Chapter 6

Single-Photon Avalanche Diodes in CMOS Technologies for Optical Communications

Edward M.D. Fisher

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.68935

Abstract

As optical communications may soon supplement Wi-Fi technologies, a concept known as visible light communications (VLC), low-cost receivers must provide extreme sensitivity to alleviate attenuation factors and overall power usage within communications link budgets. We present circuits with an advantage over conventional optical receivers, in that gain can be applied within the photodiode thus reducing the need for amplification circuits. To achieve this, single-photon avalanche diodes (SPADs) can be implemented in complementary metal-oxide-semiconductor (CMOS) technologies and have already been investigated in several topologies for VLC. The digital nature of SPADs removes the design effort used for low-noise, high-gain but high-bandwidth analogue circuits. We therefore present one of these circuit topologies, along with some common design and performance metrics. SPAD receivers are however not yet mature prompting research to take low-level parameters up to the communications level.

Keywords: single-photon, avalanche, diodes, visible, light, communications, receivers

1. Introduction

Optical technologies are routinely used in inter-continental communications, where both data rates and transmission lengths are high, or in fast networking applications such as data centres [1, 2]. At the opposite end of the distance scale, interconnects in microelectronics are moving towards photonics. This is due to the bandwidth restrictions of metals on silicon ICs. By using free-space links for personal computing, optical communication may offer an alternative to wireless standards, i.e. visible-light communications (VLC) applications [3, 4]. The radio spectrum is becoming increasingly crowded, with the UK's frequency allocation tables

(UKFAT) showing severe restrictions on wireless bandwidth, especially as 4G mobile internet is expanded towards 5G. Ultimately, the radio spectrum is limited at the high-GHz and low-THz bands by atmospheric attenuation, principally the molecular resonances of water and oxygen. There can be no doubt that as a mature technology, optical communication has become ubiquitous and is growing as bandwidth demands rise [5]. However, with that ubiquity and growth, the total electrical power, and indeed the materials and complexity of systems, is also increasing at an extraordinary annual rate.

The design of optical communication systems is highly application dependant [1, 2, 6]. The high-speed, long-distance intercontinental links can use complex, power-hungry and expensive transmitters (Tx) and receivers (Rx). This is principally as the energy per bit (J/b) and cost per bit ($/b) can be kept low through long-term installations with high throughput and multiple end-users. As a direct comparison, hardware for applications such as VLC to mobile phones and personal computers, must be simple, low-power and cheap [3, 4]. This is especially true as smart-phones or personal computers imply a single end-user and a bandwidth low enough for simple or minimal forward error correction (FEC) schemes. It also implies battery operation with an expectation of long battery life and short product lifetimes (e.g. the 2-year average for mobiles). In terms of costs, VLC transmitters and receivers must: (i) be comparable to other system level components, i.e. memory, processor or display, and (ii) be suitable for high volume production, i.e. > 100,000 units.

1.1. Basic receiver operations and research directions

Taking a high-level approach, all optical communication receivers must perform eight basic operations [1, 2, 6]:

1. Convert incoming optical signals into an electrical signal, usually electrical current.
2. Amplify and convert the current signal into a form that can be processed easily.
3. Equalize and/or modify the signal to increase bandwidth or remove unwanted artefacts.
4. Demodulate the incoming signal, depending on its transmission modulation.
5. Sample the processed signal to recover the transmitted digital bit stream.
6. Output the signal in a form suitable for the end-user's application.
7. Optionally recover the clock embedded within the data stream, and
8. Optionally perform error correction tasks.

The role of the receiver front-end covers only tasks #1 to #6, however as part of the sampling operation many receivers also perform a clock and data recovery (CDR) operation [1], i.e. task #7. As error correction (task #8) requires a memory buffer, such tasks are separated from a receiver front-end and incorporated within encoding and protocol specific units within the overall system. The above front-end and back-end split can be viewed conveniently with respect to Ethernet technologies. The front-end Tx/Rx is often denoted

as the 'PHY' as it relates to physical transmission and reception, it is also viewed as the physical layer in the Open Systems Interconnect (OSI) model. At the back-end, a Medium Access Controller (MAC) is used to implement the OSI data link layer, and contains both (i) transmit and receive buffers and (ii) framing, addressing and protocol level functions. Within the optical communication physical layer community, there are a number of key performance metrics that are driving innovation and the exploration of the current limiting factors [1, 2, 6]. These overall performance metrics are given below with current areas of innovation.

- Demand for increased bandwidth and transmission speeds
  - *Increase in bits per symbol through increased modulation complexity*
  - *Increase in channels per physical link though multiplexing in the time, frequency, phase or wavelength domains*
  - *Increased use of channel equalization techniques at both Tx and Rx*
- Reduction in energy usage and increased energy efficiency
  - *Decrease in the energy per bit*
  - *Trends towards quantum limited transmission (minimum optical power)*
  - *Increased transmission spectral efficiency for given real-world channel*
  - *Increased receiver sensitivity through (i) low-noise amplification, or (ii) optical and/or electrical coherent reception either homodyne or heterodyne detection*
- Decreased manufacture and design costs, and
  - *Increased Tx or Rx levels of CMOS integration, i.e. single-chip solutions*
  - *Decreased use of rare earth or high-cost materials*
  - *Minimization of bulky (for a particular CMOS node) analogue circuitry*
  - *Decreased design costs through Tx/Rx standardization*
- Decrease in bit errors and noise immunity
  - *Increased optical Tx extinction ratio*
  - *Decreasing Tx and Rx electrical noise*
  - *Decrease in inter-symbol interference*

1.2. Key receiver performance metrics

Before discussing the target application, a number of key receiver performance metrics need to be defined. These are (i) basic parameters at the communications level and (ii) measurements that can be made to assess the quality of a receiver, or its fitness for a particular application.

1.2.1. The data rate

The data-rate, R_D, measured in bits per second (b/s) is a measure of the serialized data throughput. As technology progresses, the throughput that is required at personal, commercial, national and international levels is increasing [5]. The data rate is intrinsically linked to the transmitter, channel and receiver bandwidth, BW, as each must pass all required frequency components in order to allow accurate, low-error decoding of the received signal back into binary data. In a communications receiver, the sampling point is chosen to be within the centre of the data symbol, thus allowing some protection from distortions on a signal's rising and falling edges. This leads to a lower frequency optimum bandwidth, BW_{-3dB}, which is related to the data rate [6], see Eq. (1). This is a mid-way point between (i) a low bandwidth that reduces input noise but produces low-pass inter-symbol interference, and (ii) a high bandwidth that captures all of a signal without distortion but also captures a large noise bandwidth and unnecessarily increases design specifications. For a 10 ns (100 Mb/s) data symbol, this would suggest a receiver bandwidth of 66.7 MHz. This is significantly below the knee frequency (Eq. (2)) of 250 MHz for a 2 ns rise time, T_R (10–90%) signal, i.e. 20% of a 10 ns 100 Mb/s data rate symbol.

$$\mathrm{BW}_{\text{-3dB}} = \frac{2\,R_D}{3} \tag{1}$$

$$F_{KNEE} = \frac{0.5}{T_R} \tag{2}$$

1.2.2. The bit error rate

The bit error rate (BER) is defined as the number of single-bit errors within a continuous bit stream [1, 2, 6]. Data links at 100 Mb/s to 10 Gb/s may utilize error correction and are thus specified with 'native', no-FEC, error rates at 1×10^{-9}, i.e. a single bit error within a transmission of one billion bits. For links with slower data rates or where the application incurs significant channel distortions (such as VLC), the use of FEC may be a requirement. For these cases, the designer aims for the best native BER performance but as this cannot be guaranteed due to unknown factors impinging on the communications channel, a worst case maximum BER is chosen. For many FEC algorithms there is a limit as to effective recovery, at an approximate BER level of 1×10^{-3}. The BER is both a specification of the link and a measurable performance parameter. As an example, a BER of 1×10^{-6} to 1×10^{-9} may be specified as part of a protocol standard, but a measured value of 1×10^{-6} at a particular incident optical power may also become improved at a higher optical power.

There are two critical issues for the BER, (i) signal noise and (ii) inter-symbol interference. Assuming the use of non-return-to-zero (NRZ) on-off-key (OOK) modulation, the error rate is related to the probabilities that the noise of a 'one' or 'zero' cross a threshold, N_{th}. As both zeros and one are subject to noise, both have amplitude distributions about their means, (N_0 and N_1), these will have standard deviations of σ_0 and σ_1 respectively. The threshold and bit error rate can be approximated, for Gaussian noise profiles by Eqs. (3) and (4).

$$N_{\text{th}} = \frac{\sigma_0 N_1 + \sigma_1 N_0}{\sigma_0 \sigma_1} \tag{3}$$

$$\mathrm{BER} = \frac{1}{2}\,erfc\left(\frac{N_1 - N_{\text{th}}}{\sigma_1 \sqrt{2}}\right) \tag{4}$$

Inter-symbol interference (ISI) is the bleed-through of symbols directly before and after, interfering with the reception of a bit. If we view the system as a low-pass filter of bandwidth, BW_{-3dB}, rectangular pulses become smeared in time with exponential rise and fall times [1]. As the signal is the supposition of multiple transitions the amplitude at the sampling point becomes corrupted by these tails. If the bandwidth decreases, these tails become longer in comparison to a bit period. Ideally, all ISI should have settled prior to sampling, however ISI acts to decrease the region within a symbol where robust sampling can occur.

1.2.3. Receiver sensitivity

For communications link budgets, it is useful to establish the receiver sensitivity [1]. This is defined as the incident optical power necessary for a receiver to reach the specified BER at the specified data rate, and is measured in dBm [1]. To ensure communication through a variety of signal attenuation factors, a high receiver sensitivity is required. We would therefore choose a receiver with a sensitivity of −20 dBm (0.01 mW) above a receiver with −5 dBm (0.32 mW). This would be particularly true if we knew the optical power was likely to be low at 0.05 mW due to perhaps low transmitter power and optical filtering. The sensitivity is dependent not only on the physical optical sensitivity, but also the received signal as a proportion of the receiver noise and inter-symbol interference. Thus, for a high-sensitivity receiver, (i) optical efficiency and electrical gain need to be high, while (ii) noise, inter-symbol interference and sampling phase-noise need to be low [1, 6].

1.2.4. The quantum limit

The receiver sensitivity can easily be reformulated in terms of the number of photons required per bit [1]. As this can be generalized to a theoretical receiver, we come across the theoretical ideal known as the quantum limit, QL. An ideal receiver would have zero noise (zero bits given by zero photon arrivals) and would receive single-photons with 100% efficiency. As photon arrivals follow Poisson statistics, the QL can be given by Eq. (5), where T_B is the symbol duration and $\gamma(T_B)$ denotes the average number of photons received per symbol. For a BER of 1×10^{-9}, approximately 20 photons are needed per symbol, which at 100 Mb/s ($T_B = 10$ ns), is 2×10^9 photons per second. For 650 nm (red) light this is 0.611 nW (−62.14 dBm).

$$\mathrm{BER}_{QL} = \frac{1}{2} e^{-\gamma(T_B)} \tag{5}$$

1.2.5. Energy per bit and power consumption

While the power dissipation of a receiver is a product of the supply voltage and receiver current, receivers with different data rates are often compared using the energy per bit, measured in Joules per bit (J/b) [1, 6]. As data rates are stretching beyond Gb/s, the energy per bit and therefore the total energy usage become significant issues [5]. Amplification and equalization circuitry often incur large energy costs; hence effort is being concentrated on efficient implementation. While reduced voltage supplies for integrated circuits (ICs) helps reduce power consumption, analogue design becomes more complex. Coupled with the bulkiness of robust analogue

circuitry in advanced CMOS nodes (<45 nm), there is a trend towards denser digital circuitry in place of analogue where this is achievable, (i.e. trends of delay-locked loops (DLLs) replacing phase-locked loops (PLLs)). A significant open question is how to achieve high receiver sensitivity while obtaining a low energy per received bit? As transmitter efficiency and overall power/bit should also be included in the transmission energy budget of data, it is in our interest to reduce the power required of the most inefficient sub-system, along with improving its efficiency.

1.2.6. Integration and silicon area usage

The degree of receiver circuit integration, directly impacts system level costs and performance. Older receivers required separate ICs for the basic receiver tasks discussed in Section 1.1, however increased integration allows higher bandwidths, lower power dissipation and significantly reduced system level design times [7]. As an example, integrating a CDR circuit with the receiver removes high-bandwidth, accurate traces from a printed circuit board (PCB). It also reduces the power that must be dissipated by the input and output pads of multiple ICs, and drastically improves jitter [1]. Likewise, by integrating a photodiode in the same substrate as a CMOS receiver, the capacitance and inductance of bond pads and wires between an external photodiode and IC can be removed and thus achieving higher bandwidth. As multi-channel and multi-port ICs become more prevalent, some circuit blocks can be shared, thereby increasing area and energy efficiency. While requiring increased configurability, and difficulty in testing and verification, there is significant drive towards reduction in the number of input/output (I/O) pads within a system, hence integration of sub-systems onto a single IC.

