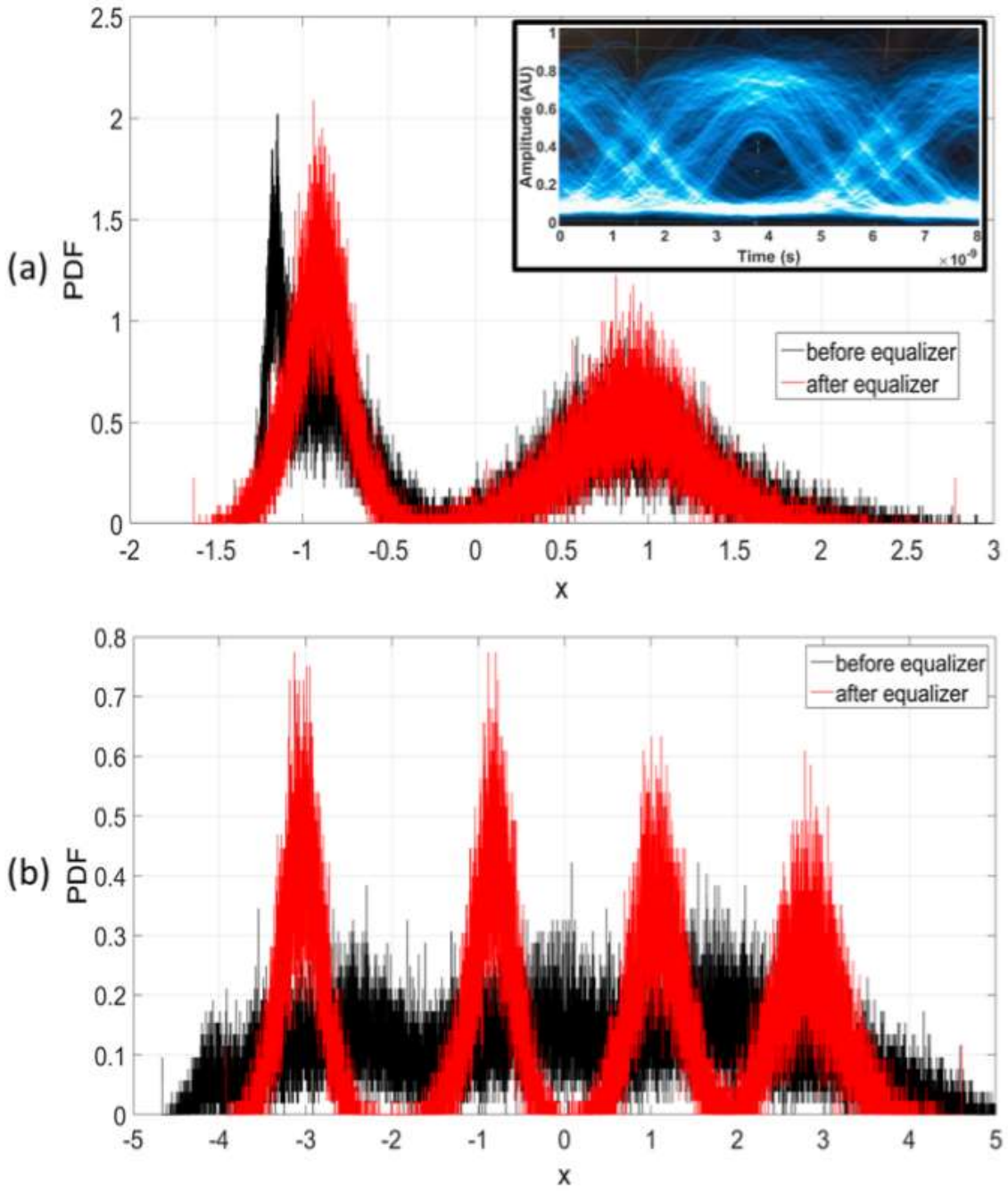

Figure 5.13: (a) OOK histogram before and after EQ at 400 Mb/s and analogue output eye diagram (before EQ) (b) 4-PAM histogram before and after EQ at 500 Mb/s.

Equalization (EQ) has marginal effect on performance at this rate, decreasing the BER from 3×10^{-3} to 1.8×10^{-3}. This is because 400 MBd symbol rate was estimated to be optimal in the simulation model in Section 3.6.3 (Fig. 3.11) – the SPAD RX is operating at the maximum sensitivity and symbol rate possible, so EQ cannot use any more power to reduce ISI. The OOK histogram (Fig. 5.13a) shows approximate Poisson distributions with variances almost, but not, equal to their mean. The lowest peak is the actual '0' probability mass function and the secondary peak is caused by ISI due to spurious counts from dead time, ER, SPAD afterpulsing (<1 %) and XOR tree jitter (0.5 ns FWHM). Fig. 5.13a also shows an eye diagram at 400 Mb/s OOK captured from the DAC output signal with a LeCroy 735Zi oscilloscope via a ZS4000 active probe.

The received 4-PAM histogram (Fig. 5.13b) has overlapping between adjacent symbols, which was negligible in the OOK histogram. We attain 500 Mb/s 4-PAM at 2×10^{-3} BER and –46.1 dBm sensitivity (Fig. 5.13b). EQ improves the BER from 7.6×10^{-3} to 2×10^{-3}. All these results meet the 3.5×10^{-3} BER threshold target required for FEC to achieve an output BER of 10^{-9} using concatenated BCH inner and outer codes with 6.69 % overhead [19].

The large SPAD array provides temporal redundancy to ease the dead time constraint by ensuring active SPADs are always available in a symbol period. This enables reception of symbols with durations (2.5 ns at 400Mb/s OOK, 4 ns at 500Mb/s 4-PAM) shorter than the individual SPAD dead time of 12 ns. Detector redundancy therefore obviates the requirement on current RX implementations that the dead time be matched to the symbol period to achieve the maximum data rate [28, 29, 37, 25]. 200 Mb/s is attained in [28] with only 4 SPADs and a short 3.5 ns dead time, but at the expense of relying on all SPADs to fire every bit period and compromised sensitivity due to crosstalk between the detectors (3 % mean, 1.5 V to 3.5 V excess bias), dark counts (1.46-13.9 kHz) and high afterpulsing probabilities up to 56 %. These spurious afterpulse avalanches are caused by the delayed release of trapped carriers around a similar timescale to the dead time [111]. Afterpulsing becomes more significant at shorter dead times and causes ISI at higher speeds. Instead, the SPAD array in this work has an optimal 1.3 V excess bias above 13.9 V breakdown, 6 kHz median dark count rate and an afterpulsing probability of around 1 %, albeit with a longer mean dead time of 12 ns. In reality, the dead time is not uniform and may vary depending on voltage and temperature, therefore, it is beneficial to reduce the influence of dead time on data rate and error performance. Reducing the susceptibility to dead time is one of the advantages of this SPAD RX.

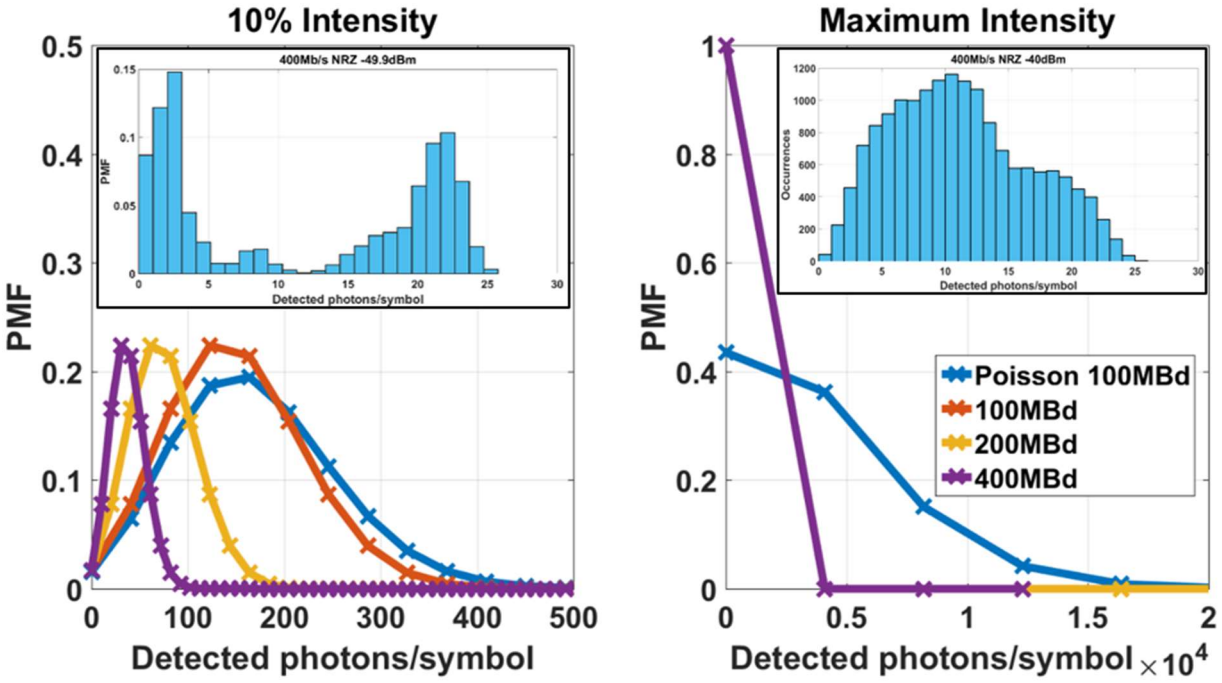

Figure 5.14: Simulated and measured PMF count distributions of a 64 × 64 array for $\tau = 12$ ns, $A = 8.33\times10^6$ s^{-1} (left) and maximum $A = 8.33\times10^7$ s^{-1} (right) at $R = 100$ MBd to 400 MBd.

