Single Photon Avalanche Diodes for Optical Communications

Danial Chitnis, Wolfson College

University of Oxford

Michaelmas Term, 2012

Supervised by Dr Steve Collins

This thesis is submitted to the Department of Engineering Science, University of Oxford, in fulfilment of requirements for the degree of Doctor of Philosophy.

Abstract

In order to improve the sensitivity of an optical receiver, the gain and the collection area of the photo-detectors within the receiver should be increased. Detectors with internal gain such as avalanche photodiodes (APD) are usually used to increase the sensitivity of the receiver. One problem with APDs is the sensitivity of their gain to their bias voltage, which makes them challenging to be fabricated in a standard CMOS process due to variations in their gain. However, when an APD is biased over its breakdown voltage, it is sensitive to a single photon, hence, referred to as a single photon avalanche diodes (SPAD). The SPADs are photon-counting detectors, which are less sensitive to their bias voltage, and can be integrated with rest of the electronic circuitry that form an optical receiver. An avalanche diode requires dedicated circuits to be operated in the SPAD mode. These circuits make the diode insensitive to an incident photon for a duration that is known as deadtime. Unfortunately, The collection area of the PD, APD, and SPADs are limited to their capacitance. Hence, a large photo-detector leads to a larger capacitance, which reduces the bandwidth of the receiver.

In this thesis, a photon counting optical receiver based on an array of SPADs is proposed which increases the collection area with a low output capacitance. The avalanche diode and peripheral circuits which operate and readout-out the SPAD array are fabricated in the commercially available UMC 0.18 μm CMOS process. Initially, the avalanche diode is tested and characterised. A high performance circuit is then designed and tested which is able to achieve short deadtimes up to 4 ns. Once the photon counting operation of the SPAD is verified, a numerical model is developed to investigate the influence of several factors, including the deadtime, on the performance of the photon-counting detector in a communication link. Based on the simulation results, which show the advantages of an array over a single detector, a prototype detector array of 64 asynchronous SPADs is designed and tested. This array uses a high-speed readout mechanism which is inspired by the current steering digital-to-analogue converters. Bit error ratio tests (BERT) verify the photon counting capability of the proposed detector, and a bit error rate of 10^{-3} has been achieved at data rate of 100 Mbps. In addition, the array of SPAD is compatible with a front-end of conventional optical receiver which uses a photodiode as a photo detector.

Acknowledgements

I would like to thank my former colleagues Dr. Bhaskar Choubey, Dr. Jingjing Liu, Dr. Hsiu-Yu Cheng, Dr. Ratnasingam Sivalogeswaran, Dr. Ross Turnbull, Dr. Chenchen Deng, and Dr. Dipayan Das. Thank you to our collaborators Prof. David Cumming, and Dr. Mohammed Al-Rawhani at University of Glasgow. Thank you to Prof. Dominic O'Brien and Mr. Ibrahim Alsolami at University of Oxford for their inspiration and encouragement for this project. A special thanks to Mr. Grahame Faulkner for his helpful technical advice.

Above all, I would like to thank Dr Steve Collins for his supervision, funding my studies at Oxford, and the unique opportunities to design four silicon chips during my studies, which I will count as an invaluable experience for rest of my career.

List of Publications

- **Broadcasting over Photon-Counting Channels via Multiresolution PPM: Implementation and Experimental Results**, Ibrahim Alsolami, Danial Chitnis, Dominic C. O'Brien, and Steve Collins, IEEE Communications Letters, IEEE, 2012
- **Wireless Capsule for Autofluorescence Detection**, Mohammed A. Al-Rawhani, James Beeley, Danial Chitnis, Steve Collins, David R.S. Cumming, Procedia Engineering, Volume 47, Pages 48-51, 2012
- **Surface plasmon resonance for digital imaging**, David RS Cumming, Qin Chen, Kirsty Walls, Timothy D Drysdale, Stephen Collins, Dipayan Das, Danial Chitnis, 12th Conference on Nanotechnology (IEEE-NANO), IEEE, 2012
- **Design and Implementation of a Wireless Capsule Suitable for Autofluorescence Intensity Detection in Biological Tissues**, Mohammed A. Al-Rawhani, Danial Chitnis, James Beeley, Steve Collins, and David R. S. Cumming, IEEE Transactions on Biomedical Engineering, IEEE, 2012
- **A CMOS image sensor integrated with plasmonic colour filters**, Q Chen, D Das, D Chitnis, K Walls, TD Drysdale, S Collins, DRS Cumming, Plasmonics, Springer, 2012
- **CMOS Photo Detectors Integrated with Plasmonic Colour Filters**, Q Chen, D Chitnis, K Walls, T Drysdale, S Collins, D Cumming, Photonics Technology Letters, IEEE, 2012
- **Photocurrent dependent response of a SPAD biased by a charge pump**, M Al-Rawhani, D Cumming, D Chitnis, S Collins, IEEE International Symposium on Circuits and Systems (ISCAS), Rio de Janeiro, 2011
- **A flexible compact readout circuit for SPAD arrays**, D Chitnis, S Collins, Proceedings of SPIE 7780, San Diego, 2010
- **Compact readout circuits for SPAD arrays**, D Chitnis, S Collins, IEEE International Circuits and Systems (ISCAS), IEEE, Paris, 2010

Contents

List of Figures

List of Tables

List of abbreviations

ADR	Automatic Diode Reset
APD	Avalanche Photodiode
BER	Bit Error Ratio
BERT	Bit Error Ratio Test
CMOS	Complementary Metal Oxide Semiconductor
CPS	Count Per Seconds
DAC	Digital to Analogue Converter
DCR	Dark Count Rate
EMI	Electro-Magnetic Interference
ER	Extinction Ratio
FEC	Forward Error Correction
FPGA	Field Programmable Gate Array
FWHM	Full Width Half Maximum
IC	Integrated Circuit
ISI	Inter Symbol Interference
LD	Laser Diode
LED	Light Emitting Diode
LVDS	Low Voltage Differential Signalling
MPPC	Multi pixel Photon Counter
OOK	ON-OFF Keying
OPC	Optical Proximity Correction
PCB	Printed Circuit Board
PD	Photo-Diode
PDP	Photon Detection Probability
PE	Probability of Error
PLL	Phase Locked Loop
PMT	Photon Multiplier Tube
POF	Plastic Optical Fibre
SiPM	Silicon Photon Multiplier
SNR	Signal to Noise Ratio
SPAD	Single Photon Avalanche Diode
STI	Shallow Trench Isolation
TIA	Trans-Impedance Amplifier

Chapter 1

Introduction

In recent years, rapid development in integrated circuit design and its associated technologies has led to the creation of compact devices with a significant computational power at an affordable cost. These devices, such as compact laptops, tablets and mobile handsets, are capable of processing a large amount of information, hence requiring larger data bandwidths. The greater demand for these devices makes their connectivity to a data network more challenging. Usually, data is transferred to these devices via copper cables such as Ethernet for fixed devices, or microwaves such as Wi-Fi for portable devices. As the number of these devices increase, the cost of cabling and the radio interference are increasingly posing problems.

Recently, optical communication has been proposed as an alternative to traditional methods of communications for consumer devices. Traditionally, optical communication was limited to expensive fibre optical installations with high-end interfaces. Recent availability of plastic optical fibres (POF) with low cost and mass manufactured interfaces is a potential alternative to the copper cables especially targeted at mass market. These plastic fibres are free from electrical interference from neighbouring cables, do not require expert installation, and have compact end-to-end transceivers [1].

As demand for wireless connectivity is sharply increasing due to growing number of portable and mobile devices, new challenges such as radio interference from various transmitters in close proximity, power consumption, and information security are emerging. Free space optical communication has the potential to provide an alternative to the traditional wireless communications [2]. These challenges could be addressed by using free space optical communication simply by focusing the transmitted light to a location where it is required to be received. This type of communication is especially beneficial to places where electrical interference is a hazard, and large

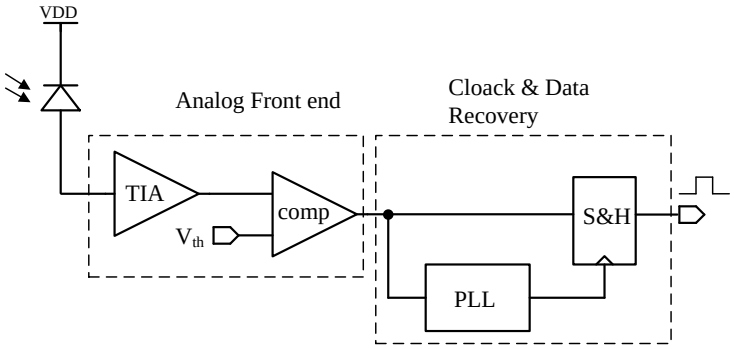


Figure 1.1: A diagram of an optical receiver

number of transceivers are in close proximity such as aircraft, and hospitals.

A major challenge with both fibre and free space optics is increasing the sensitivity of the optical receiver. The increased sensitivity leads to a lower transmitted power, a wider field of view or a longer transmission range. The main part of the optical receiver that improves the sensitivity is the photo-detector. Although sensitive photo-detectors already exist on the market, such as avalanche photo-diodes (APD), and even more sensitive such as photo-multiplier tubes (PMT), they are limited to high-end users due to their cost and availability. Some of these sensitive photo-detectors are not compatible with consumer electronic devices, and require different manufacturing process. For example, a PMT is a sensitive but fragile photo-detecor, which requires a high bias voltage in range of kilo volts, hence, it is limited to scientific and military users [3].

Recent advancement in single photon avalanche diodes, which was originally invented during a research on avalanche breakdown in p-n junctions in 1965 [4], has led to the creation of sensitive photon counting detectors which are manufactured using the same process as the consumer electronics integrated circuits are manufactured. Hence, there is a potential to integrate these photon detectors on the same IC as the rest of electronic circuits in the product, leading to a cheaper device which is affordable for the mass consumers.

1.1 An optical receiver

Figure 1.1 shows a typical optical receiver which consist of a photodetector, an analogue front end, and a clock and data recovery mechanism. The data is transmitted serially as a logic 0 and 1. The transmitter is usually a laser diode (LD) or a light

emitting diode (LED) modulated by the digital input data. The most popular and simple method of serial transmission is to turn ON the light emitting device when the input data is a logic 1, and turn it OFF when the data is a logic 0. This method of transmission is referred to as ON-OFF keying (OOK).

On the receiver side, a photo-detector generates an electrical current based on the intensity of the incident light. There are various types of photo-detectors used in systems that require detecting light. The simplest photo-detector is a photodiode (PD), which is easily fabricated, and has the ability to be integrated with other peripheral electronics. The generated photocurrent is converted to a voltage by a trans-impedance amplifier (TIA). As a result, the intensity of the incident light is proportional to the output voltage of the TIA.

In free space optical communications, the photo-detector has background noise in addition to the signal intensity. This is the incident light on the photo-detector when the transmitter is sending a logic 0. In order to distinguish between the generated voltage from the transmitter's signal and the background noise, a voltage comparator is used. If the input voltage of the comparator is larger than a threshold voltage V_{th}, then the output of the comparator is a logic 1, otherwise the output is a logic 0. The digital output of the comparator is then re-timed and aligned to the original clock of the transmitter by the clock and data recovery unit. The original clock of the transmitter is recovered usually by a phase locked loop (PLL), and a sample and hold circuit re-times the input data to the recovered clock. Now, the output of the sample and hold is prepared to be processed by following digital circuitry such as line coding and error correction blocks.

1.2 Current to voltage converter

The photodiode is the most commonly used photo detector. When an incident photon is absorbed within the depletion region of the pn junction, an electron hole pair is released. The electron-hole pair is then separated by the applied electrical field. Hence, a photo current is generated at terminals of the photodiode. In order to monitor the output current of the photodiode, this current is converted into voltage. There are two major methods to covert the current into voltage in an optical receiver circuit: a resistor, and a trans-impedance amplifier.

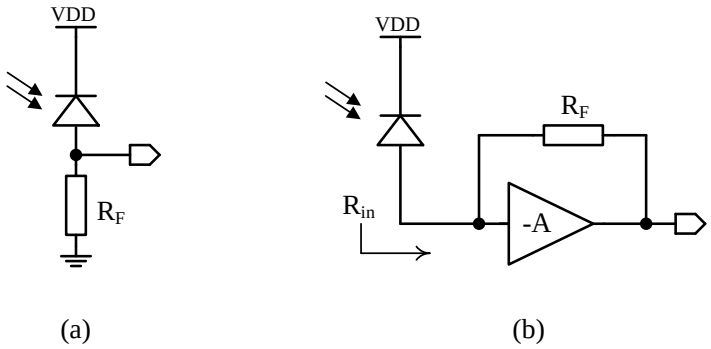


Figure 1.2: The current to voltage converters (a) a resistor (b) a trans-impedance amplifier

1.2.1 Resistor

The simplest method to convert the current into voltage is a series resistor in the current path of the photodiode, hence the voltage across the resistor R is proportional to the photo-current. Figure 1.2.a shows the schematic diagram of the resistor converter. In this case, the conversion ratio of the current-to-voltage converter is equal to R. The depletion region in the pn junction of the photodiode has an intrinsic capacitance C_d. This capacitance forms a RC circuit with the series resistor, hence the electrical bandwidth is:

$$BW_{3db} = \frac{1}{RC_d} \tag{1.1}$$

Although the conversion ratio increases by selecting a larger resistor, the electrical bandwidth decreases which leads to a longer rise/fall times. In addition, the electrical bandwidth has an inverse relationship with the capacitance of the photodiode. Hence, a larger photodiode of the same type leads to lower bandwidth.

1.2.2 Transimpedance amplifier

One of the most common methods to avoid the reduction in the bandwidth is a trans-impedance amplifier (TIA). Figure 1.2.b shows a schematic diagram of a TIA. The TIA consist of a voltage amplifier which has a negative gain of A, and a shunt-shunt feedback resistor R_F .The equivalent input resistance of a trans-impedance amplifier, R_{in} is [5]:

$$R_{in} = \frac{R_F}{A+1} \tag{1.2}$$

As a result, a higher gain of the amplifier leads to a lower input resistance. In this case, the capacitance of the photodiode C_d forms an RC circuit with the input resistance of the TIA. Hence, the electrical bandwidth of this RC circuit is:

$$BW_{3db} = \frac{A+1}{R_F C_d} \tag{1.3}$$

In comparison to the bandwidth of the series resistor described in Equation 1.1, the electrical bandwidth of a TIA has increased by $(A+1)$ times for the same value of the resistor and capacitor. The TIAs can be integrated with the photodiode in a single silicon substrate using standard CMOS technologies [6, 7].

1.3 Noise

The amount noise defines the sensitivity of the optical receiver. The lowest detectable signal is determined by the voltage noise at the output of the current-to-voltage converter. The input referred noise at the input of the TIA is expressed as the power spectral density which has a unit of A^2/Hz for a current noise. The mean square noise which is the statistical variance of the transient current is expressed as [5]:

$$\overline{i_n^2} = \int_{BW} i_n^2(f).df \tag{1.4}$$

where i_n^2 is the mean square of the current noise, $i_n^2(f)$ is the spectral density of the current noise, and BW is the bandwidth which the noise is being observed in.

In an optical receiver, three major sources of noise are the noise of the photodiode, the feedback resistor, and the voltage amplifier. The noise of the photodiode is related to statistical fluctuation of the incident photons which is also referred to as photon shot noise. The spectral density for the photon shot noise is:

$$i_{n,PD}^2(f) = 2qI_D \tag{1.5}$$

where I_D is the photo-current, and q is the electrical charge of an electron. The photon shot noise is a white noise, hence, independent of the frequency, and it increases with the photocurrent. The noise from the resistor is a thermal noise, and the spectral density is expressed as:

$$i_{n,R}^2(f) = \frac{4kT}{R} \tag{1.6}$$

where k is the Boltzmann constant, T is temperature in Kelvin, and R is the value of resistor in Ohms. Similar to the photon shot noise, the thermal noise of the resistor is a white noise, and it is inversely proportional to the value of the resistor. The total input referred noise of the system can be calculated from the noise of each component. In the case of an optical receiver's front-end with a resistor shown in Figure 1.2.a, the total noise is:

$$i_n^{rms} = \sqrt{\overline{i_{n,PD}^2} + \overline{i_{n,R}^2}} \tag{1.7}$$

where i_n^{rms} is the root mean square (RMS) of the total noise, $\overline{i_{n,PD}^2}$ and $\overline{i_{n,R}^2}$ are the mean square of the photodiode and the resistor noise respectively. As for most optical receivers, the dominate source of noise is the noise from the resistor, if the signal intensity is maintained, increasing the resistor leads to a lower total current noise. However, increasing the resistor, leads to lower electrical bandwidth due the effect of the RC circuit.

When a trans-impedance amplifier (TIA) is used as shown in Fig 1.2.b, the total noise is:

$$i_n^{rms} = \sqrt{\overline{i_{n,PD}^2} + \overline{i_{n,R_F}^2} + \overline{i_{n,front}^2}} \tag{1.8}$$

where $\overline{i_{n,R_F}^2}$ is the mean square noise of the feedback resistor R_F, and $\overline{i_{n,front}^2}$ is the mean square noise of the voltage amplifier's front-end. The spectral density of the noise of the amplifier consist of a white noise element, and a frequency dependent element [5]. The white noise element, depends on the input stage of the amplifier which is either a bipolar or field effect transistor (FET). The frequency dependent element has a direct relationship with the input capacitance of the amplifier. As a result, the total input referred noise of the TIA decreases with a higher feedback resistor and a lower capacitance of the photodiode.

1.4 Bit error ratio

During data transmission, electrical noise appears in both amplitude and timing of the received signal. Fig 1.3.a shows a random digital data at the input of the transmitter. The sources of the amplitude noise are the photodiode, resistor, and the amplifier. The sources of timing noise are mainly due to the noise in the timing circuitry. The x-axis is normalised to the bit duration T_b, starting from $t = 0$, and ending at $t = 1$. In order to assess the quality of the received signal, an eye diagram is formed by

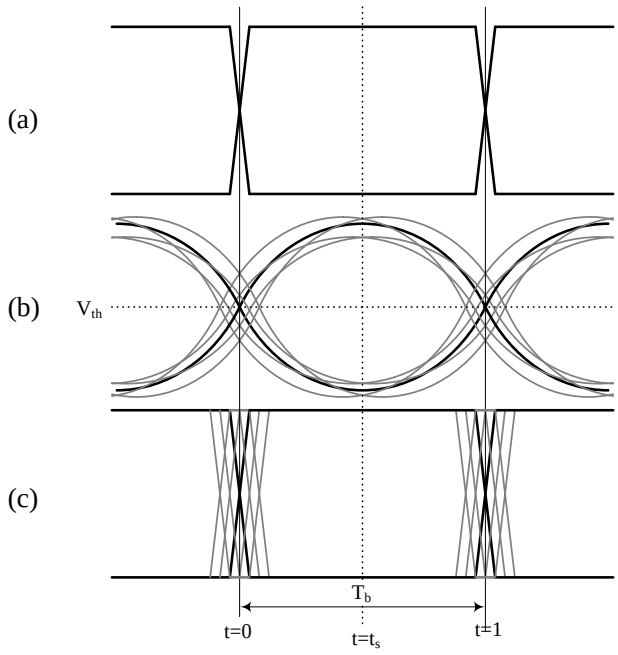


Figure 1.3: The diagram of signals during data transmission (a) input data to the transmitter (b) output of the TIA (c) output of the comparator

superimposition of the signal traces that are captured by an oscilloscope when it is triggered on the clock of the transmitted data. Fig 1.3.b shows a schematic of an eye diagram for output voltage of the TIA in an optical receiver. The amplitude noise creates a vertical uncertainty, and the timing noise creates a horizontal uncertainty. In addition, the finite electrical bandwidth of the system causes slower rise/fall times than the original digital signal.

The eye diagram is a two dimensional relative histogram. Each horizontal and vertical coordinate in the eye diagram represents a point in time and amplitude respectively. The intensity of this coordinate represents its relative frequency. In Figure 1.3.b, the most frequent coordinates are shown in a darker shade than the less frequent coordinates. When the opening of the eye has a symmetrical shape similar to the eye shown in Figure 1.3.b, the least frequent coordinate is centre of the eye.

Once the signal has passed through the comparator stage, it is converted back into a digital signal, and the amplitude noise (vertical uncertainty) disappears. Fig 1.3.c shows the output voltage of the comparator. The threshold voltage of the comparator V_{th} is equal to the amplitude of the centre of the eye. Due to the amplitude noise at the input of the comparator, some errors occur at this stage. On some occasions, the output voltage of the TIA does not rise beyond the threshold voltage when a logic 1

is transmitted. As a result, the bit is interpreted as a logic 0. A similar case occurs when the output of the TIA rises beyond V_{th} when a logic 0 is transmitted, hence interpreting it as a logic 1.

Apart from the noise of electronic circuits, one of the major contributors to the amplitude noise (vertical uncertainty) is the electrical bandwidth of the system. The finite bandwidth of the system leads to slower rise/fall times. When the rise/fall time is slow, the output voltage of the TIA does not rise/fall to its final voltage level before the next bit starts. As a result, depending on the logical value of the next bit, the output voltage may not rise/fall beyond the threshold voltage of the comparator, leading to a bit error. Therefore, a smaller bandwidth leads to a slower rise/fall time which causes more bit errors.

Although the amplitude noise is eliminated from the output of the comparator, the timing noise (horizontal uncertainty) is still present. The timing noise is also referred to as jitter. In order to obtain a clean digital signal with negligible jitter similar to the one shown in Fig 1.3.a, the clock and data recovery section eliminates the jitter at the output of the comparator. Initially a phase locked loop (PLL) circuit recovers the clock from the output signal of the comparator. This clock is aligned with the original transmitted data. The clock is then shifted by 180^o to ensure that the sampling of the signal occurs at the time coordinate of the centre of the eye t_s. In this stage, errors may occur due to the jitter at the input of the sample and hold circuit. If the jitter is too high, then the sampling t_s moves into the adjacent slots, sampling the wrong bit, and leading to a bit error.

Therefore, errors could occur both in the comparator stage and in the data recovery stage. A major performance parameter is the bit error ratio (BER) of the communication link. The BER is simply defined as:

$$BER = \frac{bit\,errors}{total\,bits} \tag{1.9}$$

If the noise is based on a Gaussian distribution, theoretical equations could be used to estimated the BER [8, 9, 5]. In cases when the distribution of the noise is non-conventional, numerical simulations are used to estimate the BER. Once the fundamental components of the optical receiver and transmitter are designed, forward error correction (FEC) could be used to improve the bit error ratio [10]. In optical communications, the major effort is focused on reducing the BER, as it indicates the quality of the communication link.

1.5 Avalanche photodiode

In applications for which sensitivity is critical, it is important to use the largest possible detector to generate the largest possible photocurrent. Although optical elements such as concentrators can be used to increases the collection area of the sensor, they limit the field of view of the photo-detector. This means that they give limited benefit in some applications and even when they are used, the sensitivity of the overall system will be improved by using larger photodetectors. Unfortunately, larger photodiodes have a larger capacitance and a lower RC bandwidth.

A popular approach to increasing the photocurrent without increasing the area of the photodetector is to use a photodiode that is specifically designed so that it can be operated at a high bias voltage. It is then possible to create a high electric field within the photodiode. If the electrical field is high enough, then each detected photon can create a large number of electron-hole pairs via a physical process known as avalanche multiplication. This process creates internal gain within the photodiode. These more sensitive detectors are known as Avalanche Photo-diode (APD).

The internal gain within an APD increases when the reverse bias voltage of the APD increases, and in particular it increases very rapidly as the applied voltage approaches the voltage at which the avalanche process becomes self-sustaining. This process is referred to as avalanche breakdown, and the voltage which the avalanche breakdown occurs is referred to as breakdown voltage.

Avalanche multiplication is intrinsically a statistical process. Hence, there are statistical fluctuations in the amount of charge generated during the avalanche process. That means the APD has an average gain of $\overline{g}$. In order to assess the randomness of the gain, the excess noise factor is defined as [9]:

$$F = 1 + \frac{var\{g\}}{\overline{g}^2} \tag{1.10}$$

The excess noise factor F is an indicator of the randomness of the gain g. In the case of a photodiode, the gain is $g = 1$ and there is no randomness in the gain, hence $F = 1$. The noise factor increases with the randomness of the gain. In an avalanche photodiode (APD) the excess noise factor F is [5]:

$$F = \gamma\overline{g} + \left(2 - \frac{1}{\overline{g}}\right)(1 - \gamma) \tag{1.11}$$

where γ is the avalanche multiplication factor, and $\overline{g}$ is the average gain. The avalanche multiplication factor is typically between $0.01 < \gamma < 0.1$ for silicon APDs.

The Equation 1.11 shows that the excess noise factor increases with the gain. The spectral noise density of an APD's current is defined as:

$$i^2_{n,APD}(f) = F.\overline{g}^2.2qI_D \tag{1.12}$$

where I_D is the photo current before the internal gain is applied. In comparison to the spectral noise density of a PD described in Equation 1.5, although the current of the APD has increased by $\overline{g}$, the noise of the APD has increased by $F.\overline{g}^2$. Because an increase in the avalanche gain leads to generation of more noise, the APD has an optimum gain which can be operated. This optimum gain, which maximises the SNR, is calculated during the design stage of the trans-impedance amplifier.

Many APDs require bias voltages well in excess of normal operating voltages available within consumer equipment. In this bias range, the gain of the APD is very sensitive to changes in the bias voltage and operating temperature. Ideally, the bias voltage applied to the APD should therefore be well controlled at a value that depends upon the operating temperature of the APD [11, 12]. Moreover, the production of an APD is generally more difficult than a conventional PD, as they require specific process fabrication to achieve a high gain and low noise. In addition, they require a higher process quality as their gain is sensitive to the bias voltage and any non-uniformity in the electrical field could severely degraded the performance of the APD. As a result, the production yield of an APD is lower than a PD, leading to higher manufacturing cost.

The widely available standard CMOS process is not suitable for APDs as the reverse bias region which avalanche gain occurs is too narrow which makes these devices particularly sensitive to changes in the bias voltage and temperature [13]. The resulting inability to integrate APDs with other components increases the power consumption, cost and size of systems [14].

1.6 Single photon avalanche diodes

Many of the problems associated with avalanche photo diodes (APDs) can be avoided if the photodiode is biased so that the avalanche process becomes self-sustaining. These devices can then be converted to single photon detectors by placing the avalanche diode in series with another device so that the current which flows in response to an avalanche process automatically reduces the voltage across the diode so that the avalanche process is stopped or quenched. After a period, the voltage across the

diode is reset to its initial value ready for the detection of the next photon. When used in this mode the detector is known as a single photon avalanche diode (SPAD), and the period when the voltage is too low to sustain avalanching which is referred to as the deadtime.

The output of the SPAD is a digital-like pulse which indicates an avalanche event. Hence, unlike PDs and APDs, the SPAD is a photon-counting detector which means the intensity of the illumination is proportional to the event rate. Unlike APDs, SPADs do not have excess noise, and are not sensitive to their bias voltage and temperature variation as the avalanche current is regarded as binary current. As a result, SPADs do not require special fabrication process, and are widely compatible with standard CMOS processes [15, 16, 17].

The quenching and recovery processes are an essential part of the SPAD operation. Simple circuits such as a resistor could be used to quench the avalanche current [18, 19], or more sophisticated active circuits which have a higher performance and a faster deadtime [20, 21, 22, 15, 23]. When SPADs are fabricated in CMOS, the avalanche diode and the peripheral circuit can be integrated into the same silicon substrate. This integration has the benefits of a smaller circuit size as well as higher speed, lower power operation due the low parasitic capacitances, and a lower production cost.

Current application for SPADs include laser range finding [24, 25], three-dimensional imaging [26, 27, 28, 29, 30], fluorescence imaging [31, 32, 33, 34] and charged particle detection [35, 36, 37, 38]. Most of these applications require operations at very low illumination, as a result most of the recent efforts have been focused on reducing the dark count rate of the SPAD [39, 40, 41, 42, 43]. The dark count rate (DCR) is the rate of spontaneous generation of avalanche events which are not related to the incident photons. These avalanche events are detected even when the detector is in the dark. This is similar to PD and APD which have a dark current. In order to reduce the DCR, special modifications to the CMOS process are required, which consequently increases the cost of manufacturing. In addition, the customised fabrication process leads to a lower reliability and production yield.

However, in high-speed optical communications, because of a very short bit durations (in a range of few to tens of nanosecond) when there is no background noise, the dark count is typically few orders of magnitudes smaller than the signal count. When the background noise do exist, the count related to the background noise is generally higher than the dark count. In this case, the DCR is not relevant to the performance of the optical receiver. As a result, SPADs which are fabricated in standard CMOS

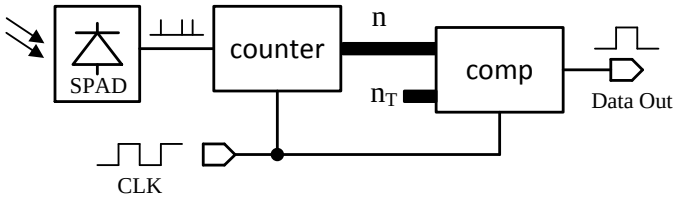


Figure 1.4: A photon counting optical communication receiver using a SPAD

processes can be used to achieve a higher reliability and a lower manufacturing cost without compromising the performance of the optical receiver.

1.7 SPADs for optical communications

Due to the photon counting nature of the SPAD and digital-like output pulses, the conventional optical receiver shown in Figure 1.1 cannot be used. Figure 1.4 shows a photon counting receiver for optical communications using a SPAD as a photo detector. The digital pulses at the output of the SPAD are random-like, and the time interval between two consecutive pulses has an exponential distribution [8, 9]. A digital counter counts theses pulses within the bit period, T_b. The input clock of the digital counter has a frequency of $1/T_b$, and is perfectly aligned with the transmitter's clock. The number photons counted within bit period, n, has a Poisson distribution, due to the exponential distribution of the pulse intervals.

In an ideal situation when there is no background noise and dark counts, no photons are detected when a logic 0 is transmitted, hence, if one or more photon is detected the received bit is declared as a logic 1. In this case, errors are caused when no photon is detected due to the statistical distribution of n when a logic 1 is transmitted. As the average number of counts, $\overline{n}$, increases, the probability of missing a photon decreases, hence the bit error ratio decreases. As a result, the BER decreases when signal intensity increases.

If there is background noise, photons are detected even when a logic 0 is transmitted. As a result, a threshold, n_T, is required to distinguish between logic 0 and logic 1. A digital comparator compares the number of detected photon in each bit period to threshold number. If $n > n_T$ then the received bit is declared as a logic1, and if $n \leq n_T$ then it is declared as a logic 0. In this case, the errors are caused when a transmitted logic 1 is received as a logic 0, and a transmitted logic 0 is received as a 1. The bit error rate now depends on the signal and background noise, and the ratio

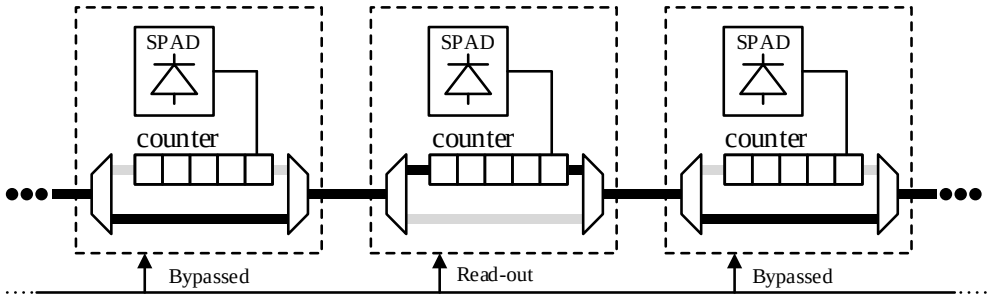


Figure 1.5: A digital array of SPADs containing a cell level counter. The disabled buses are shown in grey.

between the signal and the background noise.

A photon-counting receiver for proof of concept experiments was implemented using SPADs as a photon counting detector to assess modulation schemes for free space optical communications [44].

1.8 Arrays of SPADs

In order to increase the sensitivity of the photon counting optical receiver, a larger collection area should be used. However, similar to other detectors the maximum size of a SPAD is limited by the fact that it can only detect single photons in a predetermined period and if photon is absorbed within the device during its deadtime it will be missed. Increasing the size of the photodiode within the SPAD will also increase the dark count rate.

The collection area of the SPAD is using an array on asynchronous SPADs which operates independently, and have their own quenching circuit. The overall photon count is an aggregation of the count from each individual SPAD. One possible approach to detecting the avalanche events in each SPAD is to separately connect each SPAD to external circuits. Although this method is practical in small arrays, it is not practical in larger arrays due the large amount of digital data, and limited availability of output connections.

1.8.1 Digital arrays

One approach to avoid a large number of output connections is to incorporate a digital counter in each cell to store the number of events [45, 14, 46, 47, 48, 49, 50]. Similar to the readout method in image sensors, the array is scanned, and the number

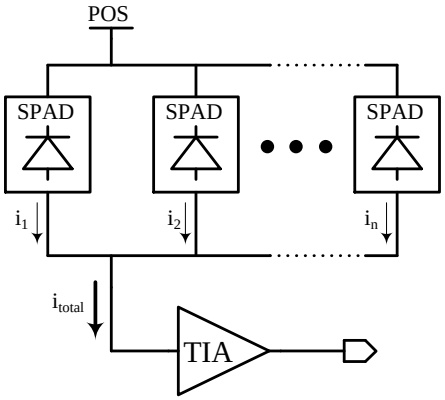


Figure 1.6: An analogue SPAD array also know as multi-pixel photon counter (MPPC)

of events stored in the counter is read-out by addressing each pixel. These counts are then added to an overall sum by a single accumulator. Figure 1.5 shows an implementation of a SPAD arrays with a digital counter in each cell [14, 48]. The output of the SPAD is connected to the digital counter so that each avalanche event increments the counter. The value of the counter is readout via a digital bus which passes through a row (or a column) of the array. A global clock addresses a cell within the row (or the column) enabling the counter to connect to the digital bus while the rest of the counters connected to the same bus are bypassed by a digital multiplexer. The digital buses are connected to a digital memory which is then read-out by an off-chip digital interface to be processed later. Due to the digital output of this readout circuit, an optical receiver similar to the receiver shown in Figure 1.4 is required.

1.8.2 Analogue arrays

Although a digital readout method benefits from operating in digital domain and avoiding system noise, in order to readout large arrays operating at high bit rates, a high frequency global clock is required which is challenging in a practical implementation. Another approach to reduce the number of output connections in a SPAD array is to share the currents which are following through each SPAD during the quenching and recovery processes. Each avalanche diode approximately generates similar amount charge during an avalanche event. Similar to a photodiode, a current to voltage converter is used to monitor the output of the array. The output voltage is proportional to the number of SPADs which simultaneously detected an avalanche event. Figure 1.6, shows this type of SPAD array that is commercially available.

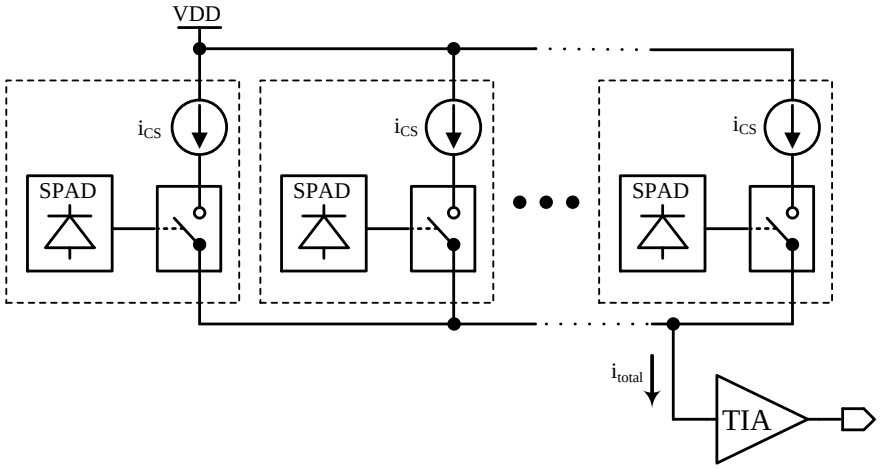


Figure 1.7: A simplified diagram of the proposed SPAD array

These arrays are referred to as multi pixel photon counters (MPPC) [51, 52, 53, 54], and silicon photo-multipliers (SiPM) [55, 56, 57, 58]. The major application for these detectors are charged particle detection, and biomedical engineering.

These arrays offer a simple and scalable physical layout. However, the total output capacitance of the array is the sum of the capacitance of individual avalanche diodes due to the parallel connection between each diode. As a result, similar to PDs and APDs, the RC bandwidth limits the performance of the detector. Moreover, the amount of charge per avalanche event may not be large enough to be above the input referred noise of the TIA, which is described in Equation 1.8. Unfortunately, this current cannot be controlled, as it is a physical characteristic of the avalanche diode.

1.9 The proposed solution

The problem with the MPPC, described in the previous section, is lack of a dedicated readout circuit within the SPAD cell. The proposed solution to the problems with the previous systems is to employ a method that has been used with high speed digital to analogue converters (DACs). Figure 1.7 shows a simplified diagram of the proposed SPAD array. In addition to the SPAD circuit, each cell contains an equally weighted current source, and a switch which enables the current of that cell to be added to a common output node. The currents from the cells are summed by sharing these outputs between cells. The resulting currents can be converted to voltages using either a resistor or a TIA as shown in Figure 1.2.

All the equivalent outputs from each building block are connected in parallel

which keeps the layout simple and easy to scale. Unlike an MPPC, the total output capacitance is sum of the capacitance of each individual switch which is significantly smaller than the capacitance of the avalanche diode. As a result, similar to high speed DACs, large number of these switches can be connected in parallel and operated at high speeds [59, 60, 61, 62, 63, 64, 65, 66].

The output of the array is a current which is proportional to the number of SPADs in which an avalanche event has recently occurred. The individual current sources can be designed and/or biased to adjust the current flowing through individual current sources. This effectively means that unlike MPPCs, the system designer and/or user controls the gain of the system. This flexibility could be used to ensure that the signal generated by a single photon is larger than the noise generated in the rest of the system. This ensures that each individual event is detected, and the photon counting behaviour of the detector is maintained. In addition, because of the current output, such a detector is compatible with the conventional optical receiver shown in Figure 1.1.

