ABSTRACT

Title of dissertation: INTEGRATED SINGLE-PHOTON SENSING AND PROCESSING PLATFORM IN STANDARD CMOS

Babak Nouri, Doctor of Philosophy, 2013

Dissertation Advisor: Dr. Pamela Abshire
Electrical and Computer Engineering Department

Practical implementation of large SPAD-based sensor arrays in the standard CMOS process has been fraught with challenges due to the many performance trade-offs existing at both the device and the system level [1]. At the device level the performance challenge stems from the suboptimal optical characteristics associated with the standard CMOS fabrication process. The challenge at the system level is the development of monolithic readout architecture capable of supporting the large volume of dynamic traffic, associated with multiple single-photon pixels, without limiting the dynamic range and throughput of the sensor.

Due to trade-offs in both functionality and performance, no general solution currently exists for an integrated single-photon sensor in standard CMOS single photon sensing and multi-photon resolution. The research described herein is directed towards the development of a versatile high performance integrated SPAD sensor in the standard CMOS process.

Towards this purpose a SPAD device with elongated junction geometry and a perimeter field gate that features a large detection area and a highly reduced dark noise has been presented and characterized. Additionally, a novel front-end system for optimizing the dynamic range and after-pulsing noise of the pixel has been developed. The pixel is also equipped with an output interface with an adjustable pulse width response. In order to further enhance the effective dynamic range of the pixel a theoretical model for accurate dead time related loss compensation has been developed and verified.

This thesis also introduces a new paradigm for electrical generation and encoding of the SPAD array response that supports fully digital operation at the pixel level while enabling dynamic discrete time amplitude encoding of the array response. Thus offering a first ever system solution to simultaneously exploit both the dynamic nature and the digital profile of the SPAD response. The array interface, comprising of multiple digital inputs capacitively coupled onto a shared quasi-floating sense node, in conjunction with the integrated digital decoding and readout electronics represents the first ever solid state single-photon sensor capable of both photon counting and photon number resolution. The viability of the readout architecture is demonstrated through simulations and preliminary proof of concept measurements.

.

INTEGRATED SINGLE-PHOTON
SENSING AND PROCESSING PLATFORM IN STANDARD CMOS

By

Babak Nouri

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
Of the requirements for the degree of
Doctor of Philosophy
2013

Advisory Committee:
Professor Pamela Abshire, Chair/Advisor
Professor Neil Goldsman
Professor Timothy Horiuchi
Professor Marty Peckerar
Professor Isabel Lloyd

© Copyright by
Babak Nouri
2013

Table of Contents

List of Figures

Chapter 1

Background & Overview

1.1 Introduction

A photon is the ultimate limit of sensitivity in optical sensing. The rise of high performance single photon detectors has led to the emergence of many new scientific fields, while enabling the continued progress of many other existing ones [1-3]. The broad spectrum of applications for this technology extends from high energy particle detection at the Large Hadron Collider to DNA sequencing and gene-associated protein detection for the human genome project.

Photons are very useful as probes for medical/biomedical imaging applications. They can travel safely through tissues, interact with them and carry environmental information from the sample back to the outside world for subsequent processing and interpretation [2]. Since the amount of light emitted from micro-scale biological structures is extremely small the optical properties of biological sample can only be exploited with optical detectors of sufficient sensitivity [3]. The set of techniques that exploit the photon-based biological interactions in order to extract specific information about the biological matter is collectively defined as Bio-photonics [4-8]. Bio-photonics constitutes a major area of advanced optical imaging. Previously the only means of observing cellular behavior was to view a frozen stained and sliced sample of the organism under a microscope. This provided no information about the active internal process of the cells and the dynamics of its interactions with the cells surrounding it [5]. Bio-photonics allows dynamic observation of biological events at the micro-scale level thus enabling real time study of biological functions at the cellular or sub-cellular levels. The granularity of such real-time observation extends to the level of a single molecule within an active cell. Ultra-sensitive optical methods are also widely utilized to identify and quantify DNA, mRNA or protein content within biological matter [6, 7].

In the clinical field, single photon sensitive imaging has enabled the study of the mechanisms involved in human health and disease at the cellular and molecular level. Optical properties of tissue such as fluorescence reveal valuable information regarding its pathology and composition [7]. The fluorescence property of biological matter is also modulated by disease-related abnormalities in the organic structure of tissues. Discerning the resulting shift in the optical properties for micro-scale biological structures would require contrast resolution on order of single photon. This marks the maximum resolution limit for optical intensity measurements.

Another major area of impact is the Imaging of extremely fast transient events such as spike-based neural signals characterized by microsecond-long action potentials travelling through neural pathways at speeds of up to 112 mph [8]. Effective imaging of such high speed phenomena requires optical detectors with a very high frame rate capability or equivalently very short integration window (small exposure time) single-photon sensitivity yields the shortest possible integration window of one photon.

Essentially the challenge for the detection system utilized in biomedical imaging application is the speed and sensitivity requirement for detection of ultra-low intensity bio-reporter signal (optical) emitted as result of bio-photonic processes [8].

LiDAR (Light Detection and Ranging) is a technology that utilizes light to probe the atmosphere in order to retrieve information about atmospheric properties such as temperature, humidity, pollution, weather pattern and the trace gases present. The detection of changes in properties of the emitted light pulse after it has interacted with the atmosphere allows for the measurement and mapping of atmospheric parameters [9]. LiDAR can also be used as Laser range finders. It can determine the distance to a remote object by detecting the reflected pulse and measuring the round trip flight time. Consideration of Cost, portability and eye safety necessitates the use of low power source for laser generation which proportionally lowers the intensity of the reflected signal. Due to the limitation of the detector active area only a small fraction of this reflected signal is captured by the detector. The low back-scattering coefficient of atmosphere also attenuates the reflected optical signal. As the result of these restrictions, LiDAR applications are critically dependent upon the ability for efficient ultrasensitive photon detection [9].

1.2 Assessment of Existing Single Photon Detection Technologies

The underlying process in signal detection is the electro-transduction of the incident energy signal into a representative electrical signal at the output. The function of a photo-detector is to convert the optical energy of light into an analog electrical signal whose magnitude is proportional to the incident intensity. The challenge presented involves fast and reliable detection of ultra-faint signals. In conventional optical detectors the free carriers generated due to the absorption of incident photons are passively integrated until the resulting electrical signal exceeds the baseline noise of the external amplification and read out electronics [10]. This method is inherently ill suited to meet the stringent sensitivity and speed requirement for detection of Ultra-weak optical signals that occur on a time scale from hundreds of nanosecond to milliseconds. The detection of such signals is contingent upon the presence of an intrinsic internal amplification mechanism that is as noiseless as possible [11-13].

The gold standard in single photon detection, for the past several decades has been the PhotoMultiplier Tube (PMT). The internal amplification process in PMT occurs on time scale of tens of nanosecond. Within that time period a single photo-generated carrier pair is converted into a pulse of current which can be directly read out. However, due to their internal mechanism for signal transduction and amplification, PMTs have high power consumption (large operating voltage) and are operationally sensitive to the presence of external magnetic field. Additionally, due to their bulky packaging, they offer poor detection granularity, and are mainly used as stand-alone point detectors. Although these limitations have been partially over come through introduction of multi-anode PMT and Micro Channel Tube, these devices still fall short of meeting the growing technological demand for greater system integration and miniaturization [14]. Many applications on the cutting edge of technology such as Lab on Chip (LoC) and Micro Total Analysis System (μTAS) must deliver ultra-sensitive and high speed integrated functionality across an implementation platform capable of very high levels of miniaturization, reliability and low power operation [14, 15]. Semiconductor technology can offer a robust implementation platform for such high performance single photon detection, provided that the speed and sensitivity performance of PMT can be realized in silicon. A basic illustration of PMT appears in Fig. 1.1a.

A modified version of the standard semiconductor CCD called Electron Multiplying CCD is capable of achieving single photon sensitivity by incorporating an amplification stage directly into the readout architecture thus eliminating the external noisy amplification step. A notable advantage over PMT is that EM-CCD devices can be manufactured into high density detector arrays capable of parallel detection. However, due to the intrinsic gradual amplification process through several stages of high voltage multiplication registers, they suffer from long response times and cannot match the detection speed of PMT device. Fig. 1.1b illustrates the basic readout architectural enhancement required to upgrade a CCD detector to an EM-CCD.

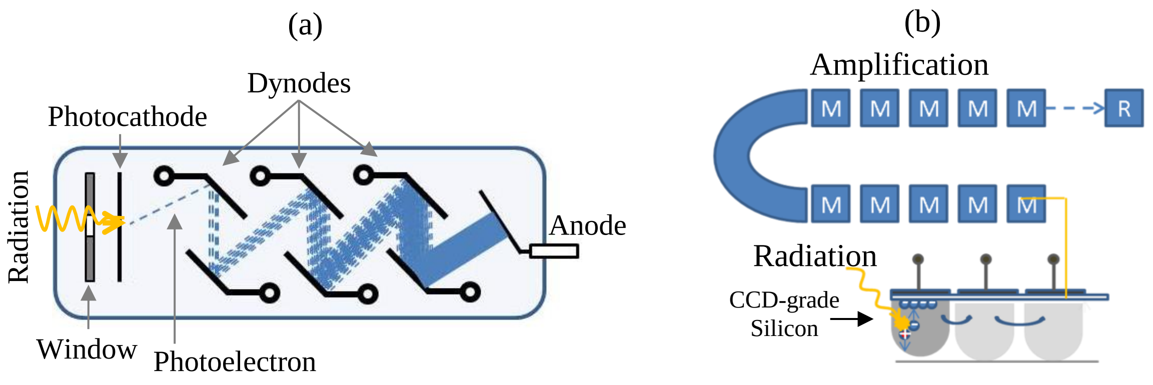


Figure 1.1 (a) Structure of a PMT, the resistive divider circuit distributes the optimum voltage to each node to induce the charge acceleration required for impact ionization. (b) EMCCD structure consisting of the CCD part and electron multiplying readout component.

Another solid state alternative for single photon detection is Single Photon Avalanche Diode (SPAD) in custom CMOS process. SPAD is essentially a p-n junction reverse-biased at above breakdown level with the maximum electric field uniformly distributed across the depletion region of the planar p-n junction, which constitutes the device multiplication region [15]. Photo-generated carriers within this region are accelerated by the high electric field and undergo a rapid multiplication process through repeated high speed collisions with the lattice atoms, resulting in successive impact ionization. The result is an avalanche current generated on picosecond time scale [16]. The charge multiplication process is driven by the high lateral electric field and occurs in a single turn. Due to the large number of electrons in the high filed region during the avalanche build up, there will be multiple impact-ionization events occurring in parallel. According to the law of large numbers, at any given time, there will be a large number of carriers working to reduce the variance of the gain within the high field region. If one electron fails to fulfill its average number of ionizations, another is likely to exceed it. Consequently the amplification noise in SPAD is negligible [17]. The rapid internal amplification by means of a run-a-way avalanche process completely eliminates many of the traditional sources of noise such as (read noise, amplification noise) typical in analog optical detectors. The main source of noise in SPADs is the Dark Count Rate (DCR), which is the rate of dark (photon unrelated) avalanche events. These non-photon related pulses can be due to thermal generation of carriers, band to band tunneling, field assisted tunneling, after pulsing or crosstalk [18]. Basic depiction of SPAD operation appears in Fig. 1.2.

Custom SPADs are capable of matching the single photon detection speed and sensitivity of PMT, however they requires front-end electronic support to enable their rapid reset and recovery for continuous operation. Since electronics components cannot be easily integrated in the custom CMOS process, hybrid systems have been proposed whereby the SPAD detectors on custom substrate are externally coupled to an electrically active standard CMOS substrate containing the necessary front end electronics.

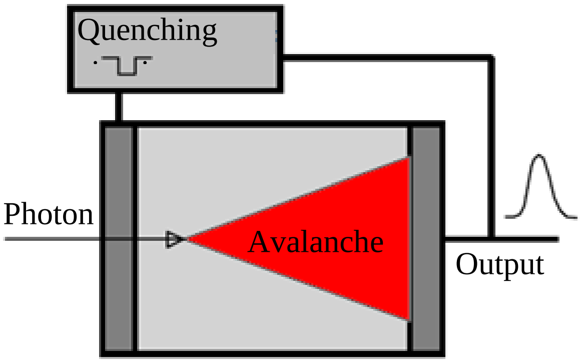


Fig 1.2: Single Photon Avalanche Diode (SPAD

The down-side is the increase in cost and complexity that results from bonding incompatible technologies. Furthermore, parasitic introduced by the external connections bring about degradation of system performance in terms of diminished sensitivity and speed [17]. Since standard CMOS offers an ideal platform for seamless integration of processing and readout electronics, it would seem that integration of SPAD in standard CMOS could offer significant advantages towards the implementation of compact high performance optical sensing systems.

1.3 CMOS Compatibility

There has been interest in a semiconductor replacement for the PMT since the 1960's [18-24]. However CMOS process is not optimized for high field conditions required for SPAD operation. The single photon electrical response of SPAD is contingent upon the volumetric breakdown of the planar depletion region. This requires effective isolation of the maximum electric field to the planar multiplication region, while counteracting the effect of electric field crowding around the curved edges of the device. Otherwise excessive electron tunneling and premature breakdown associated with the corners and curved edges will introduce noise and uncertainty in the operation of the device [25, 26]. In standard CMOS, custom features required for selective tailoring of the electric field distribution are not directly available to the designer.

An innovative layout design, introduced in [27], enabled Premature Edge Breakdown (PEB) suppression in standard CMOS by utilizing a diffused guard ring to modulate the high electric field at the periphery of the SPAD device. Despite the fill factor penalty associated with the diffused guard ring this was a breakthrough step, as it enabled, for the first time, the development of SPADs in conventional CMOS process [27-28]

1.4 CMOS Single-Photon Pixel – *Area Vs. Performance*

Implementation in conventional CMOS has opened the way for low cost development of SPAD pixels with increasing performance levels. However the added performance requires more integrated electronics which takes up additional silicon real estate resulting in a reduced fill factor. If the photosensitive area is kept constant, the loss in pixel fill factor will increase in direct proportion to its performance. The trade-off between performance and fill factor can be resolved if the active area is allowed to increase proportionally with the level of integrated functionality. However, in standard CMOS a linear increase in the SPAD active area leads to an exponential increase in its Dark Count Rate (DCR) [6]. The resulting loss in device Signal to Noise Ratio (SNR) makes the design of large active area SPADs not feasible in standard CMOS. Consequently, in practical applications, the active area of CMOS SPADs is ideally restricted to below 100 μm^2. This nonlinearity severely restricts fill factor in high performance SPAD pixels, often resulting in fill factors as low as 2 - 6% [30, 31].

The current trend is towards higher performance SPAD pixels and thus greater integration of functionality on pixel and on chip. The DCR imposed limitation on the active area creates a major tradeoff between performance and fill factor, preventing the full exploitation of the potential offered by CMOS-based SPAD detectors. The DCR associated active area constraint also presents a serious limitation for SPAD-based application as practical detection tasks often require a large photon collection surface [32].

Individual SPADs are characterized by their fill factor, bandwidth (maximum count rate), and DCR. Design methods capable of easing the trade-offs parameters at the device level would make for a more robust SPAD pixel which would naturally assemble into more robust arrays. So the first task in high performance SPAD array design must be performance optimization at the device level. This is the purpose behind the research work discussed in Chapter 2 and Chapter 3

1.5 CMOS Single-Photon System – *The Readout Challenge*

Under current implantation strategies, the effective performance of single SPAD Pixel within an array assembly is substantially lower, depending on the array size, than the performance of single standalone SPAD pixel. Due to the resulting degradation of effective single pixel response, when realized within larger multi-pixel assemblies, the development of a robust integrated large SPAD array with high performance is yet to be accomplished in CMOS and remains the primary area of research in the field [33]. The fundamental challenge has been the developments of an appropriate readout architecture that can effectively handle the massive amount of dynamic data generated by multiple SPAD elements without imposing significant bottleneck effects on the operation of the system. The conceptual illustration of the readout challenge appears in Fig. 1.3.

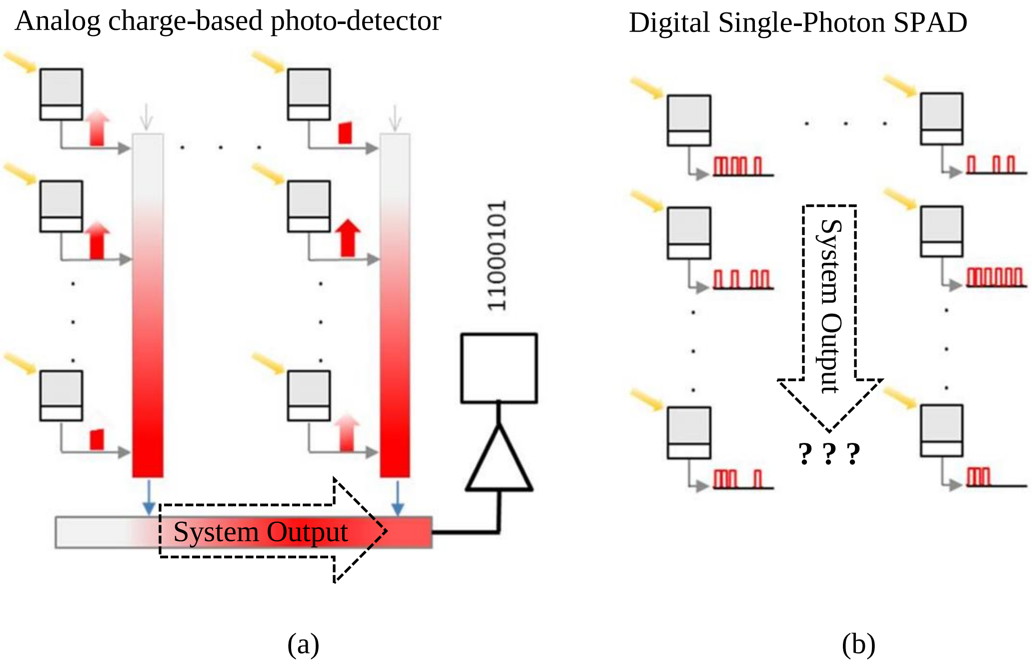

Analog charge-based photo-detector
11000101
System Output
(a)
Digital Single-Photon SPAD
System Output
? ? ?
(b)

Figure1.3: (a) Photo-signal statically stored in the pixel until it is sequentially readout and processed during the readout cycle. (b) Photo-signal must be dynamically processed in real time with each photon arrival, thus making local storage difficult

1.5.1 Digital Readout – SPAD Imager Array

The hall-mark of SPAD cutting edge operation is its single-photon sensitive digital response that can be generated with a simple in-pixel digital buffer and directly read out without amplification or pre-processing. Unlike CCD or Active Pixel Sensors, the dynamic response of SPAD cannot be statically stored as photo charge and must be processed instantaneously as either digital count value or Time Of Arrival (TOA) value, or both, before it can be stored. The otherwise excellent signal characteristics of a SPAD device pose a significant architectural challenge when it comes to reading out large arrays of such devices. The primary difficulty is characterized by the achievable efficiency in accessing pixel information as it is dynamically generated in large pixel arrays. One solution that enables massive operational parallelism and offer the best detection efficiency, is in-pixel processing of SPAD response [1, 34]. However large-area in-pixel counters or timing circuits significantly reduce the fill factor of the pixel and can prohibitively limit its photon-collection efficiency [1]. Simpler pixel-level processing improves fill factor and enable larger arrays, but would require a degree of resource sharing. Resource sharing involves multiplexing several signal paths within the same sharing group. This inhibits the dynamic operation of SPADs as digital pulses cannot share the same transmission/processing medium during overlapping time windows. The resulting performance trade-offs impose limitations on the size of the array. The fundamental challenge, and hence the bulk existing research, associated with the readout of large array of digital SPADs is fundamentally driven by the central tradeoff between efficient utilization of Silicon real estate (resource sharing) and throughput (effective transmission bandwidth per pixel).

1.5.2 Analog Readout – SiPM Detector

The severe bandwidth limitations (high rate of pulse pile up) associated with the dynamic readout of multiple digital SPAD pixels onto a common channel, described above, is eliminated by operating the array in what is known as the Silicon Photo-Multiplier (SiPM) mode. In this configuration individual current-mode pixel are parallel-connected to a common load As a result, when simultaneously fired, their signal will combine to produce an analog sum of all the avalanche signals [35-38]. Basic operation of a 4 cell SiPM macro-pixel is shown in Fig. 1.5.

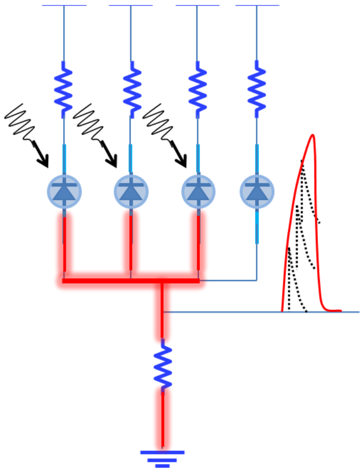


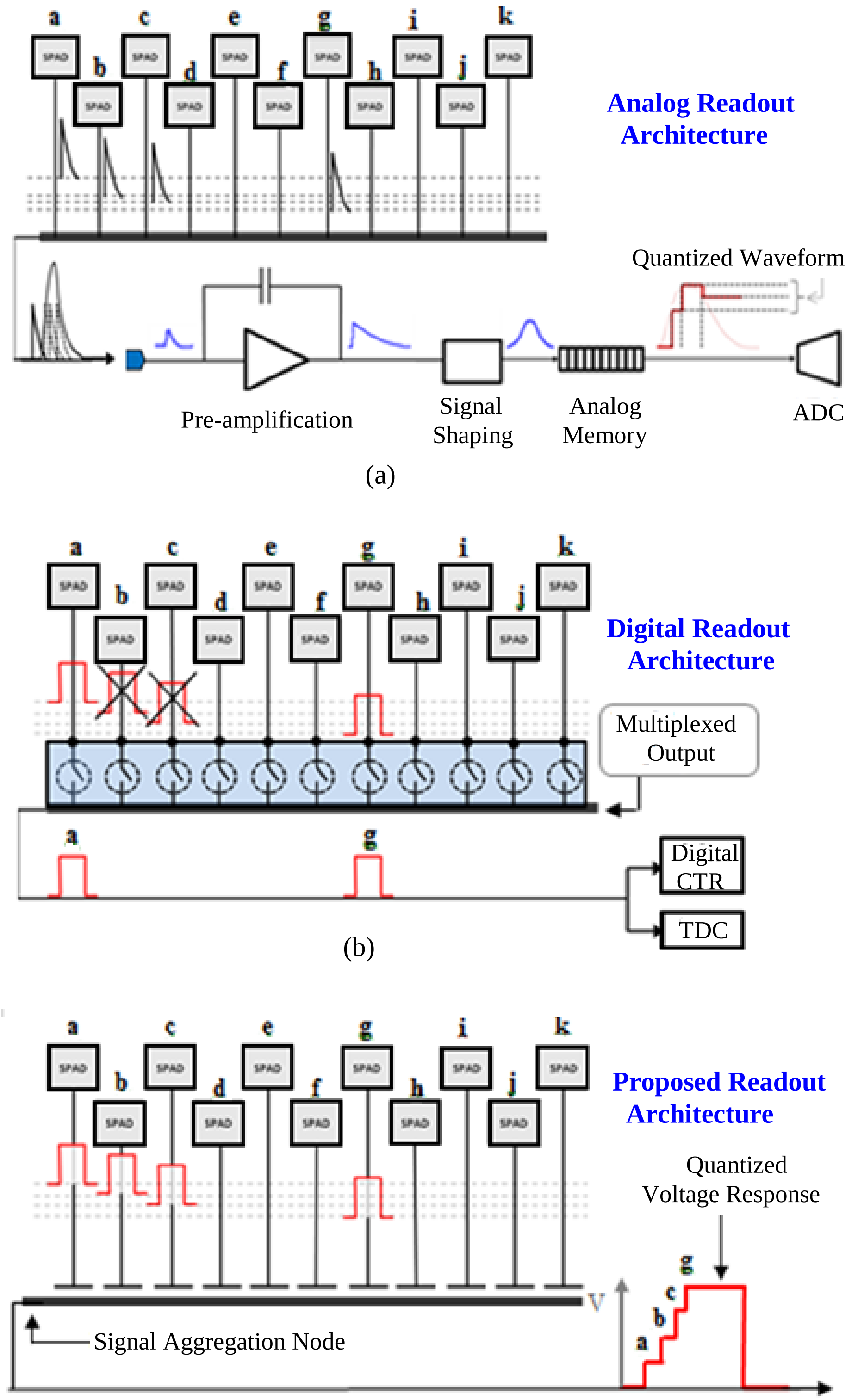
Figure 1.4: Current signal summation at the common readout node of the set of parallel connected individually quenched SPAD elements forming a single SiPM macro-pixel

In SiPM detectors, the asynchronous characteristic of analog operation enables large grouping of multiple SPAD elements without introducing performance constraints. The SiPM signal can be discerned from the background dark noise by amplitude thresholding of the summed current signal (signal amplitude) [39]. Therefore only photons that arrived within the summation window (avalanche current pulse-width) can be identified and quantified. The avalanche-gated analog summation represents the functional foundation of SiPM detector and ultimately limits the applicability of the technology. Furthermore the SiPM response signal is a faint analog current pulse with rapid transience and large active range. Digital extraction of count and TOA value room such a signal is no trivial task, involving multiple levels of signal processing [40].The electronic noise added in the process reduces the single-photon sensitivity of the detector while the foot-print area and power dissipation involved with the analog signal processing makes integrated readouts impractical.

1.5.3 Proposed Readout Architecture

SPADs are digital dynamical devices and they must be treated as such in order for their full potential as high speed single-photon detectors to be realized. The readout interface must preserve the advantage from both the dynamical (Asynchronous, high speed) and digital (noise free, direct readout) nature of SPAD response. The existing digital readout architectures preserves the digital characteristic of the SPAD device at the cost of its dynamic characteristics, while the automatically asynchronous behavior of the analog readout, preserve the dynamic nature of SPAD response at the cost of the digital feature of its operation. Chapter 5 presents a design for a novel array readout architecture that preserves both the digital and dynamical aspects of the SPAD operation. Fig. 1.5 compares the basic architecture of the analog, digital and the proposed readout architectures.

An integrated signal processing and digitization system has been designed to interface with the signal aggregation node in Fig. 1.5 c. The purpose is to establish free-



(a)

(b)

(c)

Figure 1.5: (a) SiPM analog readout based on dynamic current signal summation requires complex analog readout chain. (b) Multiplexed digital readout results in loss of overlapping pulses. (c) Proposed readout based upon digital readout at the pixel level and dynamic voltage division at the readout interface.

running dynamic digitization and readout of the detector output signal, not currently possible with the standard ADC digitization system deployed as part of SiPM analog readout chain. Fig. 1.6 illustrates the basic architecture of this integrated signal processing unit.

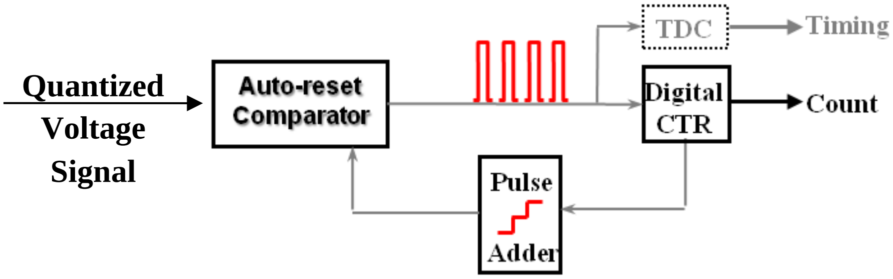


Figure 1.6: Signal processing system for the array detector interface of Fig. 1.5 c

1.6 Primary Objective and Goal

The over-arching goal for both single photon imagers and Silicon Photo-Multiplier is high sensitivity and high resolution solid state optical detection. The potential of SPAD arrays towards realization of this breakthrough seems obvious when considering the excellent single photon performance of a SPAD detector and the feasibility of their implementation in CMOS,. However, this potential is yet to be fully realized. The structure and operation of SPAD introduces several performance trade-offs and conflicting design objectives. The Development of large-scale SPAD arrays capable of reproducing the excellent performance of a single SPAD detector on a larger scale is contingent upon minimization of trade-off parameters associated with the single pixel operation and system-level operation of multi-pixel array. This feat is the primary challenge confronting research in the field. The advent of ultra compact and versatile process technologies will perhaps one day bring this goal closer to fruition – However for now the effort is in the direction of novel design concepts and new configuration techniques within the context of current low-cost, standard fabrication to realize large, high speed array of compact photon counters in a dense spatial arrangement.

1.7 Research Strategy and Organization

The focus of this thesis is to investigate a design approach to merge the best performance aspects of SPAD and SiPM design paradigm into an integrated high speed single photon counting system with greater detection sensitivity than either technology. The strategy involves a hierarchical approach to performance improving with operational requirement and performance of the next stage in mind. This means starting from the semiconductor junction level, moving up to the SPAD pixel eventually leading up to novel system design that can serve as a blue print for development of high speed SPAD array detectors featuring a scalable low-power integrated readout system with an unrestricted dynamic range, high fill factor and low noise operation.

The following brief summary of the upcoming chapters offer a conceptual organization of main research themes.

Chapter 2: This work centered around investigation of methods to improve DCR and fill factor by modifying the structure of the p-n junction. In conventional CMOS SPAD design, the fill factor degradation results from the un-scalable dead area taken up by the diffused guard ring. The alternative approach involved covering the perimeter of the junction with a poly gate, appropriately biased to deplete the area underneath it. PEB in the resulting SPAD device was achieved through reduction of the field at the curved edges or corners that constitute the periphery of planar p-n junctions, without increasing the dead area within the footprint of the SPAD pixel. We experimentally verify the effect of this field gate on lowering the DCR of the device. In addition geometrical modification of the p-n junction towards stated performance improvement has been studied and experimentally verified. Large-area SPAD pixels in standard CMOS have not been possible due to prohibitively large DCR. The purpose of this chapter is to demonstrate a large–area SPAD device in standard CMOS.

Chapter 3: This chapter focuses on performance enhancement at the pixel level which constitutes the next architectural layer in the system hierarchy. This work involved the development of front-end electronics capable of high speed quenching and recovery operation required to maximize pixel bandwidth. Towards this purpose a novel timing and logic circuit was designed to generate precisely matched and optimally synched quench and recharge signals with controllable delay between the two operations. Linearly variable delay enables fine tuning of the hold-off time required to determine the setting that optimizes the SNR (bandwidth Vs DCR). In order to standardize the readout pulse of the pixel so as to optimize the resource sharing requirement of the eventual system. For the first time an in-pixel tunable Event Pulse Generator (EPG) was integrated with the front-end electronics to act as the pixel interface. The pixel interface allows the output pulse-width of the pixel to be actively modified without affecting device operation. In clustered topology where multiple pixels use the same readout resource, the independent control of the pixel-level pulse width can improve system-level dynamic range, throughput and noise thresholding capability. The design of the Event Pulse Generator (EPG) involves re-using sections of the hold-off generation electronics in the primary AQC circuit, thus recycling some of the functionality already implemented in the primary circuit. This strategy enables a more compact design.

Chapter 4: The finite hold-off time separating the quench and recharge halves of the detection cycle introduces a non-linearity condition in the photon counting operation. In light of the associated performance non-linearity, empirically measured count rate data must be supported with an accurate theoretical description of non-linearity condition, associated with the counting process, before the measured data can be reliably interpreted. Towards this goal an analytical model describing the photon detection statistics for a detector with non-paralyzable dead time was derived. The proposed theoretical model can be used for any system that involves discrete detection of randomly arriving event in presence of a fixed dead time such as service time of a server in network traffic modeling or the refractory period of neurons in synaptic network simulation. The model was then used to theoretically predict the true count from a set of experimentally obtained corrupted counts that were measured in presence of a known dead time. The proposed model was compared against the standard model of count correction often used in the literature and shown to be more accurate, especially at higher count rates. The validity of the analytical model was experimentally verified.

Chapter 5: A novel signal collection mechanisms for a cluster of digital SPAD is introduced in this chapter. The signal accumulation scheme is based on voltage division across a floating array of minimum sized capacitors. The footprint of the interconnect interface is negligible relative to the area of the pixels. Signal accumulation scheme is architecturally simple and compact with an intrinsic asynchronous behavior and a dynamic range that is inherently matched to the number of pixels it supports. It is comparable to SiPM interconnect configuration in terms of simplicity and dynamical behavior without being marred by the SiPM associated readout obstacles and challenges.

Chapter 6: The charge redistribution principle, used in the signal collection phase to generate the quantized profile of the interface output response signal from a set of asynchronously arriving input digital pulses, is now implemented in a reverse operational order so as to bring about the disintegration of the discrete analog signal into its constituting stream of digital pulses. By sequentially processing the pulse stream with a compact digital counter, a binary code which represents the amplitude of the interface output signal, directly in terms of incident photon count can be generated and stored. Effectively a novel design for a highly compact and free running on-chip digital pulse counting system capable of sequential photon counting and photon number resolution is demonstrated

1.8 Contributions

This section will list the original contributions in the presented body of work.

a) Fill factor enhancement in standard CMOS SPAD device (Chapter 2)

b) DCR improvement in SPAD cell without area penalty, or additional processing cost.

c) New compact design for SPAD pixel front-end circuitry capable of high bandwidth operation and linear delay generation.

d) Introduction of a novel array coupling interface to the front-end electronics, The pixel interface features an independently adjustable event pulse generator and could potentially enhance the operational flexibility of the array system.

e) Derived a mathematical model for accurate prediction of fraction of measured count lost as a result of performances non-linearity in the courting process. The proposed model was able to accurately map the measured count data to the actual incident count even at high arrival rates where the standard mapping expression grossly over-estimates the true count.

f) Introduced a new acquisition interface for dynamic collection of logic pulses over a common interface method to simplify readout system for SPAD array detectors – but can just as easily be used for any system requiring readout or communication of digital voltage pulses over a single transmission line.

1.9 Significance to the Scientific Field

The proposed structural design lead to performances enhancement for large Area CMOS SPAD arrays operated at temperature. This will greatly benefit biological detection application like DNA sequencing and proteomics. SPAD-based single-photon sensing solution can be implemented with reduced constraint on the detection sensitivity due to dark count rates, and architectural design with significantly relaxed performances tradeoff and constraints. The implication of proposed sensing platform is expansion of the application scope for analog SiPM detectors and greatly reduced noise susceptibility. In the digital domain operation associated with photon-counting SPAD array SPAD the implications are lossless dynamic resource sharing without introducing additional readout complexity and throughput constraints.

Chapter 2

Performance Oriented Device Design

The photon counting efficiency of the CMOS SPAD is characterized by its fill factor, operational speed (maximum counting bandwidth) and Dark Count Rate (DCR). The Combination of these parameters determines the photon detection efficiency of the SPAD pixel in terms of its photon-collecting efficiency and Signal to Noise Ratio (SNR). The primary goal of this section is to investigate structural modifications at the device level that may lead to enhanced sensitivity and SNR. The process leading to the ignition of the characteristic avalanche phenomena starts with the absorption of a photon by the semiconductor material. Photon absorption depends on the semiconductor band structure and the wavelength of the incident light. Spectral selectivity and absorption efficiency is primarily determined by the choice of the material [41]. For electrical detection to become possible the photo-absorption process must be followed by a rapid multiplication of the generated carrier pair. It is essential therefore, that the photon is absorbed within or in the vicinity of the high field multiplication region (depletion region surrounding the on junction). Successful absorption of a photon also depends on the depth it has to penetrate before it is expected to be absorbed. The absorption efficiency of a photon as the function of the penetration depth is shown in the equation 2.1 where α represent the absorption coefficient of the material and L is the penetration depth.

$$\eta = (1 - e^{-\alpha . L}) \qquad 2.1$$

As the depth and the width of the depletion layer are determined by the parameters of the foundry process, the choice of fabrication process becomes an important consideration in meeting the spectral requirement of the application. Deep junctions are more efficient for absorption of longer wavelengths.

Another key characteristic informing the choice of the SPAD material is the Dark Rate. Impurities in the silicon create deep energy levels within the band gap structure and act as carrier trapping sites. The concentration of deep levels within the multiplication region determines the degree of after-pulsing effect as carriers, trapped in the high filed depletion region during an avalanche event are subsequently released after a statistical time delay, thus triggering false detection events [45].

When reverse biasing a basic p-n photodiodes at above the breakdown voltage, electric filed uniformity becomes an issue. Electric filed flux lines will crowd around curved surfaces and corners increasing the local electric field intensity at these regions. This effect leads to premature junction breakdown at the device edge, resulting in an uneven gain which is unsuitable for single photon detection. Suppression of premature edge breakdown (PEB) involves the reduction of electric field at the edge of the p-n junction. An effective solution to this is the localized reduction of doping concentration gradient at the susceptible regions. In dedicated CMOS process, thanks to availability of low-doped n/p well, selective shaping of the electric filed profile can be effectively achieved. One approach is based on implanting a highly doped disc at the center of the diode. The highest electric field is confined at the center of the diode for proper avalanche operation. [27]. The Same effect can be accomplished through decreasing the doping concentration at the periphery of the diode. A low doped p- region implanted at the periphery of the p-n junction decreases the electric field in this region as it diffuses into the higher doped layer, effectively lowering the local doping concentration [27, 28]. Another proposed structure for PEB suppression involves a floating field-limiting gate placed over the region separating a p-type floating guard ring and the primary p-n junction. The field gate acts to bridge the gap that may exist between the depletion regions induced by the floating ring and of the p-n junction respectively. This configuration widens the depletion region surrounding the curved edge of the device by connecting it to the depletions region of the adjacent ring. This reduces the local field intensity and counteracts the field intensifying curvature effect. Device breakdown voltage is enhanced as the result. A simple depiction of these structures is shown below in Fig. 2.1 [26].

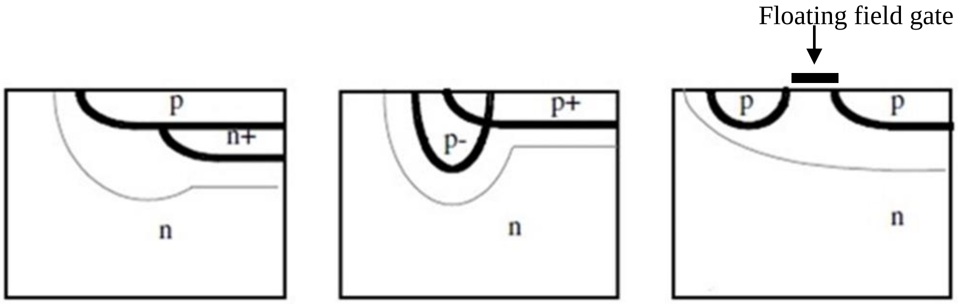


Figure 2.1: Different methods of isolating the high field region. (a) Center Field concentration. (b) Sidewall-Edge isolation (c) Low dope p adjacent ring

In dedicated fabrication technology it is possible to control the fabrication steps in such way as to inhibit PEB and minimize DCR by shaping the electric field profile at critical regions (Device surface or the curved region of the p -n junction).

2.1 CMOS SPADs

As a consequence of their customized fabrication process, other low-light semiconductor imaging technologies such as CCD and EM-CCD are able to offer low noise and efficient performance. It is also possible to develop specialized fabrication process for SPADs based on ultra-pure high resistivity silicon wafer with a non-planar proprietary technology that can yield excellent DCR and high signal to noise ratio[42]. However these gains come at the cost of high voltage operation, low fabrication yield, higher production cost, and diminished capacity for monolithic integration and system miniaturization. Implementation in standard CMOS, offers the best platform for implementation of large, highly compact array imagers with single photon sensitivity and integrated signal processing and readout [42]. Therefore focus here will be towards innovative design methods to achieve performance enhancement for Standard CMOS SPADs detectors and imagers.

2.1.1 Gain Uniformity – *Full PEB Suppression*

In a conventional CMOS process the only available implant layers are p+ and n+ diffusion, n-well (and p-well in twin tub process) and the p-type substrate. The only viable option for a planar p-n junction is a p+/n-well structure. Since substrate is common among all other devices on the chip, the p-substrate layer (p-sub) must be grounded. Therefor the alternative structure composed of a n+/p-sub junction involves a grounded cathode terminal (p-sub) with a high positive bias on the anode terminal (n+) in order to establish above breakdown revere bias condition. In this scenario, it would be impossible to connect either terminal of the diode to the ancillary (5V) CMOS electronics on the substrate. SPAD based on p+/n-well structure is the most plausible configuration in standard CMOS process.

In planar CMOS, PEB suppression is achieved by lowering the concentration gradient across the curved junction perimeter using a variety of diffusion layers available in number of CMOS processes. As an implementation work-around for the absence of a low-doped diffusion layer in the single-well vanilla CMOS process, a make-shift low-doped n-type region can be created at the curved edge of the p-n junction. This is achieved as the result of the lateral diffusion of an adjacent n-well ring surrounding the p-n junction [27, 28]. The structure is depicted in Fig. 2.2

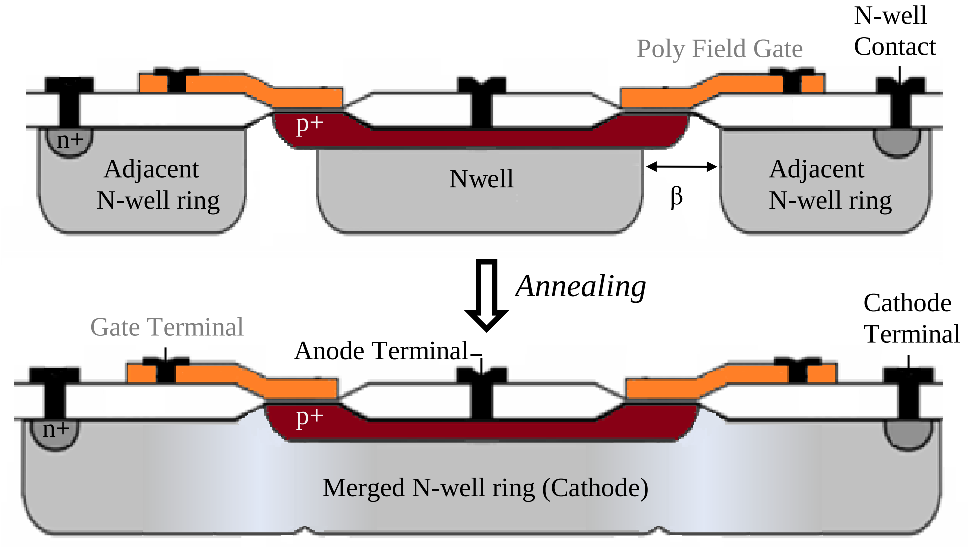


Figure 2.2: Formation of the diffused guard ring during the annealing process

Depending on the doping concentrations and the surface defect density, the charge concentration in the diffused guard ring might not have dropped enough to substantially lower the regional electric field intensity around the edges. In this case the guard ring will have failed at relocating the breakdown point sufficiently far from the surface [27, 47]. In such a scenario the center planar region might break down at only a few hundreds of mV ahead of the guard ring/edge region. Under this condition, even though it may still be possible to achieve a uniform gain (avalanche events associated with the volumetric breakdown of the planar region) the performance of the SPAD device will nonetheless suffer due to the heavy DCR contribution from the defect-rich surface edge of the device, where the statistical likely-hood of triggering an avalanche event still remains high [27].