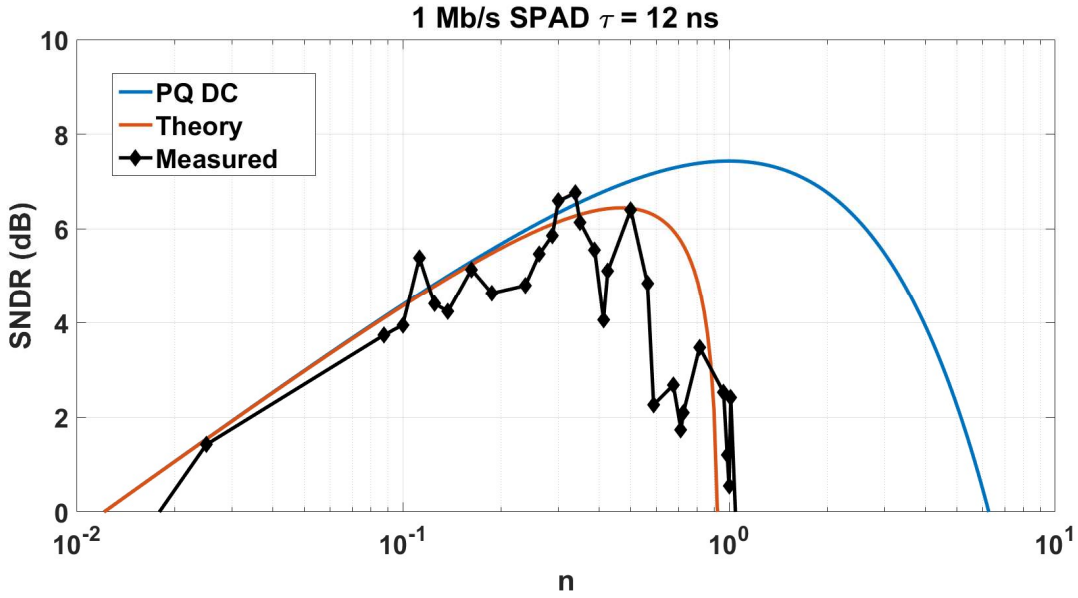

Figure 5.15: SNDR response at 1 Mb/s OOK with a single SPAD. Input waveform amplitude set by n. Peak SNDR = 6.8 dB at $n = 0.3375$.

5.4.2 Comparison to Model

The mathematical framework and simulations are compared to measurements obtained from experiments with the RX. At 400 Mb/s OOK, 3×10^{-3} BER is reached at –49.9 dBm with no equalization. The RX histogram is

overlaid with estimated distributions in Fig. 5.14. The lowest peak is the '0' PMF, with mean 4 detected photons/bit due to ISI and ER. The '1' distribution has an average of 23.1 counts/bit and so the RX is operating at around 10 % of its peak capacity of 230 Gphotons/s. At 100 %, the output is saturated as projected with the model. ER = 5.78 from the quotient of '1' and '0' means, which corresponds to 1.5 dB penalty.

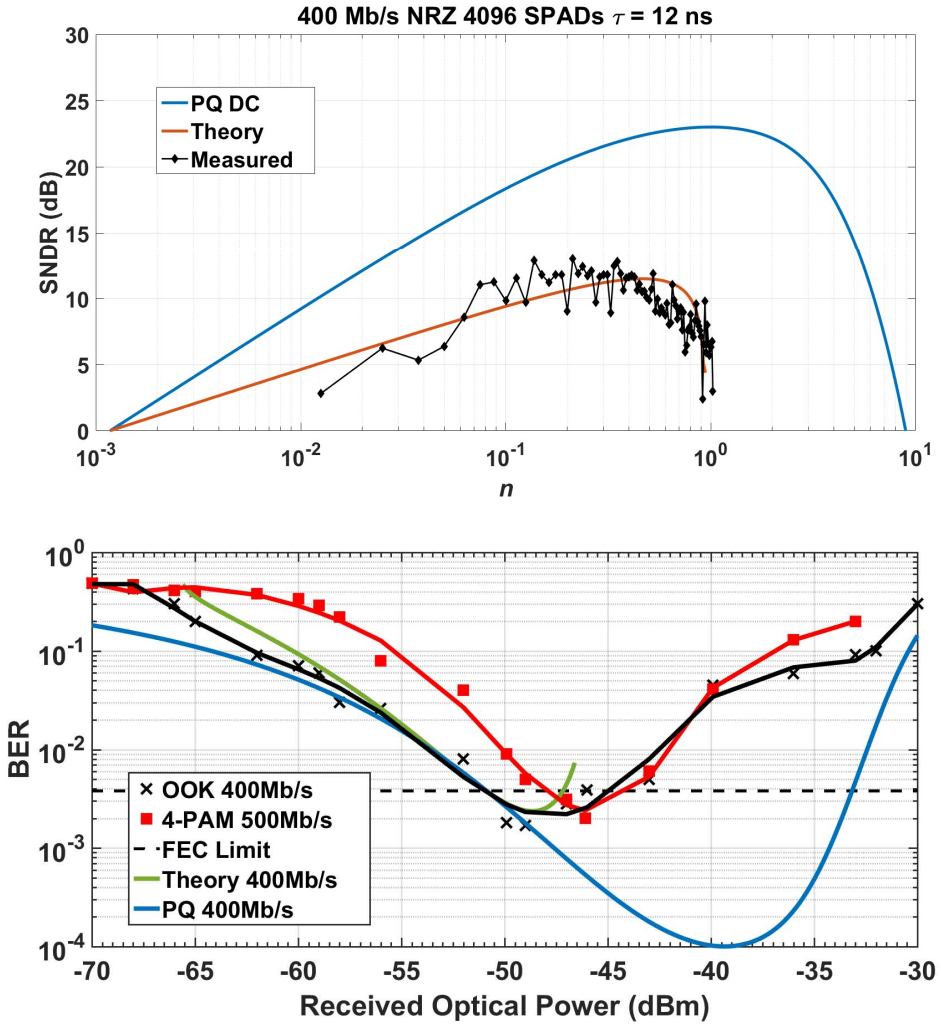

Figure 5.16: 64 × 64 array SNDR 400 Mb/s OOK against n (top). Theorized and measured BER with TX ER for 400 Mb/s OOK and 500 Mb/s 4-PAM (R = 400 MBd and 250 MBd respectively) with received power (bottom).

The intensity of the optical signal was increased, and the count rate of a single detector element was measured (Fig. 5.15). For each run, SNDR per symbol is found from the received counts and noise standard deviation. Measurements confirm increased distortion of the signal compared to a constant rate for the standard model with Poisson statistics (blue curve). Maximum SNDR = 6.8 dB at $n = 0.3375$. The sweep is repeated with all

SPADs in the array enabled (Fig. 5.16 top). Predicted BER from Eqn. (47) is plotted along with BER measurements and lines of best fit in Fig. 5.16 with incident optical power (measured with a power meter) proportional to n. As shown, the model and measurements have closely matching curves, so we attain an accurate prediction of performance.

Table 3: Comparison of estimated and measured SNDR for one SPAD and a 64 × 64 array at low (10% of the maximum count rate) and high intensities – the peak (Pk) value after distortion is added.