1.3. Target application: visible-light communications

Visible light communication is a free-space method emerging as an alternative to local wireless systems [3, 4, 8, 9]. It is intended for data rates between 100 Mb/s and 1 Gb/s and has recently been ratified by the institute of electrical and electronic engineers (IEEE) (IEEE P802.15.7) [10]. A significant thrust is the dual use of energy-efficient light emitting diode (LED) technologies for illumination and communications [3, 10] (**Figure 1**). The advantages of VLC are principally

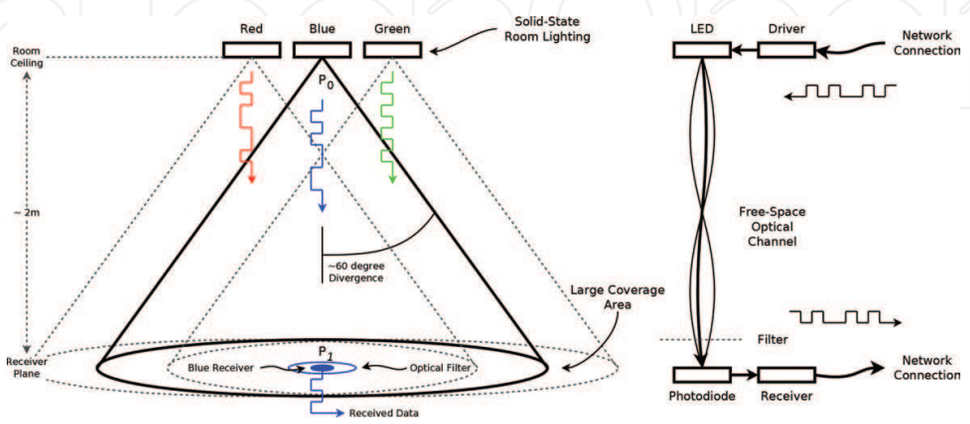

Figure 1. Visible optical communications and room lighting using red, green and blue modulated, LEDs. The wide divergence angle, presents an issue as the received optical power is low.

that the visible band is: (i) of wide bandwidth, (ii) is un-licensed and (iii) immune to electromagnetic interference from existing wireless technologies [1, 2, 6]. The primary target for end-users are mobile devices such as smart-phones and laptops where VLC can be used to supplement wireless systems for high bandwidth applications [3, 4] such as real-time, high-quality media streaming [5]. It is projected that as (i) the frequency allocation tables become increasingly dense and (ii) transmitters and receivers become increasingly difficult to produce at the high GHz radio levels, that some form of optical link—above the THz gap—will be required [3].

Despite advances in VLC, the application has a number of inherent issues that require application specific receiver designs. Firstly, the divergence angle of LEDs—particularly when wide-area lighting is needed—produces a large, diffuse illuminated area [3, 4]. When using a receiver IC of limited size, as per mobile systems, the receiver will capture a small proportion of the total transmitted optical power. Secondly, as optical sources are modulated at speeds close to their native bandwidth, both the average optical power and the modulation extinction ratio (the difference between high and low amplitudes) become smaller. Finally, in order to utilize the wide optical bandwidth, some form of optical filter is needed to perform wavelength multiplexing, i.e. multiple data channels on different colours. This has two implications, firstly that for ideally white illumination (wide-band), the receiver will receive a signal strength which is a narrow-band subset of wavelengths. A three-wavelength system (e.g. red, green, and blue, RGB), will naturally imply optical powers approximately one third of the total optical power. The second implication is that the filter itself will reduce optical power as they are neither perfect within their pass-band (signal attenuation) and have finite stop-band rejection (inter-colour interference and therefore finite separation). The effect of these issues, is that optical power incident on a receiver will be attenuated, while still being modulated at high data rates and with the user expecting both low error rates and low power consumption.

1.4. Research aims: extreme receiver sensitivity

We can consider VLC, as involving an inherent communications link budget issue, i.e. multiple attenuation factors necessitate reciprocal receiver gains. Further, the future requirements for low energy-per-bit communications would suggest an overall reduction in the link budget. The difficulty in such a scenario is twofold:

- First, the gain required in the communication budget may be extreme, especially if Gb/s links are required with multiple wavelength multiplexed channels, wide-area coverage and compact optically simple receivers.
- Second, as optical powers decrease and electrical gains increase, the signal-to-noise ratio (SNR) becomes a significant hurdle for robust detection. Johnson noise may limit analogue gain, especially if cooling is unachievable. Photon shot-noise may also become significant, without the possibility of using increased optical power at the transmitter (room lighting limited).

The work within this chapter therefore focuses on receiver designs with extreme sensitivity and high bandwidths to overcome the inherent attenuation factors of VLC, see [11, 12] and references therein. The overall method of using the avalanche multiplication of photo-generated carriers may use extra electrical power at the receiver—principally through increased biasing

voltages. But, as this removes the need for analogue amplification circuitry, represents a more economical use of both electrical power and silicon area.

2. High-sensitivity optical to electric conversion

When electromagnetic radiation is incident upon a material, there are three processes that can lead to photon interaction, and thus the detection of that photon [13]. These are the (i) photoelectric effect [14], (ii) Compton scattering and (iii) pair production. Taking the high-energy processes first, we can begin to discount processes in order to arrive at the process whereby optical, and in particular visible, sensors operate.

- In Compton scattering, the incident photon is scattered by an atomic electron. It imparts some energy to the electron, meaning that the photon energy is reduced, and thus the wavelength becomes longer. Compton scattering therefore does not destroy the photon. As noted in [13], this effect is small for energies "below tens of KeV", i.e. 1 KeV is 1.24 nm and 10 KeV is 0.124 nm, and therefore can be discounted from visible applications within the band 400–700 nm.
- In pair-production, the photon energy is high enough to result in the production of an electron-positron pair, i.e. the electron's anti-matter counterpart. The photon energy must therefore be higher than $E_p = 2m_e c^2$, where m_e is the rest-mass of an electron. As this is 1.02 MeV, i.e. 0.0012 nm, this process occurs only for X-ray and Gamma-ray interactions, and thus can be discounted from any visible light application [13, 14].

Elimination therefore leaves the photoelectric effect—which is subdivided (within textbooks, patents, company websites, whitepapers and journals) into the external and internal photoelectric effects. It is therefore the only physical process suitable for optical to electrical conversion within visible applications [13, 14].

2.1. The photoelectric effect: converting light

In both photoelectric processes, a photon is absorbed by an atom, thus destroying the photon. If the photon energy is sufficient to overcome the work function of the material, a bound-free transition takes place whereby an electron is promoted from an outer electron orbital and is ejected from the surface [13, 15]. Remaining energy is accounted for by the kinetic energy of the electron as a free particle [15]. This is the external photoelectric effect as photo-electrons physically leave the material. The photon energy is given by Eq. (6), where h is the Plank constant, c is the speed of light, λ is the wavelength and ν is the optical frequency.

$$E_P = \frac{hc}{\lambda} = h\upsilon \tag{6}$$

In contrast, for semiconductors with band-gap, E_G, between the valance, E_V, (i.e. outer-orbital bound electrons) and conduction electrons, E_C, (i.e. delocalized cloud of electrons) [16], a photon of energy greater than the band-gap ($E_P > E_G$) promotes an electron from the valence to the conduction band. As the absence of an electron in a valence state is described as a hole,

the internal photoelectric effect produces an electron-hole pair [16]. This is a bound-bound or intrinsic transition [14, 15]. The electron is still ejected from the atom; however, not from the surface. If two electrodes are placed on the material with a slight potential gradient, or if that potential exists due to a p-n doped junction, the electron-hole pair are separated and drift due to their relative charges. With many photo-generated carriers within the material due to many incident photons, the bulk conductivity of the material increases [16], allowing a photocurrent to flow through an external circuit. Photons of high energy are highly likely to cause band to band transitions, however as the wavelength increases towards a photon energy close to the band-gap, the likelihood of transition decreases, given by the absorption coefficient, α. This leads to a long-wavelength cut off, λ_C, given by Eq. (7). For silicon, this is 1.1 μm, where the absorption coefficient is 1×10^1 cm^{-1}, whereas at 400 nm it is 1×10^5 cm^{-1}.

$$\lambda_C = \frac{1.24}{E_G} \tag{7}$$

The internal photo-electric effect is therefore the mechanism whereby semi-conductor materials can be utilized for the detection of light. As materials have various band-gap energies, different materials can be used to detect optical energy with different wavelengths [16]. Typically, less than one electron-hole pair is produced per absorbed photon, fundamentally limiting the internal quantum efficiency and the spectral responsivity measured in Amperes per Watt (A/W). The optically-induced current, I_P, flowing through a photodiode is given by Eq. (8) [16], assuming a detector thickness much larger than the light penetration depth, $(1/\alpha)$. q is the electronic charge, P_{OPT}, is the incident optical power, μ_n is the electron mobility. η is the quantum efficiency, ε is the electric field within the photoconductor, L is the distance between the contacts, and τ is the carrier lifetime.

$$I_P = q\left(\eta \frac{P_{OPT}}{h\nu}\right)\left(\frac{\mu_n \tau \varepsilon}{L}\right) \tag{8}$$

Within CMOS technologies, photodiodes are fabricated using p- and n- type dopants to form a p-n junction [14, 16]. As the average depth of absorption changes with photon wavelength, the depth of the junction is chosen, if possible, to maximize the received photocurrent. The width of the p-n junction is also critical in this; however, it also has implications for photodiode bandwidth. The response speed is restricted by three phenomena. Firstly, the capacitance of the p-n junction (dictated by the junction width). Secondly, the time delay of carriers generated outside of the junction, diffusing into the junction. And thirdly, the drift or transit time, t_r, of the carriers within the junction [1, 6, 16]. The transit time is given by Eq. (9), where L can be replaced by the junction width, W, as the electrode contacts are often fabricated on the top and bottom of the junction within planar technologies.

$$t_r = \frac{L}{\mu_n \varepsilon} \tag{9}$$

Eqs. (8) and (9) indicate both limits and opportunities with respect to receiver design. If the junction width or photoconductor length are reduced, the response speed increases, although this becomes limited if a thin junction width leads to a high capacitance and therefore a slow photodiode resistance/capacitance (RC) time constant. Reducing the width also reduces the photocurrent, requiring a high receiver gain to compensate. If the electric field is increased,

both the photocurrent and diode bandwidth increase. It is here we encounter the fundamental limit and trade-off within high-speed, high-sensitivity receiver design. That transit time, and therefore bandwidth behaves in a reciprocal manner to photocurrent magnitude. As the electric field looks to be beneficial, this is a clue to guide research in the field.

2.2. The Avalanche gain mechanism: photo-carrier multiplication

A central challenge within receiver design is how to obtain the required amplification [16], or sensitivity. This must be done without undue use of (i) amplifier circuits (thus silicon area and power usage) or (ii) wide and therefore slow p-n junctions [1, 6]. Avalanche multiplication used in avalanche photodiodes (APDs), has been previously investigated for communications [1, 16–19]. This moves amplification into the diode providing initial gain without circuitry, and by increasing ε, allows both a larger photocurrent and shorter transit time.

The multiplication process uses an increased reverse bias voltage, V_R [16], creating an energy difference of $q(V_{bi}+V_R)$, where V_{bi} is the built-in potential between the p-side and n-side regions. The resultant increase in electric field, ε, accelerates a free carrier, labelled 1 in **Figure 2**, to a kinetic energy, E_K, sufficient to overcome the ionisation energy, E_G, of the material [16] (Eq. (10)). Upon a collision between a photo-carrier and the crystal lattice, the accelerated carrier ionises another carrier. An electron-hole pair, labelled 2 and 2′, is generated with those carriers then accelerated by the electric field, causing further ionisation [16, 17, 20]. This continues exponentially, creating an avalanche of carriers within the depletion region.

$$E_K = (1.5)\, E_G \tag{10}$$

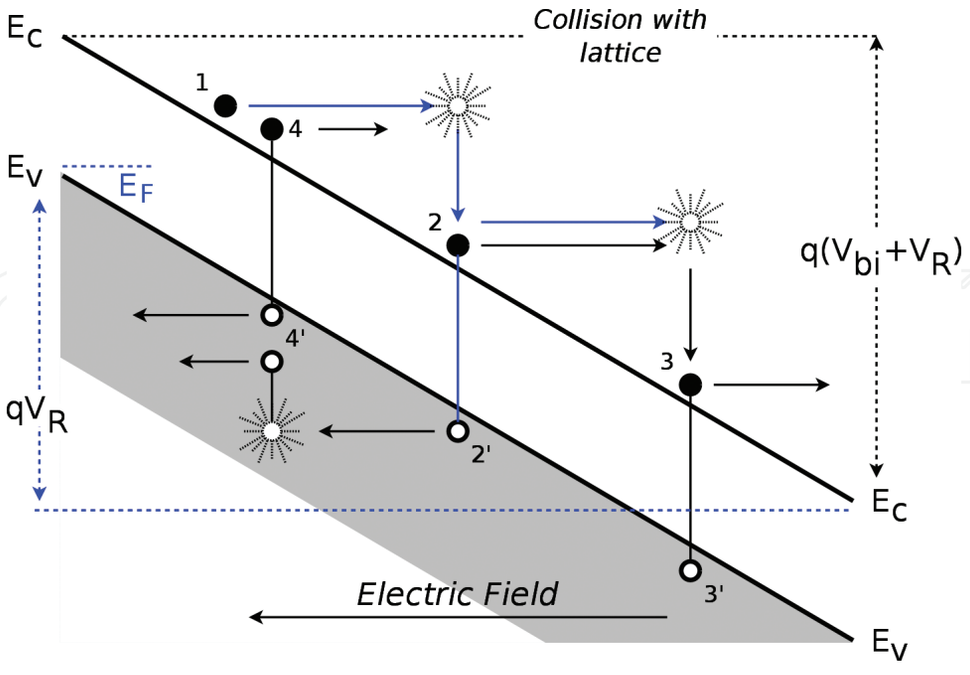

Figure 2. Band diagram showing reverse bias avalanche multiplication. When the electric field is elevated above the ionisation energy level, an accelerated carrier imparts a significant kinetic energy to a bound electron upon collision with the lattice. Adapted from Ref. [16], pp. 79 and 97.