2.1.2 Noise – *DCR Minimization*

In planar CMOS technology, due to the characteristic fabrication process (implantation through the top plane) the surface of the die is the most likely region for presence of impurities and structural imperfections that are formed during the high-energy implantation process. Another category of surface defects is the dangling bonds created due to the sudden interruption in the periodic arrangement of the silicon crystal at the surface. These surface states adversely impact the PDE of the device by promoting the recombination (annihilation) of photo-generate carriers created in the vicinity of the surface, before they can diffuse down to the multiplication region [53]. Furthermore, surface states act as active generation site otherwise known as SRH sites (Shokley Reed Hal). The high density of SRH sites at the surface heavily contribute to the overall DCR.

When there is only a small difference in break down level of the defect rich surface and the planar junction of the device, thermal and tunneling generations at the surface junction could still contribute substantially to the device DCR. Therefore, in order to achieve good SNR performance on top of preventing Premature Edge Breakdown, it would be highly advantageous to alter the breakdown voltage at the guard ring/edge region relative to that of planar junction by as much as possible [23].

A second mechanism may be used, in tandem with the diffused guard ring, to assist with further lowering of the field at this region [46]. A three dimensional model of such a device is illustrated in Fig. 2.3 (a). A negatively biased poly gate placed over the perimeter of the surface junction surface is used to deplete the underlying n-well region along the curved side-wall of the device, to a degree proportional to the applied negative voltage [46, 47]. This is conceptually illustrated in Fig. 2.3 (b). Widening of the peripheral depletion region will further suppress the electric field at the surface edge and along the curved side-wall, thus reducing the DCR contribution from these regions and improving the overall noise performance of the device.

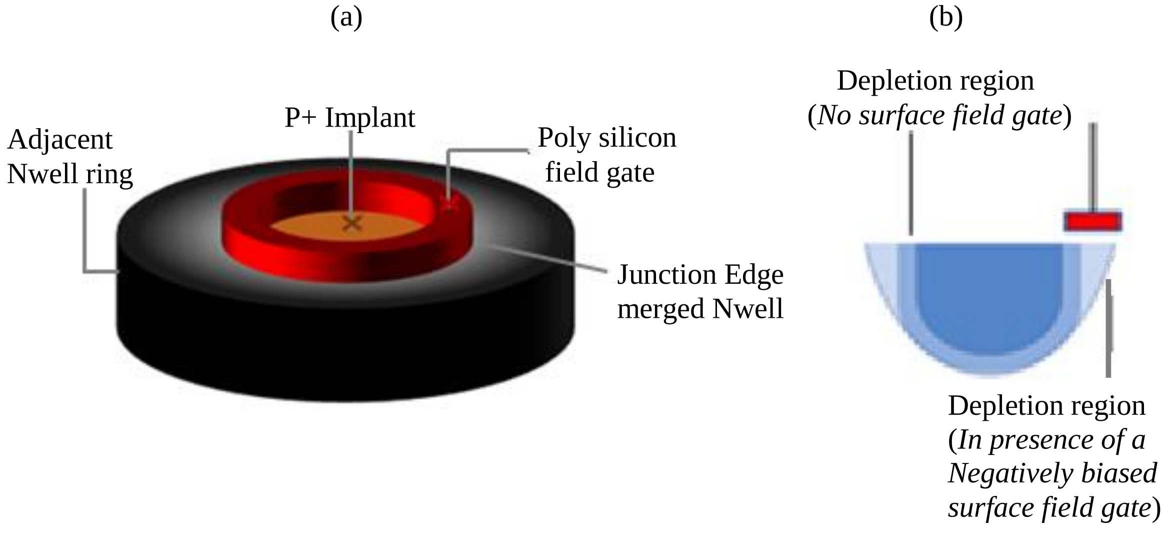

(a)
(b)
Adjacent Nwell ring
P+ Implant
Poly silicon field gate
Junction Edge merged Nwell
Depletion region (*No surface field gate*)
Depletion region (*In presence of a Negatively biased surface field gate*)

Figure 2.3: (a) Three dimensional view of the device illustrating its ring-like structure. (b) Conceptual illustration of the control gate effect in altering the shape of the depletion region around the junction perimeter and modulating the local electric field intensity at the surface and along the curved junction.

The device structure illustrated in Fig. 2.3 was designed in single-well standard CMOS process and used to study the effects of varying the field gate strength on the DCR. This measurement was performed using the 5 lambda (ring separation) structures in [44]. We experimentally demonstrated, for the first time, the strong diminishing effect of the field gate voltage on the DCR of a SPAD device in standard CMOS. DCR measurements were taken for a wide range of gate bias values [48]. Fig 2.4 shows our experimental data.

The diffused guard ring surrounding the active area of the SPAD constitutes a photon dead area. Since the minimum dimensions of the guard ring are set by the process parameters it does not linearly scale down in area as the pixel size is reduced. Consequently the resulting dead area will progressively occupy a larger ratio of the total pixel area for smaller pixels resulting in the degradation of the pixel fill factor. This will introduce a performance trade-off between sensitivity and array size and is considered an impediment towards implementation of large SPAD array detector. A clear advantage of using the field gate to lower the device edge field rather in place of the diffused guard ring, in addition to its demonstrated superior performance, is its area efficient implementation. The surface field gate does not affect the geometrical Fill Factor of the device [45, 46, 48].

DCR measurements as a function of the gate bias were recorded for SPAD devices with varying ring separations within Nwell encroachment range, ranging from widest gap possible (greatest dilution of doping density around the junction edge) to no gap at all (no change in the doping density around the junction edge junction). It was observed that at some threshold field gate value (-4V in the measurement) the effect of the field gate, in reducing the DCR, becomes the dominant factor overshadowing the impact of the diffusion ring. The DCR continues to drop as a function of the gate bias. This is a significant observation as it implies that SPAD with uniform gain and high noise performance can be developed in conventional CMOS process without the need for fill factor diminishing features such as the diffused guard ring.

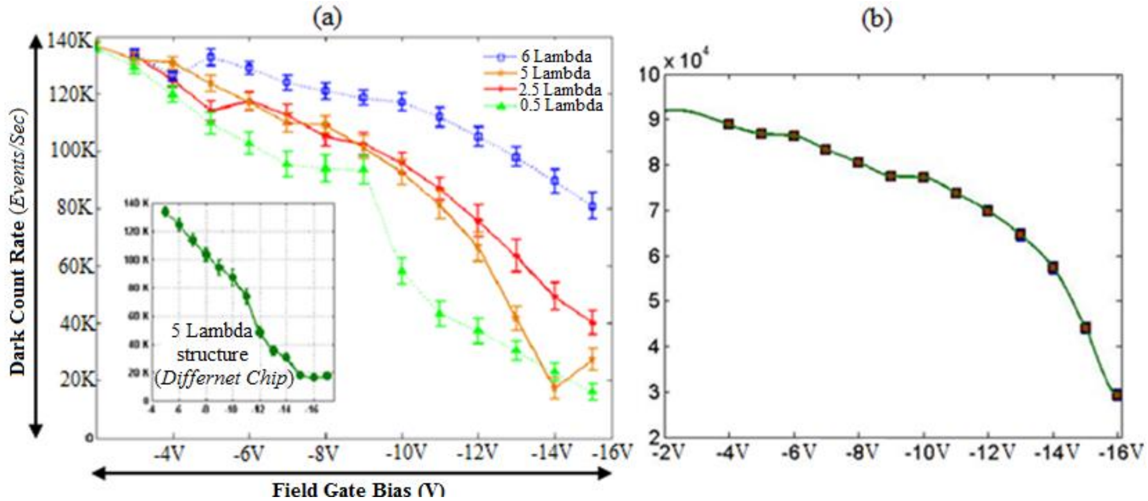


Figure 2.4: (a) Measurement of DCR for various gate bias values. The data was recorded for 3 different SPAD structures. In all cases, the DCR was observed to drop substantially with increasing the negative bias of the gate terminal. At starting voltage as low as -4V the effect of the gate becomes dominant over the diffusion ring. (b) DCR measured as a function of the field gate bias for the SPAD structure with 5λ ring gap from a separate die.

The elimination of the diffused guard ring and incorporation of a surface field gate lead to remarkable improvement in the fill factor and DCR performances of the device. The new area efficient structure represents a significant advantage for the implementation of high performance SPAD-based array detectors.

2.1.3 Low DCR Pixel Geometry

Elimination of the diffused guard ring does not completely resolve the fill factor woes in high performance SPAD design. Other non-scalable features such as quenching resistor or the front-end circuitry can significantly diminish the fill factor. Low cost development of high-performance smart SPAD pixels is the hall mark of implementation in conventional CMOS. However the added performance requires more integrated electronics which takes up additional silicon real estate and comes at the cost of fill factor reduction. The impact on the fill factor is especially problematic for pixels with a small active area. The silicon foot print of the integrated electronic is determined by the minimum feature size of the process technology and does not scale down with pixel area. Therefore if the photosensitive area is kept constant, the loss in pixel fill factor will increase in direct proportion to its performance (integrated functionality). The trade-off between performance and pixel fill factor can be resolved if the active area is allowed to increase proportionally with the level of integrated functionality.

Rows of large area high bandwidth SPADs are indispensable in high speed imaging and parallel processing of photon-based data streams in assay analysis such as DNA sequencing and parallel fluorescence correlation spectroscopy [49-54]. Detector arrays composed of limited number of SPAD pixels with large pixel area also plays a critical role in adaptive optics [55, 56].

The major performance hurdle with large area SPAD in standard CMOS process is the non-linear relationship between active area and device DCR. Theoretically speaking, the magnitude of the DCR should linearly scale with the active area, yet in actual implementation it exhibits an exponential growth profile relative to the active area [59, 60].

In standard CMOS a linear increase in the SPAD active area leads to an exponential increase in its DCR. Thus any performance gain resulting from a larger detection area is more than offset by the disproportionate degradation of device SNR performance. The resulting loss in the SNR makes the design of large active area SPADs unfeasible. Consequently, the active area of CMOS SPADs is ideally restricted to below 100 μm^2. This nonlinearity severely restricts the fill factor in high performance SPAD pixels, often reported with fill factors as low as 2 - 6% [30, 31]. The DCR imposed limitation on the active area further aggravates the existing tradeoff between performance and fill factor. The performance impact of this non-linear characteristic is especially relevant considering that the current technology trend, driven by ever-tightening performance demands of emerging applications, is towards greater integration of functionality in pixel and on chip.

The nonlinear increase in DCR is caused by a non-uniform defect density distribution across the active area. The gettering process is a fabrication step by which the substrate defects, which act as carrier generation sites, are made to diffuse out of the active area so that they no longer contribute to the active operation of the device [8]. In relation to SPAD operation this means excessive DCR contribution from a large number of generation sites is diminished. However, the gettering operation in CMOS is optimized for small active area typical of electronic components, and therefore becomes less efficient for larger active areas. In a study undertaken in [47, 59] it was reported that as a consequence of this fabrication process step, which induces doping displacement towards the junction boundary, the density distribution of generation sites is not uniform over the junction area; rather it increases radially outwards from the core center of the active area. Consequently larger active areas, especially ones with diameter in excess of 25µm (500 µm^2) contain a disproportionately larger number of generation sites as a result of their diminished gettering efficiency.

From the several alternative geometries investigated in [49, 53], a cigar-shaped topology was found to involve a more effective gettering dynamics, for the same active area, due to the stretched profile of the structure. We have investigated the effect of the perimeter field gate on the modified junction geometry and characterize the performance of the new large area SPAD featuring the perimeter field gate and the modified junction geometry in terms of its breakdown and DCR properties.

In order to characterize the reverse break-down electrical response of the new junction geometry relative to the circular one a number of capsule-shaped SPAD structures were designed and fabricated. The layout of the test structures used in the measurement is shown in Fig.2.5. The devices have different shapes but identical structural composition consisting of a p+/Nwell junction with a poly-silicon gate covering the junction boundary. Devices one, two and three incorporate a capsule-shaped geometry with identical active areas but different perimeter-area ratio. Although these devices are tailored for high gettering efficiency relative to circular device, the one with greater perimeter-area ratio should theoretically offer a greater gettering advantage as it offers a greater perimeter for the defects to diffuse through [48]. Device 4 represents the standard circular geometry.

Device characterization was performed using two different experimental techniques. The first method, as stated above, involved measurement of breakdown characteristics for all three devices. This was realized by directly connecting the anode terminal of each device to a Source Measure Unit while grounding the cathode, and setting the field gate bias.

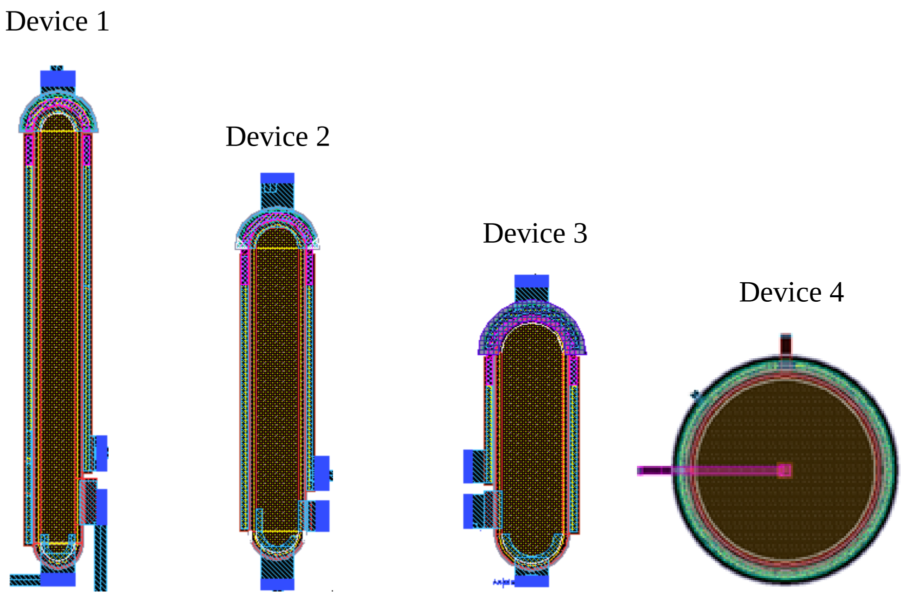


Figure 2.5: Structure of SPAD consisting of P+/Nwell junction with a poly-silicon field gate covering its perimeter is shown in 3 different implementations. Device 2 and 3 are two representations of the P+/Nwell junction with a geometrical profile that features a high gettering efficiency meant to amplify the impact of the field gate on DCR. The Standard round geometry appears as Device 1.

The anode voltage was then swept in the reverse bias direction until a sudden rise in the reverse current was detected. The process was then repeated for different values of gate terminal bias. The second method of characterization involved quantitative assessment of DCR as a function of the field gate bias for the two cigar-shaped SPAD devices. This method required the integration of an active quench and recovery circuit as part of the SPAD pixel in order to facilitate high bandwidth counting of dark events. A basic diagram of the general set up for the measurement process appears in Fig 2.6

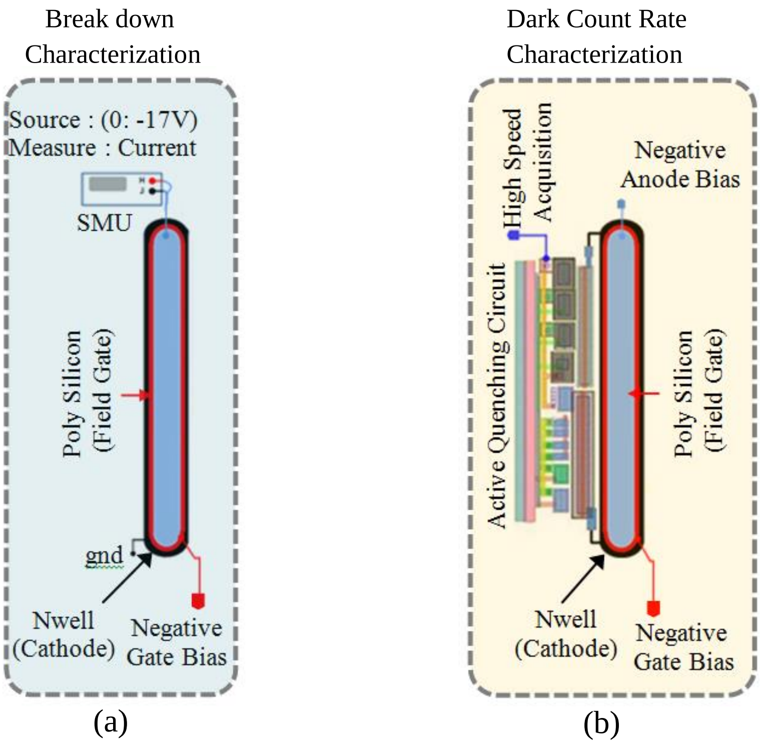


Figure 2.6: (a) stand-alone SPAD unit for breakdown characterization. (b) SPAD unit with in-pixel active quenching electronics for DCR characterization.

2.1.4 Experimental Results

The reverse biased Current-Voltage curves for device 3 at different gate bias appears in Fig 3 with the inset representing measurement for a similar round SPAD structure from [46]. As can be seen from the Fig 3 full volumetric breakdown, characterized by sudden and rapid rise in the current occurs at smaller gate voltage for the capsule-shaped SPAD relative to the round SPAD.

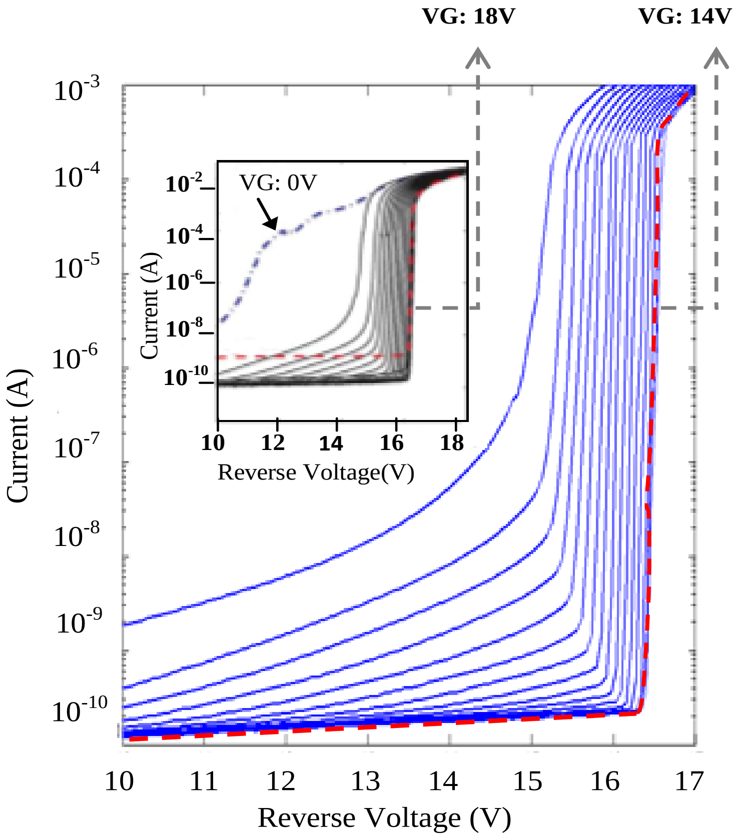


Figure 2.7: Reverse bias current-voltage relationship for device 3. The inset represent measurement results reported in [44] for the round perimeter gated junction. The Most optimal avalanche characteristics are represented by very sharp and sudden rise in current indicating full volumetric breakdown at the planar junction, indicated by the trace.

The breakdown voltage for each test structure shown in Fig. 2.7 was measured for three device instances extended across three chips. The mean value for breakdown voltage along with the standard error has been determined from the measured samples and is presented in Fig. 2.8(a). For the sake of comparison the breakdown data are displayed together in Fig. 2.8(b).

It is interesting to note that for the elongated structures the standard error associated with the mean breakdown voltage decreases with increasing field gate bias

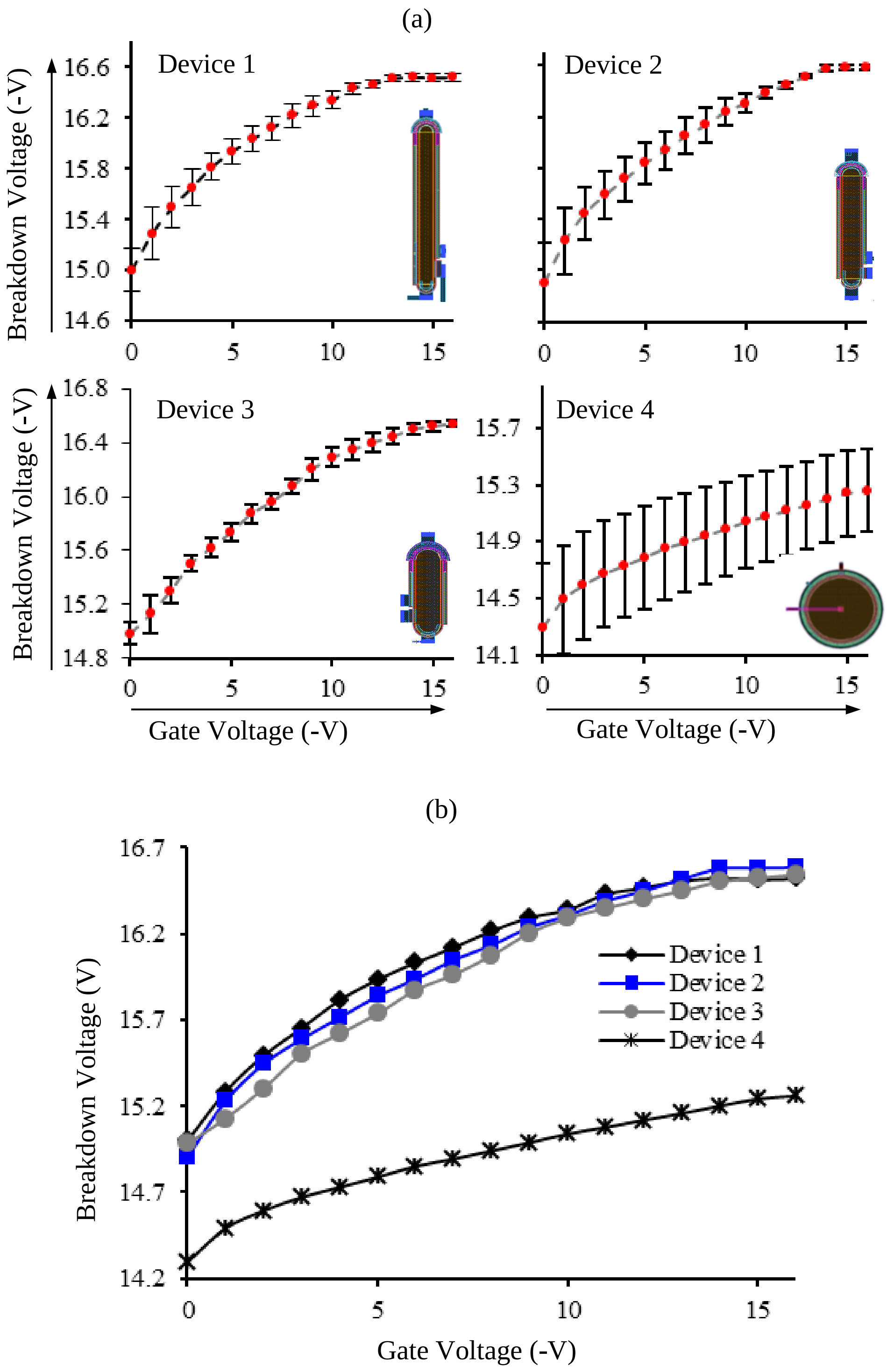


Figure 2.8: Mean breakdown voltage as a function of the perimeter field gate bias measured for devices 1, 2 and 3 representing the high gettering efficiency structures and device 4 representing the round SPAD structure.

while this pattern appears to be less prominent in the device with standard round geometry. Variability in the breakdown voltage results from the non-uniform spatial distribution of defects and impurities across the die. In the perimeter-gated SPAD structure with improved gettering dynamics, the highest concentration of active area defects are found near the periphery of the junction which corresponds to the region acted upon by the field gate. Under these conditions, increasing the gate bias strongly lowers the magnitude and hence the variance associated with the mean number of active generation sites, resulting in greater spatial uniformity of breakdown voltage. Device 4 represents a SPAD structure that is not optimized to enhance the gettering action, resulting in a spatial pattern of defect and impurity distribution that does not favor the boundary region. For such a structure, the device operation is defined by partial breakdown of the p-n junction at defect clusters and impurity precipitates that create localized regions of electric field maxima within the active area. Under these conditions, localized centers of electric field maxima fall outside of the field gate's region of influence; therefore increasing the field gate bias has no effect on the breakdown response of the device. As a result, despite full suppression of PEB effect, the breakdown response becomes diminished and largely independent of the perimeter field gate. This is observed in the breakdown characteristic of device 4 in Fig. 2.8

Another corroborating observation for the optimizing effect of elongation on the performance of the field gate is the enhanced slew rate of the breakdown response as a function of the gate bias illustrated in Fig. 2.8. The accelerated rise in the breakdown voltage relative to the rate of increase in the field gate bias suggests that the impact of the field gate is more pronounced in the capsule-shaped structures as compared to the circular structure. Another immediately obvious feature is the higher breakdown voltage of the capsule-shaped geometries at zero gate bias, observed in Fig. 2.8 (b). This indicates an enhanced electrical profile of the native diode. Also observable in Fig. 2.8 (b) is the notable similarity in the breakdown profile among the capsule-shaped structures. This finding is consistent with the DCR measurements shown in Fig. 2.9.

In order to prevent counting errors due to the finite measurement bandwidth of the system, the maximum expected DCR was capped by selecting an appropriate starting gate bias. The mean and standard error value of the DCR as a function of the gate bias were determined for each capsule-shaped structure in accordance with the same guidelines used in breakdown characterization. DCR was evaluated for reverse bias conditions corresponding to 0.3V and 0.9V of excess bias. The results presented in Fig. 2.9 (a-c) demonstrate a significant decrease in the DCR in response to the increasing field gate bias. The enhanced DCR-reductive effect of the field gate in the capsule-shaped devices is attributed to active generation sites which have shifted towards the border region where their emission activity is muted by attenuation of local electric field. Four orders of magnitude of reduction in DCR is observed for both values of excess bias parameter, for a field gate voltage of -14V relative to 0V. Fig. 2.9 (d) shows the DCR measured as a function of the reverse junction voltage at a gate bias value of -14 V (at which point the effect of the field gate has saturated, for the capsule-shaped devices). The DCR is 10 Hz for an excess bias value of 0.3 V and no discernible increase is observed as the bias is increased to 0.6V. Minimal rise in the dark count activity despite the rise in avalanche probability indicates a negligible contribution from trap-assisted carrier generation (thermal and tunneling) within the active area. The implication is that the central active area is largely void of trapping sites attributed to the presence of impurities and defects. The rise in DCR observed when the excess bias is increased to 0.9 V may be attributed to the increased likelihood of direct band to band tunneling within the active region which occurs without trap assistance. However direct band to band tunneling requires strong electric field and is unlikely in silicon due to its indirect band gap structure. It is more likely that the rise in DCR is due to the resumption of trap-assisted emission activity from the muted generation centers near the active area boundary, re-activated by the increased field strength. We speculate that deployment of a wider field gate for greater attenuation of the edge field may allow the device to operate at higher excess bias levels without the associated increase in DCR. The data illustrated in Fig. 2.9 was generated for capsule-shaped topologies, all featuring an active area of 3400 µm^2.

For comparative illustration, the mean DCR for the capsule-shaped structures associated with different perimeter-area ratio are presented together in Fig. 2.9 (d) and (e). Although all the elongated SPAD devices demonstrate similar performance, device 3 offers a more favorable profile for arrays and is therefore considered a practical option for incorporation into a SPAD pixel array.

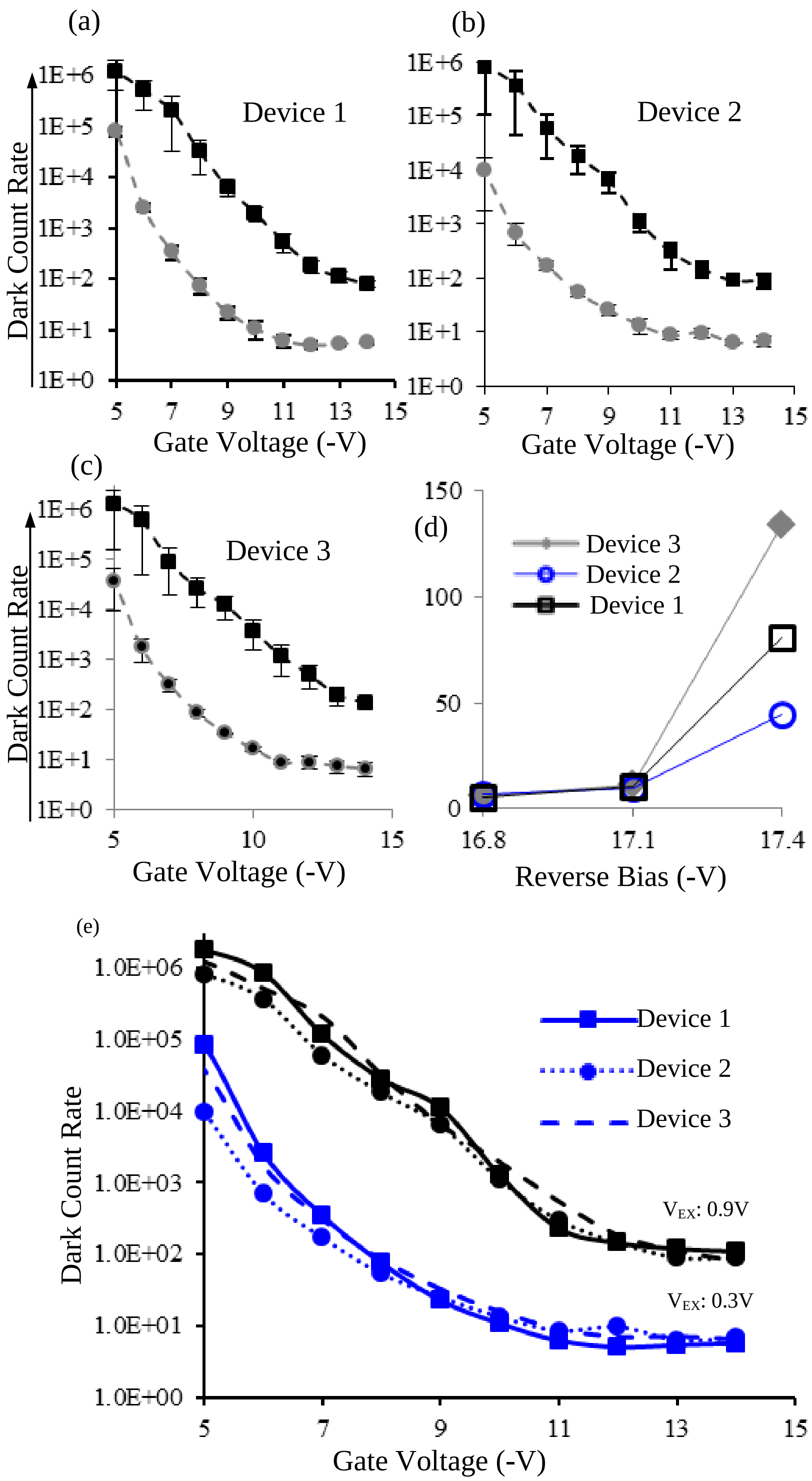


Fig. 2.9 (a-c) Measured DCR as a function of the perimeter field gate bias for devices 1, 2, and 3 operated at two different excess bias levels (0.3V and 0.9V). (d) DCR vs device reverse bias at -14V field gate bias. (e) DCR vs perimeter field gate bias for devices 1, 2, and 3 at excess bias of 0.3V and 0.9V.

2.2 Summary

The effect of poly gate placed over the p-n junction border was experimentally studied for the first time. The gate bias showed to be highly effective in reducing the DCR with the degree of suppression strongly related to the gate bias magnitude. The central performance challenge in large area SPAD implementation (particularly well suited for astronomical and biological applications) is the DCR-related active area limitation.

Large area SPADs in standard CMOS are characterized by prohibitively large DCR. Perimeter Field gate driven with a high negative voltage was previously shown to reduce the DCR by several orders of magnitude. The minimum achieved DCR was still considered high for many applications. Here we enhanced the perimeter-gated design by replacing the round SPAD design with a capsule-shaped junction geometry shown to have a better defects density profile. The Dark Count was experimentally measured for the new device. It was shown to be an improvement over the traditional circular geometry by several orders of magnitude. Additionally the magnitude of the gate bias required for the optimum effect also showed significant reduction. The DCR of SPAD device with an active area of 1200 μm^2 was reduced to as low as 10 Hz for 0.5 volt, and 180 Hz for 1 volt of above breakdown excess bias. The reported results represent an improvement of more than of 5 orders of magnitude relative to the standard design.

Devices geometries with width to high ratio close to unity (i.e. circles and squares) have the best avalanche propagation property and more practical photon collection shape profile. Subsequently the extended profile of the cigar shaped SPAD (large perimeter to area ratio) are associated with lower avalanche propagation speed and less efficient photon collection profile but offer highly improved DCR characteristics. An intermediate geometry with less extension was designed and tested. The results showed no change in the DCR response relative to the junction geometry with the maximum extension. Consequently the intermediate geometry constitutes a more optimal choice for incorporation into SPAD pixel. The DCR measured for the new SPAD device with a relatively large active areas was reduced to as low as 100 Hz. This is an improvement in excess of 4 orders of magnitude relative to the standard design.

Chapter 3

Performance Oriented Pixel Design - *Front-end Electronics*

The avalanching characteristic of a p-n junction reverse-biased above breakdown is the enabling feature for SPAD-based photon counting/timing devices. However, avalanching is a runaway destructive process and must be controlled [61, 62]. The previous section focused on structural alterations and junction design techniques to direct the avalanche breakdown process away from the junction edge and towards the planar region where uniform volumetric breakdown can take place. These techniques involved modifications to the structure of the SPAD device which involved changing the geometrical shape of the p-n junction and introducing a field gate over the surface edge of the junction. The effect of the field gate and junction shape in suppressing the surface field intensity and consequently reducing DCR were also discussed and experimentally demonstrated. In this chapter we approach the performance challenge at the next architectural level higher, which involves in-pixel integrated functionality implementation.

3.1 External Gating Signal – Hybrid Design

In a SPAD detector, charge avalanche is suppressed through lowering of the device terminal voltage to below the breakdown level (V_b). Recovery process involves restoration of the device terminal voltage to it initial over-breakdown level (V_b + V_{EX}). In the gated mode, the SPAD is biased just below its breakdown level (V_b). The bias is transiently increased to above the breakdown voltage (V_b) by a predetermined amount (V_{EX}) through application of a calibrated voltage pulse. The Device is kept at above breakdown level for short periods of time, corresponding to the gating pulse width (T_W), during which time an avalanche event may occur. The falling edge of the gate pulse lowers the device terminal voltage level and quenches the avalanche current. The duration of the pulse is kept short to limit the avalanching duration through the p-n junction, usually to around a few nanoseconds [42, 63]. Gating operation also significantly limits detector DCR by restricting the measurement bandwidth at the SPAD device level. A schematic depicting a passively controlled SPAD cell in gated mode is shown in Fig. 3.1. The gating pulse is applied to the SPAD through capacitor C_g.

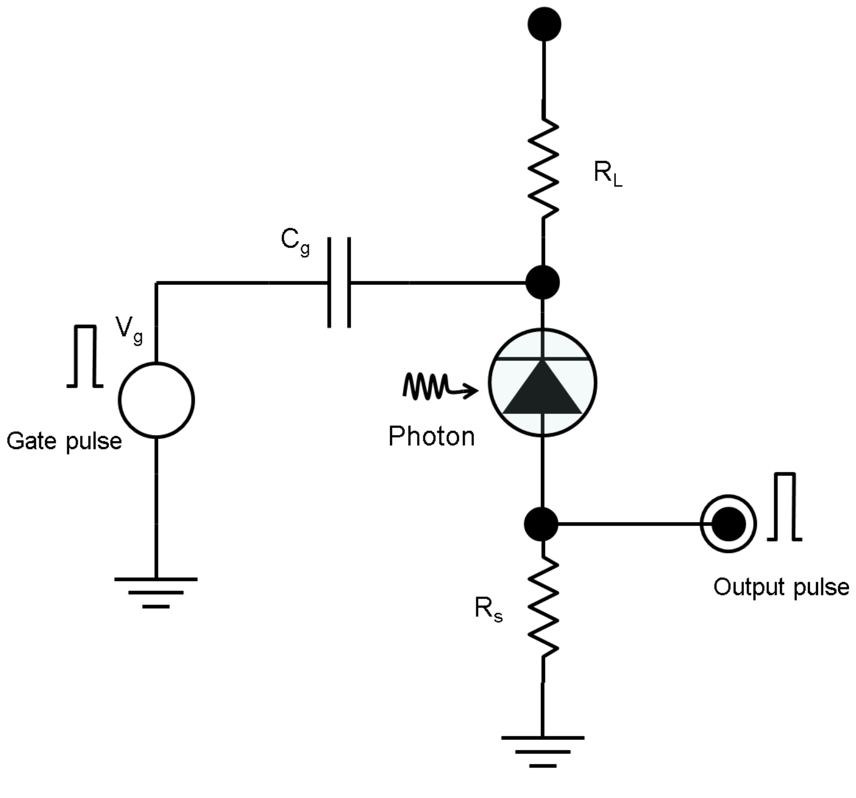


Figure 3.1: Schematic of Passive-quenched SPAD in gated-mode operation

The coupling capacitor C_g forms a parallel network with the capacitance at the SPAD cathode terminal ($C_{depletion}$ + $C_{parasitic}$) and the quenching resistor R_L. In order for the significant portion of the applied voltage to appear across the SPAD terminals, the coupling capacitance C_g should be much larger than (C_d + C_p). Since the window of operation is restricted to the immediate time window following the arrival of the optical signal at the detector, the noise contributions from after pulsing and thermally induced events can be significantly reduced.

The duration of the quench and recharge operation is determined by the slew rate of the falling and rising edge of the gating pulse respectively. The speed up in the reset and recovery operation improves the detection bandwidth and the active range of the device. However, since the gating pulse train must be synchronized with the photon arrival time, the utility of the gated operation is limited only to applications wherein the arrival time of the optical stimuli is predetermined or can be accurately attained through initial scanning.

3.2 Integrated Electronics – *Monolithic Design*

In most application, the arrival of the photon is not pre-determined or cannot be obtained with any reasonable accuracy [64]. Therefore, the detection event (subsequent rising edge of the avalanche pulse) must act as an internal dynamic timing signal to trigger the quench/recovery process. The quenching speed mediates the duration of the avalanche event. As a result, early avalanche detection followed by rapid quenching can significantly improve detection speed and SNR of the system. This action requires robust circuitry that can detect the onset of the avalanche event and actively inhibit the avalanche process followed by rapid restoration of the quiescent biasing condition.

Controlling the delay between the quench and recovery process is a critical factor in design and implementation of AQC. During an avalanche event, some of the free carriers are captured by the trapping sites within the depletion region. These carriers are subsequently released at random times following the avalanche resulting in after-pulsing phenomena. Consequently, it is sometimes desirable to keep the SPAD quenched for a period of time necessary for the trapped charges to be released before restoring the device back up to an avalanche ready state. This synchronization is not possible with passive quenching and requires more sophisticated avalanche control circuitry [65]. Before discussing the aspects of an Active control circuitry, an improved passive control scheme, utilizing a biased active device in role of the ballast resistor is introduced.

3.2.1 Passive Avalanche Control with Active Device

The performance of the SPAD unit can be improved by passive quenching with a PMOS transistor rather than a resistor. The performance improvement in the area of fill factor and dynamic range can be significant. The fill factor advantage is due to the smaller foot print associated with the PMOS transistor. The Dynamic range advantage is the consequence of faster quench/recharge operation that becomes possible with a properly biased PMOS transistor. The speed increase is especially obvious during the recharge operation.

Fig. 3.2 illustrates the difference in quench/recharge time between the two static quenching methods. The simulation was performed using the SPAD device model described in Appendix A.

Although substantial improvement is observed in the recharge time, quenching time is not affected by the same degree. Active device use in passive avalanche control also has the benefit of superior fill factor. There is a direct relation between optimizing performance and minimizing the avalanche charge [60, 66]. Therefore reducing the quenching time is a key design objective. Further improvement in the quench/recharge cycle duration is possible with active quenching systems [67-70].

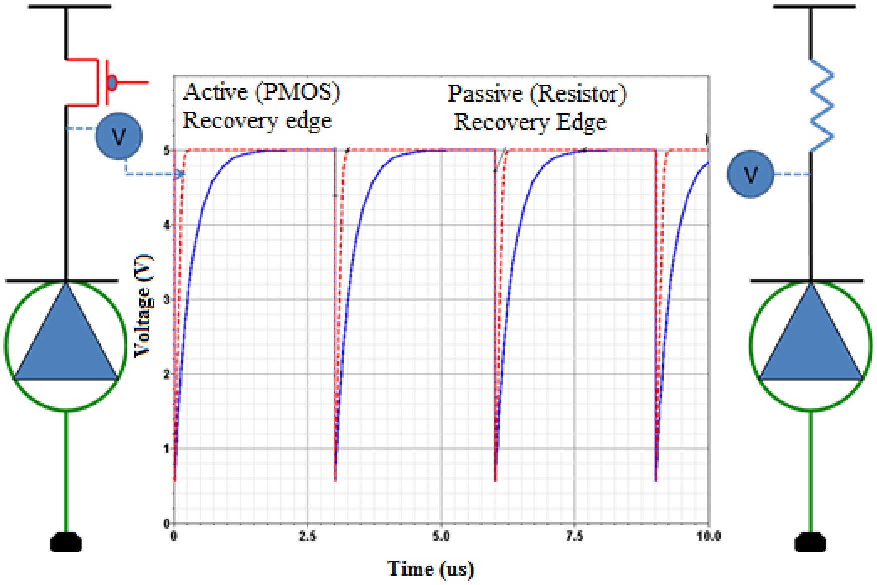


Figure 3.2: The dashed waveform represents the cathode terminal voltage during normal avalanching operation of PMOS quenched model. Solid waveform represents the cathode terminal voltage of the resistor quenched model. Recharging time of 220 ns is achieved with the PMOS component compared to 1.67 µs achieved with the ballast resistor of minimum size. The total SPAD capacitance was set to 3 pF in the simulation model

3.2.2 Active Avalanche Control

The front-end circuitry for active control of the SPAD avalanche current includes three basic functional features as illustrated in Fig. 3.3. These include active quenching of the avalanche current once it has been triggered, recharging the bias level back to the above breakdown level necessary for avalanche ignition and the timing control of the quench and recharge operations [70]. The circuit implemented for this purpose is often simply referred to as AQC (Active Quenching Circuit). In terms of performance, key feature of an AQC is the ability to shorten the duration of the avalanche current.

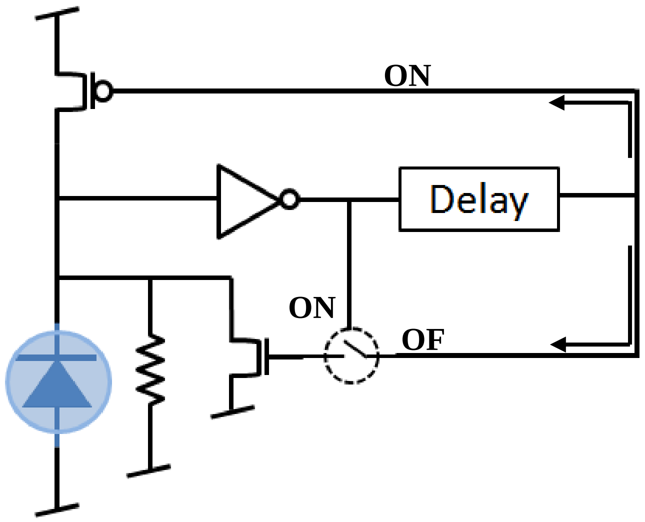


Figure 3.3: Basic block diagram of front-end system responsible for active quenching and recovery functionality in SPAD pixel.

