Edinburgh Research Explorer

Distortion Losses of High Speed SPAD Optical Receivers Approaching Quantum Sensitivity

Distortion Losses of High Speed SPAD Optical Receivers Approaching Quantum Sensitivity

John Kosman[1,2] *, Kevin Moore[2], Harald Haas[1] and Robert Henderson[1]

[1]*The University of Edinburgh, Edinburgh, EH9 3FF, United Kingdom;* [2]*STMicroelectronics, Edinburgh, EH3 5DA, United Kingdom*

Abstract

The high internal gain of single photon avalanche diodes (SPADs) operating in Geiger mode allows the quantum limit of detection to be approached. This offers a significantly improved sensitivity for optical communication over existing photodiodes. A fully integrated CMOS SPAD array receiver (RX) is presented which achieves 500Mb/s with pulse amplitude modulation in a visible light communication (VLC) link within 15.2dB of the quantum limit. However, the SPAD dead time induces around 5.7dB of transient distortion which restricts error performance and data rate. We propose a model describing a discrete photon counting system which exhibits this nonlinear behaviour and compare it to practical measurements with the RX. A unipolar intensity modulated optical signal is considered, as opposed to bipolar electric fields in conventional radio frequency wireless systems. Intermodulation between the DC and harmonic components of the data-carrying waveform is investigated, and the resulting degradation of signal-to-noise-and-distortion ratio and bit error rate is evaluated. The model is developed as a tool for understanding distortion to ultimately allow rectification through RX architecture, modulation scheme, coding and equalisation techniques.

Introduction

Integrated CMOS VLC receivers have recently been developed to enable miniaturised and low-cost links for light fidelity (LiFi) networks, underwater wireless optical communications (UWOC) and plastic optical fibre (POF) communications [1]. Sensitivity of these devices is constrained by electrical noise related to the employment of photodiodes (PDs) or linear avalanche photodiodes (APDs) and their amplification circuits [1, 2]. Fast APD RXs have improved sensitivities of –38dBm at 280Mb/s (850 nm) [2], –31.8dBm at 1Gb/s (675nm) [3] and –34.6dBm at 1Gb/s (675nm) [4] but remain two orders of magnitude above the standard quantum limit determined by photon shot noise [5]. This limit defines the minimum number of photon detections required to ensure a given bit error ratio (BER). The extremely high gain of SPADs in Geiger mode eliminates additive thermal noise in the RX chain since no transimpedance amplifier (TIA) is needed. This allows quantum sensitivity limits to be approached. A SPAD RX for fibre optic applications achieved 200Mb/s at 6.5×10^{-3} BER within 24dB of the quantum limit [6]. In this paper, we demonstrate a fully integrated 130nm CMOS SPAD RX extending this data rate by 2.5× to 500Mb/s with four-level pulse amplitude modulation (4-PAM) at 2×10^{-3}, whilst improving sensitivity to –46.1dBm, reducing the margin to the quantum limit to only 15.2dB [7].

*Author for correspondence (j.kosman@ed.ac.uk)

Address:

Scottish Microelectronics Centre, Institute for Integrated Micro and
Nano Systems, The University of Edinburgh, Edinburgh, EH9 3FF,
United Kingdom

100Mb/s on-off keying (OOK) is also reported with this RX at –55.2dBm sensitivity, corresponding to an average of 7.5 detected photons per bit, 13.4dB from the quantum limit [8]. These results meet 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 [9].

Effective detection rate is dependent on the dead time caused by quenching and recharging a SPAD after it detects a photon. During this time (3.5ns reported in [6], 12ns [7], 10.6ns [10] and 5ns [11]), 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 100Mb/s, where PD/APD RXs readily operate. Therefore, a single detector is unable to recover signals above this range with reliable BER – so multiple SPADs are required. Fig. 1 shows a representative block diagram of a photon counting RX with this principle compared to a typical PD RX. This presents design challenges to combine, count and sample the outputs of the SPADs with minimal losses and circuit area. [11] accomplishes this with an analogue silicon photomultiplier (aSiPM) of 60 SPAD currents combined via a node connected to a load resistor. Our RX architecture is digital and permits massively parallel (4096) photon event summation to be achieved at a high fill-factor (43%) and sample rate (800MHz). The RX operates in a practical, background insensitive VLC link at 1m in 1klx ambient conditions using a 450nm laser diode (LD). The large 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 (2.5ns at 400Mb/s OOK, 4ns at 500Mb/s 4-PAM) shorter than the individual SPAD dead time of 12ns. 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 [6, 10, 11]. 200Mb/s is attained in [6] with only 4 SPADs and a short 3.5ns 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, 1.5V to 3.5V excess bias), dark counts (1.46-13.9kHz) 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 [12]. Afterpulsing becomes more significant at shorter dead times and causes inter symbol interference (ISI). Instead, our SPAD array has an optimal 1.3V excess bias above 13.9V breakdown, 6kHz median dark count rate and an afterpulsing probability of around 1%, albeit with a longer mean dead time of 12ns. 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.

Another advantage of our large array is that complex modulation schemes such as PAM or orthogonal frequency division multiplexing (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. We apply adaptive modulation DC biased optical OFDM (DCO-OFDM) with 512 subcarriers and bit loading to attain 350Mb/s and 3.7b/s/Hz mean spectral efficiency at 2×10⁻³ BER. RX signal-to-noise ratio (SNR) (18dB peak near DC) is sufficient to convey 32-QAM (quadrature amplitude modulation) and 16-QAM with the lower frequency subcarriers, down to 4-QAM at subcarrier index 416. We attain 5.6b/s/Hz efficiency at a reduced rate of 200Mb/s.

With 230pJ/b at best performance, we prove the highest power efficiency of SPAD RXs published to-date [6, 10, 11]. Although the RX energy per bit is not lower than PD/APD RXs (ranging from 86pJ/b in [1] to 129pJ/b [3]), these do not include an ADC function, which is inherently realised by this SPAD RX architecture. Our value of 230pJ/b could readily be reduced by over an order of magnitude by adopting advanced CMOS nodes. Our sensitivities for 400Mb/s OOK, 350Mb/s DCO-OFDM and 500Mb/s 4-PAM all outperform that of PD, APD and SPAD RXs [1-4, 6, 10, 11] and our maximum bit rate is improved by a factor of 2.5 compared to [6] whilst maintaining 3.5×10⁻³ BER suitable for FEC with minimal bit redundancy [9]. The device reaches a sensitivity within 15.2dB of the quantum limit set by the Poisson statistics of photon arrivals [5]. This gap includes a 1.5dB power penalty due to the finite extinction ratio of the LD transmitter (TX) and an 8dB 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 [6]. The latter can be readily reduced with an optimised SPAD technology [13, 14] and structure techniques such as 3D stacking [15]. However, the remaining 5.7dB penalty is unaccounted for and attributed to the dead time inducing transient distortion and ISI. This causes a signal-to-noise-and-distortion ratio (SNDR) limitation which prevents data rate from being further improved by increasing the optical power of the signal. SPAD intrinsic parasitics evaluated in [16] are not as significant in our RX and too complex with a large array. With OFDM at low frequencies, where there is less signal attenuation, distortion is observed as effective SNR is more impaired from the total power. The dead time pile-up nonlinearity of the SPADs at high signal levels leads us to conclude that practical use of this device is to be in assistance (rather than replacement) of existing APD or PD RXs for LiFi, UWOC and POF applications; with potential to extend range or maintain connectivity in extreme conditions such as dark or diffuse environments.