APDs are biased such that multiplication achieves a constant generation rate, and thus constant gain [21] but not run-away avalanche leading to junction breakdown ([22] and references therein). The diode produces a current dependant on the incident photon flux with the gain and depletion region width being dependant on the bias [16, 21]. Some structures can be biased further into reverse bias, giving larger gains and greater sensitivities. In exceptional cases, run-away multiplication is used for specialised diodes called single-photon avalanche diodes (SPADs) [20, 22–25].

The use of avalanche gain, can be advantageous for sensitivity, as a gain of $M = 10–1000$, can be accomplished within the diode without the need of analogue amplification [1, 6]. Avalanche gain, is thus useful, however it comes with some disadvantages. Firstly, the higher bias voltage must be generated, necessitating additional power supply circuits or on-chip charge pumps. Secondly, multiplication is inherently random leading to increased noise. Finally, there is a gain-bandwidth trade off formed by the persistence of the avalanche process, whereby if the optical power decreases quickly, the avalanche caused by the previous high-optical power state takes some time to subside. To progress sensitivity and speed, through use of avalanche multiplication, circuits mitigating some of the above disadvantages are needed. For example, could we switch to smaller, lower power digital circuits and utilize very high multiplication factors by reducing the impact of multiplication noise?

2.3. Single-photon avalanche diodes: history and operation:

The Geiger region lies at the extreme end of the reverse photodiode range, beyond the linear avalanche gain region but prior to breakdown of a guard ring surrounding it ([11, 16, 21, 23, 25], and references therein). Initial research into avalanche behaviour, centered on microplasmas [21, 25]. These are small breakdown regions, corresponding to silicon defects [21, 25]. The historical study of microplasmas, resulted in artificial microplasmas [25] with guard rings to force a known breakdown region. These artificial structures later became known as Geiger-mode avalanche photodiodes (GM-APDs) or single-photon avalanche diodes (SPADs) [11, 23]. As they allow the avalanche mechanism to run-away, breaking down the p-n junction, their gain factors can be greater than 1×10^6, i.e. a single-photon, yielding a single electron-hole pair is able to produce a sizable avalanche photocurrent.

The Geiger region provides long periods of quiescent operation at a voltage of $V_{OP} = V_{BD} + V_{EX}$, where V_{EX} is the excess bias applied above the breakdown voltage, V_{BD} [20–23]. Upon the arrival of a photon, the device transitions from the quiescent point (**Figure 3**), to the I-V curve, with the avalanche quickly building up (10–100 ps). The current flows through any external series resistances, causing a significant voltage drop, designed to be of magnitude V_{EX} over that resistance. This reduction in voltage, (i) is sensed, indicating a single-photon event and (ii) reduces the voltage across the SPAD p-n junction [22, 23]. The reduced voltage, decreases the kinetic energy of carriers and halts the avalanche, a process known as quenching. The lower current flow allows the SPAD to recharge to $V_{BD} + V_{EX}$. There are two principal operating modes, passive quenching with passive reset (PQPR) and active quenching with active reset (AQAR) [23]. The passive circuit uses a large ~100 kΩ resistance (typically implemented using transistors), R_Q, and optionally a smaller 50 Ω sense resistor. The circuit recharges with a time constant given by $R_Q C_{SPAD}$, where C_{SPAD} is formed by (i) the p-n junction

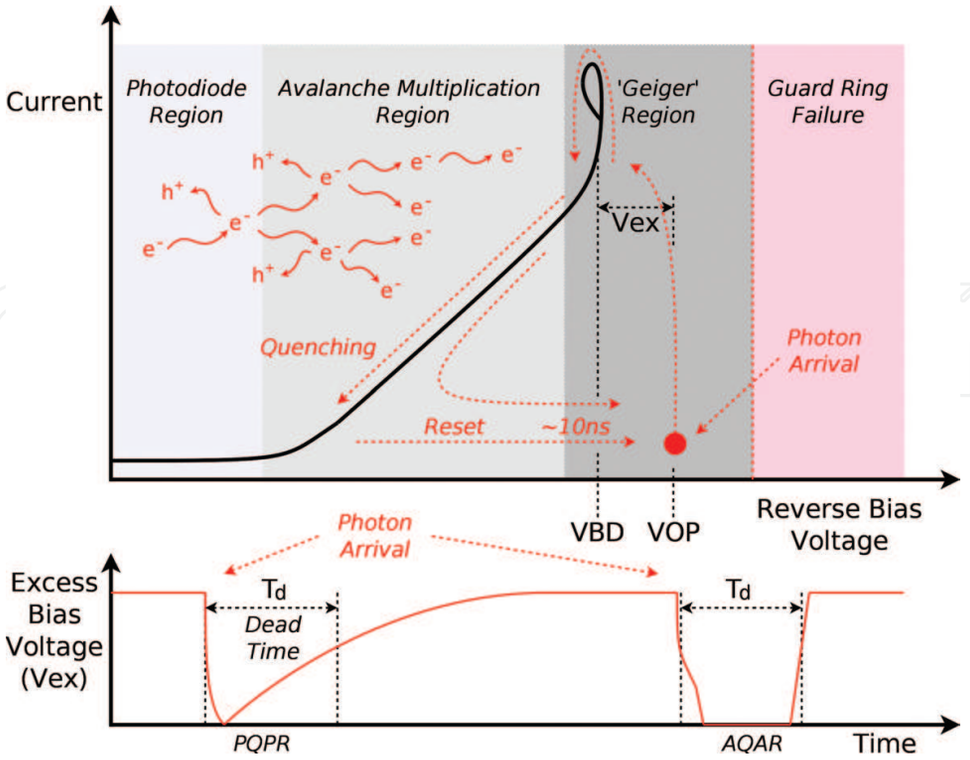

Figure 3. Top: the bias regions and I-V curve of a SPAD, showing the avalanche, quench and reset cycle. Bottom: the voltage-time pulses for both passive and active quenching circuits. Notice the steep, well defined edges of the AQAR pulse shape. The dead-time T_d is shown for both.

capacitance, and any parasitic capacitances on the diode (**Figure 4**). In the active case, the change in voltage triggers a quenching transistor which pulls V_{EX} to zero. After a short delay a separate low on-resistance transistor resets the device to V_{EX}.

In **Figure 4**, both positive-going and negative-going PQPR SPAD circuits are shown, along with the parasitic anode to substrate (C_{AS}) and cathode to substrate (C_{CS}) capacitances. The transistor gate voltage, DT_{ctrl}, can control the on-resistance, allowing adaptation of the SPAD dead-time, T_d. The n-type metal-oxide-semiconductor (NMOS) transistor circuit is preferable

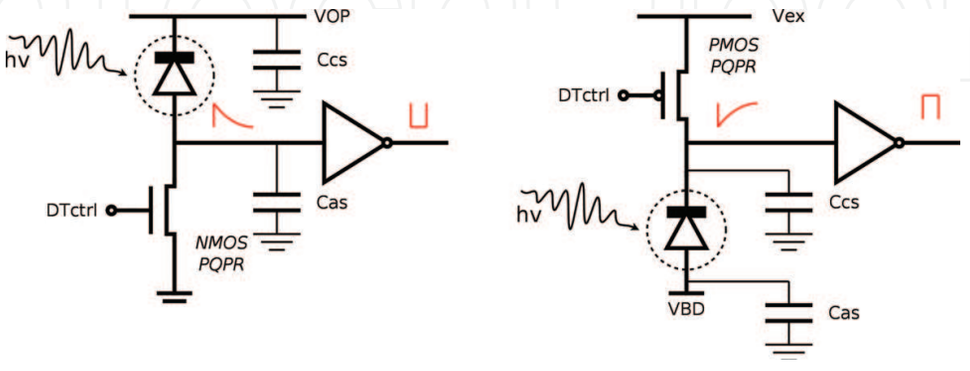

Figure 4. Two PQPR circuit topologies. Left: positive-going circuit with an NMOS quenching transistor and a small anode parasitic capacitance on the diode moving node. Right: negative-going circuit with a PMOS transistor, but with a larger cathode to substrate parasitic capacitance.

as (i) C_{AS} is typically smaller than C_{CS} and (ii) NMOS transistors can be integrated within a SPAD array without the deep N-well required for p-type metal-oxide-semiconductor (PMOS) devices.

It is with circuits such as this, that high-sensitivity receivers can be realized [11, 12]. One circuit approach would be to combine avalanche gain exceeding that used in APDs, with topologies known to give high sensitivity. The SPAD circuit can be used to replace the continuous-time photodiode, trans-impedance amplifier (TIA) and feedback network, Rf, circuitry (**Figure 5**, left). A small digital inverter, necessary to prevent loading on the SPAD, replaces any post-amplification, while the resulting full logic swing can be used with a digital counter (**Figure 5**, right). This discrete time "direct-to-digital" approach allows all gain to be implemented within the diode, removing complex analogue circuitry, and drastically reducing area requirements. Electrical power is expended within the SPAD due to a total voltage of V_{OP}, and the avalanche current flow per detection, however analogue amplifier circuits typically require continuous bias currents for each amplification stage branch [1, 6].

2.4. Single-photon avalanche diodes: key performance parameters

Before discussing SPADs within optical receivers, there are several key performance parameters that must be discussed. As the field is still under active investigation, there are as yet no definitive methods to bring low-level parameters up to the communications level, particularly the bit error rate (BER). This contrasts with mature photodiode receivers where the receiver BER can be derived [1, 2, 6]. Taking the properties of a single SPAD first, we encounter non-ideal behaviour away from true single-photon counting [23]. For receivers, the concept of the receiver power penalty within the link budget [6], can be used to estimate the effects, to first order, of various offsets and non-idealities. The power penalty, expressed in decibels (dB), models the adjustment in optical power required to compensate for an effect.

1. SPADs produce avalanche events despite the absence of light, this is the so-called dark count rate (DCR), measured in counts per second (cps) or Hertz (Hz) [20, 21, 23, 24]. This is related to (i) thermal generation of carriers, and (ii) tunnelling of carriers predominantly due to traps within the band structure. The DCR has steadily improved through (i) academic investigation of guard ring structures such as retrograde guard rings [26], (ii) modelling and understanding of the transient behaviour of the junction itself [27] and (iii) CMOS process cleanliness, and can now be as low as 1 Hz [27]. A per-SPAD rate such as this is now effectively ideal in

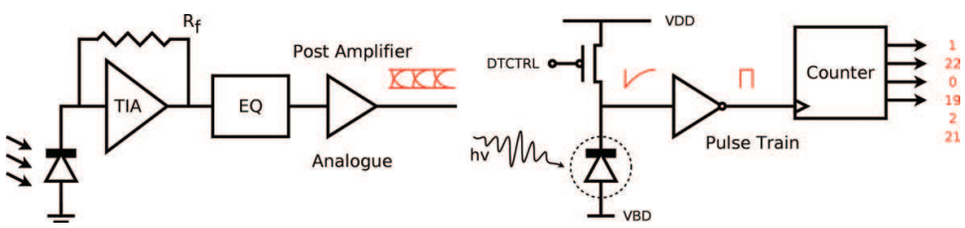

Figure 5. Left: a continuous-time conventional photodiode receiver, and right: a discrete-time replacement with single-photon sensitivity, where the SPAD and inverter allows optical to electrical conversion into a voltage (i.e. trans-impedance) with full logic swing magnitudes.

comparison to high (>100 Mb/s) data-rate communications. The DCR limits the optical photon flux at which reliable single-photon detection can be achieved. Therefore, the incident optical power would need to be increased by a power penalty, PP_{DCR}. This is given by Eq. (11), where DCR_{PC} is the DCR per channel per symbol period as a proportion of the required number of photons per symbol, similar to the voltage threshold offset power penalty ([1], p. 71). As the DCR and symbol period decrease, this penalty becomes negligible. For a DCR_{PC} of 10% of the required photons (~20), the PP_{DCR} is 0.79 dB.

$$PP_{DCR} = 10\log_{10}\left(1+\frac{2\,DCR_{PC}}{100}\right) \tag{11}$$

2. A SPAD—like a photodiode or avalanche diode—does not detect light with 100% efficiency. The photon detection efficiency (PDE), typically some 20–40%, is linked to three processes, (i) the internal quantum efficiency [16], (ii) the wavelength specific absorption depth and the width and depth of the p-n junction, and (iii) the avalanche turn-on probability [20, 23, 25]. This latter factor is a product of the kinetic energy achievable before a carrier recombines, and is thus dependent on the electric field. In combination with the DCR, this limits the lower optical power that can be reliably detected. A power penalty, PP_{PDE}, to account for the finite detection efficiency can be calculated using Eq. (12). For a SPAD with 20% PDE, this is 7 dB.

$$PP_{PDE} = -10\log_{10}\left(\frac{PDE}{100}\right) \tag{12}$$

The PDE has increased through commercial R&D into SPAD-dedicated micro-lenses. However, it has proved difficult to improve the internal quantum efficiency as increased electric fields correspondingly increase the DCR, while wider depletion regions require modifications to dopants that may not be an option for a multi-customer CMOS foundry. Correspondingly, the drive for high-PDE has led to the optimisation of the layer stack above the silicon surface, including reduced metallisation, thinner nitride layers and anti-reflection coatings, or the use of back-side illuminated structures. The combination of these design options has promoted foundries such as ST Microelectronics and TowerJazz to develop dedicated optical CMOS processes for SPAD implementation.