Output pulses associated with a passively quenched SPAD will have random amplitudes. This occurs when avalanche occurs while the SPAD terminal voltage is not at the full bias level V_{EX}. These output signals are referred to as runts [42]. Fluctuations in the height of the output pulses complicate the noise filtering process (amplitude thresholding). This complication is eliminated in SPADs with active quenching front-end due to the standardized output pulses amplitude.

A popular AQC scheme presented and patented in [71, 72] is often adopted in the literature to supplement a particular SPAD implementation. The basic structure of this circuit is shown in Fig. 3.4. It utilizes a mixed mode passive/active quenching methodology. Quenching is carried out by a passive resistor which starts to acts upon the avalanche current immediately. This quenching drive is later bolstered by a NMOS transistor acting as a current sink. Re-charging is effectively accomplished by means of a PMOS Transistor which supplies the SPAD capacitance with current from the rail voltage source during the recovery process. Timing control and delay are generated by a monostable circuit [71, 73].

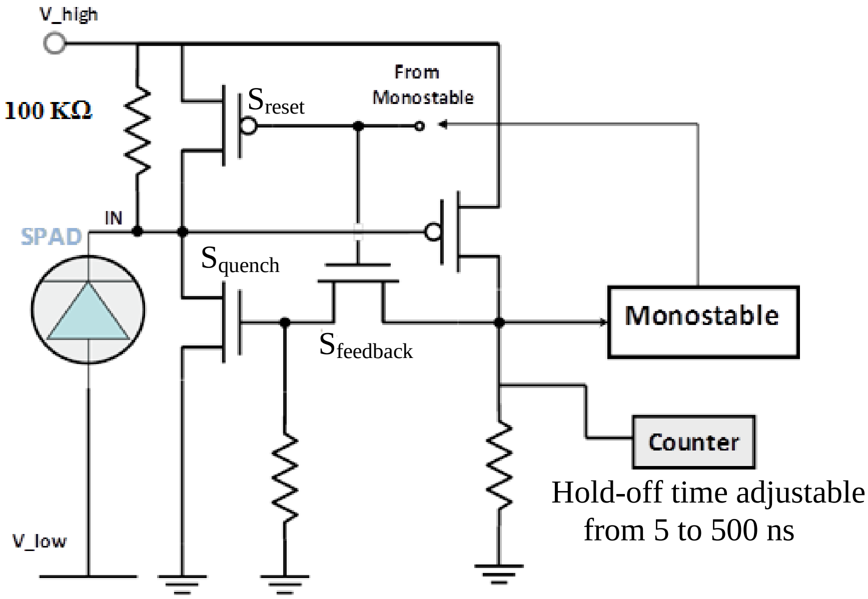


Figure 3.4: The input stage of the Active Q/R (Control) circuit. $S_{feedback}$ ensure temporal exclusivity between the charging (through S_{reset}) and quenching (through S_{quench}) operations. The reported hold-off time is adjustable from 5 to 500ns.

A monostable circuitry is characterized by a single stable operating state. Any Change, triggered by a stimulus, from the default state to an unstable state is reversed after a fixed transitory period of time, restoring the default state of the circuit. The standard monostable schematics used in [73] appears in Fig. 3.5.

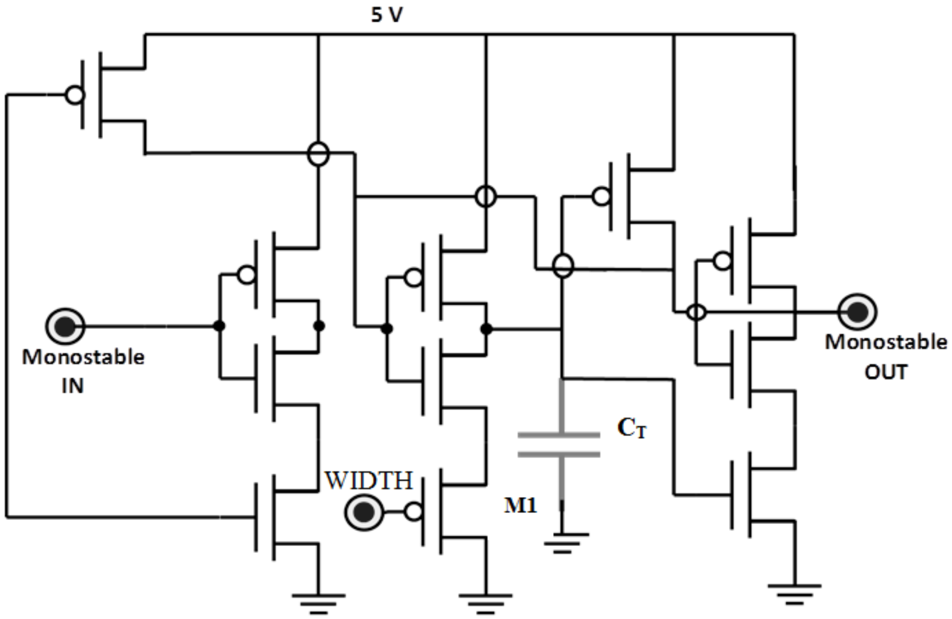


Figure 3.5: circuit diagram of a basic hold-off time generating monostable circuit. This monostable circuit is able to set a delay time of 500ns.

The signal transmitted from the avalanche sense component of the AQC circuit (Fig. 3.5) triggers the monostable. After a time interval determined by $t_{\text{hold-off}}$ the transition in the input signal reaches the output terminal of monostable, initiating a deactivation of the quenching signal and activation of the recharge transistors simultaneously. At the same the signal at the input of the monostable is turned off to restore the default state. A performance-oriented circuit design must ensure that the monostable transition time is as short as possible but long enough to ensure full recovery of the SPAD before monostable default state is restored and the recharge signal is shut off. The waiting time, $t_{hold\text{-}off}$ is determined by the RC time constant associated with capacitor C_T and PMOS transistor M1 shown in Fig. 3.5. The presence of PMOS transistor with a pinned out gate terminal ensures programmability of $t_{hold\text{-}off}$ which is reported as adjustable from 10ns to 500ns [73]. The performance trade-off associated with the value of hold-off time involves the SNR and dynamic range attributes of the SPAD pixel. The hold-off time is usually determined in accordance to the application requirements.

The AQC circuit in [71] has been improved upon in [74] by increasing the number of transistors involved in quenching and recharging of the SPAD. Here the authors report a quenching time of 2 ns and recharge time of 7ns in simulation. The performance improvement has been achieved by using a parallel network of two large transistors with an aspect ratio of 50 um/0.6 um to increase the drive strength of the quench and recharge signals. The associated monostable component involves the same design as shown in Fig. 3.5, with an addition of an extra inverter after the capacitor C_T. This is added to ensure a TTL pulse that is not distorted by the slew rate of the charge/discharge operation, as claimed to be the case in **[**71**].** By selecting a bigger capacitor the hold-off time range also has been extended from 4ns to 4.6us corresponding to gate voltage of 0.635V and 2.5V at the WIDTH terminal. The reported performance gains in the AQC reported in **[**74**]** come at the cost of fill factor degradation.

AQC design reported in **[**75**]** also operates as a monostable device, but it uses the same path for quench and recharge activation. The schematic diagram of circuit is reproduced in Fig. 3.6. The avalanche-induced voltage drop at the terminal of the SPAD device sets up a digital pulse at the output of a connecting comparator. The TTL pulse activates the quench and/or the recharge signal via a single path. The Comparator is preset to stable state of Low (0 V) output by a potentiometer network connected to its positive input, R_p in Fig. 3.6. The Avalanche signal and the ensuing cathode voltage drop (point C in Fig. 3.6) disrupts the balance of the comparator's input signal, setting its output to go high (5 V) . The high voltage at the output, through a feedback loop, reacts back onto the cathode of SPAD pulling its voltage down to below the breakdown. This action then turns of a diode connected between the SPAD cathode and the comparator input, in effect isolating the node and locking in the signal condition at the input of the comparator that had triggered the quench signal, thus holding the quench state. Eventually after a fixed time period set by the parameters of I3, a secondary path is opened to restore the signal balance at the input of the comparator and change the TTL signal at the output to LOW (0 V) which in turn will release the quench-lock on the SPAD and restore its quiescence bias setting.

An interesting feature of this active control circuit is the role of the DC un-coupling capacitor $\mathbf{C_g}$. The transient signal (AC signal) transmitted through $C_{\mathbf{g}}$ effectively provides a negative quenching voltage without the need for a negative power supply. The AC signal is generated when the Inverter I2 output signal, A, drops in response to the comparator output going HIGH (5V). Furthermore with the Capacitor C_g effectively isolating the Q/R signal drivers (I1 and I2) from the voltage at the cathode of the SPAD, the circuit can be operated with any SPAD regardless of biasing level. Minimum dead time attainable by the circuit was reported as (≈ 40ns). Which means the dynamic range is limited on the high end to 25 MHz Despite the good performance yield and interesting features and compatibility with low voltage thin Junction SPAD, the numerous

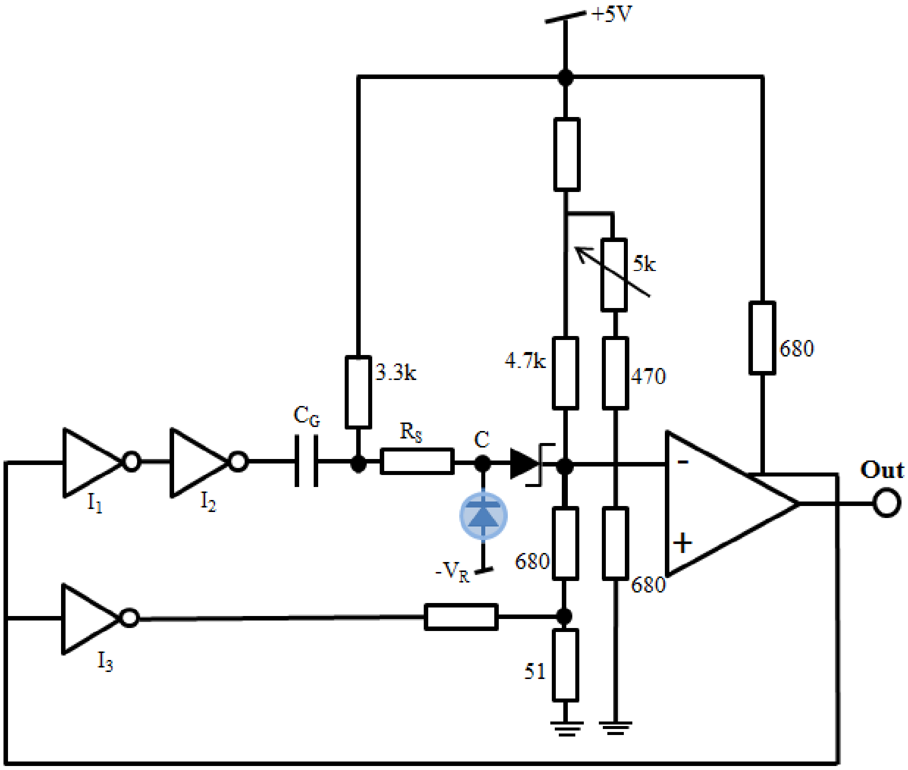


Fig. 3.6: The schematic diagram for externally implemented Active quenching unit.

components and devices such as comparator and potentiometer discourage on-chip integration, (although potentiometer can be monolithically integrated as a Transistor with the gate terminal pinned out). The substantial foot-print of the entire unit (pixel) result in considerable degradation of the fill factor rendering in-pixel integration unfeasible.

A recently reported AQC design, reported in [30] demonstrates very short quenching time using a variable load quenching to starve the avalanche current. Hold-off range 40ns to 2us is reported. This circuit also limits the High end of the dynamic range to 25 MHz The corresponding circuit implementation of the SPAD pixel front-end system is illustrated in Fig. 3.7 (SPAD cell + AQC). An NMOS is used at the anode to quench and recharge the pixel. As avalanche ignites and current increases, the operation of M1 transitions from triode into saturation. This increases the resistance in the path of the avalanche current and establishes the initial quenching drive. The active quench signal, triggered at the output of Inverter I2 eventually cuts off the NMOS transistor, voltage rises at the Anode and SPAD quenches completely.

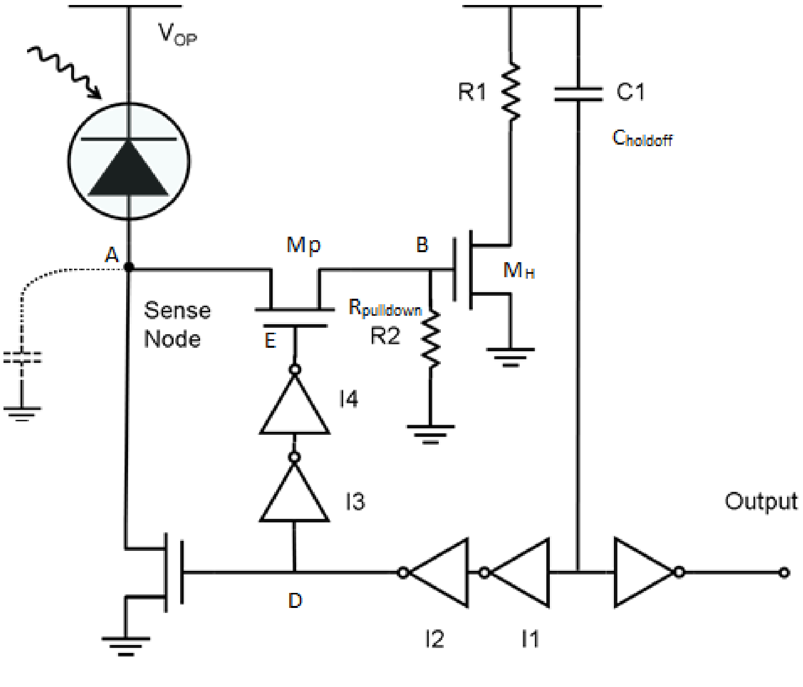


Figure 3.7: Schematic drawing of variable load AQC. Transistor Ms provides the initial quenching drive as it shifts from triode to saturation. Full quenching occurs when avalanche current is cut off as I2 goes low. Recovery occurs through discharge of Anode capacitance C_A when I2 goes high. When using a PMOS transistor in place of R1 a delay range of 4 ns to 2 µs is reported

The Precise timing required for synchronizing the operation of various signals in this setup, relies on careful transistor sizing which is a factor determined by SPAD technology in use. Also, presence of high –valued passive elements becomes a hindrance when device footprint is a priority. Furthermore, presence of transistor M1 at the Anode although advantageous in that it enables using a single transistor for both quenching and recharging rather than one transistor for quenching and one for recharging, also forces a SPAD biasing restriction. The only way to appropriately bias the SPAD above breakdown is to apply an appropriately large positive voltage to the cathode terminal of the. This causes a large reverse voltage to be dropped across the N-well/P-sub junction which, in standard plain CMOS process, could create a secondary avalanche source. An intersting capability of this design is the eliminatin of mid-recharge avalanching effect. This is accomplished by the introduction of two additional carefully sized inverters, I3 and I4. This comes at the cost of reducing the fill-factor. In situations, where the intensity of incident light is not high enough to satureate the device, which is the case in most ultra-sensitivie detection scenarios, the fill-factor trade off might not be as justified. The intricate timing to accomplish the full hold-off time is set by $R_{Pulldown}C_{holdoff}$.

After expiration of the hold-off time when node D goes high ,the recovery phase. During this phase the SPAD must continue to be kept isolated from the M_H otherwise if Mp is enabled too soon after node D goes high then the voltage at A is sensed by M_H prior to it reaching ground pulling up the voltage at node A resulting in spurious triggering and false detection events.

The use NMOS/PMOS transistors as Q/R signal drivers supplemeted with a monostable subsystem to implement synchronization and timing control is the most popular scheme in AQC design. However, different timing control methodologies have been reported. In [32] an alternate design for implementing an adjustable hold-off time is introduced that uses a digital circuit block based upon a delay line architecture rather than a monostable. The design yeilds excellent quenching time improvements, however this is accomplished through the use of a comparator designed in BICMOS technology. Its shortcoming are the cost associated with dedicated process and large complex control circuitry which renders the design unsuitable for implementation in compact arrays. Monostable seems to be a more practical venue for timing control implementation. The drawback with monostable usage, is the high limit imposed on the dynamic range due to the defualt hold-off time imposed on the detection cycle time of the circuit.

Digitally generated delay pulses using delay line, flip flops and digital gates produce accurate and linearly adjustable delay values. However they contain a large number of transistors and take up substantial amount of the pixel real estate. Another popular mechanism for generating delay in SPAD front-end circuitry is current-starved inverter chains [76]. This method of generating delay lacks precision and operational linearity in addition to high power dissipation, due to inverters operaiting in the linear rather then switching mode.

3.3 Proposed Front-end System

In order to accurately simulate the functionality and the operation of the active quench/recharge circuit a device model that can accurately simulate the behavior of the SPAD is required. This is necessary to enable analysis and evaluation of key design parameters in order to optimize the performance. For this purpose we developed a SPAD simulation model using physical parameters that closely represent the intrinsic operating conditions and specifications of a SPAD operating continuously in Geiger mode. Detailed information regarding design and verification of the SPAD model is provided in Appendix A.

The SPAD device model was used to aid in the design of front-end processing circuitry capable of implementing fast quench and reset cycles, separated by a high resolution adjustable hold-off time. The schematic diagram for the proposed processing circuit appears in Fig. 3.8.

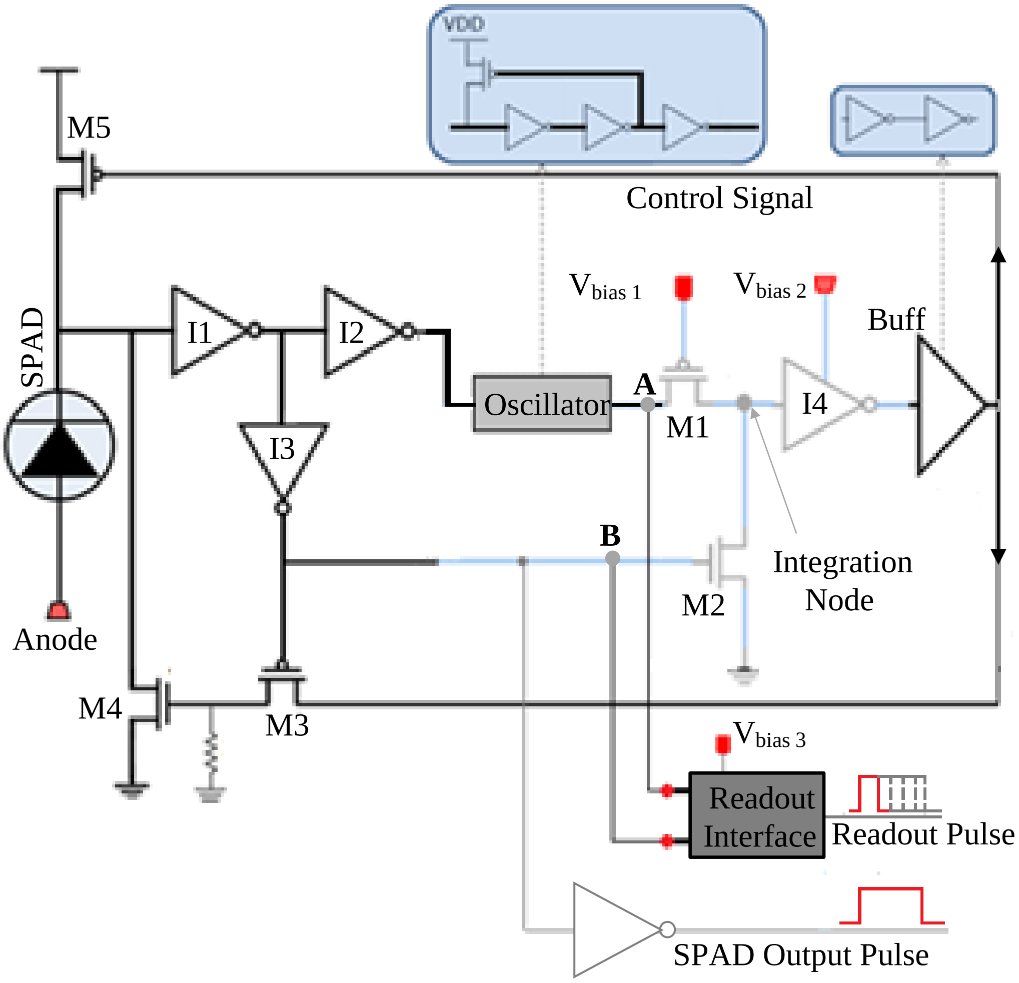


Figure 3.8: Schematic of SPAD front-end circuit implementing avalanche control, delay generation, and the readout interface.

The oscillator generates a periodic pulse train used to charge a capacitive node (integration node) following the detection of an avalanche event. The hold-off time can be controlled by regulating this charging process. Completion of the charging process signifies the end of the hold-off time, at which point the circuit automatically resets the SPAD. A comparator realized by the inverter I1 is used to sense the voltage drop at the SPAD cathode terminal signifying the onset of avalanche breakdown. Transistors M4 and M5 are used for quenching and resetting the SPAD respectively. The integration node and components I4 and M2 act together as an analog counter for the periodic pulse train generated by the ring oscillator. The comparison threshold is set by the switching threshold of I4, controlled by V_{bias2}. The initial reset is provided by transistor M2.

Initially, when there is no avalanche current both NMOS and PMOS (M4 & M5) are turned off, the oscillator is inactive and the integration node is reset to zero. At this point the voltage at the SPAD cathode terminal is charged to VDD and the SPAD device is armed. In this state, the gate of the PMOS M5 is driven high creating a high impedance path between the cathode terminal and VDD. The onset of the avalanche current causes the cathode voltage to quickly drop, turning switch M3 on, which in turn activates M4. This initiates the active quenching process by lowering the effective series resistance of the SPAD device hence speeding up the quenching process. Meanwhile the ring oscillator is activated by the falling signal transition at the cathode, and the voltage on the integration node starts to rise in discrete steps. When the voltage at the integration node exceeds the switching threshold of I4, set by the external input V_{bias2}, the output of I4 will transition low and turn off the control signal. This will shut down M4 and terminate the quenching, which marks the end of the hold-off period. At the same time, the PMOS transistor M5 is turned on and the voltage at the cathode terminal is restored to VDD. The rising cathode voltage causes the output of I1 to drop, deactivating the oscillator signal and resetting the integration node to zero. As the control signal is pulled high the PMOS transistor M5 is turned off, and the reset cycle is completed. The SPAD device is now set for the next detection event.

The holdoff time is implemented using discrete summation of small charge packets onto a capacitive node. A wide range of variable hold-off times can be generated by coarse and fine range adjustments through V_{bias1} and V_{bias2}, respectively. V_{bias1} controls the charging duty cycle through transistor M1, which determines the slew rate at the integration node. V_{bias2} determines the signal amplitude required to trigger the termination of quench and initiation of reset operation. The post-layout simulation of the integration signal obtained from the extracted netlist is shown in Fig. 3 for various V_{bias1} and V_{bias2} values. Fig. 4 shows experimental measurements of the SPAD output pulse width (representing a complete quench/hold-off/ reset activity cycle) for a set of bias values with $V_{bias1} = 2.9 – 3.5$ V and $V_{bias2} = 3.0 – 5.0$ V. The results shown in Fig. 3.9 and Fig. 3.10 demonstrate the capability of the proposed circuit in generating a wide range of hold-off times with linear resolution and adjustable granularity.

3.4 System Interface - *Event pulse Generator*

Response time of the SPAD device is described by the rise and fall time of the detector output, which is in turn determined by the RC time constant at the actively driven terminal of the SPAD plus any additional hold off time implemented for after-pulsing. The waveform representing the response maybe additionally buffered to obtain a

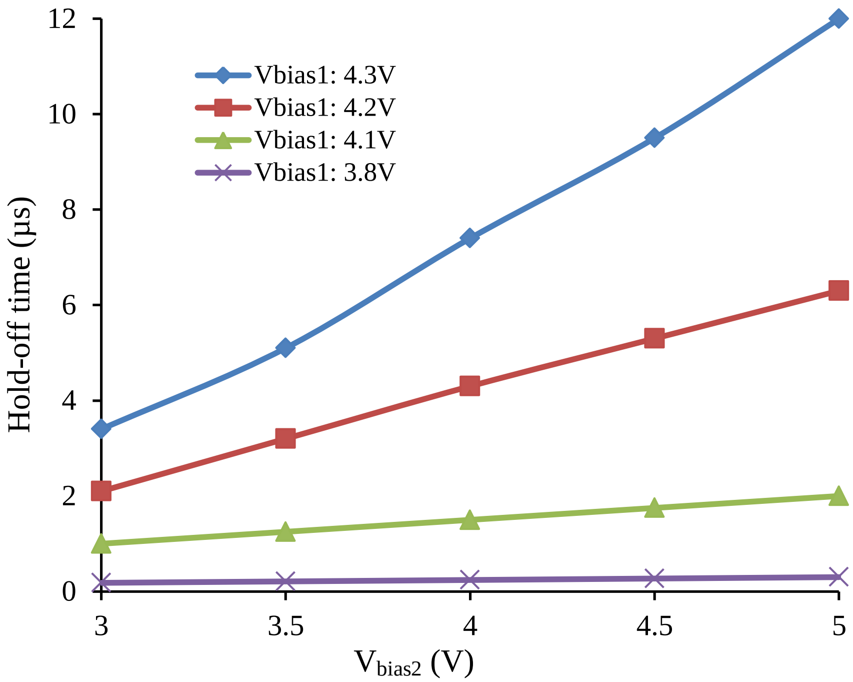


Fig. 1.9: Post-layout simulated waveform representing the transient voltage at the integration node for different slew rate and amplitude threshold levels determined by Vbias1 and Vbias2 values respectively.

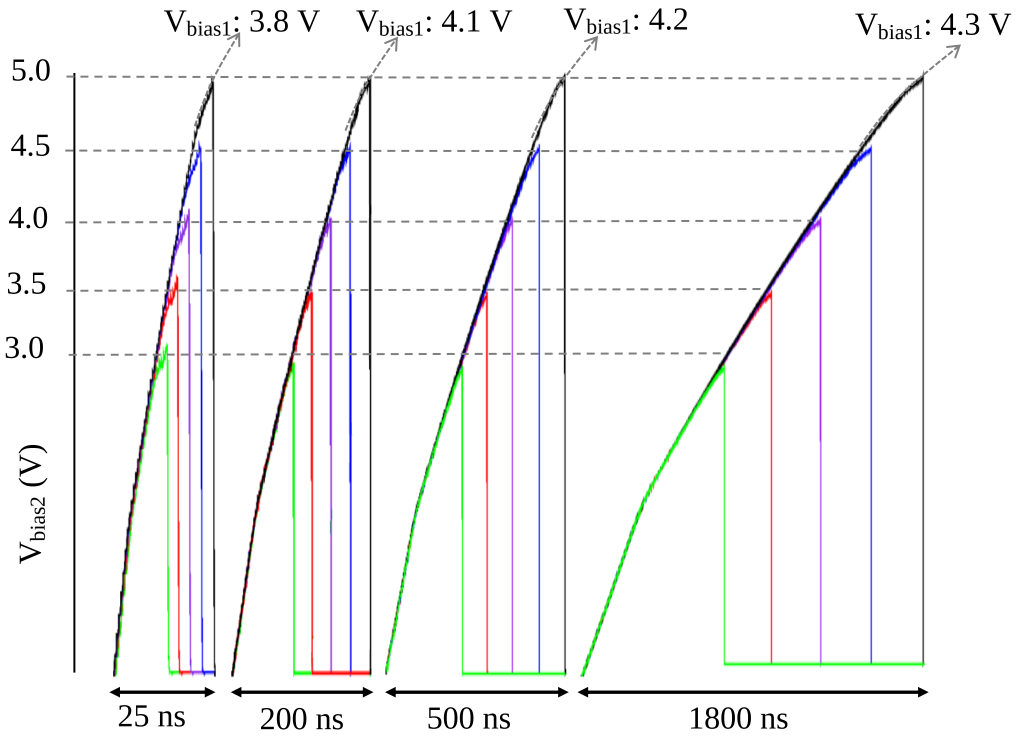


Fig. 3.10: The measured hold-off times for a set of Vbias1 and Vbias2 Values

TTL pulse which represent one detection cycle of the device. For a single SPAD device or array of devices where each device is individually read out counting these pulses provides the number of events detected. However for multi-pixel systems where by interconnects or processing module, such as counters, are shared among several devices, arbitration and timing issues associated with shared access scheme will complicate the design and reduce the readout bandwidth. In such cases, device operating parameters as the detection pulse width (dead-time) determine the bandwidth of the system as the shared medium is engaged each time for the duration of the readout pulse. For example in [77] detector pulse is used as clocking signal for an array of latches used to record the address of each device. Clock pulse width must be attuned to the timing constraints of the readout devices in order to avoid receiving a signal during the setup time of the latches. Therefore in an array implementation the hold-off time of the SPAD device becomes set in accordance to the timing requiring of the readout system rather than after-pulsing performance and device bandwidth requirements. This leads to an inflexible design. When the Device response are digitally readout over a common signaling path the readout pulse width representing a detection event time-line must be independently set and controlled in accordance to system requirements. The detection pulse must not be the same as the event pulse or the readout signal that is routed across a shared medium. In the new readout paradigm proposed in Chapter 5, the signal measurement window is determined by the readout pulse width. Therefore controlling this pulse width offers flexibility in actively configuring the detector performance to match the requirements of a target application. In order to allow for a standardized readout pulse for flexible system implementation, it is necessary that the individual detection events are represented by an event-generator that can be independently set and tuned to generate readout pulses with widths that are configurable fraction of the device detection pulse width. In this way the device output pulse width is independent of device specific parameters and performance constraints. The basic conceptual block diagram for the complete front-end system appears in Fig. 3.11

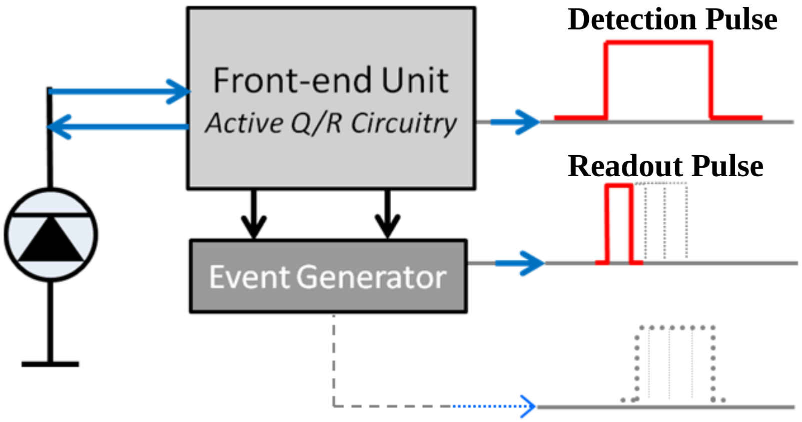


Figure 3.11: Block diagram showing the relationship between the detection and the readout pulse.

The front-end circuit in Fig. 3.8 incorporates a readout interface with an output pulse width that is an adjustable fraction of the SPAD activity cycle. The schematic for the readout interface is shown in Fig. 3.12. Hardware sharing between the hold-off generation circuitry and the readout interface, through the usage of a common oscillatory stage for delay generation, reduces the electronic footprint of the pixel and enhances the pixel fill factor. The readout pulse is initiated by the onset of the quenching action at the cathode terminal of the SPAD detector and terminated after a programmable delay controlled by V_{bias3}. The variable pulse width is implemented by the same mechanism as the holdoff time generation. At quiescent, the high voltage at the SPAD cathode terminal

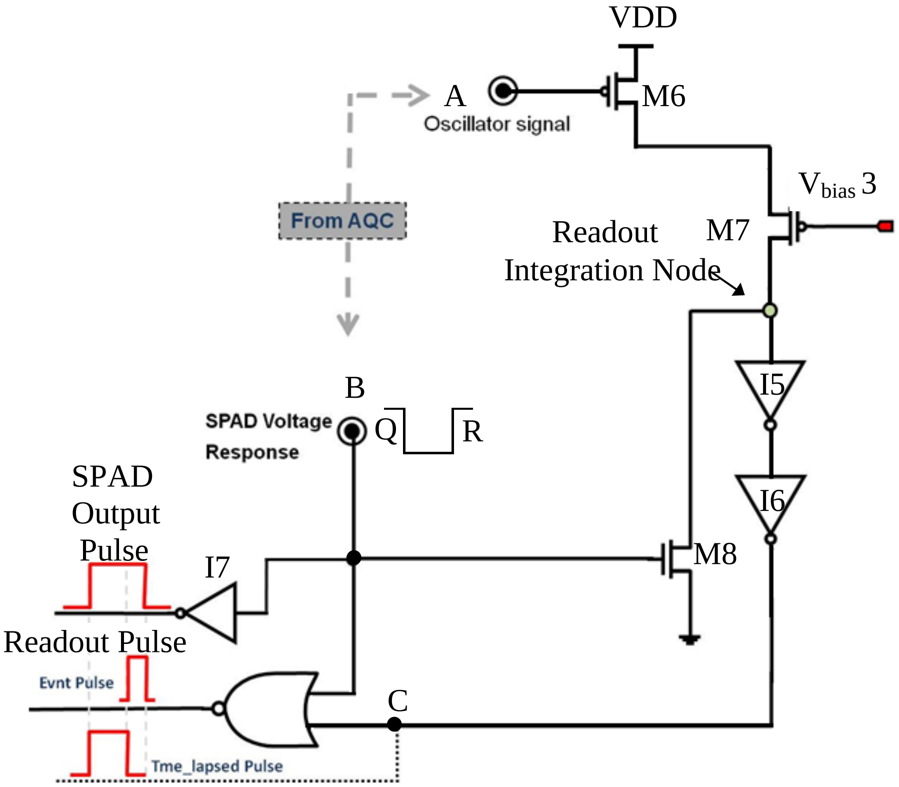


Figure 3.12: Readout interface (Event generator). The leading edge of the readout pulse is simultaneous with that of the detector pulse, its trailing edge is set by propagation delay through $V_{bias}3$ if it is set to be shorter than the detection pulse period.

causes the output of the NOR gate in Fig. 3.12 to be low. Following a detection event, the quenching action lowers the voltage on Node B, causing the output of the NOR gate to go high and triggering the rising edge of the readout pulse. The concurrent activation of the oscillatory signal at Node A and shut down of the NMOS M8 in parallel with the initiation of the readout pulse subsequently leads to the activation of Node C which turns off the output of the NOR gate and terminates the readout pulse. Resetting of the NOR gate output occurs after a delay interval determined by the value of V_{bias3}. If the value of V_{bias3} is set too high, the trailing edge of the readout pulse will coincide with the reset operation of the SPAD and the width of the readout pulse will correspond to the SPAD hold-off time. However the value of V_{bias3} can be set such that the input node C of the NOR gate is activated prior to the onset of the SPAD reset operation, resulting in a readout pulse that is a fraction of the pulse width associated with the SPAD detection pulse. Experimental measurements of the readout pulse at various V_{bias3} values are shown in Fig. 3.13.

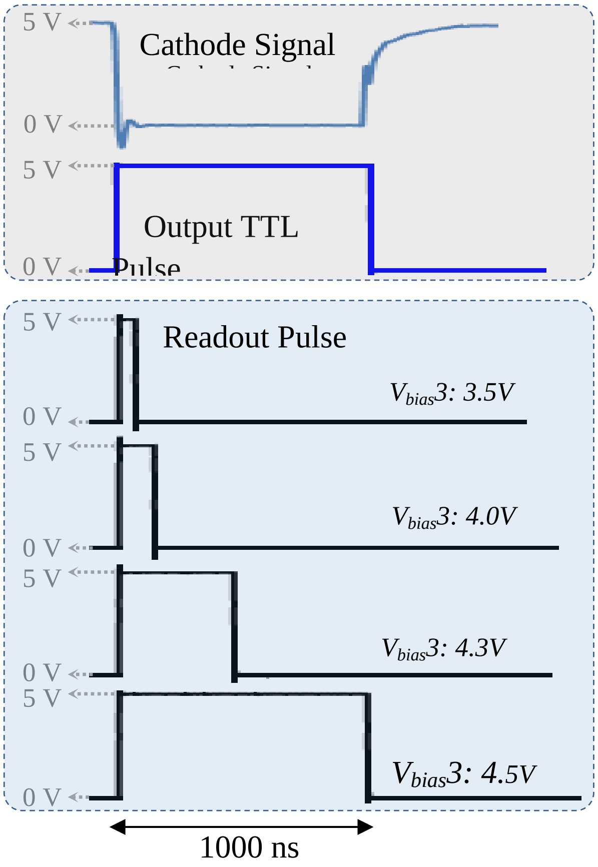
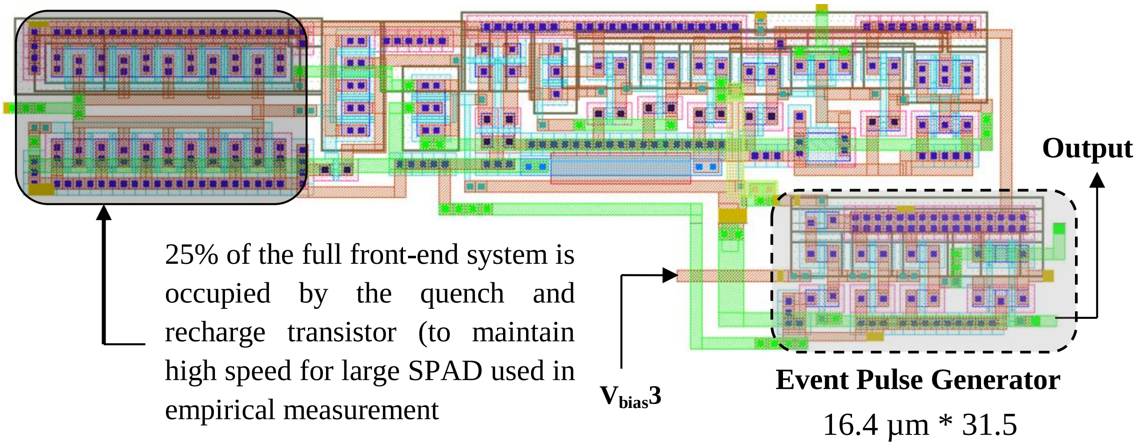


Fig. 3.13: Experimental measurement of the readout pulse with different pulse width parameters recorded at different Vbias3 values without altering the activity pulse duration.

The layout diagram of the complete front-end system appears in Fig. 3.14. In 0.5µm CMOS technology with λ = 0.35µm, the layout occupies 16.4 µm x 31.5 µm + 121.8 µm x 22.2 µm for a total area of 3222 μm^2. The proposed readout interface enables the sensor readout to operate at a reduced duty cycle relative to the operational duty cycle of the SPAD detector, even at high activity rates. This is in contrast with previous approaches in which the SPAD activity cycle directly controls the duty cycle of the sensor output signal.



25% of the full front-end system is occupied by the quench and recharge transistor (to maintain high speed for large SPAD used in empirical measurement

Figure 3.14: Layout for the front-end system integrated with each SPAD detector.

3.5 Summary

This chapter discussed the design of a fast compact front-end system for high bandwidth operation. The precisely synchronized quench and recovery cycles are temporally separated by a linearly adjustable hold-off time. A new method for linear hold-off time generation was introduced and empirically verified. Additionally a novel structure for readout pulse generation was introduced for the first time as part of the in the front-end electronics. This structure represents the pixel interface for connecting to a pixel array and is required to maintain operational transparency in both directions. It allows the SPAD readout pulse width to be configured as an adjustable fraction of its output pulse width which represents a full detection cycle. The integration of an independently adjustable pixel output interface allows for more robust and flexible system implementation that will be discussed in Chapter 5 and 6.

Chapter 4

Analytical Performance Characterization

Photon number measurements taken repeatedly from the same emission source under identical operational parameters will likely yield a different value each time. This is due to the statistical nature of optical emission [78]. In such cases a single measurement provides an instantaneous value of the desired optical parameter at the instance of measurement. This is of little value in characterization and identification of the underlying excitation phenomena or optical property. Therefore in order to obtain a meaningful statistical distribution for the desired optical characteristic, the measurements process must be repeated multiple times. The mean value obtained from the measured data can be used to characterize the desired optical property [79].

Due to random variations in statistically distributed quantities there is a possibility that the instantaneous energy signature of an incident signal, at any time, can exceed the performance limits of the detector and produce a nonlinear response [79]. Therefore the measured signal statistics cannot be considered as an accurate representation of the incident photon statistics without some accounting for the conditions of non-linearity in the detector response. In statistical measurements, the measured statistics (distribution of multiple measurements) are usually related to the input signal statistics with a transfer function that is determined using a stochastic model [80-83].

4.1 Ultra-low intensity Optical Sensing – *Dual Detection Paradigm*

An ultrasensitive optical detection platform is essentially characterized by its linear output response to photon-level variations in the incident signal at its input. The sensitivity performances are usually determined by the photo-response characteristics of the physical device and the noise floor associated with detector system. The full range of input signal levels over which the detector can maintain its sensitivity (linear response to photon-level changes at its input) defines the linear operation range or the dynamic range of the detector [84, 85].

4.1.1 Multiple Single Shot Measurements – *Photon number mode*

The photon number distribution represents the spectrum of discrete energy levels in the electromagnetic field radiation and is associated with different quantum states. The radiation field containing one, two or three photons are characterized by energies corresponding to the first, second or third excitation. The discreteness of the allowed energies (quantum states) has been demonstrated through measurement of photon statistics [86]. The advent of photon number resolving detectors with single photon resolution (one energy quanta) has made such experimental break through possible [87, 88]. Determining the photon statistic is critical in characterizing the emission source. Single shot measurement of excitation level through photon number detection is an indispensable feature for any applications associated with emission of a discrete energy optical pulse [89-91]. Application dealing with fluorescent detection and functional study of molecular dynamics, tumor diagnosis as well as scintillation detection and Astronomical observations, involve detection of low intensity excitation pulses [92-97].

4.1.1.1 Detection System – *Silicon PhotoMultiplier (SiPM)*

The binary ON/OFF response of SPAD, to stimuli as low as a single photon, makes it the ultimate optical switch. However the digital profile of its response does not allow for photon number resolution since the same ON response is triggered by single photon and multi-photon pulse events. The SiPM technology is primarily characterized by parallel readout of avalanche current from simultaneously firing SPAD microcells. This aggregated signal is read out at the output as an analog indicator of the incident photon number. SiPM is an analog detector composed of digital elements [99].