The aim of this study is to advance understanding of the performance limits of a SPAD RX, to ultimately allow Gb/s data rates to be reached whilst enhancing sensitivity. This paper 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. 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 our integrated device to evaluate the performance of a SPAD RX in a VLC link.

Methods

RX Implementation

We report a 2.8mm by 2.6mm SPAD RX integrated in 130nm CMOS imaging technology which incorporates 64 × 64 receiver elements at 21µm pitch [7]. Each element contains a single p-well/deep-n-well SPAD biased at 15.2V (1.3V excess bias V_e) with a mean dead time of 12ns, a median dark count rate of 6kHz at room temperature and a photon detection probability (PDP) of 37% at 450nm. The 1.34 × 1.34mm 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. 2 shows a schematic of the chip architecture. A receiver element comprises of a SPAD interfaced to a 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 an XOR tree into an asynchronous double data rate (DDR) sequence at up to 900MHz (1.8Gphotons/s) limited only by wiring parasitics (not SPAD dead time).

The array divides into two sets of 64-row XOR trees feeding digital readout chains positioned on the flanks of the active area. Fig. 3 shows the row parallel interface circuit, consisting of three 8b 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 cycles 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, 16b synchronous sum of SPAD events. The entire digital readout operates from a sample clock generated by an on-chip PLL with programmable frequency up to 800MHz 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 receivers 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 summation is conveyed off-chip using two D-PHY transmission blocks, each dual lane, with line rates of 400Mb/s, for a total off-chip bandwidth of 1.6Gb/s. Fig. 4 is an annotated micrograph of the manufactured chip. In addition to the D-PHY blocks, an 8b 400MHz bandwidth current-steering digital-to-analogue converter (DAC) generates an analogue output compatible with existing ADC-based RX paths for testability. The total RX power consumption is 115mW, consisting of 15mW of SPAD power and 100mW of DSP.

We generate 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. These are transmitted with an Agilent 81180A Arbitrary Waveform Generator (AWG) driving an LD (Thorlabs PL450B, 450nm). A 450nm optical bandpass filter (10nm full width half maximum (FWHM)) and two neutral density filters (31% and 10% transmission) are placed above the array to attenuate ambient light and LD power. After frame synchronization, decoding of the received waveform is performed offline. The RX confirms photon counting reception at up to 500Mb/s from 1cm to 1m. Fig. 4 shows a block diagram of the setup. Experiments are conducted with the source directly in front of the RX to minimise reflections via the channel. We operate the device 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 the RX to attain high dynamic range and Poisson SNR up to 23dB per sample in theory. RX sensitivity from 100Mb/s OOK to 500Mb/s 4-PAM is compared to other visible light RXs in Fig. 5 and a constant distortion of around 5.7dB is observed. This penalty cannot be explained with current SPAD models, so a novel method of describing nonlinearity is required.

Intensity Response Theorem

We 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 signal reaches a limit due to supply voltage, for example [17]. 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. 6) due to an undesirable dead time caused by resetting the SPAD after it detects a photon. During this time interval (in the order of 10-14ns in [7]) the SPAD is unable to respond to impinging photons. A nonlinear system with input x and output y can be represented with a power series [18]

$$y(x,t) = a_1 x + a_2 x^2 + a_3 x^3 + \cdots \qquad (1)$$

where a_1 is the linear, small signal gain (unity in this case). We intend to describe the large signal curve of a SPAD (Fig. 6) 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. For a passive quench (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 [19] as a paralyzable detector [20, 21]:

$$y_{PQ} = Ae^{-A\tau} \qquad (2)$$

This model assumes a constant (DC) mean incident photon rate A. τ governs 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:

$$G_{PQ} = \frac{Ae^{-A\tau}}{A} = e^{-A\tau} = e^{-n} \qquad (3)$$

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. 6. As a means of quantifying gain compression, the input level where G_{PQ} has dropped by 1dB is found: $G_{PQ} = e^{-n} = -1\text{dB}$ when $n = 0.2303$. Furthermore, the output drops by 4.34dB when $n = 1$ and attenuation continues to grow exponentially as n increases, with –43.4dB a decade higher at $n = 10$. DC gain is plotted in Fig. 6 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,

$$x = A\cos(\omega_1 t) \tag{4}$$

A is the amplitude rate proportional to the incident optical power. The incident photon rate now varies in time for this input. From Eqns. (1) and (4), up to the third term, the output expansion produces

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

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

Unwanted higher harmonics (second $2\omega_1$ and third $3\omega_1$) are also generated in (5) and the 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:

$$G = \frac{y_{\omega_1}}{x} = \frac{\left(a_1 A + \frac{3a_3 A^3}{4}\right)}{A} = a_1 + \frac{3a_3 A^2}{4} = a_1\left(1 + \frac{3a_3 A^2}{4a_1}\right) \tag{6}$$

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

$$10\log_{10}\left(1 + \frac{3a_3 A^2}{4a_1}\right) = -1\text{dB} \tag{7}$$

If two tones are applied to the system,

$$x = B\cos(\omega_1 t) + B\cos(\omega_2 t) \tag{8}$$

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

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

where $\cos(k\omega_{1,2} t) = \cos(k\omega_1 t) + \cos(k\omega_2 t)$ and $k = 1, 2 \text{ or } 3$. 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 at approximately a_1. As B increases, the fundamentals increase proportionally, whereas the third order intermodulation (IM3) products increase in proportion to B^3. Eqn. (9) is modified by considering an intensity modulated/direct detection (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 (8):

$$x = B\cos(\omega_1 t) + B_{DC} \tag{10}$$

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

$$y = a_1B + \frac{3a_2B^2}{2} + \frac{5a_3B^3}{2} + \left(a_1B + 2a_2B^2 + \frac{15a_3B^3}{4}\right)\cos(\omega_1 t) + \left(\frac{a_2B^2}{2} + \frac{3a_3B^3}{2}\right)\cos(2\omega_1 t) + \frac{a_3B^3}{4}\cos(3\omega_1 t) \quad (11)$$

It is seen that intermodulation between the DC and signal components occurs and the higher order terms in (9) fold down to DC and the fundamental, adding distortion. Now, distorted DC and fundamental gains are obtained:

$$G_{DC} = \frac{y_{DC}}{x_{DC}} = \frac{\left(a_1B + \frac{3a_2B^2}{2} + \frac{5a_3B^3}{2}\right)}{B} = 1 + \frac{3a_2B}{2} + \frac{5a_3B^2}{2}$$

$$G_1 = \frac{y_1}{x_1} = \frac{\left(a_1B + 2a_2B^2 + \frac{15a_3B^3}{4}\right)}{B} = 1 + 2a_2B + \frac{15a_3B^2}{4} \quad (12)$$

Gain is no longer dependant on just a_3, as in (6), but a_2 as well. This means greater distortion occurs than for just a constant rate. Computing the second and third order coefficients which delineate both the DC and fundamental distortion,

$$10\log_{10}\left(1 + \frac{3a_2B}{2} + \frac{5a_3B^2}{2}\right) = -1\text{dB} \qquad 10\log_{10}\left(1 + 2a_2B + \frac{15a_3B^2}{4}\right) = -1\text{dB}$$

$$a_2 = \frac{\left(\left(10^{-\frac{1}{10}}\right) - 1 - \frac{15a_3B^2}{4}\right)}{2B} \qquad a_3 = \frac{\left(2\left(10^{-\frac{1}{10}}\right) - 2 - 3a_2B\right)}{5B^2}$$