3. The dead time, T_d, is a discharge and recharge period, during which the bias falls below the normal $V_{EX} + V_{BD}$ [20, 21, 23]. No photon arrivals can be detected, although for PQPR as the voltage exponentially recharges there is an increasing probability of an avalanche event, and thus the dead time can be extended, an effect called paralysis [28, 29]. A short dead time is advantageous as it maximizes the detections per second, C_{MAX} (Eq. (13)). Unfortunately, an arbitrarily short dead-time, cannot be achieved as second order effects become more pronounced [23]. The dead time has been reduced within the literature, principally through the use of AQAR, and the use of smaller diodes. For AQAR, the maximum count rate and the dead time required for a target minimum number of photons per second, $N_{PHOTONS}$, for a number of SPADs, N_{SPADS} and target number of channels, N_{CHANS}, are given by Eqs. (13) and (14) respectively. For passive quenching, both equations are multiplied by 1/*e* [28, 29].

$$C_{MAX} = \frac{N_{SPADs}}{N_{CHANs} T_d} \tag{13}$$

$$T_d = \frac{N_{SPADs}}{N_{CHANs} \lfloor N_{PHOTONs} \rfloor} \tag{14}$$

4. Without a long dead-time, there is a small but finite probability of a secondary avalanche, denoted as an after-pulse, ap [20, 23]. This is related to the release of a charge carrier from a trap within the device. The number of trapped carriers is dependent on the charge flow during an avalanche, hence reducing SPAD capacitance is beneficial [22, 24]. The number of traps is a function of process cleanliness and device structure, while the trap lifetime, and hence the approximate required dead-time is a function of temperature. A minority-carrier effect, may limit the lower-bound of the dead-time with respect to after-pulsing, and hence fundamentally limits reductions in dead-times [22].

5. The Poisson distribution of photon arrivals is complemented by a similar timing variation in photon detection [20, 26]. This event timing jitter is a product of two contributions. Firstly, the statistics of avalanche build up [30], lateral spreading of the avalanche throughout the junction volume and the carrier transit time [31]. Secondly a lengthy, but low-probability jitter tail is caused by carriers deeper within the substrate diffusing towards the p-n junction. The use of silicon-on-insulator (SOI) CMOS processes may offer some mitigation, however a significant trend for jitter reduction has been SPADs of small diameter and the tailoring of the SPAD doping.

Arrays of SPADs in CMOS planar processes also have a number of non-ideal behaviours and limitations. As the SPAD dead-time or IC application may necessitate the use of multiple diodes, array level limitations have a large impact on final receiver performance [11].

6. The fill factor (FF) is the ratio of the optically receptive area to the total device area and is a key parameter in obtaining the goal of a high sensitivity receiver. For a square array of circular SPADs, the fill factor (FF_{SQ}) is given by Eq. (15). Where A_{ACT} is the total optically active area, A_{ARR} is the total square array area and N_{SPAD} is the number of SPADs within the array. D_R is the radius of the per-SPAD active area, D_G is the guard ring width, and D_S is the guard-ring to guard-ring separation. The fill factor significantly alters the receiver sensitivity as photons incident on inactive regions are not counted but non-the-less are included in link budgets. This leads to a fill-factor power penalty, PP_{FF}, given by Eq. (16) [11]. The fill factor is often limited by the diode geometries available for a given noise or capacitance specification [24], the complexity integrated onto the receiver and the CMOS process design rules [22, 24].

$$FF_{SQ} = \frac{A_{ACT}}{A_{ARR}} = \frac{\pi D_R^2 N_{SPAD}}{\left[2 D_R \sqrt{N_{SPAD}} + (2 D_G + D_S)\left(\sqrt{N_{SPAD}} - 1\right)\right]^2} \tag{15}$$

$$PP_{FF} = -10 \log_{10}\left(\frac{FF}{100}\right) \tag{16}$$

Prior to 2010, the fill-factor was often limited to 1–2% [11]. This necessitates a factor of 50–100 increase in optical power (PP_{FF} = 17 dB), to return the detected photons per data symbol to the

ideal (*FF* = 100%) theoretical value. Despite restrictions [24], the fill factor has been increasing through a number of techniques. N-well sharing, out-of-array electronics, reduced quenching circuit areas, NMOS-only logic, increased diode radii with retained low-noise, use of clustered mini arrays, non-circular geometries and micro-lenses, have all lead to fill factors reaching recent maximums of 67% [32] and 70% [33]. With a fill factor such as 70%, the optical power would need to be increased by a factor of 1.43 (PP_{FF} = 1.55 dB). To date, the fill-factor has been a particular focus of both academic and commercial R&D with several companies becoming heavily invested.

7. Detector cross talk, which is often at the 1–2% level in SPAD arrays using out of array circuitry, is induced by secondary photons in the substrate [34]. These are produced through radiative recombination during an avalanche. In-array circuitry has also been recognised as a contributor [22]. Therefore, while area within the array should be used to maximise the functionality to area cost ratio, high-speed circuitry, clock buffers and digital supplies should be placed outside the active areas. Cross talk can be reduced with increased SPAD spacing; however, this negatively impacts the fill factor.

8. As with all sensors, the temporal response departs from the ideal as characterized by the step and impulse response [11]. Upon the reception of a "zero" to "one" bit-level transition (a positive step assuming OOK), the ideal would be for the received transient to be limited by the channel. The step response however may cause a finite rise time and ripple that must settle before the output can be sampled. Likewise, upon the reception of a "zero" after a "one", the impulse response limits the speed at which the receiver settles to the new level. Upon either transition, the system must wait before sampling, and thus the symbol rate is fundamentally limited (discussed further later).

3. Circuit approaches and topologies

To date there are three distinct circuit topologies used within SPAD-based receiver designs. Two of the topologies use the innate full-logic-swing of the SPAD to produce an all-digital approach with either a sampled integrating or continuous time output. The third topology again aims for a continuous time output, but does so using the SPAD to switch current steering circuits. Each topology has advantages and disadvantages; however, the largest impactor is the level of integration within a CMOS technology as a function of the receiver functions discussed in Section 1.1. Along with these topologies, pure analogue, parallel connections have been formed into so called multi-pixel photon counters (MPPCs), these differ from SPAD arrays as the diode is not used to create a full logic level voltage. A positive feasibility study of MPPC use for communications can be found in [35].

3.1. Digital count summation: digital synchronous discrete time

One of the first all-digital multi-SPAD arrays custom designed for optical communications [12], aimed to replicate a synchronous integration mode-receiver [1]. A receiver of this type is presented in Section 3.2 as an example SPAD-based receiver. Integration prior to signal

sampling allows: (i) the output signal amplitude to be increased, i.e. allowing a doubled integration time, will approximately double the output amplitude, and (ii) noise which has a period shorter than the integration window is reduced. As SPADs produce a discrete time pulse, integrating signals is equivalent to pulse counting [11].

The topology is shown in **Figure 6A**, where the pulses from a SPAD are counted digitally using a ripple or synchronous counter. An integration window (**Figure 6B**) shorter than the symbol period can be used (i) to allow rejection of photon counts caused by the rise and fall periods of the input, (ii) to reject DCR events outside of the window, and (iii) to stop, sample and reset the digital counter. The addition of multiple high-speed synchronous data streams from multiple 'pixels', (formed from a SPAD and a counter), can be performed easily and with low latency. A multi-diode approach therefore allows multiple detection events during periods equivalent to the SPAD dead-time. A SPAD array output, in the form of a wide bit-depth synchronous digital stream can then either be (a) output to allow array diagnostics and signalling formats such as pulse amplitude modulation (M-PAM) or (b) thresholded on-chip allowing recovery of single-bit serial data streams. As the readout clock can be adjusted in phase relative to the transmitter clock, the position of the integration window can be chosen to minimize the BER, and its width can be optimized to maximize received signal amplitude.

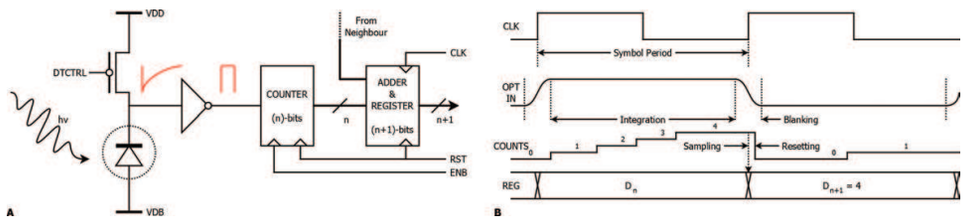

Figure 6. (A) The circuit topology with a binary counter per SPAD with synchronous binary addition of multiple SPAD counts within a defined integration window. (B) Timing diagram showing an input optical symbol, with an integration period blanking off the low-high and high-low transitions.

3.2. XOR-tree summation: digital asynchronous discrete time

To decrease the required logic, negated AND and negated OR (NAND/NOR) trees were investigated for combining multiple SPADs in small, higher fill factor mini- silicon photomultipliers (SiPMs) ([11] and references therein). However, to operate at high counting rates, the SPAD pulse (10–30 ns) must be reduced to <1 ns using per-SPAD pulse shortening. This is still smaller than per-SPAD digital counters, however it prompted the investigation of exclusive-OR (XOR) trees for mass digital pulse summation of larger SPAD arrays [36, 37] (**Figure 7A**). The XOR truth table lends itself readily to a summation task as sequential leading edges of SPAD pulses (without pulse shortening), simply toggle the output (**Figure 7B**). The limitation of this is the rate at which the final XOR within a tree can toggle its output.

As XOR gate standard cells occupy a smaller area than the digital logic required per SPAD in both digital summation and pulse-shortened NAND/NOR summation, the fill factor of the circuit can be increased significantly, in the case of [36], a fill factor of 43% was achieved. Indeed, such a receiver has demonstrated 4-PAM VLC, with highly discernible Poisson-limited levels,

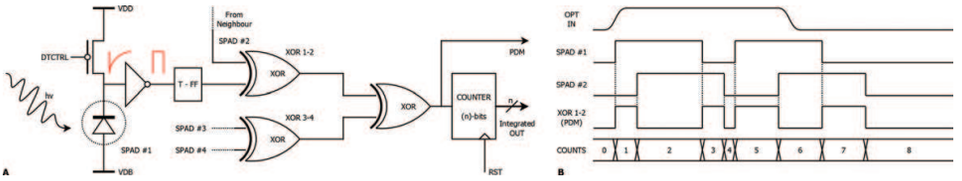

Figure 7. (A) The XOR summation of pulses, without pulse shortening, of four SPADs. The toggling of the output is a form of pulse density modulation; however, a binary counter can be used for pulse integration during a symbol. (B) Timing diagram of the XOR circuit for events on SPADs 1 and 2.

albeit with a low data rate [38]. As noted by the authors, this topology can be combined with (a) the digital summation approach, (b) on-chip modulation decoding logic, (c) signal processing logic, and (d) the adaption of the XOR tree and number of SPADs per channel. This would allow receiver optimization for power usage and optical sensitivity [38].

3.3. Current steering summation: analogue continuous time

In **Figure 8**, the current summation topology is shown [39], this uses the SPAD to switch currents onto output nodes shared by multiple SPADs, IQ− and IQ+. When the SPAD toggles, the current path changes due to the opening or closing of transistors MS1 and MS2. Transistor MCS could be implemented outside of the array, and transistors MS1 and MS2 can be far smaller than large digital circuits such as synchronous carry look-ahead counters. The authors therefore hoped for a high fill factor, however the use of PMOS devices limited fill factor through the SPAD N-well spacing rules. A fill factor of 3% was achieved however this was not optimized. A deeper issue is that the topology includes an inherent signal attenuation, whereby a full-logic swing on the SPAD (V_{EX} = 1.6 V) and inverters results in a maximum per-SPAD steered current of 100 uA. With a value of 300 Ω for RO+ and RO− this yields 30 mV per SPAD. Larger trans-impedance values could be used (e.g. 1 kΩ, yielding 0.1 V per SPAD), however trans-impedance bandwidth decreases as this resistance increases, i.e. $BW_{-3dB} = (2\pi R_O C_D)^{-1}$ [1]. It seems a shame to insert (i) a signal attenuation, and (ii) the transform {SPAD avalanche current to SPAD voltage to readout current to readout voltage}, into the topology to read out the SPAD array and communication signals.