1.10 Summary

In this chapter, the major components of an optical receivers were described. The photo-detector converts the light into a photo-current. This current is then converted into voltage by the means of a resistor or a trans-impedance amplifier. The trans-impedance amplifier has the advantage of a higher electrical bandwidth in comparison to the resistor method due to a lower input resistance. The sensitivity of the receiver depends on the noise in the system. The sources of the noise are the noise of the photo-detector which is the shot noise, the noise of the feedback resistor, and the noise of the voltage amplifier. In order to reduce the noise, a large feedback resistor is required. However, this leads to a lower RC bandwidth due to the capacitance of the photodiode. The system noise and the low electrical bandwidth contribute to an increase in the bit error ratio (BER) which is the ratio between the number of bit errors and total the number of transmitted bits.

Avalanche photodiodes (APD) are detectors which have an internal gain by means of avalanche multiplication. Their gain increases with their bias voltage, however with the cost of increased noise due to randomness in the avalanche gain. This problem can be avoided when these detectors are biased over their breakdown voltage so that the avalanche process is sustained, and the photo-diode current is regarded as a binary current. These detectors are known as single photon avalanche diodes (SPAD). Unlike

APDs, these detectors are widely fabricated in standard CMOS process.

Similar to PD, and APDs, single photon avalanche detectors are limited to their photon collection area. An array of SPAD, which are connected in parallel, increases the collection area of the detector. The output of the array is the aggregated output of each individual SPAD. However, reading the output of each SPAD is challenging in a large array. The proposed array of SPADs enables the creation of large arrays which can be read-out at a high speed based on the commonly used high-speed digital to analogue converters.

Chapter 2

The Diode Characterisation

2.1 Introduction

In a pn junction with high electrical field, the electrons and the holes gain enough kinetic energy to collide into the crystal lattice, releasing further electron-hole pairs. Continuation of this process results in the generation of a large number of electron-hole pairs. This processes is referred to as impact ionization [67, 68]. This phenomenon also known as an avalanche multiplication, is the basis of operating a diode in the avalanche photodiode (APD) mode. When a photon is absorbed in the depletion region of the diode, an electron-hole pair is released. If the diode is biased at a high voltage, the high electrical field leads to an avalanche multiplication, hence, a large number of electron-hole pairs are created which consequently lead to a larger photo-current. The gain of the APD which is the ratio of avalanche multiplication, increases by the electrical filed, hence the bias voltage. When the electrical field is beyond a certain threshold [68], a sustained avalanche breakdown occurs, where significant amount of charges are generated. The SPAD operation mode, utilises these significant gain created by the avalanche breakdown. However, in order to maintain such a high electrical fields within the diode, a special geometry and layout is required.

2.2 Avalanche diode

A planar pn junction in a CMOS process has a three dimensional structure. Figure 2.1.a shows a typical cross-section and layout view of a p+/n photodiode. The p+ is diffused into the n layer, which causes rounded edges at the sidewalls of the junction. In addition, the rectangular shape of the layout creates sharp corners at each side of

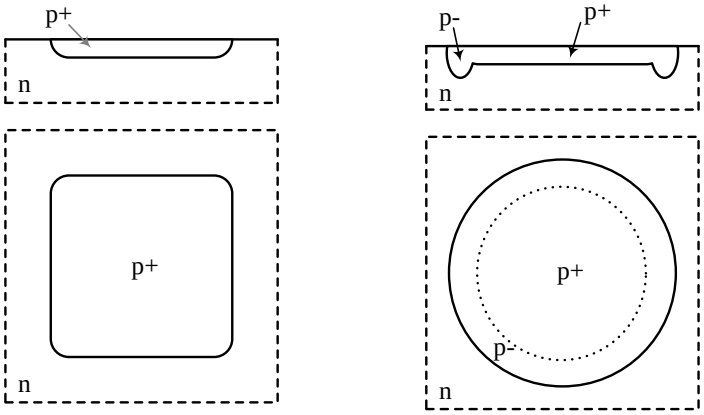


Figure 2.1: A typical cross-section and layout view of a photodiode (a) conventional diode (b) avalanche diode

the p+ layout. Each sharp edge and corner is a curvature with a small radius, hence leading to a higher electrical field at these locations [69]. As a result, the electrical field is not uniform across the pn junction due to a higher intensity in the corners and the edges than the majority of the active area of the pn junction. When the diode is close to its breakdown voltage, most of the impact ionization occurs in edges and corners rather than the active area, which is the majority of the photon collection area. The higher electrical fields lead to a lower avalanche breakdown voltage [68]. When the bias voltage of the pn junction is increased, eventually the sharp areas of the photodiode break down at lower voltages than expected.

In order to avoid the premature breakdown, the electrical field is reduced at the edges and the corners. In the edge, the electrical field is reduced by diffusing a lightly doped guarding surrounding the sidewalls of the pn junction. Figure 2.1.b shows a typical diagram of an avalanche diode with the lightly doped guard ring which is the p- surrounding the p+ active area. As a result of this guard ring, the electrical field is reduced in the edges. In addition, the layout of the pn junction is designed as a circle to avoid sharp corners with high electrical fields. As a result of these modifications, the active area now has a uniform electrical field, and the majority of impact ionization occurs where the photon collection area of the detector exist.

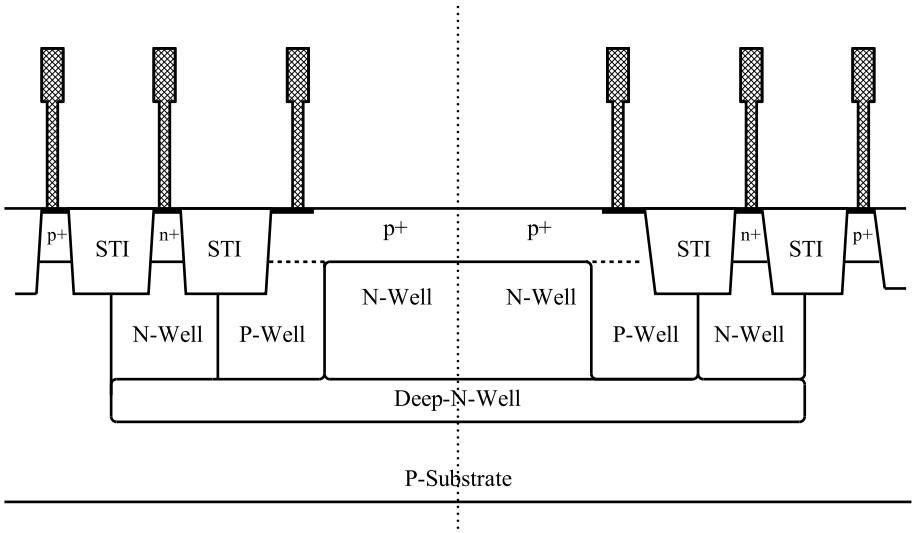


Figure 2.2: The cross-section of the implemented avalanche diode

2.2.1 Layout

Figure 2.2 shows the cross section of the implemented avalanche diode in the commercially available $0.18\,\mu m$ UMC CMOS process. The diode is a p+/Nwell junction, and the guard ring is a lightly doped p material referred to as Pwell, which is deeper than the p+ and slightly shallower than the Nwell. The N-well layer in the junction that provides noise isolation from the silicon substrate and rest of the devices within the substrate. The circular layout of the photodiode is created by minimum sized vertical and horizontal steps as the design rules restrict the use of non-45 degree angles within the layout. This is because each layout design undergoes an optical proximity correction before the masking stage. The optical proximity correction (OPC) is a correction method which compensates for the reduction in the resolution of the projected image of the mask during the photo-lithography process [70]. For non-45 degree angles, the OPC is a numerically intensive, hence, the layout is restricted only to orthogonal and 45 degree angles by the foundry. As a result, for simplicity, the circular layout is created by 10 nm horizontal and vertical steps, which after lithography is expected to appear as a smooth circle on the silicon substrate.

Based on previously reported avalanche diodes in similar processes, an active area of 10 μm diameter was selected for the avalanche diode [16, 15]. If the diameter of the active area is too small, then the inner depletion region of the surrounding guard ring could merge from either side of the junction and form a p-/n junction instead of p+/n [15]. If the diameter is too large, then a larger capacitance and a higher dark count rate are expected. Table 2.1 shows the dimensions of each layer used in the design of

#	Layer Name	Inner Radius (μm)	Outer Radius (μm)
1	p+	0	7
2	P-Well	5	9
3	Silicide Block	0	5
4	Anode contact	5.5	6.5
5	Cathode contact	10.5	11.5
6	N-Well	0	13
7	Substrate contact	14.5	15.5

Table 2.1: The dimensions of the implemented avalanche diode

the implemented avalanche diode. The active area is formed by a p+ which has a 7 μm radius. However, the Pwell guard ring is overlapped for 2 μm, hence, reducing the active area to a 5 μm radius. A silicide block layer is implemented in order to prevent deposition of the silicide on top of the active area. In recent small node technologies, silicide, which is made of metallic compounds, is used to improve the conductivity of silicon layers. Although the conductivity improves, the transparency of the active area is reduced. The entire junction area is surrounded by an N-well ring which is electrically connected to the N-well layer of the pn junction through a deep N-well layer. A layer of oxide electrically insulates each section of the avalanche diode. This oxide is a shallow trench isolation (STI) that is formed between n+ or p+ sections in the layout. As STI has been reported as a major contributor to the DCR [71], it is ensured during the design stage that the STI is not in contact with the high electrical field.

2.2.2 I-V characteristics

Figure 2.3 shows the I-V characteristics of the implemented avalanche diode for 5 fabricated chip. Each chip is illuminated with a light emitting diode (LED) driven by a constant current source. Initially, the dark current of each avalanche diode is measured. The intensity of the LED is set to ensure that the light response of the avalanche diode is approximately 10 times higher than the highest dark current among the 5 tested avalanche diodes. The avalanche breakdown occurs at the bias voltage which has a sharp increase in the current. The current of the avalanche diodes change 2.6 times from 0V to 10.0V. However, in the range 10.2V to 10.4V the current sharply increases by approximately 7 orders of magnitude, hence, the breakdown voltage in average is approximately 10.4V.The low gain until 10.2V and a very sharp increase in the avalanche gain in the range of 200mV (10.2V to 10.4V) makes these avalanche

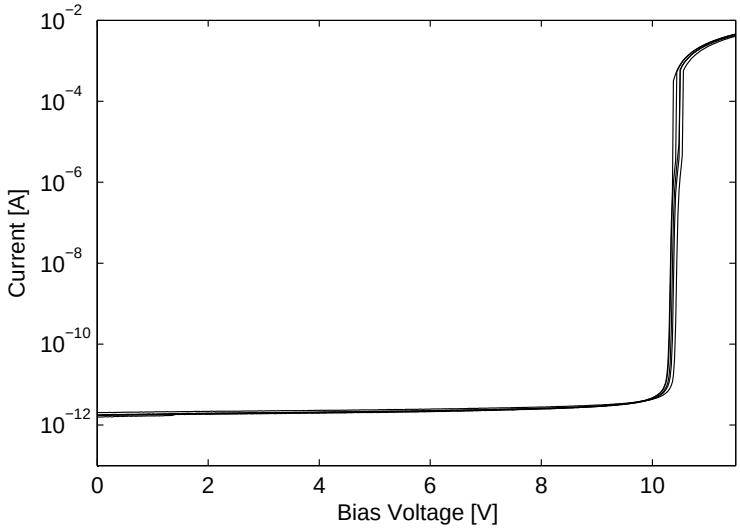


Figure 2.3: The I-V characteristics of the implemented avalanche diode illuminated with an LED (20mV steps for the bias voltage)

diodes unsuitable for use in APD mode as the gain is too sensitive to the bias voltage. In addition, the small variability in the breakdown voltage of each avalanche diode leads to a different gain when the diodes are biased at the same voltage. When the bias voltage is larger than the breakdown voltage, which is 10.4V, the diode has a sustained avalanche current which is proportional to the bias voltage due the space charge resistance.

2.3 The SPAD circuit

In the SPAD mode, when the bias voltage is larger than the breakdown voltage, the avalanche diode requires circuits to quench the avalanche current, and recover the diode after an avalanche event. These circuits can be implemented off chip or integrated into the same substrate where the avalanche diode is fabricated.

2.3.1 External quenching

In order to operate the avalanche diode in the SPAD mode, the avalanche diode should be biased over its breakdown voltage. Figure 2.4.a shows a simple schematic diagram required to operate the diode in the SPAD mode. The cathode of the avalanche diode is connected to an external resistor. The resistor is connected to a positive supply voltage and the anode of the avalanche diode is connected to a negative supply voltage. In the idle mode, the capacitance of the avalanche diode is fully charged,

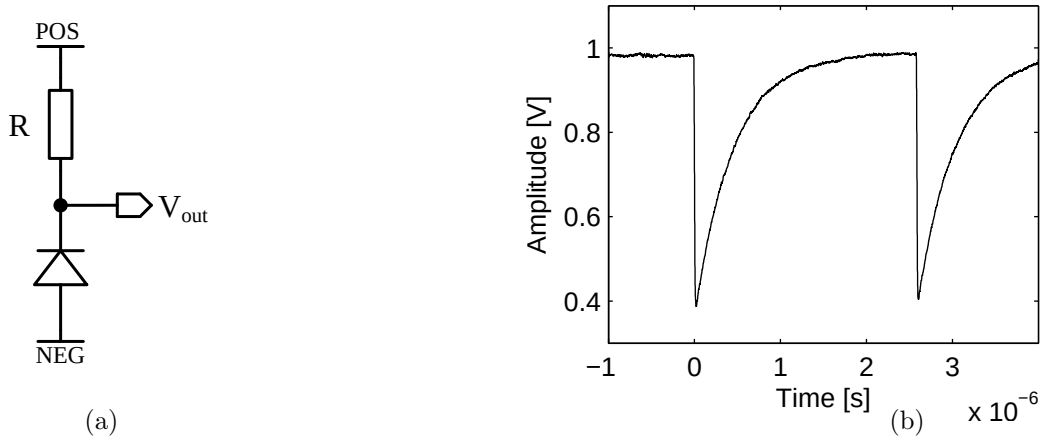


Figure 2.4: The avalanche diode with external quenching (a) the schematic diagram for the SPAD circuit (b) the measured voltage at the output node when an avalanche event occurs

and the voltage of the output node is equal the voltage of the POS terminal. The bias voltage of the diode is:

$$V_A = V_{POS} - V_{NEG} \tag{2.1}$$

where V_{POS} is the voltage of the POS terminal and V_{NEG} is the voltage of the NEG terminal. As a result, the voltage across the capacitance of the pn junction is V_A. The over voltage of the diode is defined as:

$$V_{OV} = V_{POS} - V_{NEG} - V_{br} = V_A - V_{br} \tag{2.2}$$

where V_{br} is the breakdown voltage. When an avalanche event occurs, the capacitance of the diode is rapidly discharged by the avalanche current. As a result, the voltage at the output node sharply drops to the breakdown voltage. This voltage drop is equal to the over voltage, V_{OV}. Once the voltage across the diode is close the breakdown voltage, the avalanche current is stopped. This process is referred to as quenching. In this method, the quenching is achieved by the means of an external resistor [18, 19]. Once the avalanche current is quenched, the capacitance is charged to its initial voltage V_{POS} by the resistor R. Hence, the voltage at the output node recovers with a recovery time equal to the rise time (10%-90%) of the output voltage:

$$t_r = 2.2 \times R \times C_d \tag{2.3}$$

where R is the resistance of the quenching resistor, and C_d is the capacitance of the diode. Figure 2.4.b shows the voltage at the output node when an avalanche event has

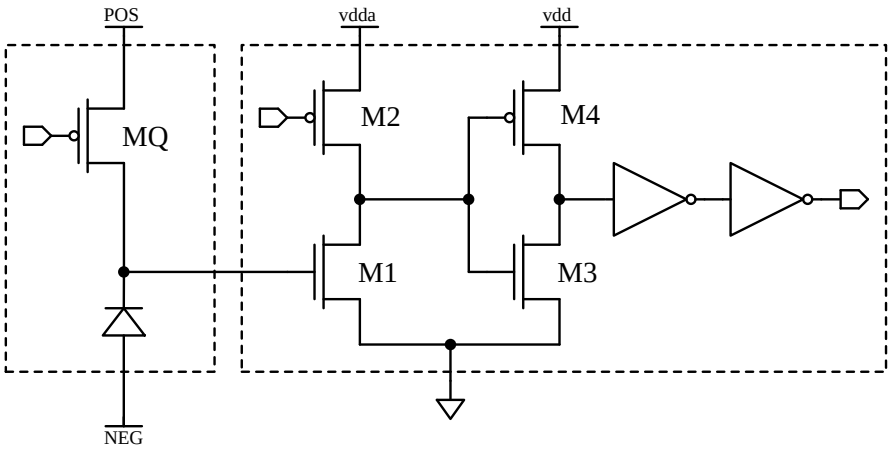


Figure 2.5: The schematic diagram of the implemented SPAD circuit

occurred. The POS terminal is connected to 1.0V, and the NEG terminal is connected to -10.0V volts. In idle mode, the output voltage is 1.0V and the over voltage is $V_{OV} = 1 + 10 - 10.4 = 0.7V$. The output of the SPAD circuit is connected to the oscilloscope by an active probe Agilent N2795A, which has a $1M\Omega$ input impedance and a 1pF input capacitance. When an avalanche occurs the output voltage drops from 1.0V to 0.4V which is the breakdown voltage of the avalanche diode relative to the ground of the circuit. The total output capacitance of the circuit which includes the capacitance of the diode and parasitic capacitance of the package is calculated from the rise time (10%-90%) in Equation 2.3:

$$C = \frac{t_r}{2.2 \times R} = \frac{0.95\mu s}{2.2 \times 27.2k\Omega} = 16pF \tag{2.4}$$

The parasitic capacitance of the package is expected to be significantly larger than the capacitance of the avalanche diode. As a result, slower recovery times are achieved when the external quenching method is used.

2.3.2 Internal quenching

Figure 2.5 shows the implemented circuit which consists of two parts: the SPAD handling circuit, and the readout circuit. The diode handling circuit quenches and recovers the SPAD upon each avalanche event. The readout circuit ensures a non-destructive voltage readout of the diode's node. The internal SPAD circuit has the benefit of integration and faster recovery times due to a lower parasitic capacitance. The diode handling circuit includes a pMOS transistor M_Q that has a controllable gate

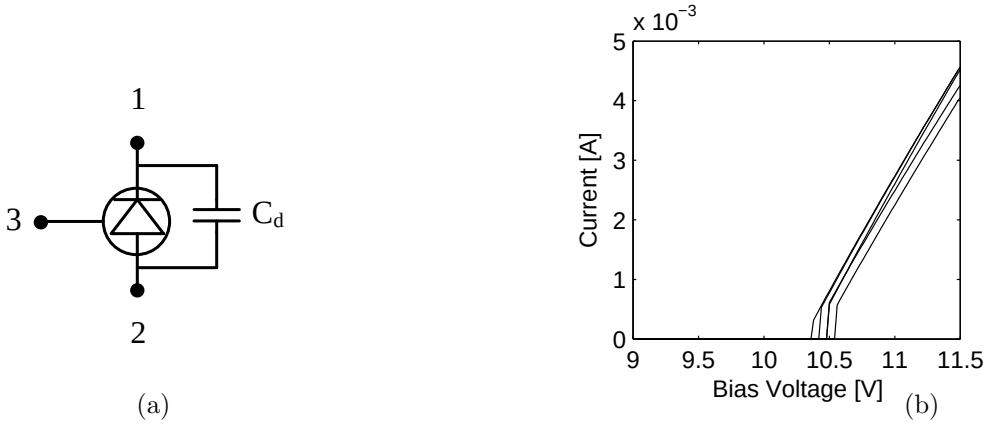
Device		1.8V	3.3V	W/L (μm)
M_Q			×	0.5 / 5
M1-M4			×	0.24 / 0.34
Inv1,2	pMOS	×		0.72 / 0.18
	nMOS	×		0.24 / 0.18

Table 2.2: The transistor dimensions for the implemented circuit in Figure 2.5

bias voltage. Similar to the resistor in Figure 2.4.a, this transistor provides the means of both quenching and recovery processes. In an idle mode, the diode is biased over its breakdown voltage by setting appropriate voltages to POS and NEG terminals. The POS terminal is connected to a positive power supply, and NEG is connected to a negative power supply. Once an avalanche event occurs, the capacitance is rapidly discharged by the avalanche current. The current limited transistor M_Q limits the amount of charge following through the pn junction. As a result, the voltage across the capacitance drops until the avalanche current is quenched. Once the avalanche process is quenched, the diode capacitance recharges to its initial bias voltage, V_A, by the current of M_Q. Therefore, the diode is recovered from an avalanche event, and ready for the detection of the next photon.

The voltage fall and rise caused by the quenching and recovery process is an indication of detection of an avalanche event. However, unlike SPAD circuit in the Figure 2.4.a, the SPAD node is not directly connected to the output of chip. A digital inverter is used to readout the SPAD node. As a result, a digital pulse at the output of the circuit indicates the detection of an avalanche effect. The implemented readout circuit includes a variable load inverter, and series of digital push pull inverters. The variable load inverter, M1 and M2, has a variable threshold voltage which is set by an external bias voltage connected to the gate of M1. An appropriate combination of the voltages of POS, NEG, and the gate of M2 ensures that the voltage swing at the diode's cathode is large enough to create a digital pulse at the output of the variable load inverter. In the next stage, a series of digital push pull inverters create a sharp digital output pulse. This digital pulse is connected to the digital output buffers which drive the parasitic capacitance of the output pin.

The speed of recovery is controlled by limiting the current of M_Q. A lower gate bias voltage of M_Q provides a higher current which results in a faster recovery time. However, as M_Q provides the means of both quenching and recovery processes, a higher current of M_Q may not provide enough voltage drop to quench the avalanche process, hence leading to a sustained avalanche current.



(a) (b)

Figure 2.6: The model for the avalanche diode (a) the schematic diagram an avalanche diode (b) The I-V characteristics of the measured avalanched diodes

Table 2.2, shows the transistor dimensions used in the implemented circuit shown in Figure 2.5. The transistors M_Q and M1 to M4 are 3.3V thick gate transistors, which have higher gate voltage tolerance than the nominal 1.8V transistors. The quenching transistor M_Q has a lower W/L ratio to ensure high resistivity in order to create a large enough voltage drop to quench the avalanche current. The inverters I1 and I2 use the nominal 1.8V transistors and have the minimum sized length.

2.3.3 A model for avalanche diode

In order to perform a simulation of the SPAD circuit, a behavioural model for the avalanche diode is developed in Verilog-A. Figure 2.6.a shows the schematic of a three terminal diode model. Terminal 1 and 2 are the cathode and anode of the avalanche diode respectively. Terminal 3 is the trigger input which emulates the start of an avalanche event. A breakdown voltage of 10.4V is set for this model based on I-V characteristics shown in Figure 3. The over voltage of the avalanche diode is:

$$V_{OV} = V_1 - V_2 - V_{br} \tag{2.5}$$

where V_1 is the voltage of terminal 1 which is connected to a positive voltage, V_2 is the voltage of terminal 2 which is connected to a negative voltage, and V_{br} is the breakdown voltage. The capacitor C_d, which is in parallel to the diode, represents the capacitance of the avalanche diode. In idle mode, C_d is fully charged and no current flows between terminal 1 and 2. In order, to trigger an avalanche event, the voltage of terminal 3 should rise beyond the threshold voltage V_{th}. Once an avalanche is triggered, the current of the diode between terminal 1 and 2 is:

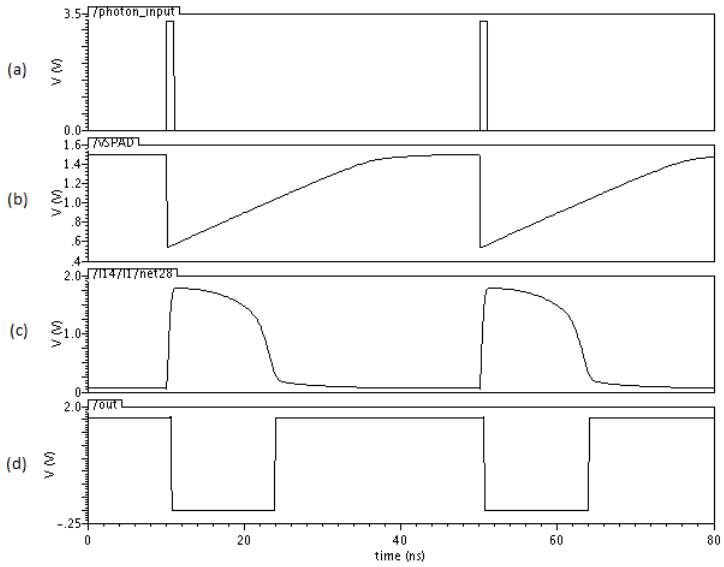


Figure 2.7: The simulation results for the implemented SPAD circuit (a) trigger terminal of the avalanche diode model (b) the diode's cathode node (c) the output of the variable load inverter (d) the final output of the SPAD circuit after the push-pull inverters

$$I_d = V_{OV}/R_{sc} \tag{2.6}$$

where V_{OV} is the over voltage, and R_{sc} is the space charge resistance of the pn junction. As a result, the capacitance C_d is discharged through the space charge resistance R_{sc}. The avalanche current is quenched once $V_1 < V_2 + V_{br}$. Hence, the diode's bias voltage is smaller than the breakdown voltage. When the avalanche is stopped it is assumed that $I_d = 0$. Figure 2.6.b shows the I-V characteristics of the measured avalanche diodes. The I-V curve shows that the diode starts to conduct after its bias voltage is beyond V_{br}. The conductivity slope is equal to the inverse of the space charge resistance R_{sc} which is approximately 200 Ω.

2.3.4 Simulation of the implement circuit

Figure 2.7 shows simulation results for the implemented SPAD circuit. The breakdown voltage of the avalanche diode in the behavioural model is set to 10.4V . The voltage of POS terminal is 1.5V and the voltage at the NEG terminal is -9.9V. Hence, the over voltage is $V_{OV} = V_{POS} - V_{NEG} - V_{br} = 1.5 + 9.9 - 10.4 = 1.0V$. By applying an appropriate bias voltage, the threshold voltage on the variable load inverter is set

to 1.0V, which is in the middle of the voltage swing of the diode's cathode. The supply voltage for both *vdda* and *vddd* terminals are 1.8V. In Figure 2.7-a, a pulse is given to the trigger terminal of the avalanche diode to indicate the start of an avalanche process. The avalanche is triggered on the rising edge of the input trigger pulse. In Figure 2.7-b, the avalanche current rapidly discharges the capacitance of the diode. Once the bias voltage of the diode is close to the breakdown voltage, the avalanche current is stopped, and the capacitance is recharged to its initial bias voltage. The current limited transistor M_Q recharges the capacitance; hence, the SPAD is recovered.

In this simulation, the recovery time is 25 ns. Figure 2.7-c shows that the variable load inverter amplifies and limits the input signal which creates a digital pulse at the output. Figure 2.7-d, shows that the push pull inverters create a sharper output signal. This signal is then connected to the output buffers which drive the output pins. The signal observed at the output of the chip is a digital signal similar to Figure 2.7.d. Although the recovery time is 25 ns, the pulse duration of the output signal is approximately 14 ns which is shorter than recovery time. This is because when cathode voltage rises beyond the threshold voltage of the variable load inverter, the output of this inverter changes from the high voltage to the low voltage, however, the recovery is still in progress.

An off-chip pulse counter is used to count the rate of output pulses of the SPAD circuit. The pulse count rate represents the rate of the detected avalanche events. The pulse counter is triggered at the falling edge of each output pulse as the falling edge indicates a detection of an avalanche event.

In high detection rates, there is a high probability that an avalanche occurs before the previous avalanche is fully recovered. Figure 2.8 shows an example of an avalanche event that has occurred 10 ns after the first event. The bias conditions are identical to simulation in Figure 2.7. While the diode capacitance is charging to its initial bias voltage, an avalanche event occurs which discharges the capacitance rapidly to the breakdown voltage. Once the avalanche is quenched, the capacitance is recovered to its initial voltage after 25 ns. In this example, the second avalanche occurs at the time when the voltage of the diode's node has not risen beyond the threshold voltage of the variable load invert, hence, the state of the output voltage does not changed, and it is maintained at low voltage.

Figure 2.8-d shows that the state of the output has not changed, and the output pulse duration is extended from 14 ns to 24 ns. Although two avalanche events have occurred in sequence, the off chip pulse counter only counts one pulse, hence, it has

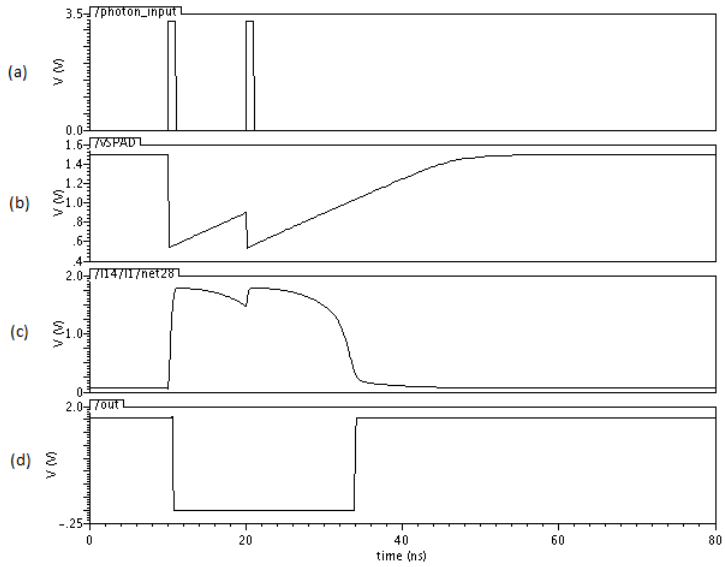


Figure 2.8: The simulation results for the SPAD circuit when an avalanche event occurs before the previous event has fully recovered (a) input terminal of the avalanche diode's model (b) the diode's cathode node (c) the output of the variable load inverter (d) the output of the circuit after the pair of digital inverters

only detected the first avalanche event and the second avalanche event is missed. As a result, the pulse counting statistics is distorted, and the rate of pulse count is no longer proportional to avalanche events especially at higher photon rate. Therefore, it is important for the characterisation of the SPAD circuit that the average time interval between each avalanche event is significantly larger than the recovery time of the SPAD circuit.

2.4 Characterisation of the SPAD circuit

The SPAD circuit is implemented in UMC 0.18 μm commercially available CMOS process. Four packaged chips are initially used in the characterisation of the SPAD circuits. These chips are numbered and referred to as C1 to C4 within this chapter.

2.4.1 Dark count rate

Figure 2.9 shows the dark count rate of four measured SPADs (C1-C4) versus the bias voltage using the pulse counting method. The voltage of the POS terminal (in Figure 2.5) is 1.25V and the voltage of the NEG terminal is varied from -9.1V to 10.8V. The voltage of the POS terminal is 100mV above the threshold voltage of the

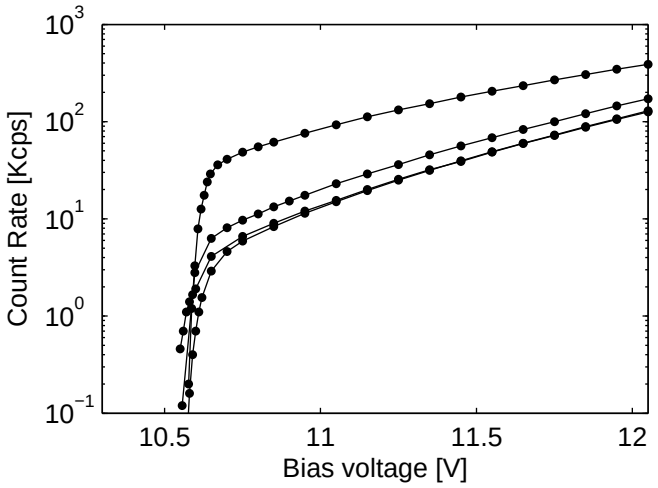
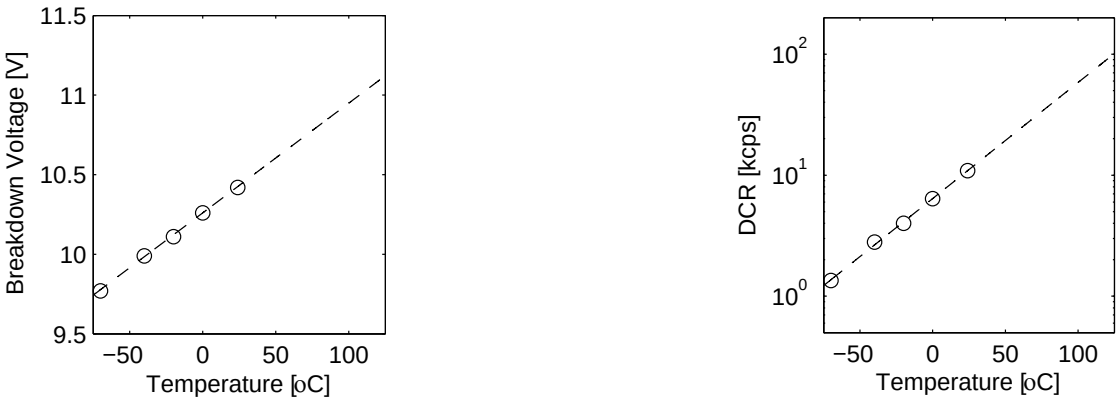


Figure 2.9: The dark count rate versus bias voltage for the SPAD circuits fabricated in chips C1-C4

variable load inverter. The gate voltage of M_Q is set to ensure that output pulse duration of 20 ns is achieved. Increasing the bias voltage while maintaining the gate voltage of M_Q results in a longer recovery time, hence, longer output pulse duration. This is because the diode's capacitance requires longer time to recharge to its initial bias voltage during an avalanche event. In order to maintain the recovery time, the gate voltage of M_Q is decreased when the bias voltage is increased. This ensures that the recovery time is always constant. Due to a relatively small recovery time, the probability of extended recovery time is negligible. The pulse rate appears to start approximately at 10.6V. Among the four tested SPAD circuits, C2 has a considerably higher DCR than the other measured chips. The higher DCR could be due to a large number of defects within the active area of the SPAD. If the over voltage is smaller than 100mV, no output pulse is detected as the voltage swing of the capacitance node is not large enough to create a digital pulse at the output of the variable load inverter.

2.4.2 SPAD versus temperature variations

The breakdown voltage of an avalanche diode and the DCR depend on the temperature variations [39, 15]. In order to investigate the effect of temperate on the SPAD operation, a sample chip was cooled in a climatic chamber, which was capable of cooling down to $-70°C$. Due to limited availability of equipment during this experiment, the avalanche diode was externally quenched with an off chip resistor shown in Figure 2.4. The entire chip, printed circuit board, and oscilloscope probes were placed into the air-sealed chamber.



(a) (b)

Figure 2.10: The effect of temperature on (a) Breakdown voltage and (b) DCR. The circles are experimental data points and the dashed line is a numerical fit based on experimental data

The breakdown voltage and DCR were calculated for selected temperatures. After changing the temperature of the chamber, enough time was allowed to ensure the temperature is stabilised. In the room temperature, the breakdown voltage is 10.4V, and with an over voltage of 0.7V the DCR is approximately 10 kcps. As the temperature decreases, the breakdown voltage is expected to decrease proportionally to the temperature [39, 15]. The breakdown voltage is the lowest voltage of the diode's node during an avalanche event. In addition, as the breakdown voltage decreases, the over voltage increases at lower temperatures. As a result, the negative supply voltage is decreased to ensure that the over voltage is always maintained at 0.7V during this experiment.

Figure 2.10.a shows the breakdown voltage versus temperature. A line is fitted to the experimental data points. As expected the breakdown voltage is proportional to the temperature of the SPAD. The slope of the fitted line is 6.9 $mV/^oC$ which is close to the reported temperature dependence of the avalanche diode in similar processes [15, 39]. Figure 2.10.b shows the DCR of the SPAD versus temperature. An exponential curve is fitted to the experimental data. As expected, the DCR increases exponentially with the temperature. The ratio of the dark count increases is 0.0095 decade count per ^oC which is close to the previously reported increase ratio[39, 15].

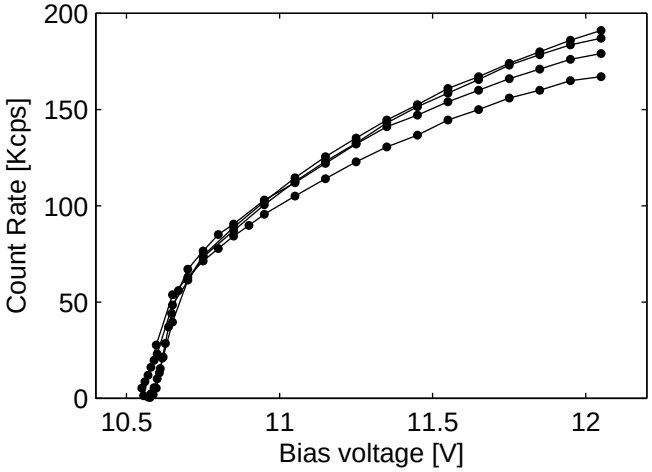


Figure 2.11: The count rate of the SPAD when illuminated with an LED versus the bias voltage

2.4.3 Relative and absolute PDP

The light response of the SPAD circuit is measured for C1 to C4 by the pulse counting method similar to section 2.4.1. In order to provide a stable and reliable light source for this experiment, a diffused epoxy blue light emitting diode (LED) with a peak wavelength of 460 nm and a full width half maximum (FWHM) of 25 nm is used. The LED is driven by a constant current source which has the advantage of providing a stable intensity against voltage variations due to temperature fluctuations. The chip and the printed circuit board were placed into an enclosed metal box to prevent any background illumination. The pulse count rate includes both the pulse count rate related to the LED illumination, and the dark count rate. Hence, the dark count shown in Figure 2.9 is subtracted from the measured count rate in this experiment. The resulting count rate is only related to the LED's illumination. Figure 2.11 shows the current versus bias voltage. These results show that the illumination count rate is relatively constant among the tested SPADs. The sample chip C2, which had a higher DCR in Figure 2.9, has an illumination count rate similar to the rest of the measured chips. This suggests all the measured chips have similar strength in their electrical field. Hence, the higher DCR in C2 is potentially due to larger number of defects within the active area of the avalanche diode.

2.4.3.1 Calculation of PDP

If the absolute photon flux on the surface of the detector is known, the PDP of the detector for the particular wavelength of the LED can be calculated. A calibrated

optical power meter is used to measure the power density, hence, the photon flux at the surface of the chip. The LED illumination generates approximately 1.0 Mph/s (ph/s = photon per second) per active area of the avalanche diode, which has a $10\mu m$ diameter. As shown in Figure 2.11, at the maximum bias voltage of 12.05V, the average count rate for the five measured chips is approximately 180 kcps. The PDP is calculated by the ratio of the pulse count rate to the number of photon arriving at the surface area of the detector. Therefore, the PDP is equal to $180(kcps)/1.0(Mph/s) = 0.18$, at wavelength of 460 nm, and bias voltage of 12.05V.

