Applied Quantum Architectures
Mekelweg 4,
2628 CD Delft
The Netherlands
http://aqua.ewi.tudelft.nl/

CAS-2019-00

M.Sc. Thesis

Design of Low-Threshold Comparator for Improved Timing-Resolution Analog/Digital SiPM

Ashish Sachdeva

Abstract

Coincidence time resolution (CTR) in time-of-flight (TOF) positron emission tomography (PET) determines the signal-to-noise ratio (SNR) in iterative image reconstruction algorithms. In PET detectors, the photodetector's single-photon timing resolution (SPTR) influences the CTR by adding uncertainty to the single photoelectron time-of-arrivals. This effect can be modelled as the convolution of the scintillation pulse shape function and the total photodetector jitter at single-photon level, before following an order statistics process. Particularly in Cherenkov-based PET detectors, SPTR has a direct impact on the CTR due to the low number of detected photons.

In this thesis, the research is focussed on the design of low threshold comparators for two specific purposes in PET. Firstly, the design of in-pixel (SPAD-cell) low-threshold comparator that improves SPAD jitter at the pixel level, since it allows the detection of photo-electron triggered avalanche at the earliest possible time, thus minimizing statistical fluctuations. The design is targeted for Cherenkov-based PET, where improvement in SPTR directly results in improvement of CTR. Secondly, for the integration of analog silicon photomultiplier (A-SiPM) on-chip. Such integration helps in realizing a high Photon Detection Efficiency (PDE) and low Dark Count Rate (DCR) A-SiPM with integrated readout electronics. A high speed comparator with a direct connection to the fast terminal of A-SiPM has been realized.

Design of Low-Threshold Comparator for Improved Timing-Resolution Analog/Digital SiPM

Thesis

submitted in partial fulfillment of the
requirements for the degree of

Master of Science

in

Electrical Engineering

by

Ashish Sachdeva
born in Delhi, India

This work was performed in:

Applied Quantum Architectures Group
Department of Microelectronics & Computer Engineering
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft University of Technology

Delft University of Technology

Delft University of Technology
Department of
Microelectronics & Computer Engineering

The undersigned hereby certify that they have read and recommend to the Faculty of Electrical Engineering, Mathematics and Computer Science for acceptance a thesis entitled **"Design of Low-Threshold Comparator for Improved Timing-Resolution Analog/Digital SiPM"** by **Ashish Sachdeva** in partial fulfillment of the requirements for the degree of **Master of Science**.

Dated: 17th December, 2018

Chairman:

prof.dr.ir. Edoardo Charbon

Advisors:

prof.dr.ir. Edoardo Charbon

ir. Esteban Venialgo

Committee Members:

prof.dr.ir. Edoardo Charbon

dr.ir. Carl Jackson

prof.dr.ir. Stephan Wong

Abstract

Coincidence time resolution (CTR) in time-of-flight (TOF) positron emission tomography (PET) determines the signal-to-noise ratio (SNR) in iterative image reconstruction algorithms. In PET detectors, the photodetector's single-photon timing resolution (SPTR) influences the CTR by adding uncertainty to the single photoelectron time-of-arrivals. This effect can be modelled as the convolution of the scintillation pulse shape function and the total photodetector jitter at single-photon level, before following an order statistics process. Particularly in Cherenkov-based PET detectors, SPTR has a direct impact on the CTR due to the low number of detected photons.

In this thesis, the research is focussed on the design of low threshold comparators for two specific purposes in PET. Firstly, the design of in-pixel (SPAD-cell) low-threshold comparator that improves SPAD jitter at the pixel level, since it allows the detection of photo-electron triggered avalanche at the earliest possible time, thus minimizing statistical fluctuations. The design is targeted for Cherenkov-based PET, where improvement in SPTR directly results in improvement of CTR. Secondly, for the integration of analog silicon photomultiplier (A-SiPM) on-chip. Such integration helps in realizing a high Photon Detection Efficiency (PDE) and low Dark Count Rate (DCR) A-SiPM with integrated readout electronics. A high speed comparator with a direct connection to the fast terminal of A-SiPM has been realized.

To the memory of my late mother Anita Sachdeva.

Acknowledgments

This masters dissertation has been able to impart a decisive contribution to the scientific research due to the contribution of many esteemed colleagues and support from friends and family.

First and foremost, I would like to begin by thanking prof.dr.ir. Edoardo Charbon for providing this great opportunity to expand on my research comprehension. This work has been made possible because of his crucial insights and guidance. Edoardo's wisdom and knowledge was the key to achieve a conclusive research.

Secondly, I thank ir. Esteban Venialgo for his major contribution in this research. Esteban's supervision at every point in the research is also the reason this contribution has been made possible. His knowledge and support has been critical for my research growth over the duration of my masters. Esteban's support and help has been a pillar strength for which I am deeply greateful.

I would like to thank SensL, and Dr. C. Jackson, for providing me the platform to work on this project.

Many esteemed colleagues and friends at TU Delft have contributed their time and knowledge in the immense discussions on numerous occasions, whom I would like to thank. I am equally thankful to Rushil Kishore Kumar, and Sriram Balamurali as they contributed great deal of time and expertise in helping me expand on my knowledge of analog IC design. I am thankful to Augusto Carimatto. Augusto's kind help and support on every occasion continue to amaze me. I would also like to thank Jeroen Van Dijk. Jeroen's key insights in the design phase of the research has helped in the realization of a meaningful contribution.

I would like to thank my friends Deniz Hofmeister, Gerard Baquer, Guillermo Ortas, Nanne Knijff, and Raoul Boedhoe.

This thesis and my masters would not have been made possible without the love of my late mother Anita Sachdeva. I also have been fortunate to have the support of my father and sisters who I cannot thank enough for their love.

Ashish Sachdeva
Delft, The Netherlands
17th December, 2018

Contents

List of Figures

List of Tables

1 Introduction

1.1 Introduction to Positron Emission Tomography

Positron emission tomography (PET) is a nuclear imaging technique utilized in medical applications where molecular probes are injected into the body (via intravenous drug) to detect abnormalities in organs or tissues of the human body. The molecular probes labeled with radioisotopes, which are atoms that decay by emitting radioactivity in order to reach a stable atomic configuration, are utilized for detecting various diseases with a very high molecular sensitivity. This is possible due to the biochemical activity of the molecular probe with the diseased tissue, due to which the unhealthy tissue absorbs the molecular probe in higher concentration. The molecular probes utilized in PET, which are labeled with a positron emitting radioisotope, has a probability to reach a targeted lesion and emit a positron (e^+) from there.

The emitted positron immediately annihilate with a surrounding electron and two back-to-back gamma photons are generated as an outcome. The respective gamma photons are detected by the PET scanner detectors. Subsequently, the detection information is utilized to generate tomographic images that corresponds to the molecular probe spatial concentration.[11]

1.1.1 Annihilation

The ejected positron propagates for a small distance before it collides with an electron, and energy is released in the form of high energy gamma photons after the occurrence of an annihilation. The total energy released in this phenomenon can be computed from Einstein's mass-energy, and is given by

$$E = mc^2 = m_e c^2 + m_p c^2 = 1.022\,\text{MeV}. \tag{1.1}$$

The electron and positron are almost at rest when the process of annihilation occurs. Henceforth, due to Newton's first law of conservation of momentum, the net momentum should be approximately zero. As a result two gamma photons are emitted in opposite direction with an energy of 511 keV each as shown in Fig. 1.1 [11][12].