While summation using Kirchhoff's current law for many SPADs is more efficient in terms of area and power than a multi-bit digital summation tree, and thus a lower energy per bit,

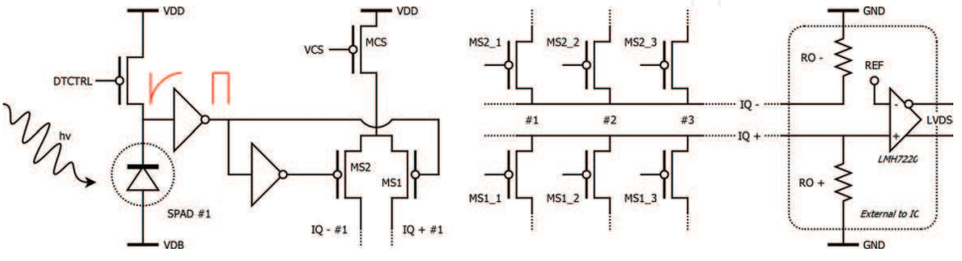

Figure 8. A current summation topology where a SPAD steers current through transistors MS1 or MS2. Multiple outputs can be added using Kirchhoff's current law on nodes IQ− and IQ+. Off-chip trans-impedance resistances RO− and RO+ are used to convert the current to a continuous-time voltage.

this readout structure introduces noise through transistor mismatch and the Johnson noise of the trans-impedance resistances. As this noise acts upon the voltage levels of "zeros" and "ones" at the analogue LMH7220 comparator, the steered current per SPAD, and thus the per-SPAD voltage has a fundamental limit, especially if the receiver is intended for M-PAM modulation schemes. This is an additive readout noise in combination with both Poisson photon arrival and noise induced by dead-time, DCR and after-pulsing effects. Further, to compare the energy per bit and the receiver area, modelling would need to include the power dissipated and area of (i) both RO resistors, (ii) the comparator, (iii) the field programable gate array (FPGA) serial-deserial (SERDES) input pads and (iv) the latching flip-flop. This would be the case as this yields the same synchronous full-logic level digital bit as the previous topologies, albeit single- rather than multi-bit. While the output is asynchronous with samples taken when the data eye is maximally open, with no blanking periods SPADs breakdown due to photons within the transitions of the optical source, thus preventing them from contributing once the optical signal has reached its correct level. This would be critical for reducing the standard deviations, σ_1, σ_2 and σ_3, upon M-PAM levels with means N_1, N_2 and N_3. Modulated OOK data streams have been received with this topology with 125 Mb/s at a BER of $1 \times 10^{-3.5}$ achievable with a dead time of 8.8 ns and before decision feedback equalization [39, 40]. We must note however, that to be fair all receivers should be compared without equalisation, although equalisation represents a logical subsequent signal processing step for low BER communications.

3.4. An exemplar receiver IC

To demonstrate the concepts and to investigate issues inherent in the design of such receivers, the digital summation topology has been investigated using a 32 × 32 array of 8 μm SPADs in 130 nm imaging CMOS (**Figure 9**) [11, 12]. This does use a hybrid approach with NAND/NOR signal addition prior to the main digital parallel summation tree, but does so mildly using 'super-Pixels' formed from four SPADs (**Figure 10**).

The design, discussed in detail within [11, 12], operates as an integration-mode receiver, with two blanking periods of 500 ps at the start and end of a symbol. These screen out SPAD events during the transitions of an optical source, and thus decreasing the variability of counts within an integration period. The digital summation tree allows clocking at 100 MHz, culminating in either a single channel 14-bit output or a 9-bit 16 channel output operating at 25 MHz per channel. As the receiver was intended as a test-harness for various concepts, the array treated the fill-factor and PDE as fixed offsets that could be improved in subsequent generations. Rather than designing to maximize known performance metrics, the array was designed to test transient issues along with unknown factors important for both communications and high-sensitivity, high sample-rate metrology.

By passing full logic swing synchronous parallel data, rather than (a) fast toggling pulse density modulated data, (b) asynchronous ripple-counter data or (c) currents requiring a trans-impedance, the receiver requires no circuitry prior to its test data being captured by an FPGA FIFO, or being thresholded. The receiver, summarized in **Table 1** (see [11] for results/modelling), represents a test harness with dedicated modes for array characterization.

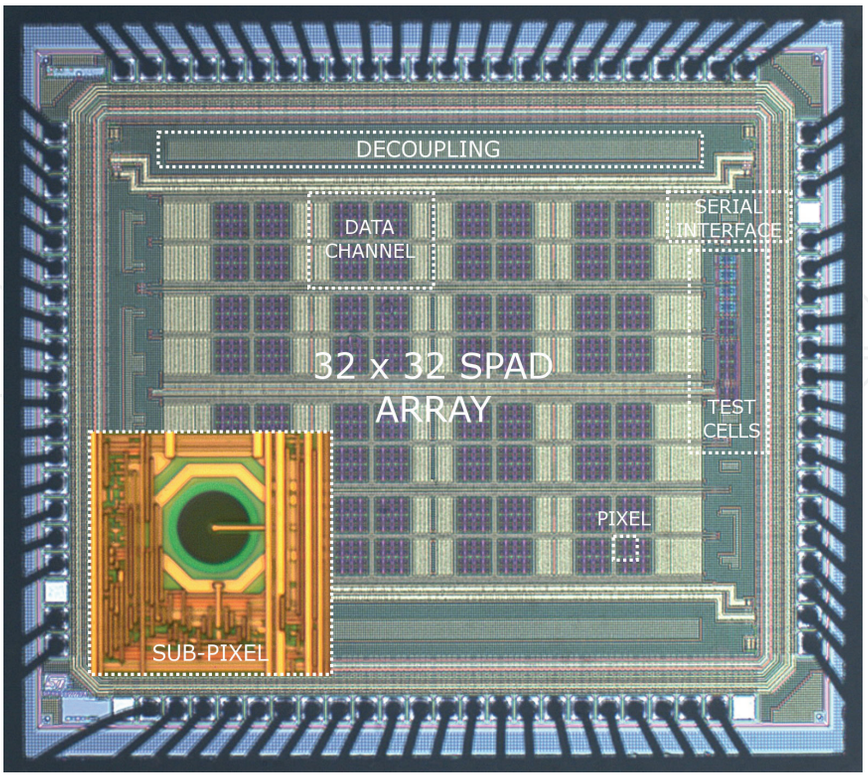

Figure 9. Micrograph of the 1024 SPAD digital summation integration-mode receiver with the SPAD inset. This has a fill factor of 2.4%, and implements up to 16 'Data Channels' each of 64 SPADs.

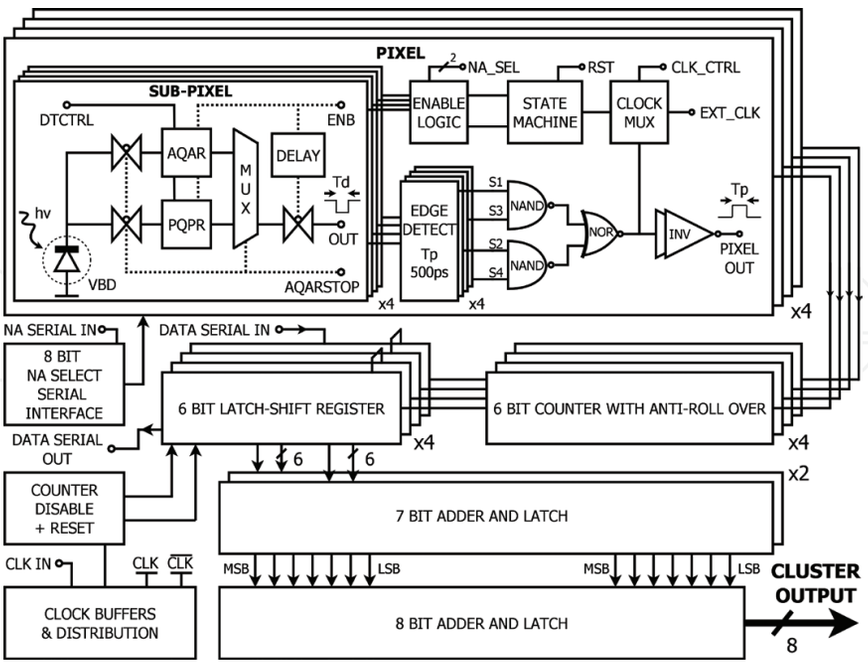

Figure 10. Block diagram of the architecture for 16 SPADs, showing per-SPAD active and passive quenching, dead time control via the analogue voltage DTCTRL, and pulse shortening. A binary counter (6 bit) is used per group of four SPADs and NAND/NOR signal addition is used.

Parameter	Value	Units	Notes
CMOS process	130 nm Imaging	—	90 nm metallization
Peak PDE	20	%	VEX = 1.2 V at 450 nm
SPAD VBD	−13.05	V	—
Dead Time	5–250	ns	Active or passive
Pulse shortening	500	ps	For NAND/NOR Sum
SPAD VEX	1.35	V	Typically, 1.25 V
Fill factor	2.42	%	Exc. I/O Pads
# SPADs	1024	—	Circular
SPAD diameter	8	μm	Active area
Quenching	AQAR or PQPR	—	Adjustable
Digital VDD	1.2 and 3.3	V	Core/IO
Pixel counter bits	6	Bit	Per 4 SPADs
Output width	14 or 9	Bit	Single- or Multi-Channel
# Channels	1 or 16	—	Configurable
Die size	2.4 × 2.1	mm	~5 mm^2
Clock frequency	108 100	MHz	Pad-Limited (Max) (Typical)
Median DCR	2.5	kHz	16°C, V_{EX} = 1.35
Max count rate	65	Gphoton/s	~10 ns Dead time
After-pulsing	0.9	%	1 μs window
Energy per bit	370	pJ/Hz	Estimated
Sensitivity estimate	−31.7	dBm	100 Mb/s BER = 1×10^{-9} 100% extinction ratio
Sensitivity estimate calibrated	−28.5	dBm	Inc. LED extinction ratio reduction

Table 1. Performance details for the 32 × 32 digital summation integration-mode receiver.

4. Results

This section will present experimental results from the example digital summation integration-mode receiver. The sensitivity curve or photon transfer curve will be shown and a model for this will be presented. The step response will also be shown as it makes a direct

impact on communications performance. This will also be modelled illuminating the choices a designer has with respect to the SPAD dead-time, number of SPADs, digital-counter read-out rate and maximum modulated input power. Finally, experimental communications results will be shown, demonstrating both feasibility and several issues we have discussed.

4.1. Photon transfer curve and optical sensitivity

In **Figure 11**, the photon transfer curve (PTC) is shown. This plots the detected photon events per second against the input optical power. The N_a modes describe a time multiplexing approach, which while beneficial for the temporal response, will not be discussed here (see [11]). The N_a = 4 mode is of interest as it turns on all SPADs within the summation. At the centre of the curve, the receiver has a linear response, however it saturates at both the low and high levels due to the combined DCR and the active or passive quenching maximum count rate, C_{MAX}, respectively. Passive quenching shows both the presence of the 1/*e* factor in its maximum counting rate, and the issue of paralysis whereby the count rate starts to decrease at high optical powers. For communications, the receiver must operate within the linear region, as nonlinearities act to modify the extinction ratio of the detected signal, and thus negatively impacts the BER.

For design of SPAD-based receivers, active quenching appears to be advantageous, however as it requires extra circuitry per-SPAD, the choice of active or passive quenching must be taken in view of top level communications performance metrics. By reducing the dead-time, a higher count rate is achievable, while an increased number of SPADs has the same effect but comes with increased total DCR. It is advantageous to reduce the DCR through both innovation and by turning high-DCR SPADs off (as shown on **Figure 11**) as this extends the lower optical power limit and pushes the receiver closer to being quantum limited.

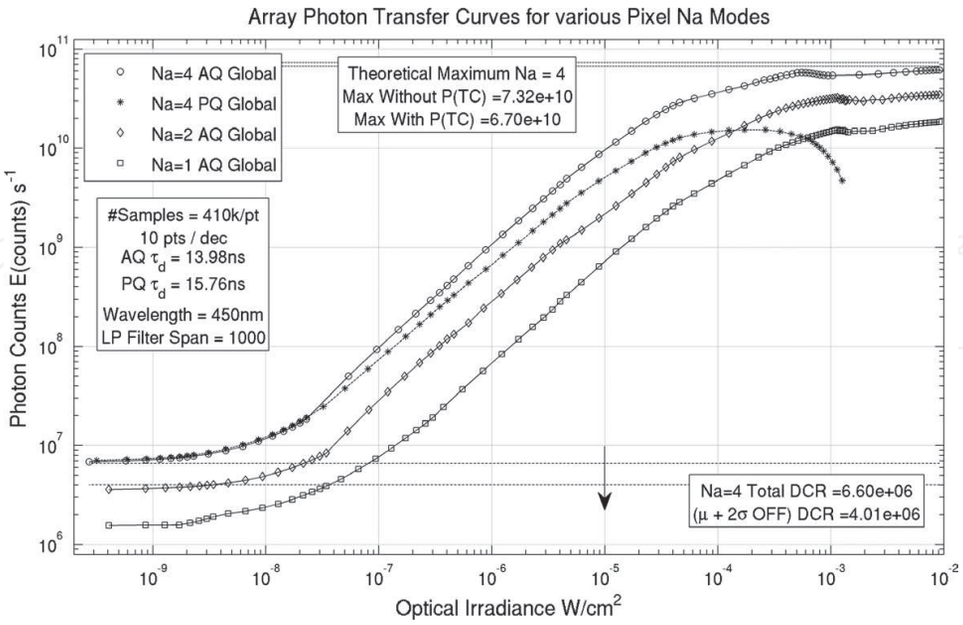

Figure 11. The photon transfer curve of photon counts per second against optical input power. The curve is linear within the centre; however, saturation occurs due to the combined DCR of all SPADs within the summation and the active or passive quenching maximum count rate, C_{MAX}.