2.4.3.2 PDP versus wavelength

The PDP is a function of wavelength because the efficiency of the photon absorption in silicon is dependent on the energy of the photon. In order to measure the PDP for various wavelengths, a monochromator with a spectral bandwidth of 10 nm FWHM is used as a source of illumination for the SPAD. The monochromator is illuminated with a tungsten light bulb driven by a constant current source, which maintains the spectrum of the light bulb during temperature variations. The grating within the monochromator has a transmission efficiency that is dependent on the wavelength. As a result, the monochromator is calibrated for every 10 nm step within the spectral range of 400 nm to 900 nm by a coefficient which is a function of the selected wavelength. As the measured count rate for each wavelength is relative to the count rate in another wavelength, the absolute PDP can be estimated when the PDP is known for a particular wavelength. Hence, the absolute PDP is calculated for the range of 400 nm to 900 nm as PDP for 460 nm is known from the previous experiment.

Figure 2.12 shows the absolute PDP of the SPAD circuit versus wavelength. This result shows that the SPAD response has a peak PDP at 490nm. The PDP is higher in the shorter wavelengths than the longer wavelengths. This is expected as the pn junction of the avalanche diode has a relatively shallow depth. The sharp variations in the PDP are due to multiple reflection layers within the oxide stack on top of the silicon substrate which leads to optical interference [72].

2.4.4 Current measurement method

An alternative method to pulse counting is to measure the steady state current which flows through the quenching transistor M_Q and the diode. The use of steady state current measurements eliminates the need of the readout mechanism, which included the variable load inverter and digital push pull inverters. Hence, M_Q is the only

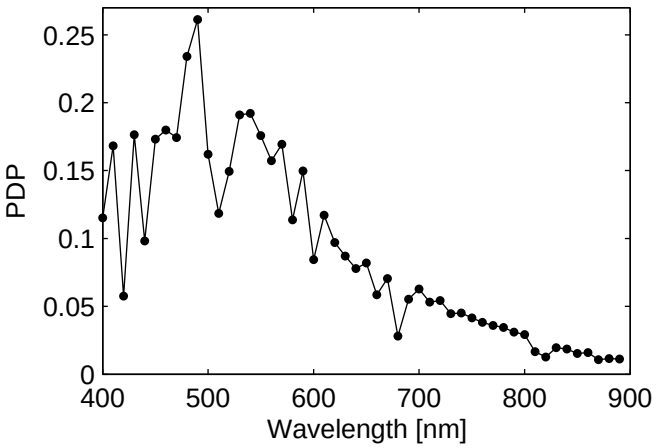


Figure 2.12: The absolute PDP of the SPAD circuit versus wavelength at every 10 nm steps

device in the current path of the diode. As a result, the diode can be biased at bias voltages which are close the breakdown voltage as a threshold voltage is no longer required. Moreover, the current measurement method is more flexible and reliable as it is automated using the provided software drivers for the current measurement equipment. An accurate current meter (Keithley 2400) is connected to POS and NEG terminals of the SPAD circuit shown in Figure 2.5. The measured current is equal to the amount of charge accumulated on the diode's capacitance during the recovery time. Each avalanche event discharges the capacitance, hence

$$I_d = C_d \times V_{OV} \times R \tag{2.7}$$

where C_d is the capacitance of the diode, V_{OV} is the over-voltage, and R is the rate of detected avalanche events.

Figure 2.13 shows the measured dark current of the avalanche diode, operating in SPAD mode, versus bias voltage for C3. In order to assess the validity of this experiment, the current of the diode is estimated by Equation 2.7 and compared to the measured current. The capacitance of the avalanche diode in Equation 2.7 is extracted by comparing the measured current data to the pulse count data shown in Figure 2.9. As a results, the estimated capacitance for C3 is $C_d = 250fF$. The circle data points in Figure 2.13 shows the estimated SPAD current using Equation 2.7 when C_d is 250 fF. As it can be observed, both data sets are in good agreement when the bias voltage is larger than 10.7V, which is 300mV above the breakdown voltage. The data mismatch for bias voltages smaller than 10.7V shows the effect

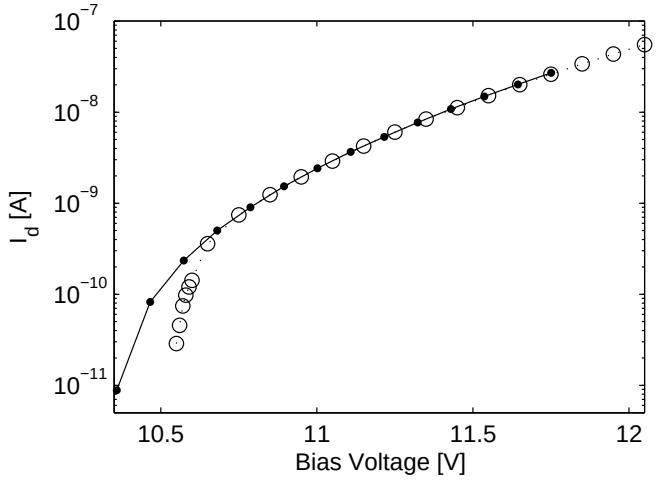


Figure 2.13: The measured dark current (dots) and estimated dark current (circles)

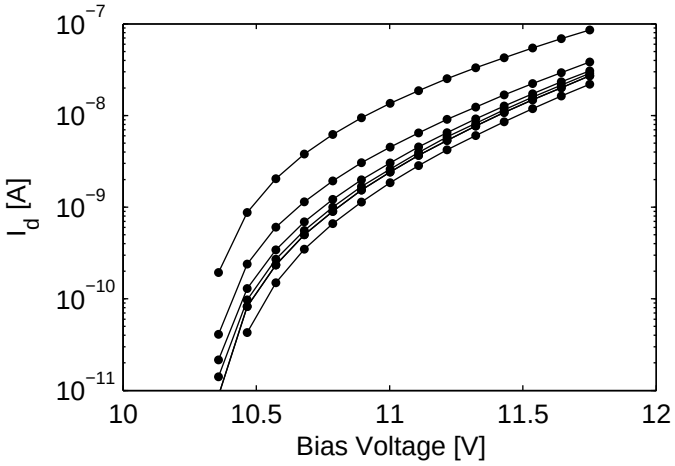


Figure 2.14: The dark current of the avalanche diode in the SPAD mode versus bias voltage. The current is measured for seven chips (C1-C7)

of the threshold voltage in the readout circuit of the SPAD, as small bias voltages produce lower voltage swings at the input of the variable load inverter. The lower voltage swings limit the capability of the readout circuit to characterise the SPAD at low bias voltages.

Figure 2.14 shows the steady state dark current of the avalanche diodes in the SPAD mode versus bias voltage. In addition to the 4 chips which were tested in section 2.4.1, two chips, C5 and C6 were added to this measurement. Similar to the DCR shown in Figure 2.9, C2 has a higher DCR than the rest of the measured chips. As expected the breakdown voltage of the SPADs are approximately 10.4V, where the current sharply increases. Below this voltage the avalanche diode is operating

	C1	C2	C3	C4
C_d	220fF	230fF	250fF	205fF

Table 2.3: The estimated capacitances of the avalanche diodes based on Equation 2.7

in APD mode. Comparing the steady state current measurements from Figure 2.14 with pulse counting measurements from Figure 2.9 shows the effect of readout circuit on the pulse counting due to insufficient voltage swing at the input of the variable load inverter.

Table 2.3 shows the estimated capacitances for samples C1 to C4 calculated using Equation 2.7. The variations in the capacitances could be due to the variability in the doping concentrations and physical formation of the pn junction. The estimated capacitance for C5 and C6 are not available as their pulse counting data is not available in Figure 2.9. It can be noticed that the capacitance of the avalanche diode estimated by the current measurement method is significantly lower than the total capacitance estimated by the rite time of the SPAD's output in section 2.3.1.

2.5 Summary

In this chapter, a 10 μm avalanche diode was fabricated in a commercially available CMOS process. The diode uses a circular guard ring in order to prevent premature breakdown due to high electrical filed in the corners and the edges of the pn junction. The fabricated avalanche diode has a breakdown voltage of approximately 10.4V, and the very sharp increase in the avalanche current makes the diode unsuitable to be operated in APD mode. However, they can be reliably operated in SPAD mode as the variation in the breakdown voltage among different measured diodes are relatively low. This is especially advantageous when an array of SPADs share the same bias voltage.

In order to operate the avalanche diode in the SPAD mode, the avalanche current should be quenched. The quenching process can be implemented with an external resistor, however, the large parasitic capacitance and lack of integration are major disadvantages of the external quenching method. The integrated SPAD circuit, provides the means of quenching and recovery by a current limited transistor. The speed of recovery is set by gate voltage of this transistor. The readout section of the SPAD circuit provides a non-destructive method of reading the diode's node. The output the SPAD circuit is a digital pulse, and the rate of these pulses is indicated the avalanche events.

The fabricated SPAD circuit has been characterised for major performance parameters. The dark count rate (DCR) of the SPAD increase with the bias voltage. It is possible that the DCR of an individual SPAD is significantly higher than the rest of the fabricated diodes. The breakdown voltage increases proportionally with temperature, and DCR increases exponentially with the temperature. The photon detection probability (PDP) is a function of bias voltage and wavelength. The PDP increases with the bias voltage. It has a maximum peak of 26% at 490 nm, and is generally higher in shorter wavelengths and decreases in longer wavelengths.

The current measurement method eliminates the need for the readout circuit, and enable characterising the SPAD at lower bias voltages. The capacitance of the avalanche diode is extracted by comparing the current of the avalanche diode in the SPAD mode to the pulse counting data. This capacitance is considerably lower than the total capacitance of the external quenching circuit using the off-chip resistor.

Although the implemented SPAD circuit in this chapter, is integrated with the avalanche diode, and can achieve relatively fast recovery times in comparison the externally quenched SPAD, the recovery speed and the extended pulse duration are obstacles to a high-speed communication link. Further modification are required to improve the speed and functionality of the SPAD circuit.

Chapter 3

Automatic Diode Reset

3.1 Introduction

In the previous chapter a circuit which enables the avalanche diode to be operated in the SPAD mode was presented. Although the circuit is effective for characterisation of the avalanche diode, it is not suitable for a high-speed performance, due to a long recovery time. The dependency of the quenching and the recovery processes in the circuit shown in Figure 2.5, limits the minimum achievable recovery time. In order to achieve a shorter recovery time, a higher recharging current is required which simultaneously leads to a lower quenching resistance. The lower quenching resistance, may not be high enough to quench the avalanche current, hence leading to a distorted SPAD operation. Moreover, the extended recovery time which is due to proximity of two avalanche events, distorts the photon counting statistics.

The automatic diode reset (ADR) circuit provides a high performance SPAD operation by separating the quenching and recovery into two independent process. Dedicated circuits disable the SPAD upon detection of an avalanche event, and enable it after a specified duration. The independence of the recovery process form quenching process allows the user to choose shorter recovery times than the circuit used for the characterisation of the SPAD.

3.2 The concept

In the automatic diode reset the avalanche diode is biased over its breakdown voltage. This bias voltage is equal to $V_A = V_{br} + V_{OV}$, where V_{br} is the breakdown voltage of the avalanche diode, and V_{OV} is the over voltage. Figure 3.1 shows a concept diagram for

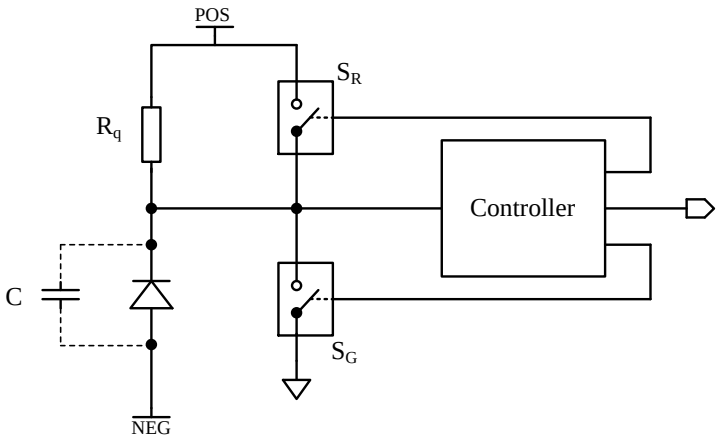


Figure 3.1: The conceptual digram of an automatic diode reset

an automatic diode reset. The diode is biased by a positive POS and a negative NEG voltage terminals. The capacitance on the diode's node is represented by a dotted capacitor. This capacitance C includes the capacitance of the diode, and the parasitic capacitance of the peripheral components connected to that node. The resistor R_q represents the quenching device which could be a resistor or an active component. The ideal switch S_G, connects the diode's node to ground, and S_R resets the diode by connecting the diode's node to the positive voltage terminal POS. The diode's node is connected to the input of a controller, which assigns appropriate timings for the switches S_G and S_R.

When an avalanche event is occurs, the diode is quenched by R_q, hence the diode's node voltage drops, and the capacitor C starts to discharge. The controller senses the voltage drop, and immediately afterwards S_G conducts to speed up the discharging of the capacitor. Once C is fully discharged, the diode's node is held at ground for a specified period. The diode is biased to ensure that its breakdown voltage is above the ground level of the circuit, hence by grounding the diode's node, the bias voltage of the diode is below its breakdown voltage. During this period, the SPAD is inactive and no avalanche event occurs. Hence, this period of inactivity is referred to as deadtime. In order to reset the diode to its initial bias voltage after deadtime has elapsed, the controller changes S_G to a non-conducting state and S_R to a conducting state. Once the capacitor C is charged, the controller changes S_R to a non-conducting state, and the SPAD is ready for the detection of the next avalanche event. This process repeats for every avalanche event. The output of the controller can be selected form one or a combination of control signals.

In addition, it is desired that the S_R switch only conducts for a short period of

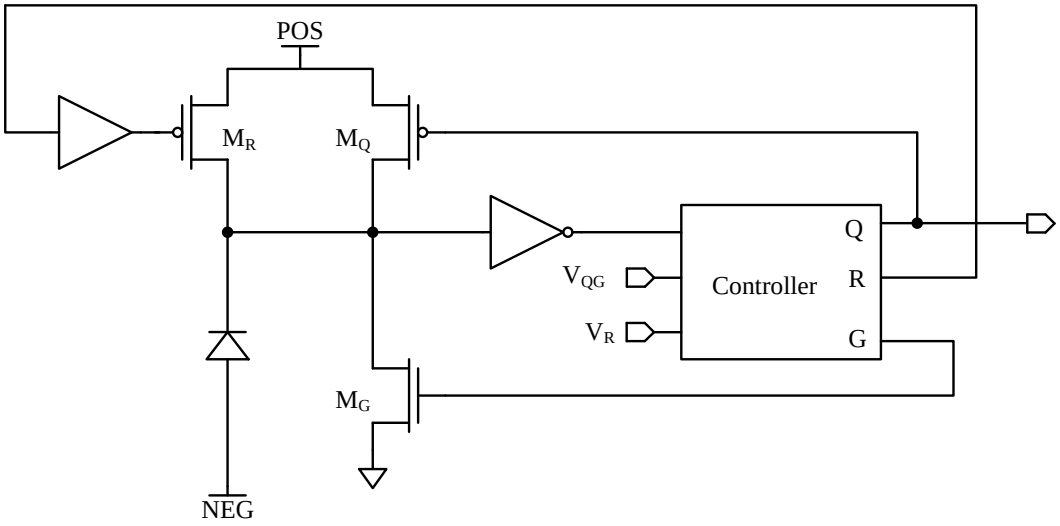


Figure 3.2: The transistor level implementation of the automatic diode reset circuit

time so that the probability of an avalanche event is negligible. If an avalanche event occurs during the recharging period, there is a probability of a sustained avalanche current due to a low resistance of the switch.

3.3 The implemented circuit

Figure 3.2 shows the transistor level implementation of automatic diode reset. The controller provides three individual control signals for the transistors. The input bias voltages V_{QG} and V_R set the appropriate timing for the transistor switches. The output of the circuit is the control signal Q. The quenching transistors is replaced with the transistor M_q, the ground switch is replaced with M_G, and the reset switch is replaced is M_R. In this implementation, the quenching transistor is disabled by connecting the inverted control signal G to the M_q. As a result, the diode is quenched using the high impedance node when none of the transistors are conducting. When the diode is charged, both M_G and M_R are not conducting, and the diode's node has a high impedance. When an avalanche event occurs, the diode is discharged by the avalanche current, and its voltage drops close to the breakdown voltage, hence, the avalanche current is quenched. The controller changes both M_q and M_G to a conducting state.

Name	type	dimensions W/L (μm)
M_R	pMOS	6/0.18
M_Q	pMOS	0.24/1
M_G	nMOS	0.24/0.18
Buffer	nMOS	0.24/0.18
	pMOS	0.34/0.18
NOT	nMOS	0.24/0.18
	pMOS	0.35/0.18

Table 3.1: The transistor dimension's for the automatic diode reset shown in Figure 3.2.

Name	Area (m^2)	Capacitance
M_R	9.94×10^{-12}	11.3 fF
M_Q	2.22×10^{-13}	0.23 fF
M_G	2.22×10^{-13}	0.23 fF
diode	7.85×10^{-11}	≈225 fF

Table 3.2: The area and capacitance of transistors connected to the diode's node

3.3.1 Dimension of the Transistors

In order to achieve highest quenching and recovery speeds, it is important to minimise the parasitic capacitance on the diode's node. The total capacitance of on the diode's node is:

$$C_{tot} = C_d + C_p = C_d + C_R + C_Q + C_G + C_{inv} + C_{wire} \tag{3.1}$$

where C_{tot} is the total capacitance of the diode's node including capacitance of the diode's pn junction C_d, and the parasitic capacitance C_p which includes, source-drain capacitance of the three transistors C_R, C_Q,C_G, the gate capacitance of the input inverter to the controller C_{inv}, and the parasitic capacitance of wires and interconnections C_{wire}.

Table 3.1 shows the transistor dimensions for the implemented automatic diode reset circuit shown in Figure 3.2. The reset transistor M_R should be wide enough to recharge the total capacitance as quickly as possible. Hence, the width of M_R is selected to ensure that the its drain capacitance is a fraction of the C_d. The grounding transistor M_G is a minimum sized, and M_q should be resistive enough so that when both M_G and M_q are conducting the voltage on he diode's node is below the breakdown voltage. As a result, M_q is 5.5 times longer than M_G, so that when both are conducting, the diode's node voltage is 0.25V.

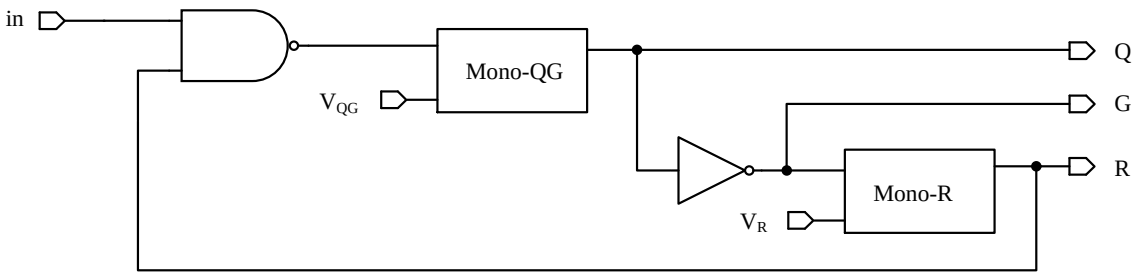


Figure 3.3: The schematic diagram of the controller used in the implemented automatic diode reset circuit

Table 3.2 shows the drain junction area and the capacitance of each transistor which are calculated using the process parameters provided by the foundry. Assuming thatC_{wire} is negligible, C_{inv} is approximately 1fF, The dominant parasitic capacitance on the diode's node apart from the diode itself is the drain capacitance of M_R which is significantly smaller than the diode's capacitance estimated in Table 2.3

3.3.2 Controller

The controller sets the appropriate signalling for the transistors M_R, M_q, and M_G. Figure 3.3 shows the schematic diagram of implemented controller. The controller consist of two sequential monostables which provide the timing signals to control the quenching Q, grounding G, and reset R signals. A monostable is a timer circuit, which upon receiving a trigger, changes the state of its output for a specified period of time, hence, it always has only one stable state. The input of the controller is connected to the diode's node via a digital inverter. The inverter ensures that the input of the controller has sharp transitions to provide a precise timing operation.

In this implementation, the monostable is configured as an active low, which means that in the idle mode the output of the monostable is a logic 1. Upon receiving a trigger, the output changes from logic 1 to logic 0 for a specified duration, which is determined by the input bias voltage of the monostable circuit. The NAND gate provides a recovery mechanism by creating a feedback loop from the reset signal R to the input. The recovery mechanism attempts to recharge the avalanche diode if it is has been failed to charge for any unexpected reason in the previous avalanche event.

The operation sequences of the active quenching circuit are: the idle phase, the QG trigger phase, the reset phase, and the recovery phase. The two sequential mono-

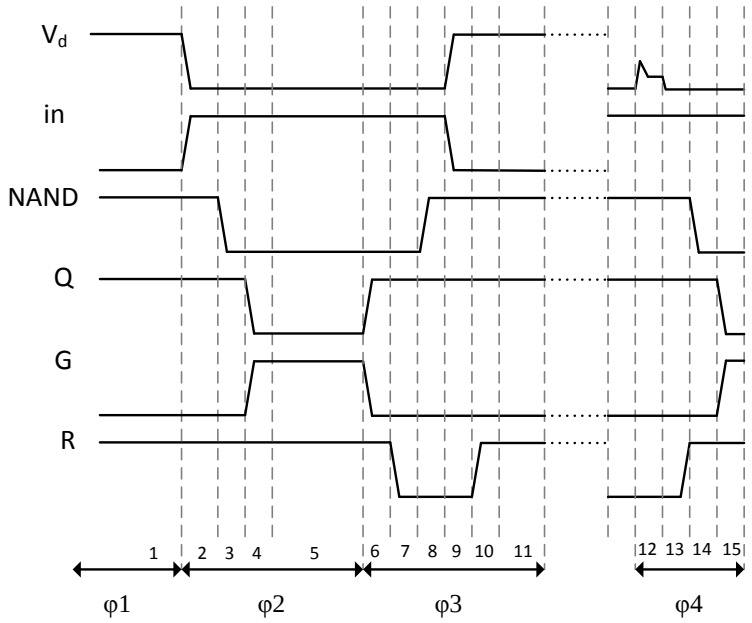


Figure 3.4: The timing diagram for the controller circuit shown in Figure 3.3

stable configuration ensures that Mono-R is triggered immediately after Mono-QG has returned back to its stable state. Figure 3.4 shows the four stages of operational sequence of the controller and its detailed transitions.

$\phi 1$: The idle phase

In the idle phase, the diode's node V_d is charged, hence the input of the inverter in is a logic 0 (stage 1). When the input of the NAND gate is 1, regardless of the second input of NAND gate its output is 1. The input of the Mono-QG is a 1, and since it is in a stable state, its output is a 1. Hence, Q is 1, and G is 0 which mean both M_q and M_G are not conducting. Although the input of Mono-R is 0, its output is in a stable state, hence it is 1. The monostable returns back to its stable state after a specified delay regardless of the state of its input. As a result, the output of Mono-R, R is 1, which means the reset transistor M_R is not conducting.

$\phi 2$: The QG trigger phase

When an avalanche events occurs, the input of the controller in changes from 0 to1 (stage 2), and it remains at 0 until it is reset back by M_R in the stage 9. Since R is a 1 from the idle phase, the output of the NAND, which is the input of Mono-QG, is $1\,NAND\,1 = 0$ (stage 3). As Mono-QG is triggered by a 1 to 0 transition, the control

signal Q changes to 1, and G changes to 0. As a result, M_q and M_G are conducting, and the V_d is maintained at 0.25V (stage 4). The bias voltage of avalanche diode is configured to ensure that its breakdown voltage is below 0.25V, hence the SPAD is inactive when the transistor pair M_q and M_G are conducting. As the signal G changes to 1, the input of Mono-R is 1, which does not make any change to its output, R, since the monostable is only triggered on falling edges (stage 5). The last transition state is maintained until the delay set by the bias voltage V_{QG} is passed. This delay defines the deadtime of the automatic diode reset.

$\phi 3$: The reset trigger phase

After the deadtime is elapsed, the output of Mono-QG, Q, returns back to its stable state 1, and G changes to 0 (stage 6). As a result, both M_q and M_G are not conducting. The input of Mono-R, G, has a 1 to 0 transition which triggers this monostable. Once Mono-R is triggered, R changes to 0 (stage 7). As the input of the controller is still 0, output of the NAND changes to 1 $NAND\, 0 = 1$ (stage 8). At this point the reset transistor M_R is conducting to recharge the diode, hence, the input of the controller returns *in* back to 0 (stage 9). After a short delay specified by the bias voltage V_R, the signal R returns back to 1 (stage 10), hence M_R is not conducting, the diode's node is back to high impedance quenching, and ready for the detection of the next avalanche event (stage 11).

$\phi 4$: The recovery phase

If the SPAD fails to recover for any unexpected reason in stage 9, the controller attempts to recharge the SPAD in a cycle which follows immediately afterwards. The main reason for a failed recovery is a detection of an avalanche event when M_R is conducting. The low resistance of M_R could lead to a sustained avalanche current. In Figure 3.4, an avalanche event is detected during the reset stage (stage 12), Hence, both the avalanche diode and M_R are conducting. In this case the space-charge resistance of the avalanche diode is smaller than resistance of M_R. As a result the voltage of the diode's node V_d remains below the threshold of the input inverter, hence the output of the inverter is remained unchanged at logic 1, and since the signal R is 0, the output of the NAND is maintained at 1 $NAND\, 0 = 1$.

Once the Mono-R timer is elapsed, the signal R returns to 1 (stage 13), hence the output of the NAND is 1 $NAND\, 1 = 0$ (s14). This 1 to 0 transition triggers Mono-QG (stage 15), which is a similar condition as described in stage 6. The rest

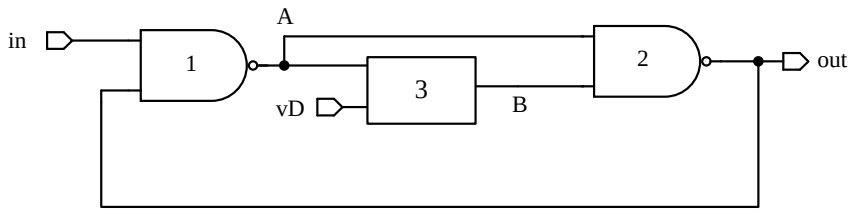


Figure 3.5: The schematic diagram of the monostable circuit implemented in the automatic diode reset

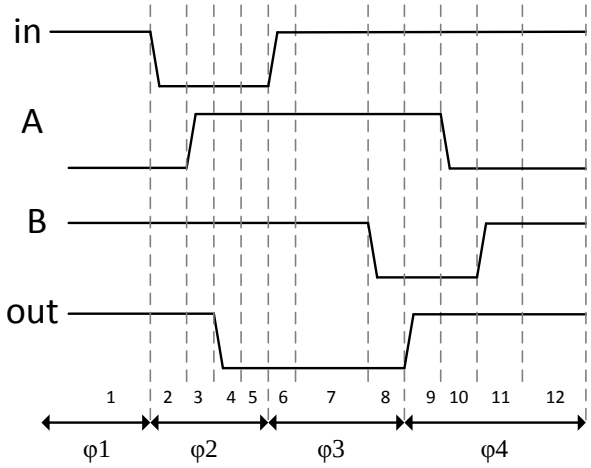


Figure 3.6: The timing diagram for the monostable operation sequence

of the signalling operation continues from stage 7 until the diode is recovered.

3.3.3 Monostable

A monostable is a timer circuit, which upon a transition at its input changes the state of its output for a specified duration, hence, it always has only one stable state. The monostable returns back to its stable state after the specified delay regardless of the state of its input. Figure 3.5 shows the schematic diagram of the monostable which is implemented in the controller shown in Figure 3.3. Components NAND-1 and NAND-2 are logical NAND gates which use minimum sized transistors (0.24 μm/0.18 μm). The component #3 is an asymmetrical delayed inverter. This inverter is configured to ensure that its output is delayed when its input has a 0 to 1 transition, however, it has a negligible delay on the 1 to 0 transition. The delay is specified by the input bias voltage V_{delay}, to the delayed inverter.

As shown in Figure 3.3, the implemented controller circuit uses two monostable,

Mono-QG, and Mono-R. These monostables are identical expect for the delayed inverter as Mono-R has a shorter delay than Mono-QG. The operation of the monostable is divided into 4 major phases: Idle phase, trigger phase, delay phase, and reset phase. Figure 3.6 shows the transition stages of the monostable in detail. These four stages are explained in respective order.

$\phi 1$: The idle phase

In this phase the monostable is in its stable state. The input *in* and output *out* of the monostable are 1 (stage 1). Hence, node A is $1\ NAND\ 1 = 0$, and node B which is the inverse of node A is 1. Since A is 1, and B is 0, the output *out* is $1\ NAND\ 0 = 1$, which ensures that monostable is in the stable state.

$\phi 2$: The trigger phase

The monostable is triggered on the falling edge of the input *in*. When the input changes from 1 to 0, the output *out* is still 1, hence node A is $0\ NAND\ 1 = 1$ (stage 2). Node A propagates instantly to the first input of NAND-2. In contrast, node A propagates with a delay specified by the asymmetrical delayed inverter to the second input of NAND-2 (stage 3). As node B is 1 form the idle phase state in stage-1, the output *out* is $1\ NAND\ 1 = 0$. As a result, *out* follows the input transition almost instantly (stage 4). Due to the change in the state of output, node A is $0 NAND 0 = 1$, hence there is no change in state of node A since stage-3 (stage 5).

$\phi 3$: the delay phase

During this phase, the state of the input *in* does not affect the monostable operation as the second input of NAND-1 is 0. As a result, node A is $X\ NAND\ 0 = 1$, where X is either 0 or 1. The input *in* can return back to 1, or kept at 0. In Figure 3.6 *in* returns back to 1 (stage 6). After a specified delayed by the bias voltage V_{delay} (stage 7), node A will propagate through the delayed inverter, and node B changes from 1 to 0 (stage 8).

$\phi 4$: The reset phase

After the delay has elapsed, The output *out* is $0\ NAND\ 1 = 1$ (stage 9). If the input *in* is still 0 from stage-6, then node A is $0\ NAND\ 1 = 1$, which does not change its previous state. However, if the input has returned to the initial state in stage-6, then node A is $1\ NAND\ 1 = 0$ (stage 10). Node A propagates instantly to the first input

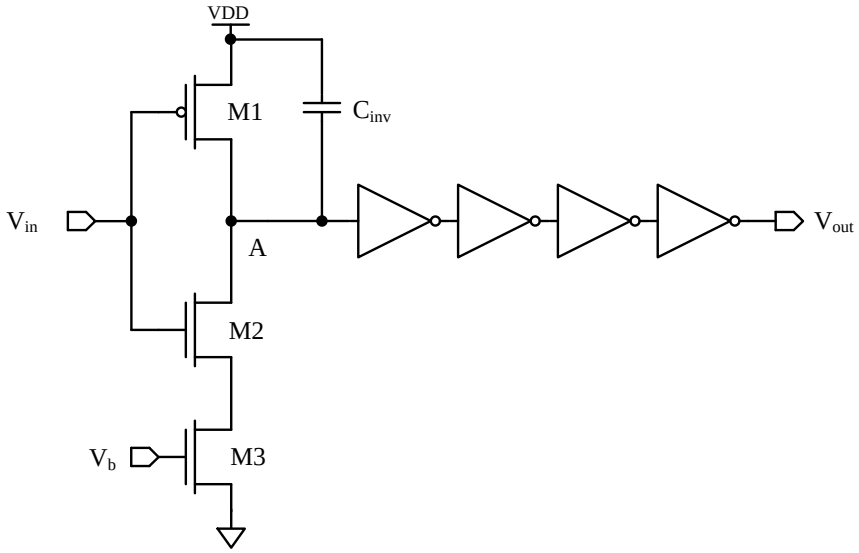


Figure 3.7: The schematic diagram for the asymmetric delayed inverter

of NAND-2. As the delayed inverter is asymmetrical and has a much shorter delay on the falling edge tan the rising edge, node B changes to 1 following a very short delay (stage 11). The output *out* is 0 $NAND$ 1 = 1, hence the monostable is in its stable state and back to the idle phase (stage 12).

3.3.4 Asymmetric delayed inverter

An asymmetric delayed inverter is a current starved digital inverter which has a delayed output only at either falling edge or rising edge of its input. The schematic diagram of the implemented asymmetrical delayed inverter is shown in Figure 3.7. The delay occurs when its input has a 0 to 1 transition (rising edge) . The delay is specified by the size of the capacitance on node A, and the discharging current, which is controlled by the gate voltage of M3. On the rising edge of the input V_{in}, when the input is a logic 1, M1 is not conducting and M2 is conducting, hence C_{inv} is discharged via M3 which is a current limited transistor by the means of its gate voltage, V_b. A higher bias voltage decreases the discharging time, hence, leading to shorter delays. The output of the inverter, V_{out} changes its state once the voltage of node A is below the threshold voltage of the next inverter. The series of inverters ensure that the output pulse has a sharp rising and falling edge. On the other hand, on the falling edge of the input V_{in}, M1 is conducting and M2 is not conducting. The capacitor C_{inv} charges rapidly via M1, which is approximately 10 times wider than M3, and has the maximum gate voltage. As a result, the falling edge transition

Component	Symbol	Value
Common	$M1$	W/L=0.24/0.18 (μm)
	$M2$	W/L=0.24/0.18 (μm)
Delay-QG	$M3$	W/L=0.24/2 (μm)
	C_{inv}	26.9 fF
Delay-R	$M3$	W/L=0.24/0.5 (μm)
	C_{inv}	not used
NOT	nMOS	W/L=0.24/0.18 (μm)
	pMOS	W/L=0.24/0.18 (μm)

Table 3.3: The specification of the components used in the asymmetric delayed inverter

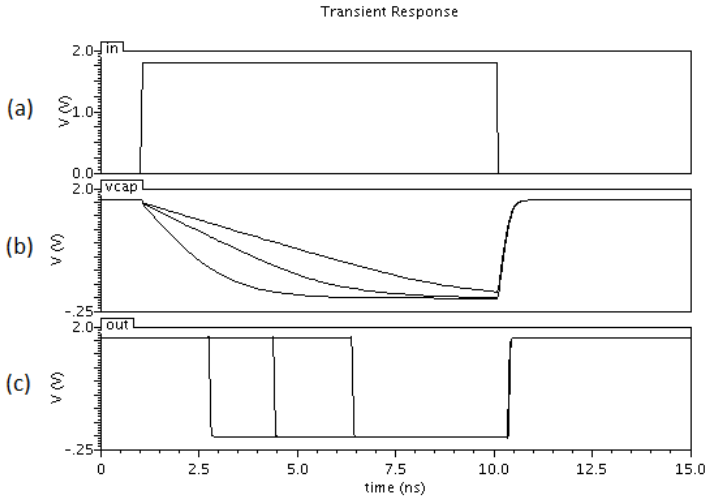


Figure 3.8: The circuit simulation for the asymmetrical delayed inverter Delay-QG, when V_{QG} is 1.7V, 1.3V, and 1.1V

creates a much shorter delay than the rising edge of the input V_{in}.

Table3.3 shows the specification of the components in the asymmetrical delayed inverter circuit. There are two types of delayed inverters designed for the ADR circuit: Delay-QG and Delay-R. The Delay-QG uses a 26.9 fF metal-insulator-metal (MIM) capacitor which creates a delay between 1.6 ns to 5.4 ns when M3 is biased above its threshold voltage. The Delay-R does not include a capacitor, hence the delay is created only by the parasitic capacitance on the node A. The small capacitance means that the delay is between 0.18 ns to 0.25 ns.

Figure 3.8 shows the circuit simulation for the Delay-QG inverter when the input bias voltage V_{QG} is 1.7V, 1.3V, and 1.1V. As expected, the delay occurs at the rising edge of the input (or falling edge of the output). A higher input bias voltage, leads to a quicker discharge of the capacitance C_{inv} , hence, causing shorter delays.

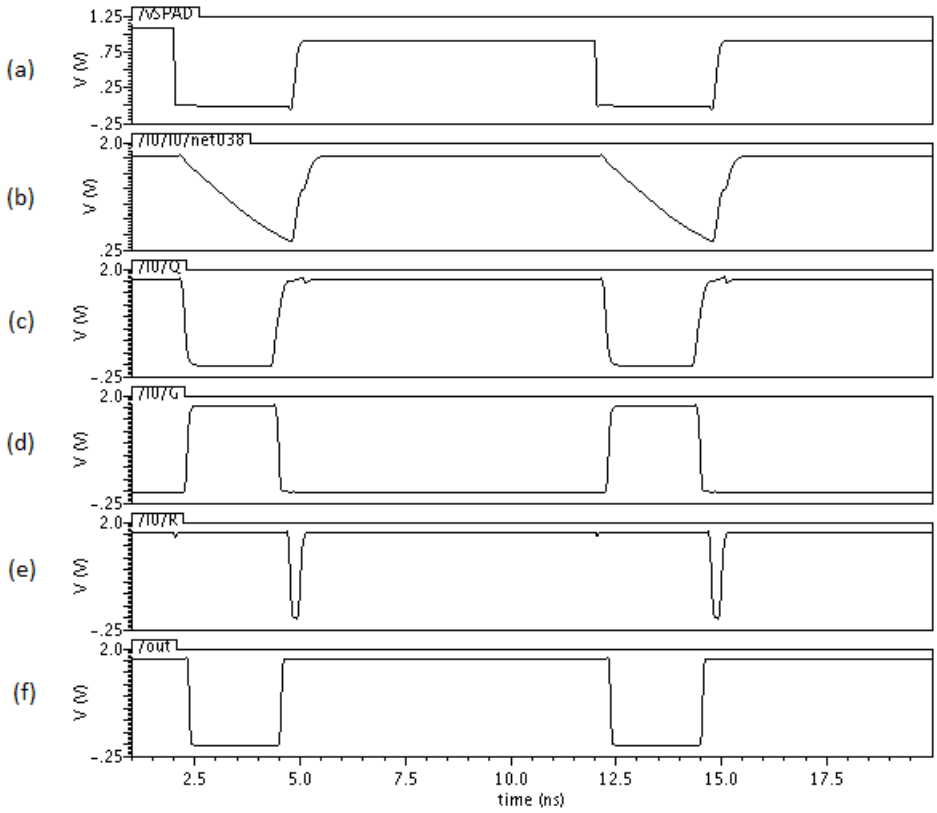


Figure 3.9: The circuit simulation for the automatic diode reset circuit shown in Figure 3.2. (a) the avalanche trigger signal, (b) the voltage of the diode's node, (c) the controls signal Q, (d) signal G, (e) signal R, and (f) the output

The simulation results for the automatic diode reset circuit are shown in Figure 3.9. The Verilog-A model developed for the avalanche diode and described in section 2.3.3 is used in this circuit simulation. In this simulation, two avalanche events occur at $t = 2\,ns$ and $t = 12\,ns$. The minimum deadtime achieved by simulation is 2.4ns, and the minimum reset pulse is 240ps when V_{QG} and V_R are equal to the supply voltage.