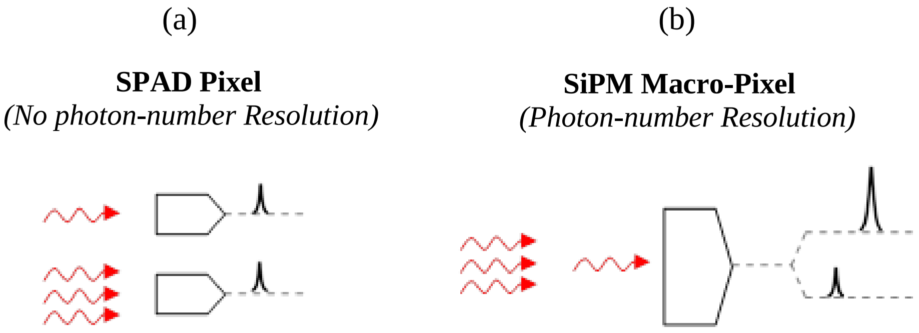


Figure 4.1: (a) Multi-photon response of single-element SPAD detector is not differentiable from its single photon response. (b) SiPM macro-pixel has linearly proportional photo-response cable of discriminating between single photon and multi photon events.

4.1.1.2 SiPM Non-linearity characterization

In single shot optical measurements, the amplitude of the SiPM output pulse provides a linear measure for the number of photons impinging upon the active area of the detector, assuming every photon in the incident pulse triggered a microcell. This assumption breaks down at higher intensity pulses as the finite granularity of the detection surface leads to the saturation of its response and loss of linearity between the its output amplitude and the number of incident photons [80, 81, 100].

According to theoretical modeling and empirical measurements reported in the literature this occurs when the number of photons is approximately 60% of the total cell count of the detector. At this point, due to the statistical distribution of photon numbers in a radiation pulse, the probability of multiple photons overlapping on a single cell (Multi-photon effect) increases enough to result in observable non-linearity in the output [101, 102, 103].

Photon number statistics in an optical pulse are characterized by a Poisson probability distribution. The purpose of single-shot measurement of the optical pulse is to retrieve the photon number statistic as accurately as possible. As mentioned earlier, reconstructing the photon number distribution involves taking multiple measurements, making reliable determination of the absolute photon number each time. In light of the detector’s response non-linearity for larger photon numbers, this requires either a calibrated set up (which sometimes requires a complicated quantitative experimental procedure) or a qualitative model for the SiPM output response. The relationship between the measured count relative to the incident photon number is illustrated in the below expression.

$$\mathrm{P}\ (\mathrm{k}|\mathrm{n}) = \sum_n \binom{n}{k} p^k q^{n-k} \times \boldsymbol{\varrho}_n \tag{4.1}$$

The first binomial term represents the conditional probability that ‘k’ photons are registered when ‘n’ photons were incident on the detector. The second term $\boldsymbol{\varrho}_n$ is the probability that ‘n’ photon were incident on the detector. The first order approximation of the expression in terms of detector and incident signal parameters appears in equation 4.2.

$$N_{\text{fired}} = N_{\text{tot}} * \left[1 - \exp\left(-\frac{N\gamma * \boldsymbol{\varepsilon}_{\mathbf{PDE}}}{\text{Ntot}}\right)\right] \qquad 4.2$$

This expression describes the mean number of fired pixels (N_{fired}) in terms of the total pixel count (N_{tot}) and the impinging number of photons (Nγ). The analytical expression for the SiPM response shown in equation 4.2 describes a first order probability model approximates the non-linearity condition stemming from finite pixel count on the performances of the array. In other words, it presents a first-order mapping between the triggered SPAD count and the incident photon count [104].

4.1.2 Continuous Measurement – *Photon Counting Mode*

Photon counting provides information on the activity of the energy source in terms of its emission intensity. In contrast to SiPM single shot measurement to determine the energy signature of a high energy event, i.e. scintillation photon, photon counting involves detection of multiple ultra-low intensity events occurring over a finite period of time. This mode of detection in the ultra-sensitive regime is characterized by continuous measurement either for the duration of the emission process or until a representative sample to perform reliable measurement has been collected. Acquisition of ultra-weak optical signals involved in applications such as low-light imaging of biological system, early detection of disease signature, molecular imaging, LiDAR 3D imaging and ranging and astronomical observations, requires true photon counting capability [25, 105]. Ultra-high speed imaging and parallel processing of photon-based data stream are also aided by this technology. In a shifting of paradigm from the SiPM and single shot measurement of photon number, in photon counting modality the detector must be fast enough to temporally resolve between individual photons at high count rate. The performance can be characterized in terms of photon counting efficiency with the non-linearity factor being the recovery time of the SPAD pixel.

4.1.2.1 **Detection System –** *Digital SPAD Array*

Active quenching front-end electronics can improve the operation speed of the SPAD by three orders of magnitude and are essential pixel components in the photon counting mode. Another useful attribute of active quenching electronics is its ability to introduce a controllable delay between the quench and reset operation which significantly reduces the after-pulsing noise of the SPAD. However, prolonging the pixel recovery cycle (Dead Time) will limit the dynamic range of the device [106, 107].

Photon arrival rate is a statistical parameter characterized by a Poisson distribution (as was photon number distribution). Therefore, in light of the dead time associated with the operation of the detector, its count rate measurement must be supported with an accurate theoretical description of its non-linearity condition before the measured data can be reliably interpreted.

4.1.2.2 Non-linearity characterization

The standard model used to correct for count loss due to non-extending detector dead time and reconstruct the True count from the measured count is shown in equation 4.3.

$$\text{True Rate} = \frac{\text{Measureed Rate}}{1-(\text{Measured Rate}*\text{DT})} \qquad 4.3$$

Since one dead time interval is associated with every detection event, the total OFF time of the detector, taken as sum of all the individual dead time intervals during the measurement window, is represented by the product of Measured Rate and dead time interval. For a unit measurement window of one second then total ON time can be represented by the denominator in equation 4.3.

Considering that all the detected events must have occurred during the ON time of the detector, the true count rate is taken as the Measured Rate divided by the Total ON time . This is the standard model used for count correction in presence of dead time in all general applications involving non-paralyzable detection of discrete events with random arrival rates. It has been used repeatedly in SPAD related literature both for device quantum efficiency characterization to compensate for band width related count loss and also to correct for the hold-off time in photon counting applications [45, 108-112].

The relation shown in equation 4.3, however, does not provide a complete theoretical description of the underlying statistical process. It does not take into account the statistics of the dropped events based on the assumption that the last detected event was the last incident event. Consequently it excludes the higher order effects that become critically important for larger arrival rates and dead Time intervals, whereby the stated assumption no longer holds true.

4.1.3 Proposed Analytical Model – *Theoretical Dead Time correction*

The proposed stochastic detection model is based on calculating the effective probability of detection for the nth photon by exploring all possibilities regarding the last detected event. The statistics of Photon arrival rate (pulses from a coherent source) follow a Poisson distribution. Hence the photon *inter-arrival time* has an exponential probability distribution. Photon Inter-arrival time is a random variable and its probability density function is shown in equation 4.4, with λ representing the mean arrival rate.

$$\mathrm{P(X = x) = \ \lambda.\, e^{-\lambda.x}} \qquad 4.4$$

$$\mathrm{P(X >} DT\mathrm{) = e^{-\lambda.DT}} \qquad 4.5$$

Equation 4.4 describes the probability that an inter-arrival time ‘X’ is greater than the dead time DT, given that the average event arrival rate is $\boldsymbol{\lambda}$. In other words it represents the detection probability of photon ‘n’ given that photon ‘n – 1’ had triggered a dead time (last detected event). If the assumption is that the last detected event is separated from the current event by one inter-arrival time (the last detected event assumed to be the last incident event), then the probability of detection for photon ‘n’ is characterized by Equation 4.4. However what if photon ‘n – 1’ had arrived during the dead time interval triggered by photon ‘n – 2’ and as such was never detected**.** Then the last detected event, ‘n – 2’, would be separated from photon ‘n’ by two inter-arrival times. The operator X in equation 4.3 and 4.4 now becomes 2(X) since the probability of detection for photon ‘n’ now requires that the sum of two inter-arrival times (X) to be greater than the DT. The sum of exponentially distributed variables follows an Erlang distribution. Equation 4.6 describes the probability that a time interval Y, (made up of ***k*** inter-arrival times; Y = kX), is greater than the length of the dead time. Operators ‘m’ and ‘n’ in equation 4.6 represent the event numbers delimiting the interval Y.

$$\mathrm{P(Y} > DT) = \mathrm{e}^{-\lambda.\mathrm{DT}}.\sum_{\mathrm{k=0}}^{\mathrm{m-n-1}} \frac{\lambda.\,\mathrm{DT}^{\mathrm{k}}}{\mathrm{k!}} \qquad 4.6$$

The index ‘k’ represents the number of arrivals within the dead time generated by the last detection event. In other words it represents the number of dropped events between the last detected event and the most recent detection event.

The detection probability of the nth event can only be calculated with respect to the last detected event, this is inherent in the operation of any non-Paralyzable (non-extending dead time) discrete detector. However the last detected event need not be the (n – 1)th event, it can be any one of the ‘n – 1’ previous events. Every dropped event improves the detection probability of the next event. Therefore the detected events are marked by a higher probability of detection depending of the mean arrival rate and dead time period as shown in equation 4.6. This effect is not incorporated in standard dead time correction model shown in equation 4.2. As a result the expression in 4.2 over-predicts the true count based on the measured count because it does not factor in the dead time induced widening effect in the probability distribution of inter-arrival times of the detected counts. Instead it accounts for the measured count magnitude in light of the total detector OFF time by inflating the True count. The probability of detection for the nth photon modeled as the product of the nth photon detection probability given k dropped events and the probability of k, summed over all possible k is shown in equation 4.7

$$P(n^{th}) = \sum_{k=1}^{n-1} P(n^{th}\,|k) * P(k) \qquad 4.7$$

The choice of n has to do with the point at which detection probability reaches a steady state after the initial state fluctuation representing different detection scenario for each new photon. The state of the detector very quickly reaches a steady state for the nth photon. The plot of detection probability over photon number from 1st to xth photon, resembles a damped sinusoid with damping factor of$(-\lambda.\mathrm{DT})$. The probability value settles to a constant value at the nth photon. The New expression for probability of detection as a function of detector non-linear parameter and signal parameter (Dead time and arrival rate) is shown in equation 4.8. Parameters J and ω represent the enumeration

$$\sum_{\omega=1}^{n-1} \sum_{j=0}^{\omega-1} \sum_{k=0}^{j} e^{-\lambda\tau d} \frac{\lambda\tau d^{k}}{k!} * P_{n-2-j} * \prod_{c=n-1-j}^{c+n-2} q_c \qquad 4.8$$

index, while pixel dead time is represented as $\boldsymbol{\tau}$ and incident arrival rate denoted by λ. The expression for quantities P and q appears below.

$$P = \left(e^{-\lambda.\tau} \right) \qquad 4.9$$

$$q = 1 - \left(e^{-\lambda.\tau} \right) \qquad 4.10$$

As a test of its prediction accuracy the model was used to reconstruct the true count representing the DCR of a SPAD device measured at three different rates (set by application of appropriate field gate bias) with and without a 10.58 µs dead time injected into the measurement. The reconstructed true count was compared against the actual true count obtained through a dead time free measurement scheme. A quantitative comparison between the proposed model represented by equation 4.8 and the standard correction model of equation 4.2 in terms of prediction error was carried out using the method described above. The measured results appear in Fig 4.2.

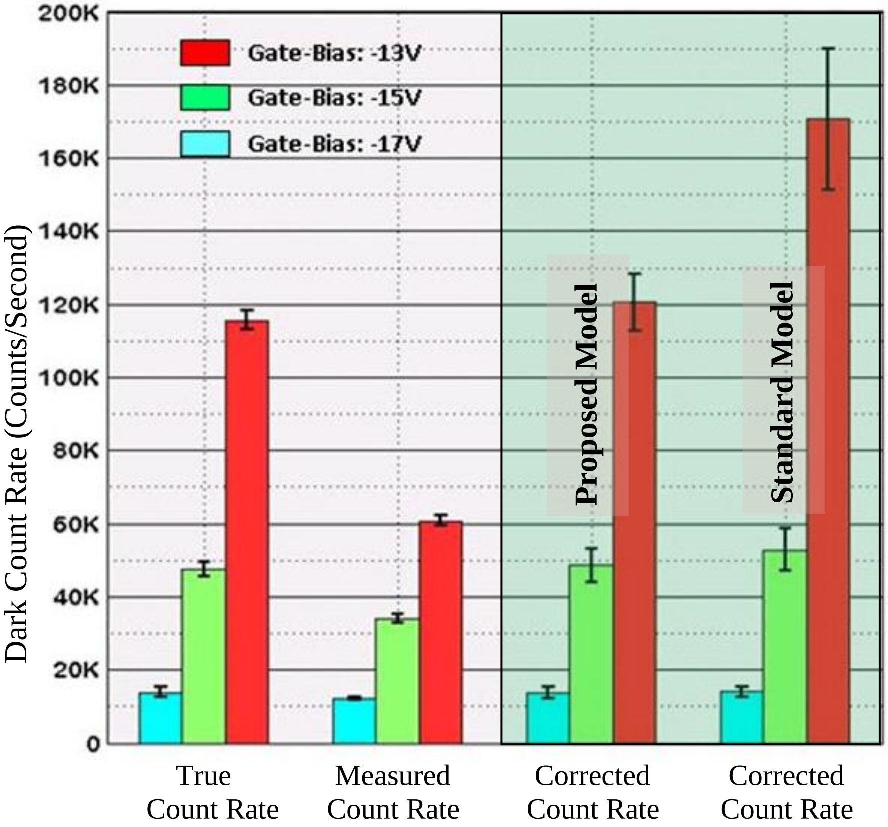


Figure 4.2: Performance comparison between the proposed and the standard model.

Chapter 5

CMOS Photon-Counting Detector Array

Main challenge in development of large integrated SPAD arrays is the readout implementation. From a detection perspective, the operation of both SiPM and SPAD sensor array is identically characterized by photo/dark triggering of the avalanche phenomena in individual SPAD elements. From an implementation perspective, however, the field has split into two distinct research and development path. The divergence is purely centered upon the analog or digital approach towards the readout of the array.

5.1 Digital Readout – SPAD Image Sensor Array

Due to the dynamic nature of SPAD response the conventional photo-detector readout scheme involving local generation and storage of photo-induced charge is inapplicable [2, 113-115]. Every detected photon is represented by a distinct event pulse that must be converted, upon arrival, to a count and/or a Time Of Arrival (TOA) value before it can be stored. Fig. 5.1 shows the conceptual nature of the problem. The processing and storage can be done in-pixel or the pulses can be read outside the array for the processing to be performed externally. In either case, all operations to include time extraction and/or count determination must be performed dynamically upon photon arrival. The main trade-off is therefore at the architectural level [115].

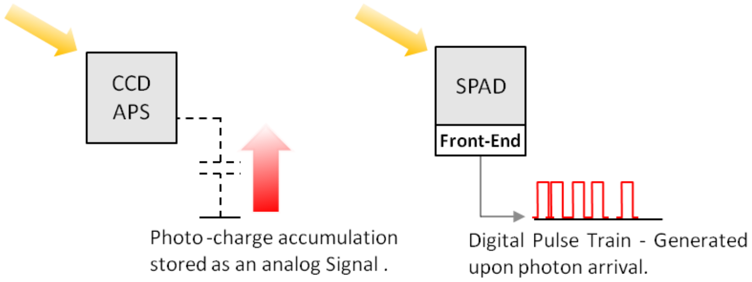

CCD
APS
SPAD
Front-End
Photo-charge accumulation stored as an analog Signal .
Digital Pulse Train - Generated upon photon arrival.

Figure 5.1: Conceptual difference in data readout requirement between standard semiconductor imagers and SPAD detector.

Applications requiring photon-level sensitivity or high resolution timing of incident photons can greatly benefit from SPAD-based array detectors. However the dynamic profile of the SPAD response makes array-level management of traffic from large number of pixels very difficult. This will restrict the system throughput and imposes practical limitations on the size of the array [116]. Ideally each SPAD element in an array must be treated as a photon-triggered digital element with a dedicated readout channel. This set up yield maximum detection accuracy as it places no restriction on system throughput. However, due to factors such as, higher power and wiring complexity, this configuration is only possible for very small array sizes [48].

Since different architectures represent different levels of detection accuracy and power optimization, performance becomes a direct function of implementation. In digital SPAD image sensors, readout implementation is solely characterized by the architectural level at which pulse processing of the SPAD response is performed. Architectural schemes proposed in the literature range from serial/random-access signal detection and readout through fully parallel signal acquisition using in-pixel counting and processing electronics [117, 118, 119].

5.1.1 Synchronous Readout – *Serial Access*

5.1.1.1 Sequential Processing and Readout – *Chip Level*

In this scheme, a single processing unit (digital counter or Time to Digital Converter) is shared among all the pixels in the array. Pixels are accessed one at a time for an arbitrary integration interval of time, at the end of which the processed quantity is read out and the processing unit is reset in order to process next pixel. Although this technique avoids the mechanical scanning process associated with single pixel SPAD imagers, it offers the same low detection bandwidth. Despite its simple architecture and high fill factor, this architecture suffers from very low throughput and poor detection efficiency. Since only one pixel can be readout at any time, photons incident on the rest of the array are lost [119].

5.1.1.2 Semi-Parallel Processing and Readout – *Column Parallel*

In this scheme parts of the readout circuits are shared among a set of pixels. The partial operational parallelism offered by this architecture results in greater throughput and photon detection efficiency, at the cost of reduced fill factor due to a greater number of processing elements utilized. The detection bandwidth still suffers as compared to In-pixel architecture (described in the next section) since the photons that are simultaneously incident on the same column are lost. Fig. 5.2 illustrates the architecture-based performance of fully sequential and column-parallel readout schemes [3, 77, 120, 121].

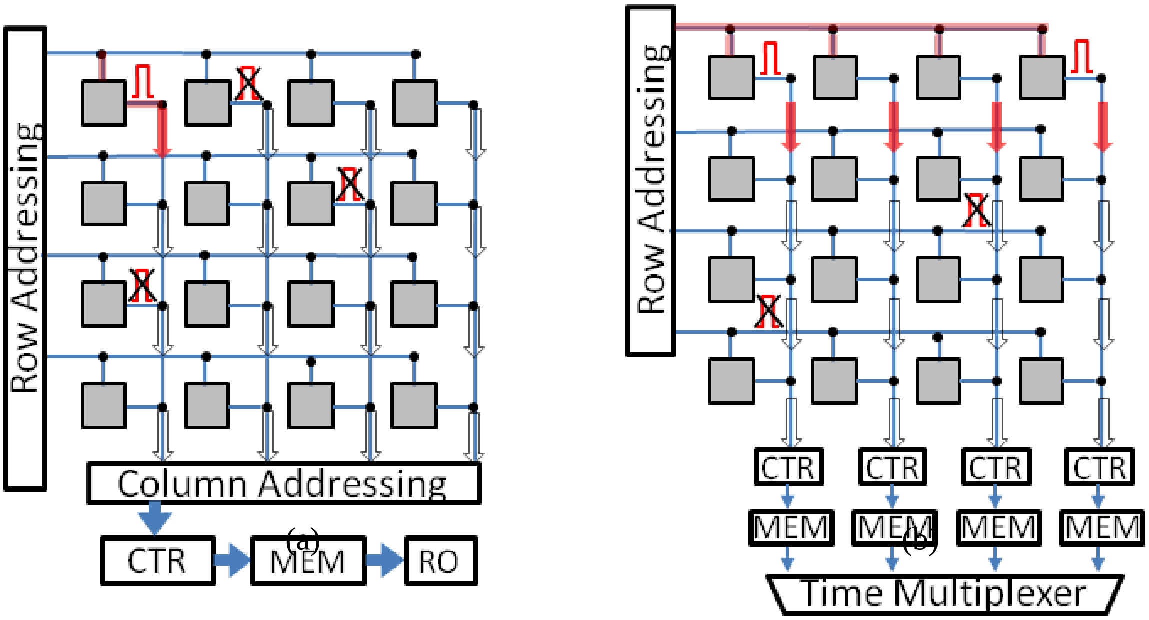


Figure 5.2: (a) Sequential architecture: offers the best power and area utilization, but the worst bandwidth and detection efficiency. (b) In-Column architecture improves the readout bandwidth but degrades the fill-factor.

5.1.1.3 Fully Parallel Processing with Serial Readout – *In Pixel*

In the fully parallel architecture all processing (counting/timing) and storage of data are performed locally (on-pixel). In this scheme every pixel has dedicated processing electronics as shown in Fig. 5.3. The stored quantities can be read out sequentially or through random access method. This architecture offers full operational parallelism and greatly improves the number of photons that can be detected and processed at the same time. No photons are lost during the detection cycle. Main limitation of this approach is that it significantly minimizes the photo-sensitive area of the sensor [118, 119, 122]. The required amount of on-chip electronics occupy a substantial portion of the silicon real estate, drastically reducing fill factor [122-124]. Existing implementations of this architecture differ according to the level of integrated functionality, ranging from a one bit counter to multi-bit counter and high precision TOA electronics.

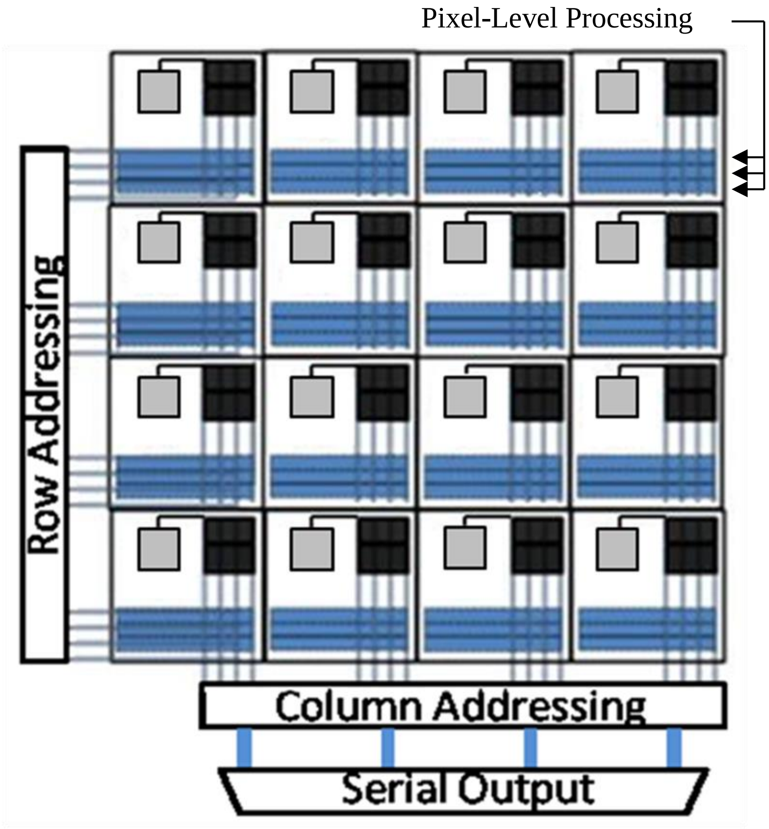


Figure 5.3: Fully parallel In-pixel processing offers the best detection bandwidth but the worst fill factor performance.

5.1.2 Asynchronous Readout – Event Driven Access

The low complexity architecture associated with synchronous readout scheme might be an adequate solution for simple applications, however the asynchronous nature of pulse arrival is best served with an on-demand or event-driven readout scheme. At the same time, the difficulty associated with in-pixel storage of information for many detected photons highlights advantages of a shared access approach. A low-complexity implementation (no access control) of such asynchronous shared-access architecture may be advantageous for low-light applications. Under this configuration an array column can be used as a digital bus that is accessed every time a photon is detected by a column element [119]. The rate of access is determined by the detection activity rate. This configuration leads to a more efficient utilization of system bandwidth, and improvement in system Signal to Noise Ratio (SNR) since brighter pixels would be favored on the transmission bus. The bus remains unavailable for the duration of each access cycle therefore multiple photons cannot be simultaneously detected on the same column. Another performance limitation associated with the low-complexity shared access protocol is the throughput bottleneck resulting from the finite bandwidth of the shared resource (counter/Timer) which would limit the bandwidth of the entire column/cluster [3, 74, 124].

In low-light applications with sufficiently low probability of simultaneous column hits, non-arbitered asynchronous (event-driven) readout scheme can improve throughput and detection sensitivity [74]. However for applications involving greater light levels some means of access control or collision avoidance such as arbitration or Time Division Multi Access (TDMA) becomes necessary. Fig. 5.4 illustrates the basic access control configurations. The desirable aspects of such protocols are sometimes offset by the resulting increase in the electronic footprint and consequently degradation of system fill factor. Implementing an arbitration scheme increases the system signaling complexity in the form of additional request and acknowledge telemetry traffic. Due to access exclusivity inherent in such asynchronous sharing architectures, system performance will not scale with the incident light intensity or the array size, thus significantly limiting the applicability of the system.

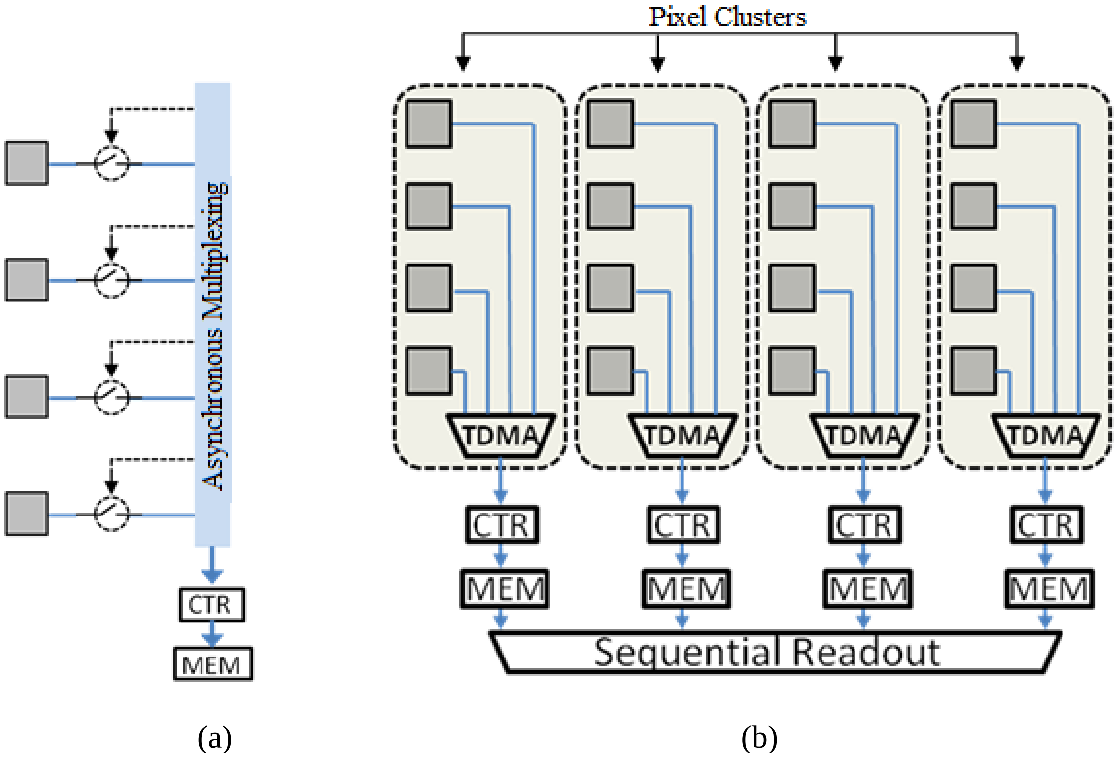


Figure 5.4: (a) Collision is avoided by verifying the state on the shared medium prior to initiating access. (b) Time division Multiplexing randomizes the transmission time window of the signals converging onto a common resource.

5.1.3 Inherent Limitations of Digital Readout

5.1.3.1 Architectural Trade-Offs – *Application Specific Design*

Large in-pixel electronics can significantly improve the throughput performance and detection efficiency of digital SPAD arrays but due to their impact on fill factor they may not always be desirable [74]. Simpler pixel-level processing will improve the pixel fill factor and enables larger arrays, at the cost of increasing system readout complexity, frame time and data loss. The resulting performance constraint limits the array size. Ideally the performance of the readout architecture in terms of bandwidth and reliability should readily scale with array size. The throughput-fill factor performance trade off, associated with the existing approach to digital readout, impose on the system a set of conflicting design objectives and architectural constraints that ultimately limit the attainable performance of the array. In the final evaluation an ideal architecture does not necessarily exist [89] – In Digital SPAD-based array detectors, the choice of readout architecture directly determine the performance trade-offs imposed on the system. As a result, architecture is strictly tied to implementation. The optimal architecture is one that combines different schemes to obtain the best performance/area trade-off for a specific application [17, 124].

5.1.3.2 No Parallel-Pattern Detection Functionality

The distribution of arriving photons in the time-domain corresponds to the range of optical frequency components in the incident signal. Therefore high frequency signals, such as radiation or certain excitation/emission signal (manifested in the time-domain as photon burst) can be characterized by the parallel or near parallel patterns of triggering activity across the SPAD array. Parallel pattern discrimination can serve to distinguish the desired signal, often characterized by a burst profile, from the random background signal or the noise inherent in the detector [127].

In the synchronous digital readout mode, the sequential nature of information retrieval hides any tell-tale signs of parallel activity patterns across the pixel array. This prevents the exploitation of signal statistic for the purpose of signal characterization and noise filtering in digital readout mode.

In the asynchronous event-driven digital readout mode, coincident triggers are either dropped or temporally re-ordered and serialized by the access-control electronics [139]. Even with excellent throughput characteristic and no readout-related data loss the detection sensitivity is still limited by the Dark Count Rate (DCR) of individual SPADs. The SNR performance can be improved by restricting the integration window with external gating techniques, thus limiting the bandwidth of detection. The minimum integration window, however, is a function of the pixel DCR.

The absence of a viable technique for systematic discrimination of dark noise in the existing digital readout schemes limits the maximum achievable frame rate in SPAD-based Photon-counting imaging arrays [77, 128]. Moreover, in cases where the desired signal information is in fact encoded by the number of coinciding photon pulses i.e. detection of radiation energy signature , quantification of photon burst front, or measurement of photon number statistics to identify an atomic emission source [48], digital readout scheme becomes entirely ineffective. In this regard the performance of the SPAD array falls short to that of PMT which is capable of an analog photon-number resolving mode of operation [128].

From a digital signaling and transport point of view, large number of independently firing SPADs multiplexed onto a common readout bus present an architecture that is fundamentally un-scalable with size and activity rate. The key design challenge for the next-generation single-photon array detectors is the implementation of a scalable readout architecture that can dynamically process both time-sequential and time-parallel photon events, over a shared signaling path, while minimizing data loss or latency, and maximizing throughput.

5.2 Analog Readout – *The Silicon PhotoMultiplier (SiPM)*

In 2003 a novel implementation of the SPAD array capable of photon number resolution was introduced. This breakthrough made possible the practical implementation of the highly sought after Silicon Photo Multiplier (SiPM). In this implementation, instead of using an in-pixel digital buffer to generate a logic pulse upon photon detection, the raw avalanche signal of each triggered SPAD is directly readout across a common load. The current signal from simultaneously triggering SPAD elements will combine to produce a sum analog signal, at the output node, with an amplitude that is linearly proportional to the number of impinging photons. Consequently, the signal peak associated with a multi-photon pulse can be easily distinguished from peaks caused by random dark events. In SiPM detector, systematic dark noise discrimination can be achieved with a simple amplitude thresholding technique.

SiPM introduced a paradigm shift in the implementation of multi-element (pixelated) SPAD detectors. The analog readout implementation represented deviation from the digital Geiger mode of operation and bears greater similarity to PMT method of operation, thus the name Silicon Photo-Multiplier.

5.2.1 Analog Signal Processing Electronics – *Associated Design Tradeoffs*

The relevant information (Photon count and Time Of Arrival) is accurately encoded in the SiPM signal response. The mark of a readout system is defined by the ability to preserve the excellent characteristics of the detector signal. In the case of SiPM signal response, characterized by a rapid, transient current pulse with a large dynamic range, this is no trivial task [129]. The encoded information cannot be directly extracted from the raw detector signal. The current signal must be integrated, converted and scaled to the operating voltage of the readout electronics, then, it must be amplified, conditioned and sampled before it can be digitized and stored [130- 134].

5.2.1.1 Pre-amplification

A preamplifier is typically the first component in the analog readout chain. It acts as an interface between the detector and the pulse processing electronics and collects the charge released in the detector. A low-noise charge Sensitive preamplifier (CSA) is widely used as the input stage of the readout chain due to the insensitivity of its gain to changes in the parasitic capacitance at its input [135].

The output voltage of the CSA is proportional to the input charge, released by the detector, divided by the feedback capacitance as represented by equation 5.1. The parameter Q_s represents the charge released in the detection process and C_f represents the feedback capacitance.

$$V_{\mathrm{pre-amp}} = \frac{Q_s}{C_f} \tag{5.1}$$

The performance of the preamplifier is defined by the speed and precision with which it can reproduce, at its output, a voltage signal with the intrinsic characteristics of the detector response signal. It is therefore critical that the CSA does not introduce excessive loading or bandwidth constraints on the output signal of the detector. With regards to SiPM operation, the key performance parameters for the CSA are dynamic range and speed. The lower limit of the dynamic range is determined by the noise floor of the preamplifier, the upper range by the size of the feedback capacitor. A large feedback capacitance improves the dynamic range of the CSA, at the cost of lower gain and diminished speed (response rise time). This is the basic tradeoff associated with the CSA design for SiPM detection systems [136].

In order to enable readout across a wider dynamic range, and still allow minimum amplitude detection necessary for calibration the CSA modules are typically implemented with a variable gain feature. Fig.5.8 illustrates the basic structure of a variable gain preamplifier. Current-mode amplifiers have been proposed as an alternative to the CSA in order to address the large dynamic range requirements. However, this comes at the cost of increased power consumption [134-136].

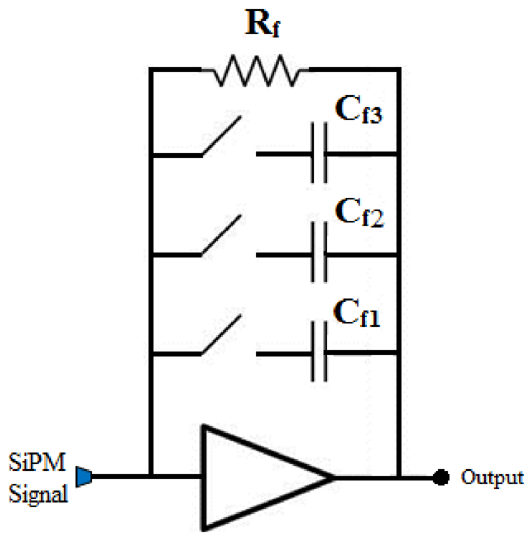


Figure 5.5**:** Variable Gain Charge Sensitive Preamplifier

5.2.1.2 Pulse Shaping

Due to the long decay time of the preamplifier (CSA) response there is an increased likelihood of overlap between signals associated with distinct input detection events. This occurs when the rising edge of a signal superimposes on the trailing edge of the previous signal. A pulse shaping amplifier is required to extract the pulse height from the consecutive rising transience in the preamplifier output signal. Fig. 5.6 illustrates the typical signal profile before and after pulse shaping [131].

Pulse shaping also optimizes the profile of the waveform for effective peak detection and subsequent digitization further up the readout chain, by scaling the amplitude and time-widening the peak of the pulse. In addition to producing a signal with a gradually rounded peak, better suited for peak detection, pulse shaping also improves the SNR by restricting the signal bandwidth (effectively acting as a filter against the electronic noise from the CSA).

A longer shaping time also result in tighter scaling of the signal amplitude hence improving the amplitude resolution and dynamic range of the measurement. Longer shaping times, however, also increase the probability of pulse pile up due to overlapping of pulse signals associated with consecutive events, which leads to errors in the measurement of the amplitude [131].

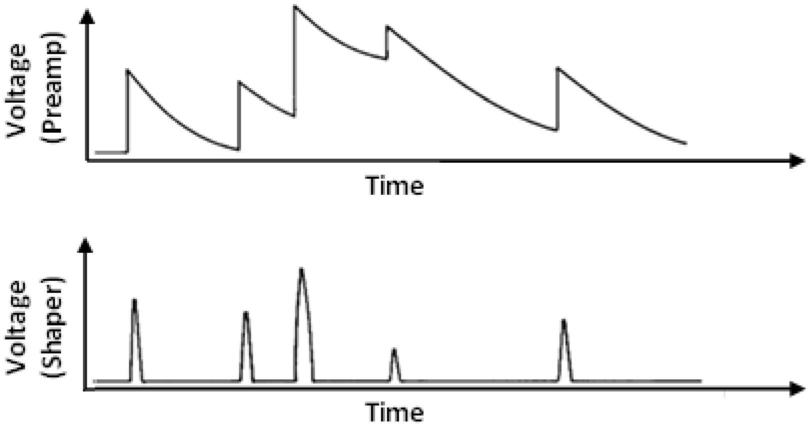


Figure 5.6: The effect of pulse shaping on the output of the preamplifier

The side-effect of improved amplitude resolution and SNR is diminished pulse pair resolution and rate capability as depicted in the Fig. 5.7. The right balance between these conflicting performance requirements must be determined according to the specific needs of the target application.

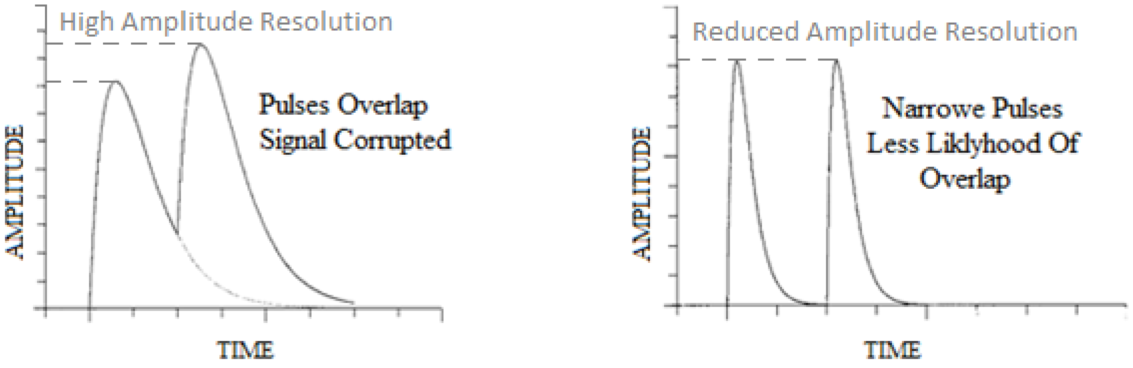


Figure 5.7: Effect of shaping time on pulse-pair resolution of the readout system.

5.2.1.3 Signal Sampling

The signal generated at the output of the pulse-shaping module has a profile optimized for sampling. The post-sampled discrete signal values are typically held in analog memory (Switched Capacitor Array (SCA)) temporarily before being read out into an Analog to Digital Converter (ADC) to undergo digitization for non-volatile storage in digital memory. The intermediate analog storage step allows a time-stretching of the signal by sampling it at a very high rate (generating a high resolution quantized signal) while reading out the quantized signal at a slower rate in order to accommodate the finite acquisition time of the ADC. As a result high speed sampling can be performed with slower ADC units (more compact, cheaper and with lower power dissipation). Sampling depth is limited by number of SCA storage cells [139-143].

A drawback of this architecture is that it can only operate in triggered mode. The signal cannot be digitized continuously as the sampling has to be stopped while the data is being read into the ADC, thus introducing additional dead time in the readout cycle.

This readout scheme is therefore best suited for applications whereby triggering events are far enough apart [142].

5.2.2 Inherent Limitations of Analog Readout

5.2.2.1 Avalanche Charge Signal is a Sub-optimal Counting Index

The avalanche current signal has a non-trivial pulse shape (due to the different carrier velocities and collection times associated with holes and electrons) so for the purpose of linear summation it is more desirable to measure the signal charge which is independent of the pulse shape. The current must be integrated over the duration of the avalanche pulse to obtain the total charge contained in the detector current pulse. The charge signal is processed as a quantity proportional to the energy deposited in the detector. The total amount of deposited energy most directly relates to the number of discrete photon quanta absorbed at the active surface of the detector, which is represented by the number of discrete avalanche events generated across the detector array. The charge signal is a RC loaded time-continuous quantity bearing a secondary relationship to the incident energy signal. It is therefore, ill-suited to act as the summation index for the inherently discrete process of photon counting. Some of the shortcomings associated with using the avalanche charge as the raw detector signal for energy measurement are briefly discussed below.

5.2.2.2 Signal Amplitude Distortion – *Sensitivity & Dynamic Range Limitation*

The magnitude of the aggregate charge released by the SiPM detector (amplitude of the induced voltage signal) is used as a linear indicator of the radiation energy. However, the magnitude of the SPAD avalanche charge is a function of operating conditions such as local temperature and the biasing voltage. Random variations in pixel characteristics or operating conditions across the array will introduce amplitude fluctuations in the output signal of the detector and lead to distortions in the measured intensity (photon count) [62, 131].

Additionally, the electronic noise from the associated processing and readout components become superimposed on the detector signal, further altering the signal amplitude and limiting the dynamic range of the measurement process. In the analog readout mode, it is often the electronic noise that determines the detection limit of the system [42, 143, 144].

Random base-line deviations in the signal also introduce uncertainty and error in the amplitude measurement process. The capacitive AC coupling between different stages of the readout system prevents transmission of DC bias signal between connected components. However, in absence of a DC component to hold the baseline value, noise related random fluctuation in the signal will cause the baseline value to fluctuate. The magnitude of the base-line signal will shift in order to make the overall transmitted charge zero [14, 142]. Base-Line Restoration (BLR) circuitry can be used to correct for this effect [140].

5.2.2.3 Signal Time-Response Distortion

Precise extraction of time of arrival, associated with the detection event, is a critical requirement in many applications. The precision with which event time can be determined is limited by the time-walk effect. The simplest way to determine the time of occurrence is to generate a fast pulse when the leading edge of the detector signal crosses a fixed discrimination threshold. However the cross-over of the detector signal is not solely the function of arrival time but also depends on signal rise time and is therefore amplitude dependent. For a fixed trigger threshold, cross-over time will shift with the signal amplitude [131, 145]. The time-walk concept is illustrated in Fig. 5.8.

The slope of the input signal is also affected by the random variations in the arrival times of individual events within the integration window. Inter-arrival times for a

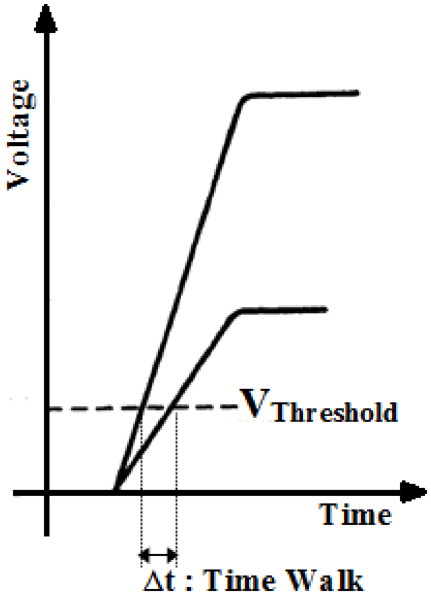


Figure 5.8: Time walk effect demonstrated for two simultaneously arriving detection events with different amplitudes.

random process are exponentially distributed. Therefore the standard deviation of inter-arrival times is the same as the mean and can be represented by the quantity $\frac{1}{N}$, N being the number of arrival per unit time. The shot noise dependent jitter in the rise time of the input signal can be considered as inversely proportional to signal amplitude.