The gamma photons generated by annihilation are then detected by the PET scanner, which measure events within a certain time coincidence window. The two gamma photons traveling in opposite directions form line-of-response (LOR), which are detected in a ring-like structure of PET detectors. Denoting the time-of-arrival (TOA) of two gamma (γ) photons, which are back-to-back, by t_1 and t_2, the location d of the annihilation along the LOR is as follows

$$d = (t_1 - t_2) \times \frac{c}{2}, \tag{1.2}$$

where c is the speed of light, as seen in figure 1.2(a.).

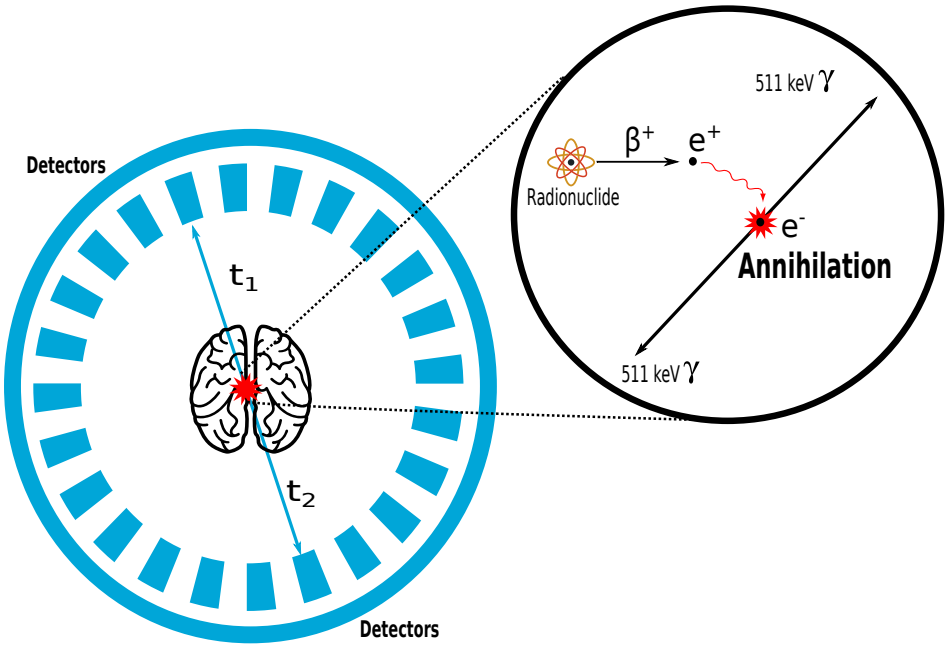

Figure 1.1: The positron and electron annihilation phenomenon

1.1.2 Gamma Photon detection

There are multiple ways to detect the γ photon radiation. The common method is based on utilizing a transparent crystalline structure called scintillator, which converts the energy from γ radiation into visible isotropic light. The released light is detected using a photodetector. The research of this thesis is centered around a sub-class of photodetectors called silicon photomultipliers (SiPMs), which essentially are arrays of single-photon avalanche diodes (SPADs). The scintillators for PET imaging in the past were designed from bismuth germanate (BGO) as it has high density, which allows it to have a high γ detection efficiency. However, BGO was not suitable for time-of-flight PET (TOF-PET) applications unless Cherenkov photons are detected [13]. The LSO and LYSO have replaced BGO due to higher light output and faster decay time.

1.1.3 Time-of-Flight PET imaging

One contributing factor in improving TOF-PET is the use of cerium-doped lutetium oxyorthosilicate (LSO) and LYSO based scintillators. The LSO/LYSO provides better performance in coincident timing resolution and thus TOF-PET reconstruction algorithm can be applied to improve the reconstructed images' signal to noise ratio (SNR) [14].

In TOF-PET reconstruction, the annihilation spatial probability can be localized within the LOR as the difference between the arrival time of the two γ photons $(t_1 - t_2)$, if the timing is precise enough measured precisely (see figure 1.2(c.)). The difference between TOF-PET and conventional PET is that the former only utilizes the time information for validating that a pair of detected γ photons belong to the same annihilation process.

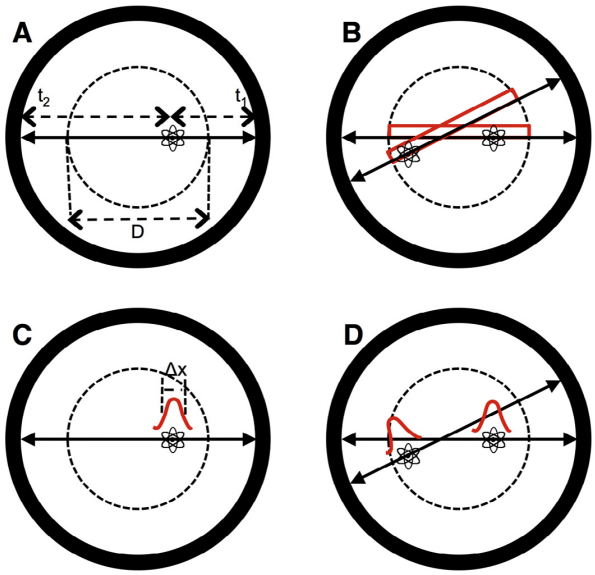

Figure 1.2: (A.) Annihilation point within the object with diameter D, which hits the detectors at t_1 and t_2 of the scanner. (B.) Image reconstruction with conventional PET by multiple LORs collection and mathematical reconstruction. (C.) The annihilation spatial probability, in TOF-PET, is restricted to Δx within the same diameter D. (D.) Localization to small spatial resolution. This image has been taken from S.Surti et. al.[1]

1.1.3.1 Coincidence Timing Resolution (CTR)

In TOF-PET, the CTR influences the precision of the annihilation spatial probability within the LOR. Therefore, the image reconstruction SNR improvement is directly influenced by the CTR. When two opposite γ photons arrive at the, say, identical scintillators within a coincidence window, the CTR is defined as follows (figure 1.3)[2]:

$$CTR_{sigma} = \sqrt{2}\sigma(t_1), \tag{1.3}$$

$$CTR_{fwhm} \approx 2.35\sqrt{2}\sigma(t_1). \tag{1.4}$$

In the previous equation, a gaussian distribution is assumed for the timing response of the PET detectors, which can be different in experimentation.

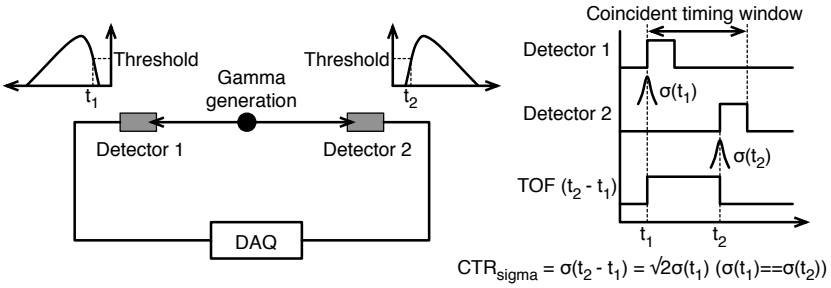

Figure 1.3: Coincidence timing resolution calculation[2]

1.2 Silicon Photomultipliers

After the scintillation process, the light photons are measured by a photodetector. In the past, the most advancements have been achieved through photomultiplier tubes (PMT) due to their high multiplication gain and fast response. However, the PMTs tag along certain drawbacks with them. Foremost, it is highly susceptible to magnetic fields, causing generated electrons to deflect in different directions under the influence of magnetic fields. Therefore, they are not suitable for PET/MRI hybrid systems.