3.3.5 ADR implementation

The automatic diode reset circuit is implemented in UMC 0.18 μm standard CMOS. Figure 3.10 shows the physical layout of the implemented ADR circuit. The total physical dimensions of the automatic diode reset is $12 \times 24\,\mu m^2$. The avalanche diode which was characterised in the previous chapter is integrated the ADR circuit. The total size of the SPAD circuit including the avalanche diode is $50 \times 30\ \mu m^2$. The

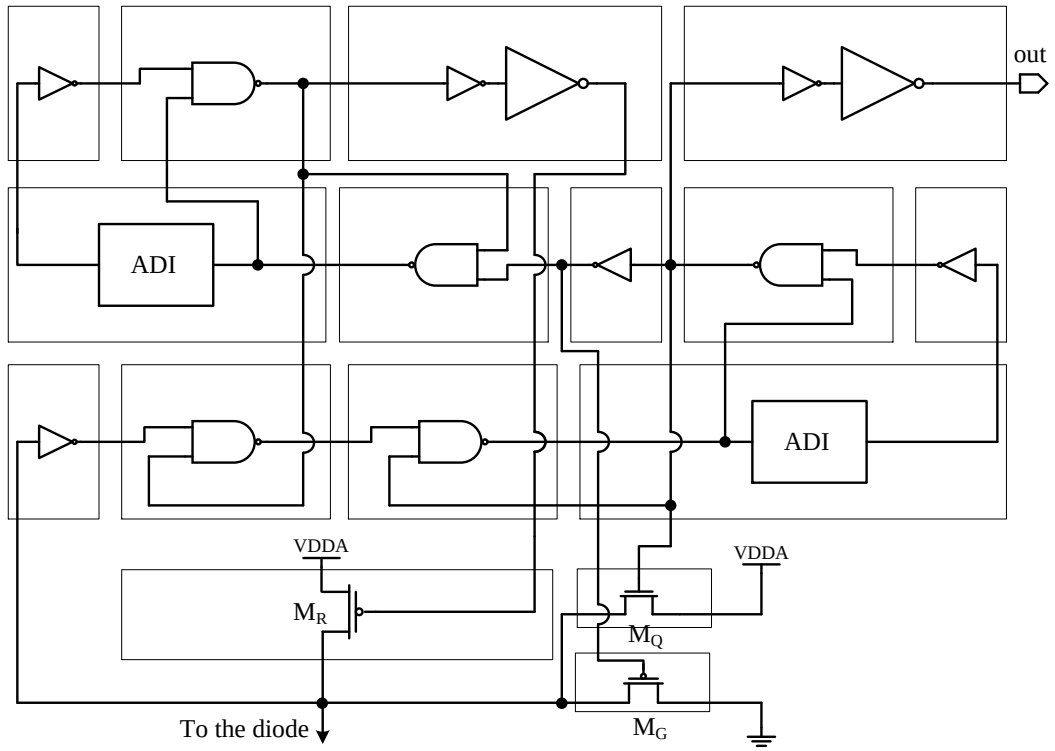


Figure 3.10: The physical layout of the automatic diode reset circuit

output of the SPAD circuit, which is the control signal Q, is locally buffered before connecting to digital output pads, which are proprietary digital push pull buffers capable of driving the parasitic capacitance of the package at high speeds. Each chip includes four SPAD circuits with individual outputs.

3.4 ADR characterisation

The fabricated SPADs were tested for their functionality and performance. The bias voltage of the avalanche diode in the SPAD circuit is $V_A = V_{br} + V_{OV}$, where V_{br} is the breakdown voltage and V_{OV}, is the over voltage. The breakdown voltage V_{br} is 10.4V based on Figure 2.3. The bias voltage of the avalanche diode is controlled by two separate voltage terminals POS and NEG shown in Figure 3.2. The voltage of the NEG terminal is chosen to ensure that the breakdown voltage is not lower that the ground of the circuit, hence, the minimum limit for NEG terminal is -10.4V. On the other hand, the maximum voltage of the POS terminal should not be more than the digital supply voltage VDD which is 1.8V.

V_{QG}(V)	Deadtime (ns)			
	#1	#2	#3	#4
1.74	3.99	3.98	4.09	3.73
1.44	5.01	5.09	5.18	5.02
1.16	6.04	6.13	6.2	6.3
1.03	7.98	9.02	8.52	8.4
0.975	9.72	11.06	10.38	10.5

Table 3.4: The experimental deadtimes versus various bias voltage V_{QG} for four SPAD circuits

3.4.1 Deadtime

The major function of the automatic diode reset is the ability to inactivate the SPAD for a specified time, which is referred to as deadtime. The deadtime is independent of biasing conditions of the avalanche diode as long as it is within the defined voltage limits. The deadtime only depends on timing circuitry which include two monostables. In order to verify the functionality of the implemented monostables, the deadtime is measured for the 4 SPADs fabricated within a chip. The output of each SPAD is a digital pulse which is at logic 1 in idle mode, falling edge when an avalanche event is detected, and logic 0 when the SPAD is inactive for the duration of the deadtime. An Agilent oscilloscope with a bandwidth of 1 GHz is used to observe the output pulse and measure the duration of deadtime. The deadtime is measured by the time difference between the rising edge and falling edge of the output pulse. The rising edge indicates that the deadtime has elapsed and the SPAD has recovered from its inactivity period. The duration of deadtime is controlled by the bias voltage V_{QG} in the monostable Mono-QG in Figure 3.3.

Table 3.4 shows of deadtimes of 4 SPADs versus various bias voltages. The deadtime increases when V_{QG} decreases as the discharge time of the capacitor of the asymmetrical delayed inverter shown in Figure 3.7 increases. There are variations in the measured deadtime among the four SPAD circuits. This could be due to variations in the threshold voltage of M3 in Figure 3.7, especially as the bias voltage approaches the threshold voltage of this transistor. The deadtime of 5ns has the lowest variation among the four SPAD, hence it chosen as the nominal deadtime in the rest of experiments this chapter.

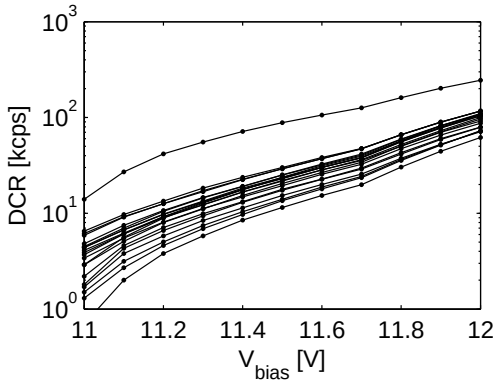


Figure 3.11: The dark count rate (DCR) versus bias voltage for 20 SPADs

3.4.2 Dark Count Rate

The dark count rate (DCR) is measured for 20 SPAD circuit, which include 5 chips, and each chip has four individual SPAD circuits. Figure 3.11 shows the DCR versus bias voltage for 20 SPAD circuits. The range of bias is 1V which is achieved by a combination of voltages on the POS and NEG terminals shown in Figure 3.2. In order to achieve a 1V voltage change across the bias voltage, the negative voltage of the NEG terminal is increased until the SPAD malfunctions due to small over-voltage. The maximum voltage of the NEG terminal which the SPAD functions is -9.7V. Similarly, the positive voltage of the terminal POS can be reduced to 1.5V. As a result, an overall bias voltage change of 1V can be achieved.

Similar to the measurement method described in section 2.4.1, the oscilloscope counter is used to measure the DCR. The increases in the DCR versus bias voltage is similar to the increases in the DCR shown in Figure 2.9 in the bias voltage range of 11V to 12V. As it can be seen from DCR measurement results, one SPAD circuit among the measured 20 SPADs, has an unusually high dark count rate which is likely to be caused by defects within the active area of the avalanche diode. The average DCR at the highest bias voltage, 12V is 102.5 Kcps (count per second).

3.4.3 Relative Photon Detection Probability

The relative photon detection probability (PDP) is measured for 20 SPADs. The measurement method for the relative PDP is similar to method described in section 2.4.3, however, PDP is normalised to the highest measured count rate among the SPADs. Figure 3.12 shows the relative PDP versus bias voltage for 20 SPADs. Similar to the previous experimental result shown in Figure 3.11, the range of the bias voltage

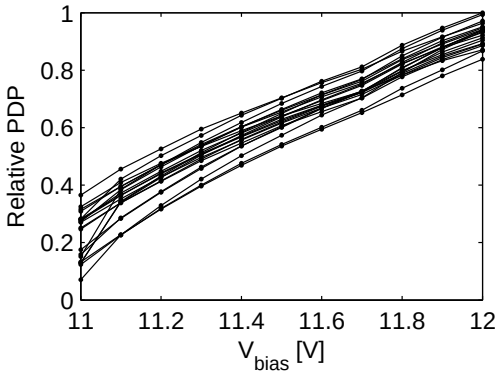


Figure 3.12: The relative photon detection probability (PDP) versus bias voltage for 20 SPADs

is from 11V to 12V. The DCR obtained in the previous experiment and shown in Figure 3.11 is subtracted from the count rate measured in this experiment. It can be observed that the relative PDP increases linearly with the bias voltage in the measured range.

3.4.4 Pulse interval statistics

In an ideal photon counting detector, the time interval between two consequent avalanche events has an exponential distribution with a mean equal to the rate of avalanche events. In the implemented SPAD circuit, an avalanche is indicated by the falling edge of the output pulse. Hence, the inter time pulse interval, which is between two falling edges should have an exponential distribution. The inter pulse statistics confirms the photon counting capability of the detector. Non-ideal effects such as after-pulsing distort the exponential distribution of the inter pulse statistics. After-pulsing, which is frequently reported in SPAD characterisation literature, is occurrence of a secondary avalanche event shortly after an avalanche event [73]. This secondary avalanche event is not related to detection of a photon, instead it is related to the trapped charges from a previous avalanche event. After-pulsing which is known to increase with shorter deadtimes and higher bias voltages [73, 74], alters the inter pulse statistics by increasing the frequency of the pulses which have shorter time intervals. As a result the inter pulse statistic is no longer has an exponential distribution.

In order to measure the inter pulse time interval statistics for the fabricated SPADs, a long sequence of pulses is captured using an oscilloscope which has a 1ns timing resolution. The time difference between each pulse is calculated and recorded.

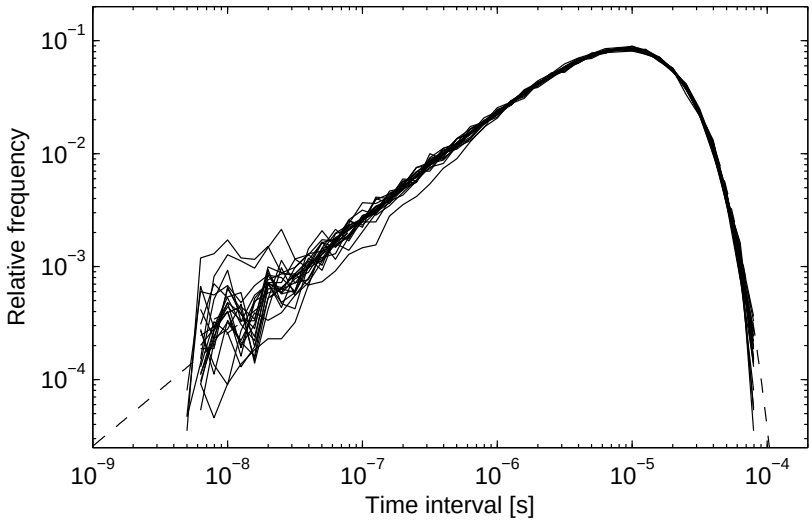


Figure 3.13: The relative frequency histograms of the inter pulse time interval statistics for 20 fabricated SPADs. Theoretical (dashed line) and Experimental (solid lines)

The experiment is repeated until enough number of samples are collected for a statistical analysis. A comparison between SPADs is possible if the average event count rate is equal for all the measured SPADs. However, as shown in Figure 3.11, each SPAD has a different DCR at a specific bias voltage.

In order to equalise the count rate of all the SPADs at a specific bias voltage, illumination is added to the SPAD during measurement to increases the DCR of each SPAD to 100 kHz at the bias voltage of 12V, except for the SPAD which has a DCR above 100 kHz. The bias voltage for the SPAD, which has a high DCR, has been reduced until a 100 kHz count rate is obtained. Figure 3.13 shows the theoretical and the experimental relative frequency histograms of the inter pulse time interval statistics for 20 fabricated SPAD circuits when average count rate is 100 kcps and the bias voltage is 12V. The time interval axis is selected to be logarithmic as it reveals more information in shorter time scales. The relative frequency axis is normalised for about 14k samples.

The theoretical histogram shown in Figure 3.13 is calculated by:

$$p(a < t < b) = F(b) - F(a) \tag{3.2}$$

$$F(x) = 1 - e^{\lambda x} \tag{3.3}$$

where $p(a < t < b)$ is the relative frequency between the range (a, b), $F(x)$ is the cumulative distribution function, and λ is the event count rate which is 100 kcps. The experimental data shows that there is no time interval data smaller than 5 ns which is the deadtime set for the SPADs. The statistical fluctuations observed in the shorter time intervals are due to limited number of samples for each measurement. As it can be observed form the histogram, the experimental data for implemented SPADs almost closely follow the theoretical histogram of the exponential distribution. This suggests that fabricated SPADs have no evidence of afterpulsing.

3.5 Summary

The automatic diode reset (ADR) provides the means of quenching the avalanche current, holding the bias voltage of the avalanche diode below its breakdown voltage for a specified duration, and recovering the diode to its initial bias voltage, ready for the detection of next avalanche event. During the period which the diode is biased below its breakdown voltage, the SPAD is inactive and no avalanche events are detected, hence this period is referred to as deadtime. Due to the independence of the means of quenching and the recovery process in the ADR circuit, relatively short deadtimes have been achieved.

Transistor switches are used to perform the quenching, holding, and recovery operations. The timing signals for these transistor switches are provided by a controller circuit which includes two monostables. A monostable is a timer circuit which creates a delay on the falling/rising of its input signal. The component within the monostable that creates the delay is an asymmetrical current starved inverter. The delay can be controlled by a bias voltage, as a result, the deadtime is set by the user. The ADR circuit has an additional mechanism which protects the SPAD from an unexpected failure to recover due to a sustained avalanche which could occur when an avalanche event is detected during the recovery stage.

The automatic diode reset circuit is implemented in the same fabrication process as the avalanche diode which was fabricated in the previous chapter. Each fabricated chips includes four SPAD circuits, and five chips have been tested, hence in total 20 SPAD circuits were used in the experiments in this chapter. A minimum deadtime of 4 ns was achieved when the ADR circuit which was integrated with the avalanche diode. The dark count rate (DCR) and the relative photon detection probability (PDP) measurements of the ADR circuit are in agreement with the results from diode characterisation chapter, hence, verifying the functionality of the quenching

and the recovery processes.

In order to verify the photon counting capability of the implemented SPAD, the inter pulse interval statistics for each SPAD is measured. In a theoretical photon counting detector, the time between to detection events is based on exponential distribution. The exponential distribution is distorted if non ideal phenomenon such as afterpulsing exist. Measurement of the fabricated SPAD with deadtime of 5ns shows that the majority of the SPADs have a time interval between avalanche events which closely follows the exponential distribution. Hence, the fabricated SPADs perform as theoretical photon counter.

Chapter 4

Numerical Modelling

4.1 Introduction

One of the performance parameters of an optical communication link is the sensitivity of the photodetector in the receiver. A photon counting detector is a sensitive photodetector which has the ability of detecting and counting individual photons. In an ideal photon counting detector, when a single photon is detected, an event is triggered which indicates the detection of a photon. In practice, similar to any other detector, not all the photons arriving at the detector's surface area are detected. Each detector has an active area which is a fraction of the detector's surface area. The photons are only absorbed in this active area. The ratio of this active are to the total detector's area is referred to as fill factor. Moreover, some the photons are absorbed before they reach the surface of the active area due to external optical elements such as lenses and glass protection layers on top of the detector. In silicon photo-detectors, light is absorbed at various depths depending on its wavelength. The photo-absorption of a silicon substrate depends on the energy of the absorbed photon. Once the photon is absorbed within the active area of the silicon detector, an electron-hole pair is created. An appropriate electrical mechanism leads to the detection of this electron-hole pair. As a result, a fraction of the arriving photons at the surface of the detector, are not detected. This leads to an overall photon detection probability (PDP) which is a specific characteristic of the detector.

In addition, in some detectors such as SPADs, not all the detected photons lead to a detection event. As these detectors are based on avalanche multiplication process which is a statistical process in nature, some of the detected photons do not generate a strong enough avalanche current which the sensing circuitry is able to detect. In this chapter it is assumed that all the receiving photons are detected, hence the detector

has a 100% photon detection probability.

Similar to other detectors such as PD and APDs, in a photon counting detector spontaneous detection events in the dark are referred to as dark counts. The dark counts degrade the performance of a communication link as the receiver is not able to distinguish between a detection event which is caused by a signal and a spontaneous event. Moreover, in free space optical communications, the background illumination that originates from sources other than the transmitter's signal further degrade the performance of the communication link. The overall background illumination and dark noise of a photon counting detector is referred to as the background noise in this chapter.

4.2 Probability of Error

The performance of the communication link ultimately depends on how many transmitted bits are successfully received. In a communication link, an error occurs when a transmitted bit is not received, leading to an error in the data transmission. Hence, the erroneous bit is required to be re-transmitted which results in degradation of the performance of the link. The Probability of Error (PE) in a communication link is [9]:

$$PE = p(R0|T1).p(T1) + p(R1|T0).p(T0) \tag{4.1}$$

where $p(T1)$ is probability of transmitting a 1, $p(T0)$ is the probability of transmitting a 0, $p(R0|T1)$ is the probability of receiving a 0 when a 1 is transmitted, and $p(R1|T0)$ is probability of receiving one when a 0 is transmitted. Figure 4.1, shows how the probability of error is calculated. The solid lines shows when the correct transmission has occurred, and the dashed lines represent the condition which an error has occurred during the transmission.

4.2.1 A receiver with no background noise

In order to assess the performance of a photon counting detector, an On-OFF Keying (OOK) modulation scheme is used. The OOK is simplest form of amplitude shift keying, and it is regarded as a binary transmission where only a single bit is transmitted at a time. In the case of an optical link, it is assumed that the transmitter is OFF when a 0 is transmitted, and ON when a 1 is transmitted. The transmitter has a specific intensity when it is ON, and the probability of detecting n photon within

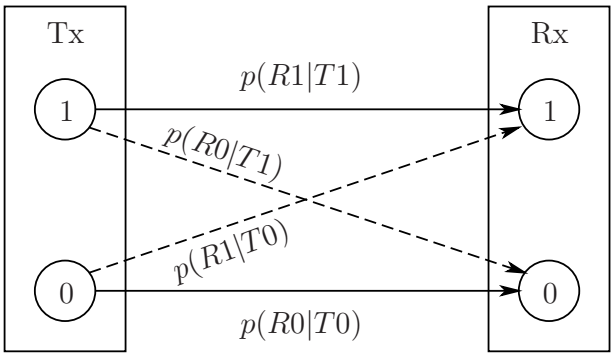


Figure 4.1: The definition of probability of error (PE).[Kartalopoulos](with permission from John Wiley & Sons)

PE	λ_s	PE	λ_s
10^{-2}	3.9	10^{-6}	13.1
10^{-3}	6.2	10^{-7}	15.4
10^{-4}	8.5	10^{-8}	17.7
10^{-5}	10.8	10^{-9}	20

Table 4.1: The Probability of Error (PE) for a photon counting detector based on Poisson distribution with no background noise

the bit period T_b has a Poisson distribution:

$$p(n) \equiv p(n = k) = \frac{(\lambda_s)^k}{k!}.e^{-\lambda_s} \tag{4.2}$$

where n is the number of detected signal photons within the time slot, $p(n)$ is the probability of having n photons within the slot duration, and λ_s is the average number of detected signal photons in T_b. In the case of an OOK transmission, according to Equation 4.1, an error occurs when a 0 is received instead of 1, and a 1 is received instead of 0. However, it is impossible for a transmitted 0 to be received as a 1 if the transmitter is fully OFF and there is no background noise. Hence, an error only occurs when no photon is detected. That is when $p(0) = p(n = 0) = e^{-\lambda_s}$, hence according to Equations 4.1 and 4.2 the PE is:

$$PE \equiv p(T1).p(R0|T1) = \frac{1}{2}.p(0) = \frac{1}{2}.e^{-\lambda_s} \tag{4.3}$$

Table 4.1 shows the average detected photon rate, λ_s, required to achieve selected PEs. Note that PE has an exponential relation with λ_s. Hence, PE is very sensitive to changes to average number of signal photons, and a small increase in λ_s leads to a significant improvement in the probability of an error.

4.2.2 A receiver with background noise

When background noise is present, photons are detected when 0s are being transmitted. As photons are being detected regardless of the transmitted data, a decision is required to distinguish between a 0 and a 1 bit. Similar to the conventional amplitude detectors, such as PDs and APDs, a decision is made based on a threshold photon count n_T. If the number of counted photons n is larger than the decision threshold, n_T, the received bit is perceived as a 1, otherwise a 0. In the case of background an additional error is caused when a 0 bit is perceived as a 1. If the average background noise count within a slot is λ_b, the probability of error described in Equation 4.1 is re-written as:

$$PE = p(T1).p(n < n_T|T1) + p(T0).p(n > n_T|T0) \tag{4.4}$$

where the first term is related to the transmitted 1s which are detected as 0s, and the second term is related to the transmitted 0s which are detected as 1s. When a 0 is transmitted, the average detected photon rate is the background noise rate, λ_b. However, in a slot when a 1 is transmitted the average photon rate is the sum of detected signal and noise photon rates, $\lambda_{sb} = \lambda_s + \lambda_b$. The background noise rate is always present regardless of what bit is transmitted. In a random bit sequence, the probability of having a 1 is equal to the probability of having a 0, hence $p(T1) = p(T0) = 0.5$. Therefore the PE can be calculated as [9]:

$$PE = \frac{1}{2}.\sum_{k=0}^{n_T} \frac{(\lambda_s + \lambda_b)^k}{k!}.e^{-(\lambda_s+\lambda_b)} + \frac{1}{2}.\sum_{k=n_T}^{\infty} \frac{(\lambda_b)^k}{k!}.e^{-\lambda_b} \tag{4.5}$$

Notice that if there is no background noise, then $n_T = 0$, and the PE is equal to the PE from Equation 4.3. If $\lambda_b > 0$, then according to Equation 4.5, PE is a function of λ_s, λ_b, and n_T. Unlike the first two variables, the decision threshold n_T is set by the receiver system. Having a constant λ_s and λ_b, the receiver always selects n_T to ensure that the lowest PE is achieved. As a result, the PE for a receivers with an adaptive threshold is function of only λ_s, λ_b. Figure 4.2 shows the PE versus average detected signal photon rate λ_s for selected background photon rate λ_b. These traces are calculated from equation 4.5. The abrupt changes seen in PE traces in Figure 4.2, are caused by changes in n_T. Figure 4.3 shows the corresponding n_T for each trace in Figure 4.2. It can be observed that the λ_s which PE changes abruptly, is the same λ_s where n_T is increased by one. As n_T is an integer value, the changes to PE appear to be abrupt. For example, when $\lambda_b = 0.01$, the first abrupt jump occurs at

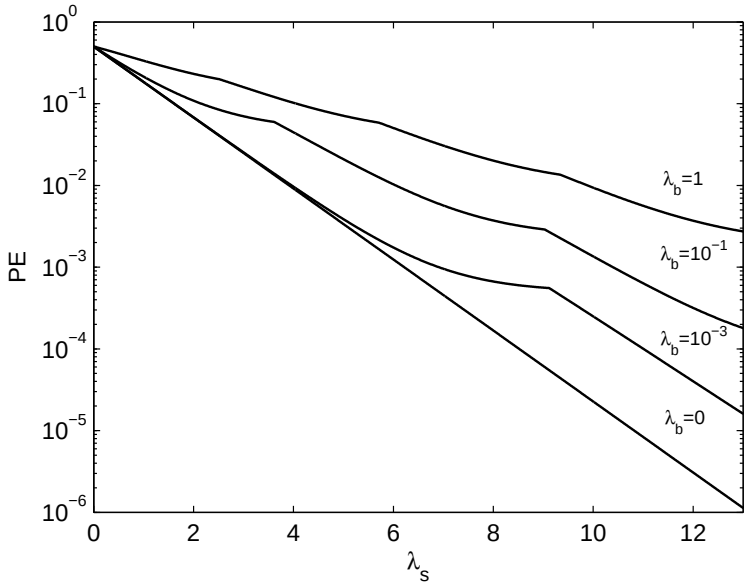


Figure 4.2: The PE versus average detected signal photon rate λ_s, for selected average background noise λ_b

$\lambda_s = 6.5$, where n_T changes from zero to one.

4.2.3 Power penalty for background noise

Two major observations are made from the PE calculations shown in Figure 4.2: the power penalty required to maintain a target PE with increasing background noise, and the maximum background noise when the power penalty is negligible. As expected, when background noise increase, more signal intensity is required to maintain the PE. For example without background noise, 6.2 photons are required to achieve a

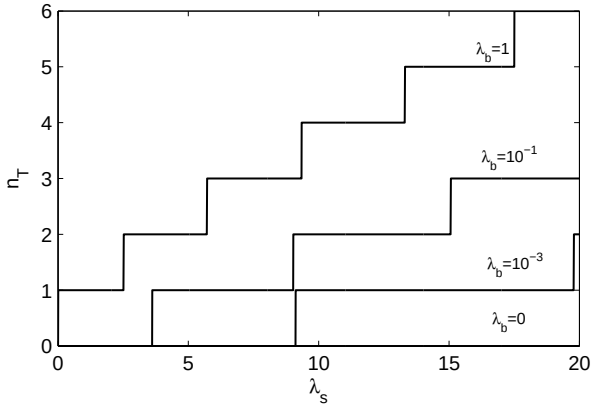


Figure 4.3: The threshold level n_T for figure 4.2

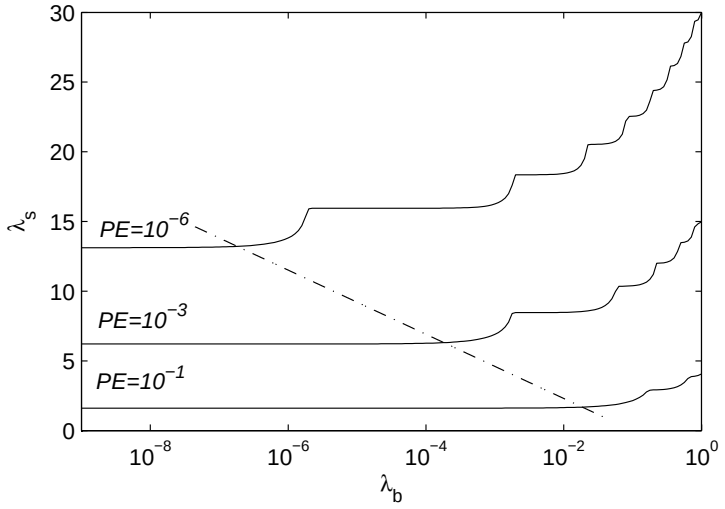


Figure 4.4: The power penalty for background noise in an ideal detector

$PE = 10^{-3}$. However when $\lambda_b = 10^{-2}$, about 8.4 average detected signal photons are required to achieve the same PE. As a result, 2.2 more photons in average are required to compensate for the effect of background noise. These extra signal photons are the power penalty for maintaining the PE. Figure 4.4 shows selected PE contour lines based on Equation 4.5. Similar to the photon counting receiver described in section 4.2.2, the PE is calculated for a receiver with adaptive decision threshold. It can be observed that when the background noise increases, the average signal photon required to maintain the BER increases, hence the power penalty increases. Similar to Figure 4.2, the abrupt changes in PE contour lines are as a result of decision threshold change.

As it can be seen in Figure 4.4, beyond a certain limit, the BER is independent of the background noise. This limit is represented by a dashed line in Figure 4.4. For the background noise smaller than the dashed line the BER is only a function of λ_s. The relation for this estimation is:

$$\lambda_b < \frac{1}{10}.e^{-\lambda_s} \qquad \Rightarrow \qquad \lambda_b < \frac{1}{5}.PE(\lambda_s) \tag{4.6}$$

The dashed line estimates that when the average background noise is smaller than the fifth of the ideal PE (without background noise), the power penalty is negligible. In general, the condition which $\lambda_b \ll PE(\lambda_s)$ is a reasonable estimation for negligible power penalty for the background noise. In practice, when the probability of a background photon is much smaller than probability of missing a signal photon, the PE is almost independent of the background noise.

4.3 Detector with deadtime

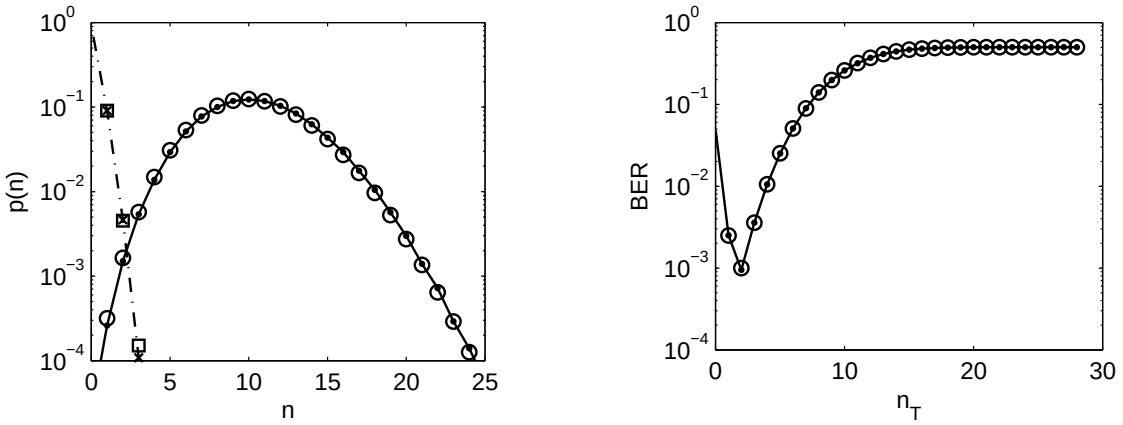
In this discussion so far it was assumed that the photon counting detector counts the photons instantaneously, however practically a detector requires a period after each detection event to recover to its initial state. For example, in SPADs with automatic diode reset a photon is detected when an electron-hole pair is generated. The detected photon triggers an avalanche process in the diode which leads to a detection of an avalanche event. The avalanche is quenched and the detector is recovered to its initial state after a period of time. This time is referred to as deadtime, which means the detector is inactive during this period. Although an electron-hole pair is generated, they do not lead to an avalanche event because the avalanche process is disabled. Therefore, here after, detected photons are distinguished form detected events.

4.3.1 The numerical model

In order to investigate the effect of deadtime on photon counting detectors, a numerical simulation model for a photon counting detector with deadtime has been developed. The simulation starts from a pseudo random bit stream representing the transmitted data sequence. Each bit within this sequence is processed individually. Photon arrival times are generated according to the corresponding transmitted bit. If the transmitted bit is 0, then the average photon rate $\lambda = \lambda_b$, and if the transmitted bit is 1, then $\lambda = \lambda_b + \lambda_s$, where λ is the average detected photon rate for that particular slot. The photon arrival times are calculated based on Poisson statistics [75]:

$$\tau_i = -\frac{1}{\lambda}.ln(u).T_b \tag{4.7}$$

where τ_i is the inter-time interval of the ith photon within the slot, u is a uniformly distributed random number, and T_b is the time lot for each bit. For simplicity during all the simulations , it is assumed that $T_b = 1$, hence the results are normalised for the slot duration. Only τ_1 is the time between the first photon and the beginning of the bit period where $t = 0$. The valid photons are when $\tau_i < T_b$. Equation 4.7 is based on exponential distribution of the time of arrival of photons. As each photon is generated, they do not have any correlation with the previous photons. This independence in photon generation results in an exponential distribution between generation time of each photon which also leads to a Poisson statistics for the number of counted photons. For each bit, the number of detected photon, n, is the number of photons which are detected within T_b. the value of n is stored for all the received bits within



(a) (b)

Figure 4.5: The simulation results for a photon counting detector based on the developed numerical Model and theoretical equations shown in Equation 4.5. (a) relative frequency histogram for 0s (□model, ×theory) and 1s (◯model, • theory), when $\lambda_s = 10.4$, and $\lambda_b = 0.1$. The total photon rate in a slot when a 1 is transmitted is $\lambda(theory) = 10.4 + 0.1 = 10.5$, and $\lambda(sim) = 10.503$. (b) when $\lambda_s = 10.4$, and $\lambda_b = 0.1$. The BER (theory) is 0.995×10^{-3} @ $n_T = 2$, and The BER (simulation) is 0.950×10^{-3} @ $n_T = 2$

the data sequence. Once n for each bit is calculated, a decision threshold is applied to declare received 1s and 0s. Then the received data sequence is compared to the transmitted data sequence to identify the erroneously received bits. The total number errors divided by the length of the data sequence determines the bit error ratio. The bit error ratio (BER) is major performance parameter of in a receiver's performance. The BER is basically an estimation of the probability of error (PE). In the rest of this chapter, BER is used instead of PE as majority of the results are based on numerical simulations.

The simulation results when $\tau_d = 0$

Figure 4.5 shows results from the simulation model, and a comparison with the theory based on Equation 4.5. Figure 4.5.a shows the probability distribution for n when of $\lambda_s = 10.4$ and $\lambda_b = 0.1$. As expected it is based on a Poisson distribution for received 0s with an average of 0.1, and received 1s with an average of 10.5. The total photon rate when a 1 is transmitted includes both signal and background noise which is $\lambda = \lambda_s + \lambda_b = 10.4 + 0.1 = 10.5$. Figure 4.5.b shows the BER with various n_T calculated from theory and simulation for the result in Figure 4.5.a. As expected, in presence of relatively high background noise, the best BER is achieved when $n_T = 2$,

hence $BER \approx 10^{-3}$. As it can be seen in 4.5, the simulated values are in close agreement with the theoretically calculated values.

The simulation results when $\tau_d > 0$

In order to implement the deadtime in the simulation model of the photon counting detector, an inactivity timer is introduced to the model. At the beginning of each simulation, this timer is set to zero. Once a photon is detected the inactivity timer is triggered. If a photon in the generated photon stream is within the inactivity period, it is counted as a detected photon but not as a detected event. As a result, the number of detected events is always equal or smaller than number of detected photons. Once the deadtime is elapsed, the timer is reset, and the next photon is counted both as a detected photon and a detected event. If the bit slot duration is finished before the deadtime is elapsed, the remaining deadtime time is forwarded into the next adjacent slot. In this case the next slot is inactive until the forwarded deadtime has elapsed. The continuity of the timer from the slot to the next slot ensures that the detector is always inactive for the period deadtime regardless of when the photon has been detected within the photon stream. In a photon counting detector, the relative deadtime τ_d is the ratio of deadtime to the bit slot duration T_b:

$$\tau_d = \frac{DT}{T_b} \tag{4.8}$$

where DT is the deadtime, and T_b is the slot duration. The number of events n_e, that can be recorded within a bit slot is unlimited in a detector with no deadtime, however, in a detector with deadtime, the maximum number of events that can be recorded within a slot duration is [76]:

$$n_e(max) = \left\lfloor \frac{1}{\tau_d} \right\rfloor + 1 \tag{4.9}$$

For example when $\tau_d = 0.1$ a maximum of 11 events are detected regardless of the number of photons detected. Hence, in relatively high photon rates, many of the detected photon do not lead to a detected event. As a result, the rate of detected photons has a non-linear relationship with the detected events. If λ_e is the average rate of detected events, and $\tau_d \ll T_b$ then [77]:

$$\lambda_e = \frac{\lambda_s}{1 + \lambda_s.\tau_d} \tag{4.10}$$

Figure 4.6, shows the non-liner relation between λ_e as a function of λ_s for selected

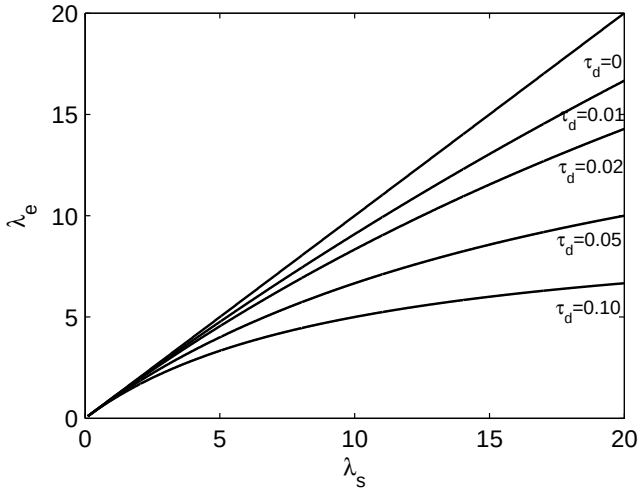


Figure 4.6: The effect of deadtime on the event rate λ_e versus detected photon rate λ_s

τ_d, calculated from Equation 4.10. When there is no deadtime, the detected event rate is equal to detected photon rate, however as deadtime increases, the event rate will eventually converge to $n_e(max)$.