5.2.2.4 Static Integration Window – *Limited Applicability, No Adaptability*

In the conventional analog SiPM design the signal summation window is inherently determined by the width of the avalanche current pulse. This inherent self-gating proves to be highly effective for noise hit rejection in cases where individual photons in the incident optical signal, arrive at the detector within one avalanche cycle. i.e. within the narrow time margin between avalanche propagation and extinction. Otherwise no detection will be registered by the readout system. As a result, due to the static nature of its integration window, the application scope of SiPM is primarily limited to large optical bursts of very short duration.

Another drawback stemming from this limitation is performance adaptability of SiPM detector. Most performance metrics including SNR, active range, detection bandwidth, DCR and power consumption can be improved by minimization of the avalanche duration. From a performance point of view it is always desirable to shorten the avalanche duration as much as possible [45]. However any modifications to the SPAD micro-cell, whether at the implementation level (front-end active quenching electronics) or physical level (alteration in SPAD active area), that affect the duration of its avalanche signal will directly impact the detection criteria of the SiPM macro-pixel.

Therefore device level changes may arbitrarily alter the system level response. This could mean that avalanche events corresponding to a target burst signal might no longer overlap to produce a detection signature at the readout node of the SiPM even if the signal is temporally distinctive over the noise.

Existing analog SiPM systems are unable to adapt to enhancement in micro-cell performance characteristics, smart pixel designs and creative system-level configurations such as nearest neighbor clustering for dark noise minimization which currently is only possible by reducing the preamplifier integration time and/or operating the detector at cryogenic temperatures. The limitation of the analog SiPM in scaling with micro-pixel level performance is depicted in Fig. 5.9.

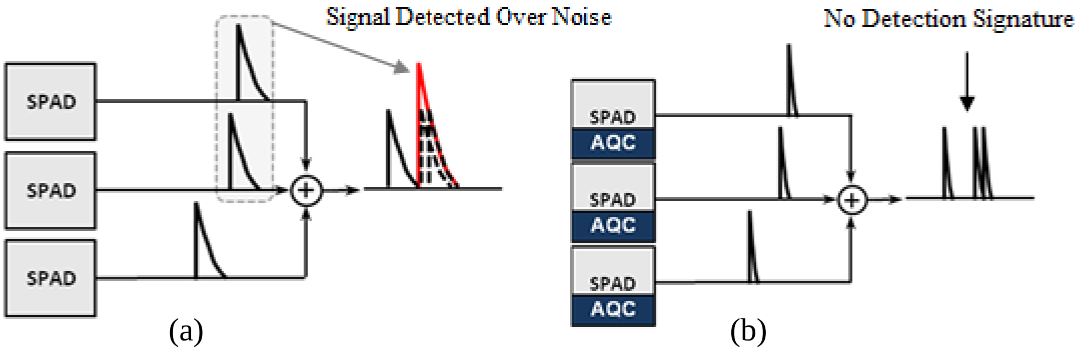


Figure 5.9: Shortening the avalanche pulse, through active quenching, is highly desirable due to its enhancing effect on SNR and active range performance of the device. However such pixel-level improvements interfere with the detection criteria at the SiPM macro-pixel level.

5.2.2.5 Design Complexity and Performance Trade-off

Preserving the analog characteristics of the SiPM output signal (dynamic range and rise time) places stringent requirements on the analog readout electronics. The set of conflicting design objectives associated with the analog readout electronics were previously discussed in section 5.2.1. It was shown that performance improvement in one area often involves compromise in another. For example, faster pulse shaping time improves the readout speed and the throughput of the system but degrades the SNR (larger band width, thus less rejection of injected electronic noise) and the dynamic range. The basic block diagram of the standard analog signal processing and readout chain used for SiPM detectors appears in Fig. 5.10. The SiPM macro-pixel is represented, in the figure, as a multi-cell detector with 11 micro-cells labeled ‘a’ through ‘k’.

Time of arrival and intensity (photon number) are both critical parameters that must be extracted with as much speed and precision as possible. This action requires precise time pick-off and high resolution amplitude measurement, and is associated with conflicting readout solutions as it simultaneously requires fast and slow pulse shaping of the preamplifer signal. The standard architecture for SiPM readout is shown in Fig. 5.10.

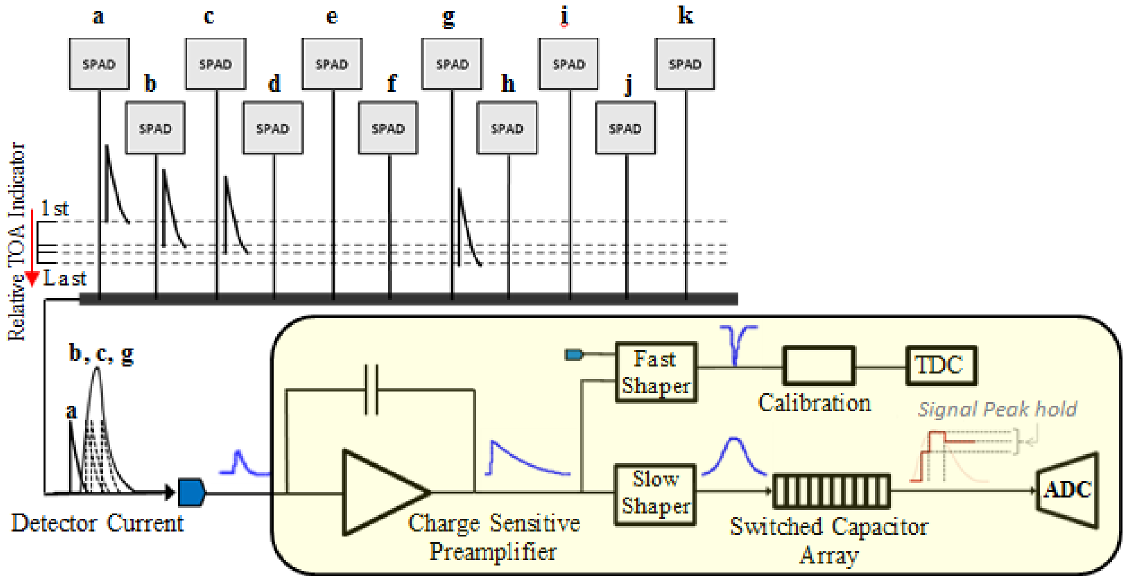


Figure 5.10: Typical Analog readout chain used in conjunction with SiPM detectors

In most SiPM applications timing precision and amplitude resolution are both critical parameters and neither may be compromised in favor of the other. Therefore the standard solution as shown in Fig. 5.10 is to split the pre-amplified detector signal and perform each operation independently This of course leads to added design complexity and cost as well as increased power dissipation.

Note that in Fig. 5.10, even though the pulse response from SPAD labeled ‘a’ does not fall within the summation window, and is therefore considered random noise, it is nonetheless processed by the readout electronics across the amplification, scaling, shaping and time extraction stage before it is recognized as noise and discarded. Due to its stringent characteristic, threshold comparison directly on the raw detector signal is not possible. The signal must be sufficiently processed before noise or signal determination is made. The resulting impact on the performance of the system is increased power dissipation (due to constant electrical activity in the readout chain) and reduced throughput due to the additional dead times from spurious triggers caused by noise (while the system is busy processing noise through the analog readout chain, it will be unavailable to the incoming signal) [148].

Due to multiple design trade-offs SiPM readout optimization is only possible for a subset of performance parameters pertaining to specific requirements of a particular application. No off the shelf, standardized electronic readout module is commercially available for SiPM detectors. However, multiple groups across academia and industry are involved in design and development of readout ASIC for these detectors [149-160].

The level of power dissipation involved and the amount of signal processing hardware required for high performance analog readout implementation inhibits the on chip integration of the readout chain. The resolution of conventional multi-channel SiPM readout systems is limited by the wire spacing between different ASIC circuit board channels, which is typically on order of 1mm. Performance demand of most advanced applications in niche SiPM markets (Nuclear Medical Imaging and high energy particle detection) has been steadily moving towards higher readout granularity, requiring greater number of readout channels packed closer together [171].

Although high density integration of SiPM macro-pixels in CMOS is now easily achievable, bringing the same level of integration to the electrical signal path with a large number of readout channels pose a significant challenge. Development in this area is currently the dominant trend in design and development of SiPM sensors for scintillation-based radiation detectors and PET imagers [161]. In order to eventually maximize performance and versatility of the detection system and fully exploit the advantages of SiPM, full integration of SiPM detector and readout electronics is required.

5.3 Digital SiPM – *Displaced Complexity*

Digital SiPM represents the most recent innovation in SPAD-based detection field. In this implementation, individual SiPM micro-cells , each with integrated 1-bit counter are connected to a shared multi-bit counter with a synchronous data bus. The signal intensity is measured by sequentially counting across the triggered 1-bit counters, thus eliminating the analog readout chain all together. In-pixel electronics also couple to a separate trigger-detect network that implements statistical trigger detection based on repeated logical OR operations on output of individual SPAD cells [162, 163]. The system implements a more complicated, digital version of the analog amplitude thresholding process by controlling the counting and timing operation of the sensor using signals from the trigger-detect network. The noise discrimination feature, achieved automatically in the conventional analog configuration is implemented digitally with an embedded refresh logic to zero out the output of the 1-bit counter if no trigger detect signal is received within a pre-set time window [164,165]. In effect Digital SiPM is a variation of the parallel In-pixel SPAD image sensor architecture with an integrated 1-bit counter as was mentioned in section 5.1.1.3. The only difference is that it is modified for single-shot rather than continuous detection. In other words, rather than being synchronously clocked the readout operation is dynamically triggered using a trigger-detect network.

The digital operation renders the system less susceptible to gain variations and electronic noise resulting in improved photon-count and TOA measurements. The performance gain, however, comes at the cost of lower pixel fill-factor and greater operational and structural complexity.

5.4 General Summary and Critique of the Existing Readout Paradigms

Single SPADs essentially behave as digital single-photon switches with external reset. Accordingly a digital architecture would seem the most obvious choice for the readout of a SPAD array. Yet the performance of such a readout architecture is severely limited. Transmission of multiple digital pulses across a shared medium is a time-exclusive operation as digital pulses are discrete events and cannot occupy the same space at the same time. Resource sharing in digital systems is fundamentally a synchronous process. SPADs however, are inherently asynchronous devices with dynamic response profile. Digital readout across a shared medium essentially results in loss of temporally correlated signals, with the loss mediated by the degree of sharing implemented. Digital technology is therefore intrinsically ill-suited as a transmission platform for a system featuring a large array of autonomous, dynamically firing elements, such as a SPAD array.

Analog technology, on the other hand, is inherently asynchronous and can support dynamic data transmission without rate or timing constraints. Conceptually an analog architecture would represent an optimal signal encoding platform for an array of independently firing dynamic nodes. SiPM configuration exploits the inherent advantage of analog technology as it relates to the dynamic aspect of SPAD operation, especially relevant in array-based implementation. In SiPM detectors, the analog readout architecture allows constraint-free dynamic signal transfer from multiple micro-pixels to a common output node across a shared path, thus offering maximum operational parallelism using a single readout channel. The effective transmission bandwidth is comparable to having a dedicated readout channel for each SPAD device. The ability to dynamically represent different patterns of parallel activity, across multiple micro-cells, on a single readout node has been the key attribute responsible for making SiPM the SPAD-based single-photon detector of choice in radiation detection. Yet the analog approach has fundamental shortcomings. Although it enables a very simple and precise scheme for dynamic encoding of the signal at the array-level, it displaces all the imprecision and complexity from the signal encoding level, to the signal decoding level. The robust digital voltage response of the SPAD device is replaced by analog charge signal that is highly susceptible to fluctuation and noise. The electronic extraction of the relevant information from the analog response of the SiPM detector introduces many challenges and trade-offs that ultimately constraint the performance of the detector and limits its applicability.

It was understood early on that, as with any high performance pixel, the tremendous potential of SPAD technology lies in development of large monolithic arrays. The SPAD response can be described, highlighting its most impressive features, as dynamic in nature and digital in profile. The digital readout approach takes advantage of the digital nature of SPAD response at the device level but fails to support the dynamic aspect of SPAD operation in context of a large array. The analog implementation supports the dynamic aspect of the large SPAD arrays at the cost of the robust digital response at the device level.

The dominant strategy to address the performance limitations of the analog architecture has been signal processing brute force. Analog SPAD arrays (SiPM) operate in conjunction with large power intensive readout ASIC. In the digital domain resource sharing has emerged as an implementational compromise in face of trade-off between performance and fill factor.

5.5 Proposed Readout Architecture – *A New Encoding Paradigm*

It seem clear that an all-digital or an all-analog solution to the readout problem of monolithic CMOS SPAD arrays inevitably runs into limitations stemming from fundamental constraints of the underlying signaling technology. Addressing readout challenges and limitations constitute the bulk of the existing research in the field and dominant approach has been through increasing structural complexity.

If SPAD operation is characterized as dynamic in nature and digital in profile with each functional characteristic best supported by a different signaling technology, as described in section 5.4, it would seem that a hybrid system offers the most apt readout solution. A digital model adequately describes the instantaneous state of the detector at the micro-pixel level and is the most robust signaling scheme at this level. However, considering that the state of the detector at the array level can be represented, at any given instance of time, by the collective states of its constituting elements (the number of active elements), an all-digital approach, that is binary in nature, would have clear limitations in representing the instantaneous state of the detector array. Array-level response is most optimally represented by a discrete-time analog signal.

5.5.1 Discrete-Time Representation of Detector response

Instead of compromising between an all-analog or all-digital approach, a new paradigm for representation of the array response is proposed that mergers the best characteristics of digital mode (generation of a logic pulse with standard size and shape at the pixel level) with those of the analog mode (dynamic linear summation of individual SPAD signals over a common readout wire to represent the array-level signal). In this way the system benefits from the architectural simplicity and dynamic nature of analog readout, while minimizing noise and additional signal processing by early digitization of the SPAD signal. In a nut-shell, the proposed array signal encoding methodology incorporates the advantages of both readout schemes without suffering the drawbacks of either.

This is made possible across a sense interface that supports multiple digital input ports and a single DC-biased floating output node. The digital outputs of the SPAD cells are capacitively coupled to a shared floating terminal which comprises the output port of the detector interface. The interface dynamically converts the simultaneous digital activity on its input ports into a representative discrete-time signal at its output port, by direct weighed summation of temporally correlated input pulses. The basic schematic for the detector interface is illustrated in Fig. 5.11.

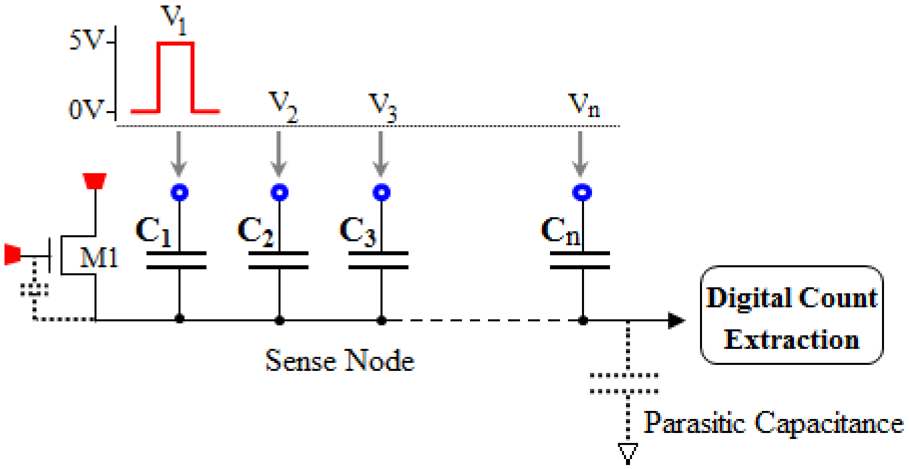


Figure 5.11: General representation of the detector readout interface

The layout diagram for a 10-input detector interface, utilizing minimum size poly2/poly capacitors is shown in Fig. 5.12. The silicon area occupied by the interface is insignificant relative to the area taken by the connected SPAD pixels.

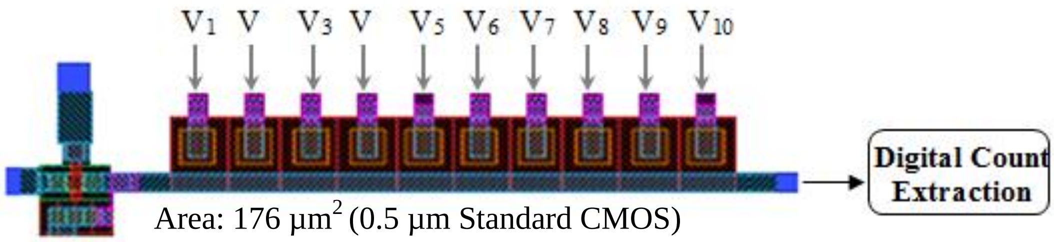

V1 V V3 V V5 V6 V7 V8 V9 V10
Digital Count Extraction
Area: 176 µm2 (0.5 µm Standard CMOS)

Figure 5.12: Compact layout structure of a ten input readout interface.

The bottom plates of the input coupling capacitors are connected together to form the floating output node of the detector interface. According to the charge conservation law, the charge trapped on an isolated (floating) node is redistributed among all the connected capacitors in response to signal transitions on a capacitively coupled neighboring nodes, but a the total amount of charge remain the same before and after the signaling. Therefore, in reference to Fig. 5.11, when a SPAD device fires, activating its corresponding input ports, the electrical energy of the signal is converted to an electrostatic charge induced on the top plate of the respective coupling capacitor. Accordingly, an equivalent amount of charge, of opposite polarity, is re-distributed on the floating bottom plate [165, 166]. A system of charge distributes itself in such a way to minimize the electrostatic potential energy of the system and this equates to a uniform distribution where charges are equally distanced apart from one another across the surface of the conductor. Assuming the initial trapped charge, Q_0, is zero and the total parasitic capacitance from all transistor driven by the floating sense node is $C_{P,}$ the resulting charge relation at the floating node after triggering of its input ports is expressed by equation 5.2 [167, 168].

$$Q_0 = C_1(V_{fn} - V_1) + C_2(V_{fn} - V_2) + \cdots + C_n (V_{fn} - V_n) + C_P V_{fn} \qquad (5.2)$$

The change in the floating node voltage in response to the signaling activity on the coupled input ports is represented by the parameter V_{fn}. C_1 through C_n are the equal-sized input coupling capacitances and V_1 through V_n represent the corresponding voltage transitions on the input cross-connects. The equivalent capacitance seen at the output floating node is equal to the sum of all the input coupling capacitances, C_T plus the parasitic capacitance (Fig. 5.11).

$$C_1 + C_2 + \cdots + C_n + C_P = C_T + C_P \qquad (5.3)$$

Using equation 5.2 and 5.3, the voltage on the floating node can be expressed as

$$V_{fn} = \frac{C_1V_1 + C_2V_2 + \cdots + C_kV_k}{C_T + C_P} \qquad (5.4)$$

The subscript k represents the number of triggered inputs out of n total connections. Therefore induced voltage on the floating node can be expressed as linear sum of voltage transitions on the input coupling ports scaled by the total number of cross-coupled connections, minus the voltage from the induced charge on the parasitic capacitance. Equation 5.5 illustrates the basic expression for the induced voltage on the floating node with ‘n cross-coupled connection in response to parallel activation of k inputs [170].

$$\mathrm{V_{fn}} = \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{k}} \mathrm{V_i} - \frac{\mathrm{Q_P}}{\mathrm{C_T}} \tag{5.5}$$

Form equation 5.4 it is observed that the parasitic capacitance reduces the induced voltage on the floating node by a factor represented by the expression 5.6. The floating

$$\frac{\mathrm{C_T}}{\mathrm{C_T} + \mathrm{C_P}} \tag{5.6}$$

sense node essentially acts as a charge sharing line onto which input signals can be coupled capacitively. The voltage at the sense node is a function of charge division between the capacitors connected to the floating node minus the charge induced on the parasitic capacitance [170].

Note that the scaled voltage summation at the floating node is not restricted to simultaneous signals and can be applied for different time intervals. This window can be actively adjusted by varying the pulse-width of the output response associated with the connected SPAD pixels.

The quiescent charge on an isolated (floating) node is not predictable or controllable (charge trapped on the node during fabrication process) and its DC quiescent state (No activity on the inputs) is typically undefined. Therefore, in order to set the initial voltage condition of the floating node to a known reference value (typically zero), a high impedance path to a reference voltage (typically ground) is connected to the output floating node. This is accomplished by the cut-off transistor M1 in Fig. 5.11. The gate of the NMOS is grounded to set the high impedance path – The anchor-bias terminal connects the drain of the NMOS to the desired baseline DC bias that will serve as the reference voltage level for the induced voltage variations on the floating node.

5.5.1 Quantized Detector Response – *Performance Aspects*

Since the proposed method of detector signal encoding does not involve integration of charge as a proxy representation for a detection event, the dynamic range of the output detector signal is not restricted by the charge handling capacity of the interface electronics. Rather the active range of the detector signal is inherently tuned to operating voltage (gnd – VDD) and cannot exceed it. Furthermore the detector signal amplitude is a direct indication of the number of the photons detected and not a proxy quantity proportionally related to the detected photon count. The raw detector signal exhibits a flat top profile that is directly available for digitization without requiring further shaping or conditioning. This significantly simplifies the readout chain by excluding the several signal conditioning components required in the analog mode.

In the proposed method the slope of the detector signal is constant regardless of the signal amplitude. As a result, the time walk effect, a major source of timing uncertainty in the standard analog readout, is eliminated. Additionally the performance of the proposed architecture is unaffected by the performance constraint typically associated with increasing the number of interconnections in a shared access configuration, such as longer data transfer delay, greater power consumption and decreased bandwidth.

5.5.2 Adjustable Measurement Window – *Utility Aspect*

The noise hit rejection by signal amplitude thresholding is valuable feature in analog implementation of SPAD arrays. The inability to achieve the same in digital mode has been a key limiting factor in performance and utility of digital SPAD arrays. The proposed architecture enables a digital implementation of noise filtering based on amplitude thresholding. The Overlap rate (integration window) is controlled by the adjustable duration time of the SPAD readout pulse**.** As described in Chapter 3 all SPAD devices have been equipped with an actively adjustable event pulse generator as part of their integrated front-end circuitry. Since pulse summation occurs during the overlap window between the readout pulses, a variable pulse-width feature enables active configuration of system detection window in accordance to the application requirements or target signal characteristics**.** The adjustable gating feature enables variable energy windowing and more effective system-level noise suppression. In general it extends the detection scope of the system. The simulation of 4-input interface for two different SPAD readout pulse-width is shown in Fig. 5.13.

The simulation illustrates the relationship between the pixel output pulse-width (the integration interval of the cluster signal) and the induced interface output voltage. As can be seen from the simulated results varying the gating window modifies the detection criteria of the system.

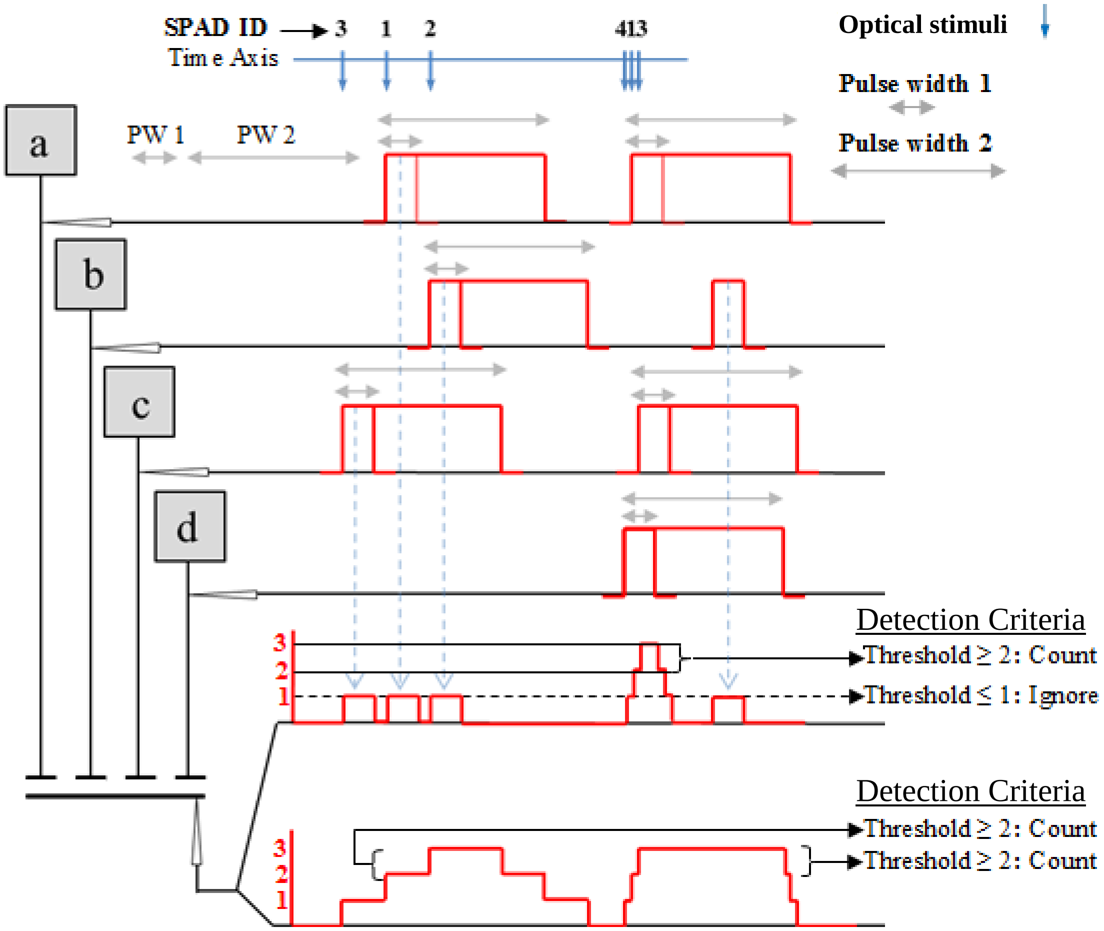


Figure 5.13: Simulation results showing the response at the output port of a 4-input interface for two different input pulse widths PW1 and PW2.

5.6 Summary

Digital SPAD array implementations suffer from architectural constraints that limit their throughput. Analog implementations benefits from the inherent asynchronous nature of analog signaling; however at the device level the analog output is highly prone to amplitude and timing variations. The extraction of relevant information from the analog detector response requires complex and power-intensive signal processing electronics. As such the utility of each approach has been limited to specific applications. In this chapter a novel paradigm for representation of SPAD-based array response has been proposed whereby the robust nature of digital single photon SPAD output can be exploited in an array configuration capable of parallel detection for resolution of multi-photon packets. The photon resolving capability for an array composed of digital SPADs has been made possible across a multi-input detector interface with single output port comprised of a DC-biased floating node. The resulting platform enables dynamic summation of multiple digital pulses with the sum signal represented as a discrete voltage waveform with clearly defined discrete amplitude levels. Existing digitally implemented shared access architectures are based on first-in take–all sharing scheme. The proposed architecture offers a platform for multiple digital SPADs to dynamically share a resource, i.e. digital counter, without data loss or delay. The hybrid architecture that operates in both analog and digital domain improves the performance of SPAD-based array detectors in both photon counting and photon number resolving modes of operation.

Chapter 6

Fully Integrated Asynchronous Digital Decoding System

6.1 System Architecture Overview

Digital circuits offer robust performance in processing and storage of electronic data, thus there generally is a strong trend towards integrating digital electronics with the analog front-end of optical sensors and imagers. Analog to Digital Converter (ADC) is typically used as an interface between the analog core and the digital processing electronics in conventional image sensors. In the standard SiPM design the final stage of the analog readout chain consists of an ADC that performs the final digital decoding of the relevant data. The accuracy of the digitization process depends on the analog sampling rate and the bit resolution of the ADC. Performance improvement, in terms of ADC signal conversion accuracy and speed comes at the cost of larger silicon real estate, and greater power dissipation, typically in the range of tens of mW [171, 172]. ADC modules generally require additional digital support circuitry to provide timing and readout control. As such monolithic integration of ADC presents a few problems. There is a great deal of research being conducted into more compact higher performance analog processing and readout systems including compact, low power ADCs that offer adequate accuracy and speed for effective digitization of the analog SiPM signal [173, 174].

The proposed readout system consists of two main functional units – operationally intertwined. The First unit is the multi-access conversion interface responsible for *dynamic* encoding of the incoming digital data. The design and operation of the encoding interface unit was described in chapter 5. The interface signal does not preprocessing and can be directly read into an ADC. However, since the signal on the floating output node of the interface is highly sensitive to parasitic capacitance external processing with an Off-chip ADC is not an option. In fact, parasitic effect of long on-chip interconnects should also be avoided. Hence it would be best to perform the digitization as close as possible to the detector interface.

This chapter proposes alternative digitization architecture to comprise the second functional unit of the readout system. This method involves deconstruction of the main detector signal into a set of serially generated pulses, equal in number to the corresponding parallel input set. The pulse train is sequentially processed by a digital counter to reproduce the incident photon count in binary format. The conceptual block diagram of the digitization system is shown in Fig. 6.1.

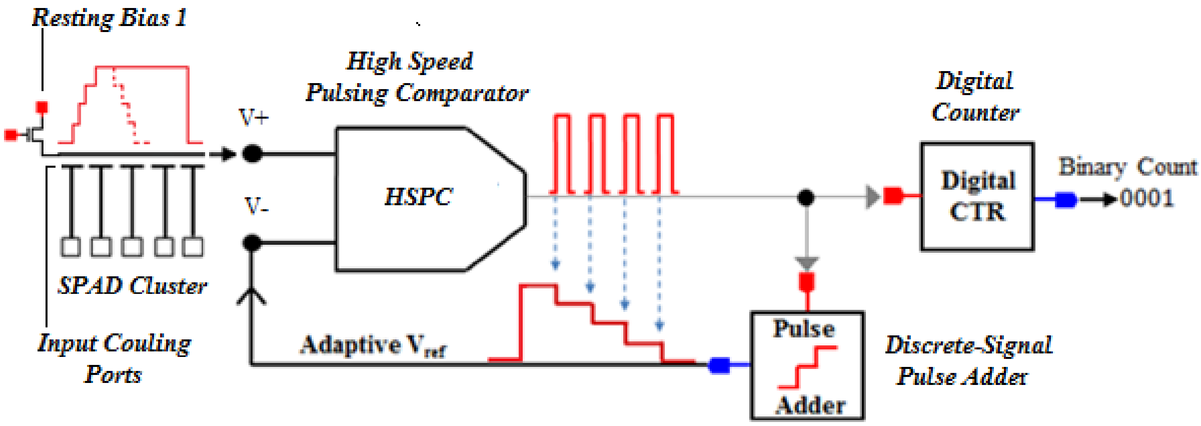


Figure 6.1: Conceptual signaling block diagram of the readout system showing the primary internal signals and the operational flow of the overall system.

The High Speed Pulsing Comparator (HSPC) is triggered high when the voltage signal at its positive input terminal, coupled to the detector interface, exceeds a predetermined threshold value. The HSPC generates a continuous stream of narrow pulses while the voltage on its positive terminal remains higher than the reference voltage on its negative terminal. The design of the HSPC is discussed in section 6.2.1. The generated pulses are subsequently added by the Discrete-Signal Pulse Adder (DSPA). The summation is carried out by the same method of charge redistribution used to generate the signal on the detector interface. The difference is that instead of the multiple parallel input lines of the interface circuit, the pulse adder only has a single serial input line. The input line routes each incoming pulse to an immediately adjacent open coupling port by sequentially stepping through the capacitor array.

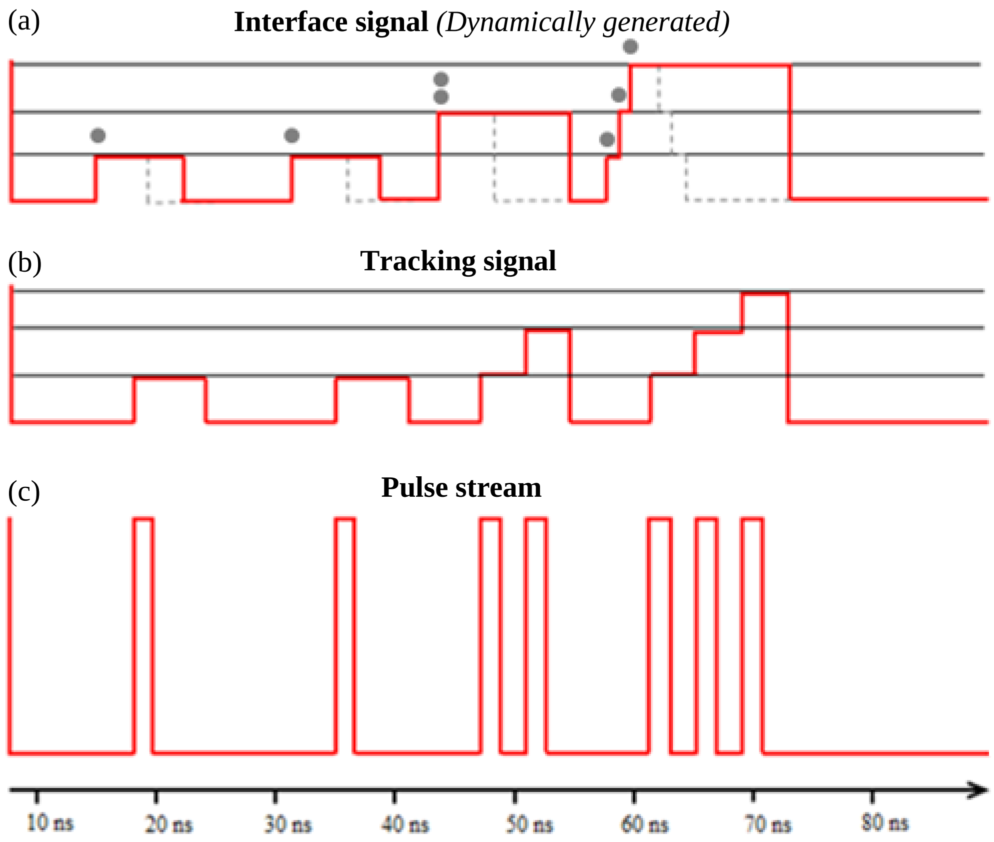
As illustrated in Fig. 6.1 the output of the DSPA is connected to the negative terminal of the HSPC module. Therefore the DC bias of the DSPA unit provides the discrimination threshold for filtering noise events. The output of DSPA represents an adaptive reference voltage that successively tracks the signal on the positive terminal of HSPC in discrete increments. Once the signal level on the positive terminal is exceeded by the tracking signal, the HSPC terminates the pulse stream.

The interface output voltage moves in accordance to the signal transitions on its input lines. However since events are represented by digital pulses each step rise in the interface voltage in response to a rising transition on one of its inputs will eventually be followed by a step reduction from the falling edge of the corresponding pulse. This is useful in the case of temporally uncorrelated noise events that do not trigger a counting cycle, as it ensure that the associated interface response is quickly diminished.

On the other hand when multiple SPAD cell fire in parallel the rising edge of their corresponding output pulse lines up to induce a proportional rise in the interface voltage (biased floating node) that may exceed the detection threshold and trigger the counting process. As stated above the counting process terminates when the amplitude of the interface signal is exceeded by the tracking signal. Therefore any drop in the interface voltage level (in response to a falling pulse edges), prior to the completion of the counting process could prematurely terminate the serial pulse stream and distort the count measurement. As a result, while the count extraction process is in progress the interface signal level must not be allowed to drop. The implementation of this feature is further discussed in section 6.2.3.

The profile of the main internal signals of the system is shown in Fig. 6.2. The incident optical stimuli are represented by the grey dots in Fig. 6.2a. The waveform edge represented by dashed gray line represents the falling edge of interface voltage as it would proceed if the detection threshold was not crossed. By observing the system primary signal in Fig 6.2 it becomes clear that the system schematic shown in Fig 6.1 is not complete. There is no circuitry for generation of a timing signal required to hold and release the interface signal and reset the tracker signal in synch with the start and end of the counting cycle.

The time boundaries of the HSPC signal mark the start and end of the counting process, however the HSPC signal is a uniform pulse train and cannot be used as a clocking signal to indicate the start and end of the count cycle. Therefore a second High Speed Comparator (HSC) is required to generate a clocking pulse whose rising and falling edge signals the start and end of the digital count extraction process respectively.



(a) **Interface signal** *(Dynamically generated)*

(b) **Tracking signal**

(c) **Pulse stream**

Figure 6.2: Primary system internal signals. (a) Dynamic signal on the output port of the detector interface. (b) Tracking signal at the output of the DSPA module. (c) Serial pulse stream.

The width the control pulse would represent the duration of the counting process and hence the magnitude of the extracted count. Since count conversion is performed sequentially, longer counts take a proportionally longer to complete. The complete system level block diagram is shown in Fig. 6.3.

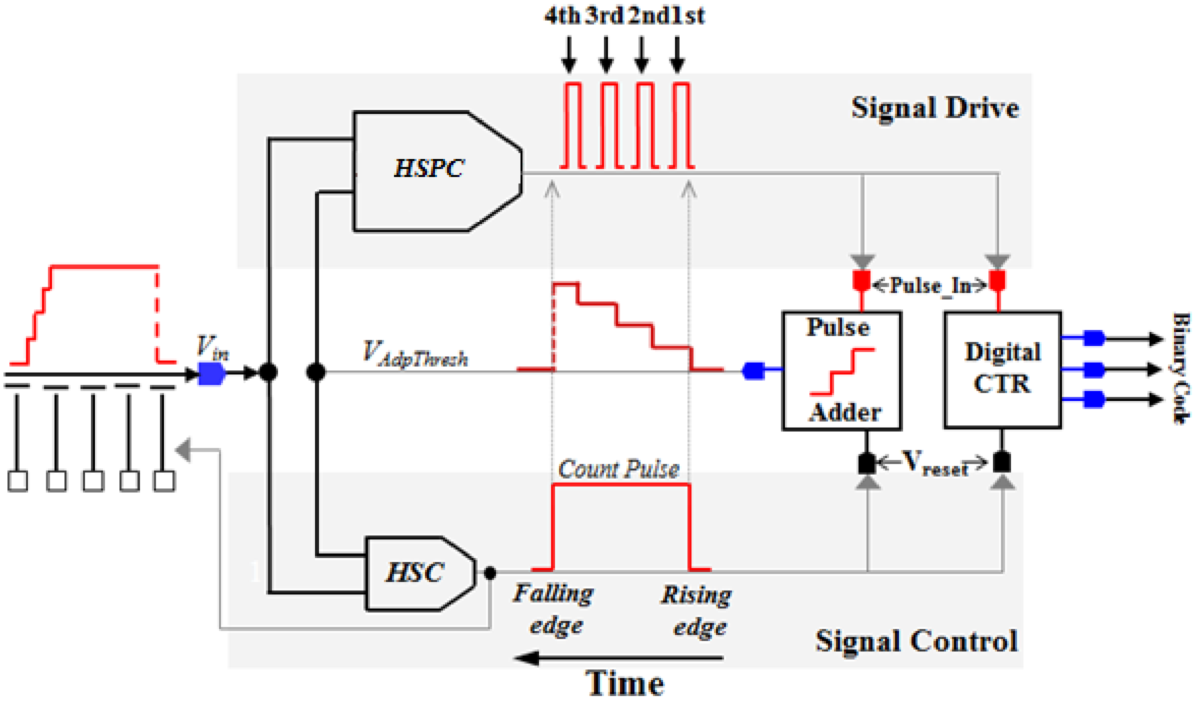


Figure 6.3: Conceptual block diagram of the complete readout system, featuring the drive signaling portion and the control signaling portion of the overall architecture.

6.2 System Electronics

6.2.1 High Speed Comparator

The key performance parameters for a comparator are gain and response time. Gain of the comparator determines the smallest signal deviation at the input that would trigger the output of the comparator to switch. It determines the smallest ΔV_i that the comparator will respond to. Response time is the time interval between input and output signal transition. Here I use a conventional three stage comparator comprised of an input buffer, decision stage and output buffer as depicted in Fig. 6.4.

A key requirement for the input stage is to minimize the loading effect on the detector interface signal. Therefore very low input capacitance and very high input

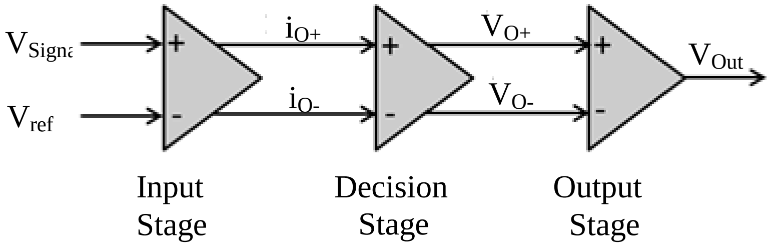


Figure 6.4: Conceptual block diagram of the standard three stage high speed comparator.

resistance are required parameters. The decision stage is composed of cross coupled NMOS latch driven by the current output of the input stage. Cross-coupled connection in the decision circuit results in a positive feedback effect which will enable faster switching speeds. The schematic for the regenerative comparator is shown in Fig. 6.5.