All in dB	10% SNDR	SNDR Pk	Array 10% SNDR	Array SNDR Pk
Theory	4.5	6.5	9	12
Simulation	4.7	5.4	9	12.7
Measured	4	6.8	9.9	13.4

It can be seen that BER degradation occurs earlier (at –45 dBm) than the paralyzable model (upper transfer curve) would indicate because of transient dead time distortion of the signal. This reduces SNDR to approximately 12 dB a decade below the maximum incident rate, despite Poisson SNR = 18 dB at this level (6 dB penalty). SNDR values from numerical estimates, simulations and RX measurements are summarised in Table 3. Maximum expected SNR = 23 dB at $n = 1$, assuming Poisson statistics and no data-carrying signal. Peak (Pk) effective SNDR is 9.6 dB lower than this, at 13.4 dB (when $n = 0.21$), which defines the BER inflection point. Therefore, 9.6 dB of distortion occurs on the received OOK signal. –49.9 dBm sensitivity at 400 MBd could potentially be reduced to –55.9 dBm if the 6 dB penalty due to nonlinear distortion is mitigated. This would allow a higher symbol rate or modulation depth to improve overall bit rate. A similar penalty occurs with 4-PAM, albeit at a higher average power.

5.4.3 O-OFDM Performance

Another advantage of a large SPAD array is that complex modulation schemes such as OFDM can be applied for high spectral efficiency and multipath mitigation. Furthermore, the SPAD RX is direct-to-digital so no analogue-to-digital converter (ADC) is required. DC biased optical OFDM (DCO-OFDM) waveforms are generated and adaptive modulation is applied with bit loading in each of the 512 subcarriers to attain 350 Mb/s

and 3.7 b/s/Hz mean spectral efficiency at 2×10^{-3} BER. Fig. 5.17 is the frequency response of the SPAD RX determined through channel estimation. The OFDM analogue bandwidth of the SPAD RX is 200 MHz.

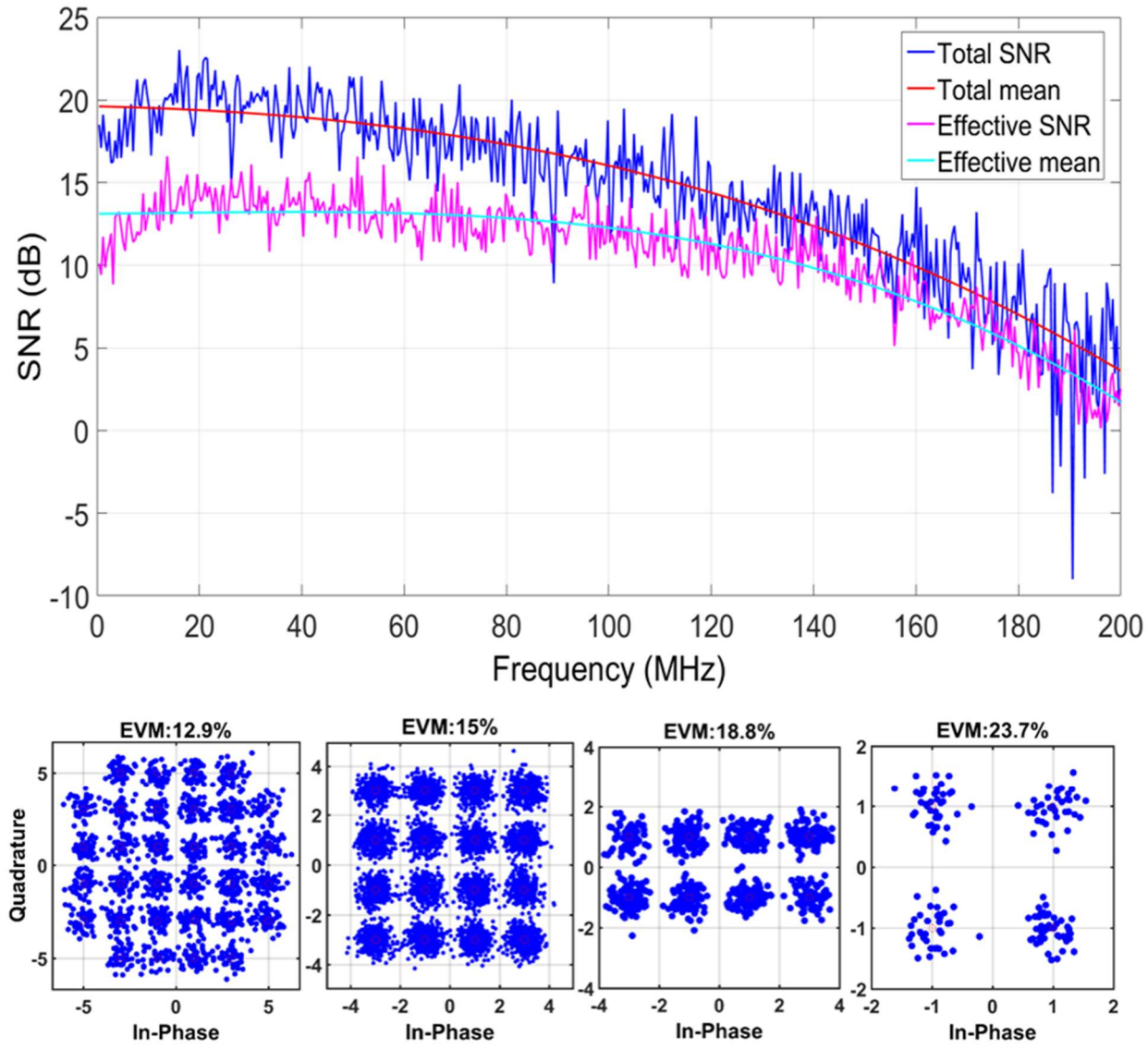

Figure 5.17: SNR (blue line) and SNDR (pink line) measured at each subcarrier from channel estimation (top). 32-QAM, 16-QAM, 8-QAM and QPSK received constellations (bottom left to right).

At low frequencies, where there is less input signal attenuation, the effective SNDR is more impaired from the total power SNR. There is a distortion penalty of approximately 6 dB between SNR and SNDR at 40 MHz. The penalty is 7 dB at subcarriers <40 MHz and around 2 dB for subcarriers from 150 MHz to 200 MHz. This distortion is due to dead time nonlinearity at higher incident photon rates. SNR (17 dB peak near DC) is sufficient to convey 32-QAM and 16-QAM with the lower frequency subcarriers, down to QPSK at subcarrier index 416. The SNDR is high enough to distinguish up to 32 transmitted QAM symbols and therefore increase the average number of

bits conveyed per symbol, which provides high spectral efficiency. I attained a spectral efficiency of 5.6 b/s/Hz at a reduced rate of 200 Mb/s.

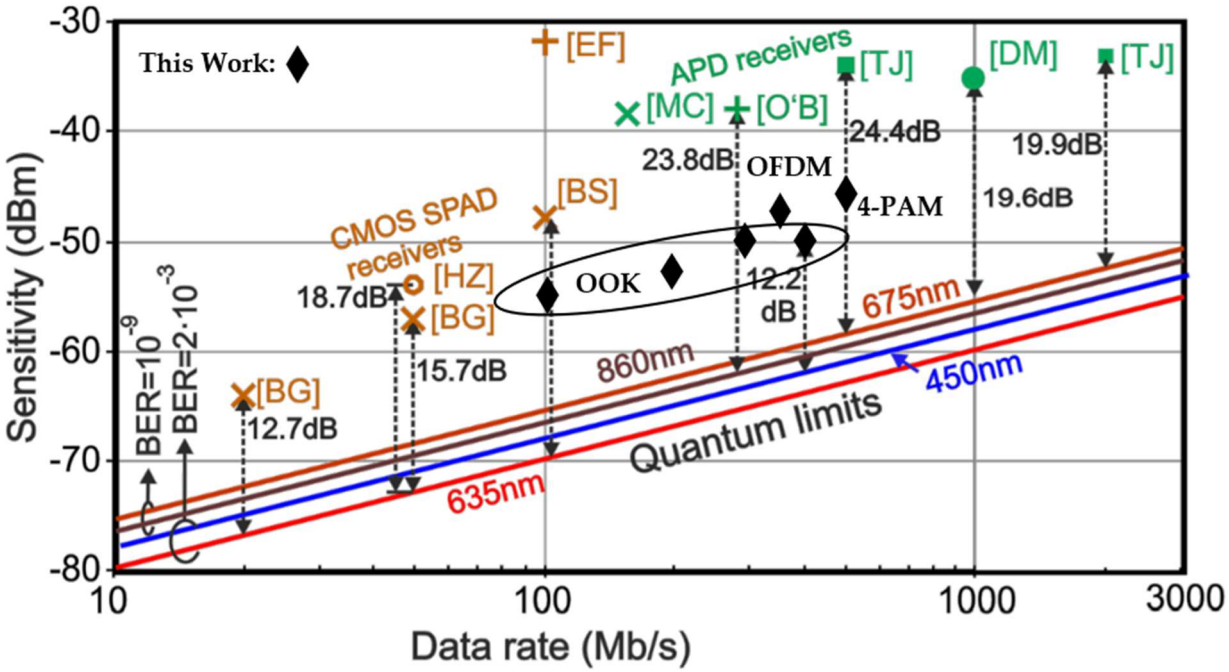

Figure 5.18: Comparison of PD, APD and SPAD visible light RX devices to the QL. Plot from [106] annotated with results from this work. SPAD RXs are shown in orange and APD RXs are shown in green. 400 Mb/s OOK with the 130 nm CMOS device is 12.2 dB from the QL.