The dynamic range can be calculated as: (i) the ratio of maximum (i.e. C_{MAX}) to minimum (i.e. $\sum(DCR)$ observable signals (approx. 80 dB in Figure), or (ii) the ratio of the maximum to the noise of the dark count floor rather than its absolute value, i . e. $\sqrt{\sum(DCR)}$ (approx. 149 dB). If we accept that between two modulated amplitude levels, a suitable gap must be left in order to achieve the required BER, the size of the linear region gives an estimate as to the theoretical number of M-PAM levels that could be implemented. With the required offsets, due to the finite fill factor and photon detection efficiency known through measurement of the PTC, other power penalties can be assessed and ranked as a guide to future design.

4.1.1. A competitive count model of SPAD avalanche initiations

The PTC can be modelled using competitive interaction, matching the measured experimental data (**Figure 12**). The arrival of a photon, (assuming AQAR), predicates loss of a photon, dark count or after-pulse if one occurs during the dead time. This creates competition at high count rates, in which (i) photons prevent dark counts or after-pulses, (ii) dark counts prevent photon counts or after-pulses, and (iii) after-pulses prevent photon counts or dark counts.

A simple model, describing the number of output counts as a function of input photons is given in Eq. (17). Where FF is the fill factor, DCR is the dark count rate, ΔT represents the time that counts are collected and $P(ap)$ is the after-pulsing probability. $PDE(\lambda)$ is the photon detection efficiency at a particular wavelength and $\psi(\lambda)$ is the number of photons, of wavelength λ, incident per second. This does not account for the dead-time, although it could be combined with hybrid paralysable models of dead-time limited counting [28, 29]. This would modify the output counts, N_{counts}, but would not include the concept of competition due to events other than photon arrivals. The model assumes that photons, dark counts and after-pulses can be separated, which while achievable with careful experimentation, cannot be performed during normal operation. The question therefore arises of what is the DCR level when the receiver is being illuminated, knowing that some DCR events will not be detected as the SPAD has already broken down? As noted within [23], a photon arrival and an after-pulse, or a dark count and an after-pulse are not statistically independent.

$$N_{counts} = \left[1 + P(ap)\right]\left(PDE(\lambda)\psi(\lambda)FF + DCR\right)\Delta T \tag{17}$$

As high counting rates prevent the separability, independence and static contribution assumptions being used, the concept of an enable time, T_{enb}, as a proportion of ΔT, can be introduced, i.e. an avalanche induced dead time will remove a period, T_d, from ΔT. The time in which a SPAD is in its receptive state, is then given by Eq. (18), and is the first step towards a count competitive model. The values ψ_{act} and DCR_{act} are the actual numbers of photon and dark detections, however these are intermediate output values in Eq. (18) as they must be corrected for count competition [11].

$$T_{enb} = \Delta T - \left[\left(1 + P(ap)\right)\left(\psi_{act} + DCR_{act}\right)\left(T_d\right)\right] \tag{18}$$

A regression can be made backwards to the input values for photons, ψ_0, dark counts, DCR_0, and after-pulses, AP_0, that can be calculated from the incident optical power, fill factor or

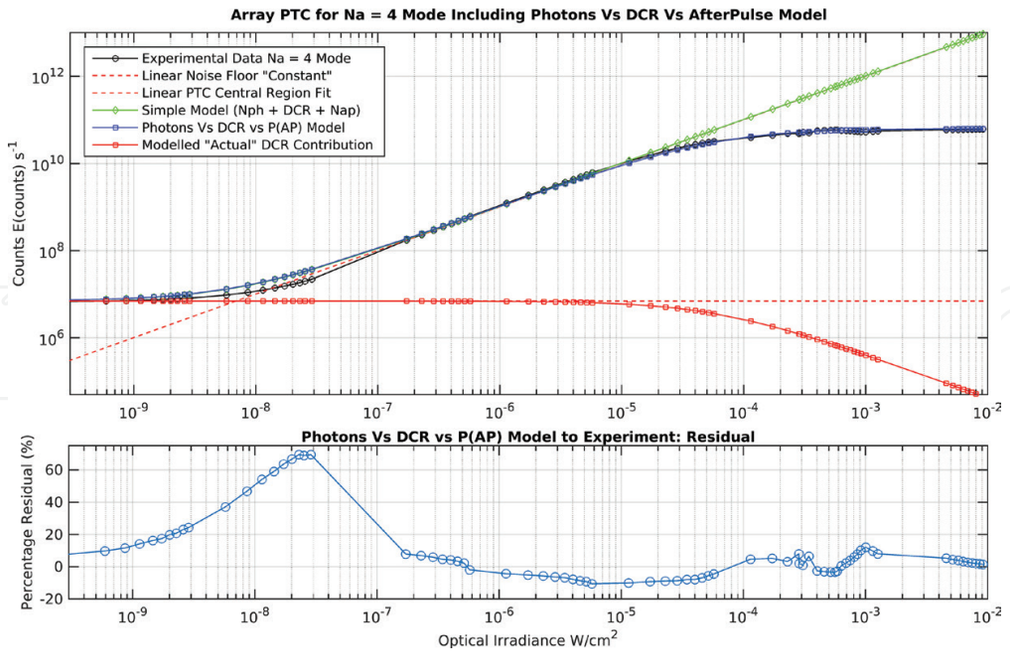

Figure 12. Count competitive and simple addition models for the photon transfer curve. The fitting parameters are: λ = 450 nm, PDE = 23%, FF = 2.4%, $P(ap)$ = 1%, T_d = 16 ns, DCR = array average DCR. Two linear fits are included to both the central region of the experimental PTC and the constant noise floor. The competition implies a reduced DCR as the PTC saturates, this is also included.

carrier generation rate etc. The final counts N_{counts}, including competitive interaction and the finite dead time, is then given by Eq. (19).

$$N_{counts} = \left(\frac{\Delta T}{1+(1+P(ap))(\psi_0+DCR_0)(T_d)}\right)(\psi_0+DCR_0) + \left(\frac{\Delta T}{1+(1+P(ap))(\psi_0+DCR_0)(T_d)}\right)^2(\psi_0+DCR_0)P(ap) \quad (19)$$

In **Figure 12**, this model shows a low overall residual error, while incorporating the saturation at the level of C_{MAX}. The deviation from experimental values in the region of 1×10^{-8} W/cm^2 is due to small inconsistencies in the measurement approach as the optical power approached the limits of the optical power meter. The power of the model however, is that the instantaneous DCR or after-pulsing counts can be estimated given count competition rather than assuming their calculated, or stand-alone measurement values. For example, DCR_0 may be measured at a mean of 100 cps, however through the introduction of some photon detections and the corresponding loss in T_{enb}, the actual detected dark count rate, DCR_{act} may instead be 90 cps, i.e. 10 thermal generation events were ignored as they happen to coincide with SPAD dead-times caused by photon events. The value of DCR_{act} is therefore plotted showing that as a SPAD approaches count saturation, the number of detected dark events falls. This effect must therefore be included in any assessment of count noise at a particular optical power, as would be needed for the BER of OOK or M-PAM of finite extinction ratio.

4.2. Step and impulse response

The step and impulse responses play a key role in the effectiveness of communication. In [11, 12], the OOK communications experiments are severely impacted by both (i) a step

http://dx.doi.org/10.5772/intechopen.68935

response initial peak, ripple and lower than expected settled count rate and (ii) an impulse response slow decay. Both act to close the data eye, thereby increasing the BER. For the step response, the inter-symbol interference effect is directly linked to the dead-time of SPADs as a function of the array read-out-rate, and has also been observed experimentally in [39, 40].

Figure 13 shows a large change in the peak and ripple as the input power increases [11]. For the 5 μW step, the peak is 29% above the steady state. This will force a separate peak in histograms of the data amplitude levels depending on the data sequence. The following data point shows a value 10% below the steady state, indicating that a received 'one' at this point would have an artificially lowered amplitude. In comparison, lower optical powers such as the 0.5 μW step, give peaks only 10% above the steady state, followed by a minimal trough. When combined with the observation that the lower light level responses have a steady state closer to the initial, (and correct incident), number of photons, there will be a corresponding reduction in the splitting of the N_1 data distribution and an improved BER (explored further in [11]). As the steady state is below the initial number of counts, the relative decrease in steady state as optical power increases acts to decrease the received extinction ratio of the modulated OOK or M-PAM signal, thereby increasing errors.

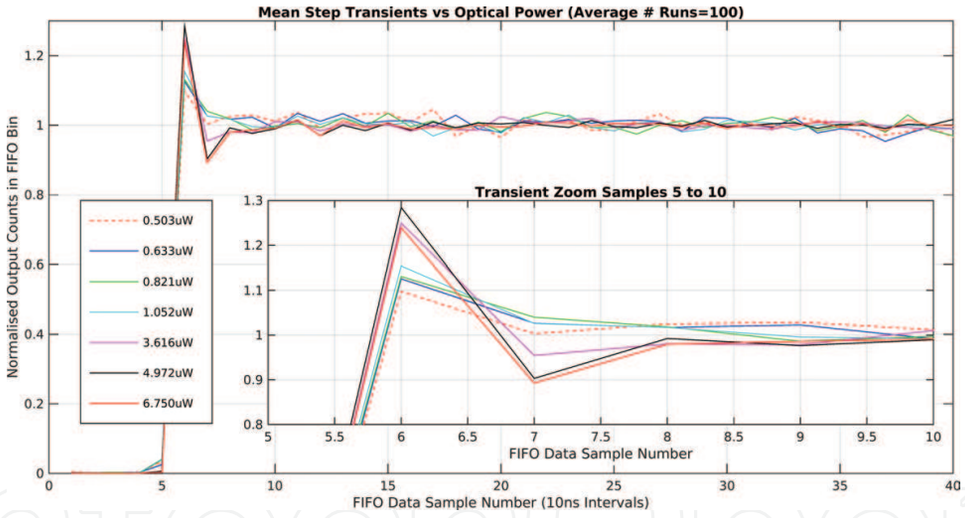

Figure 13. Normalised average step response over a sweep of optical power. The peak and ripple in the number of counts per 10 ns bin, is highlighted in the inset zoom (50 ns) after the start of the laser step.

4.2.1. A detector pooling model of SPAD array step response:

The SPAD-array step response can be modelled allowing optimisation of future designs. Initially, the receiver has a full complement of SPADs (e.g. N_{array} = 1024). At some subsequent time, a number of diodes have broken down, giving a number at the output. These have been removed from the pool, and hence the effective total efficiency of the array, (i.e. the PDE and array fill-factor for an array of N_{array} SPADs), η_{det}, for subsequent photon arrivals reduces as there is a new number of available SPADs, N_{avail}. The removal of SPADs from the pool, and the reduction in detection probability, leads to decreased output counts

despite continuous input levels. Sometime later, the SPADs that were removed will be recharged back into the available pool, bringing the effective detection efficiency back towards unity.