Around the beginning of this century, silicon-based matrix devices which employ photodiodes that operate in reverse bias region beyond breakdown voltage, gained interest in the field of LIDAR and TOF-PET imaging. These devices are often called as silicon photomultipliers (SiPMs), because of their single-photon detection capability, and precise timing response. The gieger mode devices placed in the SiPM matrix are called single-photon avalanche diodes (SPADs).

1.2.1 Single-photon Avalanche Diode

Single-photon avalanche diode (SPAD) is a pn diode biased beyond the breakdown voltage and operates in gieger mode. When the light photon hits the depletion region of the SPAD, ionized carriers are introduced in the device. Due to high electric field, impact ionization causes a rapid multiplication of electron-hole pairs. This state is called avalanche build-up.

The SPAD would be permanently destroyed if the avalanche is not rapidly quenched. In order to not damage the device and not allow the current to flow in the diode for a long period of time, the SPADs are provided with a quenching circuit, which brings its bias voltage below its breakdown.

1.3 Research Motivation

1.3.1 From SPTR to CTR

The equations presented in this section model the timing resolution of the SPAD and SiPM, and many factors like shot noise, dead time, after pulsing are assumed to be not present in estimating the full width half maximum (FWHM) timing resolution. That being said, the analysis shown below explains how single-photon timing resolution (SPTR) and CTR are related.

As the scintillation photons time-of-generation have a probabilistic distribution function (PDF), $f_a(t)$ which can be approximated as a double-exponential function,

$$f_a(t) = \begin{cases} 0 & : t < 0, \\ \frac{exp(\frac{-t}{\tau_d}) - exp(\frac{-t}{\tau_r})}{(\tau_d - \tau_r)} & : t > 0, \end{cases} \tag{1.5}$$

where τ_d and τ_r are the decay and rise time constant respectively[3][15]. The timing jitter for a single photon from a SPAD in the D-SiPM array can be considered to follow a normal distribution $\mathcal{N}(\mu, \sigma^2)$ [16]. Here, μ is the mean and σ is the standard deviation,

or timing jitter of the SPAD. For the case of simplicity let us plot equation 1.5 as shown in figure 1.4(a.), using arbitrary values of τ_d as 40 ns and τ_r as 0.5 ns [3].

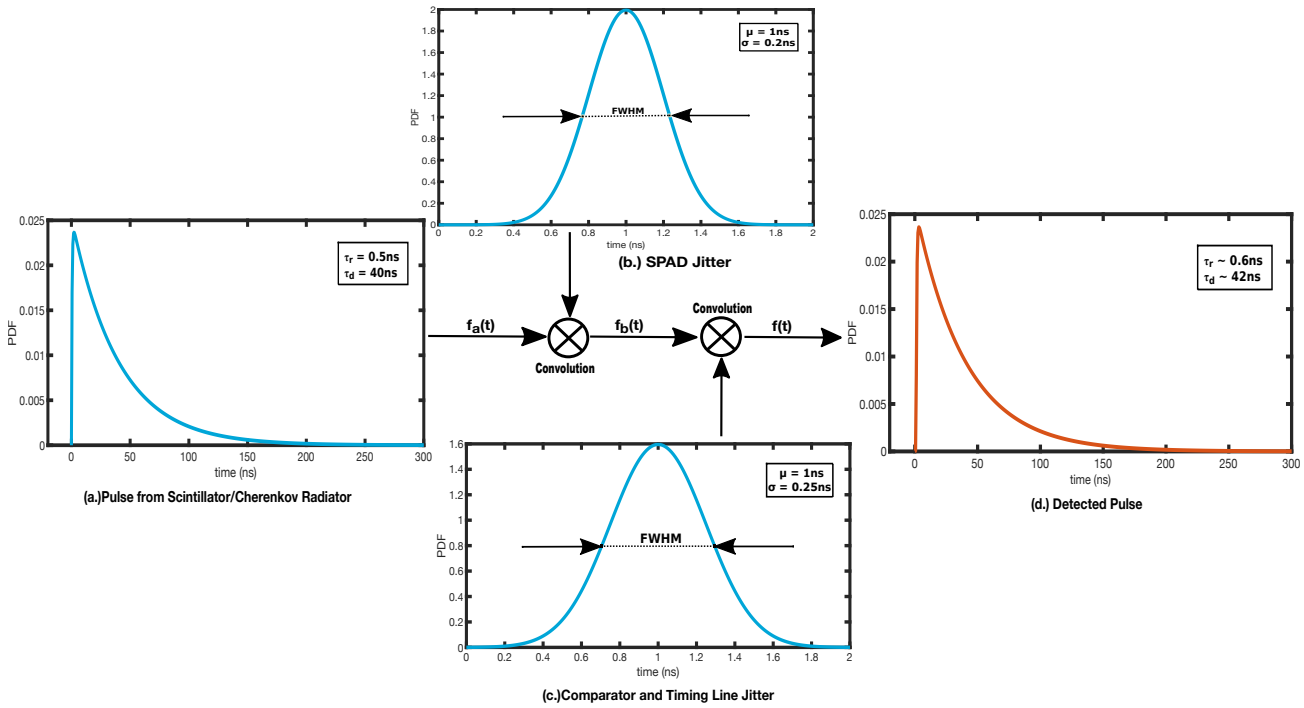

Figure 1.4: Block Diagram depicting the scintillation detection model. Constructed using [3][4]

Adding the SPAD jitter into the calculations, the arrival time PDF at the pixel level becomes a convolution of $f_a(t)$ with SPAD's timing jitter represented by its normal distribution $\mathcal{N}(\mu_s, \sigma_s^2)$. Let us consider, for better understanding σ_s equal to 0.2 ns, and μ_s equal to 1 ns, and the normal distribution PDF is plotted in Fig. 1.4(b).

$$f_b(t) = f_a(t) \circledast \mathcal{N}(\mu_s, \sigma_s^2). \tag{1.6}$$

The measurement electronics (comparator, timing lines, and time to digital converter (TDC) in later stage) further add timing degradation in the form of timing jitter. The comparator timing jitter can be due to multiple reasons such as noise, offset, and other non avoidable causes. Presuming for the sake of simplicity, the timing jitter of the comparator and timing lines can also be represented in the form of a gaussian distribution,$\mathcal{N}_{Comp,tlines}$ as shown in Fig. 1.4(c). The numbers assigned in this case are σ_c as 0.25 ns, and μ_c as 1 ns. The measured time PDF at the output of the photodetector, D-SiPM, is the convolution of $f_b(t)$ and $\mathcal{N}_{Comp,tlines}(\mu_c, \sigma_c^2)$ as shown below and also plotted in Fig. 1.4(d):

$$f(t) = f_b(t) \circledast \mathcal{N}_{Comp,tlines}(\mu_c, \sigma_c^2). \tag{1.7}$$

In this specific case of gaussian distributions, the total jitter of the photodetector, which is defined as the SPTR of the photodectector, can also be modelled as a subsequent

convolutions. Also, it is possible to replace the subsequent convolutions just by increasing the overall jitter sigma by adding in quadrature all the contributions (σ_c) of the comparator, timing line jitter, and the SPAD jitter. However, the convolution modelling is not limited to gaussian distributions and is true for any distribution type to reach the SPTR of the photodetector. Henceforth we reach to a conclusion that the SPTR can be achieved by,

$$\mathcal{N}(\mu_{SPTR},\, \sigma^2_{SPTR}) = \mathcal{N}_{SPAD}(\mu_s,\, \sigma^2_s) \circledast \mathcal{N}_{Comp,tlines}(\mu_c,\, \sigma^2_c). \tag{1.8}$$

As seen from Fig. 1.5, the magnified pulse from the scintillator, and the detected pulse at the comparator show that due to added timing jitter, the detected pulse has deteriorated rise time compared to the original pulse.