Substituting,

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

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

$$a_3 = \frac{0.1645}{B^2} = \frac{0.7143n}{B^2} \qquad a_2 = -\frac{0.4113}{B} = -\frac{1.7859n}{B} \quad (13)$$

Unlike the conventional polar case (6), 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:

$$HD_2 = \frac{harmonic\ amplitude}{a_1B} = \frac{\left(\frac{a_2B^2}{2} + \frac{3a_3B^3}{2}\right)}{a_1B} = \frac{a_2B}{2} + \frac{3a_3B^2}{2} = 0.0411 = 0.1786n$$

$$HD_3 = \frac{\left(\frac{a_3B^3}{4}\right)}{a_1B} = \frac{a_3B^2}{4} = 0.0411 = 0.1786n \quad (14)$$

Substituting the coefficients (13) into the system polynomial (1),

$$y = x - \frac{1.7859n}{B}x^2 + \frac{0.7143n}{B^2}x^3 \quad (15)$$

With this general equation, the response can be determined for any input x, such as a sinusoid or data carrying OOK signal.

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 [18]. Consider a baseband signal $w(t)$ given by

$$w(t) = \sum_p a(p)s(t - pT) \quad (16)$$

where $a(p)$ is the information symbol sequence, p denotes symbol index, $s(t)$ is the pulse shape and T is symbol period. $a(p)$ is a stream of randomly generated bits in Matlab. $s(t)$ sets the duty cycle and rectangular NRZ signalling is considered, where the pulse amplitude is held constant throughout the symbol period: $s(t) = 1$ for $0 \leq t \leq T$ 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 via the means of an LD. Alternatively, a light emitting diode (LED) could be used. Assuming an ideal channel, let the peak-to-peak amplitude be A, such that the received signal x is

$$x = Aw(t) \tag{17}$$

This can be separated into a DC component B_{DC} and an AC swing of B. A TX with finite ER means symbol rate affects the 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 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 [22]:

$$\int_{-\infty}^{\infty} |x(t)|^2 dt = \int_{-\infty}^{\infty} |X(f)|^2 df \tag{18}$$

where $X(f)$ is the Fourier transform of $x(t)$, the output response can be determined this way. Photon noise is introduced to the model, but standard Poisson statistics cannot be assumed because the output count distribution is distorted. 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 [23]:

$$p_0(k) = \frac{(\lambda T)^k e^{-\lambda T}}{k!} \tag{19}$$

where λ is the 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 [24]:

$$\lambda = \frac{P_r PDE}{h\nu} \tag{20}$$

where PDE is the 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 [25]. 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 [26]:

$$p_K(k) = \sum_{i=k}^{k_{max}-1} (-1)^{i-k} \binom{i}{k} \frac{\lambda^i (T - i\tau)^i e^{-i\lambda\tau}}{i!} \tag{21}$$

for $k < k_{max}$ and $p_K(k) = 0$ for $k \geq k_{max}$. i is an integer index. PDE is defined as the product of PDP and fill factor C_{FF} and is treated as a constant 8dB loss (at 450nm, $V_e = 1.3$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. (21) 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 [27]

$$p_X(x) \sim \mathcal{N}(\mu_X, \sigma_X^2) \tag{22}$$

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 principle 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 $1 \rightarrow 0$ 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:

$$p_0(x) \sim \mathcal{N}(\mu_0, \sigma_0^2) \qquad\qquad p_1(x) \sim \mathcal{N}(\mu_1, \sigma_1^2) \qquad\qquad (23)$$

with variances σ_0^2 and σ_1^2. Decoding is implemented by comparing the received counts to a threshold x_T [27],

$$x_T = \frac{\mu_1 \sigma_0 + \mu_0 \sigma_1}{\sigma_1 + \sigma_0} \qquad\qquad (24)$$

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

$$BER \cong Q\left(\frac{\mu_1 - \mu_0}{\sigma_1 + \sigma_0}\right) = Q\left(\sqrt{SNDR}\right) \qquad\qquad (25)$$

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.

SPAD Simulator

We develop a simulation model, first presented in [28], to estimate the behaviour of a SPAD RX and compare this to the mathematical framework above. Matlab is used to simulate the stochastic events that occur in the photon counting system. With statistical analysis (21), it is shown that the counting process does not follow a Poisson distribution. As in (10), 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. Our 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 which describes the statistical detection process of a PQ SPAD. The physical parameters of the SPAD including dead time, PDP and V_e are coded in this 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 1b ADC. The block is initialised in a rest state at zero. Fig. 7 displays a block diagram of the simulation method. 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 800MHz 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 [29].

Results

Output Response

Using the simulator, the response of a PQ SPAD RX is estimated. Fig. 7 shows input and output spectra for a unipolar 1MHz cosine input signal at 10% and 100% of the maximum peak-to-peak event rate $A = 8.33 \times 10^7$ s^{-1}. Gain is determined from the difference between the input and output tones. 7dB output SNDR is computed from the output signal power and noise floor at 10% intensity, whilst arrival SNR = 9dB. A second harmonic (2MHz) appears at maximum input intensity and DC (–3dB) and signal gains (–5dB) are found from subtracting the corresponding magnitudes. At this intensity, SNDR = 4.7dB, arrival SNR = 19dB. Fig. 8 shows the output derived from Eqn. (15) together with simulated DC and signal gains found for different values of n. These estimates are plotted against the standard paralyzable model in Fig. 8. The corresponding gain is computed from the quotient of the output response calculated in (15) 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. Near matching between unipolar theory 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.

BER Performance

The photocount distribution of a SPAD as described in (21) is simulated and plotted in Fig. 9 with mean incident rate $A = \lambda = 8.33 \times 10^6$ s^{-1}, a decade below the maximum of 8.33×10^7 s^{-1}. Symbol rate R is increased from 1MBd ($T = \frac{1}{R} = 1\mu s$) to 10MBd ($T = 100ns$). $AT = 8.33$ photons/symbol for a logical '1' bit at 1MBd, which ensures 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 (21) simplifies to a Poisson form with mean λT when τ is set to zero. At 1MBd, the '1' level mean is 7 counts/symbol, which is slightly lower than the arrival mean, and this translates to –0.757dB average loss. This continues with means of 3 counts/symbol, 1 count/symbol and 0.8 counts/symbol at 2MBd, 5MBd and 10MBd 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 2MBd, 5MBd and 10MBd 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. Similar analysis is carried out at maximum intensity $A = 8.33 \times 10^7$ s^{-1}. At 1MBd, the '1' level mean is 31 counts/symbol, –4.29dB distortion loss from $AT = 83.3$ photons/symbol arrival mean. The distribution deviates from Poisson form by –4.29dB, –4.16dB, –3.77dB and –3.19dB with respect to the corresponding photons/symbol at each R. SNDR=8dB at 4Mb/s (25), so it is deduced that this is the maximum OOK bit rate of a single SPAD with these parameters to sustain BER for FEC. Fig. 10 shows simulated PMF distributions of a 64 × 64 array with $\tau = 12ns$, $A = \lambda = 8.33 \times 10^6$ s^{-1}, $C_{FF} = 0.43$ and $R = 100MBd$ to 400MBd (100Mb/s to 400Mb/s NRZ). 400MBd is chosen because it is the highest baud rate demonstrated by the SPAD RX in practice. Since the number of array elements is large, the output count distribution approaches a Gaussian form (23), as according to the Central Limit Theorem [27]. SNDR is calculated from Eqn. (25) with 5.78 ER (1.5dB penalty) added to the model. PMFs shown (Fig. 10 right) with the same parameters, but at maximum intensity $A = 8.33 \times 10^7$ s^{-1}. 100MBd 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 [30], since T is shorter than the τ of each SPAD.