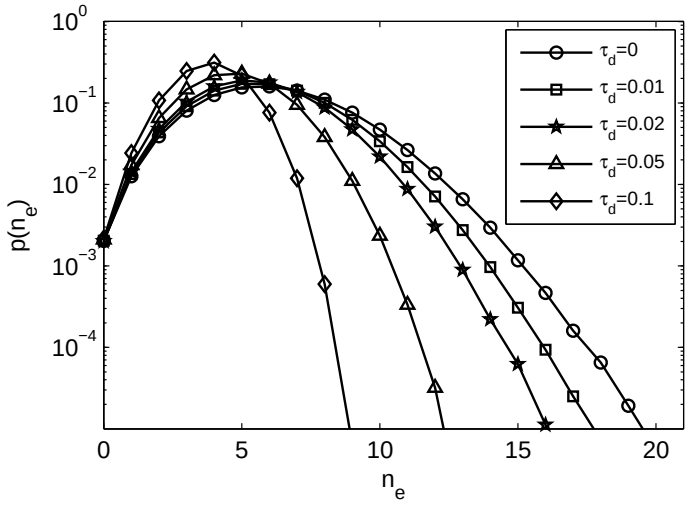


Figure 4.7: The distribution of detected events for selected deadtimes when $\lambda_s = 6.2$

Figure 4.7 shows the simulation results for selected deadtimes when $\lambda_s = 6.2$, and no background noise. n_e is the detected event count within the slot duration, and $p(n_e)$ is the relative occurrence frequency of n_e . When there is no deadtime the relative frequency histogram of the detected events is identical to the detected photon which is based on Poisson distribution, hence $\lambda_e = \lambda_s$. Table 4.2 shows calculated λ_e

for deadtimes shown in Figure 4.7, by both Equation 4.10, and the simulation model. As expected when deadtime increase, the rate of detected events decreases. In the simulation model, λ_e is mean of each distribution shown in Figure 4.7. which are in close agreement with the calculated λ_e based on Equation 4.10.

τ_d	λ_e (sim)	λ_e (theory)	n_{max}
0	6.201	6.2	∞
0.01	5.838	5.838	101
0.02	4.520	5.516	51
0.05	4.747	4.733	21
0.10	3.865	3.827	11

Table 4.2: The comparison of event rate based on the theoretical and the simulation results for small deadtimes

As there is no background noise, and no deadtime the BER is $p(n = n_e = 0)$ and it is calculated by Equation 4.3, which results in $BER = 10^{-3}$. However, when $\tau_d > 0$ the distribution of the detected events is no longer based on Poisson distribution. As a result, the Equation 4.3 is no longer valid for non-zero deadtime. In the example shown in Figure 4.7, although the distribution of detected events changes with deadtime, the probability of missing and event, $p(n_e = 0)$ is approximately independent of deadtime. As a result, for this particular example, the $BER = 10^{-3}$ for $\tau_d < 0.1$.

However, when relative deadtime is close to the slot duration, the deadtime frequently overflows into the next adjacent slot, which leads to a form of inter symbol interference (ISI). The ISI increases the BER by blocking detected photon to be detected as events, hence the bit error rate is no longer independent of deadtime.

4.3.2 Inter Symbol Interference

The overflow of the deatime from a previous slot into the next slot causes an inter symbol interference (ISI). In order to parametrised ISI, τ_{isi} is defined as the amount of deadtime overflowed into the beginning of the next bit slot. If there is no ISI in a bit slot, then $\tau_{isi} = 0$, However, if $\tau_{isi} > 0$, the detector is inactive at the beginning of the slot, hence reducing the chances of detecting an event. This phenomenon may increase the BER due to the number of missed detected events.

4.3.3 Probability of ISI

The probability of ISI, $p(\tau_{isi})$, in a photon counting detector is defined as the probability of an inter symbol interference occurring after a 1 is transmitted. An ISI occurs only when $\tau_{isi} > 0$. For simplicity, it is assumed that there is no background noise. The introduction of $p(\tau_{isi})$ enables the investigation of the effect of ISI on the bit error ratio. The probability of ISI, $p(\tau_{isi})$, can be defined as:

$$p(\tau_{isi}) \equiv \frac{p(\tau_{isi} > 0 \,|\, T1)}{p(T1)} \tag{4.11}$$

where $p(T1)$ is the probability of a transmitted a 1 bit, and $p(\tau_{isi} > 0 \,|\, T1)$ is the probability of ISI occurring after a transmitted 1 bit. The τ_{isi} is ignored for the 0 bits because no photon are being detected when a 0 is transmitted as there is no background noise. As a result the probability of ISI is a function of detected signal photon rate λ_s and deadtime τ_d. The probability of ISI is a conservative estimation for occurrence of the ISI, and does not provide any information about the amount deadtime overflowed into the next slot. As a result, it is expected that for relatively low probability of ISI when $p(\tau_{isi}) \ll 1$, the BER is independent of deadtime because ISI is small enough to not influence the bit error rate. Figure 4.8 shows the contours for selected $p(\tau_{isi})$ as a function of λ_s and τ_d. These contours are calculated by Equation 4.11 using the simulation model of the photon counting detector with deadtime. The results in Figure 4.8 shows that for small $p(\tau_{isi})$, the contours can be estimated by:

$$\lambda_s.\tau_d = \frac{1}{\alpha} \tag{4.12}$$

where α is a function of $p(\tau_{isi})$. It can be seen from Figure 4.8 that α has an inverse relationship with $p(\tau_{isi})$. In addition for a specific α, average signal photon rate and deadtime are inversely proportional. It is expected that for relatively small probability of ISI, the BER is independent of deadtime. Hence, based on the simulation results in Figure 4.8, a conservative estimate for $p(\tau_{isi}) < 0.1$ shows that:

$$p(\tau_{isi}) < 0.1 \quad \Rightarrow \quad \alpha > 4 \tag{4.13}$$

replacing α from Equation 4.12:

$$\lambda_s.\tau_d < \frac{1}{4} \tag{4.14}$$

That means in order to achieve a higher BER which is independent of deadtime,

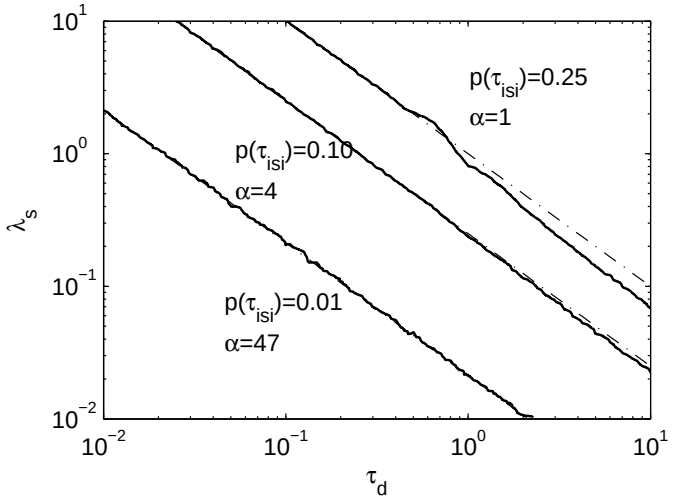


Figure 4.8: The contours for $p(\tau_{isi})$ as a function of relative deadtime and average signal photon rate. Calculated based on the simulation model (solid lines) and the estimation (dashed lines)

a higher λ_s, with a smaller τ_d is required. For example, when the target BER is 10^{-3}, the average signal photon rate required is 6.2. Hence the relative deadtime which satisfies the condition in Equation 4.14 is $\tau_d < 0.04$. Knowing the actual deadtime of the detector, the maximum slot duration, T_b which BER is independent of deadtime can be found. For example, for a SPAD with deadtime of 5 ns, the maximum slot duration which BER is independent of deadtime is 125ns, hence a data rate of 8 Mbps is achieved.

4.3.4 Average of ISI

The average of τ_{isi} throughout the bit stream $\langle\tau_{isi}\rangle$, is defined as the average deadtime that overflows into the next slot. Similar to $p(\tau_{isi})$, the average ISI only considers the transmitted 1 bits, as the 0 bits have no influence on the performance of the detector since there is no background noise. The average ISI also includes the situations when $\tau_{isi} = 0$. Figure 4.9 shows$\langle\tau_{isi}\rangle$ as a function of deadtime for selected average signal photon rate. The $\langle\tau_{isi}\rangle$ is calculated using the simulation model for a photon counting detector with deadtime. As can be seen from this figure, $\langle\tau_{isi}\rangle$ does not have a direct relationship with deadtime, as it has peaks and troughs when the deadtime changes.

It can be seen from Figure 4.9, these features occur at deadtimes when $n_e(max)$ changes from one integer value to the next one based on Equation 4.9. As the deadtime approaches a value at which $n_e(max)$ changes, the$\langle\tau_{isi}\rangle$ increases, however just before $n_e(max)$ changes, the$\langle\tau_{isi}\rangle$ sharply reduces and rises afterwards when $n_e(max)$

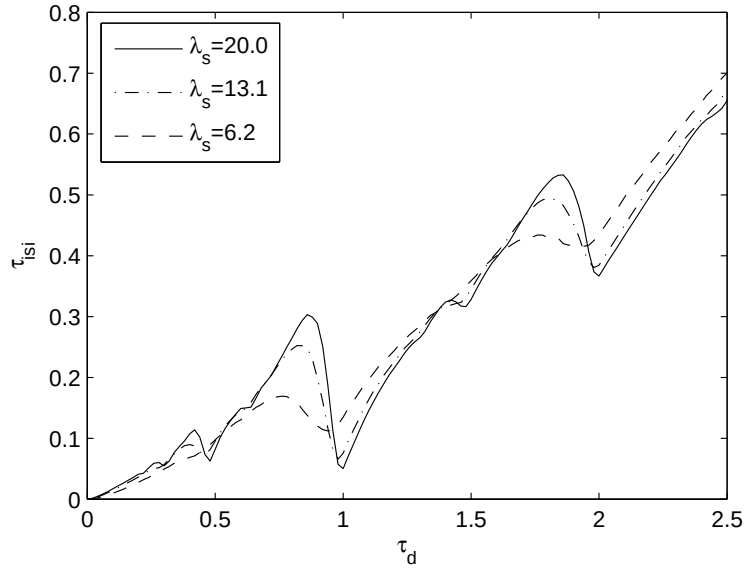


Figure 4.9: Average ISI as a function of deadtime

has jumped to the next integer. In addition, as it can be seen, whenλ_s increases the amplitude of these features also increases.

These features can be best described by an example which has a pattern of "011" as show in Figure 4.10. The x-axis represents time and the y-axis represent the inactivity of the detector. The first zero within this pattern ensures that no ISI occurs at the second bit, as it is not affected by ISI as long as $\tau_d < 2$. In order to exaggerate the features observed in Figure 4.9, it is assumed that the detected photon rate is high enough that the first photon within the slot arrives very close to the beginning of the slot. Given the probability of detecting one or more photon in the slot is:

$$p(n > 0) = 1 - p(0) = 1 - e^{-\lambda_s} \tag{4.15}$$

for example when $\lambda_s = 20$, there is 90% chance that a photon is detected at the first 11% of the slot. This probability increases as the λ_s increases. In Figure 4.10-a when $\tau_{d1} \approx 0.5$, a maximum of two events can be detected within the slot. The first photon which is detected at the beginning of the slot makes the detector inactive for approximately half of a time slot. The remaining time to the end of slot, τ_{g1}, is long enough for another event to occur. Once an event is detected during this remaining time, an ISI occurs for the third slot. As the deadtime increases, τ_{isi} further increases in the third slot. In Figure4.10-b when $\tau_{d2} \approx 0.7$, the remaining time τ_{g2} has decreased, however the probability of detecting a photon within τ_{g2} is

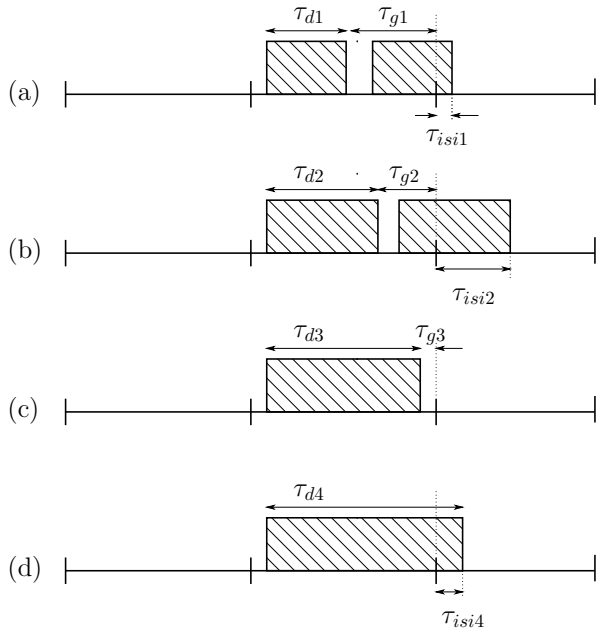


Figure 4.10: An example effect of ISI with increasing deadtime

approximatively equal the probability of detecting a photon in τ_{g1}. As a result, the increased deadtime leads to an increased ISI in τ_{isi2} in the third slot. In Figure 4.10-c when $\tau_{d3} \approx 0.9$, the reducedτ_{g3} decreases the probability of detecting a photon , as a result, the average ISI is significantly reduced as the third slot does not have an ISI. In in Figure 4.10-d, once $\tau_{d4} > 1$, the deadtime overflows into the third slot and τ_{isi4} starts to increase.

In the example in Figure 4.10, the error ratio in the second bit is not affected by deadtime as always $\tau_{isi} = 0$, unless the deadtime is larger than a slot duration. The error in the third bit is effected by the deadtime, as inactivity of the detector at beginning of the slot reduces the probability of detecting an event.

4.3.5 Effect of deadtime on BER

Figure 4.11 shows the BER as a function of deadtime for selected average signal photon rate when there is no background noise. As Figure 4.11.a shows, for relatively small deadtimes the BER is mostly unaffected. The condition which fulfils Equation 4.14 is shown as a dashed line. It should be noticed that this condition is only an approximation. Figure4.11.b shows the same results in Figure 4.11.a but in a linear scale of deadtime. The BER has a direct relationship with deadtime, until it approaches unity. The deadtimes where abrupt changes occur in the BER, correspond to the features in the average ISI trend shown in Figure 4.9. For example, when

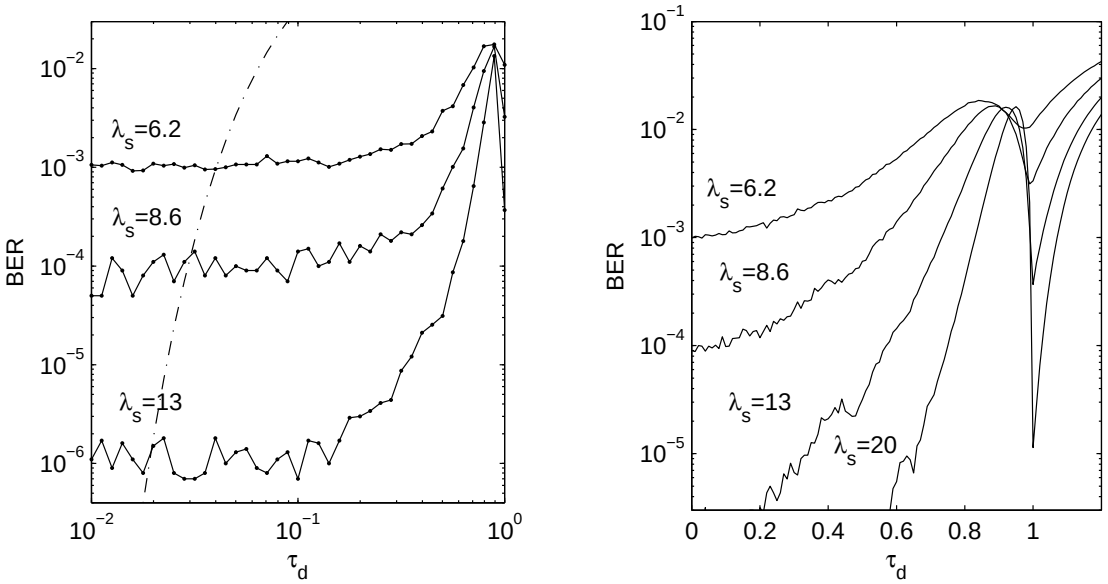


Figure 4.11: The BER as a function of relative deadtime for selected λ_s. the relative deadtime is in (a) logarithmic scale and (b) linear scale. The dashed line is the condition in Equation 4.14.

$\tau_d \approx 0.5$ the rate of change in BER increases. This point correspond to the small feature at the same deadtime in Figure 4.9. When $0.9 \lesssim \tau_d \lesssim 1$, the BER sharply decreases, and increases again when $\tau_d > 1$. This range of deadtime is where the large change occurs in $\langle \tau_{isi} \rangle$in Figure 4.9.

It can be seen in Figure4.11 that in order to maintain a particular BER, more photons are required for a longer deadtime. Figure 4.12 shows contour traces of BER for variable deatime and signal photon rate. As expected when the deadtime increases the average photon rate required to maintain the BER also increases. The difference between the photon rate required to maintain the BER with a deadtime, and the photon rate required to get the same BER with no deadtime, can be referred to as the power penalty for that deadtime. For a specific BER, a larger deadtime requires a higher power penalty. As the deadtime approaches $\tau_d = 1$, the power penalty sharply increases, and for $\tau_d > 1$, it abruptly reduces and increases again in a narrow range of deadtime. This is similar to the features seen in Figure 4.11.b.

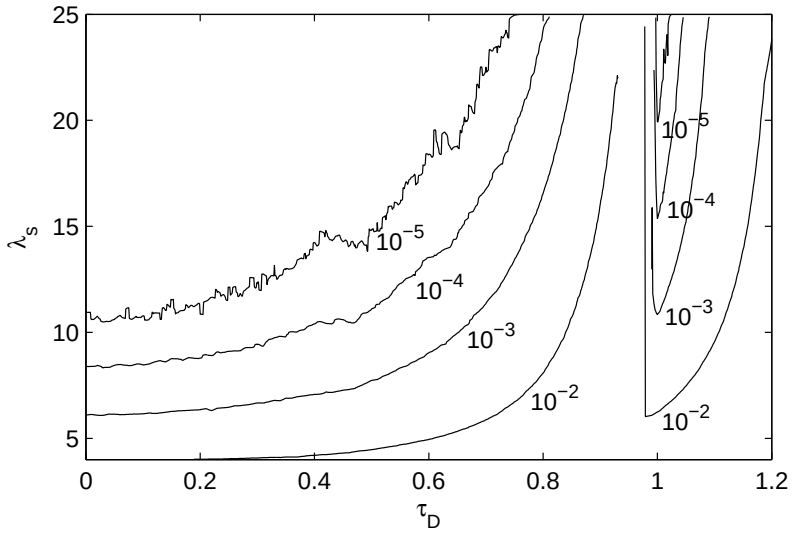


Figure 4.12: The BER contours as a function of relative deadtime and average signal photon rate

4.3.6 Deadtime and background noise

In the case when background noise exist, $\lambda_b > 0$, then the BER is function of λ_s, λ_b, and τ_d. Figure 4.13 shows the contour traces for the $BER = 10^{-3}$ as function of relative deadtime and average signal photon rate for selected background noise. When $\lambda_b = 0$, it is assumed that no photons are detected in a transmitted 0 bit, hence no ISI is created within 0 bit, hence the BER contour trace is identical to the trace shown in Figure 4.12 when $BER = 10^{-3}$. For small λ_b which satisfies the condition in Equation 4.6, the effect of background noise on the BER is negligible, hence the BER is only function of deadtime. In Figure 4.13 although $\lambda_b = 10^{-3} > \frac{1}{5}.BER$, the contour line has a small difference with the contour which has no background noise.

However, when background noise is much larger than the condition in Equation 4.6, The background noise further increases the BER due to the detected events in a 0 bit, which may lead to an overflow of deadtime into the next adjacent slot. The contour line for $\lambda_b = 10^{-2}$ and $\lambda_b = 10^{-1}$ show that a power penalty is required to maintain a specific BER. The overall power penalty required to compensate for deadtime and background noise, is an combination of both the power penalty for noise described in Figure 4.4, and the power penalty for deadtime described in Figure 4.12.b. As a result, when neither the condition in Equation 4.6 for background noise, nor the condition in Equation 4.14 for deadtime is satisfied, then overall the power penalty for both deadtime and background noise is larger than the power penalty required for either of them individually.

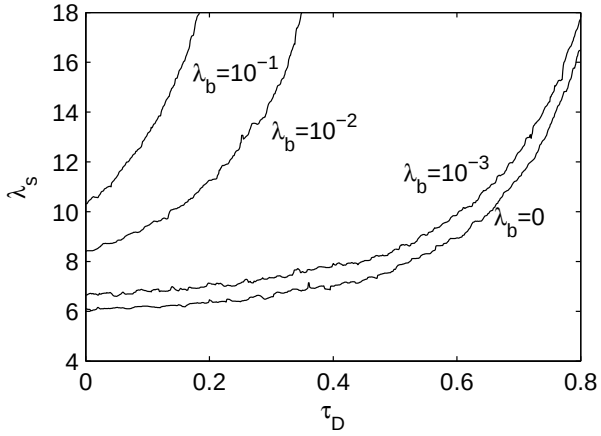


Figure 4.13: Effect of deadtime on the BER with background noise: (a) The relation between BER and average signal photon rate when $\lambda_b = 10^{-2}$ for selected deadtimes. (b) The contours for $BER = 10^{-3}$ as a function of relative deadtime and average signal photon rate for selected background noise level.

4.4 Array of detectors

The sensitivity of a photo-detector is increased with a larger collection area. However in practice, enlarging a detector's collection area poses few problems. As the active area of a conventional detectors increases its capacitance also increases. The increased capacitance leads to a longer rising and falling times, hence it is required to operate at lower speeds. In addition, enlarging the collection area also poses a reliability issue especially in APD and SPADs. As the surface area of the detector increases, there is higher probability that a defect appears in the collection area of the detector. These defects could alter the normal characteristics of the detector such as a higher dark count rate or afterpulsing. As a result, the production yield decreases as the chances of finding a detector with negligible defects decreases.

Therefore, in order to achieve larger collection area, and maintain the speed and the reliability of a smaller detector, an array is created of single detectors which operate independently. The output of the array is an aggregated output of each individual detector. In a photon counting detector, the output of each detector, which is the counted events within a time slot, n, is summed to form a larger photon counting detector, therefore:

$$N = \sum_{i=1}^{M} n_i \tag{4.16}$$

where n_i is the output of each detector, M is size of the array, and N is the total photons counted in the entire array within the time slot T_b. However, the benefit of

an array is not only limited to its increased surface area.

4.4.1 Array of detectors with no deadtime

Assuming a uniform intensity on the surface area of the detector, the average detected signal photon rate of a detector array, λ_{AS}, is defined as

$$\lambda_{AS} = \lambda_s \times M \qquad \Rightarrow \qquad \lambda_s = \frac{\lambda_{AS}}{M} \tag{4.17}$$

where λ_s is the average detected signal photon rate of a single detector within the array, and M is the size of the array. If there is no background noise, then the BER only depends on:

$$BER = \tfrac{1}{2}.e^{-\lambda_{AS}} \tag{4.18}$$

$$= \tfrac{1}{2}.e^{-M.\lambda_s} \tag{4.19}$$

As a result, if the transmitter's intensity is constant, then λ_s is constant, hence when the size of the array increases, λ_{AS} also increase, leading to a better BER. On the other hand, if a constant BER is desired, a specific λ_{AS} is required to achieve that BER. Hence, when the size of the array increase, the average detected photon rate for each detector within the array decreases, leading to a lower transmitter's intensity. Therefore, increasing the size of the array either improves the BER if the intensity is constant, or requires less intensity per single detector if the BER is constant.

For example , for a single detector with $\lambda_s = 6.2$, the BER is 10^{-3}. If an array of $M = 4$ is made of this single detector, then $\lambda_{AS} = \lambda_s \times M = 6.2 \times 4 = 24.8$. Hence, the BER calculated by Equation 4.18, is $\approx 10^{-11}$.

However when $\lambda_b > 0$, the overall background noise increases when the size of the array increases. As discussed in section 4.2, the background noise is a combination of background illumination and dark counts of the detector. Assuming a uniform background intensity, and identical DCR for each detector within the array, similar to Equation 4.17, the overall background noise for an array, λ_{AB}, is defined as:

$$\lambda_{AB} = \lambda_b \times M \tag{4.20}$$

where λ_b is the average background noise for a single detector. As a result, the

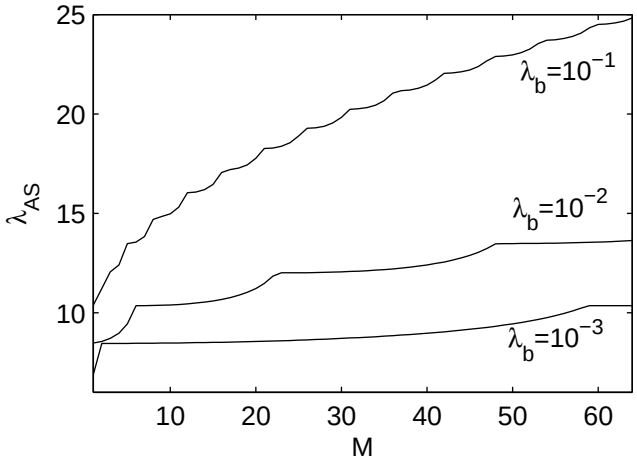


Figure 4.14: The $BER = 10^{-3}$ contours as a function of total photon rate and size of the array for selected background noise of a single detector

BER is a function of overall detected signal photons, and background noise:

$$BER = f(\lambda_{AS}, \lambda_{AB}) \quad (4.21)$$

$$= f(\lambda_{AS}, \lambda_b, M) \quad (4.22)$$

The Equation 4.22 shows that the BER can be estimated by the overall detected signal photons λ_{AS}, background noise of a single detector λ_b, and the size of the array, M. The parameter λ_{AS} is selected rather than λ_s because the former parameter directly leads to the estimation of BER based on Equation 4.18 or Table 4.1 independent of the size of the array.

For example, for a single detector with background noise of $\lambda_b = 10^{-3}$, the average number of detected photons for a single detector to achieve a BER of 10^{-3} is $\lambda_s = 6.9$. If an array is made of this single detector with a size of $M = 100$, then the overall background noise is $\lambda_{AB} = \lambda_b \times M = 10^{-3} \times 100 = 0.1$. In order to maintain the BER, from Figure 4.2 the required λ_{AS} is 10.4. As a result, although the overall background noise has increased by 100 times, the overall detected signal photon rate has increased by less than 2 times. That means the new λ_s is $\lambda_{AS}/M = 10.4/100 = 0.104$, hence the intensity per single detector is reduced by ≈ 66 times.

Figure 4.14, shows the $BER = 10^{-3}$ contours as a function of λ_{AS} and M for selected λ_b, based on Equation 4.22. It can be observed that although the background noise increases proportionally to the size of the array, the amount of overall photon rate λ_{AS} required to maintain the BER is less than $\lambda_s \times M$. Therefore, even with background noise, a better performance is achieved either by using an array of photon

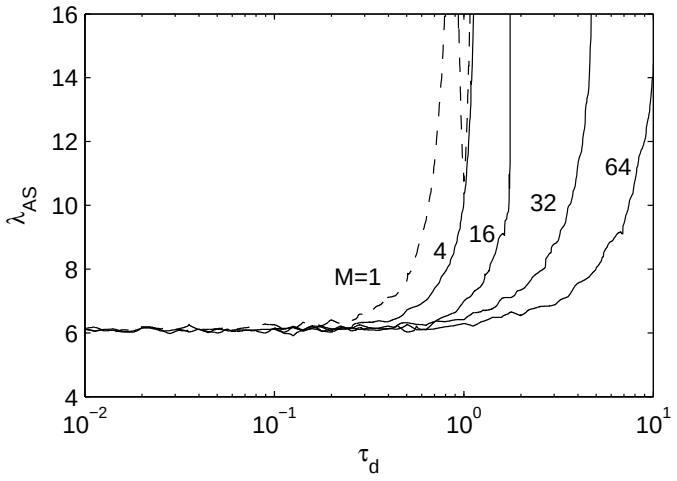


Figure 4.15: The $BER = 10^{-3}$ contours as a function of average detected photons of the array and deadtime of a single detector for selected size of the array.

counting detectors due to decreased the BER or reduced intensity per single detector.

4.4.2 Array of detectors with deadtime

It was found in Equation 4.12 that the probability of inter symbol interference is related to the product of average detected photon rate, λ_s and the relative deadtime, τ_d. In a high data rate communication link which $T_b \lesssim 5\,ns$, the relative deadtime approaches unity, or is larger than unity. For example, a SPAD with a deadtime of 5 ns, for a data rate of 1 Gbps $\tau_d = 5$. In order to maintain a low probability of ISI, Equation 4.14 estimates that $\lambda_s \ll 1$. For such a low intensity, although the probability of ISI is small, it is impossible to achieve a reasonable BER. However, in an array of photon counting detectors with a relative deadtimes of τ_d, the BER depends on the total detected photons within the array. Hence, in order to maintain a particular BER, the larger the array, a smaller λ_s is required which leads to fulfilling the low probability ISI condition. As a result, it is possible to achieve a reasonable BER with an array of detectors when each detector has a relative deadtime which is larger than unity.

Practically when deadtime is larger than the bit duration, $\tau_d > 1$, a single detector could be inactive for several bit slot durations, hence it is not able to detect the following transmitted data. When another detector is added in parallel, it is able to detect photons during the inactivity period of the first detector. As a result, the chances of detecting a signal photon increases while the size of the array increases. Figure 4.15 shows the $BER = 10^{-3}$ contours as a function of relative deadtime, τ_d,

and average array detected photon rate, λ_s, for selected array sizes, M. It is assumed that there is no background noise. When $M = 1$, the contour is identical to the $BER = 10^{-3}$ contour in Figure 4.12.

It can be observed that as the size of the array increases, the maximum deadtime at which the BER is independent of deadtime, also increases. In addition, as the size of the array increases the maximum deadtime that the power penalty for that particular deadtime decreases. For example, the power penalty to achieve a $BER = 10^{-3}$ with a single detector with $\tau_d = 0.5$ is 1.6 photons. When $M = 4$ the power penalty to achieve the same BER is 0.5 photons. On the other hand, for a power penalty of 1 photon, with an array of 4 detectors, a deadtime of $\tau_d = 0.44$, with an array of 16 detectors, a deadtime of $\tau_d = 1.1$, and with an array of 64 detectors, a deadtime of $\tau_d = 7.2$ can be achieved.

4.5 Experimental results

In order to experimentally investigate the effect of inter symbol interference on the bit error ratio, a series of proof of concept experiments are performed. These experiments use the SPAD circuit with the automatic diode reset described in the previous chapter, as a photon counting detector with a variable deadtime.

4.5.1 Experimental configuration

Figure 4.16 shows the experimental configuration for bit error ratio tests (BERTs). A fixed or random bit pattern is generated and transmitted optically in free space using ON-OFF keying modulation. The response of the chip is processed and compared to the originally transmitted data to find the bit error ratio of the communication link.

The bit pattern is generated by a Xilinx Spartan-6 FPGA kit. Fixed patterns are per-defined and programmed into the FPGA. The bit duration is controlled by internal clock management units which multiply and dived the original on-board clock to create the desired bit rate. The pseudo random bit pattern is generated in real-time by a linear feedback shift register (LFSR) [78]. The implemented 31-bit LFSR consist of a 31 bit shift registers with an XNOR in the feedback loop which connects the output of the 28th and 31st shift registers to the input of the first shift register. The LFSR has a repetitive pattern with a maximum sequence length of $2^{31} - 1$ bits. In addition to the data signal generated by the pattern generator, a sync signal is also generated in order to assist the data capture system to detect the beginning of

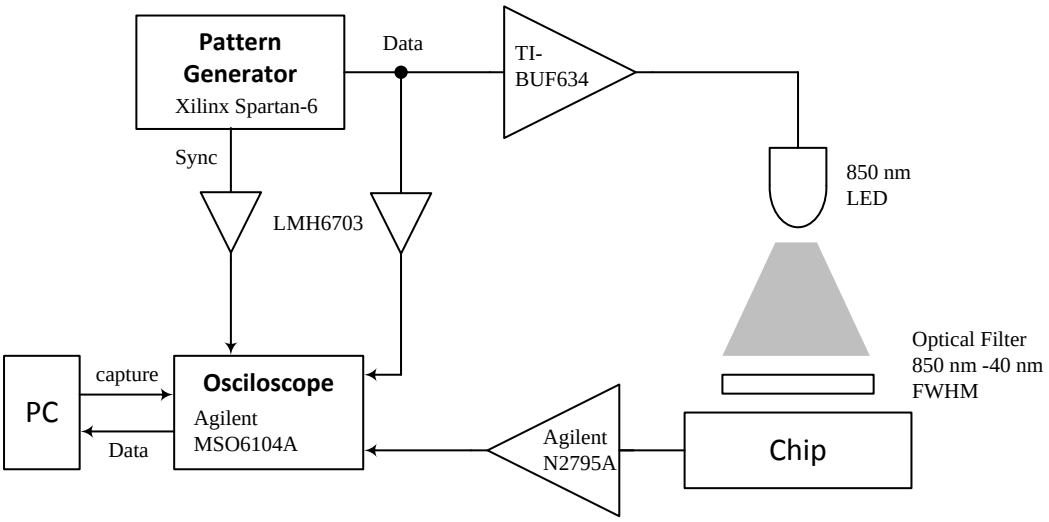


Figure 4.16: The experimental configuration for bit error ratio tests

a bit slot.

The output data signal of the pattern generator is connected to a Texas Instrument BUF634 unity gain voltage buffer. This buffer has a maximum drive capability of 250 mA, and a 180 MHz bandwidth. As a result, this buffer is suitable for driving LEDs which require large currents due to their large capacitance. The output of the voltage buffer is connected to a communication wavelength LED which has 850 nm centre wavelength, and a rise/fall time of 10 ns. The LED is driven by a 10 mA current when it is transmitting a logic 1, and no current when it is transmitting a logic 0.

A band pass optical filter with 850 nm peak wavelength transmission, and a 40 nm full width half-maximum (FWHM) transmission bandwidth is used to minimise the background noise during the experiments. This filter has a peak transmission of 70%, and up to four orders of magnitude attenuation in the blocking band. The chip, which was described in Chapter 3, is mounted onto a generic printed circuit board. The output of the chip is connected to an Agilent N2795A active probe which has a bandwidth of 1 GHz, input impedance of 1 $M\Omega$, and input capacitance of 1.0 pF. The high input impedance and low input capacitance of the probe ensure minimum drive current for the output buffers of the chip, which leads to a better signal integrity of the output signal. The data is captured by an Agilent MSO6104A oscilloscope which has a 1 GHz bandwidth and a 4 GSps (= samples per second) sampling rate. In addition to the output of the chip, the sync signal and the original data signal, which are generated by the pattern generator, are connected to the oscilloscope. The

oscilloscope applies appropriate delays to each signal in order to compensate for the delay in the signal path.

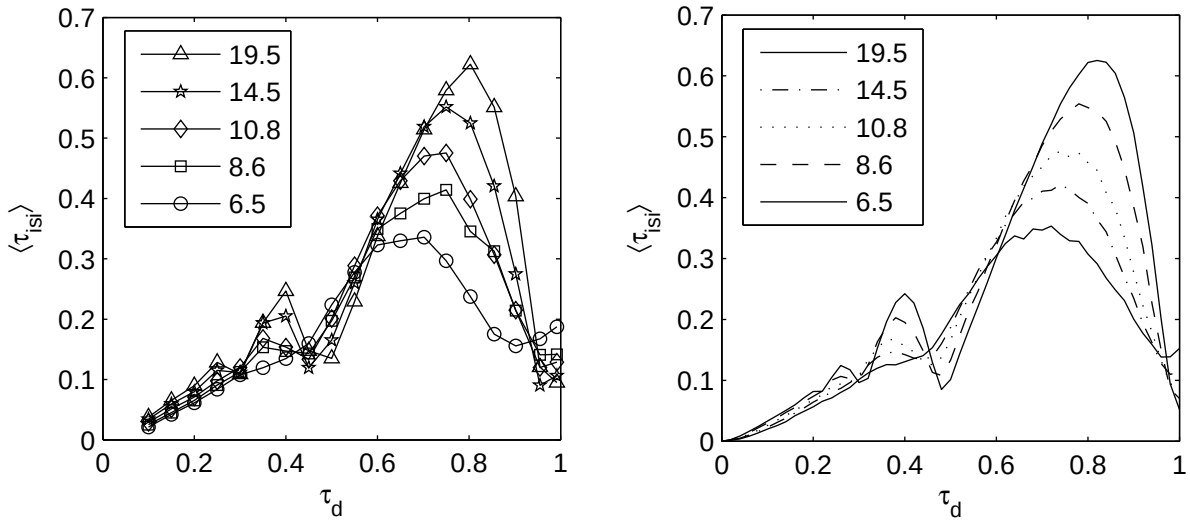
The calculation of BER is performed using an offline method that means only sections of the generated bit sequence are captured by the oscilloscope. A Labview script running on a personal computer (PC) sends a capture signal to the oscilloscope to capture one bit or more in the sequence. Each captured frame contains the sync signal, original data, and output of the chip. The captured frame is then processed in Labview and Matlab, and the original data is compared to the output of the chip to determine the erroneously received bits. The capturing process is repeated until enough number of errors are detected to provide and accurate measurement results for the BER test.

4.5.2 Average ISI

The average ISI,$\langle\tau_{isi}\rangle$, is the average deadtime overflowing into the next bit slot. In this experiment, a fixed pattern of 010 is used to investigate the effect of deadtime on the average ISI. By varying deadtime the amount deadtime which overflows into the third bit slot of the 010 pattern changes. Duration of each bit slot is 100 ns, and the deadtime is varied form 10 ns to 100 ns in approximately 5 ns steps. Hence, the relative deadtime, τ_d, changes from 0.1 to 1. Each 010 pattern is captured as an oscilloscope frame. An avalanche event is detected by the falling edge of the output signal of the chip. The result discussed in Section 4.3.4 are without background noise. However, the SPAD circuit in the chip under test has a background noise of approximately 12 kcps, that means the average background noise in duration of a bit slot is $\approx 10^{-3}$. In order to eliminate the effect of background noise on the measurement results, if an avalanche is detected in the first bit slot of the pattern, the entire captured frame is dropped and will not be processed.

The avalanche events which are detected in the second slot in the pattern (the 1) may overflow into the third slot. In each captured frame, the time difference between the start of the third slot and the rising edge of the SPAD output is measured. The rising edge indicates that the deadtime is elapsed, and the SPAD is out of its inactivity period. This time is the τ_{isi} for that particular frame. The average ISI is calculated by repeating the frame capture. If no rising edge is detected in the third slot, that means no deadtime has overflowed, hence $\tau_{isi} = 0$.

Figure 4.17.a shows the experimental average ISI versus relative deadtime for selected intensities. Having the event rate from the experimental results, the detected



(a) (b)

Figure 4.17: The average ISI versus relative deadtime for selected photon rates (a) experimental (b) simulation

photon rate is estimated using Equation 4.10. As expected the features which was discussed in Section 4.3.4 and Figure 4.9 also appears in the experimental results. As the photon rate increases, the intensity of these features also increases. Figure 4.17.b shows the simulated average ISI for the estimated photon rates extracted from the experimental results. The experimental and simulation result are in close agreement.