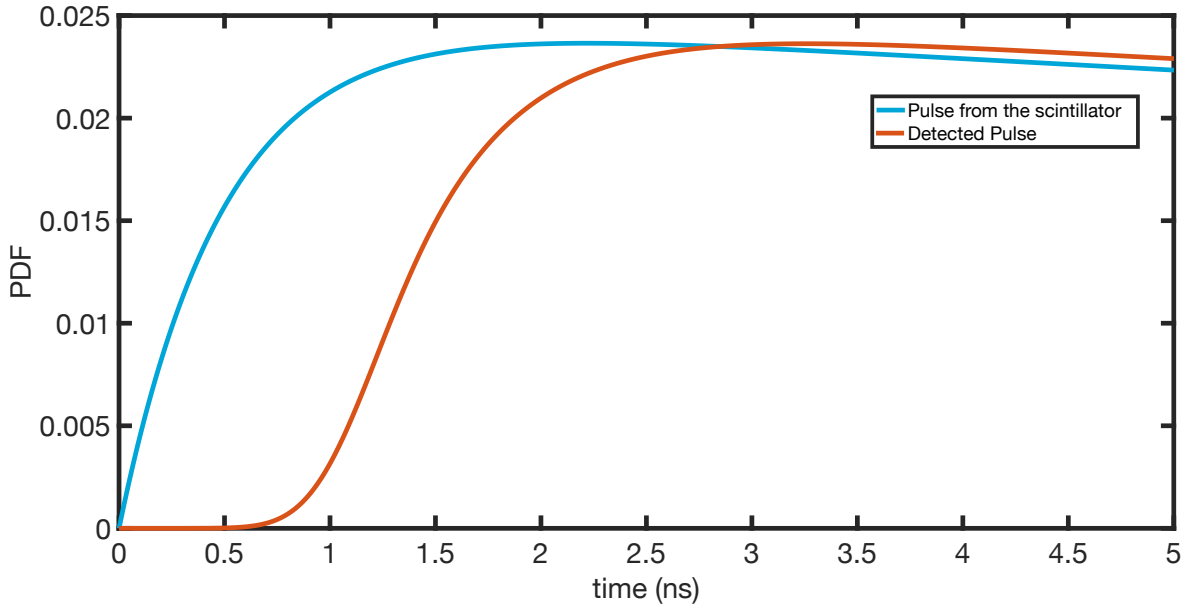

Figure 1.5: Magnified view of the pulse from the scintillator and the detected pulse

The time statistical evaluation on the photo-electron pulse emission from scintillator shows that for any n number of time ordered samples, the order statistics of PDF can be described as [17],

$$f_{k:n}(t) = n \binom{n-1}{k-1} f(t) F(t)^{k-1} [1 - F(t)]^{n-k}. \tag{1.9}$$

Here, k is the k^{th} order statistics. Hence, equation 1.9 approximates the timestamp distribution of detected photoelectrons, if f(t) is the resulting convolution of the scintillation pulse-shape function and the overall jitter of the system (see equation 1.7). Here F(t) is cumulative distribution function (CDF). We reach CTR from the photodetector SPTR just by simply taking the standard deviation of $f_{k:n}$. Assuming the case of identical detectors, (i.e. $\sigma(t_1) == \sigma(t_2)$), the timing resolution becomes:

$$CTR_{sigma} = \sigma_{f_{k:n}(t)}, \tag{1.10}$$

$$CTR_{fwhm} \approx 2.35\sqrt{2}\sigma_{f_{k:n}(t)}. \tag{1.11}$$

Again, equation 1.11 is only valid if $f_{k:n}$ can be approximated as a gaussian distribution, with sufficient accuracy. In Cherenkov's radiators, the number of photons which contain any information are very small, the SPTR influences the CTR more strongly, and can be approximated to

$$CTR_{fwhm} \cong \sqrt{2} \cdot SPTR_{fwhm}, \tag{1.12}$$

in the case of detecting, in average, a single-light photon per gamma-photon detection.

1.3.2 How to improve SPTR?

The previous section helped us understand the relationship between SPTR and CTR. Now, the question arises, how one goes on about improving the timing jitter of a SPAD?.

First, let us analyze the research done by Lacaita et. al. [18]. In this research letter it was found out that larger the SPAD area, larger it results in timing jitter. A comparator with a set threshold level can be used to detect the avalanche current of a SPAD. In a circular SPAD, for instance, the avalanche process occurs with a seed emerging at a random location and propagating to the center of the SPAD. The closer to the center the light-photon arrives in the SPAD, and where the seed will occur, the faster the pulse rises as the seed takes less time to reach the center. This results in a timing jitter from the fact that the light-photon can arrive at different places inside the SPAD leading to risetime variation. The arrival peak of the avalanche current shows statistical timing fluctuations based on where the seed occurs on the SPAD [18].

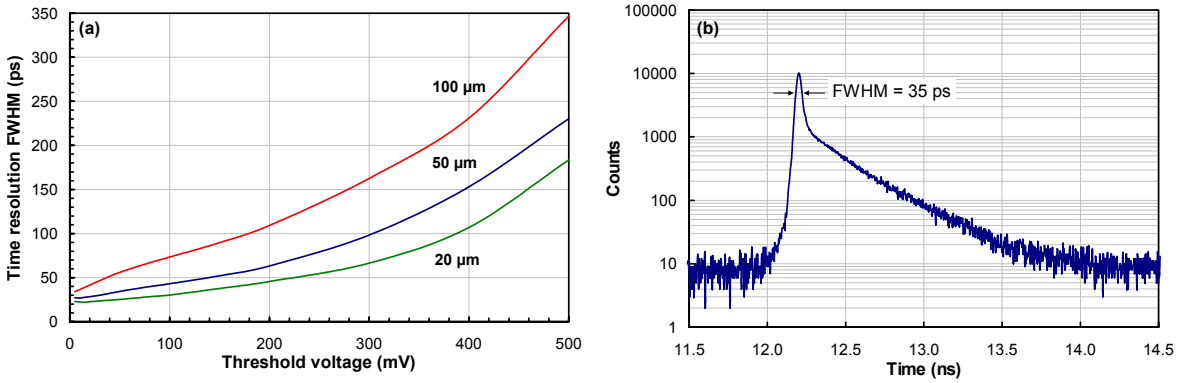

Figure 1.6: (a.)FWHM time resolution versus threshold of voltage comparator for different diameter SPADs as presented by Gullinatti et. al in [5], (b.) Time resolution test results by Gullinatti et. al [5] for 200μm SPAD

It has been observed that for large diameter SPADs, the worsening of the time resolution can be mitigated with detecting the avalanche at the stage when the multiplication of avalanche has not spread to a large area (see figure 1.6(a.)) [19]. All of these observations were analyzed by Gulinatti et. al. in his article [5][20], where they achieved SPAD timing resolution of 35 ps as shown in figure 1.6(b.), with an implementation and testing of a low-threshold comparator. However, the front end electronics (a low-threshold comparator in this case) were not on-chip. The foundation of implementing an on-pixel

low-threshold comparator is based on this analysis. Our goal in this thesis was aimed to improve timing jitter of a SPAD by implementing an in-pixel low-threshold comparator. The current state of the art D-SiPM arrays implement an inverter which has a threshold of around $V_{DD}/2$. With this implementation of a low-threshold comparator for every SPAD in D-SiPM, the timing resolution is expected to improve drastically.