We compare our mathematical framework and simulations to measurements obtained from laboratory experiments with our RX. At 400Mb/s OOK, 3×10^{-3} BER is reached at –49.9dBm with no equalization. The RX histogram is overlaid with estimated distributions in Fig. 10. 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 230Gphotons/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 1.5dB penalty. We sweep the intensity of the optical signal and measure the output count rate of a single detector element (Fig. 11). 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 = 6.8dB at $n = 0.3375$. The sweep is repeated with all SPADs in the array enabled (Fig. 12 top). Predicted BER from Eqn. (25) is plotted along with BER measurements and lines of best fit in Fig. 12 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. It can be seen that BER degradation occurs earlier (at –45dBm) than the paralyzable model (upper transfer curve) would indicate because of transient dead time distortion of the signal. This reduces SNDR to approximately 12dB a decade below the maximum incident rate, despite Poisson SNR = 18dB at this level (6dB penalty). SNDR values from numerical estimates, simulations and RX measurements are summarised in Table 1. Maximum expected SNR = 23dB at $n = 1$, assuming Poisson statistics and no data-carrying signal. Peak (Pk) effective SNDR is 9.6dB lower than this, at 13.4dB (when $n = 0.21$), which defines the BER inflection point. Therefore, 9.6dB of distortion occurs on the received OOK signal. –49.9dBm sensitivity at 400MBd could potentially be reduced to –55.9dBm if the 6dB 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.

Conclusion

The 130nm CMOS SPAD RX demonstrates 500Mb/s and a 15.2dB 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. Dead time remains as a 6dB limitation and performance is highly sensitive to changing signal intensity conditions, so a model is established to aid understanding and predict the resulting distortion. 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 our analytical model. Techniques such as predistortion at the TX [31] could be implemented to mitigate this distortion, with the metrics determined from our 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.
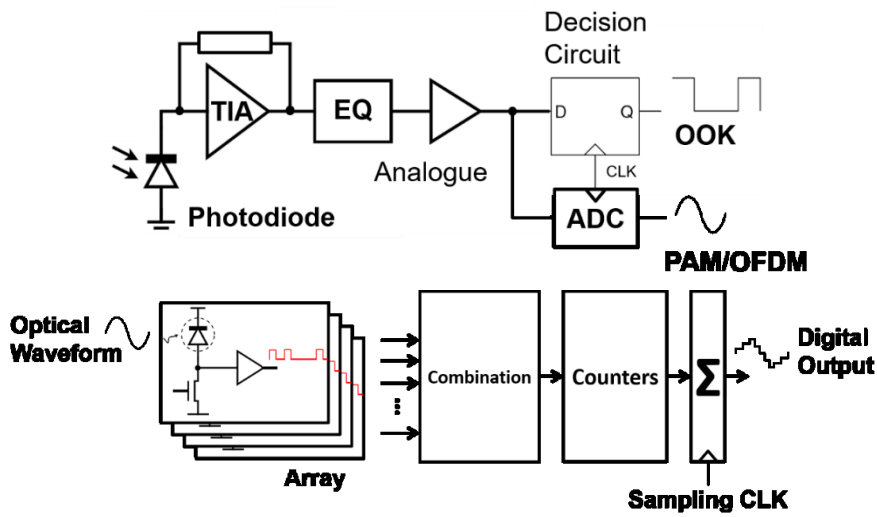
Figure 1: Block diagram of a conventional PD RX (top) and a photon counting RX with summed SPAD array (bottom).
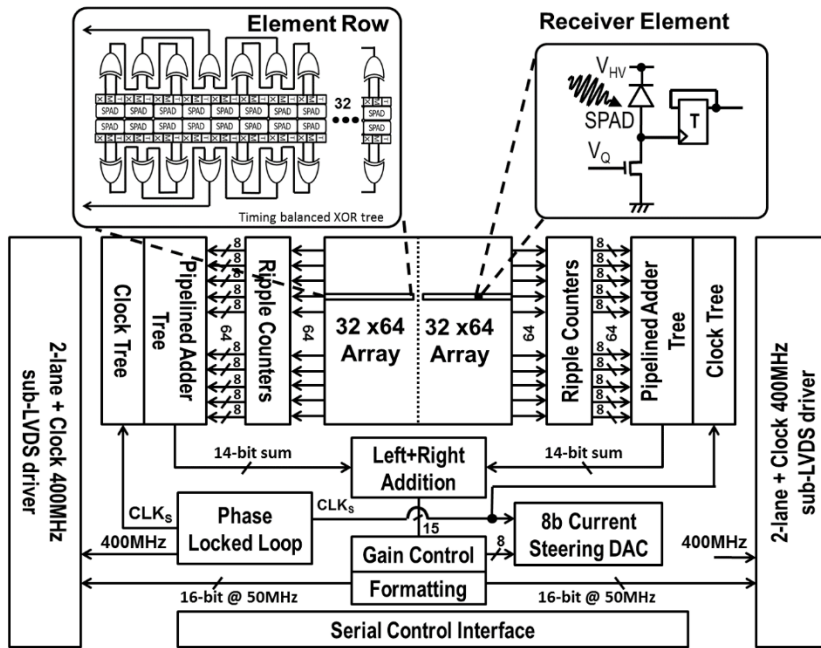

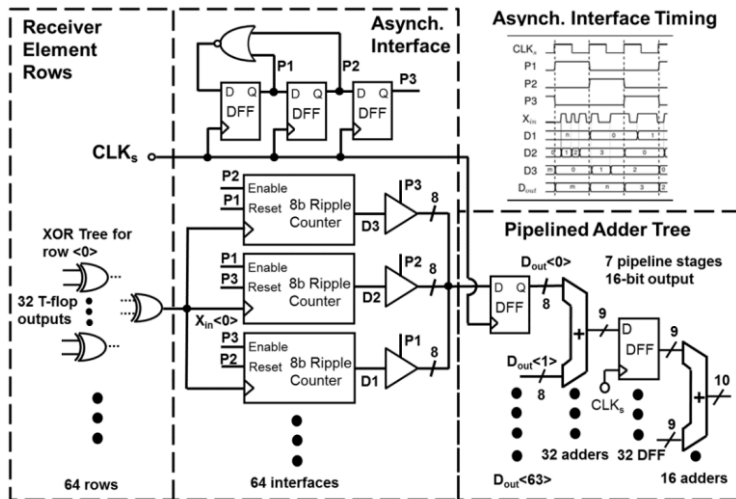Figure 2: Simplified chip block diagram.