5.5 Performance Comparison

The device is operated below the maximum rate of the XOR tress to avoid readout saturation. The array offers redundancy where detector elements are continually firing and recovering in steady state conditions. This enables high RX dynamic range and Poisson SNR up to 23 dB per sample in theory. RX sensitivity from 100 Mb/s OOK to 500 Mb/s 4-PAM is compared to other RXs in Fig. 5.18 and a distortion of around 5.7 dB is observed. This penalty cannot be explained with current SPAD models, so a novel method of describing nonlinearity is established in Chapter 3. Fig. 5.18 also shows the corresponding quantum limits for 450 nm and 635 nm wavelengths at 2×10^{-3} BER. Experiments are conducted both in darkness and with a background illuminance of 1 klx and results show that ambient light has negligible effect on performance when narrow bandpass optical filters are used.

Table 4: Performance comparison of PD, APD and SPAD visible light RX devices. The 130 nm RX has the best data rate, sensitivity, and power efficiency of SPAD ICs (highlighted in green).

Ref.	Fahs	O'Brien	Jukic	Fisher	Zhang	Steindl	This Work
Type	PD	APD	APD	SPAD	SPAD	SPAD	SPAD
Technology	0.35μm	N/A	0.35μm	130nm	180nm	0.35μm	130nm
Elements	1	1	1	1024	60	4	4096
Fill Factor	N/A	N/A	N/A	2.42%	3.2%	48%	43%
Sensitivity (dBm)	-23	-38	-34.6	-31.7	-53	-43.8	-46.1
Modulation Type	OOK	OOK	OOK	OOK	OOK	RZ-OOK	4PAM, OFDM
Bit Rate	2.5Gb/s	280Mb/s	1Gb/s	100Mb/s	100Mb/s	200Mb/s	500Mb/s, 350Mb/s
BER	10^{-3}	10^{-9}	10^{-9}	10^{-9}	10^{-3}	$6.5x10^{-3}$	$2x10^{-3}$
Consumption (pJ/bit)	86	N/A	244	800	N/A	248	230

A SPAD RX for fibre optic applications achieved 200 Mb/s at 6.5×10^{-3} BER within 24 dB of the quantum limit [28]. The fully integrated 130 nm CMOS SPAD RX SoC extends this data rate by 2.5× to 500 Mb/s with 4-PAM 2×10^{-3} BER, whilst improving sensitivity to –46.1 dBm, reducing the margin to the quantum limit to only 15.2 dB. The lowest margin to the QL achieved with this device is 12.2 dB at 400 Mb/s OOK. 100 Mb/s OOK is also reported with this RX at –55.2 dBm sensitivity with a time-correlated demodulation scheme, corresponding to an average of 7.5 detected photons per bit, 13.4 dB from the quantum limit [129].

With 230 pJ/b at best performance, this work proves the highest power efficiency of SPAD RXs published to-date [25, 28, 29]. Although the RX energy per bit is not lower than PD/APD RXs (ranging from 86 pJ/b in [24] to 129 pJ/b [102]), these do not include an ADC function, which is inherently realised by this SPAD RX architecture. Our value of 230 pJ/b could readily be reduced by over an order of magnitude by adopting advanced CMOS nodes. The sensitivities for 100-400 Mb/s OOK, 350 Mb/s DCO-OFDM and 500 Mb/s 4-PAM all outperform that of PD, APD and SPAD RXs [24, 25, 28, 29, 101-103] and the maximum bit rate is improved by a factor of 2.5 compared to [28] whilst maintaining 3.5×10^{-3} BER suitable for FEC with minimal bit redundancy [19]. The device reaches a sensitivity within 15.2 dB of the quantum limit set by the Poisson statistics of photon arrivals [130].

This gap includes a 1.5 dB power penalty due to the finite extinction ratio of the LD TX and an 8 dB loss from the SPAD photon detection probability (37%) and fill factor (43%). The former is not an RX issue and can be practically eliminated with the use of an external modulator at the source [28]. The latter can be readily reduced with an optimised SPAD technology [131, 57] and structure techniques such as 3D stacking [132]. However, the remaining 5.7 dB penalty is unaccounted for and attributed to the dead time inducing transient distortion and ISI. This causes an SNDR limitation (explored in Chapter 3) which prevents data rate from being further improved by increasing signal power. SPAD intrinsic parasitics evaluated in [119] are not as significant in our RX due to the low APP. Furthermore, these parasitics are too complex to analytically model with such a large array. With OFDM at low frequencies, where there is less signal attenuation, distortion is observed as effective SNDR is more impaired from the total power. This is caused by the high PAPR of OFDM, causing it to be more vulnerable to nonlinearity than single-carrier modulation formats.

5.6 Summary

The 130 nm CMOS SPAD RX demonstrates 500 Mb/s and a 15.2 dB margin to the quantum limit in a VLC link. Although the device attains the best SPAD RX performance published to date, this is accomplished with a large array of detectors, providing redundancy to ease the dead time constraint. BER degradation occurs earlier (at –45 dBm) than the SPAD large signal transfer curve would indicate because of transient dead time distortion of the signal at high photon rates, which reduces SNDR. Dead time remains as a 6 dB ISI limitation and performance is highly sensitive to changing signal intensity conditions, so a model is established to aid understanding and predict the resulting distortion (Chapter 3). Nonlinearity at high intensities infers that a practical use of this device is to supplement, rather than replace, existing PD/APD RXs. SNDR and BER is estimated and close matching with measurements validates the effectiveness of the simulation algorithm and the analytical model.

6. Conclusions

6.1 Outcome

The ideal statistics of light quantify the performance limit of a photon counting RX. A model is created to focus on the parameters of a practical photon counting RX and aid in understanding the inconsistencies between theory and practical measurements. The model concludes that the fixed dead time τ and higher maximum count rate benefit of AQ over PQ is only marginally useful, because the SPAD should not operate in its saturation region anyway with an incident modulated signal. Therefore, the added circuit complexity and power consumption brought with AQ is not worth it at this point in research. The accuracy of this method is confirmed by experiments. Results with the 40 nm CMOS IC show that 1 Mb/s OOK is the maximum bit rate achievable with a single PQ SPAD detector. OOK RZ with 50% duty cycle is shown to be the optimum signalling method for SPADs, however, further optimisation of the pulse profile is a potential direction for future work. 100 Mb/s RZ is attained with 112 SPADs. To the author's knowledge, this is the first time an almost linear scaling of detector elements to symbol rate has been demonstrated. This means the sensitivity slope of the RX is parallel to the QL. Ideally, a photon counting RX could be designed this way such that all the counts from the detectors are combined into a single channel without losses. The differences observed between the experimental results and the expected results are explained. Parasitics such as afterpulsing are shown to be negligible for high-speed communication if dead time is not reduced below 5 ns [28] and if SPAD processes are optimised [109].

The 130 nm CMOS RX architecture permits massively parallel (4096) photon event summation to be achieved at a high fill-factor (43%) and sample rate (800 MHz). Detector redundancy obviates the requirement on current RX implementations that the SPAD dead time be matched to the symbol period to achieve the maximum data rate [3, 4, 5]. Another advantage is that complex modulation schemes such as OFDM or PAM can be applied for high spectral efficiency and multipath interference mitigation. 400 Mb/s OOK is close to the target rate predicted in the model with a similar number of SPADs. It is demonstrated that the RX can operate in a practical, background insensitive VLC link at 1 m in 1 klx ambient conditions using the 450 nm LD and optical filtering. The RX confirms photon counting reception at up to 500 Mb/s from 1 cm to 1 m. 350 Mb/s is attained with DCO-OFDM at -48 dBm – this is 15dB from the quantum limit. RF technologies support the 802.11ac and 802.11n standards [45, 46] with net data rates between 54 Mb/s and <7 Gb/s. This work is the first CMOS optical RX to

support a wireless link with a similar data rate, modulation standard, and received irradiance (–49.9 dBm to –46.1 dBm) similar to that of WiFi.