1.4 Project Requirements and Specifications

The thesis specifications are outlined in two phases, first phase is for high speed comparator design for analog SiPM (A-SiPM). Here, the major challenge is the speed of the comparator, utilizing a mature (0.35 µm) CMOS process that was optimized for high PDP and low DCR SPADs. The second phase is the low-threshold in-pixel comparator designed for D-SiPMs. The specifications are even further laborious than the first phase of the design, mainly because of area and power restrictions in order to integrate a comparator per SPAD pixel.

1.4.1 Design I: High speed voltage Comparator for Integrated A-SiPM

The major goal of this design was to integrate the fast output of a specific A-SiPM with the comparator for a digital readout [21]. The SiPM models used for the design and simulations are SensL C-series and J-series 30035 with 35 µm pitch and optimized for TOF-PET imaging. The MicroFJ-30035-TSV (J-Series) A-SiPM provides photon detection efficiency (PDE) of approximately 50% over the wavelength of 420 nm with an excess bias voltage of 6 V. The total chip size for this project $3 \times 3\,\mathrm{mm}^2$. The requirement for the first phase design is summarized in the table below:

Technology	350 nm
Supply	3.3 V
Bandwidth	1-1.2 GHz
Comparator Maximum Power	< 1 mW
Comparator Threshold	To be able to detect 2.5% of the micro-cell of SiPM
Maximum Output Rise time	1 ns

Table 1.1: Specifications of the proposed Design I

1.4.2 Design II: In-pixel low-threshold Comparator for D-SiPM

The current state of the art for D-SiPM employs an inverter in the pixel itself, to detect a SPAD's avalanche, which has a minimal static power consumption. The major challenge to forthcoming in this design was to design a comparator for SPAD, which not just provides better performance than an inverter in terms of threshold (ideally less than 50 mV), but consumes a reduced static power and does not degrade the SPAD-pixel fill factor. The design requirements derived in this section, are derived from the Design Phase I. Henceforth, this design in essence is a continuation and improvement on the first design. Each specification requires its own justification. Keeping in mind the design

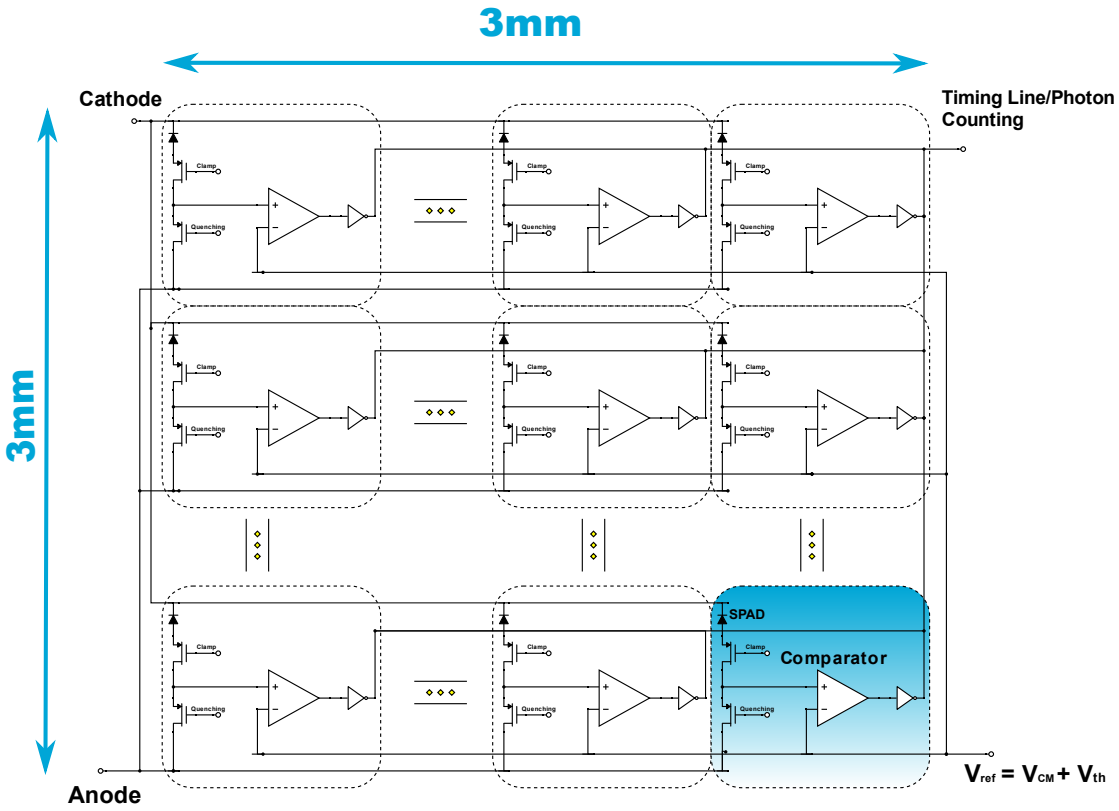

Figure 1.7: $3 \times 3\text{mm}^2$ proposed D-SiPM array

of a comparator per pixel is integrated for a $3 \times 3\,\text{mm}^2$ D-SiPM as shown in figure 1.7. Henceforth, we define them as follows:

1.4.2.1 Area

The pixel area is divided in the following table:

Area per pixel	$50 \times 50\,\mu\text{m}^2$
Pixel Active Area	70 %
Guard Ring + SPAD circuitry	20 %
Comparator Area	$10\,\% = 0.1 \cdot 50 \times 50\,\mu\text{m}^2 = 16 \times 16\,\mu\text{m}^2$

Table 1.2: Area Specifications

1.4.2.2 Power and Other Specifications

The total number of pixels in $3 \times 3\,\text{mm}^2$ area with a pitch of 50 µm is 3600 pixels (SPADs and comparators). The total comparator power budget for all of the pixels was defined as lower than 200 mW. Hence, the power of a single comparator becomes 55.5 µW.

The technology used in this design is 180 nm optimized for SPADs with a nominal supply voltage of 1.8 V. In this technology, we performed DC level simulation on NMOS

and PMOS and characterized the technology parameters V_{th}[mV] and μC_{ox}[µA V^{-2}].