4.5.3 Error ratio of pattern 011

In the previous experiment, although the third bit slot in the 010 pattern is affected by the ISI, no error is caused as a 0 is transmitted in the third bit slot. An error is caused when the third slot is a transmitted 1, hence a 011 pattern. The 0 in the pattern ensures that the second slot within the pattern (the first 1) is not affected by ISI as the relative deadtime is smaller than unity. However, the third slot within the pattern (the second 1) is affected by ISI due to the overflow of the deadtime from the second slot (the first 1). As a result, the SPAD is inactive at the beginning of the third slot leading to a lower probability of detecting an event. Hence, more errors occur in third slot due to ISI. In this experiment, the effect of deadtime on the error ratio of the third slot is investigated.

Similar to the previous experiment in section 4.5.2, if an event is detected in the first slot (the 0), the entire frame is dropped, hence the experiment is not affected by

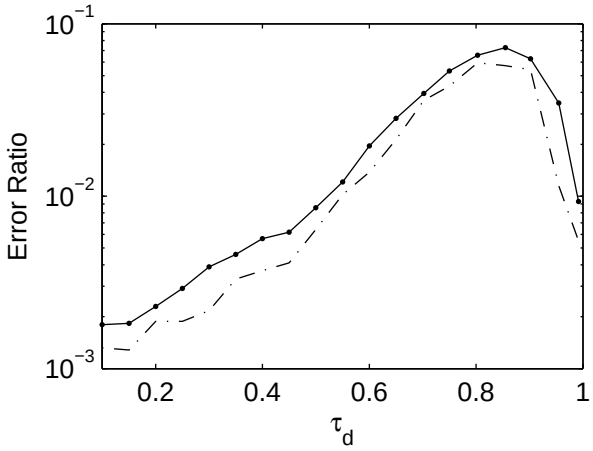


Figure 4.18: The error ratio for the third bit in the pattern 011,ER_3, versus relative deadtime (solid) experimental (dashed) simulation

background noise. The probability of error in second slot (the first 1) is not affected by deadtime, and it is same as a detector with no deadtime. Having the event rate from the experimental results, for an estimated $n = 6.5$, (based on Equation 4.2) the error ratio for the first 1 is:

$$ER_2 = p(n = 0) = e^{-6.5} = 0.0015 \tag{4.23}$$

However, for the third slot (second 1) as the ISI increases the error rate increases. Figure 4.18 shows the experimental and simulation error ratios for the third slot (second 1) ER_3 versus relative deadtime. When relative deadtime is small as $\tau_d = 0.1$, the error ratio is similar to the error ratio of the second slot ER_2. However, as the deadtime increase, the error ratio increase until $\tau_d \approx 0.9$, when error ratio sharply drops. This feature is consistent with the average ISI results discussed in section 4.5.2.

The difference between the experimental and simulation results are likely due to the experimental inaccuracy. Each output pulse of the SPAD has a rise time and fall time of 6.8ns (10%-90%). Due to the relatively slow rise time and fall time, when two pulses are too close to each other, they merge and form an extended pulse instead of two individual pulses. The extended pulse is approximately twice longer than the deadtime. However, this pulse extension is different with pulse extension shown in Figure 2.8, which occurs if the diode is quenched by a resistive method (a resistor or current limited transistor). In the case of automatic diode reset, the pulse duration is always equal to the deadtime. As the pulse detection is based on the edges of the signal, an extended pulse distort the detection of an avalanche event. As

a result, a complementary algorithm is deployed that in addition to detecting edges of each pulse, the length of each pulse is also taken into consideration. Knowing the deadtime, if a pulse is longer than the deadtime, the edge of the pulse which has been disappeared due to the slow rise/fall times is estimated by the post-processing algorithm. This algorithm significantly improves the count loss due to insufficient slew rate of the output signal.

The fluctuations observed in both experimental and simulation results in Figure 4.18 are due to limited number of runs for both experimental and simulation results. In order to have more accurate results, more runs are required, however, more runs require more time to perform the experiment or the simulation. As a results, there is trade off between number of runs and the time to perform.

4.5.4 Error ratio of random pattern

The effect of deadtime on the error rate of a random pattern is less significant than the fixed pattern of 011. The ISI only affects the bit error rate if a transmitted 1 is flowed by another 1. In the case of the 011 pattern, the pattern is deliberately defined so that ISI increases the error ratio in third slot (second 1). However, in a random pattern, a transmitted 1 has equal probability of being followed by a 0, which will not affect the error ratio, and a 1, which will affect the error ratio.

Unlike previous experiments, the effect of background noise cannot be eliminated using the previous method due to the randomness of the pattern. However, as described in Equation 4.1, the term $p(T1).p(R0|T1)$ is related to signal and the term $p(T0).p(R1|T0)$ is related to the background noise. As a result, in this experiment only the probability which is related to the signal $p(R0|T1)$ is considered. The Average event rate is the same as the previous experiment estimated at $n = 6.5$, hence as a result the error ratio of a detector with no deadtime is:

$$ER(R0|T1) = e^{-6.5} = 0.0015 \tag{4.24}$$

Figure 4.19 shows the experimental error rate for transmitted 1s which are received as 0s, $ER(R0|T1)$, versus relative deadtime. Similar to the simulation results shown in Figure 4.18, as the deadtime increases the probability of errors increases. When relative deadtime is close to unity, the error ratio slightly improves as the average ISI decreases. As expected the intensity of the rise and fall in the error ratio is less than the error ratio in the fixed pattern 011 shown in Figure 4.18 due to randomness of the pattern. Similar to the previous experiment, the difference between experimental

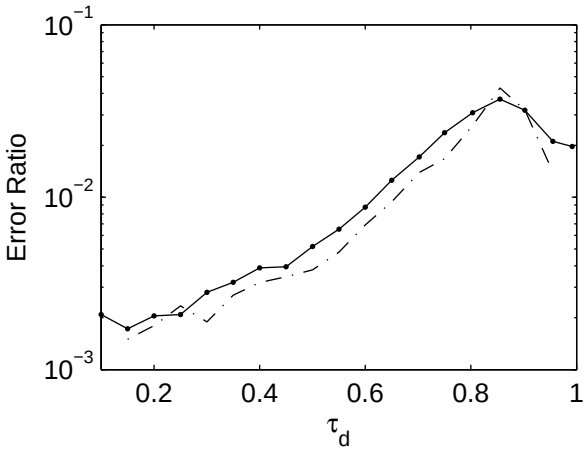


Figure 4.19: The error ratio for transmitted 1s which are received as 0s, $ER(R0|T1)$ for a random pattern versus relative deadtime. (solid) experimental (dashed) simulation

and simulation results is due to experimental inaccuracy and insufficient rise/fall time of the output signal of the SPAD circuit. The same correction algorithm which compensates for the slow rise/fall times is also applied in this experiment.

4.6 Summary

In this chapter, the performance of a communication link for photon counting detectors was assessed. Initially, the performance of a detector with no deadtime based on Poisson distribution was investigated. In this case the BER only depends on the average detected signal photon rate and the average background photon rate. The power penalty for the background noise which is the amount of extra intensity required to compensate for the increase of the BER was described. In addition, the maximum background noise which does not affect the BER was estimated. In practice in photon counting detectors, the overflow of deadtime into the next bit slot leads to an inter symbol interference, hence increasing the BER. When the deadtime is a small fraction of the bit slot duration, the ISI is negligible, hence the BER is independent of the deadtime. When the deadtime is comparable to the slot duration, similar to the background noise, a power penalty is required to maintain a target BER.

An array of single counting photon detector improves the photon collection area which leads to a greater sensitivity or better BER. The output of an array is an aggregated photon count of individual detectors. It was found that when the size of the array increases, although the total background noise increases, an overall improve-

ment in the performance is achieved. In addition to an increases in the collection area, larger arrays have the capability of utilizing single detectors with longer deadtimes while maintaining their BER performance. A series of experiments were conducted to investigate the features which appear in the simulation results. These proof of concept experiments show that the simulation results are in close agreement with the experimental results.

Chapter 5

Current Steering Read-out Array

In the previous chapter, it was found that an array of SPAD improves over a single SPAD by increasing the collection area, and reducing the effect of deadtime on the bit error ratio. The array is made of asynchronous SPADs cells which operate independently by having their own automatic diode reset circuit. The overall photon count of the array is an aggregation of the counts from each individual cells. A basic method to readout the number of detected events in each cell is to separately connect each cell to external circuits. This creates a digital output bus from the chip which contains the information from each cell. As a limited number of output connections are available, a digital bus is only practical for relatively small numbers of arrays.

In order to overcome the limited number of connections and to provide a scalable array of SPADs, a readout mechanism is required. This readout mechanism could include a digital counter added in each cell to store the number of events of the SPAD. A global readout circuit within the chip then reads out the value of each counter in a synchronous readout operation to the external digital interface. The obtained counts for each SPAD are then later added to an overall sum by a digital processor [46, 50]. In order to readout large arrays at high speeds using this mechanism, high speed global clocks are required which will be challenging to design in the digital domain.

In this chapter, an analogue readout method for the SPAD arrays is proposed which is based on high speed digital to analogue converters (DACs). The proposed analogue readout circuits (AROC) is explained after discussing basic analogue readout alternatives. The analogue readout mechanism enables high speed operation in large arrays.

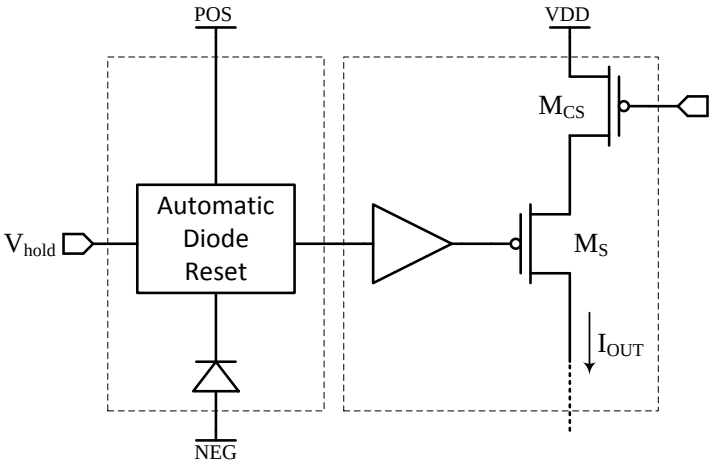


Figure 5.1: A basic SPAD cell with an analogue current output

5.1 Analogue readout-circuits

5.1.1 Basic method

Figure 5.1 shows a schematic diagram of a SPAD cell with an analogue readout. The cell contains an automatic diode reset (ADR) which was described in Chapter 3 and a analogue readout circuit. In the idle mode the output voltage of the ADR is a digital high signal. When an avalanche event occurs, the output of the ADR is a digital low signal for the period of the deadtime. As a result, the transistor switch M_S is conducting a current which is limited by the gate bias voltage of the pMOS current source M_{CS}. The output node is terminated by an off-chip resistor to convert the output current into voltage. The amplitude of the current is controlled by the gate voltage of M_{CS}. The gate voltage of the transistor which provides the current source is chosen to ensure that the transistor operates in its saturation region, hence the generated current is independent of the output voltage. When deadtime has elapsed, Ms is non-conducting, and the current flow through the external resistor is stopped. By monitoring the voltage across the terminating resistor, an avalanche event is indicated as a discrete voltage step which is proportional to the current of M_{CS}, and the value of the terminating resistor.

The independence of the readout circuit from the diode handling circuit ensures that the readout mechanism is not dependent on the diode's characteristics. In a previous implementation but with a resistive quenching circuit, a readout mechanism based on Figure 5.1 was utilised in an imaging application [17]. The current output was connected to a capacitor to integrate charge over a specific period of time. A reset

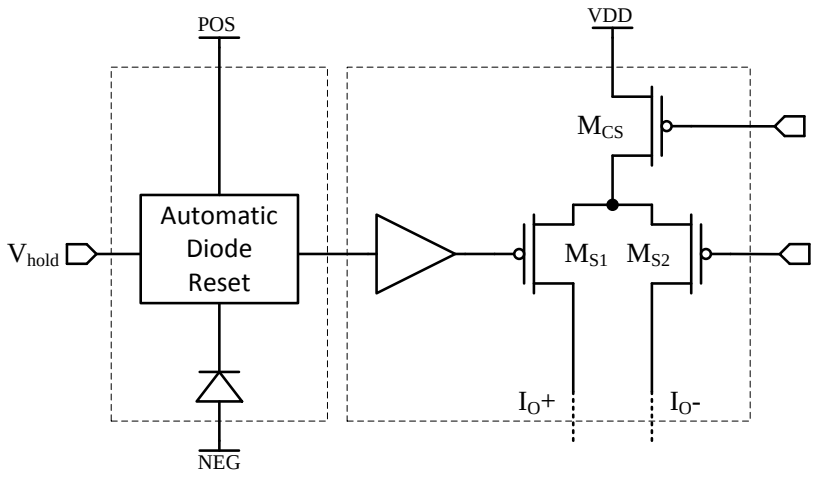


Figure 5.2: A SPAD cell with a basic differential current output

transistor was added to recharge the capacitor. The number of event was proportional to the amount charge stored in the capacitor at the shared output node.

However, in this design, all SPAD cells within the array, share the same output, hence the total output current is sum of the individual currents of each SPAD cell. The voltage steps at the output indicate the activity of a SPAD cell. When the output voltage steps up that means an avalanche event is detected. When it steps down a SPAD cell has recovered. The amplitude of the output voltage at any time indicates the number of SPAD cells that are currently inactive because they have recently detected an avalanche event.

5.1.2 Differential method

Although the circuit in the Basic AROC is suitable for applications such as imaging and sensing, it is not optimal for high speed performance. In particular, as each current source is connected to the output the current flowing into the circuit shown in Figure 5.1 fluctuates. This constant switching action causes a signal integrity problem on the power supply rails. In order to provide a constant current load on the power supply rails, a differential output mechanism is used. Figure 5.2 shows a basic SPAD cell with a differential current output. A second transistor switch is added to constantly sink the current provided by M_{CS}. When an avalanche event occurs, MS1 conducts, and the current of M_{CS} is shared between both M_{S1} and M_{S2}. By selecting an appropriate bias voltage for the gate of M_{S2}, current fluctuations on the power supply rails are minimised.

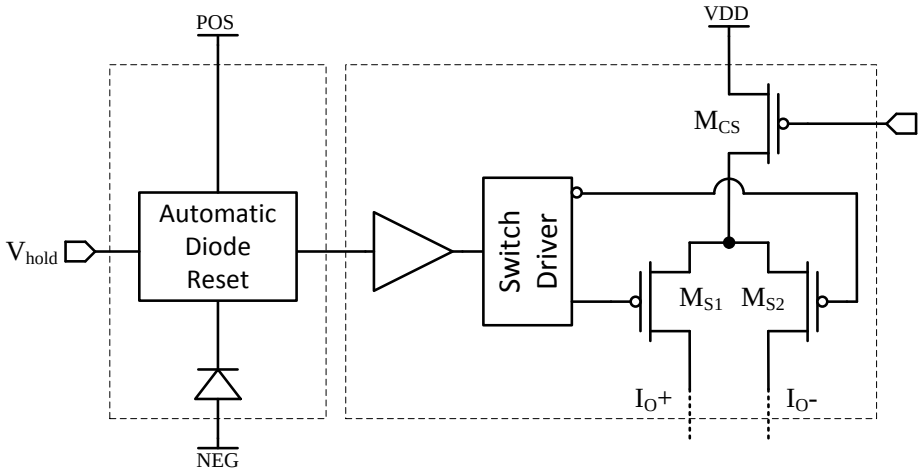


Figure 5.3: A high performance current steering read-out circuit

5.1.3 Current steering method

In high speed applications, the impedance of the switches and the asymmetry of the switching signals in the previous circuit shown in Figure 5.2 causes glitches at the output of the circuit. To avoid these problems, differential current steering switches are widely used in high speed digital-to-analogue converters. These switches are made of a matched pair of transistors, hence providing similar switching transient characteristics. A complementary differential signal drives these switches providing a complementary switching operation, and a current steering mechanism. The complementary current steering action minimises the glitch which occurs during high speed switching operation. A switch driver circuit provides the symmetrical complementary switching signals required for the switches [79, 63, 80, 81].

Figure 5.4 shows the signalling operation of the Switch Driver. The input voltage to the switch driver is a single ended output from the output buffer of the automatic diode reset. In order to avoid a high impedance node, the amplitude of the signal which drives the switching transistor is limited to a low voltage. This ensures that the transistor is never fully turned off. As a result, the output current swings between a minimum and a maximum current, which is determined by the Switch Driver unit.

Figure 5.3 shows the schematic diagram of the proposed SPAD cell which includes an automatic diode reset and a current steering readout mechanism. The output of the buffer is connected to the switch driver which creates a differential pair of signals to drive the current steering switches. When an avalanche event occurs, the current is steered from the negative output to the positive output. The differential current output is usually terminated by a pair of resistors which converts the current into

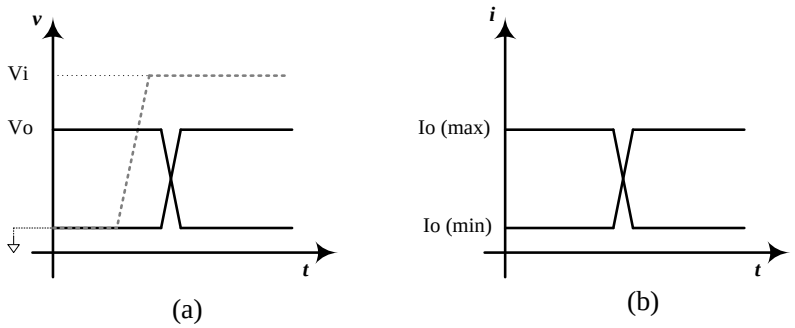


Figure 5.4: The role of the switch driver in signalling

voltage.

When the SPAD cell is in idle mode, the gate voltage of M_{S1} is high and the gate voltage of M_{S2} is low. As a result, less current flows through M_{S1} than through M_{S2}, and the majority of the current is flowing through the negative output. When an avalanche event occurs the switch driver causes more current to flow through M_{S1} than through M_{S2}. As a result, the current is steered from the negative output to the positive output. Once the deadtime is elapsed, the state of the switch driver changes, and the current is steered back from the positive output to the negative output.

5.2 Implementation of the current steering readout circuit

The current steering readout system which is implemented in the SPAD cell is based on the commonly developed equally weighted current steering digital-to-analogue converters (DACs) [66]. The static performance of these DACs is related to matching of the current sources within in each cell. The accuracy of the current source is limited by random and structural mismatch in the fabrication process. In addition the overall output has a finite output impedance which leads to a non-linearity in the output. Their dynamic performance is limited by the glitches which occur during the switching operation at the output. These glitches are important during high speed operating conditions as they distort the output of the array.

5.2.1 The current source

The current through the current source transistor is controlled by its gate voltage. The current of these transistors can sources or sinks current to the output, however,

for the intended application sourcing the current is more convenient as the output resistors are terminated to the ground. Additionally, it can benefit from utilising a trans-impedance amplifier which improves the performance of the current to voltage conversion at the output. Hence, a pMOS transistor sources the current from the supply rail to the output through the current switches. The pMOS current source within each cell is operated within its saturation region, hence the output current is weakly dependent on the output voltage. The output current equation for a long channel MOSFET is then:

$$I_D = \frac{\mu_p C_{ox}}{2}.\frac{W}{L}.(|V_{gs}| - |V_{th}|)^2 = \beta(|V_{gs}| - |V_{th}|)^2 \tag{5.1}$$

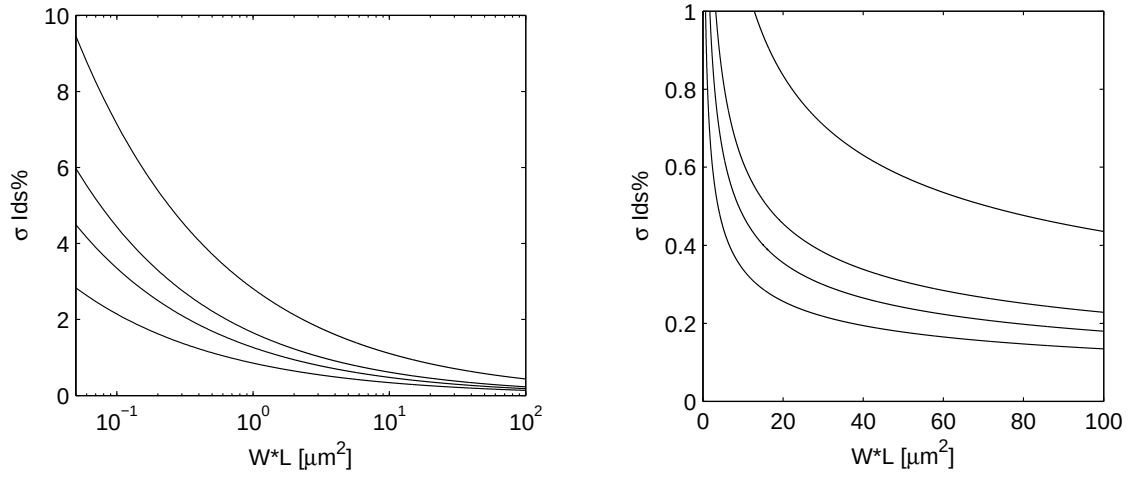
where μ_n is the carrier mobility, C_{ox} is the gate capacitance per unit area, W is the width, L is the length, V_{gs} is the gate-source voltage, and V_{th} is the threshold voltage. Although the current of each current source within the SPAD cell is intended to be equal, the statistical variations in the transistor characteristics leads to variation in each current source. When operating in the saturation region, the statistical variation in the drain current is dependent on the statistical variation of all the parameters within Equation 5.1. Within an array of SPAD cells, each cell is physically located in various positions in the silicon substrate. Several factors including variations in gate oxide thickness across the substrate causes threshold voltage variations and gate capacitance variations in each transistor. In addition, differences in each transistor's dimension especially due to rounded corner in smaller dimensions by lithography process, causes variation in the current. The variation in output current is estimated as [82, 83]:

$$\sigma^2\left(\frac{\Delta I}{I}\right) = \frac{2\sigma_I}{\bar{I}}\ , \qquad \sigma^2\left(\frac{\Delta I}{I}\right) = \sigma^2\left(\frac{\Delta\beta}{\beta}\right) + \frac{4\sigma^2(V_{th})}{(V_{gs} - V_{th})^2} \tag{5.2}$$

and

$$\sigma^2\left(\frac{\Delta\beta}{\beta}\right) = \frac{A_\beta}{W.L} \qquad \sigma^2(\triangle V_{th}) = \frac{A_{V_{th}}}{W.L} \tag{5.3}$$

where $\sigma_I/\bar{I}$ is the relative standard deviation of the current, $\sigma^2(x)$ is interpreted as the variance of the stochastic parameter x [83], A_β and $A_{V_{th}}$ are process dependent parameters. Equations 5.2 and 5.3 show that the relative standard deviation of current has an inverse relationship with the transistor area, and with the gate-source voltage. Figure 5.5 shows the relative standard deviation of the current versus transistor size of a 1.8V pMOS transistor provided by the experimental data from the



(a) (b)

Figure 5.5: The effect of transistor size on the relative standard deviation of the current. This data is for a 1.8V pMOS transistor.(a) the variation data plotted for a wide range of transistor sizes starting from the minimum size transistor (b) the possible transistor size is based on the available silicon area. From top to bottom $|V_{gs}|$ is 0.7, 1.0, 1.2, 1.8.

foundry. It can be seen that the minimum size pMOS transistor with a dimension of $W \times L = 0.24 \times 0.18\,\mu m$ has a relative standard deviation of approximately 3% at its highest gate-source voltage. As the silicon area dedicated for a SPAD cell is limited, it is reasonable to limit the size of the current source transistor comparable to the size of the other blocks within SPAD cell including the diode and automatic diode reset circuit. Figure 5.5-b shows the relative standard deviation for a practical transistor dimensions. It can be seen that when $W \times L < 20\,\mu m$ the σ_{Ids} significantly improves with increasing size. However when $W \times L > 20\,\mu m$ the increase in dimension has negligible improvement on σ_{Ids}. While the size is increased by 5 times, σ_{Ids} improves only twice. As the transistor is required to be in saturation the maximum gate-source voltage which is 1.8V cannot be used. On the other hand, a low gate-source voltage which is close to the threshold voltage is not desired due to a very large standard deviation in the current. Therefore, a transistor dimension which is about $20\,\mu m$ is a reasonable choice which has a maximum 0.5% relative standard deviation of current.

Additionally, each current source has a finite output impedance due to channel length modulation in the current source transistor. As a result, the output current is dependent on the output voltage which leads to non-linearity and degradation in the static performance of the readout circuit. Therefore the length of the current

source transistor should be taken into consideration as the channel length modulation decreases when a longer channel length is selected.

5.2.2 The current switches

The current switches M_{S1} and M_{S2} in Figure 5.4 steer the current of the current source M_{CS} into a common output node. Both of these transistors are a pMOS pair. The similar characteristics of both switches ensure a symmetrical performance during a switching operation. The transient switching characteristics of this pair determines the dynamic performance of the readout circuit.

During the switching operation the switching transistors change from a conducting state to a non-conducting state. This causes a voltage fluctuation at the drain of the current source M_{CS} . As a consequence, the parasitic capacitance on the output node of M_{CS} is charged and recharged by the voltage fluctuations. This parasitic capacitance is dominated by the drain capacitance of the M_{CS} which is proportional to the transistors width. These voltage fluctuations cause current fluctuations at the output node leading to glitches at the output which reduces the dynamic performance of the readout system. Although a current source with large dimensions leads to a better static performance, the large parasitic capacitance of the drain reduces the transient response of the readout system.

In addition, the switching signals which are applied to the gate of the current switches feed through the output due to the gate capacitance of the switches. As this degrades the dynamic performance of the readout circuit, it is desired to minimise the size of these transistors to ensure a minimum parasitic capacitance.

5.2.3 Selection of the transistor dimensions

The W/L ratio of the current switches should be large enough to avoid a high resistance switches which lead to a voltage drop, and drive the current source transistor out of saturation. As a result of considerations on reducing the parasitic capacitance of the current switches, and having a large enough W/L ratio a minimum length transistor with W/L=1/0.18 μm is used for the current switches . In order to minimise the channel length modulation in the current source transistor, a gate length of 2 μm is selected. Considering the effect of the current switches on the current of M_{CS}, a transistor width of 4.5 μm is chosen to ensure that a maximum current of 100 μA per current source is sunk form the output. As a result the entire array of 64 SPAD cell can sink up to 6.4 mA.

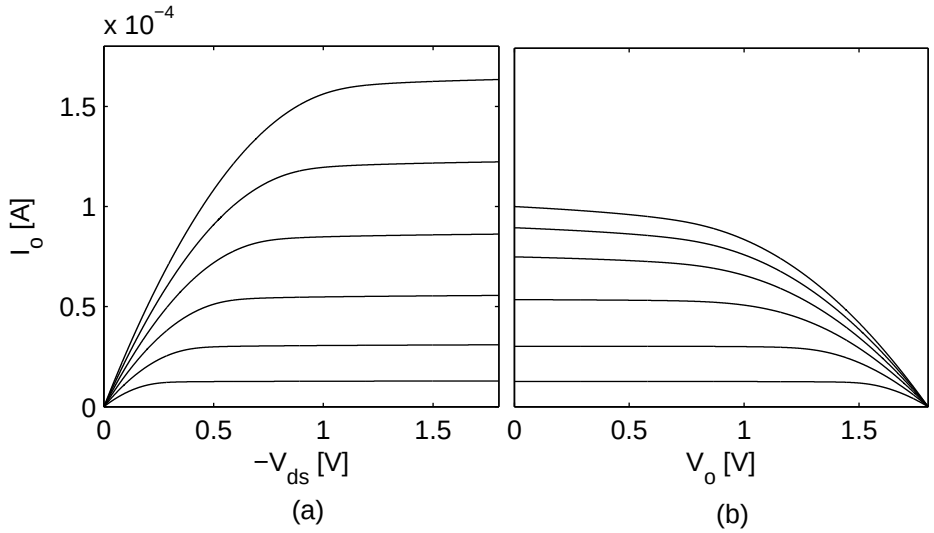


Figure 5.6: The I-V traces for (a) the current source transistor M_{CS} and (b) when M_{CS} is in series with the current switches (from top to bottom $|V_{gs}|$ is equal to 1.8, 1.6, 1.4, 1.2, 1.0, 0.8)

Figure 5.6-a shows the relationship between the current and voltage of the current source transistor. The current source transistor has a maximum 160 μA when biased at its maximum gate voltage. However, when in series with the current switches, the current of the current source is limited to maximum 100 μA. the solid line in Figure 6-b show the current voltage relationship when M_{CS} is in series with M_S. This figure shows that for gate voltages $|V_{gs}| > 1.4\,V$, the current source is not into saturation. As a result it is best to choose$|V_{gs}| \leq 1.2\,V$.

As a result of the selection of the transistor sizes, the output capacitance of a SPAD cell is equal to the drain capacitance of the transistor switch M_S, which is approximately 0.6 fF based on the process parameters provided by the foundry. This capacitance is significantly smaller than the capacitance of the avalanche diode estimated in Table 2.3.

5.2.4 The switch driver

The switch gate driver is made of two sections: the differential converter and the level shifter. The differential converter converts the single ended output of the automatic diode reset to a pair of complementary signals. The level shifter reduces the amplitude of the switching signals to achieve a better performance.

The output of the automatic diode reset is a single ended digital pulse. Both switches require a complementary switching signal. The symmetry in switching sig-

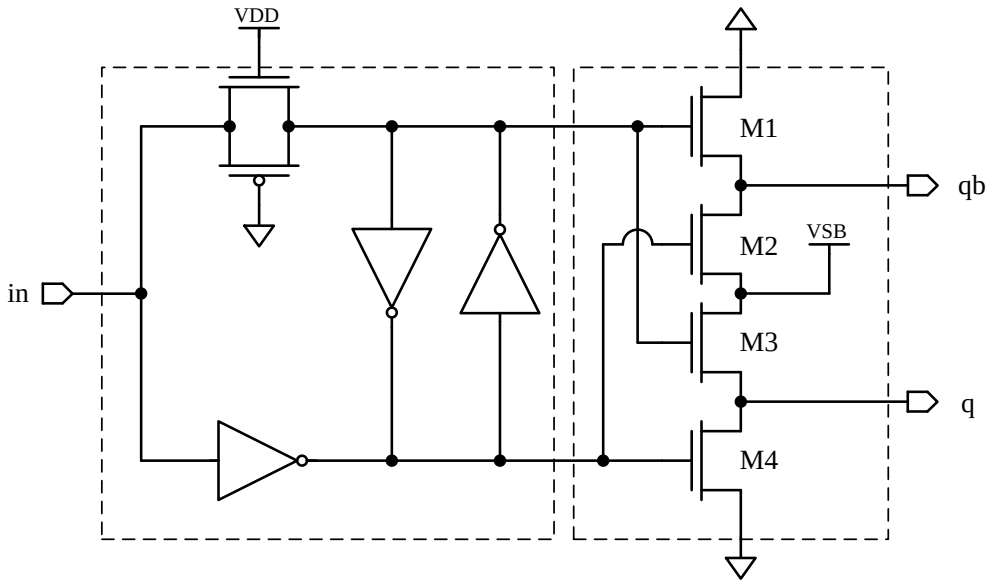


Figure 5.7: The switch driver: a differential converter stage, and level shifter stage

nals ensure a low distortion switching as both switches change their state simultaneously, hence the voltage variations on the output node of the current source is minimised. Figure 5.7 shows the schematic diagram of the switch driver. The differential converter includes an inverter in parallel with an always conducting transmission gate. The purpose of the transmission gate is to compensate for the delay created by the inverter. The cross-coupled inverter latches which are widely used in switch gate drivers ensure signal stability by creating a positive feedback loop [84, 79].

The output of the differential converter is a digital signal with a full swing from the power supply rail to the ground level. When the gate signal of the pMOS switch is close to the power rail, the switch is not conducting. This causes a voltage drop on the output of M_{CS} which leads to a current glitch at the output. In order to avoid the voltage drop, a level shifter stage reduces the swing of the differential signals. As a result, instead of a full swing between V_{DD} to ground, the signal has a swing between V_{SB} and ground [80]. The bias voltage V_{SB} which can be controlled as an external bias voltage to prevent the pMOS transistors from being in a non-conducting state.

When the input voltage is digitally high, $M1$ and $M3$ are conducting and $M2$ and $M4$ are not conducting. As a result, the output q is at V_{SB} level and qb is at ground level. Hence, the switch M_{s2} in Figure 5.3 is fully conducting and the majority of the current is flowing through the negative output. When the input is digitally low, $M2$ and $M4$ are conducting and $M1$ and $M3$ are not conducting. As a result, the output

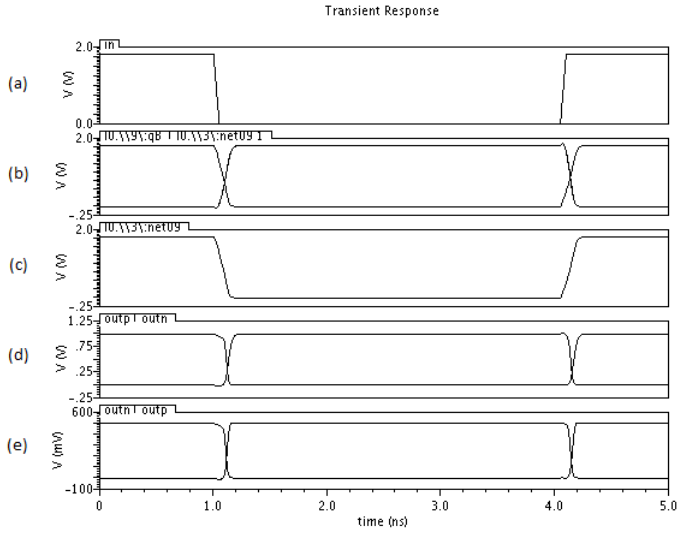


Figure 5.8: The simulation results for the switch driver (a) input (b) output of the differential converter (c) output of the level shifter when $V_{SB} = 1.0V$ (d) when $V_{SB} = 0.5V$

Device	type	diemnsion (W/L) μm
Transmission gate	nMOS	0.24/0.18
	pMOS	0.5/0.18
Inverters	nMOS	0.24/0.18
	pMOS	0.5/0.18
M1-M4	nMOS	0.24/0.18

Table 5.1: The dimensions of the transistors for the switch driver shown in Figure 5.7

qb is at V_{SB} and q is at ground level. Hence, the switch M_{s1} in Figure 5.3 is fully conducting and the majority of the current flows through the positive output.

The reduced voltage swing causes current for each output to swing between a maximum and a minimum current. As V_{SB} decreases, the voltage swing of the switch transistors M_{s1} and M_{s2} decreases, which leads to a higher minimum current . Although the reduced voltage swing improves the dynamic performance, it reduces the current swing of the output.

Figure 5.8 shows the SPICE simulation results for the switch driver. The dimensions of the transistors in the switch driver circuit are shown in Table 5.1. The shape of the input voltage is similar to the output of the automatic diode reset when an avalanche event occurs. The output of the differential converter is a pair of symmetrical complementary signals which has a full voltage swing. The level shifter limits the voltage swing to 1.0V and 0.5V.

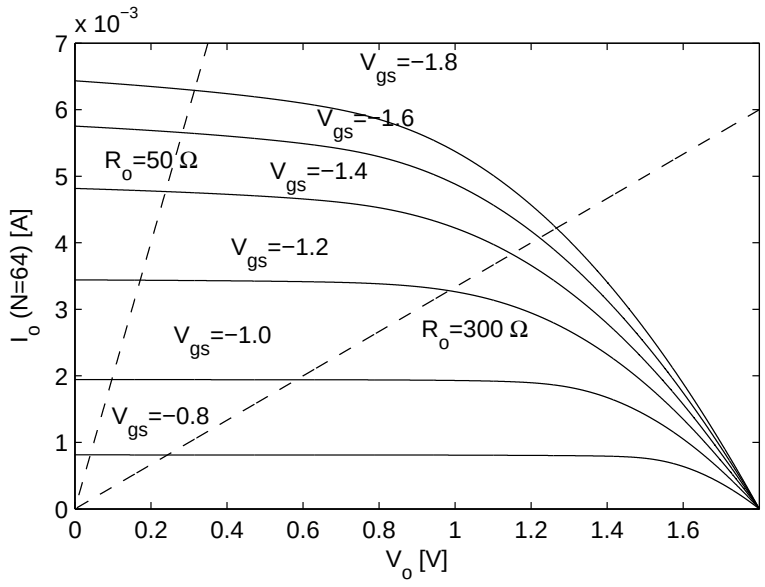


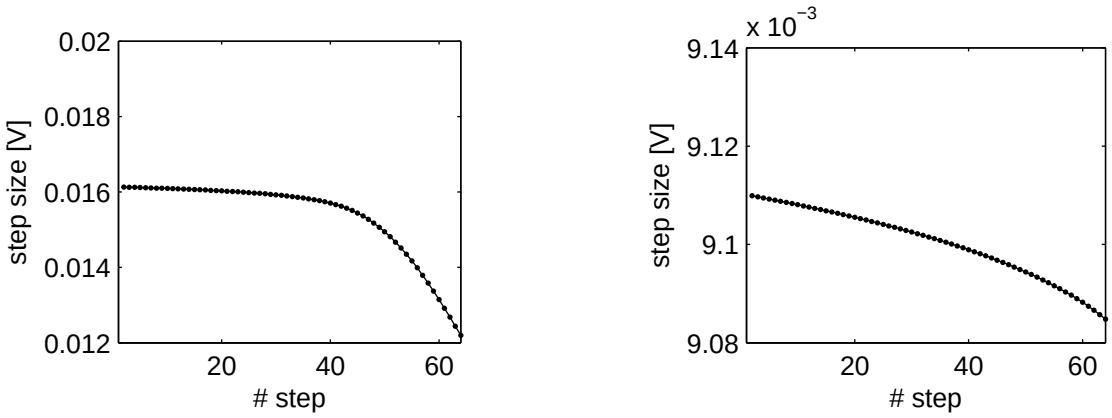
Figure 5.9: The I-V characteristics of the array's output including two load lines 50 ohm and 300 ohm

5.3 The Readout cell

5.3.1 Static performance

The output of the readout circuit can be terminated with an external resistor. The value of this resistor depends on the desired voltage step, speed, and impedance matching. In order to achieve linearity at the output, the value of the resistor should ensure that the current source is always in the saturation. Figure 5.9 shows the load line of a 300 Ω and a 50 Ω resistors alongside the I-V characteristics of the readout circuit which represent all the 64 SPAD cells. The intersection between the load line and the I-V trace of the readout circuit is the maximum voltage and the maximum current, when all current sources are steered into one output of the two differential outputs. With a particular load, as $|V_{gs}|$ increases the maximum output voltage increases, hence the size of each output voltage step increases. With a particular gate voltage, as the load resistance R_o decreases the maximum output voltage decreases, hence the output voltage step decreases. The relationship between the maximum output voltage and the gate voltage of the current source determines that whether M_{CS} is operating in saturation region. Having selected the load, a maximum gate voltage can be selected to keep M_{CS} in saturation. As shown previously in Figure 5.6, when $|V_{gs}| > 1.4$ then M_{CS} is not in saturation.