The higher power consumption and dead time pile-up nonlinearity of the SPADs at high signal levels leads to concluding that the practical use of this device to be in assistance (rather than replacement) of existing APD or PIN RXs for LiFi, UWOC and POF applications towards extended link range or maintaining a low-rate link in highly scattering environments. In an improved RX implementation, the XOR tree could be replaced by fast summation at the exterior of the array, reducing photon cancellation effects and improving fill factor further. Ultimately, 3D stacking would allow further improvements to SPAD PDP (also in near-infrared) [59, 109] and the placement of combination and summing electronics under the detectors without compromising fill factor [58]. These improvements move towards eliminating almost all of the 8 dB penalty due to PDP and FF losses. In addition, the 1.5 dB penalty from the LD extinction ratio could be mitigated if the cost of shutters such as SLMs goes down [103]. Recently, a PTID algorithm has been developed to mitigate the ISI induced by dead time and improve RX sensitivity by up to 7 dB [42]. This would repair the 6 dB distortion loss predicted in the model (Chapter 3) and confirmed in the experiments (Chapter 5).

While the RX data rate is much lower than the fastest rate reported with an ASiPM to-date (3.45 Gb/s), we prove a photon counting architecture is more sensitive and closer to the QL. The main advantage of a DSiPM RX over an ASiPM RX is lower power consumption since no ADC is required in the receive path. In addition, a smaller number of detector elements are required, which means that each SPAD in the 130 nm IC is used more efficiently than in a large ASiPM. This means the total active area can be adjusted to optimise sensitivity.

6.2 Future Work

The assumptions in this work should be addressed to create a more practical system suitable for use outside of the laboratory. The 450 nm LD used has a high output power and is not Class 1 or 1M eye safe, making it unsuitable for VLC-based consumer applications [21]. For example, a diffuser could be used to create a secondary source and diverge the beam to ensure eye safety. Alternatively, a blue μLED array would have sufficient BW (~980 MHz [17]) for the symbol rates (250 MBd to 400 MBd) and OFDM BW (200 MHz) demonstrated in this thesis. Another assumption, that the 55 % transmission loss of the bandpass filter at the SPAD RX is counteracted by the nonideal spectral emission of the LD (allowing a band of wavelengths to be

detected), should be more accurately modelled with and without different ambient light sources. Furthermore, precise alignment in the high-speed experiment is not practical and should be relaxed. Mircolenses on the SPAD array could solve this issue [109].

There are potential routes towards higher speeds and improved error performance for SPAD RXs. Techniques such as predistortion at the TX [134] could be implemented to mitigate this distortion, with the metrics determined from the model used for calibration. Alternatively, a digital equivalent of automatic gain control could be employed at the RX to reduce the output level when the input signal intensity is too high. BER could potentially be reduced by optimising and automatically adjusting the decision threshold for OOK and PAM depending on received count statistics. In this thesis, adaptive EQ was proven very effective with the nonlinear nature of SPADs and there is further work for ASIC-based EQ and adaptive threshold implementations. For SPAD-based OFDM, recent work shows a promising solution to rectifying nonlinear distortion due to dead time by applying pre-equalization at the TX [135]. Underwater wireless optical communication is a promising use case for the very high sensitivity of the digital SPAD RX in this work. The SPAD-based system solves the sensitivity-area bottleneck at the analogue front-end whilst maintaining a compact form-factor suitable for portable devices, but further research into field of view, multipath, and mobility could be done. SPAD RXs could be integrated with current ToF imaging systems and emerging technologies such as LiDAR [47]. Future practical work would involve extending wireless or POF link range and maintaining connectivity in extreme conditions such as dark or diffuse environments.

References

[1] The United Kingdom Frequency Allocation Table, OFCOM UK Std. 17, June 22, 2017. [Online] Available: http://stakeholders.ofcom.org.uk/spectrum/information/uk-fat/

[2] Cisco, “Cisco Annual Internet Report (2018–2023),” White Paper, March 9, 2020. [Online] Available: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.pdf

[3] D. Karunatilaka, F. Zafar, V. Kalavally, and R. Parthiban, “LED based indoor visible light communications: state of the art”, *IEEE Communications Surveys & Tutorials*, vol. 17, no. 3, pp. 1649-1678, 2015.

[4] H. Burchardt, N. Serafimovski, D. Tsonev, S. Videv, and H. Haas, “VLC: Beyond point-to-point communication,” *IEEE Commun. Mag.*, vol. 52, no. 7, pp. 98–105, Jul. 2014.

[5] T. Komine, M. Nakagawa, “Fundamental analysis for visible-light communication system using LED lights”, *IEEE Trans. Consumer Electronics*, vol. 50, no. 1, pp. 100-107, 2004.

[6] H. Elgala, R. Mesleh, and H. Haas, "Indoor Optical Wireless Communication: Potential and State-of-the-Art," *IEEE Commun. Mag.*, vol. 49, no. 9, pp. 56-62, 2011, ISSN: 0163-6804.

[7] L. Shi, W. Li, X. Zhang, Y. Zhang, G. Chen and A. Vladimirescu, "Experimental 5G New Radio integration with VLC," *2018 25th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, Bordeaux, 2018, pp. 61-64.

[8] Vishwaraj and L. Ali, "Hybrid MIMO-OFDM System for 5G Network Using VLC-A Review," *2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, Coimbatore, India, 2019, pp. 1-5.

[9] D. Tsonev, S. Videv, and H. Haas, "Towards a 100 Gb/s visible light wireless access network," *Opt. Express 23*, pp. 1627-1637 (2015).

[10] D. Tsonev, H. Chun, S. Rajbhandari, J. J. D. McKendry, S. Videv, E. Gu, M. Haji, S. Watson, A. E. Kelly, G. Faulkner, M. D. Dawson, H. Haas, and D. O’Brien, “A 3-Gb/s single-LED OFDM-based wireless VLC link using a gallium nitride μLED,” *IEEE Photon. Technol. Lett.* 26, pp. 637–640 (2014).

[11] D. V. Dinh, Z. Quan, B. Roycroft, P. J. Parbrook, and B. Corbett, “GHz bandwidth semipolar (112−2) InGaN/GaN light-emitting diodes,” *Opt. Lett.*, vol. 41, no. 24, pp. 5752–5755, 2016.

[12] E. Xie et al., “High-speed visible light communication based on a III-nitride series-biased micro-LED array,” *J. Lightw. Technol.*, vol. 37, no. 4, pp. 1180–1186, Feb. 15, 2019.

[13] Y. Zhou, J. Zhao, M. Zhang, J. Shi, and N. Chi, “2.32 Gbit/s phosphorescent white LED visible light communication aided by two-staged linear software equalizer,” in *Proc. 10th Int. Symp. Commun. Syst., Netw. Digit. Signal Process. (CSNDSP)*, Jul. 2016, pp. 1–4.

[14] M. S. Islim et al., “Towards 10 Gb/s orthogonal frequency division multiplexing-based visible light communication using a GaN violet micro-LED,” *Photon. Res.*, vol. 5, no. 2, pp. A35–A43, Apr. 2017.

[15] J. Grubor, S. C. J. Lee, K.-D. Langer, T. Koonen, and J. W. Walewski, “Wireless high-speed data transmission with phosphorescent white-light LEDs,” in *Proc. 33rd Eur. Conf. Exhib. Opt. Commun. Post-Deadline Papers*, 2008, pp. 1–2.

[16] R. X. G. Ferreira et al., “High bandwidth GaN-based micro-LEDs for multi-Gb/s visible light communications,” *IEEE Photon. Technol. Lett.*, vol. 28, no. 19, pp. 2023–2026, Oct. 1, 2016.

[17] E. Xie et al., "Over 10 Gbps VLC for Long-Distance Applications Using a GaN-Based Series-Biased Micro-LED Array," in *IEEE Photonics Technology Letters*, vol. 32, no. 9, pp. 499-502, May 1, 2020.

[18] R. Bian, I. Tavakkolnia and H. Haas, "15.73 Gb/s Visible Light Communication With Off-the-Shelf LEDs," in *Journal of Lightwave Technology*, vol. 37, no. 10, pp. 2418-2424, May 15, 2019.

[19] Telecommunication Standardization Sector, ITU-T. 2004. Forward Error Correction for High Bit Rate DWDM Submarine Systems. *ITU Tech. Rep.*, G.975.1. [Online] Available: http://www.itu.int/rec/T-REC-G.975.1-200402-I/en

[20] Vishay, “Ultrabright 0402 ChipLED,” Malvern, PA, USA, Feb. 2017.

[21] R. Singh *et al.*, "Design and Characterisation of Terabit/s Capable Compact Localisation and Beam-Steering Terminals for Fiber-Wireless-Fiber Links," in *Journal of Lightwave Technology*, vol. 38, no. 24, pp. 6817-6826, Dec.15, 2020.

[22] A. Abhishek, Z. Zeya, P. Suraj and R. K. Badhai, "Design of Beam Steering Antenna for 5G at 28GHz Using Butler Matrix," *2020 5th International Conference on Computing, Communication and Security (ICCCS)*, Patna, India, 2020, pp. 1-4.