Figure 3: Asynchronous interface between (right) SPAD XOR tree and pipelined added tree. Inset: interface timing diagram.
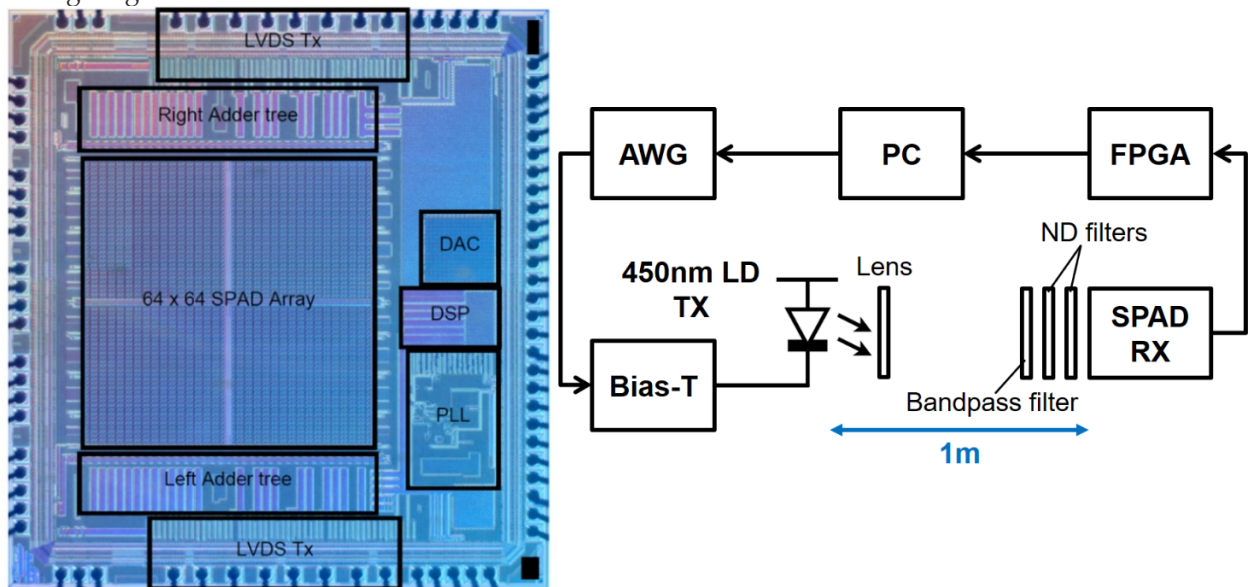

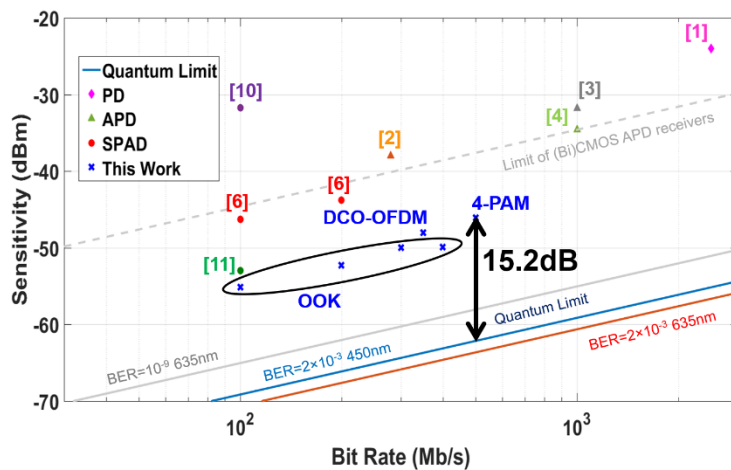Figure 4: Annotated chip micrograph and measurement setup.

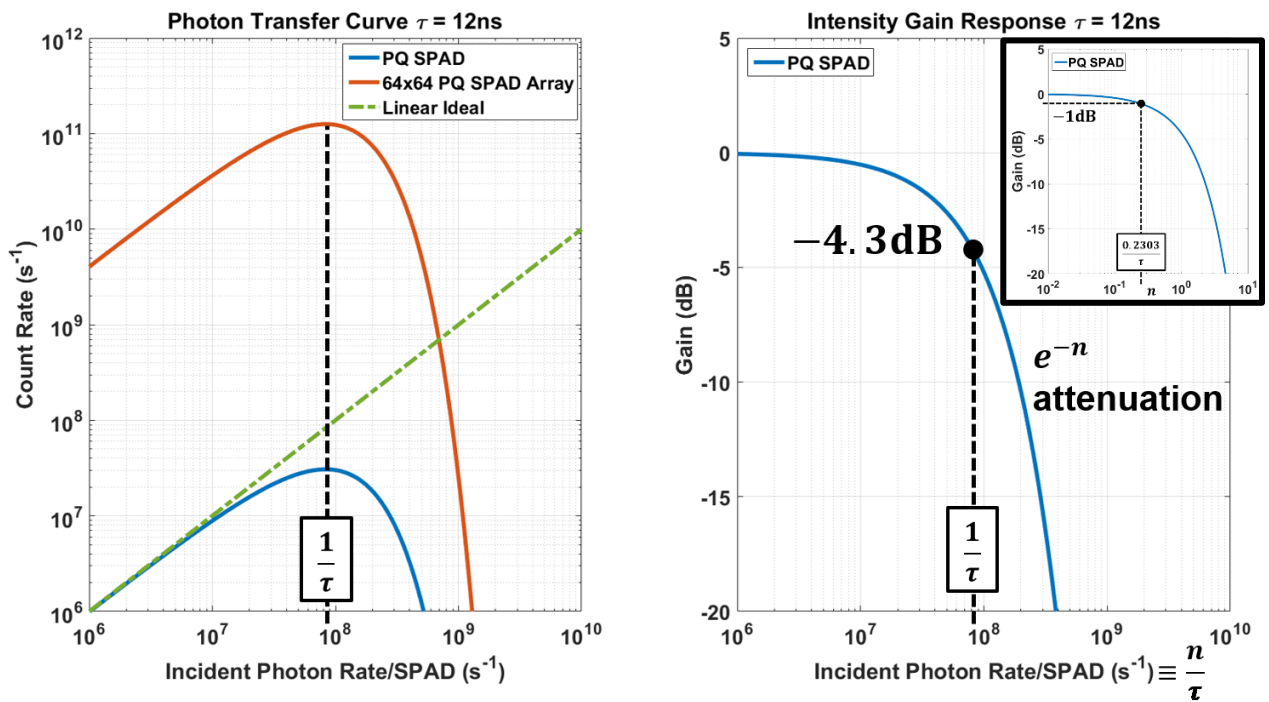Figure 5: Comparison of PD, APD and SPAD visible light RX devices.