All transistors are sized 4/2

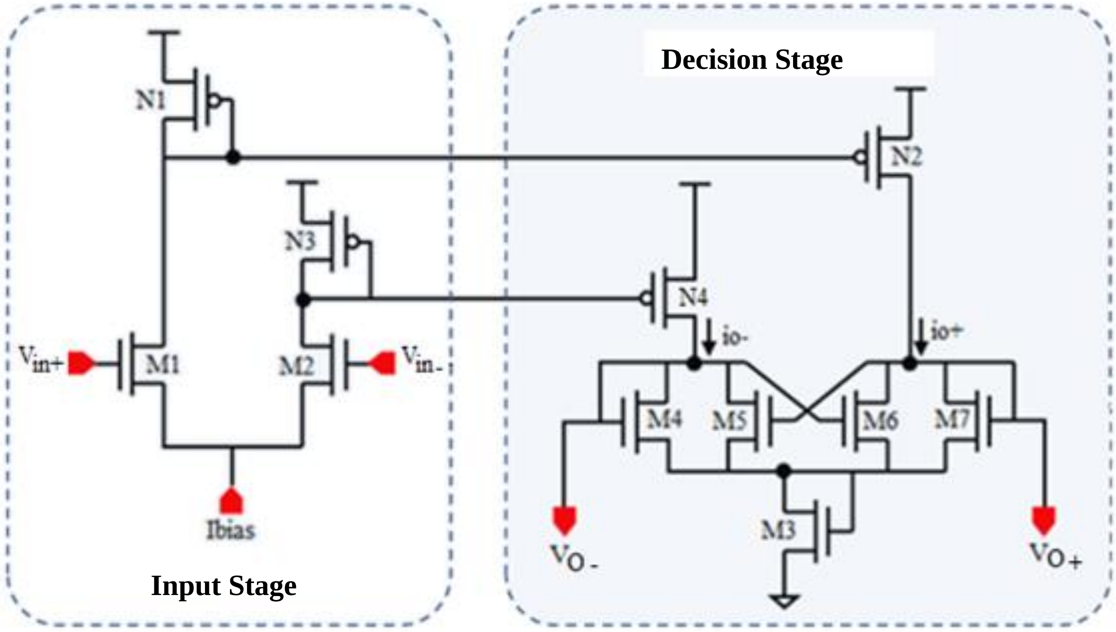


Figure 6.5: Input and the decision stage of the High Speed Comparator

The gain of the input stage is given by the equation 6.1. The output current is composed of the small signal current of the driving NMOS (M1) and the bias current component. As shown in equation 6.2 [174].

$$\mathrm{A_v} = \frac{\mathrm{g_{m(M1)}}}{\mathrm{g_{m(N1)}}} \tag{6.1}$$

$$\mathrm{i_{o+}} \quad = \quad \mathrm{g_m}\left(\frac{\mathrm{V\,in_+ - V\,in_-}}{2}\right) + \frac{\mathrm{Ibias}}{2} = \mathrm{Ibias} - \mathrm{i_{o-}} \tag{6.2}$$

In the Decision circuit the sum of the input differential currents equals the total current I_B. Referencing Fig. 6.5, if io- is raised well above i_(o+) the gate-drain voltage of M7 start to drops to reflect the magnitude of i_(o+). As a result transistor M5 starts to shut down. The current (io-) at this point can be represented by the current through M5. This is shown in the equation 6.3.

$$\mathrm{i_{o-}} = \mathrm{i_{M5}} = \frac{\beta_\mathrm{x}}{2}\left(\mathrm{v_{o-}} - \mathrm{v_{th(Nmos)}}\right)^2 \tag{6.3}$$

Where $\beta_\mathrm{x} = \mathrm{K}'\frac{\mathrm{W}}{\mathrm{L}}$ relates to transistors M5 and M6 while $\beta_\mathrm{y} = \mathrm{K}'\frac{\mathrm{W}}{\mathrm{L}}$ relates to transistors M4 and M7 in Fig 6.5. As i_{o+} increases, due to the conservation of the total latch current (I_B), i_{o-} will correspondingly decreases. The reduction in i_{o-} will induce the output voltage v_{o-} to decrease according to the equation 6.3. Since v_{o-} is realized with reference to the voltage at the common source node of the latch circuit in Fig. 6.5, it can be represented as the gate source voltage of transistor M6 (v_{gs6}). The decrease in v_{o-}

will eventually shut off transistor M6. At the switching point, the current i_{o+} can be represented by the saturation current through M6, shown in equation 6.4.

$$\mathrm{i_{o+}} = \mathrm{i_{M6}} = \frac{\beta_x}{2}\left(\mathrm{v_{o-}} - \mathrm{v_{th(nmos)}}\right)^2 \tag{6.4}$$

An expression for the differential current can be derived by dividing the expression in 6.4 and 6.3. This expression is shown in equation 6.5

$$\mathrm{i_{o+}} = \frac{\beta_x}{\beta_y}\,\mathrm{i_{o-}} \tag{6.5}$$

Similarly the differential output voltage of the latch circuit can be expressed in terms of the β parameter by substituting the expression for $\mathrm{i_{o-}}$ and $\mathrm{i_{o+}}$ into equation 6.3 and 6.4 respectively. The differential output response of the decision circuit can be expressed by equations 6.6 and 6.7. By combining equations 6.2 and 6.5 an expression for the input

$$\mathrm{v_{o-}} = \sqrt{\frac{2\mathrm{i_{o-}}}{\beta_y}} + \mathrm{v_{th(nmos)}} = \sqrt{\frac{2}{\beta_y}} + \mathrm{v_{th(nmos)}} \tag{6.6}$$

$$\mathrm{v_{o+}} = \sqrt{\frac{2\mathrm{i_{o+}}}{\beta_x}} + \mathrm{v_{th(nmos)}} = \sqrt{\frac{2}{\beta_x}} + \mathrm{v_{th(nmos)}} \tag{6.7}$$

The maximum output voltage of the decision circuit is equal to:

$$v_{o+(\mathrm{max})} = \sqrt{\frac{2Ibias}{\beta_x}} + v_{th(nmos)} \tag{6.8}$$

trigger signal ($\mathbf{\Delta v_{in}}$) can be derived in terms of the bias current ($Ibias$) and transistor aspect ratios. This is represented by equation 6.9. The minimum input trigger signal sets the limit for the smallest signal level that can be detected by the HSC

$$\Delta v_{in} = \frac{\text{Ibias}}{\text{gm}} \left(\frac{\beta_x - \beta_y}{\beta_x + \beta_y} \right) \tag{6.9}$$

The output stage of the HSC serves as a high speed buffer for the output signal of the decision stage generating a fast rising digital output pulse. A complementary Self Biased Differentia Amplifier (CSBDA) is used to implement the output stage of the High Speed Comparator. The schematic of the output stage appears in Fig. 6.6. Self-biasing

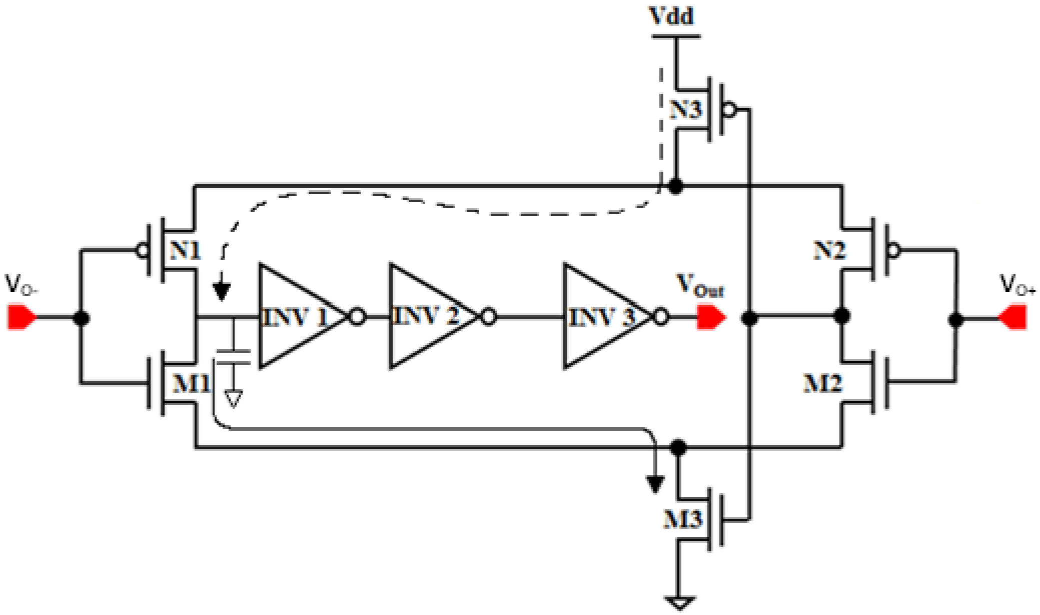


Figure 6.6: Output buffer stage of the High Speed Comparator, generally deployed for slew rate improvement and generation of an output response with a digital profile.

allows the CSBDA to generate a switching current much greater than its quiescent current. This will improve the slew rate of the output signal and increase the speed of the overall device. The schematic diagram of the High Speed Comparator in pulse mode appears in Fig. 6.7. The grayed connection comprises the modification required for the pulsing operation.

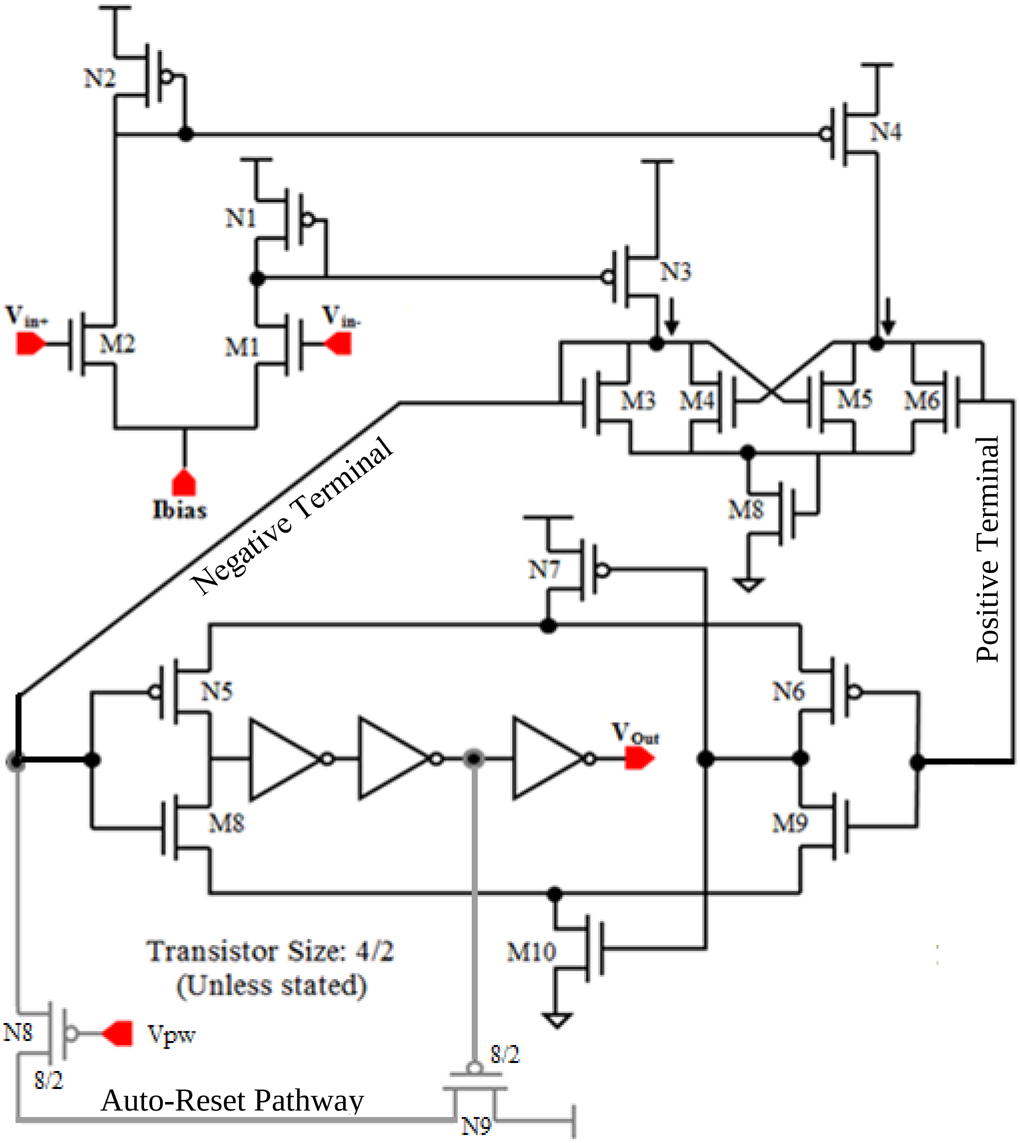


Figure 6.7: Schematic diagram of the complete High Speed Comparator with the auto pulsing feature.

The pulsing feature is enabled by modifying the output stage of the HSC. Every time the output signal of the HSC (V_{out}) rises, a connection is made, through transistor N9, between the Negative Terminal of the output buffer and VDD. This will reset the output signal to zero and turn N9 off at the same time. As long as the signal at the positive input of the HSC (V_{in+}) is greater than the reference voltage (V_{in-}) the process will repeat resulting in a continuous pulse stream. The biasing signal V_{pw} sets the impedance of the Auto-Reset Pathway thus controlling the duty cycle of the generated pulse stream. The layout diagram for the pulse mode HSC is shown in the Fig. 6.8. The HPSC is the input stage that directly couples to the detector interface. It is the unit responsible for the internal conversion of the parallel pulse activity at the input of the interface into a serial pulse stream of equal pulse count at the output of the interface.

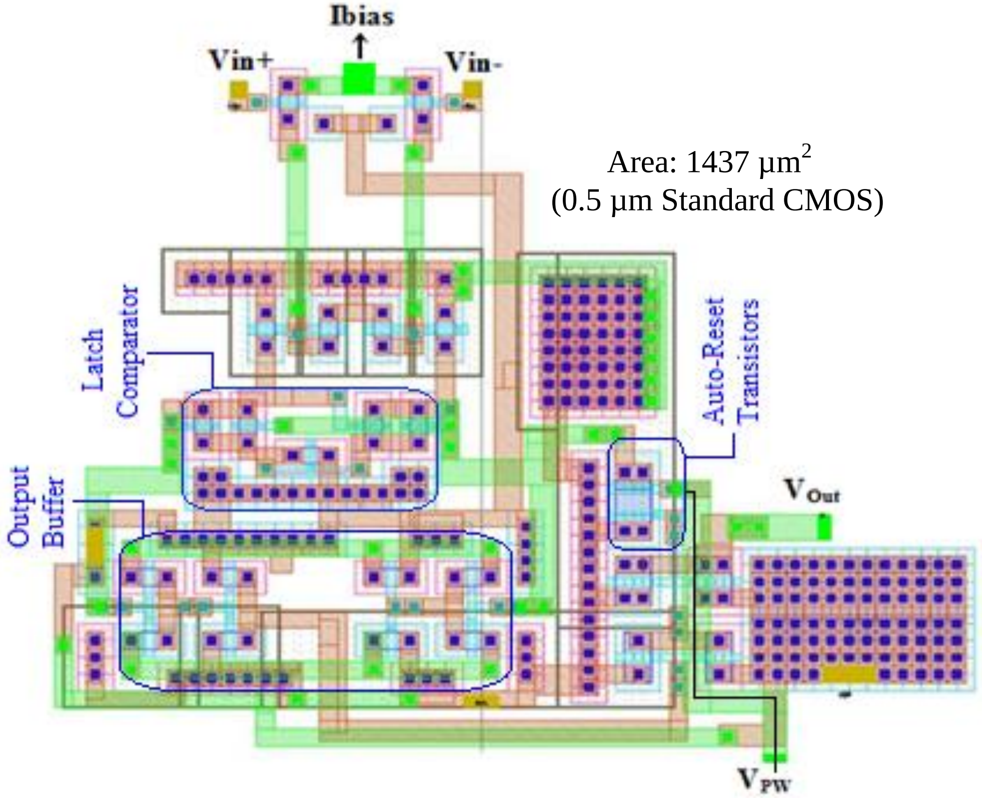


Figure 6.8: Layout diagram of the High Speed Comparator, highlighting the compact electronic footprint of the module.

The pulse stream is subsequently processed by a digital counter and converted into a binary count value. In this architecture noise discrimination can be performed efficiently in the analog domain, directly on the raw detector signal rather than at the end of the signal processing chain as is the case with the analog SiPM readout architecture. If the discrete analog waveform on the main interface is less than the Adaptive reference voltage the HSPC unit remains in the quiescent mode and no signaling activity takes place. The adaptive reference voltage signal is represented by the DC bias value of the DSPA module and can be externally set and actively adjusted. A serial pulse stream is only generated when the optical activity, within a set time interval, at the main interface exceeds this reference value. As a result the ancillary electronics is not triggered by noise thus yielding lower power consumption and greater signal throughput.

6.2.2 Discrete-Signal Pulse Adder (DSPA)

The Discrete-signal Pulse Adder monitors the output pulse stream of the HSPC unit by tracking the voltage level at the positive terminal of the HSPC which is connected to the detector interface. The schematic diagram of the DSPA module is shown in Fig. 6.12. The coupling nodes labeled ‘a’ through ‘e’ on Fig. 6.12 are driven high successively in response to the incoming pulse stream. Since all the coupling capacitors are of equal size (within the mismatch parameter of the technology process) each incoming pulse induces approximately the same step rise on the reference floating node.

The rising edge of each incoming pulses successively activates the next available coupling port, at the same time the falling edge is blocked from propagating down to the coupling node and deactivating it**.** In Fig. 6.9 the first rising transition of the input pulse

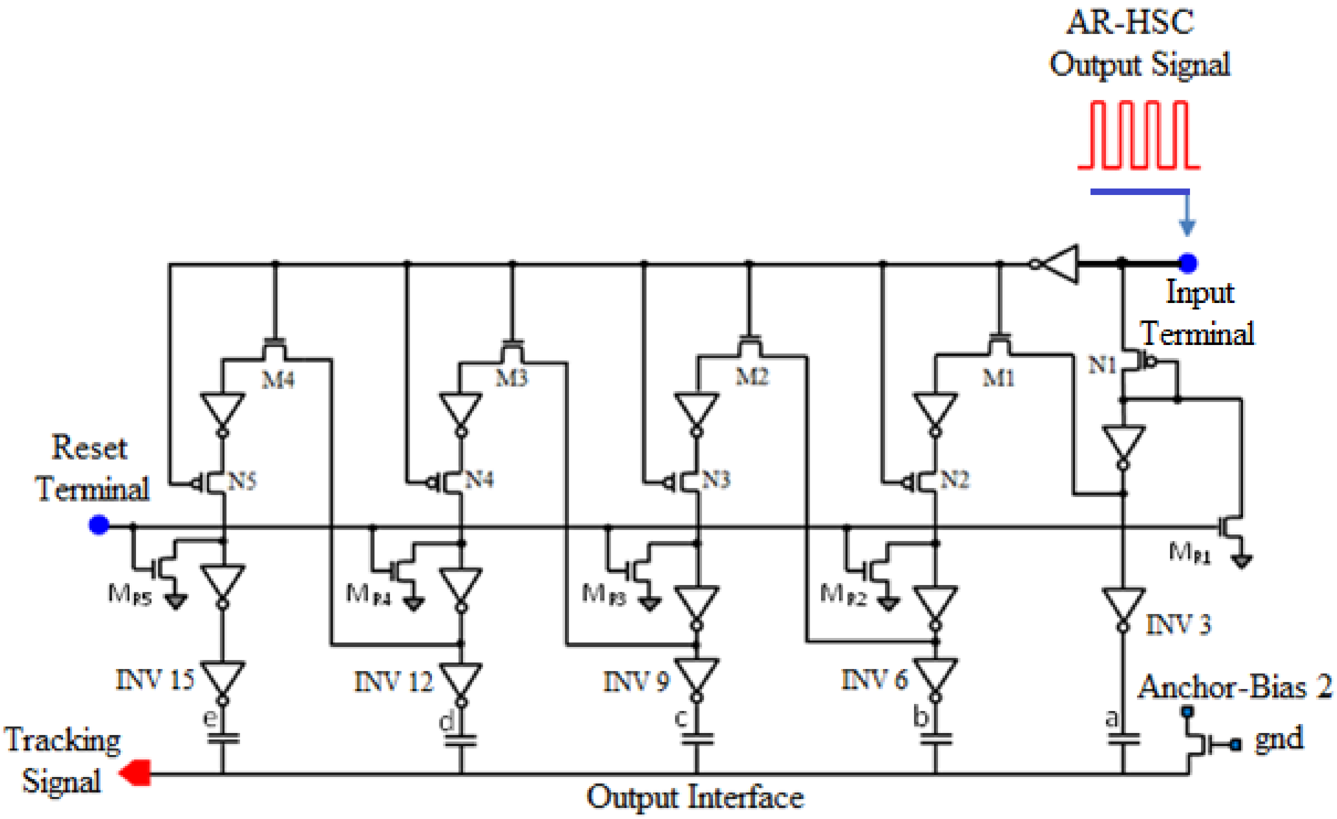


Figure 6.9: Schematic diagram of the Discrete-Signal Pulse Adder.

stream pulls the first internal coupling node ‘*a’* high while the Diode-connected transistor N1 isolates the node from the subsequent falling edge of the associated pulse. This ensures that, once activated, node ‘a’ will remain high until the activation of the Reset terminal. Transistor M1 and N2 (routing transistors for node ‘b’) ensure that node ‘b’ is driven high at the next rising transition of the Input signal following the activation of node ‘*a’*, not sooner or later. Transistors M2 and N3 do the same thing for node ‘c’ and so on. The successive incrementation of the tracking signal continues in this way until its magnitude exceeds the magnitude of the voltage signal on the detector output interface. At that point, the input pulse stream stops and the Reset signal is initiated, deactivating the internal coupling ports in preparation for the next extraction event.

A more compact version of the DSPA module that is driven by the Least Significant Bit (LSB) signal from the digital counter, rather than the HSPC output pulse stream, is presented in Fig. 6.10. Since each signal transition of the digital counter LSB signal corresponds to a distinct event, the DSPA output voltage, in this configuration, must increment on every transition of the input signal.

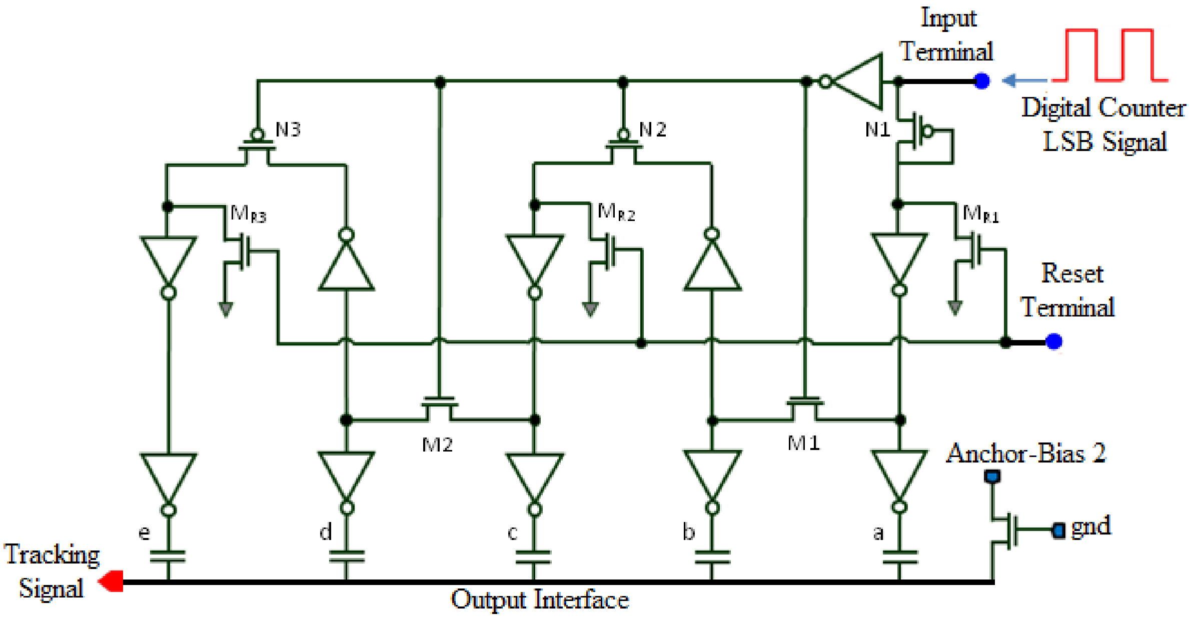


Figure 6.10: Schematic diagram of the edge-driven Discrete-Signal Pulse Adder featuring improved operational speed and a more compact structure.

The layout diagram for a 10 count edge-driven SPA module appears in Fig. 6.11. The edge driven DSPA uses half as many gating transistors (N1, N2, M2, M3) for routing hence it presents less loading on the transitioning input signal. It also utilizes fewer reset transistors. Consequently there is also less loading on the control signal which manages the reset function of the DSPA unit.

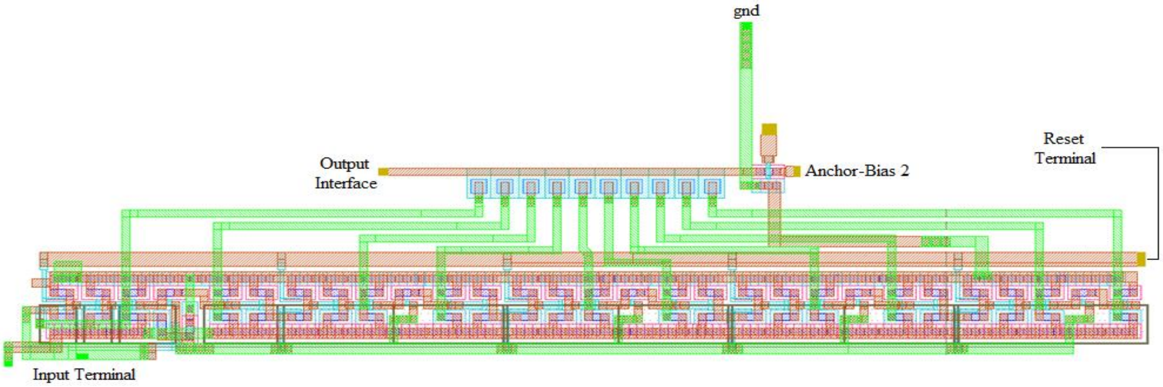


Figure 6.11: Layout diagram of the 10-Count pulse-driven Discrete-Signal Pulse Adder

6.2.3 SPAD Signal Path Configuration

As discussed earlier, successful digital extraction of the count is contingent upon preventing amplitude level drops at the interface due to falling edge of the event pulses in the counting widow. One approach is to hold the signal directly at the floating output node of the detector interface using a transistor switch in series with a storage capacitor, typically used as a part of track and hold circuitry for peak detection in standard ADC system. However, this approach suffers from inherent non-ideal effects such as capacitive feed-through and charge injection which can limit the performance of the system. This is particularly a problem when dealing with sensitive analog signals that require precise measurement [175, 176].

The proposed alternative approach exploits the digital nature of the raw input signal (the TTL output pulse of the triggered SPAD devices), by performing the gating function in the digital regime (directly on the signal path of individual SPAD elements) rather than the discrete-analog signal at the detector interface. This means that after a detection event has triggered the count extraction process, the falling edge of all count-contributing input pulses must be prevented from reaching their corresponding coupling port until the count extraction is complete. It is important that falling-edge filtering is only initiated after a detection event, otherwise events that do not fall within the counting window (noise events) will be counted as well and no noise filtering would be possible. The simulation of the detector interface signal, using in-path gating of the SPAD signal, for a five element SPAD cluster is shown in Fig. 6.12. The simulation was run for two

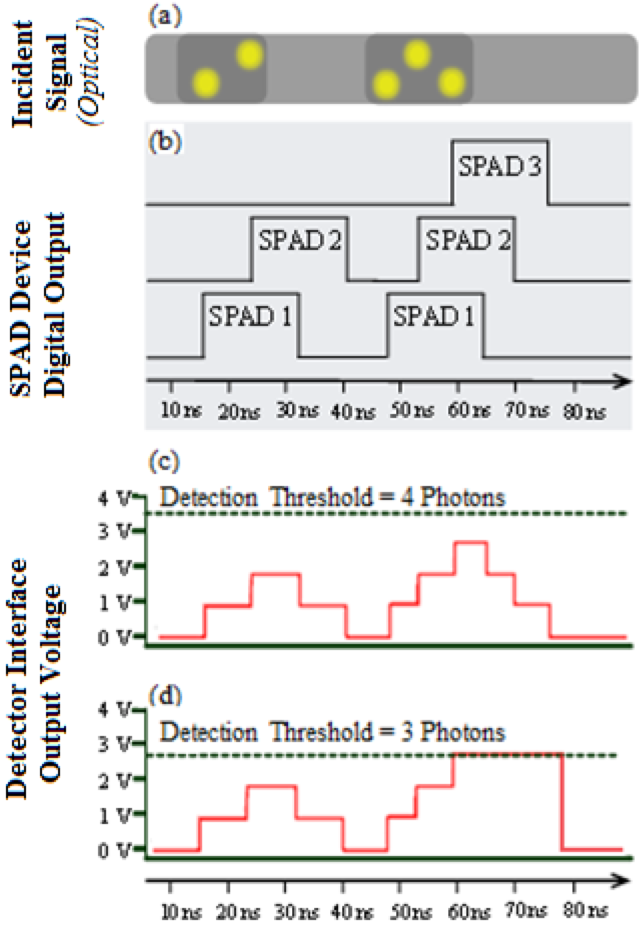


Figure 6.12: Simulation of the interface signal response. (a) Optical input stimuli used for the simulation. (b) SPAD device digital output. (c) Interface response for a 4-photon detection threshold. (d) Interface response for a 3-photon detection threshold.

different detection thresholds to demonstrate the effect of in-path gating on the interface signal. The SPAD model used in the simulation is described in Appendix A.

The incident photon signal, shown in Fig. 6.12a is applied to the simulation model as short voltage spikes which trigger the model response. The important characteristics of the optical signal are the number and relative timing of the individual photons. As can be observed from the simulation results in-path gating of individual event pulses accomplishes the intended function without distorting the interface signal. The in-path gating can be realized by simply adding a switching transistor along the path connecting the SPAD output to its dedicated coupling port as shown in Fig. 6.13.

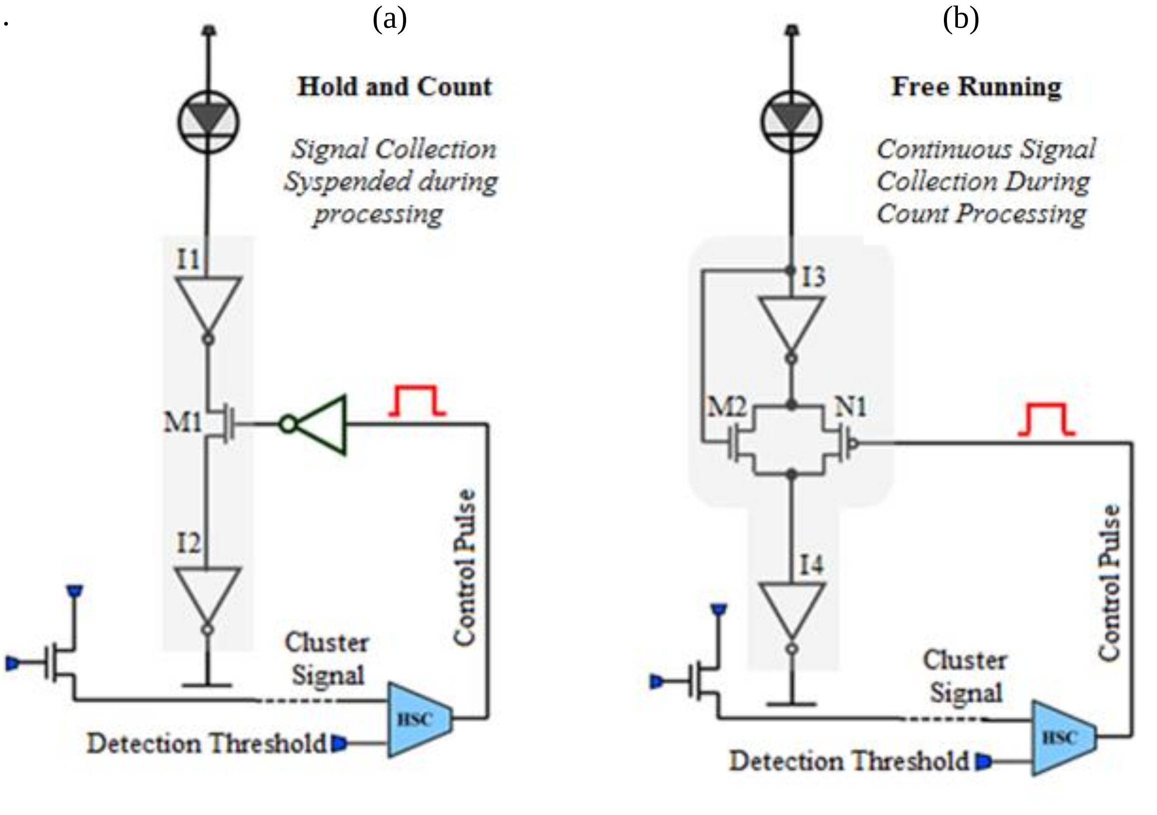


Figure 6.13: SPAD Signal path electronics. (a) Hold and Count Mode electronics. (b) Free Running Mode electronics

In Fig 6.13(a) transistor M1 acts as a switch driven by the control pulse (also referred to as the Count Pulse in Fig. 6.3). The inverter, I2 isolates the coupling node from the switching action on the gate of M1. In the Hold and Count mode, shown in Fig. 6.13a, in addition to the falling edge of event pulses within the counting window, the output pulse from SPAD triggered after the start of the count cycle are also blocked until the current counting cycle is complete. Therefore, in the Hold and Count configuration no activity is allowed on the interface during the hold phase.

The main limitation of the Hold and Count format in Fig. 6.13a is that the dead time following each detection event applies to all the SPAD elements of the cluster. This configuration is sub-optimal for resolution of extended light pulses or continuous photon counting applications. Similar to the analog SiPM detectors, this implementation is best suited for Photon number resolution tasks involving high intensity, short pulses of light.

The signal path configuration presented in Fig 6.13b eliminates the dead time associated with the counting process. In this circuit transistor M2 automatically blocks the falling edge of the SPAD pulse, while always allowing the rising edge to propagate down the signal path to the input coupling node. A secondary path through N1 remains open to the coupling port for both rising and falling pulse edge, before the count cycle is triggered and count pule is initiated (detection threshold not met). The N1 path enables the falling edge, associated with a SPAD output pulse, to extinguish signals that are below the triggering threshold (associated with non-detection events). The rising signal transitions from SPAD element triggered during a counting cycle continue to pass through M2 and dynamically increment the interface signal during the count extraction process. In the Free Running mode the digitization process continuously tracks the dynamically varying interface signal in order to update the digital count extraction process with any incoming events. The simulation results in Fig 6.14 are generated for a 5 element cluster using burst mode and continuous input stimuli.

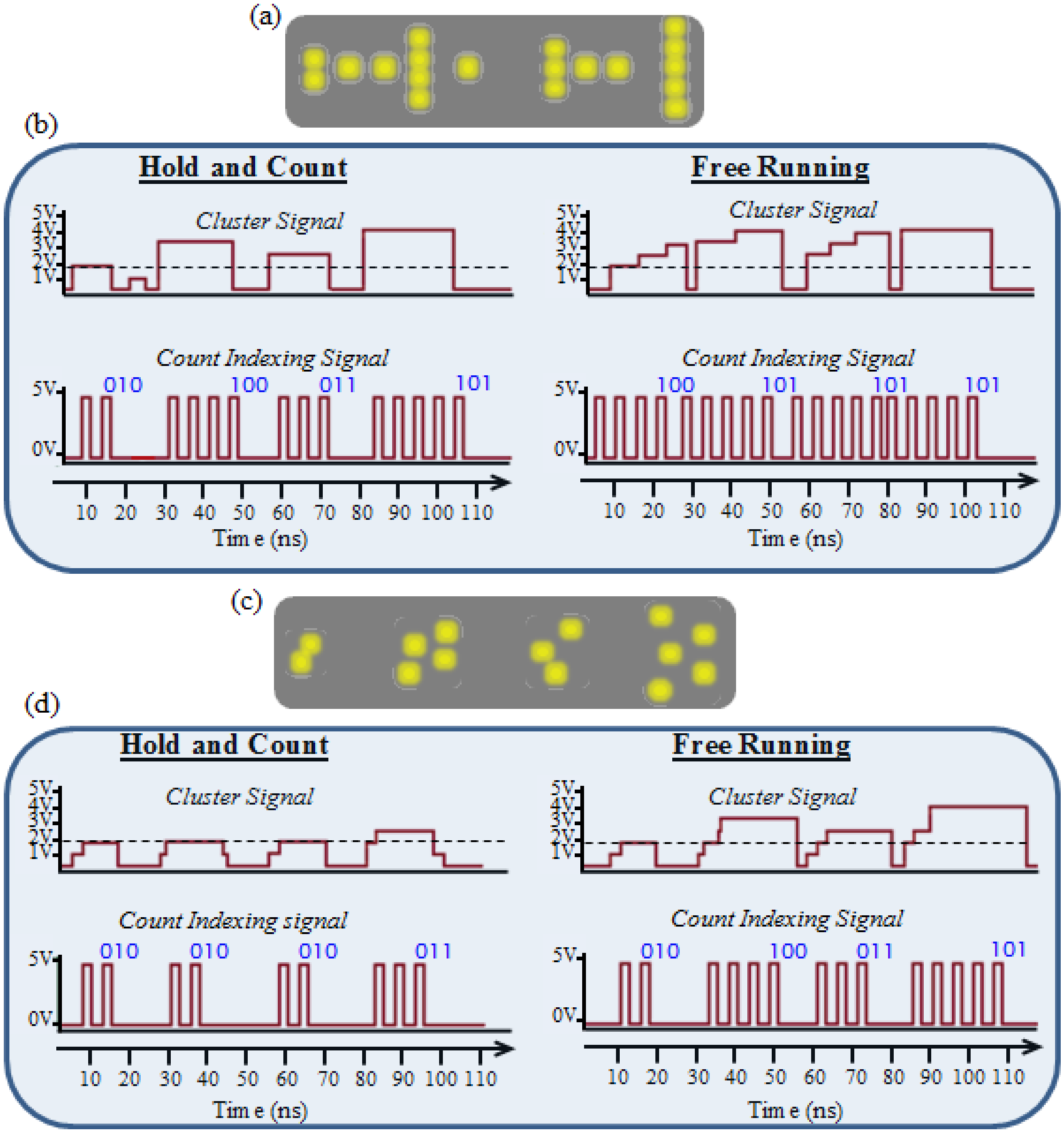


Figure 6.14: Simulation results of primary internal signals. (a) Optical stimuli representing short multi-photon optical pulses. (b) Internal system response to short multi-photon stimuli profile under both signal path configuration. (c) Optical stimuli representing longer optical pulses or a continuous photon stream. (d) Internal system response to continuous optical stimuli.

As illustrated in Fig. 6.14, the Free Running configuration results in a more accurate representation of the photon count data as it involves no counting dead time. However, for certain optical profiles represented by Fig 6.14a, the increased counting accuracy of the Free-Running mode might inversely affect the SNR. In general this applies to detection scenarios characterized by short burst of multi-photon packets where any subsequent solitary events are regarded as noise or background signal. Photon number characterization for an atomic emission source and detection of scintillation photons are examples of applications that can benefit from the Hold and Count mode, while for photon counting applications, the Free-Running mode is the most the optimal choice. An adjustable-mode signal path design, is shown in Fig 6.15.

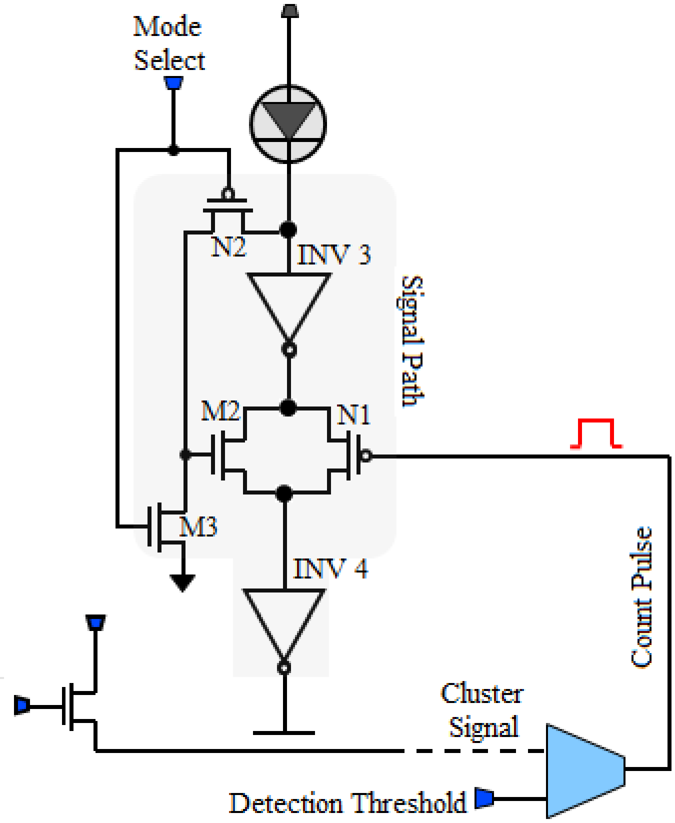


Figure 6.15: The schematics for the Configurable signal path electronics

6.2.4 Highly Compact Asynchronous Digital Counter

Internally the photon count is represented by a serial pulse stream equal in number to the parallel pulse output of the SPAD cluster. For external digital representation, the pulse sequence must be processed with a digital counter and coded into a binary word for storage. The internal parallel to serial conversion process enables constraint free resource sharing among pixels. In this case a single digital counter is used to accurately process the count from all SPADS in the cluster. An integrated digital counter represents the main external interface of the readout system. Standard digital counters are large devices typically involving up to 46 transistors per bit [178]. In a fully integrated system where fill factor is a design constraint, a more compact realization would be highly desirable. The signal routing circuitry utilized in DSPA may be modified alternate between high and low output transitions according to the transition pattern associated with binary counting. The schematic diagram for a two bit compact digital counter, realized in this way, is shown in Fig, 6.16.

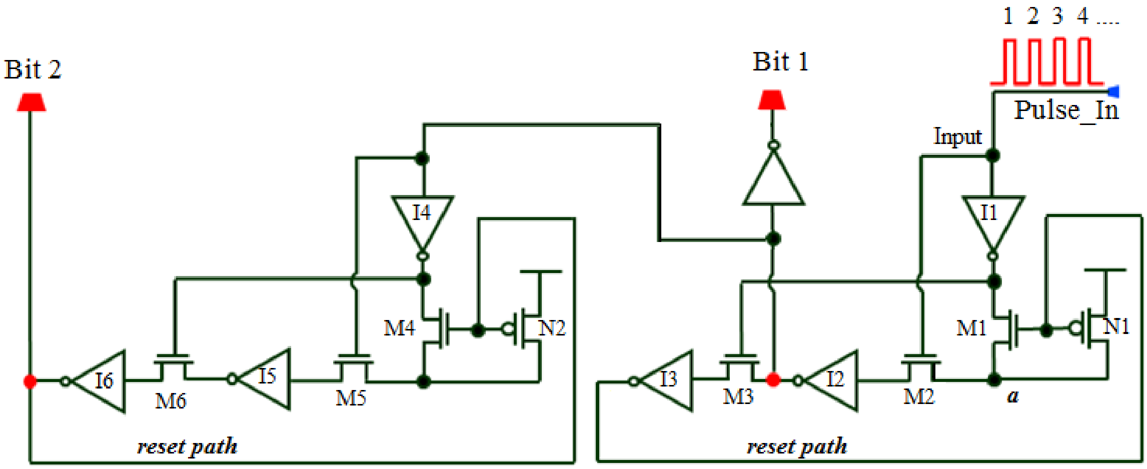


Figure 6.16: Schematic diagram showing the compact structure for a 2-bit Digital counter

The output of Bit 1 switches at every positive transition of signal at the input. When the input is low Bit 1output is low and N1 is off while M1 is on. Signal rise at the Input node turns M2 on while driving node ‘a’ low and turning M3 off to prevent the activation of the reset path. Consequently Bit 1 output goes high to register the event arrival. The subsequent falling edge associated with pulse 1 turns M2 off in order to prevent the signal transition at node ‘a’ from affecting the output of Bit 1. Immediately following the deactivation of M2, M3 is turned on. This action will active the N1transistor and pull the voltage on node’ *a*’ high, thus arming the reset signal. The next rising transition at the Input node activates M2, allowing the voltage at node ‘a’ to propagate through and reset the output of Bit 1 back to zero, indicating the arrival of pulse 2. The input of the Bit 2 circuit is driven by the inverse of the Bit 1 output signal. In this way resetting of Bit 1 coincides with the activation of Bit 2 and so on all the way to the Most Significant Bit (MSB) of the counter. When all the Bit outputs are high and the maximum count is reached the next incoming event will cause the counter to self-reset, returning the count back to zero. As previously described, the readout system is capable of sequential photon counting and parallel multi-photon detection. Both modes are simultaneously active. This is made possible due to the internal serial encoding of parallel input by the system. It is essential for the internal sequential processing of parallel events to remain transparent at the bit output of the counter. In this way the readout of the counter will accurately represent both the single-photon and the multi-photon events as they are dynamically represented on the main detector interface, while avoiding excessive power dissipation from unnecessary output signal transitions. In order to achieve this, the digital count readout should be gated by the pulse output of the HSC module. The schematic of a 3-bit latched counter with dynamic reset appears in Fig. 6.17. The layout for the 3-bit counter implemented in 0.5µ CMOS process is shown in Fig. 6.18.