For example when Ro=300 ohm, and$|V_{gs}| = 1.2$, ($V_g = 0.6$), as shown in to



(a) (b)

Figure 5.10: The voltage step at the output when $R_o = 300\Omega$, and (a) when $|V_{gs}| = 1.2$ (b) when $|V_{gs}| = 1.0$

Figure 5.6, the current of each unit current source is approximately 54 μA, hence each voltage step is 16.2 mV. As shown in Figure 5.9, the maximum voltage when all current sources are steered to one output is $0.98V$, and the maximum current is 3.28 mA.

Even though the saturation condition is satisfied, the channel length modulation and voltage dependence of the current close to the linear region leads to a non-linearity in the output voltage steps. Figure 5.10-a shows the value of each voltage step when $|V_{gs}| = 1.2V$ and $R_o = 300\,\Omega$. The initial voltage step is approximately 16.2 mV. As the number of current sources which are steered to one output increases, the output voltage increases, hence according to I-V characteristic of the current source shown in Figure 5.9, the current of the next current source which is being steered to that output decreases. As the M_{CS} of each current source approaches the linear operation region, the voltage steps decreases sharply. If a greater level of linearity is required, a lower $|V_{gs}|$, or a less resistive load should be chosen. For example, in Figure 5.10-b, the relative difference between the initial voltage step and the final voltage step is 24%. If $|V_{gs}|$ is reduced to $1.0V$, then the relative difference between the initial and the final voltage steps is approximately 0.3%. However, due to the reduced $|V_{gs}|$, the initial voltage step is now reduced to approximately 9.1 mV. As a result, a greater linearity is achievable with the cost of a smaller initial voltage step.

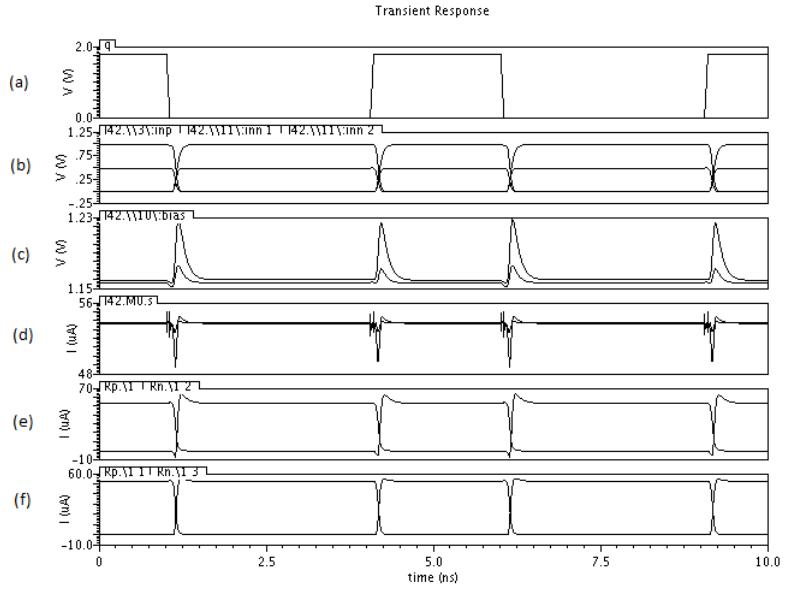


Figure 5.11: The simulation results for the current steering readout circuit (a) the input from automatic diode reset, (b) the gate voltage of current switches when $V_{SB} = 1.0$, and $V_{SB} = 0.5$ (c) voltage variations on the output of M_{CS} (d) current variation of M_{CS} (e) output currents when $V_{SB} = 1.0V$ (f) output currents when $V_{SB} = 0.5V$

5.3.2 Dynamic performance

Figure 11 shows the transient response of the readout circuit to an input pulse shape which is similar to the output of automatic diode reset circuit. The readout circuit in Figure 5.3 is simulated for $R_o = 300\,\Omega$, $|V_{gs}| = 1.2V$ ($Vg = 0.6V$), $V_{SB} = 1.0V$ and $V_{SB} = 0.5V$. As simulation results in Figure 5.11-c show, the current steering operation causes voltage fluctuations on the output node of M_{CS}, which leads to current glitches in Figure 5.11-d. These glitches also appear in the output as shown in Figure 5.11-e. As expected, when V_{SB} is reduced, the amplitude of the glitches reduces which leads to a better dynamic performance of the readout system shown in Figure 5.11-f. However, duration of these glitches is less than $0.3ns$ which is negligible for applications which do not require operating at very high speeds.

5.3.3 Readout array

Figure 5.12 shows simulation results for an example of an array of 4-cell readout circuit, when the $R_o = 300\,\Omega$, $|V_{gs}| = 1.2V$, and $V_{SB} = 0.5V$. The two traces show the positive and negative current outputs. Initially all the 4 current sources are steered into the negative output. An input signal to the switch driver of each cell,

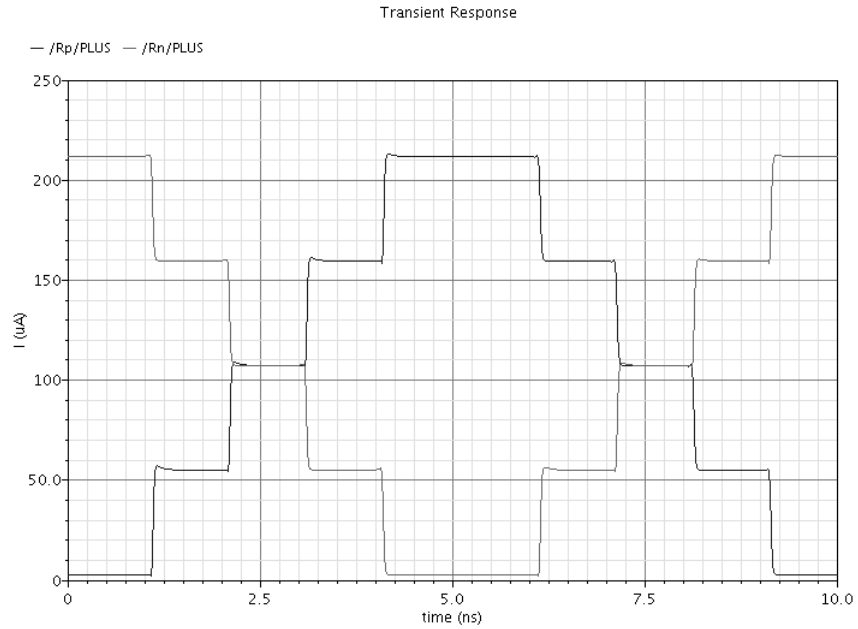


Figure 5.12: An example of a 4-cell operation of current steering readout circuit when $R_o = 300\,\Omega$, $|V_{gs}| = 1.2V$ ($V_g = 0.6V$), and $V_{SB} = 0.5V$

steers the current from the positive output to the negative output. The steering signal is sent to each cell incrementally at every 5 ns, hence the output current increases step by step. After all the currents are steered into the positive output, the signal into switch drivers steers back the current form the positive output to the negative output. Each current step is approximately $52\,\mu A$, the differential signals are equal when 2 cells are steered at the positive output, and the remaining 2 are steered to the negative output. The minimum current in each output is $3.1\mu A$ when $V_{SB} = 0.5V$.

5.3.4 Implemented SPAD cell

An implementation of 64-cell array was fabricated in UMC 0.18 μm commercially available standard CMOS process. The SPAD cell in Figure 5.3 consists of the avalanche photodiode, the automatic diode reset, and the current steering readout circuit.

In order to increase the flexibility of the system, the diode's bias voltages, the deadtime, and gain of the readout system are controlled by external bias voltages. The bias voltage for the avalanche diode has a positive voltage and a negative voltage which biases the either side of the diode. These two voltages determine the overall bias voltage of the avalanche diode operating in SPAD mode. Similar to the chip in Chapter 3, the deadtime can be changed by varying an external bias voltage which changes the timing of the hold monostable. The gain of the readout system is con-

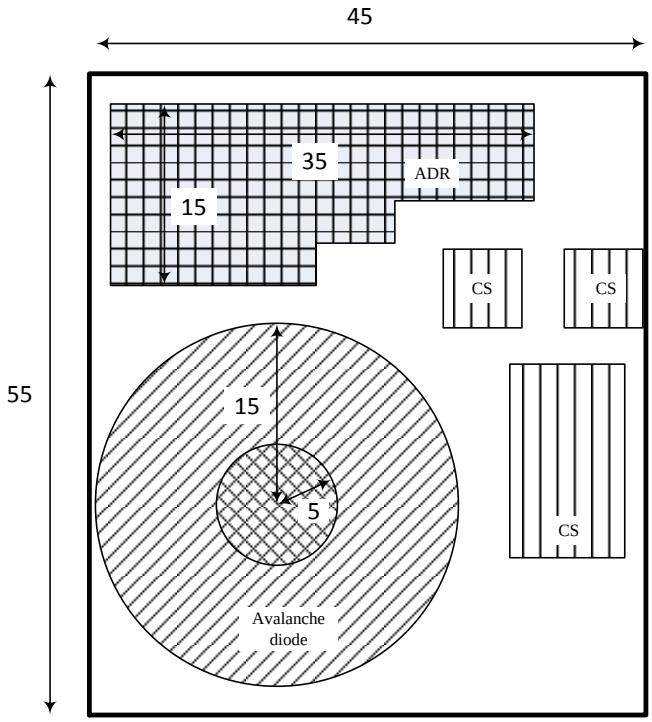


Figure 5.13: The diagram of the physical layout of the implemented SPAD cell showing the relative size of the different parts of the cell. Including the avalanche diode, automatic diode reset (ADR), and current steering readout circuit (CS). Units are in micrometer

trolled by the gate bias voltage of M_{CS} which determines the current of each cell.

Figure 5.13 shows the physical layout of the SPAD cell. This layout is similar to the previous implementation in Chapter 3 , however, the new current steering readout circuit has been added to the layout. The area consuming components in the new readout circuit are the current source and the pair of current switches. The switch driver includes 12 minimum sized transistors which relatively occupies a small area in comparison to the rest of the readout circuit. The current source and the pair of differential switches are surrounded by n+/NWell and p+ guard rings. The inner n+/Nwell ring is connected toV_{DD}, and the outer p+ ring is connected to ground. These guard rings are included in the layout to enhance electro-static discharge (ESD) protection, and provide better substrate noise isolation according to the foundry recommendations. The substrate noise could occur especially when neighbouring circuits are switching at high speed. The parasitic capacitance of source and drain couples the noise into substrate. The guard rings provide a low noise performance of the readout circuit by reducing the noise coupling between devices which are located

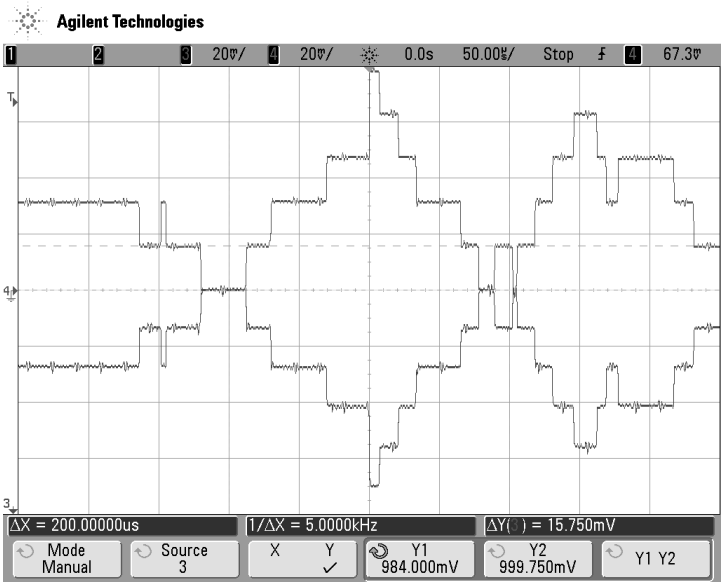


Figure 5.14: The positive and negative voltage output when Ro=300 ohms, overvoltage 1V and deatime is maximum. $VSB = 1.0V$. Voltage step is 16mV. Both signals are complementary

in a close proximity. According to the information provided by the foundry, a device with a p+ guard ring provides approximately -25 dB, and a device with p+, n+/Nwell guard ring provides approximately -40 dB isolation in comparison to a device with no guard ring, at a frequency of 1GHz.

5.4 Experimental Results

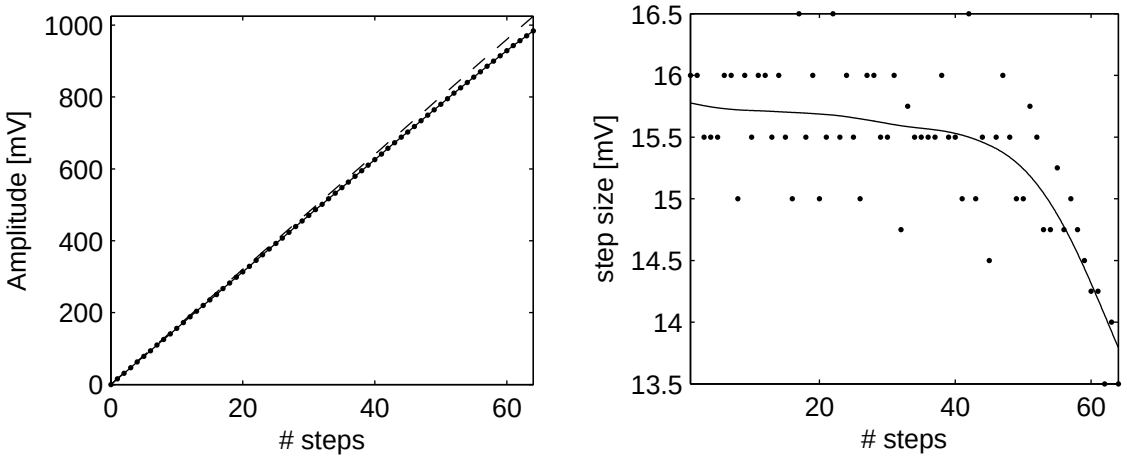
5.4.1 Assessment of the readout system

The implemented current steering readout system has been tested for its functionality. The goal of the experiment is to verify output voltage steps created by the new readout system. The amplitude of the output signal at any time depends on the number of SPAD cells which have detected an avalanche event, and they are within their inactivity period. As the bias voltage of the diode increases, both dark count rate (DCR) and photon detection probability increases. This increase leads to more cells detecting an avalanche event. Similarly, an increase the deadtime leads to an increased number of cells to be in their inactivity period and supplying current to the output. As a result, the average number of SPAD cells which are simultaneously within the inactivity period can be controlled by the bias voltage of the diode and the deadtime. This leads to a control over average amplitude of the output signal.

Figure 5.14 shows a typical differential output signal from the SPAD array when the diode is biased at 1V over its breakdown voltage and the output resistor is 300 Ω. The gain of the readout system is set to ensure that the first voltage step is approximately 16 mV. The deadtime is set to the longest possible value within the circuit limits. As a result of a long deadtime, the rate of step change in the amplitude is in range of tens of micro seconds which leads to clear and sharp output signal due to sufficient measurement bandwidth. An oscilloscope with 4 GSps (sample per second), 1 GHz bandwidth, and a passive 1:1 probes are used to measure the output signal. When the chip is disconnected the measured noise level at the output is 23mV pk-pk and 2mV RMS, hence the system noise level is larger than the intended voltage step. As the transitions do not occur at high frequencies due to long deadtimes, a low pass filter could be used to reduce the amount noise.

When the internal low pass filter of the oscilloscope is used, the noise level reduces to 12mV pk-pk and 1 mV RMS. Furthermore, the oscilloscope applies a special "High Resolution" mode which averages several adjacent samples to achieve a higher resolution plot which contains less amount of noise. The trace in bottom half of the Figure 5.14 shows the positive output signal, and the trace in top half shows the negative output signal. The maximum of the positive signal which indicates the level of the last voltage step is 0.98 V, and it is close to the predicted maximum voltage shown in Figure 5.9 which is 0.98V . The minimum of the negative channel is at the ground level. These voltage levels indicate that most of the time all the 64 cells within the array are sourcing current to the positive output. However, occasionally some of these cells recover from their deadtime, and the number of cells sourcing current from the positive output drops below 64. Hence, the current in the recovered SPAD cells is steered from the positive output to the negative output. This results in a variable but discrete voltage at the both differential outputs. The small glitches seen in Figure 5.14 which appear on both traces are likely to be due to electromagnetic interference (EMI) as they appear on both outputs approximately at the same time and amplitude.

Figure 5.15-a shows the amplitude for all the 64 voltage steps. The initial step is set to be 16mV. The dashed line shows the ideal linear response when all the steps are 16mV. As expected there is a non-linearity especially towards the higher amplitudes at the output. Figure 5.15-b shows the size of each voltage step. Due to limited voltage resolution in the oscilloscope, the measure data is noisy. A smoothed spline curve has been fitted to the experimentally measured data. Similar to Figure 5.10-a, as the output voltage increase the voltage step decreases due to channel length



(a) (b)

Figure 5.15: (a) The amplitude of the all the 64 voltage steps. Initial step is 16 mV. The dashed line is the ideal linear response. the doted line is the experimental data. (b) The value of each voltage step. The dots are the measurement data points. A smoothed spline curve is fitted to the experimental data

modulation, and proximity to the linear operating region.

If a greater linearity is required, the gate-source voltage $|V_{gs}|$ of the current source transistor could be reduced. As shown in Figure 5.10-b, when an initial voltage step of approximately 9mV is used, the relative difference between the first and last step is less than 0.3%. Figure 5.16 shows accumulated oscilloscope traces of the both differential outputs when the initial voltage step is set to be 9mV. A combination of over voltage and deadtime is selected to ensure that the positive output has its minimum amplitude at ground level. Hence, the positive output is fluctuating at lower end of output voltage swing. As a result, the negative output is fluctuating at the higher end of the output voltage swing. Both differential outputs are offset to ensure that they overlap on the oscilloscope screen. The overlapping of both outputs suggests that the voltage steps at the lower and higher end of the voltage swing have negligible difference. As a result, the linearity has significantly improved when the initial step is 9mV in comparison to the situation when the initial step was 16mV.

5.4.2 Bit error ratio tests

Bit error ratio tests (BERTs) were performed on the implemented SPAD array. The goal of these experiments are to prove the photon counting capability of the new array using the conventional readout method for an optical communication receiver shown

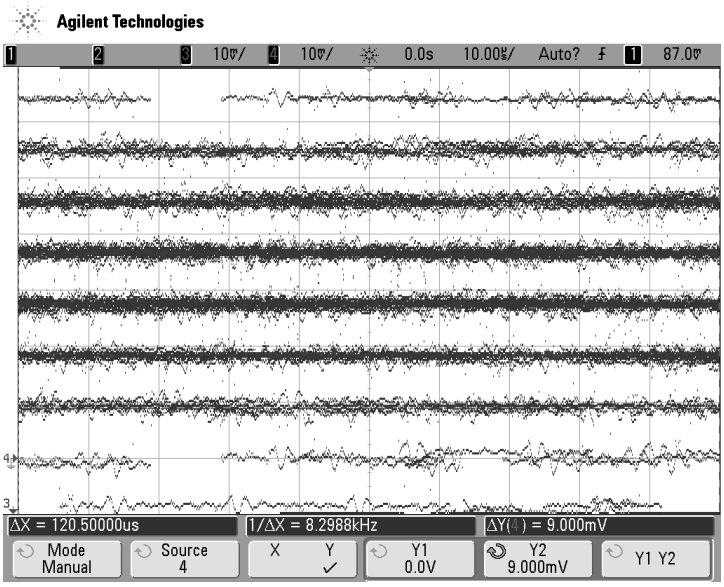


Figure 5.16: The accumulated oscilloscope traces of the both differential outputs when the initial voltage step is set to 9mV

in Figure 1.1. Due to the current output of the new array, the array has replaced the conventional photodiode. Figure 5.17 shows the experimental configuration for the bit error ratio test. This configuration is similar to the setup shown in Figure 4.16. However, due to a higher speed operation, a higher performance transmitter is used [85]. This transmitter is driven by a low voltage differential signalling (LVDS), and has an 860 nm modulated laser diode (LD) with an extinction ratio (ER) of approximately 15. The extinction ratio is the ratio between the transmitter's illumination intensity when a logic 1 is transmitted to the intensity when a logic 0 is transmitted.

Due to a higher clock frequency and fast slew rates the electromagnetic coupling (EMC) is significant between the clock signal and the output of the array. Because the output is an analogue voltage, the noise on this signal should be minimised. In order to decrease the EMC, LVDS signalling [86] is used instead of the standard 3.3V output at the FPGA kit. The LVDS outputs are terminated with a 120Ω resistor, and only the voltage of the positive output of the LVDS signal is connected to the voltage buffers. This results, in an approximately 0.5V voltage swing which significantly reduces the electromagnetic interference on the output of the array.

5.4.2.1 The measurement procedure

For each BER test an eye diagram is formed. Initially the data signal and output of the array is aligned with the clock signal, and their delays are compensated by the

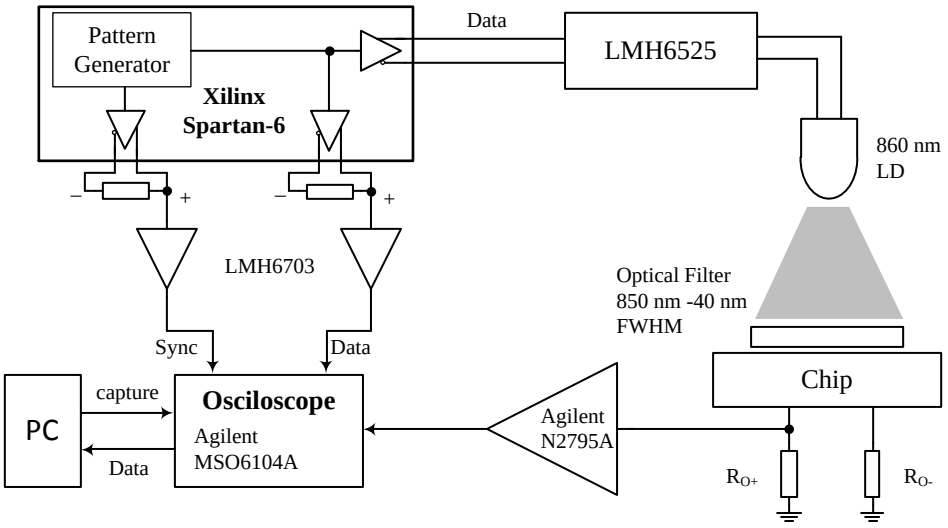
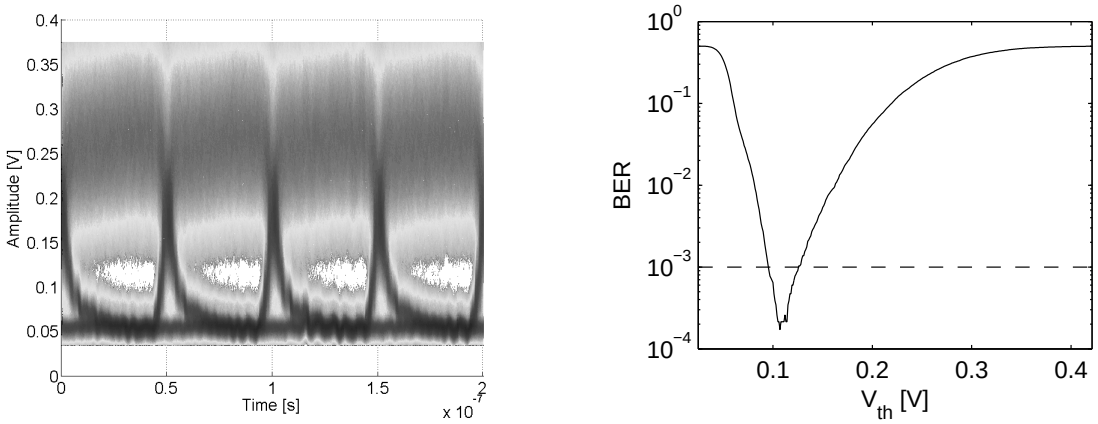


Figure 5.17: The experimental configuration for the bit error ratio test (BERT)

oscilloscope. The eye diagram is formed by superimposing a large number traces into an image memory using a Labview script which capture oscilloscope's data. Once the eye diagram is formed, similar to Figure 1.3.b the sampling time t_s is determined by finding the coordinate within the eye which has the least relative frequency. When the sampling time is determined, another experiment records the output voltage of the array at the sampling. The BER is then calculated by comparing the recorded voltage again a threshold voltage.

5.4.2.2 The optimum deadtime

The amplitude of the output signal at any time is the number inactive SPAD which is the product of the signal intensity and the deadtime. When the deadtime is much smaller than the bit duration, only few SPADs are inactive simultaneously, as the deadtime is elapses shortly after an even is detected. However, when deadtime is comparable to the bit duration, each SPAD within the array is inactive for more most of the bit period. As more events are detected in the array, the number of inactive SPADs increases, as a result, the output amplitude which is an accumulation of current sources of each cell increases. The maximum amplitude is when all current sources are steered into the positive output. If the deadtime is too long, then few SPAD are inactive after the bit duration has finished. This results in an inter symbol interference. Therefore it is expected that the maximum amplitude and minimum inter symbol interference is achieved when the deadtime is equal to the bit period.



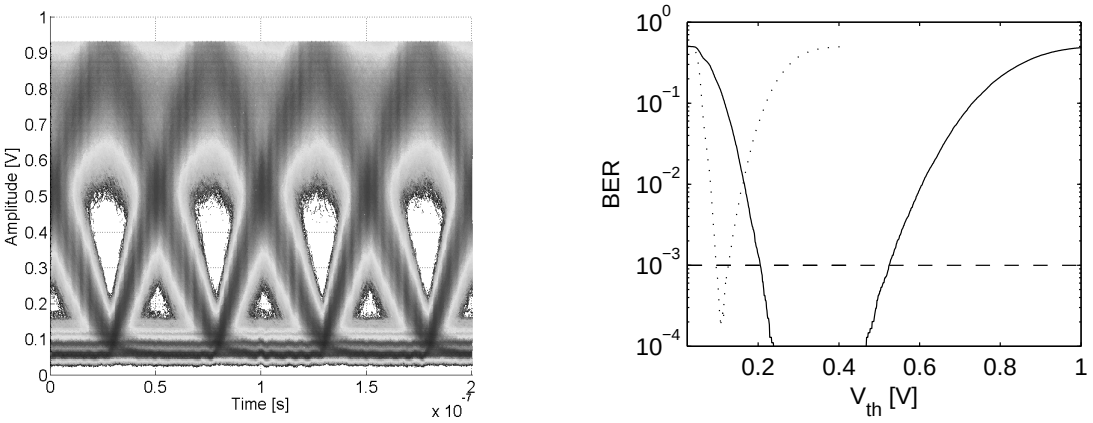
(a) (b)

Figure 5.18: The BERT when $T_b = 50\,ns$, $\lambda_{AS} = 42.8$, and $\tau_d = 0.2$ (a) the eye diagram (b) the BER versus decision threshold voltage

In order to investigate the effect of deadtime on the output amplitude, initially a deadtime of 10 ns is selected. The bit duration is $T_b = 50\,ns$, hence the relative deadtime is $\tau_D = 0.2$. To achieve a maximum gain, the current of current sources M_{CS} is set to the maximum current. Although the linearity of the output is affected by setting the maximum current, it is not important in this experiment to have a linear output as long as the decision threshold is within the linear region of the output. The bias voltage is 11.9V and $V_{SB} = 1.0V$.

The intensity of signal and background noise are measured by a calibrated power meter. The photon detection probability (PDP) at the transmitter's wavelength, which is 860 nm, is approximately 0.015 based on Figure 2.12. Hence, the average photon count per entire array during the bit duration is $\lambda_{AS} = 42.8$, and the average count when the transmitter is transmitting logic 0 is 1.8 due to the effect of finite extinction ratio in the transmitter. Based on Figure 3.11, the average dark count rate (DCR) for a SPAD cell is approximately 90 kcps (count per seconds). As there are 64 cells within the array, the total DCR is expected to be equal to $64 \times 9 \times 10^4 = 5.7 \times 10^6$. Hence the counts related to the DCR is $5.7 \times 10^6 \times 50 \times 10^{-9} = 0.28$, and the total background noise count is background noise is $\lambda_{AB} = 1.8 + 0.28 = 2.08$. In this case the background noise is dominated by the illumination of the transmitter when it is transmitting logic 0. The power penalty for deadtime when $M = 64$ and $\tau_d \leq 1$ is negligible based on Figure 4.15. As a result, the theoretical BER based on Equation 4.5 is $BER = 3.4 \times 10^{-8}$.

Figure 5.18.a shows the experimental eye diagram. The white area in the middle



(a) (b)

Figure 5.19: (a)The BERT when $T_b = 50\,ns$, $\lambda_{AS} = 42.8$ and $\tau_d = 1.0$ (a) the eye diagram (b) the BER versus decision threshold voltage. The dotted curve is the BER from Figure 5.18

of the eye shows the coordinates with the least relative frequency, hence the minimum BER is achieved in this area. The BER experiment is performed when the sampling $t_s = 37.5\,ns$ after the start of bit duration. Figure 5.18.b shows the BER versus decision voltage threshold. The minimum BER is 1.7×10^{-4} which occurs when $V_{th} = 0.1\,V$. As a result, the experimental BER is significantly higher than the theoretical BER.

However, a better BER is expected when the deadtime is equal the bit duration. Figure 5.19.a shows the experimental eye diagram when deadtime is approximately 50ns. The conditions of the experiment except the deadtime, which is $\tau_d = 1$, is the same as the previous experiment shown in Figure 5.18. It can be observed that the eye diagram is now symmetrical, and the rise/fall time has increased due to a longer deadtime. The BER experiment is performed when the sampling time $t_s = 25\,ns$, which is centre of the eye. Figure 5.19.b shows the experimental BER versus decision threshold voltage. As a results, the BER has significantly improved, however due to the limited length of the received data, the minimum BER cannot be calculated. The improvement in the BER, which is due to increases amplitude, is due to an increases in the deadtime.

5.4.2.3 The photon counting capability

In order to verify the photon counting capability of the implemented SPAD array, an achievable BER of 10^{-3} is targeted. The deadtime and biasing conditions are the

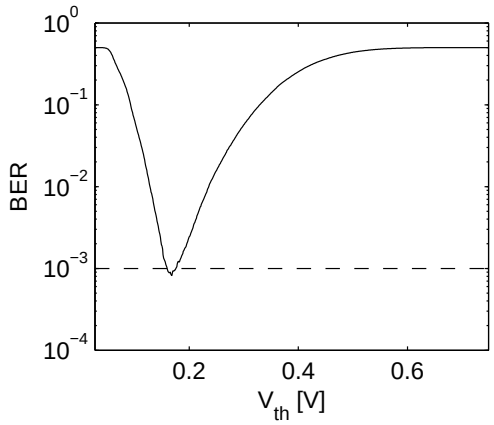
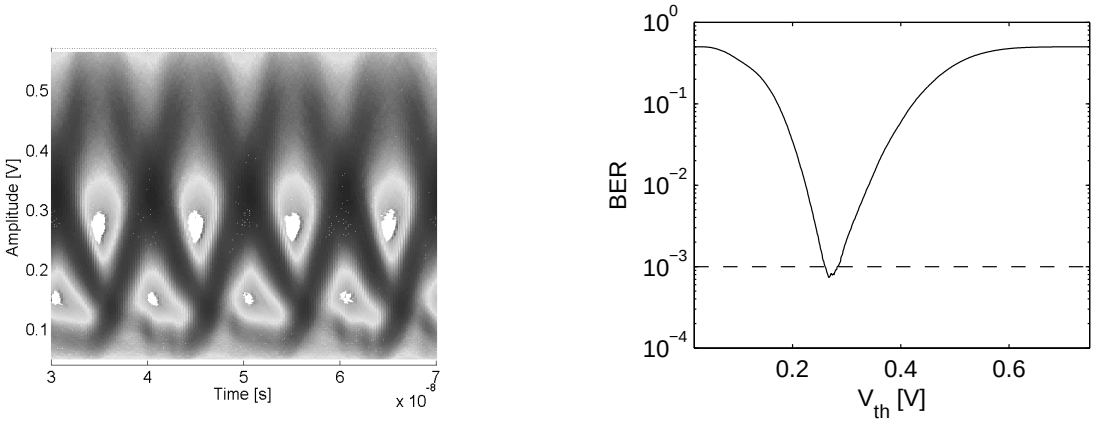


Figure 5.20: The BER versus decision threshold voltage when $T_b = 50\,ns$, $\tau_d = 1.0$, and $\lambda_{AS} = 13.6$

same as the previous experiment shown in Figure 5.19. The transmitter intensity is changed, and the BER experiment is performed repeatedly until a minimum target BER is obtained. Figure 5.20 shows the experimental BER versus threshold voltage when the average signal count per entire array is $\lambda_{AS} = 13.6$, and the background noise including the DCR is $\lambda_{AB} = 0.6$. The minimum BER of 0.8×10^{-3} is achieved when $V_{th} = 0.17V$. The theoretical BER based on Equation 4.5 is 0.9×10^{-3}. The difference between the theoretical and the experimental is likely to be due to inaccuracy in measuring the transmitter's intensity and estimation of the PDP. Since the BER is sensitive is to the intensity due to an exponential relationship, small inaccuracy could lead to a noticeable difference in the BER. Therefore, results from this experiment suggest that by using the array in a conventional optical receiver, the photon counting ability is achieved when the deadtime is equal to the bit period.

Figure 5.21-a shows the eye diagram when the bit duration is $T_b = 10\,ns$, and $\tau_d = 1.0$. The average signal count per entire array is $\lambda_{AS} = 19.8$, and the average background noise including the DCR is $\lambda_{AB} = 0.55$. It can be observed form the eye diagram that the fall time has extends for two bits because of a shorter bit duration in comparison to the eye diagram shown in Figure 5.19, hence it is expected that the BER results are affected by the limited electrical bandwidth. Figure 5.21 shows the BER versus decision threshold voltage. The minimum BER of 0.7×10^{-3} is achieved at $V_{th} = 0.27V$, however the average photon count required to achieve this BER based on Equation 4.5 is $\lambda_{AS} = 14$. The 41% relative difference between the theoretical and experimental average signal count is potentially due to the effect of limited electrical bandwidth within the measurement system. Future improvement are required in the experimental configuration in order to the increases the overall



(a) (b)

Figure 5.21: The BERT when $T_b = 10\,ns$ and $\tau_d = 1.0$ (a) the eye diagram (b) the BER versus decision threshold voltage

bandwidth and improve the signal integrity of the system.

5.5 Summary

In this chapter, the proposed readout circuit for a SPAD cell is implement based on current steering digital-to-analogue converters. This enables implementation of large arrays operating at high speeds. The current steering readout circuit includes a current sources which sources a constant current to the output through a differential pair of switches. Based on the input of the readout out circuit, these switches steer the current to either of the differential current output. At the design stage, the dimensions of the current source is selected based on the acceptable variation among the current sources in each cell within the array, and the available physical silicon area within the cell due to the inverse relationship between variability and the dimensions of the transistor. The width of the current switches are selected large enough to ensure that have the least resistance, however, a longer width leads to a larger parasitic capacitance which causes capacitive coupling between digital signal at the gate of the current switch and the analogue current output. The output capacitance of each cell is the drain's capacitance of the current switch.

A dedicated circuit is required to drive the gates of the current switches. This circuit referred to as switch driver, provides complementary switching signals to the gate of the switches, and reduces the voltage swing on the gate of the current switches in order to minimise the glitches which occur during the switching operation. The

output of the array is the shared output of each SPAD cell. Similar to a conventional photodiode, this output current is then converted to a voltage by an external resistor or a trans-impedance amplifier (TIA). The value of the external resistor (or the feedback resistor when a TIA is used) is selected based on the maximum amplitude and the amount of linearity required at the output. The amplitude of the output signal at any time is proportional to the number of inactive SPAD cells due to detection of an avalanche event.

An array of 64 SPAD cells was implemented in UMC 0.18 μm standard CMOS process. A SPAD cell includes the avalanche diode, automatic diode reset (ADR), and the current steering readout circuit. The bias voltage of the avalanche diode, the deadtime of the ADR circuit, and the gain of the readout circuit are set by external bias voltages. The current output of the array is terminated with an off-chip resistor. The functionality of the proposed readout circuit is verified by measuring voltage steps at the output the array when different gain is set for the readout circuit.

Appropriate modifications have been made to the experimental configuration in order to minimise the electromagnetic interference (EMI) on the analogue output of the array due short rise/fall times of the digital signals. The experimental results prove that with a constant signal intensity, the lowest bit error rate (BER) is achieved when deadtime is equal to the bit duration. As a result the photon counting capability of the implemented array is achieved by using the conventional setup for an optical receiver. A bit error rate of 10^{-3} at a bit rates of 100 Mbps has been achieved with the implemented array.

Chapter 6

Conclusion and Future Work

6.1 Summary

The front end of an optical receiver includes a photo-detector, which converts the illumination into a photo-current, and a current-to-voltage converter which converts the photo-current into a voltage. The most popular photo-detector is a photodiode (PD) which generates a photo-current proportional to the incident light. A simple method for converting the generated photo-current into a voltage is to use a resistor. The capacitance of the photodiode and the resistor form an RC circuit, which limits the electrical bandwidth of the receiver. A more common approach for converting the photo-current into voltage is a trans-impedance amplifier (TIA), which uses a voltage amplifier with a shunt-shunt feedback resistor. The TIA has the benefit of a higher electrical bandwidth in comparison to the resistor method due to a lower input resistance of the TIA.