Figure 6: Transfer curves of a single PQ SPAD and a 64 × 64 array (left). Corresponding gain (right). Inset: gain as a function of multiples n of $\frac{1}{\tau}$. −1dB compression point (highlighted) at $n = 0.2303$ for a DC arrival rate.
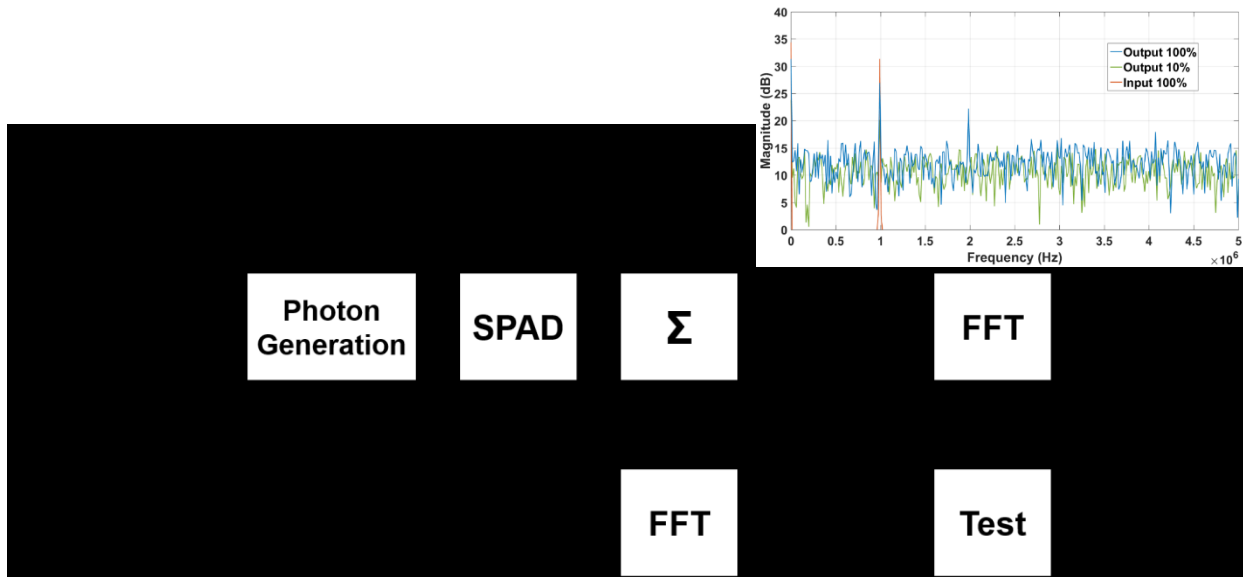
Figure 7: Simulation methodology. Input and output spectra for a unipolar 1MHz cosine input at 10% and 100% of maximum intensity (top right).
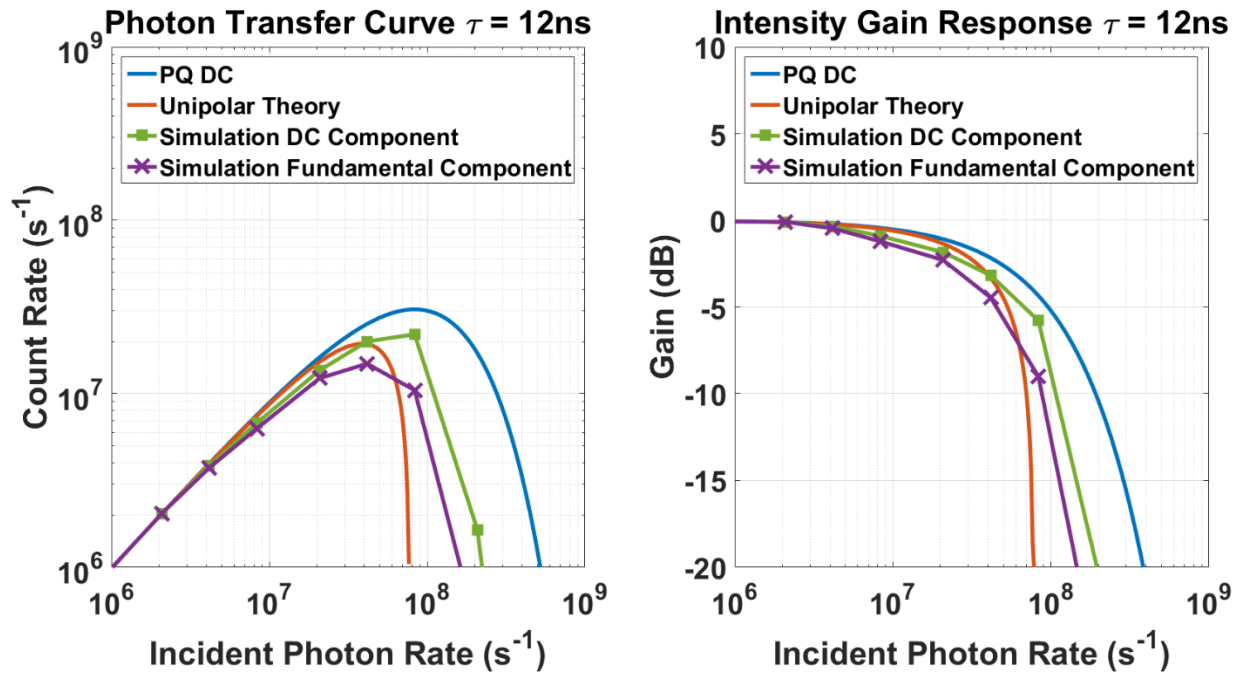
Figure 8: Comparison between the paralyzable SPAD model (with DC rate *A* equal to the peak-to-peak amplitude of the unipolar waveform), analytical prediction and simulation results.
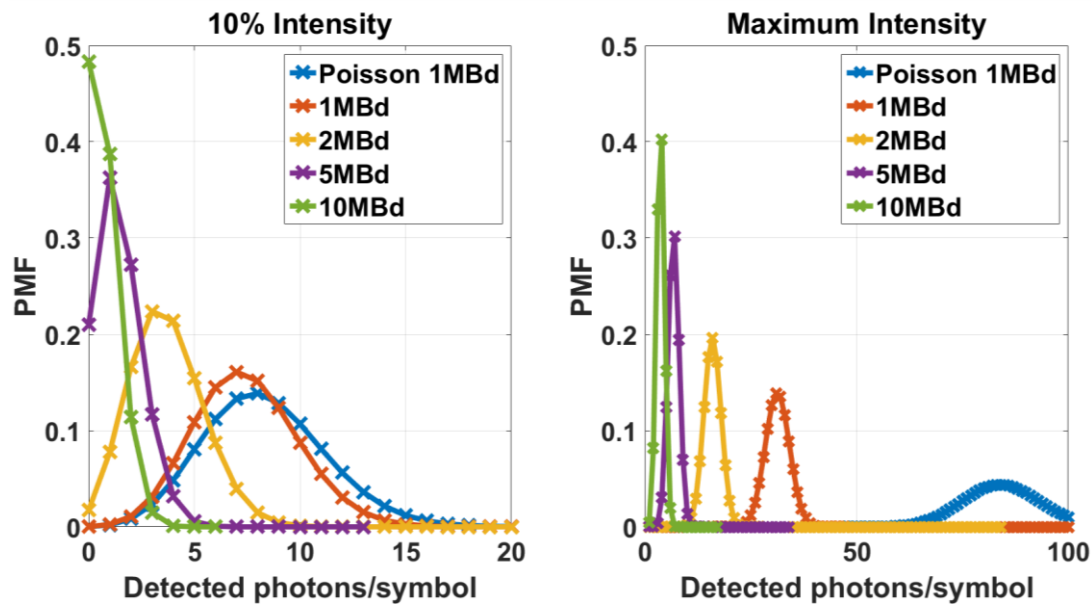