The maximum current allowed for a comparator in a pixel resulted as 30 µA. Furthermore, we define the input common mode maximum (ICMR+), input common mode minimum (ICMR-), and input common mode range (ICMR). For our design using thin oxide MOSFET, we use ICMR+ of 900 mV, ICMR- of 730 mV, leading to the ICMR of 170 mV. Similarly for thick oxide MOSFET, we have, ICMR- of 1 V, ICMR+ of 1.1 V, leading to the ICMR of 100 mV.

Design I: High speed voltage Comparator for Integrated A-SiPM

2

2.1 Brief Overview of A-SiPMs and D-SiPMs

The output of a SPAD in gieger mode operation results in a binary on/off state. As shown in figure 2.1, multiple SPADs are joined together in the form of a matrix. Each SPAD contains its own separate quenching resistor. Multiple SPAD outputs are added together in the case of A-SiPM. In the case of SensL, a fast output is also available as it provides high speed pulse, which is conventionally processed off-chip using a comparator and TDC, to extract the timing information. The pulse generated in the conventional cathode/anode provides a magnitude related to the number of SPADs fired due to the scintillation activity.

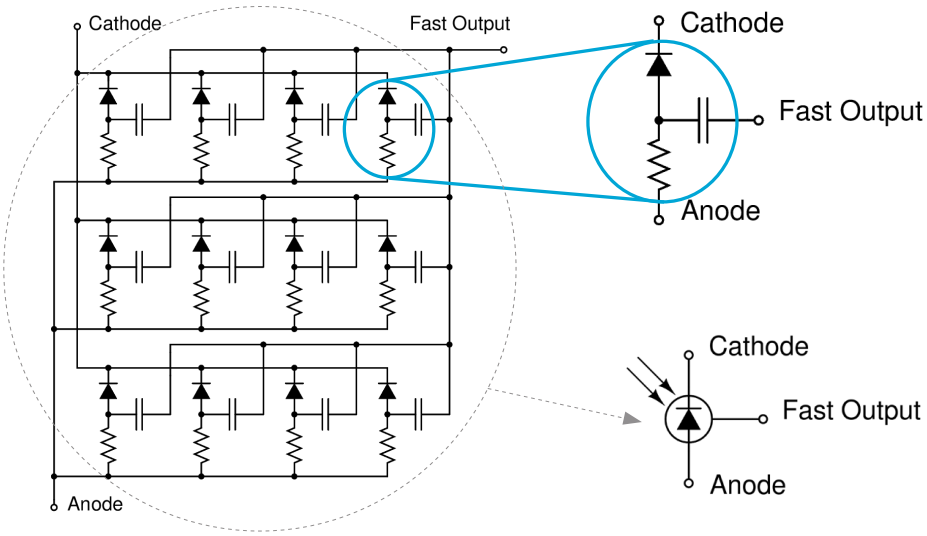

Figure 2.1: SensL's A-SiPM with fast output [6]

D-SiPM, on the other hand, employs a buffer or a comparator for each SPAD and process the single SPAD data on-chip using a TDC and pixel counters. As seen from figure 1.7, each SPAD pixel in D-SiPM matrix produces its own digital output based on the firing of the SPAD. In an extreme scenario, we can implement a TDC for every SPAD pixel, however it affects the fill factor of the pixel and drastically increases the power consumption. A more practical approach is to share number of SPADs per TDC to have a trade off between fill factor and better statistical information, as implemented by S. Mandai et al. in [4].

2.2 Comparator Fundamentals

This section explains comparator basics before we move onto the specific design. The introduction explained here is inspired from [22]. A comparator in its simplest form can be described as a 1-bit analog-to-digital converter(ADC). The symbol diagram of comparator is shown in figure 2.2, with positive input as v_p and negative input as v_n with an output as v_o. Let us define the maximum and minimum output levels achieved as V_{OH} and V_{OL}. For an ideal comparator, v_o is V_{OH} if $v_p - v_n$ is positive even for an infinitely small positive value. Similarly vice-versa is true, as v_o is V_{OL} if $v_p - v_n$ is negative, irrespective how smaller is the magnitude. The gain, A_v from the input difference to the output is infinitely large, which means for the ideal case the input difference must be zero, and is given by

$$A_v = \frac{V_{OH} - V_{OL}}{v_p - v_n}. \tag{2.1}$$

In a real comparator with finite gain, the input difference, which is defined as the minimum input difference required for the A_v to be sufficient enough for the output to be able to switch, is called the resolution of the comparator. Furthermore, a comparator is broadly divided into two major types, which are asynchronous comparators, without a clock signal, and synchronous comparators, where a clock signal is necessary to make a decision at the output. For our design, we restrict ourselves to asynchronous comparators, as in our application of TOF-PET imaging the pulse provided to the comparator input is asynchronous.

2.2.1 Threshold Setting in a Comparator

In both of our comparator designs for A-SiPM and D-SiPM, we provide a fix threshold voltage (V_{th}) at the non inverting terminal (v_p) of the comparator, and an input signal (V_{in}) from SPAD at the inverting terminal (v_n). Both the input and threshold are biased above the common mode (CM) voltage instead of ground, which means that

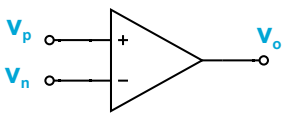

Figure 2.2: Comparator symbol

$$v_p = V_{ref} = V_{th} + V_{CM}, \tag{2.2}$$
$$v_n = V_{in} + V_{CM}, \tag{2.3}$$

where V_{ref} is known as reference voltage. V_{CM} is required to keep all of the transistors in saturation region and V_{th} is the threshold voltage, which is a minimum value that the noise level allows to set. Also, the threshold is defined by the resolution of the comparator. The threshold of a comparator can be set using analysis performed by Rivetti et al. [23], which can be summarized below based on the noise frequency. Noise might cause spurious hits at the comparator output even when the input has not been applied. If we set threshold to zero, we can find that the frequency of zero crossing due

to noise f_{n0} is given as

$$f_{n0}^2 = 4\frac{\int_0^\infty f^2 H(f)^2 df}{\int_0^\infty H(f)^2 df}. \tag{2.4}$$

Where $H(f)$ is the transfer function of the comparator in frequency domain [24, 25]. Henceforth, due to the presence of noise in a zero threshold environment, the baseline level is crossed f_{n0} times per second in either direction, which implies the output rectangular pulse frequency under the influence of noise is $f_{n0}/2$. Furthermore, for the case of a non-zero threshold voltage, V_{th} and a certain noise voltage, V_n the frequency of noise hits f_n becomes a function of the zero crossing frequency f_{n0}. And, it is given by

$$f_n = \frac{1}{2} f_{n0} e^{-\frac{V_{th}^2}{2V_n^2}}. \tag{2.5}$$

Finally as discussed in [23], we can also define noise hits in a given time frame as $f_n \Delta t$ as noise occupancy. For asynchronous systems, such as the design presented in this thesis, Δt can be defined by the unit time interval or the reciprocal of bandwidth (BW). From equation 2.5 we can solve for V_{th}/V_n as,

$$\frac{V_{th}}{V_n} = \sqrt{-2ln f_n 2\pi\tau}, \tag{2.6}$$

which implies that there is a trade-off between bandwidth and frequency of noise hits. In other words, if we increase the BW of the system, or reduce the τ, for the same V_{th}/V_n, we will have more noise hits reaching at the output.