Eq. (20) is an initial recursive discrete time model for the number of SPADs available at a time step, n, where $C_{in}(n)$ is the expected number of input counts (photons, dark counts and afterpulses). The model makes several assumptions, however fitting to experimental results shows that: (i) the assumptions hold for the 32 × 32 digital summation integration mode receiver and (ii) these can be tackled in subsequent revisions of the model. The model assumes that the SPADs break down at the start of a discrete time step, and that the array readout frequency giving rise to these discrete time steps is initially equal to the SPAD dead time. This latter assumption is expanded within the modelling below to include dead times integer multiples of the readout period, i.e. $2n$, $3n$ etc. It should be noted that the quantity, $PDE_0FF_0C_{in}(n)$ models the number of SPAD detections received by the SPAD array within the discrete period and thus can be combined with the count competitive model (Eq. (19)).

$$N_{avail}(n) = N_{avail}(n-1) - \frac{PDE_0 FF_0 N_{avail}(n-1)\, C_{in}(n-1)}{N_{array}} + \frac{PDE_0 FF_0 N_{avail}(n-2)\, C_{in}(n-2)}{N_{array}} \tag{20}$$

The steady state and the reduction of the extinction ratio due to this effect, can be estimated by Eq. (21), (the 'alpha model'), or a simple model (Eq. (22)) using the initial input counts (see [11]). The 'alpha model' uses the long-term average input counts, $\langle C_{in} \rangle$, expected for a light level, i.e. N_{counts} from Eq. (19), along with a fitting factor, α, which has been found through minimisation of the residual to be 0.725 i.e. approximately $1/\sqrt{2}$.

$$C_{out}(n) = \langle C_{in} \rangle \left[1 - \frac{\langle C_{in} \rangle}{N_{array}} \left(1 + \alpha \frac{\langle C_{in} \rangle}{N_{array}} \right) \right] \tag{21}$$

$$C_{out}(n) = \langle C_{in} \rangle \left[1 - \frac{\langle C_{in} \rangle}{N_{array}} \right] \tag{22}$$

The discrete time model (Eq. (20)) is (i) uncalibrated, (ii) departs from noisy experimental results and (iii) is not yet combined with the steady state model of Eq. (21). Despite this it can be used as an indication as to future modelling and device requirements. The critical point of both time and steady state models is that they allow preliminary investigation into the splitting of the OOK N_1 distribution that is seen clearly in experimental communication results (next section). In **Figure 14**, an array of 1024 SPADs receives a step change in optical power. The input is fixed at an example 100 photons per symbol as ultimately the BER must remain static (Eq. (5)) [1, 6]. With the dead time fixed at 10 ns, the discrete time 'readout' period is changed as a proxy for different data rates. As the data symbol or 'readout' period becomes shorter than the dead time, there is a severe increase in the ripple, the number of symbols prior to SPADs recharging into the detector pool and a decrease in the steady state. While readout periods of 20 ns for the 10 ns dead time produce minimal decreases in the effective detection efficiency and step-response ripple, we must remember that there may be a limit to how short the dead time can be. As designs push the data rate in the first instance, a dead-time equal to the readout or symbol period is preferable.

http://dx.doi.org/10.5772/intechopen.68935

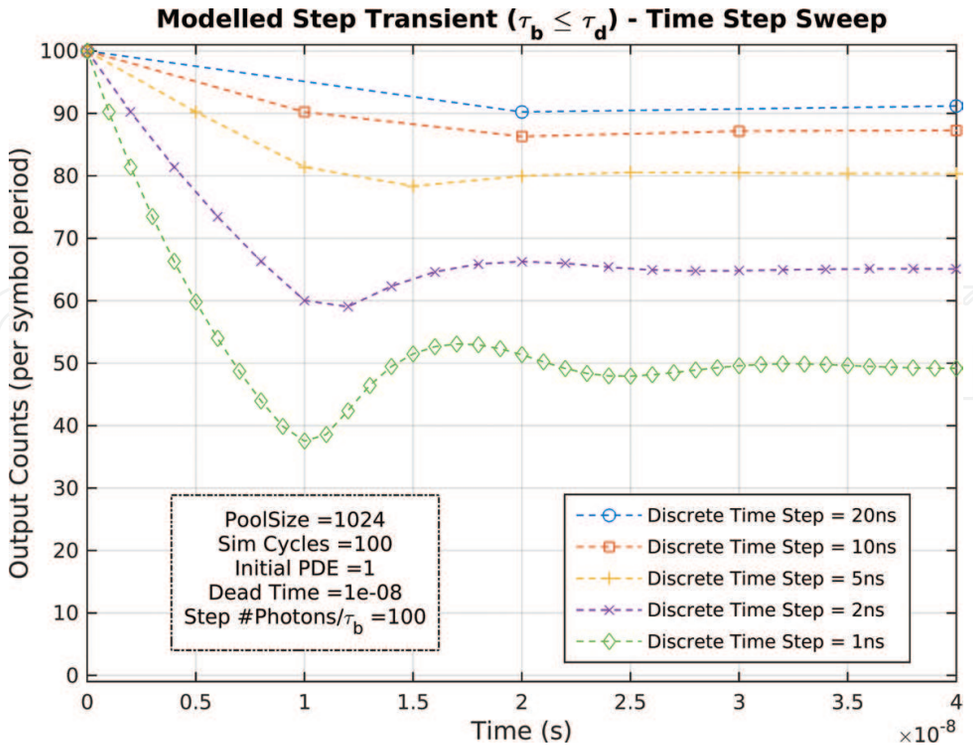

Figure 14. Modelled (uncalibrated) step response with an input step of 100 photons per symbol, a pool of 1024 SPADs and a SPAD dead time of 10 ns. The discrete time step or array readout speed is swept as a proxy for data rates between 50 Mb/s and 1 Gb/s.

In **Figure 15**, a readout period of 1 ns is chosen to highlight observations within the model. The step is held at 100 photons per symbol and the dead time is again set to 10 ns, however here the size of the array is changed from 2048 to 256 SPADs. It is possible to decrease the severity of the step response by increasing the number of diodes. However, clearly it is the ratio of the expected counts per symbol to the pool size that is critical. Reduce the number of diodes too much and the step response ripple becomes the limiting factor for both communications (BER) and optical metrology (for 100 photons per sample, the steady state of the 256 SPAD array is approx. 20 counts per sample). For completeness, the model, (once calibrated), can be used to estimate the number of diodes necessary, at a specified dead time (10 ns) and symbol period (1 ns), for a step response ripple and steady state that is within the Poisson noise of the expected input count level. For 100 counts per symbol, the standard deviation, ($\sqrt{100} = 10$), leads to ~9750 SPADs required for a steady state of 90.

These parameter sweeps, although chosen to visually highlight SPAD array issues, point to three heuristic design rules:

- First, that the dead-time should be equal to the data rate, (10 ns for 100 Mb/s etc.),
- Second, that the number of SPADs should be high in comparison to the expected photon counts per data symbol, and
- Third, if T_d is limited, the number of diodes can be increased to reduce the detrimental effects on the step response if we wish to operate at a data rate higher than $1/T_d$.

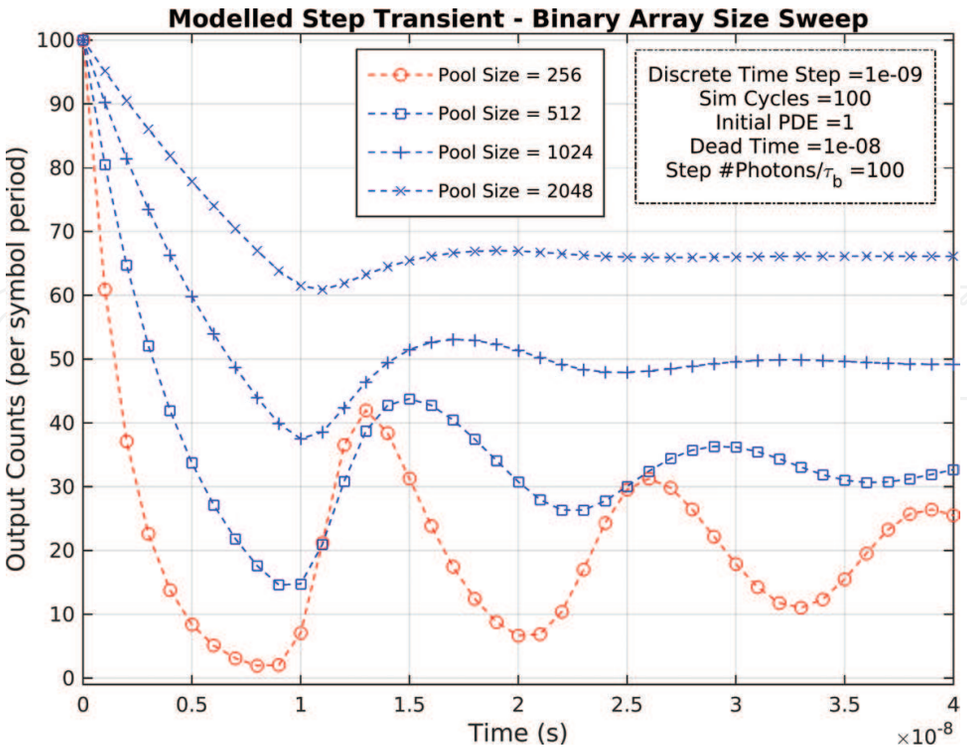

Figure 15. Modelled (uncalibrated) step response with an input step of 100 photons per symbol, a SPAD dead time of 10 ns and an array readout period of 1 ns (for illustration purposes). The number of SPADs in the pool is swept highlighting that if a dead time equal to the symbol period is not possible, the number of SPADs must be large in comparison to the expected photon arrivals per symbol.

Of course, these rules are a product of only the step response model. Increasing the number of diodes within the array may have a detrimental effect on the array fill factor. Hence research must concentrate on models that once combined, can be used for overall receiver optimisation with respect to communication level metrics (sensitivity, BER, data rate, etc.).

4.3. Communication results and feasibility

Early communications results have been achieved using the 32 × 32 SPAD-based integrating receiver, demonstrating that such receivers are feasible. **Figure 16** shows the eye diagram of a 50 Mb/s on-off key non-return-to-zero data stream. The transmitter in this case is a Gallium-Nitride 34 μm diameter 450 nm micro-LED, with a bandwidth of 220 MHz. The data eye is open at this optical power (20 μW/cm^2) and data rate, however several issues become apparent.

- Firstly, upon a 'zero' to 'one' transition, there is a large peak 75–100 counts higher than the main N_1 level (approx. 250). This can be directly attributed to the step response discussed above, and can be seen clearly near sample 110 in the time traces of **Figure 17**. Indeed, it leads to a split N_1 distribution, with the severity of the splitting of the ideally two distributions (N_0 and N_1) being evident in the count histogram.
- Secondly, the impulse response, (measured in [11]), creates a slow 'one' to 'zero' transition. This is clearly visible near time sample 90 (**Figure 17**). For some high frequency transitions, the impulse response prevents the transmitted 'zero' from obtaining the low count average

(approx. 45) as the impulse response is still decaying when the next 'one' bit is transmitted (see near sample 150). This therefore leads to the splitting of the N_0 distribution.

While excellent agreement is shown between the transmitted and received pseudo random bit streams (**Figure 17**), the closure of the eye can be attributed to (a) the steady state of the step response, (b) the bleeding of previous 'one' bits into subsequent 'zero' bits due to the impulse response, and (c) the finite extinction ratio of the micro-LED transmitter. The 32 × 32 SPAD receiver, as a testing harness, has provided crucial details for the future development of SPAD-based receivers in that clearly, the transient response which

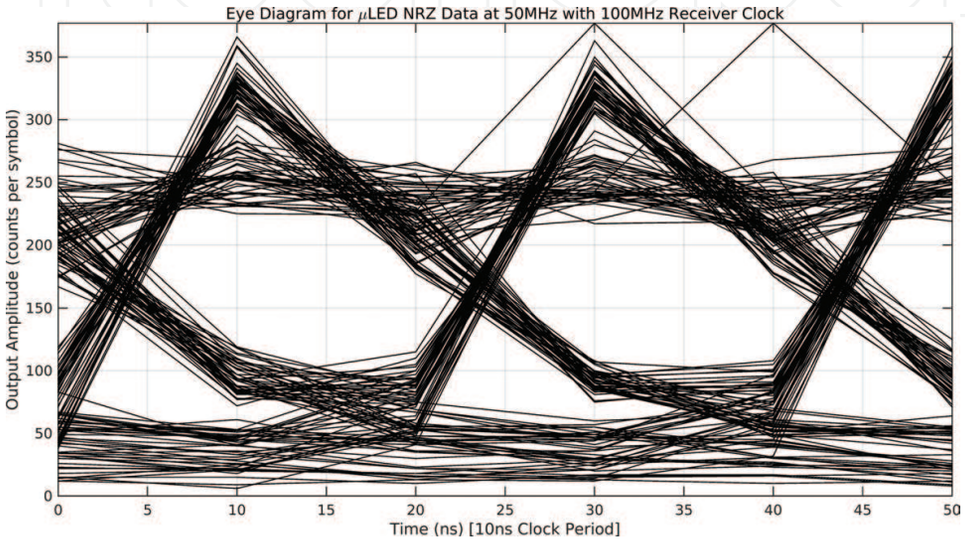

Figure 16. A measured eye diagram for 50 Mb/s OOK-NRZ data. The dead time was approx. 13.5 ns.