Figure 9: Simulated photocount distributions of a SPAD for $\tau = 12$ns, fixed mean $A = 8.33 \times 10^6$ s^{-1} (a decade below the maximum incident rate) on the left and $A = 8.33 \times 10^7$ s^{-1} on the right. $R = 1$MBd to 10MBd. 1MBd Poisson distribution without dead time (blue) for reference.
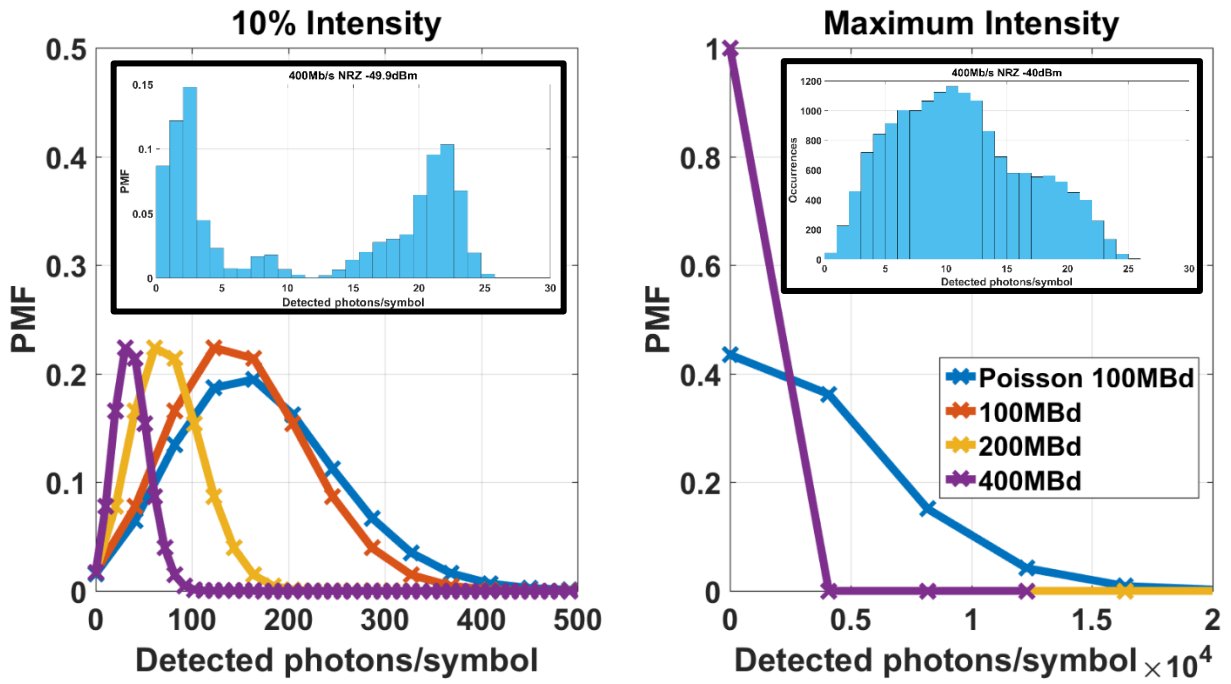
Figure 10: Simulated and measured PMF count distributions of a 64 × 64 array for $\tau = 12$ns, $A = 8.33\times10^6$ s^{-1} (left) and maximum $A = 8.33\times10^7$ s^{-1} (right) at $R = 100$MBd to 400MBd.
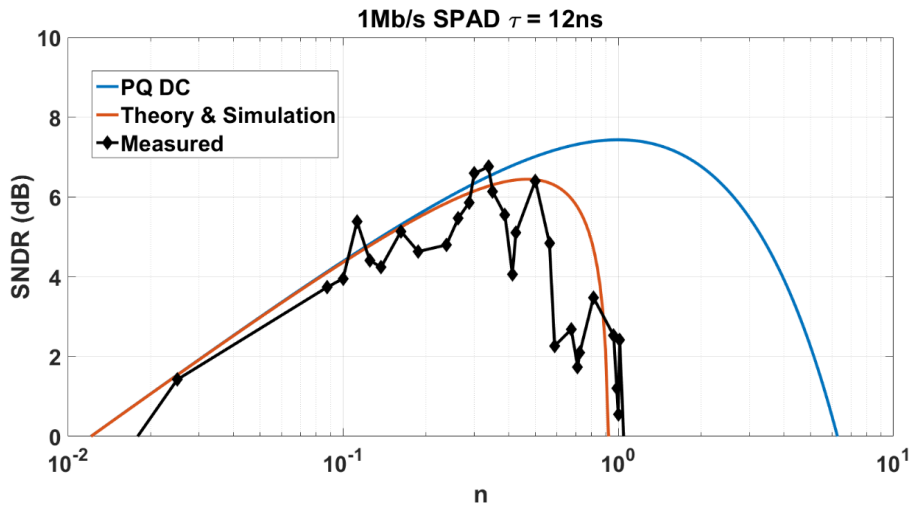
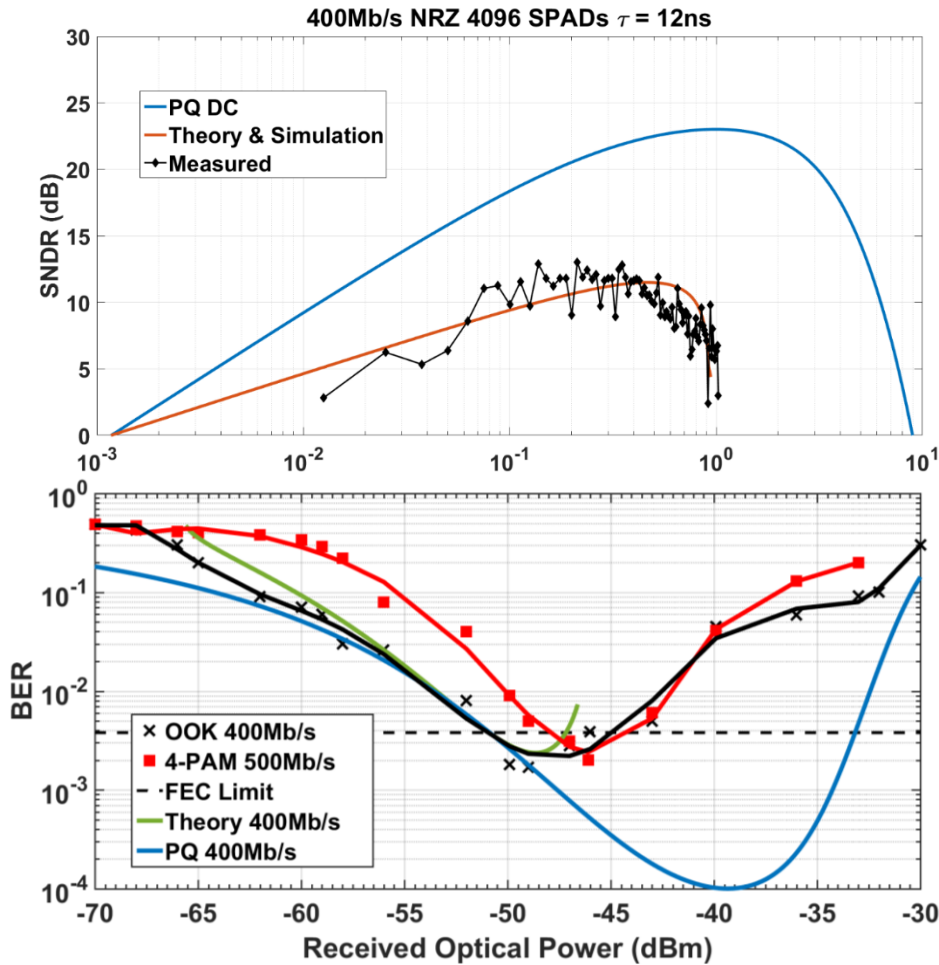Figure 11: SNDR response at 1Mb/s OOK with a single SPAD. Input waveform amplitude set by n. Peak SNDR = 6.8dB at $n = 0.3375$.