[23] H. Huang *et al.*, "100 Tbit/s free-space data link using orbital angular momentum mode division multiplexing combined with wavelength division multiplexing," *2013 Optical Fiber Communication Conference and Exposition and the National Fiber Optic Engineers Conference (OFC/NFOEC)*, Anaheim, CA, USA, 2013, pp. 1-3.

[24] B. Fahs, A. J. Chowdhury and M. M. Hella, "A 12-m 2.5-Gb/s Lighting Compatible Integrated Receiver for OOK Visible Light Communication Links," in *Journal of Lightwave Technology*, vol. 34, no. 16, pp. 3768-3775, Aug. 15, 2016.

[25] L. Zhang *et al.*, "A Comparison of APD- and SPAD-Based Receivers for Visible Light Communications," in *Journal of Lightwave Technology*, vol. 36, no. 12, pp. 2435-2442, June 15, 2018.

[26] J. M. Kahn and J. R. Barry, "Wireless infrared communications," in *Proceedings of the IEEE*, vol. 85, no. 2, pp. 265-298, Feb. 1997.

[27] D. O'Brien, S. Rajbhandari, and H. Chun, ''Transmitter and receiver technologies for optical wireless,'' *Phil. Trans. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 378, no. 2169, Apr. 2020, Art. no. 20190182.

[28] B. Steindl, M. Hofbauer, K. Schneider-Hornstein, P. Brandl and H. Zimmermann, "Single-Photon Avalanche Photodiode Based Fiber Optic Receiver for Up to 200 Mb/s," in *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 24, no. 2, pp. 1-8, March-April 2018.

[29] E. Fisher, I. Underwood and R. Henderson, "A Reconfigurable Single-Photon-Counting Integrating Receiver for Optical Communications," in *IEEE Journal of Solid-State Circuits*, vol. 48, no. 7, pp. 1638-1650, July 2013.

[30] S. Huang and M. Safari, "Hybrid SPAD/PD Receiver for Reliable Free-Space Optical Communication," in *IEEE Open Journal of the Communications Society*, vol. 1, pp. 1364-1373, 2020.

[31] Onsemi.com. 2020. *J-Series SiPM Sensors Datasheet*. [Online] Available: https://www.onsemi.com/pub/Collateral/MICROJ-SERIES-D.PDF, Accessed 29 December 2020.

[32] W. Matthews, Z. Ahmed, W. Ali and S. Collins, "A SiPM-based VLC Receiver for 3.45 Gigabits/s Communication Using OOK Modulation," *2020 IEEE Photonics Conference (IPC)*, Vancouver, BC, Canada, 2020, pp. 1-2.

[33] Z. Ahmed, R. Singh, W. Ali, G. Faulkner, D. O'Brien and S. Collins, "A SiPM-Based VLC Receiver for Gigabit Communication Using OOK Modulation," in *IEEE Photonics Technology Letters*, vol. 32, no. 6, pp. 317-320, March 15, 2020.

[34] Y. Li, M. Safari, R. Henderson, and H. Haas, "Optical OFDM with single-photon avalanche diode," *IEEE Photon. Technol. Lett.*, vol. 27, no. 9, pp. 943–946, May 2015.

[35] Y. Li, M. Safari, R. Henderson, and H. Haas, "Nonlinear distortion in SPAD-based optical OFDM systems," in *Proc. IEEE Globecom Workshops*, 2015, pp. 1–6.

[36] E. Sarbazi, M. Safari and H. Haas, "Statistical Modeling of Single-Photon Avalanche Diode Receivers for Optical Wireless Communications," in *IEEE Transactions on Communications*, vol. 66, no. 9, pp. 4043-4058, Sept. 2018.

[37] H. Zimmermann, B. Steindl, M. Hofbauer, and R. Enne, “Integrated fiber optical receiver reducing the gap to the quantum limit,” *Sci. Rep.*, vol. 7, no. 1, 2017, Art. no. 2652.

[38] D. Chitnis and S. Collins, "A SPAD-Based Photon Detecting System for Optical Communications," in *Journal of Lightwave Technology*, vol. 32, no. 10, pp. 2028-2034, May 15, 2014.

[39] D. Chitnis *et al*., "A 200 Mb/s VLC demonstration with a SPAD based receiver," *2015 IEEE Summer Topicals Meeting Series (SUM)*, Nassau, Bahamas, 2015.

[40] L. Zhang, H. Chun, Z. Ahmed, G. Faulkner, D. O'Brien and S. Collins, "The Future Prospects for SiPM-Based Receivers for Visible Light Communications," in *Journal of Lightwave Technology*, vol. 37, no. 17, pp. 4367-4374, Sept. 1, 2019.

[41] J. G. Proakis, *Digital Communications*. New York, NY, USA: McGraw Hill, 2007.

[42] S. Huang, S. M. Patanwala, J. Kosman, R. K. Henderson and M. Safari, "Optimal Photon Counting Receiver for Sub-Dead-Time Signal Transmission," in *Journal of Lightwave Technology*, vol. 38, no. 18, pp. 5225-5235, Sept. 15, 2020.

[43] A. Griffiths *et al.*, “Temporal encoding to reject background signals in a low complexity, photon counting communication link,” *Mater.*, vol. 11, no. 9, 2018, Art. no. 1671.

[44] J. Kosman *et al*., "A 500Mb/s -46.1dBm CMOS SPAD Receiver for Laser Diode Visible-Light Communications," *2019 IEEE International Solid-State Circuits Conference - (ISSCC)*, San Francisco, CA, USA, 2019, pp. 468-470.

[45] Perahia, E., and R. Stacey. *Next Generation Wireless LANs: 802.11n and 802.11ac*. 2nd Edition. United Kingdom: Cambridge University Press, 2013.

[46] "IEEE Standard for Information technology--Telecommunications and information exchange between systems—Local and metropolitan area networks--Specific requirements--Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications--Amendment 4: Enhancements for Very High Throughput for Operation in Bands below 6 GHz.," in *IEEE Std 802.11ac(TM)-2013 (Amendment to IEEE Std 802.11-2012, as amended by IEEE Std 802.11ae-2012, IEEE Std 802.11aa-2012, and IEEE Std 802.11ad-2012)* , vol., no., pp.1-425, 18 Dec. 2013.

[47] S. Patanwala, I. Gyongy, N. Dutton, B. Rae, and R. Henderson, “A reconfigurable 40 nm CMOS SPAD array for LiDAR receiver validation,” in *Proc. Int. Image Sensor Workshop*, 2019, pp. 1–4.

[48] I. Glover and P. Grant, *Digital Communications*, 3rd ed. Pearson, 2010.

[49] B. Razavi, *Design of Integrated Circuits for Optical Communications*, 1st ed. McGraw-Hill, 2003.

[50] L. N. Binh, *Digital Optical Communications*, 1st ed. CRC Press, 2008.

[51] J. Campbell, “Recent Advances in Telecommunications Avalanche Photodiodes,” *Lightwave Technology, Journal of*, vol. 25, no. 1, pp. 109–121, 2007.

[52] M. Zuffada, “The Industrialization of the Silicon Photonics: Technology Road Map and Applications,” in *ESSCIRC, Proceedings of*, 2012.

[53] E. Sackinger, *Broadband Circuits for Optical Fiber Communication*. Wiley, 2005.

[54] 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.

[55] 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.

[56] M. Lee and E. Charbon, "Progress in single-photon avalanche diode image sensors in standard CMOS: From two-dimensional monolithic to three-dimensional-stacked technology," *Japanese Journal of Applied Physics*, vol. 57, no. 10, 1002A3, Sept. 10, 2018.

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

[58] K. Ito *et al.*, "A Back Illuminated 10μm SPAD Pixel Array Comprising Full Trench Isolation and Cu-Cu Bonding with Over 14% PDE at 940nm," *IEEE International Electron Devices Meeting (IEDM)*, Dec. 2020.

[59] O. Kumagai et al., "A 189x600 Back-Illuminated Stacked SPAD Direct Time-of-Flight Depth Sensor for Automotive LiDAR Systems," *2021 IEEE International Solid-State Circuits Conference - (ISSCC)*, pp. 110-111, 2021

[60] C. Lee *et al.*, “4Gb/s Direct Modulation of 450nm GaN Laser for High-speed Visible Light Communication”, *Opt. Express*, vol. 23, no. 12, pp. 16232-16237, 2015.

[61] Image [Online] Available: https://www.spiedigitallibrary.org/ContentImages/Journals/OPEGAR/55/6/060501/FigureImages/OE_55_6_060501_f001.png, Accessed 2 February 2019.