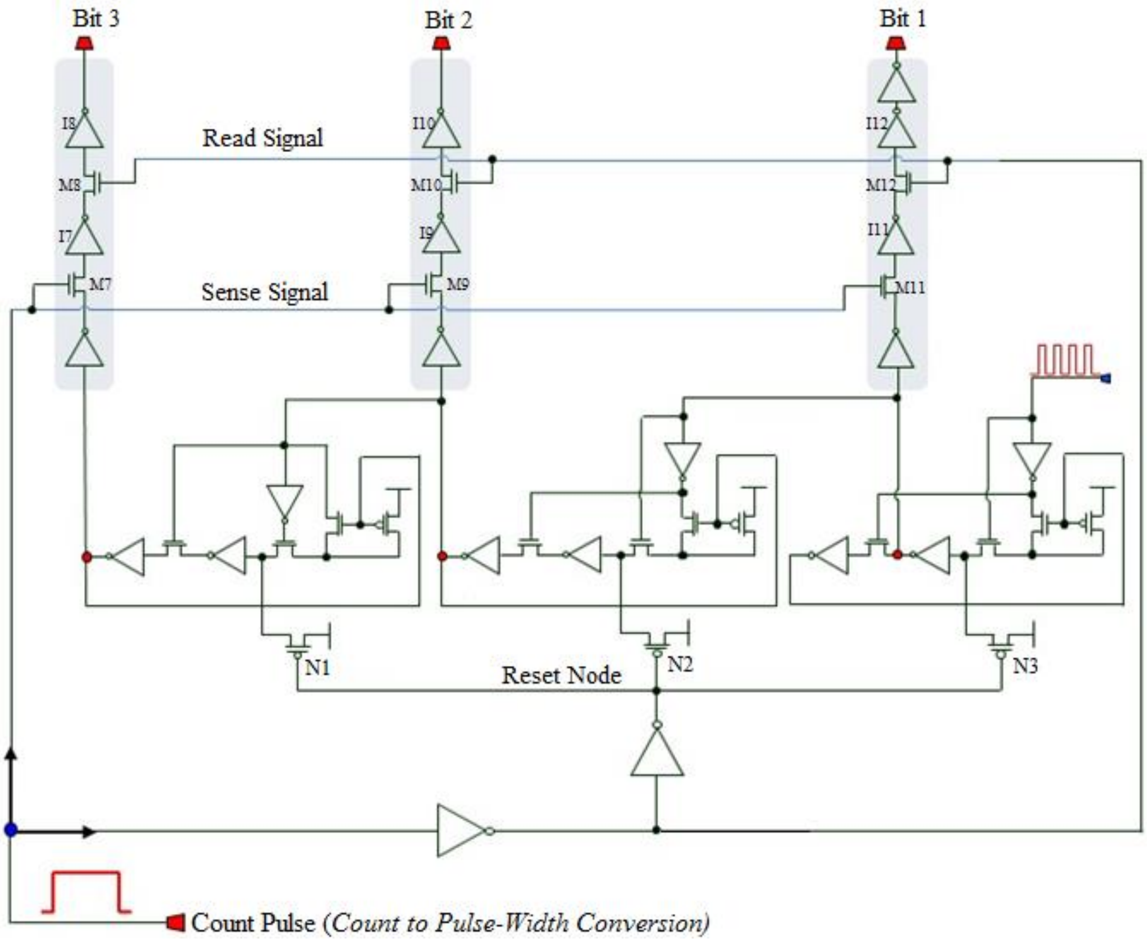


Figure 6.17: 3-bit compact digital counter with active reset.

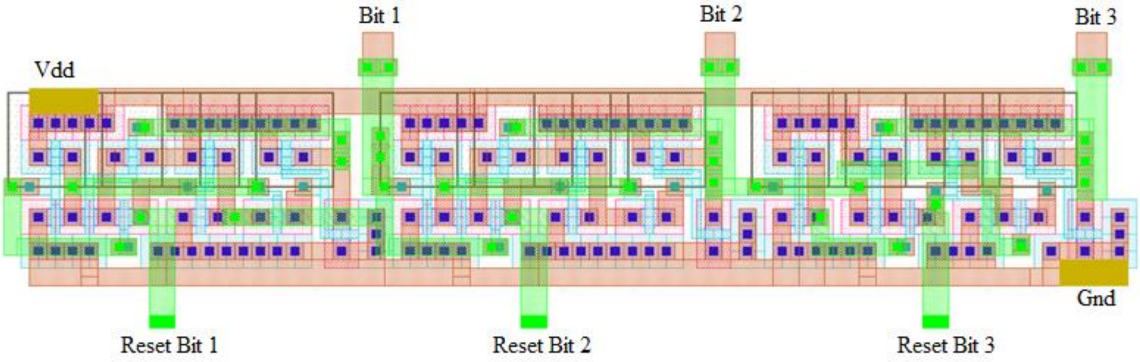


Figure 6.18: Layout diagram of the 3-bit compact digital counter with active reset

The count pulse indicates the start and end for each sequential-processed counting cycle. If the readout and reset signals of the counter are gated by the count signal the counter will latch the extracted binary count at the output before resetting the bit signals. In this way the digital output represents the number of events within each subsequent measurement window at the end of its respective counting cycle.

6.3 Integrated Readout System

The schematic and layout diagram in Fig 6.19 and Fig. 6.20 represent a readout system with gated read and reset, along with SPAD signal path connection to each input coupling port for a 5-SPAD cluster.

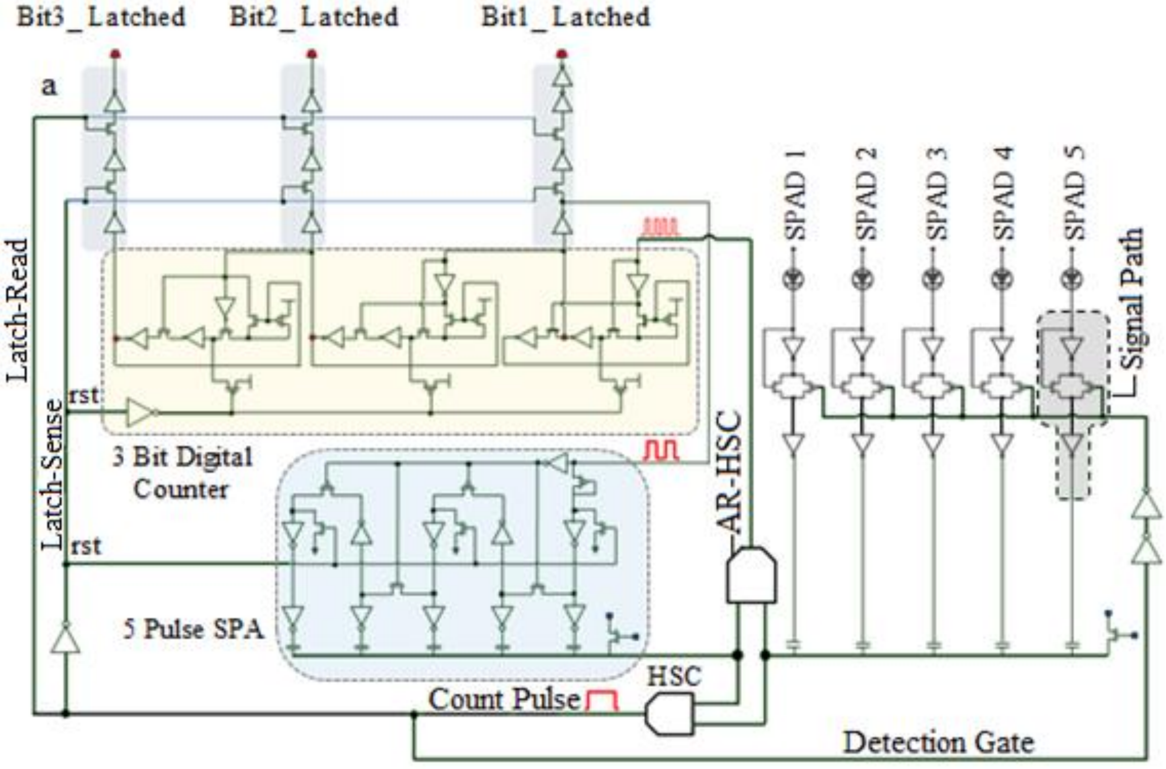
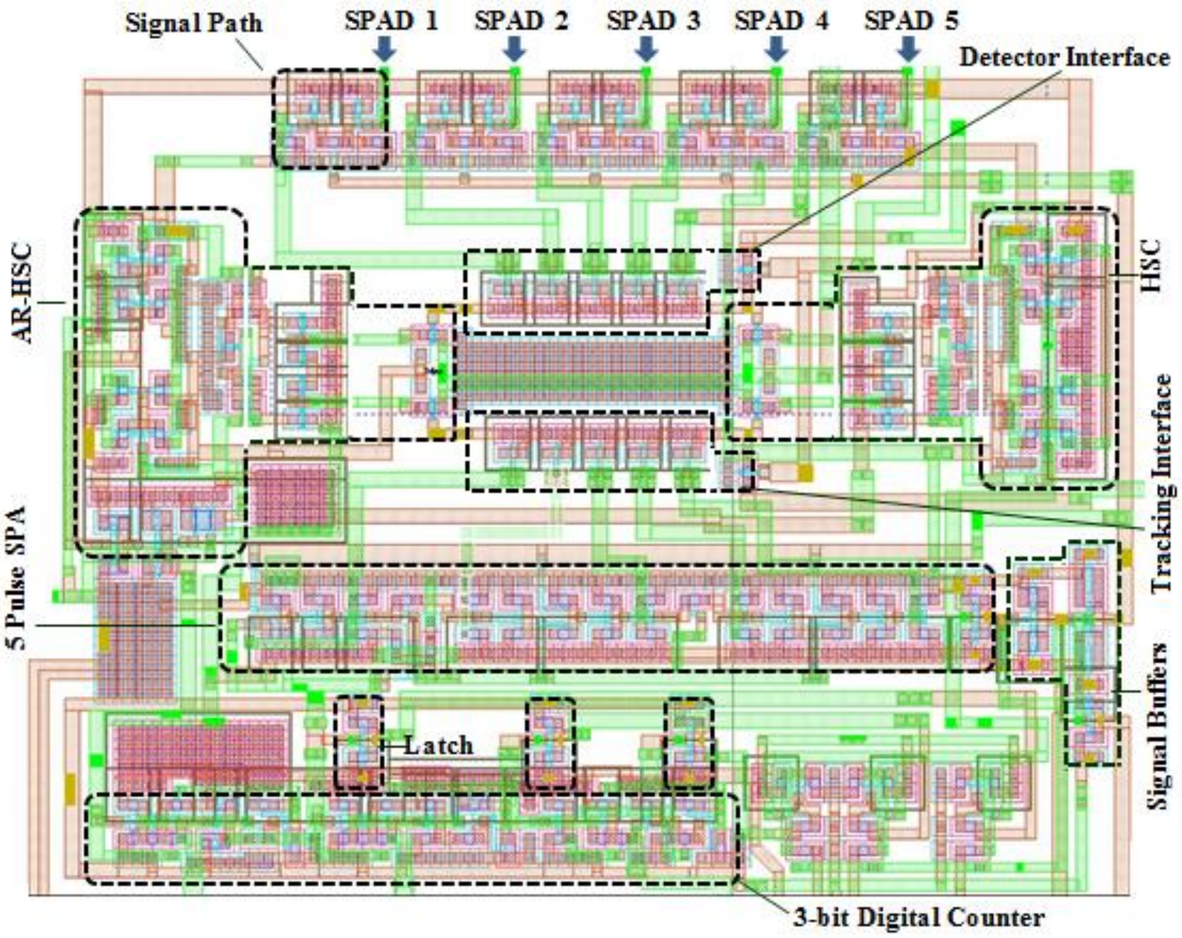


Figure 6.19: The schematic diagram of the readout system using free-running architecture



6.20: Layout diagram of the full readout architecture for a 5-SPAD cluster

The post layout simulation of the primary internal and external signals of the readout system with gated readout and active reset features, along with the optical profile for the input simulation stimuli used for the simulation is shown in Fig 6.21.

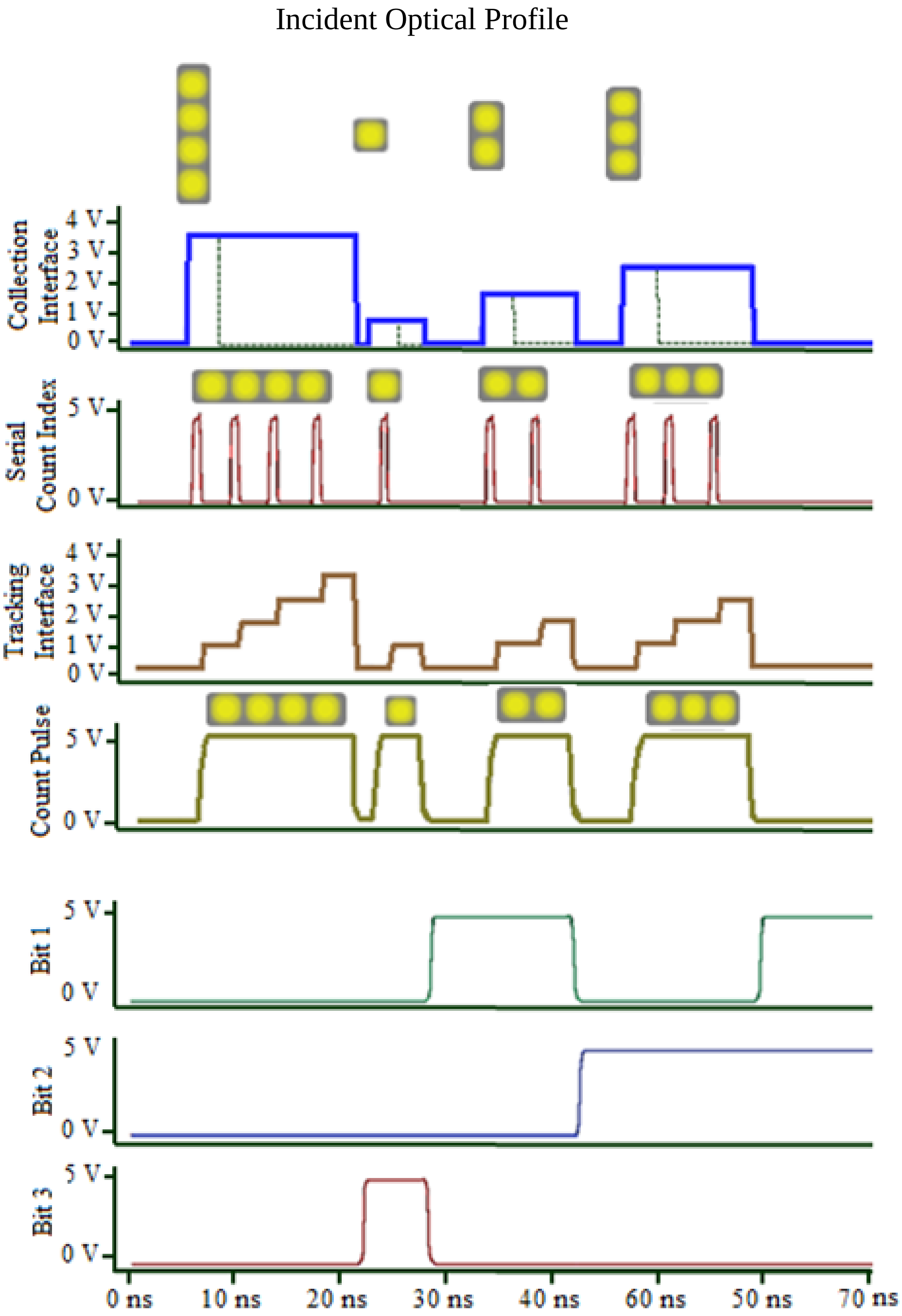


Figure 6.21: Post-layout simulation results showing the primary internal and the external signals

6.3.1 Experimental Results – *Proof Of Concept*

As previously discussed, the amplitude of the interface signal (magnitude of the detection event) can be represented by the pulse width of the count pulse signal. This is a fundamental system feature. In order to asses this functionality a test version of the system with the detector interface replaced with an Input/output pad connection to facilitate externally generated interface signal, was designed and tested. The layout diagram and the measured input/output waveforms for the test unit appear in Fig. 6.22.

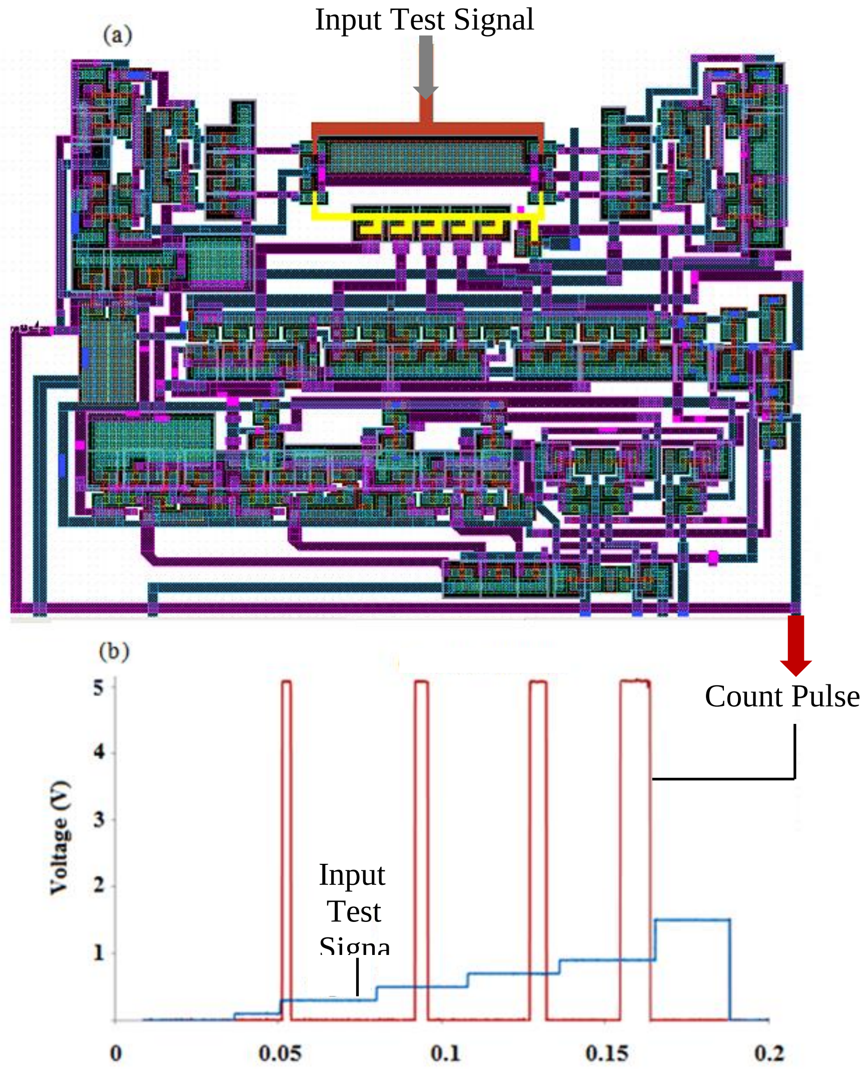


Figure 6.22: (a) Proof of concept test structure. (b) Measured waveforms for the user-defined input test signal and the generated count pulse response of the test system.

The input and output terminals of the structure are highlighted in the layout diagram in Fig. 6.22(a). Fig. 6.22 (b) shows the measured count pulse waveform generated in response to the test signal. As can be observed from the measurements the width of the count pulse is a function of the interface signal amplitude. An important observation is the significant drop in the maximum range of the amplitude signal in the measured data. This is due, in a large part, to the parasitic bottom plate capacitance at the floating node. Since the aggregate cluster signal is generated on the shared bottom plate of the capacitor-based sense interface, the large parasitic capacitance between the Nwell and the substrate (in case of MOSCAP based interface) will impact both the linearity and the dynamic range of the interface signal [175, 176]. Several techniques such as geometrical manipulation of the layout structure, bootstrapping and bottom-plate shielding can used to reduce the bottom plate parasitic effect [177].

Another selected criterion for proof of concept demonstration is the effect of SPAD count rate on the activity level of the cluster. Since the count signal represents the combined real time activity of all the SPADs within the cluster, increasing the count rate of each associated SPAD pixel should increase the probability of overlap among independent output pulses, resulting in greater arrival rate of detection event. A five input readout system in free running mode coupled to cluster of 5 passive SPADs was used to test the cluster response as a function of pixel activity rate. This test structure is shown in Fig. 6.23. For testing purposes, the SPAD activity rate was modified using the field gate bias setting. Multiple measurements of the count pulse signal in 5ms time window were taken. A representative 5 ms capture waveform for the count pule signal at different gate bias values is shown in Fig. 6.24.

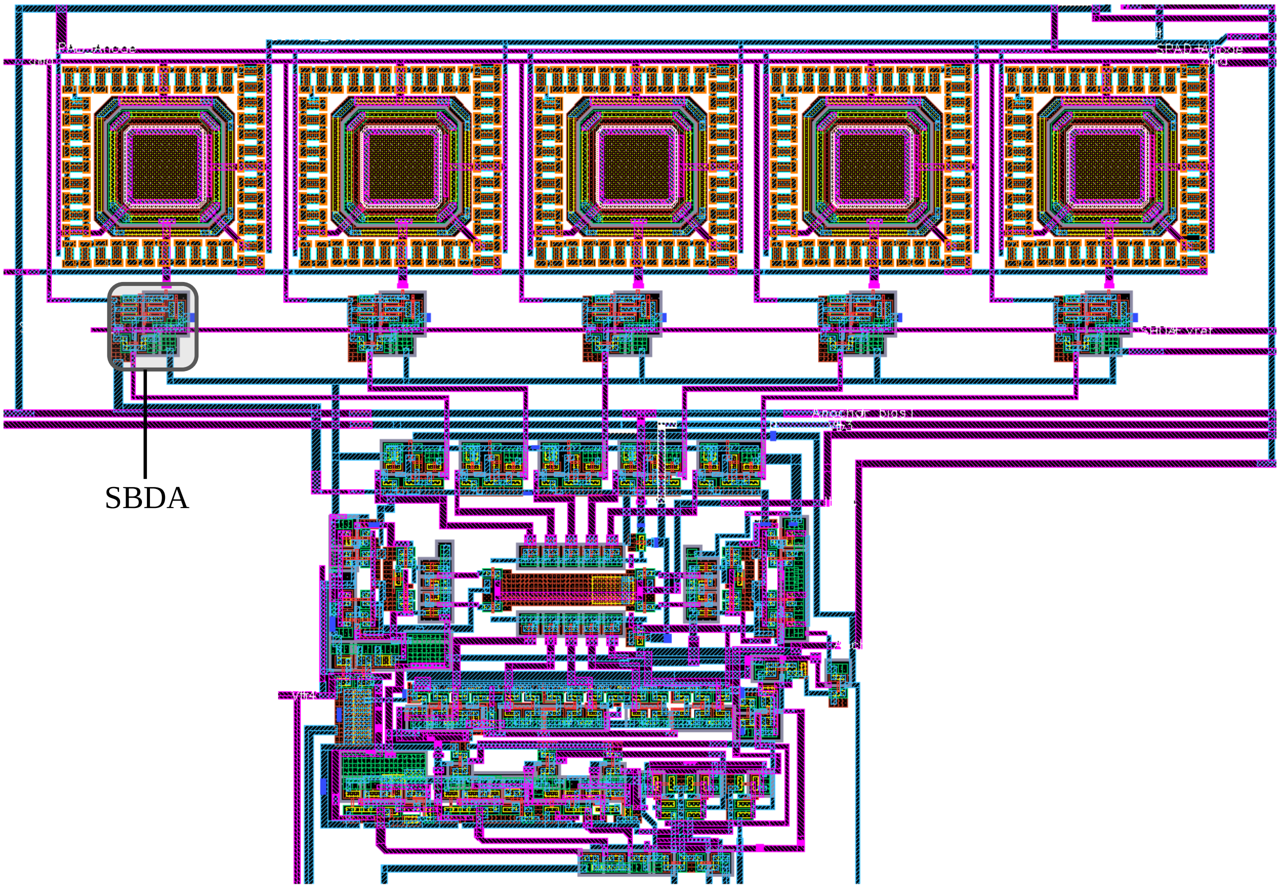


Figure 6.23: Test Structure showing a full cluster system for 5 passively quenched SPADs

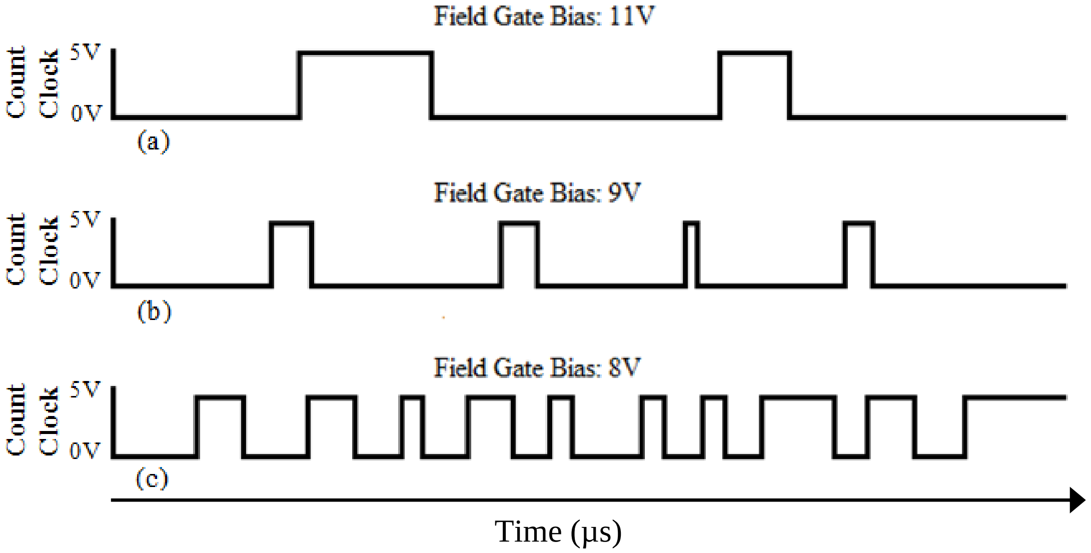


Figure 6.24: Representative snapshot of count pulse signal captured for Proof Of Concept demonstration (a) Count signal at gate bias of 11V corresponding to low overlap probability hence fewer detection events.(b) Cluster response at gate bias of 9V showing increased detection activity (c) Lowest gate bias leading to highest measured activity rate

A Self-Biased Differential Amplifier (SBDA) was used as the digitization interface for the passively quenched SPAD pixel in Fig. 6.23. The bias voltage on the reference input of the SBDA controls the width of the output TTL pulse produced by the SPAD pixel, such that larger reference bias values (Vref) results in a wider pixel output pulse and hence a larger aggregation window at the cluster level. A longer integration/summation interval will accommodate a greater number of arrival events which will lead to longer counting cycles (sequential count processing). The duration of the counting cycle is represented by the width of the count pulse, therefor increasing the magnitude of Vref at the SPAD pixel level should reflect in the pulse width of the count pulse signal at the cluster level. Fig. 6.25 shows the average pulse width of the representative count pulse as a function of Vref. The representative count pulse is determined based on the highest rate of occurrence observed in ten consecutive measurement windows of 5 ms duration.

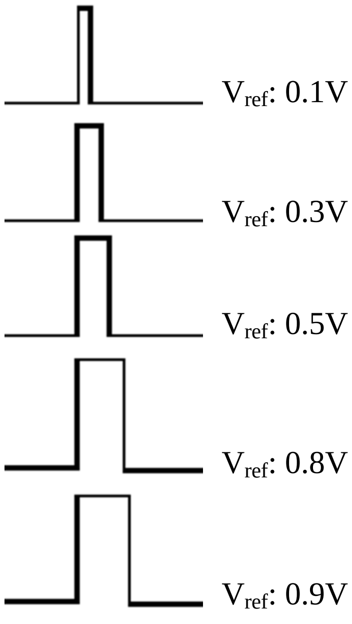


Figure 6.25: Count pulse for various integration intervals determined by the value of the Vref signal.

Since the count pulse signal can dynamically represent, with a single waveform, the operational response of the system in response to both single photon and multi photon events, it is the most convenient and useful measurement index for proof of concept demonstration. However the final output of the system, ready for final storage in digital memory, is provided by the parallel bit outputs of the digital counter. The digital readout functionality is examined by verifying bit signal transition relative to the edges of the count pulse signal. Since the bit readout of the digital counter is latched only at the end of each counting cycle, bit signal transitions should only follow the falling edge of count pulse signal. This has been shown to be the case for the represented measurements, simultaneously recorded for the count pulse and Bit 1 signal, over three consecutive 20 ms time window and shown in Fig. 6.25.

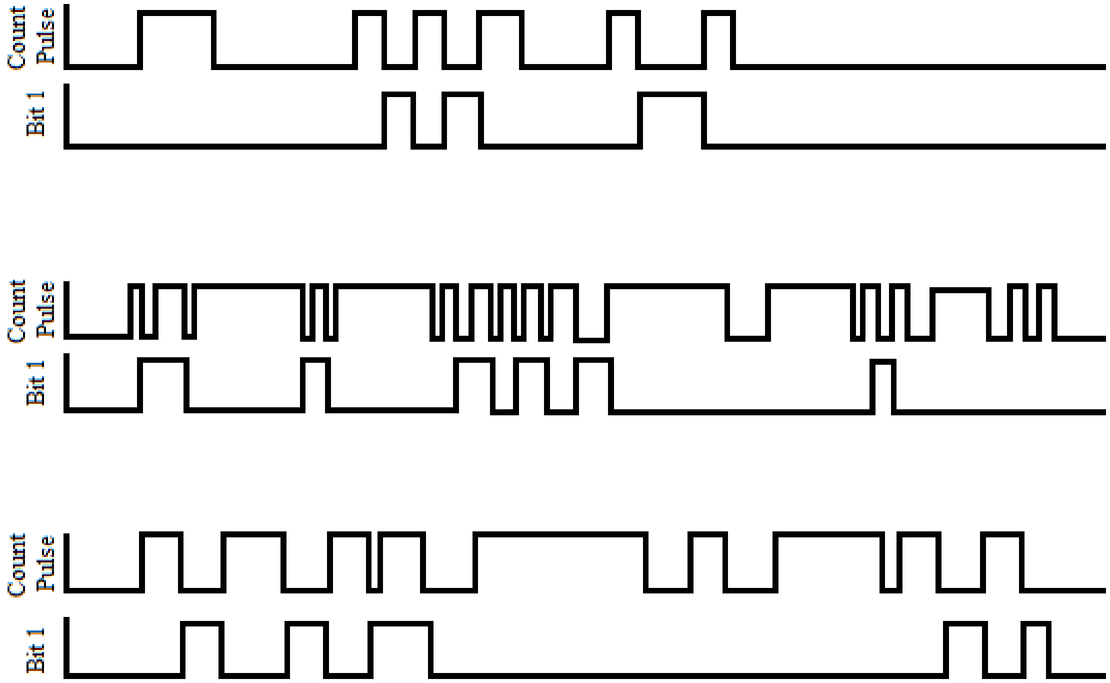


Figure 6.26: Digital decoding of the pulse-width modulated count signal. The waveform trace represents the count pulse and the Least Significant Bit transition respectively, at the end of each counting cycle as implemented by the SPAD-cluster readout system shown in Fig. 6.23.

Dynamic transitions of Bit 1 and Bit 2 are shown in Fig. 6.27. The measured waveform, captured over 20 ms measurement window, represents the first two bits of the photon count value. These measurements are limited in scope and meant only for functional verification of the system possible with the available 2-channel signal acquisition board.

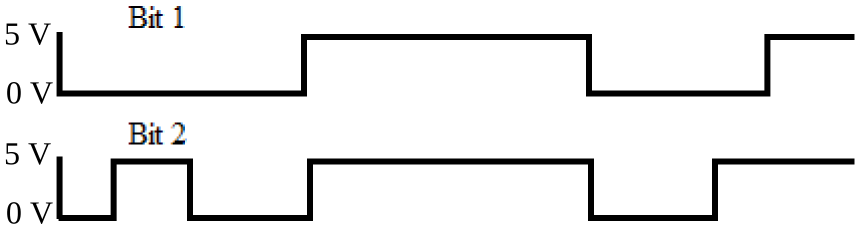


Figure 6.27: Digital representation of photon count associated with an optical event, determined through simultaneous measurement of the latched Bit 1 and Bit 2 signal.

6.4 Summary

Existing ADC-based digitization systems, used in conjunction with single-photon analog detectors, do not digitally encode the photon count information directly. Instead, they generated a digitally coded quantity proportional to the photon count. In the last two chapters, a new paradigm for signal processing and readout for high performance single-photon pixel arrays was defined and developed. The novel architecture implemented is capable of direct extraction of digital photon count data from the raw detector signal.

This chapter conceptualized, designed and demonstrated a Signal Digitization Unit (SDU) for digital decoding of the detector interface signal. The digitization methodology introduced is particularly well-suited for the signal collection scheme described in Chapter 5. It is capable of representing both sequential and parallel events.

In the case of multi-photon events, characterized by parallel triggering of multiple SPAD elements, the system internally generates an equal-weighted serial pulse stream and digitally process the serial input.

The precision-related difficulties are inherent in the analog approach taken towards the discrete quantification process which underlies the operation of ADC systems. There should be no analog processing in intensity measurement through a fundamentally discrete process of photon counting. In direct adherence to this performance guideline the proposed system operates in the discrete-signal regime, utilizing a self-generated quantized reference signal for direct conversion of detector response into digital count values. In this way the discrepancy and precision shortcomings associated with the standard signal conversion method of ADC systems is avoided.

The operation of the readout system optimizes power consumption by eliminating the electrical processing of unwanted signals (noise) and excessive signaling activity on the external readout terminals. The minimization of unnecessary electrical activity streamlines the operation of the system and improves its throughput.

Since the signal collection/aggregation interface is part of the readout architecture, the system acquisition mode can be actively controlled by the digitization electronics directly at the individual SPAD signaling path. This has the advantage of enabling active configuration of detection parameters. A mode-select bit can transition the operation of the system from trigger-based mode best suited for discrete detection of short multi-photon packets to a free running photon counting mode capable of continuous digitization on a dynamically incrementing peak signal. This eliminates the dead time and performance bottle-necks associated with digitization and the readout process.

In digital SPAD arrays, current schemes for effective utilization of hardware resources such as counters and TDCs are based on a first-in take–all sharing model. This architecture results in loss of system throughput, resulting in performance trade-of and application limitations for the digital SPAD arrays. The proposed readout system offers a method for multiple SPADs to dynamically share a single counter without introducing additional throughput limitation or performance trade-offs.

Furthermore, speed optimization in SPAD imager arrays in a large part has to do with readout of individual counters. By aggregating the signal, using the scaled summation technique and processing the signal with the proposed signal Digitization Unit, multiple counters can be replaced with a single slightly larger shared counter. This can simplify the readout architecture and improve the readout speed while significantly reducing the occupied silicon footprint.

The integrated digital count extraction electronics is comprised of modules that offer unique performance advantages and can be utilized independently in variety of application. An itemized list of these module and the novel features they implement is provided below.

1) A Comparator configuration that combines the functionality of a high speed comparator with that of a Voltage Controlled Oscillator (VCO) with externally adjustable duty cycle.

2) A Discrete-signal pulse counter that maintains an internal reference signal with a constant rise time for each successive incrementation step. This results in the elimination of the slew rate constraint which is a fundamental limiting factor in analog counters.

3) A novel design for a highly compact digital counter, and a method of generating a pulse-width modulated signal that encodes both the arrival time and the absolute photon count of associated with a detection event. In this way both single photon and multi-photon events can be simultaneously represented using a single output waveform. This functionality is currently missing in both CMOS-based photon counting pixel arrays and PMT-based optical detectors.

Chapter 7

Conclusion

Available design and development strategies for SPAD-based pixel arrays in standard CMOS offer limited optimization possibilities. The existing trade-offs afflicting different operational aspects of the system only allows selective enhancement of performance parameters. A multi-objective optimization strategy for a general single-photon sensing solution in CMOS does not currently exist nor is it being pursued. Current research and development efforts are primary focused towards application specific design and targeted implementation while a general solution is postponed pending the advent of future ultra-compact and versatile process technologies that can alleviated some of the inherent limitation and performance trade-offs. The underlying motivation behind this research is to bridge the existing performance gap in development of large CMOS integrated SPAD arrays by introducing novel design concepts and innovative implementation schemes, at every architectural level from device to system, within the context of existing low-cost, conventional CMOS process.

Towards this purpose, at the device level, the structure of the p-n junction has been modified with a surface field gate and the geometrical profile of the junction altered to increase the perimeter to area ratio. The proposed structure enables a SPAD device with a large detection area and high SNR performance. The enhancing effect of the incorporated features on the fill factor has been established. The resulting improvement on the performance was demonstrated through measurement of the device DCR.

At the pixel level, functional support to enhance performance and utility is achieved using a novel design for the front-end pixel interface that optimizes speed and noise performance of the device while providing an adjustable output interface signal in order to accommodate array level operation. The incorporation of an independent readout interface at the pixel level enables functional flexibility at the system level. In order to ascertain valid interpretation of pixel measurement data in presence of dead time related system nonlinearities, a theoretical detection model was developed and verified against controlled empirical measurements.

Finally at the array level the favorable functional features exclusive to the digital readout architectures and the performance advantages unique to the analog method are brought together across a readout interface that resolves the associated intrinsic implementational conflicts and incorporates key featural advantage from each readout paradigms into unified monolithic readout architecture. An integrated digitization sub-system was also designed and developed for digital information extraction. The operational profile of the digitization unit is customized to the profile of the detector response and the method of signal collection used at the detector interface. The unified system aims to present a complete blueprint for CMOS integrated single-photon sensing and processing platform.

Analog SPAD array readout architecture, used in SiPM detectors, suffers from the complexity and power consumption associated with the required analog signal processing and readout circuitry. This obstacle has limited the applicability of the SiPM system and prevented monolithic system integration of the readout. Exploiting the digital nature of SPAD response is deemed a necessary future enhancement for the SiPM detector array only then can full integration of SPAD detector and readout system be achieved in SiPM. The proposed readout system presents an architectural platform whereby the digital nature of SPAD is fully exploited at the pixel level within a design framework that also exploits the dynamic and asynchronous nature of analog signaling at the array level. This can improve the current SiPM systems by enabling full integration of sensing and processing units towards a true single-chip solution.

Digital SPAD array readouts, used in SPAD image sensors, are based on a first-in take –all sharing scheme. The proposed readout system offers a method for multiple SPADs to dynamically share a single counter without introducing additional throughput limitation. Speed optimization in SPAD imager arrays in a large part has to do with readout of individual counters. By aggregating the signal, using the scaled summation technique and processing the signal with the proposed signal Digitization Unit, multiple counters can be replaced with a single slightly larger shared counter. This can simplify the readout architecture and improve the readout speed while significantly reducing the occupied silicon footprint.

The proposed readout system comprises the first generation design. The empirical measurement and testing performed on the readout system are limited to proof of concept demonstration and identification of areas for targeted future design improvement. Gross verification of system operational has been achieved through qualitative demonstration of system response as a function of key operational parameters. Parasitic capacitance available at the floating sense adversely affects the system operational linearity and range node and has been identified as a major candidate for feature enhancement in later design revision. Next generation design should focus on shielding techniques and more effective layout strategies in implementing the floating node of the detector interface.

APPENDIX A

SPAD Device Modeling

In order to adequately reflect the behavior of a SPAD device, a circuit model has to include functional elements for triggering, self-sustaining and self quenching the avalanche process. The SPAD device has been traditionally modeled as a series combination of a resistor, representing space-charge, neutral and contact resistances in series with a DC voltage source representing the breakdown voltage. The intrinsic and parasitic capacitance of the diode device is represented via a parallel connected capacitor. [64]. A transistor controls the triggering of the avalanche. Fig A.1 shows a schematic of the model.

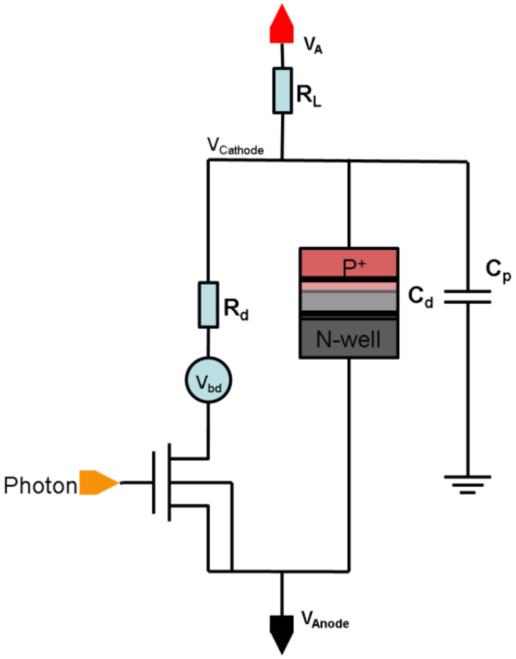


Figure A.1: SPAD device model

In this model, photon arrival is represented by a voltage pulses at the gate of the NMOS. The rising edge of the pulse turns on the NMOS and triggers the avalanche current, subsequently the falling edge of the pulse turns the NMOS off and interrupts the avalanche current. Therefore quenching is controlled by the simulation stimulus photon and is a function of the stimulus pulse width rather than the design. A model proposed to deal with this issue, uses a combination of current and voltage controlled switches to control the triggering of the avalanche [65].A schematic representation of the model appears in Fig A.2.