Figure 12: 64 × 64 array SNDR at 400Mb/s OOK against n (top). Theorized and measured BER with TX ER for 400Mb/s OOK and 500Mb/s 4-PAM (R = 400MBd and 250MBd respectively) with received power (bottom).

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	12	13.4

Table 1: Comparison of estimated and measured SNDR for one SPAD and a 64 × 64 array at low and high intensities.

Authors' Contributions
John was involved with the acquisition of data from the RX, which was designed by Robert and those listed in the acknowledgements. Robert also assisted John with the analysis of data. John drafted the paper and worked

on the theory and simulations with support and supervision from Kevin and Robert. Harald and Robert contributed to revising the article and final approval of the version to be published.

References

1. Fahs B *et al.* 2016. A 12m 2.5Gb/s Lighting Compatible Integrated Receiver for OOK Visible Light Communication. *Journal of Lightwave Tech.* **34**, 3768-3775. (10.1109/JLT.2016.2587598)

2. O'Brien D *et al.* 2012. High-speed optical wireless demonstrators: Conclusions and future directions. *Journal of Lightwave Tech.* **30**, 2181-2187. (10.1109/JLT.2012.2193874)

3. Brandl P, Enne R, Jukić T and Zimmermann H. 2015. OWC using a fully integrated, highly sensitive optical receiver with large-diameter APD. *IEEE Photon. Tech. Lett.* **27**, 482-485. (10.1109/LPT.2014.2382333)

4. Jukić T, Steindl B and Zimmermann H. 2016. 400μm Diameter APD OEIC in 0.35μm BiCMOS. *IEEE Photon. Tech. Lett.* **28**, 2004-2007. (10.1109/LPT.2016.2578979)

5. Ebeling KJ. 1993. *Integrated Optoelectronics* Berlin, Germany: Springer.

6. Zimmermann H, Steindl B, Hofbauer M and Enne R. 2017. Integrated fiber optical receiver reducing the gap to the quantum limit. *Scientific Reports* **7**, no. 2652. (10.1038/s41598-017-02870-2)

7. Kosman J *et al.* 2019. A 500Mb/s –46.1dBm CMOS SPAD Receiver for Laser Diode Visible-Light Communications. *IEEE ISSCC* **62**, 468-470. (10.1109/ISSCC.2019.8662427)

8. Griffiths AD *et al.* 2019. High-sensitivity free space optical communications using low size, weight and power hardware. (arXiv:1902.00495)

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

10. Fisher E, Underwood I and Henderson R. 2013. A Reconfigurable Single Photon-counting Integrated Receiver for Optical Communications. *IEEE J. Solid-State Circuits* **48**, 1638-1650. (10.1109/JSSC.2013.2253222)

11. Zhang L *et al.* 2018. A Comparison of APD and SPAD-Based Receivers for Visible Light Communications. *Journal of Lightwave Tech.* **36**, 2435-2442. (10.1109/JLT.2018.2811180)

12. Fishburn MW. 2012. *Fundamentals of CMOS SPADs* Ph.D. Dissertation, Dept. Elect. Eng., Delft Univ. Technol., Delft, The Netherlands.

13. Palubiak DP and Deen MJ. 2014. CMOS SPADs: Design Issues and Research Challenges for Detectors, Circuits and Arrays. *IEEE J. Selected Topics in Quantum Electronics* **20**, 409-426. (10.1109/JSTQE.2014.2344034)

14. Bronzi D, Villa F, Tisa S, Tosi A and Zappa F. 2016. SPAD Figures of Merit for Photon-Counting, Photon-Timing, and Imaging Applications. *IEEE Sensors Journal* **16**, 3-12. (10.1109/JSEN.2015.2483565)

15. Charbon E, Bruschini C and Lee M. 2018. 3D-Stacked CMOS SPAD Image Sensors: Technology and Applications. *25th IEEE ICECS*, 1-4. (10.1109/ICECS.2018.8617983)

16. Mahmoudi H, Hofbauer M, Steindl B, Schneider-Hornstein K and Zimmermann H. 2019. Statistical Study of Intrinsic Parasitics in an SPAD-Based Integrated Fiber Optical Receiver *IEEE Trans. Elect. Devices* **66**, 497-504. (10.1109/TED.2018.2882344)

17. Dogancay K. 2005. Blind compensation of nonlinear distortion for bandlimited signals. *IEEE Trans. on Circuits and Systems* **52**, 1872-1882. (10.1109/TCSI.2005.852936)

18. Njabeleke IA, Bhattacharrya PK and Leigh JR. 1993. Mathematical Modelling of Nonlinear Systems using Power Series. *IFAC Proceedings* **26**, 93-96. (https://doi.org/10.1016/S1474-6670(17)49082-3)

19. Eisele A *et al.* 2011. 185MHz Count Rate, 139dB Dynamic Range Single-Photon Avalanche Diode with Active Quenching Circuit in 130nm CMOS Technology. *Proc. Int. Image Sensor Workshop*, 278-281.

20. Lee SH and Gardner RP. 2000. A New GM Counter Dead Time Model. *Appl. Radiat. Isot.* **53**, 731-737. (10.1016/S0969-8043(00)00261-X)

21. Lee JH, Kim IJ, and Choi HD. 2009. On the Dead Time Problem of a GM Counter. *Appl. Radiat. Isot.* **67**, 1094-1098. (10.1016/j.apradiso.2009.01.074)

22. Mulgrew B, Grant P and Thompson J. 2002. *Digital Signal Processing: Concepts and Applications* Basingstoke, United Kingdom: Palgrave Macmillan.

23. Saleh BEA and Teich MC. 1991. *Fundamentals of Photonics* New York: Wiley.

24. Gagliardi RM and Karp S. 1995. *Optical Communications* New York: Wiley.

25. Sarbazi E and Haas H. 2015. Detection Statistics and Error Performance of SPAD-based Optical Receivers. *IEEE 26th Int. Symp. Personal, Indoor, and Mobile Radio Commun.*, 830-834. (10.1109/PIMRC.2015.7343412)

26. Sarbazi E, Safari M and Haas H. 2018. Statistical Modelling of Single-Photon Avalanche Diode Receivers for Optical Wireless Communications. *IEEE Transactions on Commun.* **66**, 4043-4058. (10.1109/TCOMM.2018.2822815)

27. Sarbazi E, Safari M and Haas H. 2015. Photon Detection Characteristics and Error Performance of SPAD Array Optical Receivers. *Proc. IEEE 4th Int. Workshop on Optical Wireless Commun.*, 132-136. (10.1109/IWOW.2015.7342281)

28. Gnecchi S *et al.* 2016. A Simulation Model for Digital Silicon Photomultipliers. *IEEE Transactions on Nuclear Science* **63**, 1343-1350. (10.1109/TNS.2016.2550494)

29. Lavrador P,de Carvalho N and Pedro J. 2004. Evaluation of signal-to-noise and distortion ratio degradation in nonlinear systems. *IEEE Trans. Microwave Theory and Techniques* **52**, 813-822. (10.1109/TMTT.2004.823543)

30. Sarbazi E, Safari M and Haas H. 2018. The Impact of Long Dead Time on the Photocount Distribution of SPAD Receivers. *IEEE GLOBECOM*, 1-6. (10.1109/GLOCOM.2018.8647814)

31. Waegemans R. 2010. *Digital electronic predistortion for optical communications* Doctoral Thesis, Dept. Elect. Eng., Univ. College London, United Kingdom. (http://discovery.ucl.ac.uk/id/eprint/20007)