2.3 A-SiPM fast output pulse modelling at the Comparator's input

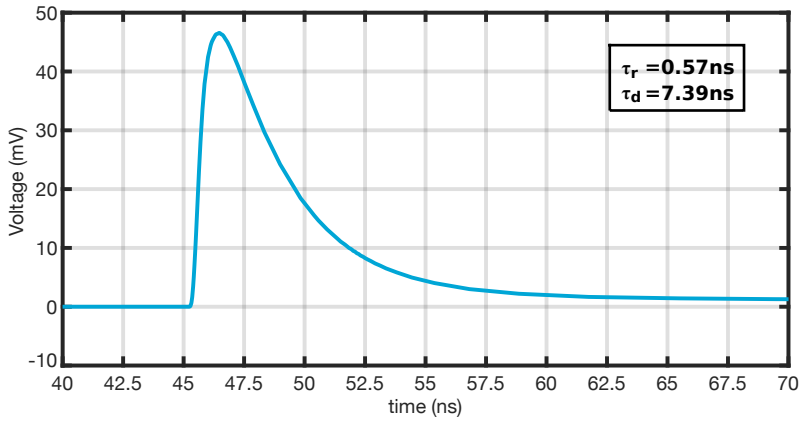

Figure 2.3: J-Series SPICE model pulse with 2.5% microcells fired[7]

We started the design analysis by modelling the fast output pulse for comparator's input. The specifications of input rise time, input capacitance, and AC coupling can be

extrapolated from such analysis. The information provided in SensL's MicroFJ-30035-TSV (J-Series) datasheet gives us the pulse shape for a specific number of microcells (SPADs) fired, which is 2.5 % , that is 142 microcells out of 5676 microcells in total. The SPICE modelled pulse used as an input of the comparator is shown in figure 2.3.

Furthermore, the output capacitance of MicroFJ-30035-TSV SiPM array's fast output is 40 pF. The achieved fast output modelled pulse, which serves as the input of the comparator, is necessary for the design of a comparator which meets the specifications accurately.

2.4 Asynchronous Comparator Design based on Complementary Self Biased Differential Amplifier

The selected comparator architecture is a modified design presented in [26], which is based on the foundation of complementary self biased differential amplifier (CSDA). This concept was introduced by Bazes et. al. [8] and the self-biased differential amplifier (SDA) was introduced in [27].

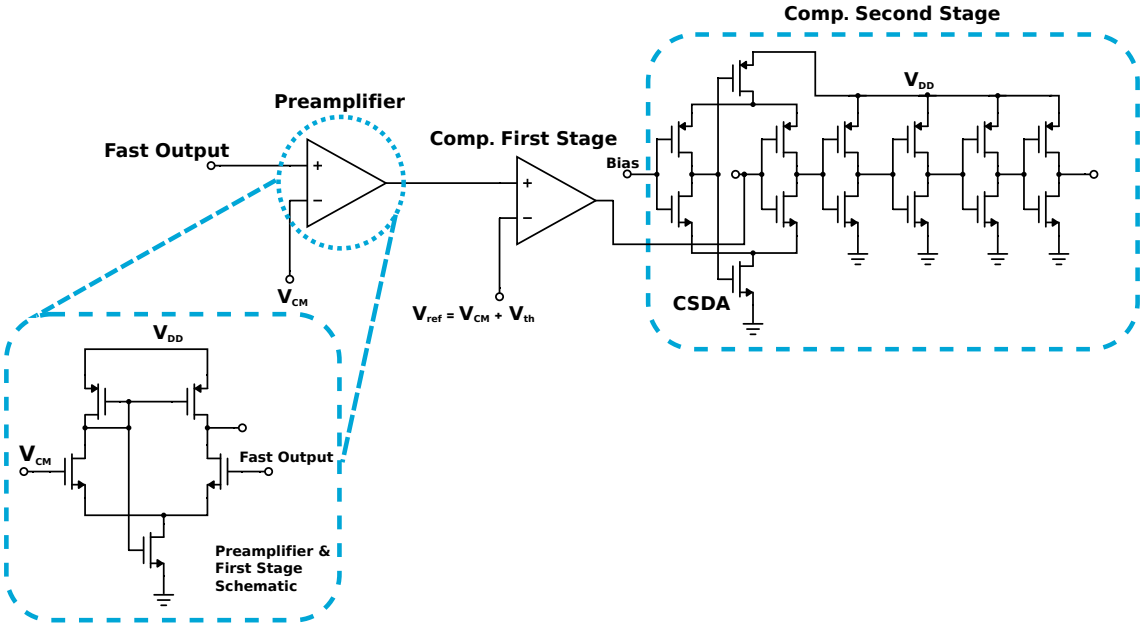

Figure 2.4: Proposed design of high-speed asynchronous comparator

A preamplifier stage was first added to increase the absolute threshold resolution with respect to the non-amplified input signal range. Furthermore, the preamplifier stage helps in improving the slew rate. The design used for the preamplifier stage and first stage of the comparator is SDA as shown in figure 2.4 [27]. We did not implement the first stage using CSDA directly as it offers better performance in output buffer stage, as will be discussed later. Furthermore the CSDA common mode must be biased at $V_{DD}/2$ for optimal operation. The SDA allows near threshold voltage (NTV) CM

setting, hence it is implemented in the first stage. The design of the self biased and external biased differential amplifier is discussed in the next chapter. Here, we move forward with the second stage analysis, which is a CSDA stage followed by buffers.

2.4.1 Complementary Self Biased Differential Amplifier

A CSDA can be derived and understood by considering two complementary differential amplifiers with similar sizes but opposite in terms of NMOS and PMOS. Also, the active loads are removed and connected together as shown in figure 2.5(a.)(b.). Such an amplifier contains two equal bias voltages for the current sources M_3 and M_4, and producing same bias voltage for both NMOS and PMOS is a critical task. Since any shift in bias voltage would result in unequal current. This requirement is met by removing the external biases and connecting the gates of M_3 and M_4 to one of the drains of M_{1A} and M_{2A} after connecting them together as shown in figure 2.5(c.). The self-bias structure forms a negative feedback loop which stabilizes the bias voltage.

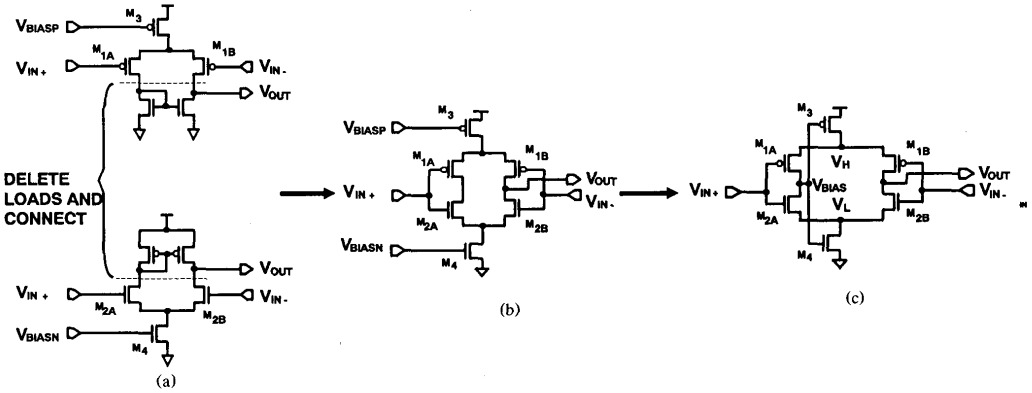

Figure 2.5: (a.) Deleting active loads of two complementary differential amplifiers, (b.) Merging the two differential amplifiers with two similar bias voltages, (c.) Constructing a self bias by removing the external bias [8]

As $V_{in,p}$ increases the node V_{bias} starts to decrease, as Q1 turns on and Q2 turns off, and we see an inverter type phenomenon as shown in figure 2.6(a.). Consequently, Q5 turns on and Q6 turns off, which allows current path from V_{DD} till $V_{lin,p}$. At the same time, as $V_{in,n}$ starts to decrease, Q4 turns on and Q3 turns off. Henceforth, the current can sink via V_{DD} to the load C_L. Thus, the output node reaches to $V_{DD} - I_{on,p}(R_{on,Q4} + R_{on,Q5})$. This analysis proves that CSDA provides large current sourcing and sinking property without the need of high quiescent current. Henceforth, the speed of switching is high. Furthermore, the Q5 and Q6 are sized to be in the triode region, so the available swing at the output is high[8].