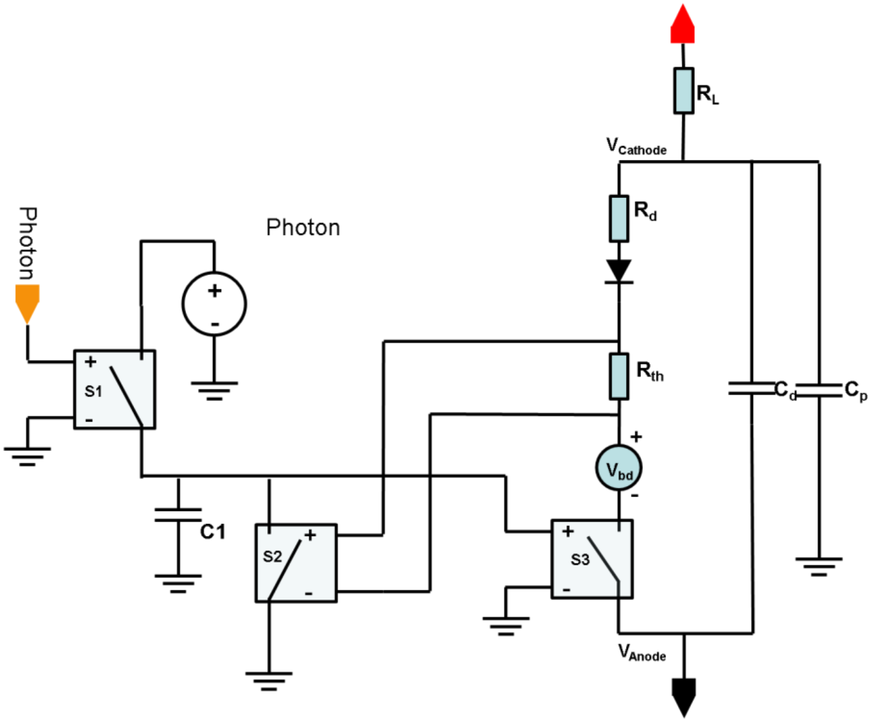
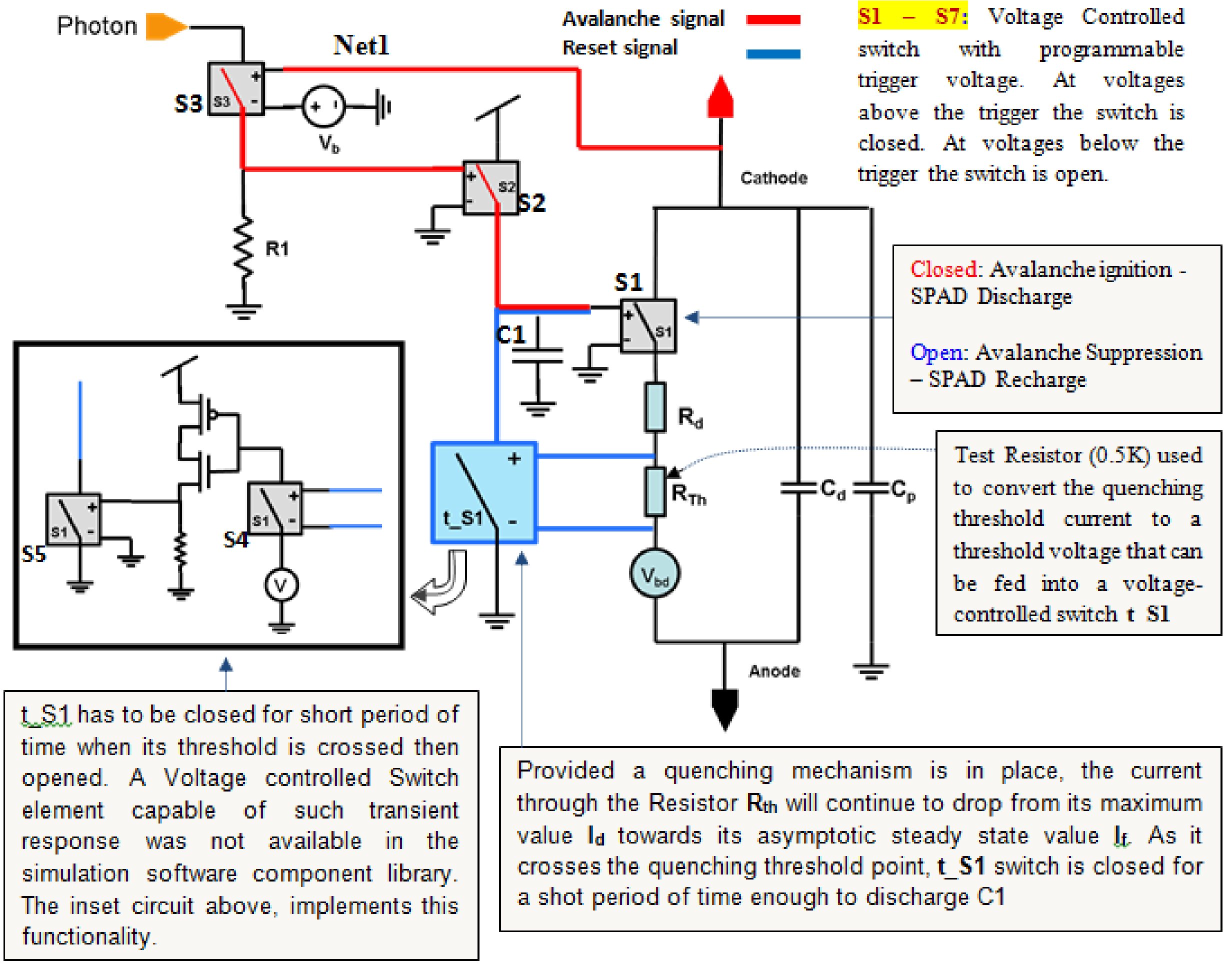


Figure A.2: State of art SPAD device model

In this model, avalanche quenching is dissociated from the simulation photon signal. A capacitor stores and applies the voltage necessary to keep the avalanche switch

S3 on even after switch S1 is opened in synch with the falling edge of the simulation photon pulse. Switch S2, usually open, is set to close when its voltage drop below a threshold. If there is a quenching mechanism in the circuit , the current through the accessory resistor R_{th} drops below the threshold of S2 at a point that can be set through selection of S2 threshold and R_{th} size. This will close S2, discharging C1 and hence, opening S3 which terminates the avalanche. In this model, if a photon pulse arrives at S1 before the diode has been charged to a voltage above its breakdown a reverse avalanche current pulse is triggered. In order to prevent an ignition of reverse avalanche pulse a diode is placed between model terminals. However, the avalanche trigger switch, S3, is still activated despite the sub-avalanche voltage conditions. This does not accurately reflect Geiger mode Avalanche Photodiode behavior. A different model [66] overcomes this problem by controlling the path of the photon pulse by a separate switch, with threshold set to V_{bd}. Therefore, the photon pulse triggers an avalanche only when the appropriate voltage requirement is met. The model then uses a current-controlled switch to sustain and quench the avalanche. Although this is more accurate reflection of SPAD behavior, the implementation uses a current controlled switch which is not available in standard simulation tools. Here we propose a device model for photodiode operation in Geiger mode which does not require modification of the SPAD structure [66] or inclusion of nonstandard circuit component to ensure (correct behavior). The proposed model incorporates avalanche trigger mechanism based on presence of a photon and correct bias condition, utilizing standard voltage controlled switches in conjunction with passive/active components. The signal for ignition and quenching of the avalanche are represented by two separate paths.



S1 – S7: Voltage Controlled switch with programmable trigger voltage. At voltages above the trigger the switch is closed. At voltages below the trigger the switch is open.

Closed: Avalanche ignition - SPAD Discharge

Open: Avalanche Suppression – SPAD Recharge

Test Resistor (0.5K) used to convert the quenching threshold current to a threshold voltage that can be fed into a voltage-controlled switch t S1

t_S1 has to be closed for short period of time when its threshold is crossed then opened. A Voltage controlled Switch element capable of such transient response was not available in the simulation software component library. The inset circuit above, implements this functionality.

Provided a quenching mechanism is in place, the current through the Resistor R_{th} will continue to drop from its maximum value I_d towards its asymptotic steady state value I_f. As it crosses the quenching threshold point, **t_S1** switch is closed for a shot period of time enough to discharge C1

Figure A.3: Proposed SPAD device model with embedded text box describing the functionality of select components. The inset shows the actual circuit implementation of t_S1 component.

Incoming photons are represented by a voltage pulse train. When S3 is closed and a photon is incident (Net1 is high) S2 closes, activating the Avalanche Signal Path and hence closing S1, discharging the SPAD. This indicates avalanche ignition. The falling edge of voltage pulse representing a photon opens S3, thus deactivating S2. However, the capacitor C1 and hence the Avalanche Signal Path remains charged (self-sustaining) until the activation of the Reset Signal Path which discharged C1. The Reset Signal Path should only be activated long enough for C1 to discharge and t_S1 to open. If it does not disengage at that point the positive terminal of the S1 will stay grounded, the switch will stay open, and the SPAD will stay quenched through the subsequent avalanche events. For this reason t_S1, has to disengage automatically, shortly after it is activated, in order to release the signal path. Circuit implementation of t_S1 is shown in the inset of fig 3.3. Momentary shorting of the reset path (activate and release) is accomplished by feeding a trigger signal (voltage v, in the inset) to an inverter and using current pulse of the inverter across a small resistor as the Activate/Release signal for the reset path (blue wire). Current of an inverter rises and drops in response to a high voltage at the input. This provides activate (rising edge) and release (falling edge) signal characteristic required. By setting the NMOS aspect ratio much larger than that of the PMOS, the inverter's current pulse output in response to a low input voltage can be kept below the threshold of S5. Therefore only a high signal indicating that the latch current threshold has been crossed, will enable the quenching operation.

Diode junction capacitance (C_d) is function of over bias voltage V_{EX}. However in situations where the capacitance of the diode is dominated by interconnections and pad capacitance C_d can be kept constant without a loss of accuracy. This is especially applicable for small V_{EX} values. However, the intrinsic series resistance of the diode represented as R_d in the above device models is a function the excess voltage across the multiplication region (V_{Ex}). In almost all device models in the literature this value is set as a constant, usually 1K [65], [66], [67]. The model proposed in [68] accounts for the non-linear I/V characteristic in avalanche mode by programming empirically obtained I/V values of a SPAD device into a voltage source. The voltage source was modeled in

Verilog-A and the was placed in parallel with the capacitors C_d (junction) and C_p (parasitic). The model then compares the diode voltage V (t) with the hardcoded values and appropriately computes a value for R_d at each step. Although accurately modeling the dynamic space-charge behavior of diode, this approach requires pre-testing the device with a source-measure unit and can be used only as a model for the particular device. In the proposed model of fig A.3, the dynamic feature of avalanche operation will be incorporated by using Cadence Spectre bsource functionality. R_d can be modeled as bsource, with its resistance set to appropriate function of the avalanche current, once a representative expression is worked out. It should be noted that for small excess bias voltage V_{EX} and hence small avalanche current, the effect of diminishing avalanche current on R_d has much less of an impact on the performance of the diode. The proposed model in [68] is simulated with a large excess bias. The diode capacitance has been modeled as a function of diode terminal voltage. Verification of the model was carried out by comparing the simulated waveform of the model with the measured waveform of a SPAD test device. The simulations parameters were calibrated to match the bias condition and parasitic loading of the test device, as closely as possible. The simulation input photon stream was built by matching it to the avalanche event time of the measured waveform. In Fig A.4 a, the depletion capacitance of SPAD model was calculated using the test device area, doping profile and the depletion region voltage at quiescent. In Fig. A.4 b, same procedure follows, except that the time-varying value of depletion voltage is wired into the expression for the capacitance. The simulations result closely matches the measured data, particularly when the fluctuating SPAD terminal voltage is taken into account when computing the depletion capacitance.

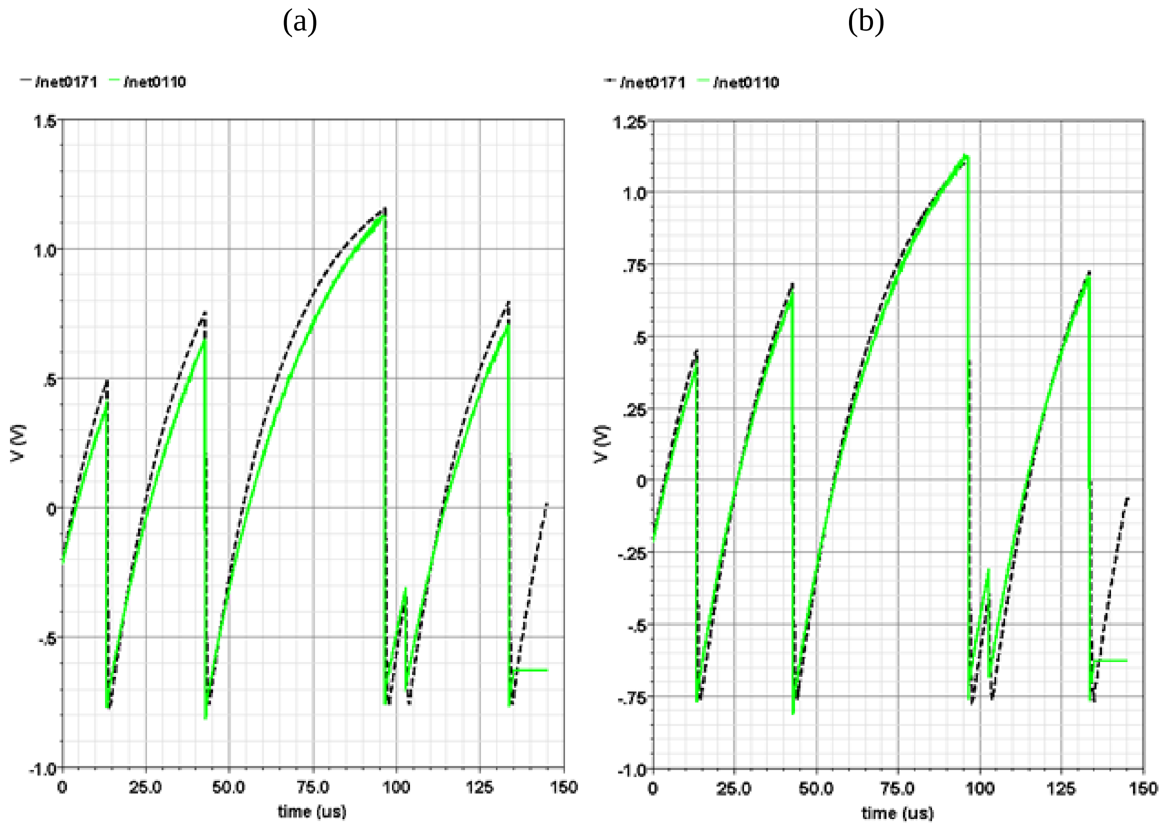


Figure A.4: (a) Solid Green trace represents the measured waveform using data acquisition system. The black trace is the simulation out using the SPAD model in Figure 4.3. Depletion Capacitance was set at its quiescent value. (b) Solid green trace represents the measured waveform using data acquisition system. The black trace is the simulation out using the SPAD model in Figure 4.3. Depletion Capacitance was set to change according to the time varying SPAD terminal voltage (depletion region voltage drop)

Bibliography

[1] E. Charbon, Precision Imaging for Biosensing, Invited paper, EPFL.

[2] Cristiano Niclass, et al., A Single Photon Avalanche Diode Array Fabricated in Deep-Submicron CMOS Technology, , Ecole Polytechnique Proceedings of the conference on Design, automation and test in Europe, 81-86, 2006.

[3] I. Prochazka, et al. SPAD Detector Package for Space Born Applications, CTU technical press.

[4] J. O'Keeffe et al., New Developments in Photon Counting Modules, SensL, 2000

[5] I. Rech et al, Multipixel single-photon avalanche diode array for parallel photon counting applications, Journal of Modern Optics, Vol. 56, Issue 2 & 3, 2009 .

[6] P. Froehlich, Use of Luminescence Spectroscopy in Clinical and Biological Analysis, Applied Spectroscopy Reviews, Vol. 12, Issue 1, 1976.

[7] R. Kohling, et al.et al. Spatio-temporal patterns of neuronal activity: analysis of optical imaging data using geometric shape matching, J Neurosci Methods, Vol.114, N0.1, 2002.

[8] S. Hernandez-Marin, Multilayered 3D LiDAR Image Construction Using Spatial Models in a Bayesian Framework, IEEE Transaction on Pattern Analysis and Machine Intellegence, Vol. 30, No. 6, 2008.

[9] E. Culurciello, et al. Biomorphic Digital Image Sensor IEEE Journal Of Solid-State Circuits, Vol. 38, No. 2, 2003

[10] S. Cova, et al, Single-Photon Avalnche Diodes For Quantum Key Distribution., Proc. of Single Photon Workshop. 27-28, 2007.

[11] Y. C. Tai, et al. Instrumentations Aspects Of Animal PET, Annu. Rev. Biomed. Eng, Vol. 7, 255-285, 2005.

[12] F. Guerrieri, et al. Fast Single-Photon Imager acquires 1024 pixels at 100 kframe/s , IS&T/SPIE Electronic Imaging, 2009.

[13] I. Rech, et al. High performance silicon single-photon avalanche diode array, Proc. of SPIE Vol. 7320, 2009.

[14] E. Charbon, High Speed CMOS Imaging: Four Years Later, International Conference on Solid-State and Integrated-Circuit Technology, 2008.

[15] F. Brian, et al. Geiger-Mode Avalanche Photodiodes for Three-Dimensional Imaging, Lincoln Laboratory Journal, Vol. 13, No. 2, 2002.

[16] H. W. Lee, Solid-State Impact Ionization Multiplier, Ph.D. Thesis. Bringham Young University, 2006.

[17] S. M. Sze, Physics of Semiconductor Devices. Wiley-Interscience, 1981.

[18] Ch. Spanoudaki, et al. Photo-Detectors for Time of Flight Positron EmissionTomography (ToF-PET), Sensors 10, 10484-10505, 2010.

[20] D.A. Shushakov, et al. New solid state photomultiplier, Proc. SPIE Int. Soc. Opt. Eng.Vol. 2397, 544-554, 1995.

[21] V.E. Shubin, et al, New avalanche device with the ability of analog few-photon pulse

detection, Proc. SPIE Int. Soc. Opt. Eng. Vol. 2550, 284-293, 1995.

[22] P.P. Antich, et al., Avalanche photo diode with local negative feedback sensitive to UV, blue and green light, Nucl. Instrum. Meth. A, Vol. 389, 491-498, 1997.

[23] N. Bacchetta et al., MRS detectors with high gain for registration of weak visible and UV light fluxes, Nucl. Instrum. Meth. A, Vol. 387, 225-230, 1997.

[24] G. Bondarenko et al., Limited Geiger-mode silicon photodiode with very high gain, Nucl. Phys. B, Vol. 61, 347, 1998) .

[25] Jackson, J. C., D. Phelan, A. P. Morrison, R. M. Redfern and A. Mathewson. Toward integrated single photon-counting microarrays. Opt. Eng., Vol.42:112–118, 2003.

[26] CL. Niclass, et al. A CMOS Single Photon Avalanche Diode Array for 3D Imaging”, ISSCC Digest of Technical papers, 120-121, 2004.

[27] A.Rochas, Single Photon Avalanche Diodes in CMOS Technology. Ph.D. Thesis, EDFL, 2003.

[28] A. Rochas, First Fully Integrated 2-D Array of Single-Photon Detectors in Standard CMOS Technology, IEEE PHOTONICS TECHNOLOGY LETTERS, VOL. 15, NO. 7, JULY 200

[29] M. Ghioni, A.Gulinatti, I. Rech, F. Zappa and Sergio Cova, “Progress in Silicon Single-Photon Avalanche Diodes,” IEEE J. Select. Topics Quantum. Electron., Vol. 13, no. 4, pp. 852-862, ,July/Aug., 2007.

[30] Simone Tisa, Fabrizio Guerrieri and Franco Zappa. Variable-Load Quenching Circuit for single-photon avalanche diodes. OSA 4 February 2008 / Vol. 16, No. 3 / OPTICS EXPRESS 2232

[31] J. Richardson, R. Walker, L. Grant, D. Stoppa, F. Borghetti, E. Charbon, M. Gersbach, and R. K. Henderson, “A 32x32 50 ps resolution 10 bit time to digital converter array in 130nm CMOS for time correlated imaging,” in Proceedings Of The IEEE 2009 Custom Integrated Circuits Conference (IEEE, New York, 2009), pp. 77–80.

[32] V H. Dhulla, Single photon counting for ultra-weak fluorescence detection: system design, characterization and application to DNA-sequencing. Ph.D. Thesis. Stony Brook University, 2007

[33] C. Niclass, A. Rochas, P.A. Besse, E. Charbon, “Towards a 3D Camera Based on Single Photon Avalanche Diodes” IEEE Journal of Selected Topics in Quantum Electronics, vol. 10 (4), pp. 796-802, 2004.

[34]] D. Stoppa, F. Borghetti, J. Richardson, R. Walker, L. Grant, R. Hen-derson, M. Gersbach, and E. Charbon, “A 32 32-pixel array within-pixel photon counting and arrival time measurement in the analogdomain,” in Proc. Eur. Solid-State Circuits Conference (ESSCIRC),2009, pp. 204–207.

[35] Renker, D. Geiger-mode avalanche photodiodes, history, properties and problems. Nucl. Instr.Meth. Phys. Res. A 2006, 567, 48–56.

[36] Mazillo, M.; Condorelli, G.; Sanfilippo, D.; Valvo, G.; Carbone, B.; Fallica, G.; Billota, S.Belluso, M.; Bonanno, G.; Cosentino, L.; Pappalardo, A.; Finocchiaro, P.; Silicon Photomultiplier Technology at STMicroelectronics. IEEE Trans. Nucl. Sci. 2009, 56, 2434–2442.

[37]. Piemonte, C.; Battiston, R.; Boscardin, M.; Dalla Betta, G.-F.; Del Guerra A.; DInu N.; Pozza A.Zorzi N. Characterization of the First Prototypes of Silicon Photomultiplier Fabricated at ITC-irst.IEEE Trans. Nucl. Sci. 2007, 54, 236–244.

[38] Stewart, A.; Saveliev, V.; Bellis, S.; Herbert, D.J.; Hughes, P.J.; Jackson, J.C. Performance of 1-mm 2 Silicon Photomultiplier. IEEE J. Quantum Electron. 2008, 44, 157–164.

[39] C. Ponchut, et al., Evaluation of medipix-1 in X-ray scattering and X-ray diffraction applications. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment",2003,1-2,29-34

[40] F. Corsi et al, ASIC development for SiPM readout, JINST 4, 2009,23-26

[41] R. B. Merrill, "Color separation in an active pixel cell imaging array using a triple-well

structure." USA: Foveon, Inc., 1999.

[42] S. Cova, M. Ghioni, A. Lacaita, C. Samori, F. Zappa, Avalanche photodiodes and quenching circuits for single-photon detection, Appl. Opt.35 (12) (1996) 1956–1976.

[43] A. S. Grove, O. Leistiko, Effect of Surface Fields on the Breakdown Voltage of Planar Silicon p-n Junctions, IEEE Transactions On Electron Devices, Vol. 14, NO.3 (1967)

[44] J. N. Haralson, II, and Kevin F. Brennan, Novel Edge Suppression Technique for Planar Avalanche Photodiodes, IEEE Journal Of Quantum Electronics, Vol. 34, NO. 12 (1999)

[45] I. Rech, M.Chioni, Progress in Quenching Circuits for Single Photon Avalanche Diodes, IEEE TRANSACTIONS ON NUCLEAR SCIENCE, VOL. 57, NO. 6, DECEMBER 2010

[46] M.dandin et al, Single Photon Avalanche Detectors in Standard CMOS Proceeding of IEEE conference on Sensors 2007 1889-1892

[47] M.Dandin et al, Characterization of Single-Photon Avalanche Diodes in a 0.5 um Standard CMOS Process Part 1: Perimeter Breakdown Suppression, IEEE Sensors Journal (2010).

[48] B.Nouri et al, Large-Area Low-Noise Single-Photon Avalanche Diodes in Standard CMOS, Sensors 2012

[49] A. Zanchi et al, On-chip probes for silicon defectivity ranking and mapping, Proceeding. 38th IEEE International Reliability Phuysics Symposium, 370 – 376, 2000

[50] B. Nouri et al, Structural optimization of CMOS Single Photon Avalanche Diodes, Proceeding of IEEE conference on Sensors 2009

[51] L. Alaverdian, et al, " A family of novel DNA sequencing instruments based on single photon detection" Electrophoresis 23 (2002), p. 2804.

[52] G. Godkuv et al, 32-Channel Single Photon Counting Module For Ultra-sensitive Detection of DNA Sequences, Proc. of SPIE Vol. 6372, 63720C, 2006

[53] A. V. Agronskaia, L. Tertoolen, H. C. Gerritsen, Fast Fluorescence Lifetime Imaging of Calcium in Living Cells, Journal of Biomedical Optics, Vol. 9, N. 6, pp. 1230-1237, Nov./Dec. 2004.

[54] P. Schwille, U. Haupts, S. Maiti, W. W. Webb, Molecular Dynamics in Living Cells Observed by Fluorescence Correlation Spectroscopy with One- and Two-Photon Excitation, Biophysics Journal, Vol. 77, pp. 2251-2265, 1999.

[55] G. MacBeath, "Protein microarrays and proteomics," Nat. Genet. 32 Suppl., 526–532 (2002).

[56] J. B. Pawley, Handbook of Biological Confocal Microscopy (Springer Press, 2006).

[57] D. Bonaccini, S. Cova, M. Ghioni, R. Gheser, S. Esposito, andG. Brusa, ''Novel avalanche photodiode for adaptive optics,'' in Adaptive Optics in Astronomy, M. Ealey and F. Merkle, eds, Proc. SPIE 2201, 650–657 119942.

[58] D. Bonaccini, F. Rigaut, A. Glindemann, G. Dudziak, J-M Mariotti, and F.Paresce "Adaptive Optics for ESO VLT Interferometer," Proc. SPIE 3353, 224-232 (1998).

[59] R. H. Haitz, "Mechanisms contributing to the noise pulse rate of avalanche diodes," J. Appl. Phys. 36, 3123–3131, 1965.

[60] S. Cova, A. Lacaita, and G. Ripamonti, "Trapping phenomena in avalanche photodiodes on nanosecond scale," IEEE Electron. Devices Lett. 12, 685–687, 1991.

[61] D. M. Taylor, et al, Characterization of novel active area silicon avalanche photodiodes operating in the Geiger mode, journal of modern optics, Vol. 51, NO. 9-10, 1323-1332 2004

[62] H. Lee, solid state impact ionization multiplier, Ph.D. Thesis, BYU, 2006.

[63] B F. Aull, et al, Geiger-Mode Avalanche Photodiodes for Three- Dimensional Imaging, Lincoln Laboratory Journal, vol 13, 2, 2002.

[64] A. Gallivanoni et al, Progress in Quenching Circuits for Single Photon Avalanche Diodes, IEEE Transactions On Nuclear Science, vol. 57, 6, 2010

[64] M A. Itzler et al, Geiger-Mode APD Single Photon Detectors, Optical Fiber Communication Conference, 2008

[65] M. Ghioni, Resonant-Cavity-Enhanced Single Photon Avalanche Diodes, IEEE Photncs Technology Letters, Vol. 20, 6, 413-415, 2008

[66] D. S. Bethune, W. P. Risk, and G. W. Pabst, J. Mod. Opt., vol. 51, no. 9-10, pp. 1359–1368, 2004.

[67] F. Zappa, Principles and features of single-pohoton avalanche diode array, Jour Sensors and Actuators A: Physical, Vol. 140, 1, 103-112, 2007

[68] GF. D Betta, Avalanche Photodiodes in Submicron CMOS Technologies for High-Sensitivity Imaging, Phtodiodes, 2011

[69] M.Liu et al, Reduce Afterpulsing of Single Photon Avalanche Diodes Using Passive Quenching With Active Reset, IEEE J. of Quant. El., vol 44, no 5, 430 ,2008

[70] P. Antognetti, et al, ‘A study of the operation and performances of an avalanche diode as a single photon detector,’’ in Proceedings of the Second Ispra Nuclear Electronics Symposium, EURATOM Publ. EUR 537e. 453–456, 1975.

[71] F. Zappa et al, Complete single-photon counting and timing module in a microchip, Optics Letters, Vol. 30, No. 11, 2005.

[72] S. Tisa et. Al, “Electronics for Single photon avalanche diode arrays, Sensors and Actuators A 140, 113-122, 200.

[73] F. Zappa et al, Fully integrated Active-Quenching circuits for single-photon detection, ESSCIRC, 355-358, 2002.

[74] D. Cronin, et al, Simulated monolithically integrated single photon counter, in Proc. IEEE High Frequency Postgraduate Student Colloquium’04, 9 -14, 2004.

[75] M. Stipcevic, Active quenching circuit for single-photon detection with Geiger mode avalanche photodiodes, Applied Optics, Vol. 48, No. 9, 2009.

[76] J. Richardson, et al, Dynamic Quenching for Single Photon Avalanche Diode Arrays, International Image Sensor Workshop, 2007.

[77] C. Niclass, et al, A CMOS 64x48 Single Photon Avalanche Diode Array with Event-Driven Readout, IEEE ESSCIRC, 2006.

[78] Christine Silberhorn, Detecting quantum light, Contemporary Physics, Vol. 48, No. 3, 143-156, 2007.

[79] Daryl Achilles et al, Photon number resolving detection using time-multiplexing, Journal of Modern Optics, Volume 51, Issue 9 & 10 June 2004, 499 - 1515

[80] James F. Christian et al, Advances in CMOS solid-state photomultipliers for scintillation detector applications, Nuclear Instruments and Methods in Physics Research A

[81] A.G. Wright, The statistics of multi-photoelectron pulse-height distributions, Nuclear Instruments and Methods in Physics Research A 579 (2007) 967–972

[82] N. Otte, The Silicon Photomultiplier - A new device for High Energy Physics, Astroparticle Physics, Industrial and Medical Applications, SNIC Symposium, Stanford, California -- 3-6 April 2006

[83] V Spanoudaki, Photo-Detectors for Time of Flight Positron Emission Tomography (ToF-PET), Sensors Oct, 10484-10505, 2008

[84] P. Buzhan et al., Silicon photomultiplier and its possible applications, Nucl. Instrum. Methods Phys, vol. 504, no. 1–3, 48–52, 2003

[85] P. Buzhan, et al, An advanced study of silicon photomultiplier, ICFA Instrumentation Bulletin, vol. 23, 28, 2001.

[86] I. Afek, et al. Quantum state measurements using multipixel photon detectors. Phys. Rev. A, 79(4):043830, 2009.

[87] C. Silberhorn, Detecting quantum light, Contemp. Phys.48, 143-156,

[88] S. Castelletto, et al, Theoretical aspects of photon number measurement," Metrologia 37, 613–616, 2000

[89] G. Zambra, et al, Reconstruction of photon statistics using low performance photon counters, arXiv:quant-ph/0607052, vol. 1, 2006.

[90] Li G, et al, Photon statistics of light fields based on single-photon-counting modules, Phys Rev A, 2005

[91] D. Achilles et al, Direct loss- tolerant characterization of nonclassical photonstatistics, Phys. Rev. Lett. 97, 043602 , 2006

[92] P. Schwille et al., "Molecular Dynamics in Living Cells Observed by Fluorescence Correlation Spectroscopy with One- and Two-Photon Excitation", Biophysics .l, Vol. 77, 2251-2265, 1999.

[93] C. Situma, et al, Merging microfluidics with microarray-based bioassays, Biomol.Eng, 23(5), 213-231, 2006

[94] J. Petrik, Diagnostic applications of microarrays, Transfus. Med. 16(4), 233-247, 2006

[95] S. Nagl, et al, Fluorescence analysis in microarray technology, Mikrochim. Acta 151, 1-21, 2005

[96] Gambhir SS. 2002. Molecular imaging of cancer with positron emission tomography. Nat. Rev. Cancer 2:683–93

[97] Phelps ME. 2000. PET: the merging of biology and imaging into molecular imaging. J. Nucl. Med. 41:661–81

[98] B. Valeur. Molecular Fluorescence: Principles and Applications. Wiley-VCH, 2002.

[99] P. Finocchiaro et al., “Characterization of a novel 100-channel silicon photomultiplier-Part I: Noise,” IEEE Trans. Electron Devices, vol. 55, pp. 2757–2764, 2008

[100] S. Vinogradov et al, Probability Distribution and Noise Factor of Solid State Photomultiplier Signals with Cross-Talk and Afterpulsing

[101] B. Dosgoshein, et al., Status report on silicon photomultiplier development and its applications, Nucl. Instr. Meth. A, vol. 563, pp. 368-376, 2006

[102] M. Ramilli, et al, “Photon-number statistics with silicon photomultipliers,” J. Opt. Soc. Am. B 27(5), 852–862 , 2010

[103] H B Coldenstrodt-Ronge, et al. Avalanche photo-detection for high data rate applications, J. Phys. B: At. Mol. Opt. *Phys*, Vol. 40, No. 19,

[104] D. Renker, et al. Advances in solid state photon detectors, J. Instrumentation, Vol. 4, No. 4, P04004, 2009.

[105] I. Rech, et al. Multipixel single-photon avalanche diode array for parallel photon counting applications, J.Mod. Opt. 56(2), 326–333,2009

[106] V. Ch. Spanoudaki, et al, Use of single photon counting detector arrays in combined PET/MR: Characterization of LYSO-SiPM detector modules and comparison with a LSO-APD detector, J. Instrum, P12002, 2007

[107] V. Ch. Spanoudaki, et al.Photo-Detectors for Time of Flight Positron Emission Tomography (ToF-PET, Sensors 10, 10484–10505, 2010

[108] K. Kleinknecht. Detectors for Particle Radiation. Cambridge University Press, second edition, 1998.

[109] L. Neri, et al. Dead time causes and correction method for single photon avalanche diode devices, Review of Scientific Instrumentation, 81, 086102, 2010.

[110] L. Neri, Time Resolved Photon Imaging Sensor with Single Photon Avalanche Diode. Ph.D. Thesis, 2011.

[111] S. Vinogradov, et al. Efficiency of solid state photomultipliers in photon number resolution, IEEE Nuclear Science Symposium Conference Record, N28- 3, 2009.

[112] K. Jahoda, et al. Peak luminosities of bursts from GRO J1744–28 measured with the RXTE PCA. Nucl. Phys. B 69, 210–215, 1998.

[113] CJ. Stapels, et al. CMOS-based avalanche photodiodes for direct particle detection, Nucl Instrum Methods Phys Res Sect A, 2007.

[114] HG Moser, "Silicon detector systems in high energy physics, Progress in particle and nuclear physics, Vol 63, 186 – 237, 2009.

[115] C. Niclass , et al. A 4 µs integration time imager based on CMOS single photon avalanche diode technology, Sens. Actuators A, Phys., vol. 130-131, 273 -281, 2006.

[116] C. Niclass, et al. A 128 × 128 Single-Photon Image Sensor With Column-Level 10-Bit Time-to-Digital Converter Array" IEEE J. Sol.-St. Circ. 43, 2977–2989, 2008.

[117] E. Charbon, Highly sensitive arrays of nano-sized single-photon avalanche diodes for industrial and bio imaging, in Proc. Nano-Net, 4th Int. ICST Conf.,161–168, 2009.

[118] E. Charbon, Techniques for CMOS Single Photon Imaging and Processing", IEEE ASICON, Oct. 2005.

[119] E. Charbon. Single-photon Imaging in CMOS, Proc. SPIE, vol. 7780, 77801D-1 -77801D-15, 2010

[120] S. J., et al., A 256x256 CMOS imaging array with wide dynamic range pixels and column-parallel digital output. IEEE J. of Solid State Circuits, Vol. 33, 2081-2091, 1998.

[121] C.J.M. Basset, CMOS Imaging Technology with Embedded Early Image Processing,. Ph.D. Thesis., California Institute of Technology, 2007.

[122] D. Stoppa, et al., A 32x32-Pixel Array with In-Pixel Photon Counting and Arrival Time Measurement in the Analog Domain, IEEE European Solid-State Device Conf, 2009.

[123] F. Guerrieri, et al. Fast single-photon imager acquires 1024 pixels at 100 kframe/s, Proc. SPIE, vol. 7249, 72490U, 2009.

[124] E. Charbon. Towards large scale CMOS single-photon detector arrays for lab-on-chip applications," J. Phys. D: Appl. Phys., vol. 41, no. 9, 2008.

[125] G. Giraud, et al. Fluorescence lifetime biosensing with DNA microarrays and a CMOS-SPAD imager, Biomedical Optics Express 1(5), 1302–1308, 2010.

[126] LQ. Li. Single photon avalanche diode for single molecule detection, Rev. Sci. Instrum.á64, 1524-1529, 1993.

[127] P. Gatt, S. Johnson, and T. Nichols, Geiger-mode avalanche photodiode ladar receiver performance characteristics and detection statistics, Appl. Opt. 48, 3261–3276, 2009.

[128] C. Lotto. Synchronous and Asynchronous Detection of Ultra-low Light Levels Using CMOS-Compatible Semiconductor Technologies. Ph.D. Thesis. University of Neuchatel, 2010.

[129] F. Corsi et al., "ASIC development for SiPM readout," JINST 4, P03004, 2009.

[130] S.E. Derenzo, Scintillation counters, Photodetectors and Radiation Spectroscopy, Nuclear Science Symp, IEEE Short Course Radiation Detection and Measurement, 1997.

[131] H. Spieler, Pulse Processing and Analysis, Nuclear Science Symposium, IEEE NPS Short Course, Radiation Detection and Measurement, 2002

[132] A. Baschirotto et al. A CMOS high-speed front-end for cluster counting techniques in ionization detectors; Proc. of International Workshop on Advances in Sensors and Interface, 2007.

[133] ML. Woodring, Multiplexed avalanche photodiode arrays for radiation imaging, Proc. SPIE 4784, X-Ray and Gamma-Ray Detectors and Applications IV, 2003.

[134] M. Bouchel et al., SPIROC (SiPM Integrated Read-Out Chip): Dedicated very front-end electronics for an ILC prototype hadronic calorimeter with SiPM read-out, IEEE Nuclear Science Symposium Conference Record, 2007.

[135] G. Gramegna, Low-noise CMOS Preamplifer-Shaper for Silicon Drift Detectors, IEEE Transactions on Nuclear Science, Vol. 44, No. 3, 1997.

[136] F. Corsi et al., “ASIC development for SiPM readout,” JINST 4, P03004, 1-10, 2009.

[137] Sun B, et al. A new differential CMOS current pre-amplifier for optical communication in Proc.of IEEE ISCAS’03, vol. 1, 341–344, 2003.

[138] F. Corsi et al., CMOS analog front-end channel for silicon photo-multipliers, Nucl. Instr. and Meth A617, 319-320, 2010.

[139] S. Kleinfelder, Gigahertz Waveform Sampling and Digitization Circuit Design and Implementation, IEEE Trans. Nucl. Sci. Vol.50, No. 4, 2003.

[140] C. Brönnimann, R. Horisberger and R. Schnyder, Nucl. Instr. Meth. A420, 264, 1999.

[141] S. Ritt, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, Nucl. Instr. Meth. Vol. 518, 470-471, 2004.

[142] S. Ritt. Development of high speed waveform sampling ASICs, NSNI-2010 Mumbai, 2010.

[143] G. Varner, et al. The first version buffered large analog bandwidth (BLAB1) ASIC for high luminosity collider and extensive radio neutrino detectors, Nuclear Instruments and Methods in Physics Research, Vol. 591, 534-545, 2008.

[143] P. Finocchiaro, et al..Characterization of a novel 100-channel silicon photomultiplier-part II: charge and time, IEEE Transactions on Electron Devices, Vol. 55, No. 10, 2765-2773, 2008.

[144] H. Spieler, Silicon Detectors, Part 2, SLUO Lectures on Detector Techniques, Stanford Linear Accelerator Center, 1998.

[145] D.G. Maeding, et al. Readout Electronics for Nuclear Applications (RENA) IC, Proc. SPIE 3445, 364, 1998.

[146] H. Finkelstein, et al. Performance trade-offs in single photon avalanche diode miniaturization, Review of scientific instruments, Vol 78, 2007

[148] P.D. Olcott, et al. Pulse Width Modulation: a Novel Readout Scheme for High Energy Photon Detection, IEEE Nuclear Symposium Conference Record, 4530-4530, 2008.

[149] C. de La Taille, et al. Front-end chip for SiPM readout of ILC Analog HCAL, International Linear Collider Workshop, 2005.

[150] P. Barrillon, et al. MAROC: Multi-Anode Readout Chip for MaPMTs, IEEE Nuclear science symposium, 809, 2006.

[151] S. Callier ,et al, Silicon Photomultiplier integrated readout chip (SPIROC) for the ILC: measurements and possible further development, IEEE Nuclear Science Symposium , 2009.

[152] P. Jarron, et al. Time based Readout of silicon photomultiplier (SiPM) for Time Of Flight Positron Emission Tomography (TOF-PET), IEEE Nuclear science symposium, 2009.

[153] F. Powolny, A Novel Time-based Readout Scheme for a Combined PET-CT Detector Using APDs, IEEE Nuclear science symposium, 2008.

[154] M.Ritzert, et al. Compact SiPM based Detector Module for Time-Of-Flight PET/MR, 167th IEEE-NPSS Real Time Conference, 163, 2009.

[155] P.Fischer, et al. Fast Triggered Multi Channel Readout ASIC for Time and Energy Measurement, IEEE Transactions on Nuclear Science, Vol. 56, No. 3, 1153, 2009.

[156] F. Corsi, et al. BASIC: An 8 Channel Front-end ASIC for Silicon Photomultiplier Detectors, IEEE Nuclear Science Symposium, 1082, 2009.

[157] A. Delguerra, et al. Silicon Photomultipliers (SiPM) as novel photodetectors for PET – NIM, Nuclear Instruments and Methods in Physics Research A, Vol. 648, 232-235, 2011.

[158] P.S. Marrocchesi, et al. Test of front-end electronics with large dynamic range coupled to SiPM for space-based calorimetry, 30th International Cosmic Ray Conference, 2007.

[159] R. Mos, et al. Front-End electronics for Silicon Photomultipliers Implemented in CMOS VLSI, 16th International Conference on Mixed Design of Integrated Circuits and Systems, 266, 2009.

[160] W. Kuecewicz, et al. The two channel CMOS converter for silicon photomultiplier, International conference on signals and electronic Systems, 165, 2008.

[161] M. Micuda, et al. High-efficiency photon-number-resolving multichannel detector, Phys.Rev. Vol. 78, 2008

[162] C. Degenhardt, et al. The Digital Silicon Photomultiplier– A Novel sensor for the Detection of Scintillation Light, Proceedings of Nuclear Science Symposium , 2383-2386, 2009.

[163] T. Frach, et al. The Digital Silicon Photomultiplier–Principle of Operation and Intrinsic Detector Performance, Proceedings of Nuclear Science Symposium, 1959–1965, 2009.

[164] T. Frach, et al. Digital silicon photomultiplier for TOF-PET. Patent No. US7723694, 2010.

[165] A. Hayrapetyan, et al. New Digital SiPMs from Philips: Applications and First Tests, Contribution to the 3 rd Detector Workshop of the Helmholtz Alliance "Physics at the Terascale", Heidelberg, 2010.

[166] R. Chunara, et al. Low-Noise electronic Readout for high-Throughput, Portable Bimolecular Detection in Micro-Channel Arrays, Ph.D. Thesis, Massachusetts Institute of Technology, 2006.

[167] S. Louise, A 1-V CMOS D/A Converter with Multi-Input Floating-Gate MOSFET, IEEE Journal of Solid-state Circuits, Vol. 34, No. 10, 1999.

[168] D. Arumi, et al. Diagnosis of Full Open Defects in Interconnect Lines with Large Fan-out, XXIV Conference on Design of Circuits and Integrated Systems, 2009.

[169] K. D. Layton, Low-Voltage Analog CMOS Architectures and Design Methods,. Ph.D. Thesis, Brigham young University, 2007.

[170] S. Y. Peng. Charge-based Analog Circuits For Reconfigurable Smart Sensory Systems. Ph.D. Thesis, Georgia Institute of Technology, 2008.

[171] C. Jansson, A High-Resolution, Compact, and Low-Power ADC Suitable for Array Implementation in Standard CMOS, IEEE Transactions On Circuits And Systems – I: Fundamental Theory and Application, Vol. 42, No. 11, 1995.

[172] H. Chen, et al. A 13-bit, low-power, compact ADC suitable for sensor applications, ISCAS, 2414-2417, 2010.

[173] G. S. Varner, et al., Compact, low-power and precision timing photodetector readout, PoS, 62, 2008.

[174] T. Ohnuki, et al.,Development of an Ultra-fast Single-Photon Counting Imager for Single-Molecule Imaging, Proc. of SPIE Vol. 6092, 1605-7422, 2006)

[174] T. Xia, et al., On-chip short-time interval measurement system for high-speed signal timing characterization., Asian Test Symposium, 326-331, 2003.

[175] Z. Ning, et al., A Simple and Accurate Capacitance Ratio Measurement Technique for Integrated Circuit Capacitor Arrays, Proc. IEEE Inf. Conference on Microelectronic Test Structures, Vol. 18, 2005.

[176] J. L. McCREARY, et al., Precision Capacitor Ratio Measurement Technique for Integrated Circuit Capacitor Arrays., IEEE Transactions On Instrumentation Ans Measurement, Vol. 28, No. 1, 1979.