[62] A. Perallos, U. Hernandez-Jayo, I. J. G. Zuazola, and E. Onieva, *Intelligent Transport Systems: Technologies and Applications*, John Wiley & Sons, 2015.

[63] T. Koonen, K. A. Mekonnen, Z. Cao, F. Huijskens, N. Q. Pham and E. Tangdiongga, "Beam-Steered Optical Wireless Communication for Industry 4.0," in *IEEE Journal of Selected Topics in Quantum Electronics*, 2021.

[64] B. Fahs, M. Romanowicz, J. Kim and M. M. Hella, "A Self-Alignment System for LOS Optical Wireless Communication Links," in *IEEE Photonics Technology Letters*, vol. 29, no. 24, pp. 2207-2210, 15 Dec., 2017.

[65] Image [Online] Available: https://lh3.googleusercontent.com/HV2-VL5vn_UIo2QFRUir7b01Ea-VXrmQZ64Wa25ezj74bkQsgbG_r4QzWPRKPYywhrv4eQ=s93, Accessed 2 February 2019.

[66] M. Elamassie, M. Karbalayghareh, F. Miramirkhani, R. C. Kizilirmak, and M. Uysal, “Effect of fog and rain on the performance of vehicular visible light communications”, in *IEEE 87th Vehicular Technology Conference (VTC2018-Spring)*, Porto, Portugal, 2018.

[67] Y. H. Kim, W. A. Cahyadi, and Y. H. Chung, “Experimental demonstration of VLC-based vehicle-to vehicle communications under fog conditions”, *IEEE Photon. J.*, vol. 7, no. 6, pp. 1-9, Dec. 2015.

[68] M. Karbalayghareh, F. Miramirkhani, M. Safari and M. Uysal, "Vehicular Visible Light Communications with SPAD Receivers," *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, Marrakesh, Morocco, 2019, pp. 1-5.

[69] S. W. Hutchings *et al.*, "A Reconfigurable 3-D-Stacked SPAD Imager With In-Pixel Histogramming for Flash LIDAR or High-Speed Time-of-Flight Imaging," in *IEEE Journal of Solid-State Circuits*, vol. 54, no. 11, pp. 2947-2956, Nov. 2019

[70] R. K. Henderson *et al.*, "A 192x128 Time Correlated SPAD Image Sensor in 40-nm CMOS Technology," in *IEEE Journal of Solid-State Circuits*, vol. 54, no. 7, pp. 1907-1916, July 2019.

[71] A. R. Ximenes, P. Padmanabhan, M.-J. Lee, Y. Yamashita, D. N. Yaung, and E. Charbon, “A 256×256 45/65 nm 3D-stacked SPAD-based direct TOF image sensor for LiDAR applications with optical polar modulation for up to 18.6 dB interference suppression,” in *IEEE ISSCC Dig. Tech. Papers*, San Francisco, CA, USA, Feb. 2018, pp. 96–98.

[72] S. Lindner, C. Zhang, I. M. Antolovic, M. Wolf, and E. Charbon, “A 252×144 SPAD pixel flash Lidar with 1728 dual-clock 48.8 PS TDCs, integrated histogramming and 14.9-to-1 compression in 180nm CMOS technology,” in *Proc. IEEE VLSI Symp.*, Honolulu, HI, USA, Jun. 2018, pp. 18–22.

[73] D. Bronzi, *et al.*, "Automotive Three-Dimensional Vision Through a Single-Photon Counting SPAD Camera", *IEEE Transactions on Intelligent Transportation Systems*, 2016.

[74] The Engineering Toolbox, "Recommended light levels – illuminance, " [Online] Available: https://www.engineeringtoolbox.com/light-level-rooms-d_708.html, Accessed 2 February 2019.

[75] C. Niclass *et al.*, "A 100-m Range 10-Frame/s 340x96-Pixel Time-of-Flight Depth Sensor in 0.18µm CMOS", *IEEE JSSC*, 2013.

[76] G. Cossu, R. Corsini, A. M. Khalid, S. Balestrino, A. Coppelli, A. Caiti, and E. Ciaramella, "Experimental demonstration of high speed underwater visible light communications," *International Workshop on Optical Wireless Communications (IWOW)*, Newcastle upon Tyne, UK, pp. 11-15, Oct. 2013.

[77] T. Shafique, O. Amin, M. Abdallah, I. S. Ansari, M. S. Alouini, and K. Qaraqe, “Performance analysis of single-photon avalanche diode underwater VLC system using ARQ”, *IEEE Photon. J.*, vol. 9, no. 5, pp. 1-11, Oct. 2017.

[78] C. Wang, H. Y. Yu, and Y. J. Zhu, “A long distance underwater visible light communication system with single photon avalanche diode”, *IEEE Photon. J.*, vol. 8, no. 5, pp. 1-11, Oct. 2016.

[79] Image [Online] Available: https://aemstatic-ww1.azureedge.net/content/dam/lfw/online-articles/2012/07/1207LFW07f2web.gif, Accessed 2 February 2019.

[80] PCI-SIG, "PCIe 6.0 Specification: The Interconnect for I/O Needs of the Future," June 4, 2020. [Online] Available: https://pcisig.com/sites/default/files/files/PCIe%206.0%20Webinar_Final_.pdf, Accessed 29 December 2020.

[81] "The USB 3.2 Specification released on September 22, 2017 and ECNs". USB.org. USB Implementers Forum. September 2017.

[82] H. Kaeslin, *Digital Integrated Circuit Design: From VLSI Architectures to CMOS Fabrication*, Cambridge University Press, 2008.

[83] M. Cvijetic, *Optical Transmission Systems Engineering*. Artech House, 2003.

[84] M. Atef and R. Swoboda "1.25 Gbit/s over 50m step-index plastic optical fiber using a fully integrated optical receiver with an integrated equalizer", in *Journal of Lightwave Technol.*, vol. 30, no. 1, pp.118-122, 2012.

[85] Y. Koike, *Fundamentals Plastic Optical Fibers*, New York, NY, USA: Wiley, 2015.

[86] Y. Koike and T. Ishigure, “High bandwidth plastic optical fiber for fiber to the display,” *J. Lightw. Technol.*, vol. 24, no. 12, pp. 4541–4553, Dec. 2003.

[87] Y. Koike and A. Inoue, "High-Speed Graded-Index Plastic Optical Fibers and Their Simple Interconnects for 4K/8K Video Transmission," in *Journal of Lightwave Technology*, vol. 34, no. 6, pp. 1551-1555, March 15, 2016.

[88] A. Polley and S. E. Ralph, “40 Gbps links using plastic optical fiber,” *Optical Fiber Communication Conf.*, Montreal, QC, Canada, 2007.

[89] S. R. Nuccio, L. Christen, X. Wu, S. Khaleghi, O. Yilmaz, A. E. Willner, and Y. Koike, “Transmission of 40 Gb/s of DPSK and OOK at 1.55 μm through 100 m of plastic optical fiber,” *European Conf. Optical Communication*, Brussels, Belgium, 2008.

[90] C. Gimeno, C. Aldea, S. Celma and F. Aznar, "Cost-Effective 1.25-Gb/s CMOS Receiver for 50-m Large-Core SI-POF Links," in *IEEE Photonics Technology Letters*, vol. 24, no. 6, pp. 485-487, March 15, 2012.

[91] S. Loquai, F. Winkler, S. Wabra, E. Hartl, B. Schmauss and O. Ziemann, "High-Speed, Large-Area POF Receivers for Fiber Characterization and Data Transmission >10-Gb/s Based on MSM-Photodetectors," in *Journal of Lightwave Technology*, vol. 31, no. 7, pp. 1132-1137, April 1, 2013.

[92] Y. Dong and K. W. Martin, "A 4-Gbps POF Receiver Using Linear Equalizer With Multi-Shunt-Shunt Feedbacks in 65-nm CMOS," in *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 60, no. 10, pp. 617-621, Oct. 2013.

[93] Y. Dong and K. W. Martin, "A High-Speed Fully-Integrated POF Receiver With Large-Area Photo Detectors in 65 nm CMOS," in *IEEE Journal of Solid-State Circuits*, vol. 47, no. 9, pp. 2080-2092, Sept. 2012.

[94] X. Meng, *Transmission Capacity Improvement for Single Photon Avalanche Diode-based Optical Communication Using PAM Modulation* (Doctoral thesis), University of Cambridge, 2019. https://doi.org/10.17863/CAM.47343

[95] F. Tavernier and M. Steyaert, *High-Speed Optical Receivers with Integrated Photodiode in Nanoscale CMOS*, Springer-Verlag New York, 2011.

[96] S. M. Sze, *Semiconductor Devices: Physics and Technology*, 2nd Edition, 2nd ed. John Wiley and Sons, Inc, 2001.