The sensitivity of the detector is limited by the collection area of the PD and the noise of the front end of the optical receiver. This noise has an inverse relationship with the feedback resistor and a direct relationship with the capacitance of the photodiode. Increasing the resistance, reduces the noise, however, it reduces the RC bandwidth. The collection area of the PD is proportional to its capacitance, hence, a larger PD increases the noise and leads to a lower RC bandwidth. The noise and the slow rise/fall times due to the limited RC bandwidth leads to bit errors in the receiver. The bit error ratio (BER), which is a major performance parameter for a communication link, is simply defined as the ratio between the bit errors to the total transmitted bits.

Avalanche photo diodes (APD) enhances the sensitivity of the photo-detector by an internal gain without increasing the capacitance. The gain of APDs, which is due

to avalanche multiplication, increases with their bias voltage, and to achieve higher gains they should be biased in a region where the gain is very sensitive to the bias voltage. This make the APD sensitive to voltage and temperature variation. In addition, the randomness in the avalanche multiplication process leads to statistical fluctuations in the gain. These fluctuations lead to a higher noise, hence, there is a trade-off between the gain and the noise of an APD. The avalanche photodiodes are mainly targeted at high-end optical communications due to their custom fabrication process which leads to a discrete device with a relatively high cost.

The randomness of the gain in APDs can be avoided when the avalanche diode is biased over its breakdown voltage so that the avalanche current is regarded as a binary current. In this case, the detector is sensitive to a single photon, hence, it is referred to as a single photon avalanche diode (SPAD). The output of the SPAD is a digital-like pulse indicating detection of an avalanche event, and the rate of the digital output pulses is proportional to the intensity of the incident light. Unlike avalanche photodiodes, SPADs are easily integrated with other peripheral circuits in the optical receiver in a standard CMOS process. However, similar to a PD and an APD, the collection area of a SPAD is limited to its junction capacitance. In addition in SPADs, each device can only detect one photon per deadtime.

In order to use the avalanche diode in the SPAD mode, a circuit is required to quench the avalanche current and recover the SPAD after an avalanche event has occurred. An integrated quenching and recovery circuit is implemented which uses a transistor to quench and recover the SPAD.

An asynchronous array of SPADs, which are connected in parallel and operate independently, increases the overall collection area of the detector and avoids the problems of using a single large device. The output of the array is an aggregated output of each SPAD. However, reading out large arrays is challenging. One approach is to connect all the diodes in parallel so that the current generated during the avalanche event in each SPAD passes through a common node, which is the output of the array. Similar to a conventional photo-detector the current is then converted into a voltage. The major problem with this approach is that although a higher sensitivity can be achieved, connecting the diodes in parallel creates a larger capacitance which limits the bandwidth of the detector. In addition, the output current is dependent on the characteristics of the avalanche diode, hence, reducing the flexibility of photo-detector. Another approach, that has been proposed recently, is to connect the digital output of the SPAD to a digital counter which is then read-out by a global clocked system. Although all the readout mechanism is performed in the digital domain,

the size of the array is limited by the frequency of the readout clock. Hence, implementing large arrays is practically challenging due to high speed clock generation and distribution.

The proposed readout mechanism in this thesis is inspired by the architecture of high speed digital to analogue converters (DAC). This readout mechanism is based on steering the current of a constant current source within the SPAD cell to either of differential outputs. Implementation of large arrays is possible as the output capacitance of each SPAD cell within the array is equal to the output capacitance of the switches which steer the current. This capacitance is significantly smaller than the capacitance of the avalanche diode. In addition, the gain of the array which is the current of the current source can be controlled by an external bias voltage. This can be used to ensure that the output of the array is beyond the noise margin of the subsequent current-to-voltage converter. Moreover, the avalanche diode and the peripheral circuits which handle the diode are independent of the proposed readout mechanism. Hence, this readout mechanism is able to integrate with future development in SPADs based on CMOS technology.

The aim of this study was to design, characterise, and test a prototype array of SPADs based on the proposed readout concept. Initially the avalanche diode was implemented and characterised. The diode was fabricated in a commercially available 0.18 μm standard CMOS process. The avalanche diode has a more complex layout than a conventional photodiode. A circular guard ring surrounds the high electrical field region of the avalanche diode in order to provide a uniform electrical field, and to prevent premature breakdown. The breakdown voltage of the diode was measured at 10.4V.

The avalanche diode is characterised for performance parameters such as dark count rate (DCR) and photon detection probability (PDP). The dark count is the generation of spontaneous avalanche events, which also occur in the dark and are not related to incident photons. The DCR increases with the bias voltage, and has an exponential relation with the bias voltage in most of the voltage range. The photon detection probability, is the ratio of the avalanche events to the incident photons, and depends on the bias voltage and the wavelength of the incident photons. The PDP increases linearly with the bias voltage in most of the bias voltage range, and is higher in shorter wavelengths of the light spectrum due to the shallow pn junction in the avalanche diode.

The SPAD circuit, which was used for characterisation of the fabricated avalanche diode, is not suitable for a high-speed performance. The transistor, which quenches

the avalanche current, also provides the means of the recovery process. A faster recovery time means that the quenching transistor should have a lower resistance. However, the low resistance of the quenching transistor may not quench the avalanche current leading to a sustained avalanche current and malfunction in the SPAD operation.

An automatic diode reset (ADR) circuit has been designed to provides the means of quenching and recovering the avalanche diode independently. The ADR circuit includes transistor switches to handle the bias voltage of the avalanche diode, and a timing circuitry, which controls the timing of these switches. When an avalanche event occurs the avalanche current is quenched, then the bias voltage of the diode is held below its breakdown voltage for a specified delay which is referred to as deadtime. Once the deadtime is elapsed, the diode is quickly recovered to its initial bias voltage. The deadtime is controlled by an external bias voltage. A minimum deadtime of 4 ns was achieved with the implemented ADR circuit which was integrated with the avalanche diode. The DCR and PDP measurement confirmed that the avalanche diode is operating in the SPAD mode as expected from the results previously obtained by characterising the avalanche diode.

Once the SPAD circuit was tested, a numerical model was developed in order to assess the performance of the SPAD for optical communications. A SPAD is a photon counting detector with a photon count which has Poisson statistics, hence, the numerical model is developed based on a photon counting detector with deadtime. The simulated communication link is based on on-off keying (OOK) method which means that the transmitter is ON when sending a logic one, and OFF when sending a logic zero. The simulation results show that the BER improves when more signal photons are counted as the probability of missing a photon decreases. The background noise includes the dark count of the detector and background light which is not related to the transmitter. The photon counts related to the background noise when the transmitter is sending a logic zero, degrades the performance the detector.

The modelling results show that the BER of a photon counting detector depends on the average signal photon count, the average background noise, and the deadtime. The BER increases with the background noise. The increase in the BER can be avoided by increasing the number of signal photons, an increase which is referred to as the power penalty for the background noise. If the background noise is below a certain limit, then the power penalty is negligible. In practice, the DCR of the fabricated SPADs are generally smaller than this limit. Similarly, the deadtime decreases the performance of detector by causing inter-symbol-interference (ISI). As a result, the BER generally increases with deadtime, except when the deadtime is close to the bit

duration. In order to compensate for the increase in the BER, and extra signal photon count is required which is the power penalty for the deadtime. When deadtime is smaller than a fraction of the bit duration, the power penalty for the deadtime is negligible.

An array of photon counting detector is modelled by aggregating the output of each photon-counting detector. The simulation results confirmed that although the background noise increases with the size of the array, the power penalty for the background noise does not increase at the same rate as the background noise. Hence, there is an advantage in increasing the size of the array even in the presence of the background noise. Moreover, the power penalty for the deadtime decreases with the size of the array, as more detectors are available within the array to detect an avalanche event.

Once the benefits of the array of photon counting detectors were established, the proposed readout mechanism, which is based on high-speed current steering digital to analogue converters, was implemented and integrated alongside the avalanche diode and the automatic diode reset circuit in order to form a SPAD cell. The new readout circuit contains an equally weighted constant current source with a controllable gate voltage. The current of this current source is steered to one of the differential outputs depending on the status of the SPAD. The output of the automatic diode reset is connected to a switch driver circuit which provides a complementary switching signal for the gate of the transistor switches. When the SPAD is inactive due to detection of an avalanche event, the current is steered by a pair of transistor switches from the negative output to the positive output.

A prototype 64-cell array of SPAD was implemented and tested. Each SPAD cell included the avalanche diode, the automatic diode reset circuit, and the current steering readout circuit. Initial experiments verified the functionality of the proposed readout circuit by comparing the measured output voltage steps to the simulation results of the circuit. A bit error ratio testing platform was configured based on a conventional optical receiver. A series of tests proved that although the array was connected as a conventional photo-detector, it was able to maintain its photon counting characteristic when the deadtime was equal to the bit duration. The final test achieved a reasonable bit error rate of 10^{-3} at a bit rate of 100 Mbps.

6.2 Future work

6.2.1 The testing platform

Immediate enhancements can be made on the testing platform in order to utilise the full potential of the fabricated chip. In the test platform shown in Figure 5.17 which was used for bit error ratio tests (BERT), the signal integrity can be improved by using a dedicated printed circuit board (PCB). It is important to use shortest track lengths with appropriate impedance matching to achieve a better signal integrity, which prevents slow rise/fall times, and ringing effects in the signal.

A trans-impedance amplifier (TIA) can be used instead of the resistor to improve the electrical bandwidth of the system. A discrete operational amplifier with a high signal slew rate could be used as the voltage amplifier in the TIA. Special considerations are required in the PCB layout design for the TIA when operating at a high speed. A lack of ground plane and impedance mismatching could lead to instability at the output of the TIA.

In the configuration which was developed for BERT, although the implemented array has a differential output, only a single-ended output was used. Differential signalling is advantageous over single-ended signalling due to the common mode rejection in the differential signalling. The digital equipment especially the field programmable gate array (FPGA) kit and digital interfaces generate electromagnetic interference (EMI) which appears as noise on the analogue output of the chip. If the differential output of the chip is used, the noise related to EMI is minimised as approximately an equal noise signal appears on both of the differential output lines, hence, the common mode is subtracted from the original signal. The differential signalising requires two channels of the oscilloscope, as each channel is single-ended.

These hardware improvements have the potential to improve the signal integrity and the electrical bandwidth of the measurements system which leads to shorter fall/rise times, hence, improving the BER at high bit rates. Moreover, improvements are also possible by software techniques. Forward error correction (FEC) methods have the potential to improve the bit error ratio. FEC is a coding method which corrects for bit errors by adding an overhead to the original data without a need for communicating back to the transmitter [10]. Standard coding schemes such Reed-Solomon which are recommended in the ITU-T G.709 standard for long-haul fibre optics transmission can improve the bit error ratio of 10^{-3} to 10^{-6} with 7% data overhead. More advanced FEC codes, which are implemented in hardware, are able to improve the BER of 10^{-3} to 10^{-12} with 7% data overhead [87]. However, FEC codes,

which are developed and tested for conventional optical receivers, require further investigation to be used with photon counting receivers.

Another major improvement to the BER test is possible by accelerating the speed of experiment. The BER tests performed for this thesis used an offline processing mechanism which only random segments of the generated bit stream was captured by the oscilloscope. In offline processing, the data was then transferred to a personal computer for further processing. As a result, the speed of experiment was limited to the delay caused by the capture and process mechanism regardless of the data rate. On the other hand, in real-time processing, the data is captured and compared with the originally transmitted data using a hardware which shares the same clock signal as the bit stream generator. Hence every bit is captured and compared, and the speed of the experiment is limited by the data rate. The higher speeds in the test enables measuring lower bit error ratios, and specifically improve the resolution of results which require parametric sweeps.

Traditionally, real-time BERT hardware is limited in availability and expensive to obtain. Recently, field programmable gate arrays (FPGA) have been used to create BERT platforms due to their flexibility in digital design and wide availability [88]. In the BERT experiments which were performed for this thesis, only the bit stream generator was implemented in the FPGA as the data transmitter. However, the data capture and the comparison with the original data stream could be implemented within the FPGA as the data receiver. A global on-board clock is used for both the data transmitter and the data receiver. The FPGA has only digital input/output interfaces, hence, the analogue output of the TIA should be connected to the comparator before connecting to the FPGA. The comparator stage shown in Figure 1.1, applies the decision threshold voltage, and can be implemented with a high bandwidth operation amplifier.

A more realistic approach to the optical transceiver design is to separate the receiver's and transmitter's clock, hence, both units operate independently. As a result, a clock and data recovery (CDR) system is required as shown in Figure 1.1. The CDR can be implemented with the integrated phase locked loop (PLL) system which is integrated within the FPGA. The PLL extracts the clock form the input data signal. The extracted clock is then used to synchronise the clock of the receiver with the transmitter. Additionally, peripheral logic including FEC codes [89], experiment automation, and data interfaces can be implemented into a single FPGA alongside the BERT logic.

6.2.2 The SPAD array

Improvements are possible on the SPAD array to increase its sensitivity. In order to improve the sensitivity of the fabricated array, the fill factor of each cell should be increased. The fill factor is the ratio between the active area of the avalanche diode to the entire area of the SPAD cell. A major challenge is to increase the fill factor of the avalanche diode itself. The guard rings surrounding the avalanche diode can be reduced and made narrower. In addition, the size of the active area can be increased while the thickness of the guard rings remains constant, hence, increasing the fill factor. However the size of the active area is limited by the capacitance of the avalanche diode.

Recently, near infra-red avalanche diodes have been developed in CMOS process [90]. These diodes use deeper junctions with lower levels of doping leading to a reduction in the capacitance of the diode and generally a lower dark count rate when operated in SPAD mode. The deeper pn junction intrinsically enhances the near infra-red response of the SPAD especially in the popular wavelength for infra-red wireless communication which is 850 nm [2].

Moreover, some of the layers within the avalanche diode which share the same terminal could be shared to increase the fill factor of the avalanche diode. In the fabricated avalanche diode shown in Figure 2.2 and 3.2, the voltage connected to the NEG terminal which is the anode of the diode is shared among the diodes within the SPAD array. However, this layer cannot be physically shared in the silicon substrate as the anode is surrounded by an outer N-well ring which is the cathode of the diode. In this case the node which the bias voltage of the diode is observed is the outer N-well ring. An alternative bias method uses the inner p+ (anode) as the node which the bias voltage is observed [91], the outer N-well can be shared among the SPAD within the array. The sharing of N-wells leads to an avalanche diode with a smaller size and higher fill factor [92]. However, sharing the N-wells is likely to increases the cross talk and coupling between neighbouring avalanche diodes as the P-sub/N-well junction which was an isolating layer between the diodes has been eliminated.

In order to further improve the overall performance of the detector, the TIA and the comparator can be integrated with the SPAD array. The dominant component in limiting the electrical bandwidth of the detector is the parasitic capacitance of the chip package. Integrating the analogue components of the optical receiver with the SPAD array eliminates the effect of the package capacitance. As a result, a high speed serial interface can be used to connect the digital output of the comparator to off-chip components [93].This is possible by using a low voltage differential signal

(LVDS) interface which provides high speed digital connections [86].

Once these changes have been made a demonstrator is required. This demonstrator enables an implementation of a fully functional communication link, and will be thoroughly tested as a consumer-end product. During these tests the shortcomings and challenges of the system will be identified, and further solutions to these challenges will be proposed. One of these challenges is to select the transmission wavelength which depends on several factors including eye safety, efficiency and cost of the optical system in that wavelength.

Bibliography

[1] Y. Dong and K. Martin, "Gigabit communications over plastic optical fiber," *Solid-State Circuits Magazine, IEEE*, vol. 3, no. 1, pp. 60–69, 2011.

[2] J. Kahn and J. Barry, "Wireless infrared communications," *Proceedings of the IEEE*, vol. 85, no. 2, pp. 265–298, 1997.

[3] K. Hamamatsu Photonics, E. Committee *et al.*, "Photomultiplier tubes-basics and applications," *Edition 3a*, 2006.

[4] R. Haitz, "Mechanisms contributing to the noise pulse rate of avalanche diodes," *Journal of Applied Physics*, vol. 36, no. 10, pp. 3123–3131, 1965.

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

[6] K. Schrodinger, J. Stimma, and M. Mauthe, "A fully integrated CMOS receiver front-end for optic Gigabit Ethernet," *Solid-State Circuits, IEEE Journal of*, vol. 37, no. 7, pp. 874–880, 2002.

[7] W. Chen, S. Huang, G. Wu, C. Liu, Y. Huang, C. Chin, W. Chang, and Y. Juang, "A 3.125 Gbps CMOS fully integrated optical receiver with adaptive analog equalizer," in *Solid-State Circuits Conference, 2007. ASSCC'07. IEEE Asian.* IEEE, 2007, pp. 396–399.

[8] S. Alexander, *Optical Communication Receiver Design: Stephen B. Alexander*, ser. IEE Telecommunications series. Society of Photo Optical, 1997.

[9] R. Gagliardi and S. Karp, *Optical communications*, ser. Wiley series in telecommunications and signal processing. Wiley, 1995.

[10] A. Tychopoulos, O. Koufopaulou, and I. Tomkos, "FEC in optical communications-A tutorial overview on the evolution of architectures and the future prospects of outband and inband FEC for optical communications," *Circuits and Devices Magazine, IEEE*, vol. 22, no. 6, pp. 79–86, 2006.

[11] P. Casper and W. Ashley, "Temperature compensated avalanche photodiode optical receiver circuit," 1984, US Patent 4,438,348.

[12] J. Richard, P. Gilliland, E. Chamness, E. Anguelov *et al.*, "Single chip ASIC and compact packaging solution for an avalanche photodiode (APD) and bias circuit," 2005, US Patent 6,894,266.

[13] J. Youn, H. Kang, M. Lee, K. Park, and W. Choi, "High-speed CMOS integrated optical receiver with an avalanche photodetector," *Photonics Technology Letters, IEEE*, vol. 21, no. 20, pp. 1553–1555, 2009.

[14] S. Verghese, J. Donnelly, E. Duerr, K. McIntosh, D. Chapman, C. Vineis, G. Smith, J. Funk, K. Jensen, P. Hopman *et al.*, "Arrays of InP-based avalanche photodiodes for photon counting," *Selected Topics in Quantum Electronics, IEEE Journal of*, vol. 13, no. 4, pp. 870–886, 2007.

[15] N. Faramarzpour, M. Deen, S. Shirani, and Q. Fang, "Fully Integrated Single Photon Avalanche Diode Detector in Standard CMOS 0.18 μm," *Electron Devices, IEEE Transactions on*, vol. 55, no. 3, pp. 760–767, 2008.

[16] M. Marwick and A. Andreou, "Fabrication and Testing of Single Photon Avalanche Detectors in the TSMC 0.18 μm CMOS Technology," in *Information Sciences and Systems, 2007. CISS'07. 41st Annual Conference on.* IEEE, 2007, pp. 741–744.

[17] D. Chitnis and S. Collins, "A flexible compact readout circuit for SPAD arrays," in *SPIE NanoScience+ Engineering.* International Society for Optics and Photonics, 2010, pp. 77 801E–77 801E.

[18] S. Cova, M. Ghioni, A. Lacaita, C. Samori, and F. Zappa, "Avalanche photodiodes and quenching circuits for single-photon detection," *Applied optics*, vol. 35, no. 12, pp. 1956–1976, 1996.

[19] S. Cova, M. Ghioni, A. Lotito, I. Rech, and F. Zappa, "Evolution and prospects for single-photon avalanche diodes and quenching circuits," *Journal of Modern Optics*, vol. 51, no. 9-10, pp. 1267–1288, 2004.

[20] M. Ghioni, S. Cova, F. Zappa, and C. Samori, "Compact active quenching circuit for fast photon counting with avalanche photodiodes," *Review of scientific instruments*, vol. 67, no. 10, pp. 3440–3448, 1996.

[21] F. Zappa, M. Ghioni, S. Cova, C. Samori, and A. Giudice, "An integrated active-quenching circuit for single-photon avalanche diodes," *Instrumentation and Measurement, IEEE Transactions on*, vol. 49, no. 6, pp. 1167–1175, 2000.

[22] F. Zappa, A. Lotito, A. Giudice, S. Cova, and M. Ghioni, "Monolithic active-quenching and active-reset circuit for single-photon avalanche detectors," *Solid-State Circuits, IEEE Journal of*, vol. 38, no. 7, pp. 1298–1301, 2003.

[23] S. Tisa, F. Guerrieri, and F. Zappa, "Variable-Load Quenching Circuit for single-photon avalanche diodes," *Optics express*, vol. 16, no. 3, pp. 2232–2244, 2008.

[24] J. Massa, G. Buller, A. Walker, S. Cova, M. Umasuthan, and A. Wallace, "Time-of-flight optical ranging system based on time-correlated single-photon counting," *Applied optics*, vol. 37, no. 31, pp. 7298–7304, 1998.

[25] G. Buller, R. Harkins, A. McCarthy, P. Hiskett, G. MacKinnon, G. Smith, R. Sung, A. Wallace, R. Lamb, K. Ridley *et al.*, "Multiple wavelength time-of-flight sensor based on time-correlated single-photon counting," *Review of scientific instruments*, vol. 76, no. 8, pp. 083 112–083 112, 2005.

[26] C. Niclass, A. Rochas, P. Besse, and E. Charbon, "Toward a 3-D camera based on single photon avalanche diodes," *Selected Topics in Quantum Electronics, IEEE Journal of*, vol. 10, no. 4, pp. 796–802, 2004.

[27] C. Niclass and E. Charbon, "A single photon detector array with 64× 64 resolution and millimetric depth accuracy for 3D imaging," in *Solid-State Circuits Conference, 2005. Digest of Technical Papers. ISSCC. 2005 IEEE International*. IEEE, 2005, pp. 364–604.

[28] C. Niclass, A. Rochas, P. Besse, and E. Charbon, "Design and characterization of a CMOS 3-D image sensor based on single photon avalanche diodes," *Solid-State Circuits, IEEE Journal of*, vol. 40, no. 9, pp. 1847–1854, 2005.

[29] G. Buller and A. Wallace, "Ranging and three-dimensional imaging using time-correlated single-photon counting and point-by-point acquisition," *Selected Topics in Quantum Electronics, IEEE Journal of*, vol. 13, no. 4, pp. 1006–1015, 2007.

[30] D. Stoppa, L. Pancheri, M. Scandiuzzo, L. Gonzo, G. Dalla Betta, and A. Simoni, "A CMOS 3-D imager based on single photon avalanche diode," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 54, no. 1, pp. 4–12, 2007.

[31] D. Schwartz, E. Charbon, and K. Shepard, "A single-photon avalanche diode array for fluorescence lifetime imaging microscopy," *Solid-State Circuits, IEEE Journal of*, vol. 43, no. 11, pp. 2546–2557, 2008.

[32] L. Pancheri and D. Stoppa, "A SPAD-based pixel linear array for high-speed time-gated fluorescence lifetime imaging," in *ESSCIRC, 2009. ESSCIRC'09. Proceedings of*. IEEE, 2009, pp. 428–431.

[33] D. Tyndall, B. Rae, D. Li, J. Richardson, J. Arlt, and R. Henderson, "A 100Mphoton/s time-resolved mini-silicon photomultiplier with on-chip fluorescence lifetime estimation in 0.13 μm CMOS imaging technology," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2012 IEEE International*. IEEE, 2012, pp. 122–124.

[34] Y. Maruyama and E. Charbon, "A Time-Gated 128x128 CMOS SPAD Array for on-Chip Fluorescence Detection," in *Proc. Int. Image Sens. Workshop*, 2011, pp. 270–273.

[35] P. Finocchiaro, A. Campisi, D. Corso, L. Cosentino, G. Fallica, S. Lombardo, M. Mazzillo, F. Musumeci, A. Piazza, G. Privitera *et al.*, "Test of scintillator readout with single photon avalanche photodiodes," *Nuclear Science, IEEE Transactions on*, vol. 52, no. 6, pp. 3040–3046, 2005.

[36] M. Fishburn and E. Charbon, "System tradeoffs in gamma-ray detection utilizing SPAD arrays and scintillators," *Nuclear Science, IEEE Transactions on*, vol. 57, no. 5, pp. 2549–2557, 2010.

[37] A. Therrien, B. Berube, C. Thibaudeau, S. Charlebois, R. Lecomte, R. Fontaine, and J. Pratte, "Modeling of single photon avalanche diode array detectors for PET applications," in *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2011 IEEE*. IEEE, 2011, pp. 48–53.

[38] E. Wilman, S. Gardiner, A. Nomerotski, R. Turchetta, M. Brouard, and C. Vallance, "A new detector for mass spectrometry: Direct detection of low energy ions using a multi-pixel photon counter," *Review of Scientific Instruments*, vol. 83, no. 1, pp. 013 304–013 304, 2012.

[39] C. Niclass, M. Gersbach, R. Henderson, L. Grant, and E. Charbon, "A single photon avalanche diode implemented in 130nm CMOS technology," *Selected Topics in Quantum Electronics, IEEE Journal of*, vol. 13, no. 4, pp. 863–869, 2007.

[40] M. Gersbach, J. Richardson, E. Mazaleyrat, S. Hardillier, C. Niclass, R. Henderson, L. Grant, and E. Charbon, "A low-noise single-photon detector implemented in a 130nm CMOS imaging process," *Solid-State Electronics*, vol. 53, no. 7, pp. 803–808, 2009.

[41] J. Richardson, L. Grant, and R. Henderson, "Low dark count single-photon avalanche diode structure compatible with standard nanometer scale CMOS technology," *Photonics Technology Letters, IEEE*, vol. 21, no. 14, pp. 1020–1022, 2009.

[42] C. Niclass and M. Soga, "A miniature actively recharged single-photon detector free of afterpulsing effects with 6ns dead time in a 0.18 μm CMOS technology," in *Electron Devices Meeting (IEDM), 2010 IEEE International*. IEEE, 2010, pp. 14–3.

[43] D. Bronzi, F. Villa, S. Bellisai, B. Markovic, S. Tisa, A. Tosi, F. Zappa, S. Weyers, D. Durini, W. Brockherde *et al.*, "Low-noise and large-area CMOS SPADs with Timing Response free from Slow Tails," in *Solid-State Device Research Conference (ESSDERC), 2012 Proceedings of the European*. IEEE, 2012, pp. 230–233.

[44] I. Alsolami, D. Chitnis, D. C. O'Brien, and S. Collins, "Broadcasting over Photon-Counting Channels via Multiresolution PPM: Implementation and Experimental Results," *Communications Letters, IEEE*, vol. 16, no. 12, pp. 2072 –2074, december 2012.

[45] P. Hopman, P. Boettcher, L. Candell, J. Glettler, R. Shoup, and G. Zogbi, "An end-to-end demonstration of a receiver array based free-space photon counting communications link," in *Optics & Photonics*. International Society for Optics and Photonics, 2006, pp. 63 040H–63 040H.

[46] J. Mendenhall, L. Candell, P. Hopman, G. Zogbi, D. Boroson, D. Caplan, C. Digenis, D. Hearn, and R. Shoup, "Design of an optical photon counting array receiver system for deep-space communications," *Proceedings of the IEEE*, vol. 95, no. 10, pp. 2059–2069, 2007.

[47] J. Glettler, P. Hopman, S. Verghese, J. Aversa, L. Candell, J. Donnelly, E. Duerr, J. Frechette, J. Funk, Z. Liau *et al.*, "InP-based single-photon detector arrays with asynchronous readout integrated circuits," *Optical Engineering*, vol. 47, no. 10, pp. 100 502–100 502, 2008.

[48] B. Aull, M. Renzi, R. Reich, and D. Schuette, "CMOS readout architecture and method for photon-counting arrays," 2010, US Patent App. 12/730,048.

[49] J. Frechette, P. Grossmann, D. Busacker, G. Jordy, E. Duerr, K. McIntosh, D. Oakley, R. Bailey, A. Ruff, M. Brattain *et al.*, "Readout circuitry for continuous high-rate photon detection with arrays of InP Geiger-mode avalanche photodiodes," in *Proceedings of SPIE*, vol. 8375, 2012.

[50] E. Fisher, I. Underwood, and R. Henderson, "A reconfigurable 14-bit 60GPhoton/s Single-Photon receiver for visible light communications," in *ESSCIRC (ESSCIRC), 2012 Proceedings of the*. IEEE, 2012, pp. 85–88.

[51] K. Yamamoto, K. Yamamura, K. Sato, T. Ota, H. Suzuki, and S. Ohsuka, "Development of multi-pixel photon counter (MPPC)," in *Nuclear Science Symposium Conference Record, 2006. IEEE*, vol. 2. IEEE, 2006, pp. 1094–1097.

[52] S. Gomi, H. Hano, T. Iijima, S. Itoh, K. Kawagoe, S. Kim, T. Kubota, T. Maeda, T. Matsumura, Y. Mazuka *et al.*, "Development and study of the multi pixel photon counter," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 581, no. 1, pp. 427–432, 2007.

[53] M. Szawlowski, D. Meier, G. Maehlum, D. Wagenaar, and B. Patt, "Spectroscopy and timing with Multi-Pixel Photon Counters (MPPC) and LYSO scintillators," in *Nuclear Science Symposium Conference Record, 2007. NSS'07. IEEE*, vol. 6. IEEE, 2007, pp. 4591–4596.

[54] A. Nassalski, M. Moszynski, A. Syntfeld-Kazuch, T. Szczesniak, L. Swiderski, D. Wolski, T. Batsch, and J. Baszak, "Multi pixel photon counters (MPPC) as an alternative to APD in PET applications," *Nuclear Science, IEEE Transactions on*, vol. 57, no. 3, pp. 1008–1014, 2010.

[55] P. Eraerds, M. Legré, A. Rochas, H. Zbinden, and N. Gisin, "SiPM for fast photon-counting and multiphoton detection," *Optics Express*, vol. 15, no. 22, pp. 14 539–14 549, 2007.

[56] N. Dinu, R. Battiston, M. Boscardin, G. Collazuol, F. Corsi, G. Dalla Betta, A. Del Guerra, G. Llosa, M. Ionica, G. Levi *et al.*, "Development of the first prototypes of Silicon PhotoMultiplier (SiPM) at ITC-irst," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 572, no. 1, pp. 422–426, 2007.

[57] N. Dinu, Z. Amara, C. Bazin, V. Chaumat, C. Cheikali, G. Guilhem, V. Puill, C. Sylvia, and J. Vagnucci, "Electro-optical characterization of SiPM: A comparative study," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 610, no. 1, pp. 423–426, 2009.

[58] L. Braga, L. Pancheri, L. Gasparini, M. Perenzoni, R. Walker, R. Henderson, and D. Stoppa, "A CMOS mini-SiPM detector with in-pixel data compression for PET applications," in *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2011 IEEE*. IEEE, 2011, pp. 548–552.

[59] A. Van Den Bosch, M. Borremans, M. Steyaert, and W. Sansen, "A 12b 500 MSample/s current-steering CMOS D/A converter," in *Solid-State Circuits Conference, 2001. Digest of Technical Papers. ISSCC. 2001 IEEE International*. IEEE, 2001, pp. 366–367.

[60] K. O'Sullivan, C. Gorman, M. Hennessy, and V. Callaghan, "A 12-bit 320-MSample/s current-steering CMOS D/A converter in 0.44 mm^2 ," *Solid-State Circuits, IEEE Journal of*, vol. 39, no. 7, pp. 1064–1072, 2004.

[61] Q. Huang, P. Francese, C. Martelli, and J. Nielsen, "A 200MS/s 14b 97mW DAC in 0.18μm CMOS," in *Solid-State Circuits Conference, 2004. Digest of Technical Papers. ISSCC. 2004 IEEE International*. IEEE, 2004, pp. 364–532.

[62] K. Doris, J. Briaire, D. Leenaerts, M. Vertreg, and A. Van Roermund, "A 12b 500MS/s DAC with 70dB SFDR up to 120MHz in 0.18 μm CMOS," in *Solid-State Circuits Conference, 2005. Digest of Technical Papers. ISSCC. 2005 IEEE International*. IEEE, 2005, pp. 116–588.

[63] D. Mercer, "Low-power approaches to high-speed current-steering digital-to-analog converters in 0.18-μm CMOS," *Solid-State Circuits, IEEE Journal of*, vol. 42, no. 8, pp. 1688–1698, 2007.

[64] S. Sarkar and S. Banerjee, "An 8-bit 1.8V 500 MSPS CMOS Segmented Current Steering DAC," in *VLSI, 2009. ISVLSI'09. IEEE Computer Society Annual Symposium on*. IEEE, 2009, pp. 268–273.

[65] Z. Cui, H. Piao, and N. Kim, "A 10-bit current-steering DAC in 0.35 μm CMOS process," *Trans. Electr. Electron. Mater.*, vol. 10, pp. 44–8, 2009.

[66] K. Doris, A. van Roermund, and D. Leenaerts, *Wide-Bandwidth High Dynamic Range D/A Converters*, ser. The Springer International Series in Engineering and Computer Science. Springer, 2006.

[67] R. McIntyre, "A new look at impact ionization-Part I: A theory of gain, noise, breakdown probability, and frequency response," *Electron Devices, IEEE Transactions on*, vol. 46, no. 8, pp. 1623–1631, 1999.

[68] S. Sze, *Physics of Semiconductor Devices*. Wiley, New York, 1969.

[69] S. Ghandhi, *Semiconductor power devices: physics of operation and fabrication technology*. Wiley New York, 1977.

[70] M. Garza, N. Eib, and K. Chao, "Optical proximity correction method and apparatus," 1998, US Patent 5,723,233.

[71] H. Finkelstein, M. Hsu, and S. Esener, "STI-bounded single-photon avalanche diode in a deep-submicrometer CMOS technology," *IEEE Electron Device Letters*, vol. 27, no. 11, pp. 887–889, 2006.

[72] Q. Chen, D. Chitnis, K. Walls, T. Drysdale, S. Collins, and D. Cumming, "CMOS photodetectors integrated with plasmonic color filters," *Photonics Technology Letters, IEEE*, vol. 24, no. 3, pp. 197–199, 2012.

[73] A. Giudice, M. Ghioni, S. Cova, and F. Zappa, "A process and deep level evaluation tool: afterpulsing in avalanche junctions," in *European Solid-State Device Research, 2003. ESSDERC'03. 33rd Conference on*. IEEE, 2003, pp. 347–350.

[74] C. Niclass, M. Sergio, and E. Charbon, "A single photon avalanche diode array fabricated in 0.35μm CMOS and based on an event-driven readout for TCSPC experiments," in *Proc. SPIE*, vol. 6372, 2006.

[75] M. Schroeder, "How to generate thermal photons on the computer," in *On Klauder's Path: A Field Trip*, 1994, pp. 197–202.

[76] C. Chen, "Effect of detector dead time on the performance of optical direct-detection communication links," in *The Telecommunications and Data Acquisition Report*, vol. 1, 1988, pp. 146–154.

[77] M. Larsen and A. Kostinski, "Simple deadtime corrections for discrete time series of non-poisson data," *Measurement Science and Technology*, vol. 20, no. 9, p. 095101, 2009.

[78] P. Alfke, "Efficient Shift Registers, LFSR Counters, and Long Pseudo-Random Sequence Generators," 1996, XAPP052.

[79] M. Usama and T. Kwasniewski, "Design and comparison of CMOS current mode logic latches," in *Circuits and Systems, 2004. ISCAS'04. Proceedings of the 2004 International Symposium on*, vol. 4. IEEE, 2004, pp. IV–353.

[80] D. Mercer, "Digital to Analog Converter Design," 2008.

[81] P. Aliparast, H. Bahar, Z. Koozehkanani, J. Sobhi, and G. Karimian, "A 12-Bit 1-Gsample/s Nyquist Current-Steering DAC in 0.35 μm CMOS for Wireless Transmitter," *Circuits and Systems*, vol. 2, no. 2, pp. 74–84, 2011.

[82] J. Bastos, M. Steyaert, and W. Sansen, "A high yield 12-bit 250-MS/s CMOS D/A converter," in *Custom Integrated Circuits Conference, 1996., Proceedings of the IEEE 1996*. IEEE, 1996, pp. 431–434.

[83] M. Pelgrom, A. Duinmaijer, and A. Welbers, "Matching properties of MOS transistors," *Solid-State Circuits, IEEE Journal of*, vol. 24, no. 5, pp. 1433–1439, 1989.

[84] D. Mercer and L. Singer, "12-b 125 MSPS CMOS D/A designed for spectral performance," in *Proceedings of the 1996 international symposium on Low power electronics and design*. IEEE Press, 1996, pp. 243–246.

[85] D. O'Brien, R. Turnbull, H. Le Minh, G. Faulkner, M. Wolf, L. Grobe, J. Li, O. Bouchet, and E. Gueutier, "A 280Mbit/s infrared optical wireless communications system," in *SPIE Optical Engineering+ Applications*. International Society for Optics and Photonics, 2011.

[86] A. Boni, A. Pierazzi, and D. Vecchi, "LVDS I/O interface for Gb/s-per-pin operation in 0.35 μm CMOS," *Solid-State Circuits, IEEE Journal of*, vol. 36, no. 4, pp. 706–711, 2001.

[87] K. Seki, K. Mikami, A. Katayama, S. Suzuki, N. Shinohara, and M. Nakabayashi, "Single-chip FEC codec using a concatenated BCH code for 10 Gb/s long-haul optical transmission systems," in *Custom Integrated Circuits Conference, 2003. Proceedings of the IEEE 2003*. IEEE, 2003, pp. 279–282.

[88] A. Xiang, T. Cao, D. Gong, S. Hou, C. Liu, T. Liu, D. Su, P. Teng, and J. Ye, "High-speed serial optical link test bench using FPGA with embedded transceivers," in *Proceeding of the topical workshop on electronics in particle physics (TWEPP), Paris, France*, 2009.

[89] M. Francis, "Forward Error Correction on ITU-G. 709 Networks using Reed-Solomon Solutions," *Xilinx Application Note XAPP952*, 2007.

[90] E. Webster, J. Richardson, L. Grant, D. Renshaw, and R. Henderson, "A single-photon avalanche diode in 90-nm CMOS imaging technology with 44% photon detection efficiency at 690 nm," *Electron Device Letters, IEEE*, vol. 33, no. 5, pp. 694–696, 2012.

[91] M. A. Al-Rawhani, D. Chitnis, J. Beeley, S. Collins, and D. R. S. Cumming, "Design and Implementation of a Wireless Capsule Suitable for Autofluorescence Intensity Detection in Biological Tissues," *Biomedical Engineering, IEEE Transactions on*, vol. 60, no. 1, pp. 55 –62, jan. 2013.

[92] R. Henderson, E. Webster, R. Walker, J. Richardson, and L. Grant, "A 3× 3, 5μm pitch, 3-transistor single photon avalanche diode array with integrated 11V bias generation in 90nm CMOS technology," in *Electron Devices Meeting (IEDM), 2010 IEEE International*. IEEE, 2010, pp. 14–2.

[93] A. Athavale and C. Christensen, "High-speed serial I/O made simple," *Xilinx Inc*, vol. 4, 2005.