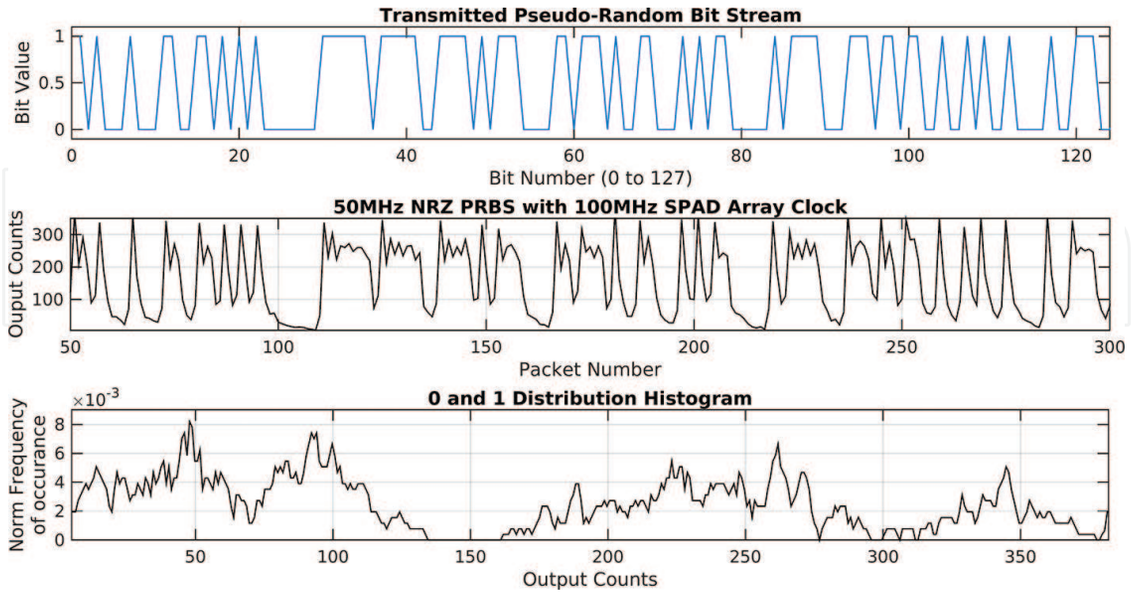

Figure 17. The transmitted OOK NRZ data stream, a time trace of the SPAD-receiver output and the histogram of received amplitudes. A sampling threshold at 150 counts shows separability of the N_0 and N_1 distributions, however the impulse and step responses have acted to split the distributions and close the data eye. With a reduced impulse and step response, the transmitter should render Poisson-limited distributions near 50 and 350 counts.

was only evident once empirically measured, must be addressed for low bit error rate communications.

5. Conclusions

SPAD-based optical receivers increase the prospective methodologies for applications requiring extremely high optical sensitivity. This chapter has covered the specifics of both optical detection using the internal photoelectric effect, and how avalanche multiplication can be extended to run-away avalanche producing full logic-level voltage pulses with single-photon sensitivity. The application, visible light communications, highlights that such receivers require further study before all noise sources and all circuit topologies are understood fully with respect to the higher-level communication performance metrics.

Three SPAD receiver topologies were introduced, the digital summation, the XOR summation and a current-summation. The performance metrics of both (i) optical communications (data rate, bit error rate, energy per bit etc.) and (ii) single SPAD and SPAD array circuits (fill factor, PDE, DCR), were also discussed. At present, there is little receiver theory to link low-level SPAD metrics with higher-level performance, although some work as already been done. For example, both the SPAD array fill factor and photon detection efficiency are known to directly impact optical sensitivity, and appropriate modelling has been developed. The receiver BER has also been investigated with respect to the SPAD dead time and the output photon-countingstatistics of a digital summation SPAD-receiver. To date however, there is little to link the SPAD after-pulsing probability with the BER, although one would assume its inherent delay would act as a form of ISI.

One requirement for future research and commercialization, would be the use of application specific design methodologies fully linked to higher level communications performance. How should the designer balance the SPAD fill factor and dead-time for example, when attempting to optimize a receiver for a required data rate and sensitivity? At present, too few high to low performance relations exist to allow such an interplay to be optimized. Likewise, if the application speed requires the dead-time to be traded off with after-pulsing, the effect that after-pulsing has needs to be quantified to allow comparison with dead-time ISI effects.

Overall, SPAD-based receivers are an active field of research, aiming to provide optimized designs with both extreme sensitivity and high-speed. Data-rates and energy per bit will likely remain modest and far removed the highest speed links of today. However, such receivers may offer options in (i) room-lighting VLC applications, supplementing existing Wi-Fi for download intensive tasks such as high-quality, real-time video streaming, and (ii) niche applications where optical attenuation is significant enough to warrant increased receiver gain and a higher energy per bit.

6. Student exercise questions

1. Calculate the quantum limit number of photons required per symbol, assuming an ideal, noiseless receiver, at a data rate of 50 Mb/s, a BER of 1×10^{-6}, and a wavelength of 600 nm.

2. Calculate the minimum number of SPADs within a receiver array, for the conditions in question #1, if all SPADs have a dead time of 10 ns.

3. If the SPADs are circular, and laid out in a square array, calculate the approximate array fill factor (10 μm SPAD diameter), assuming (i) no CMOS circuitry is incorporated within the array and (ii) that D_s and D_g are 2 μm and 5 μm respectively.

Author details

Edward M.D. Fisher

Address all correspondence to: e.fisher@ed.ac.uk

Institute of Digital Communications (IDCOM), The School of Engineering, The University of Edinburgh, Scotland, UK

References

[1] Razavi B. Design of Integrated Circuits for Optical Communications. 1st ed. McGraw-Hill, New York, USA; 2003.

[2] Binh LN. Digital Optical Communications. 1st ed. CRC Press, Boca Raton, Florida, USA; 2008.

[3] Hanzo L, Haas H, Imre S, O'Brien D, Rupp M, Gyongyosi Y. Wireless myths, realities, and futures: From 3G/4G to optical and quantum wireless. Proceedings of the IEEE. 2012;**100**:1853-1888 (Special Centennial Issue)

[4] Elgala H, Mesleh R, Haas H. Indoor broadcasting via white LEDs and OFDM. IEEE Transactions on Consumer Electronics. 2009;**55**(3):1127-1134.

[5] Cisco Systems Inc. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016-2021 White Paper [Internet]. 09/02/17 [Updated: 09/02/17]. Available from: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.pdf [Accessed: 24/02/17]

[6] Sackinger E. Broadband Circuits for Optical Fiber Communication. Wiley, New Jersey, USA; 2005.

[7] Carusone A, Yasotharan H, Kao T. Progress and trends in multi-Gbps optical receivers with CMOS integrated photodetectors. In: Custom Integrated Circuits Conference (CICC); IEEE, San Jose California, USA; 2010:1-8.

[8] Zeng L, O'Brien D, Minh H, Faulkner G, Lee K, Jung D, Oh Y, Won ET. High data rate multiple input multiple output (MIMO) optical wireless communications using white LED lighting. IEEE Journal on Selected Areas in Communications. 2009;**27**(9):1654-1662.

[9] O'Brien D, Le Minh H, Zeng L, Faulkner G, Lee K, Jung D, Oh Y, Won ET. Indoor visible light communications: challenges and prospects: free-space laser communications VIII. Proceedings of the SPIE. 2008;**7091**:709106-709109.

[10] IEEE Computer Society—WG802.15—Wireless Personal Area Network (WPAN) Working Group. P802.15.7 —IEEE Approved Draft Standard for Short Range Wireless Optical Communication Using Visible Light [Internet]. April 2011 [Updated: Rev. P802.15.7]. Available from: https://standards.ieee.org/develop/project/802.15.7.html

[11] Fisher E. A parallel, reconfigurable single-photon avalanche diode array for optical communications [Thesis]. The School of Engineering, The University of Edinburgh, Edinburgh, Scotland, UK; 2015.

[12] Fisher E, Underwood I, Henderson R. A reconfigurable single-photon-counting integrating receiver for optical communications. IEEE Journal of Solid-State Circuits. 2013;**48**(7): 1638-1650.

[13] Henley E, Garcia A. Sub-Atomic Physics. 3rd ed. World Scientific, Toh Tuck Link, Singapore; 2007.

[14] Durini D, editor. High Performance Silicon Imaging: Fundamentals and Applications of CMOS and CCD Sensors. 1st ed. Elsevier Ltd, Woodhead Publishing (for Elsevier), Cambridge, UK; 2014.

[15] Roychoudhuri C, Kracklauer AF, Creath K, editors. The Nature of Light: What is a Photon?. 1st ed. CRC Press, Boca Raton, Florida, USA; July 2008. 452 p.

[16] Sze SM. Semiconductor Devices: Physics and Technology. 2nd ed. Wiley and Sons, New Jersey, USA; 2001.

[17] Campbell J. Recent advances in telecommunications avalanche photodiodes. Journal of Lightwave Technology. 2007;**25**(1):109-121.

[18] Yang B, Schaub J, Csutak S, Campbell J. 10 Gbps all silicon APD optical receiver. Lasers and Electro-Optics Society. The 15th Annual Meeting of the IEEE. 2002;**2**:681-682.

[19] Takahashi S, Tajima A, Tomita A. High-efficiency single photon detector combined with an ultra-small APD module and a self-training discriminator for high-speed quantum cryptosystems. In: Proc. 13th Micro-Optics Conference, Japan. 2007

[20] Ghioni M, Gulinatti A, Rech I, Zappa F, Cova S. Progress in silicon single-photon avalanche diodes. IEEE Journal of Selected Topics in Quantum Electronics. 2007;**13**(2):852-862.

[21] Haitz RH. Mechanisms contributing to the noise pulse rate of avalanche diodes. Journal of Applied Physics. 1965;**36**(10):3123-3131.

[22] Webster E. Single-photon avalanche diode theory, simulation, and high performance CMOS integration [Thesis]. The School of Engineering: University of Edinburgh; 2013.

[23] Cova S, Ghioni M, Lacaita A, Samori C, Zappa F. Avalanche photodiodes and quenching circuits for single-photon detection. Applied Optics. 1996;35(12):1956-1976.

[24] Richardson J, Webster E, Grant L, Henderson R. Scaleable single-photon avalanche diode structures in nanometer CMOS technology. IEEE Transactions on Electron Devices. 2011;**58**(7):2028 -2035.

[25] Goetzberger A, McDonald B, Haitz R, Scarlett RM. Avalanche effects in silicon p-n junctions. II. Structurally perfect junctions. Journal of Applied Physics. 1963;**34**(6):1591-1600.

[26] Richardson J, Grant L, Henderson R. Low dark count single-photon avalanche diode structure compatible with standard nanometer scale CMOS technology. IEEE Photonics Technology Letters. 2009;**21**(14):1020 -1022.

[27] Webster E, Richardson J, Grant L, Henderson R. Single-photon avalanche diodes in 90 nm CMOS imaging technology with sub-1 Hz median dark count rate. Proceedings of the International Image Sensor Workshop. 2011:262-265.

[28] Lee S, Gardner R. A new G-M counter dead time model. Applied Radiation and Isotopes. 2000;**53**(4-5):731-737.

[29] Neri L, Tudisco S, Musumeci F, Scordino A, Fallica G, Mazzillo M, Zimbone M. Note: dead time causes and correction method for single photon avalanche diode devices. Review of Scientific Instruments. 2010;**81**(8): 086102-1 to 086102-3

[30] Blazej J, Prochazka I. Avalanche photodiode output pulse rise-time study. In: SPIE 7355, Photon Counting Applications, Quantum Optics, and Quantum Information Transfer and Processing II; May; SPIE Headquarters: Bellingham, Washington, USA; 2009. pp. 73550M-11.

[31] Spinelli A, Lacaita A. Physics and numerical simulation of single-photon avalanche diodes. IEEE Transactions on Electron Devices. 1997;**44**(11):1931-1943.

[32] Vilella E, Alonso O, Montiel A, Vila A, Dieguez A. A low-noise time-gated single-photon detector in a HV-CMOS technology for triggered imaging. Sensors and Actuators A: Physical. 2013;**201**(1):342-351.

[33] C. Niclass, M. Soga, H. Matsubara and S. Kato, "A 100m-range 10-frame/s 340×96-pixel time-of-flight depth sensor in 0.18μm CMOS," 2011 Proceedings of the ESSCIRC (ESSCIRC), Helsinki, 2011, pp. 107-110. doi: 10.1109/ESSCIRC.2011.6044926

[34] Rech I, Ingargiola A, Spinelli R, Labanca I, Marangoni S, Ghioni M, Cova S. Optical crosstalk in single photon avalanche diode arrays: a new complete model. Optics Express. 2008;**16**(12):8381-8394.

[35] Zhang G, Yu C, Zhu C, Liu L. Feasibility study of multi-pixel photon counter serving as the detector in digital optical communications. Elsevier Optik. 2013;**124**(22):5781-5786. DOI: http://dx.doi.org/10.1016/j.ijleo.2013.04.060

[36] Dutton N, Gnecchi S, Parmesan L, Holmes A, Rae B, Grant L, Henderson R. A Time-correlated single-photon-counting sensor with 14GS/s histogramming time-to-digital converter. In: Technical Digest International Solid-State Circuits Conference (ISSCC); 22 Feb; San Francisco. All IEEE references share the same IEEE headquarters location of: New York, USA; 2015. pp. 1-3.

[37] Gnecchi S, Dutton N, Parmesan L, Rae B, Pellegrini S, McLeod S, Grant L, Henderson R. Digital silicon photomultipliers with OR/XOR pulse combining techniques. IEEE Transactions on Electron Devices. 2016;**63**(3):1105-1110.

[38] Almer O, Dutton N, Abbas T, Gnecchi S, Henderson R. 4-PAM visible light communications with a XOR-tree digital silicon photomultiplier. In: Summer Topicals Meeting Series (SUM); 13 July; IEEE; 2015. pp. 41-42.

[39] Chitnis D, Collins S. A SPAD-based photon detecting system for optical communications. Journal of Lightwave Technology. 2014;**32**(10):2028-2034.

[40] Chitnis D, Zhang L, Chun H, Rajbhandari S, Faulkner G, O'Brien D, Collins S. A 200Mb/s VLC demonstration with a SPAD based receiver. In: Summer Topicals Meeting Series (SUM), 2015; July; All IEEE Publisher Locations are New York, USA; 2015. p. 226-227.