[97] Image. [Online] Available: https://www.teamwavelength.com/wp-content/uploads/pin-photodiode-cross-section-1024x486.png, Accessed 1 December 2020.

[98] F. Tavernier and M. Steyaert, “High-Speed Optical Receivers With Integrated Photodiode in 130 nm CMOS,” *IEEE Journal of Solid-State Circuits*, vol. 44, no. 10, pp. 2856 –2867, 2009.

[99] 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.

[100] S. Csutak, J. Schaub, W. Wu, R. Shimer, and J. Campbell, “CMOS-Compatible High Speed Planar Silicon Photodiodes Fabricated on SOI Substrates,” *IEEE Journal of Quantum Electronics*, vol. 38, no. 2, pp. 193 –196, Feb. 2002.

[101] T. Jukic, B. Steindl and H. Zimmermann, “400µm Diameter APD OEIC in 0.35µm BiCMOS,” *IEEE Photon. Tech. Lett.*, vol. 28, no. 18, pp. 2004-2007, Sept. 2016.

[102] P. Brandl, R. Enne, T. Jukic, and H. Zimmermann, “OWC using a fully integrated, highly sensitive optical receiver with large-diameter APD,” *IEEE Photon. Tech. Lett.*, vol. 27, no. 5, pp. 482-485, Mar. 2015

[103] D. O’Brien, *et al.*, “High-speed optical wireless demonstrators: Conclusions and future directions,” *Journal of Lightwave Tech.*, vol. 30, no. 13, pp. 2181-2187, July 2012.

[104] Thor Labs, "APD430A2/M - Si Variable-Gain Avalanche Detector Manual," [Online] Available: https://www.thorlabs.com/thorproduct.cfm?partnumber=APD430A2/M, Accessed 17 December 2020.

[105] Hamamatsu, "S12023-02 Si APD Datasheet," [Online] Available: https://www.hamamatsu.com/resources/pdf/ssd/s12023-02_etc_kapd1007e.pdf, Accessed 1 December 2020.

[106] H. Zimmermann, "APD and SPAD Receivers: Invited Paper," *2019 15th International Conference on Telecommunications (ConTEL)*, Graz, Austria, 2019, pp. 1-5.

[107] S. Cova, M. Ghioni, A. Lacaita, C. Samori, and F. Zappa, “Avalanche photodiodes and quenching circuits for single-photon detection,” *Appl. Opt.*, vol. 35, no. 12, pp. 1956–1976, Apr. 1996.

[108] A. Eisele *et al.*, “185 MHz count rate, 139 dB dynamic range singlephoton avalanche diode with active quenching circuit in 130 nm CMOS technology,” in *Proc. Int. Image Sensor Workshop*, Hokkaido, Japan, Jun. 2011, pp. 278–281.

[109] S. Pellegrini *et al*., "Industrialised SPAD in 40 nm technology," *2017 IEEE International Electron Devices Meeting (IEDM)*, San Francisco, CA, USA, 2017, pp. 16.5.1-16.5.4.

[110] J. A. Richardson, L. A. Grant and R. K. Henderson, "Low Dark Count Single-Photon Avalanche Diode Structure Compatible With Standard Nanometer Scale CMOS Technology," in *IEEE Photonics Technology Letters*, vol. 21, no. 14, pp. 1020-1022, July 15, 2009.

[111] M. W. Fishburn, *Fundamentals of CMOS SPADs*, Ph. D. thesis, TU Delft, The Netherlands, 2012.

[112] S. Gnecchi *et al*., "Digital Silicon Photomultipliers With OR/XOR Pulse Combining Techniques," in *IEEE Transactions on Electron Devices*, vol. 63, no. 3, pp. 1105-1110, March 2016.

[113] Z. Ahmed, L. Zhang, G. Faulkner, D. O'Brien and S. Collins, "A Shot-Noise Limited 420 Mbps Visible Light Communication System using Commerical Off-the-Shelf Silicon Photomultiplier (SiPM)," *2019 IEEE International Conference on Communications Workshops (ICC Workshops)*, Shanghai, China, 2019, pp. 1-5.

[114] L. Zhang, H. Chun, G. Faulkner, D. O'Brien and S. Collins, "Efficient pulse amplitude modulation for SPAD-based receivers," *2018 Global LIFI Congress (GLC)*, Paris, France, 2018, pp. 1-5.

[115] B. E. Saleh, M. C. Teich, *Fundamentals of photonics*, John Wiley & Sons, Feb. 27, 2019.

[116] E. Sarbazi, M. Safari and H. Haas, "Photon detection characteristics and error performance of SPAD array optical receivers," *2015 4th International Workshop on Optical Wireless Communications (IWOW),* Istanbul, Turkey, 2015, pp. 132-136.

[117] Yichen Li; Videv, S.; Abdallah, M.; Qaraqe, K.; Uysal, M.; Haas, H. "Single photon avalanche diode (SPAD) VLC system and application to downhole monitoring", *Global Communications Conference (GLOBECOM)*, 2014 IEEE, pp. 2108-2113.

[118] R. M. Gagliardi, S. Karp, ***Optical communications***, Wiley, 1995.

[119] H. Mahmoudi, M. Hofbauer, B. Steindl, K. Schneider-Hornstein and H. Zimmermann, "Statistical Study of Intrinsic Parasitics in an SPAD-Based Integrated Fiber Optical Receiver," in *IEEE Transactions on Electron Devices*, vol. 66, no. 1, pp. 497-504, Jan. 2019.

[120] K. Dogancay, "Blind compensation of nonlinear distortion for bandlimited signals," in *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 52, no. 9, pp. 1872-1882, Sept. 2005.

[121] I. A. Njabeleke, P. K. Bhattacharrya, J. R. Leigh. “Mathematical modelling of nonlinear systems using power series - An overview,” in *IFAC Proceedings Volumes*, vol. 26, no.2, pp. 93-96, Jul. 1, 1993.

[122] Image [Online] Available: https://cdn.everythingrf.com/live/1570708165422_637063049661030727.png, Accessed 3 June 2019.

[123] S. H. Lee, R. P. Gardner. “A new G–M counter dead time model,” in *Applied Radiation and Isotopes*, vol. 53, no. 4-5, pp. 731-7, Nov. 15, 2000.

[124] J. H. Lee, I. J. Kim, H. D. Choi. “On the dead time problem of a GM counter,” in *Applied Radiation and Isotopes*, vol. 67, no. 6, pp. 1094-1098, Jun 1, 2009.

[125] B. Mulgrew, P. Grant, J. Thompson, *Digital signal processing: concepts and applications*, Macmillan International Higher Education, Nov. 11, 1999.

[126] E. Sarbazi and H. Haas, "Detection statistics and error performance of SPAD-based optical receivers," *2015 IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC),* Hong Kong, China, 2015, pp. 830-834.

[127] S. Gnecchi et al., "A Simulation Model for Digital Silicon Photomultipliers," in *IEEE Transactions on Nuclear Science,* vol. 63, no. 3, pp. 1343-1350, June 2016.

[128] P. M. Lavrador, N. B. de Carvalho and J. C. Pedro, "Evaluation of signal-to-noise and distortion ratio degradation in nonlinear systems," in *IEEE Transactions on Microwave Theory and Techniques*, vol. 52, no. 3, pp. 813-822, March 2004.

[129] A. D. Griffiths, J. Herrnsdorf, O. Almer, R. K. Henderson, M. J. Strain, M. D. Dawson, *High-sensitivity free space optical communications using low size, weight and power hardware.* Ithica, N.Y. Feb 4, 2019.

[130] K.J. Ebeling, *Integrated optoelectronics: waveguide optics, photonics, semiconductors*, Springer Science & Business Media, Dec 6, 2012.

[131] D. P. Palubiak and M. J. Deen, "CMOS SPADs: Design Issues and Research Challenges for Detectors, Circuits, and Arrays," in *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 20, no. 6, pp. 409-426, Nov.-Dec. 2014, Art no. 6000718.

[132] E. Charbon, C. Bruschini and M. Lee, "3D-Stacked CMOS SPAD Image Sensors: Technology and Applications," *2018 25th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, Bordeaux, France, 2018, pp. 1-4.

[133] E. Sarbazi, M. Safari and H. Haas, "The Impact of Long Dead Time on the Photocount Distribution of SPAD Receivers," *2018 IEEE Global Communications Conference (GLOBECOM)*, Abu Dhabi, United Arab Emirates, 2018, pp. 1-6.

[134] R. Waegemans, "Digital electronic predistortion for optical communications." PhD diss., UCL (University College London), 2010.

[135] C. He, Z. Ahmed and S. Collins, "Signal Pre-Equalization in a Silicon Photomultiplier-Based Optical OFDM System," in IEEE Access, vol. 9, pp. 23344-23356, 2021.