In figure 2.6(b.), we see the small signal model at low frequencies. In [28] gain equation for CSDA has been derived in the next paragraph.

As Q5 and Q6 are operating in triode region, we can model them as a current source and resistor in parallel. Also, for a simplified gain analysis let us consider $V_{lin,p}$ and $V_{lin,n}$

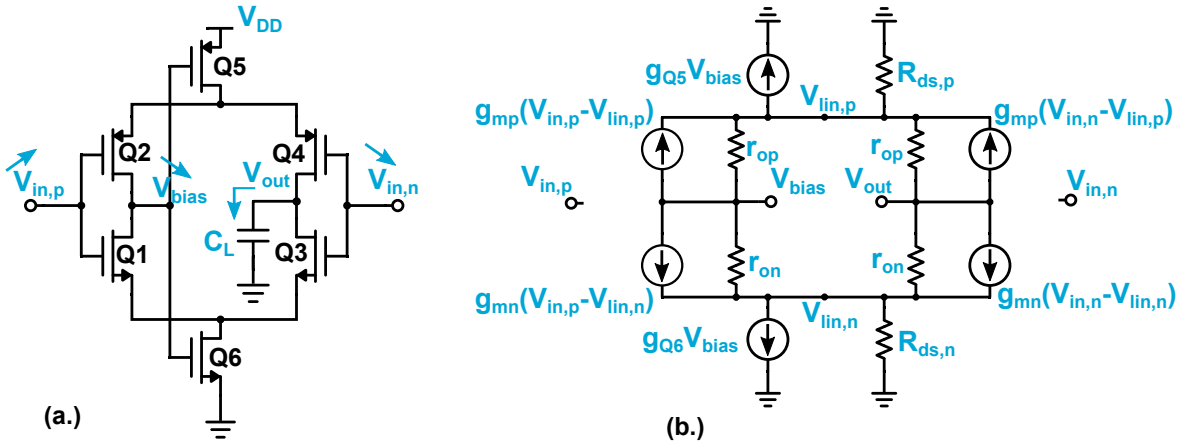

Figure 2.6: (a.) CSDA architecture, (b.) small signal analysis of CSDA

close to zero, as these operate in triode region and neglect the drain-source resistance ($R_{ds,p} = R_{ds,n} = 0$) of Q5 and Q6. The gain is given by

$$A_v = \frac{V_{out}}{V_{in,p} - V_{in,n}} = \frac{g_{mn} + g_{mp}}{g_{on} + g_{op}}, \tag{2.7}$$

$$V_{bias} = 0 \tag{2.8}$$

[28]. Furthermore, with a more accurate output voltage analysis performed by [28], we predicted that more gain is reaching from $V_{in,n}$ to V_{out} compared to $V_{in,p}$. Henceforth, CSDA resulted more suitable for output buffer stage, as suggested by [22] as well. Such analysis also confirms our choice of using CSDA to produce better performance as the CSDA is part of the second stage in our design as shown in figure 2.4.

[

Design II: In pixel low-threshold Comparator for D-SiPM

3

3.1 Pixel Farm

As discussed in chapter 1 we can improve the timing jitter of a SPAD by implementing a low-threshold comparator inside a pixel. In this regard, we provide a pixel farm design, which is implemented in $1 \times 1\,\mathrm{mm}^2$ 180 nm CMOS technology optimized for SPADs. The floorplan of the chip is shown in figure 3.1. The top and bottom area are kept for 48 bonding PADs, that provide voltage supplies and pixel biasing. The middle region has been assigned to the pixel farm, which contains twelve different types of pixels and six different types of comparator structures. As shown in figure 3.2 we use six comparator architectures, each of which is implemented for two types of SPAD, with diameter of 12 µm and 40 µm active area , making a total of 12 pixels. The aim to use multiple architectures in the pixel farm is to test the timing resolution of multiple comparator structures, in order to obtain the optimum SPAD-pixel low-threshold readout circuit.

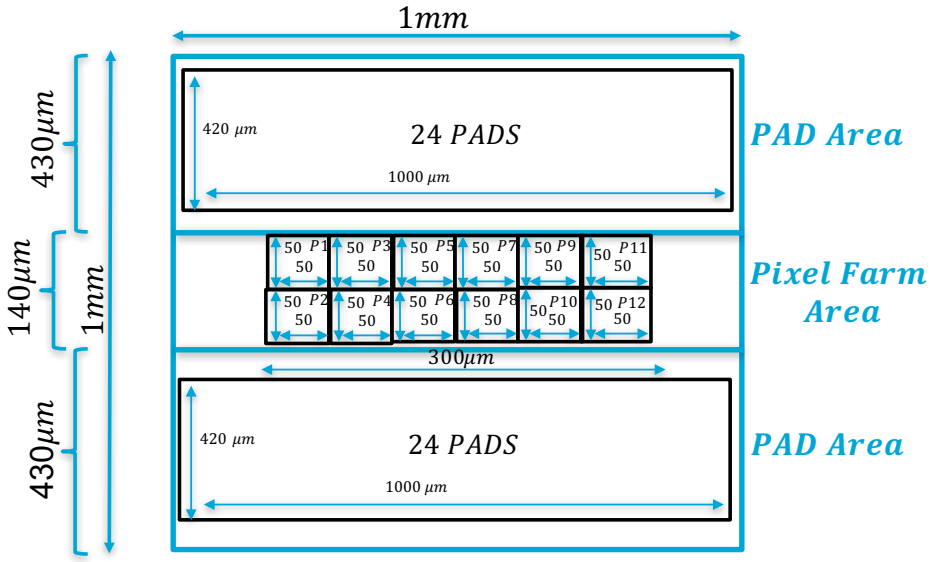

Figure 3.1: Floorplan in $1 \times 1\mathrm{mm}^2$ designed chip

When the SPAD fires, the anode can reach a high voltage value, which might cross the gate's breakdown voltage of a thin oxide NMOS. Henceforth, we implemented a thick oxide NMOS based differential amplifier in our design as shown in figure 3.2(c.) and 3.2(d.), in order to prevent damage to the pixels. In this design, we also implemented self bias differential amplifier architectures, which reduces the number of pads require for biasing the tail current source. We also designed comparator variations, in order

to study the impact of several parts of the circuits (see figure 3.2). The variations are explained in detail when discussing the results.

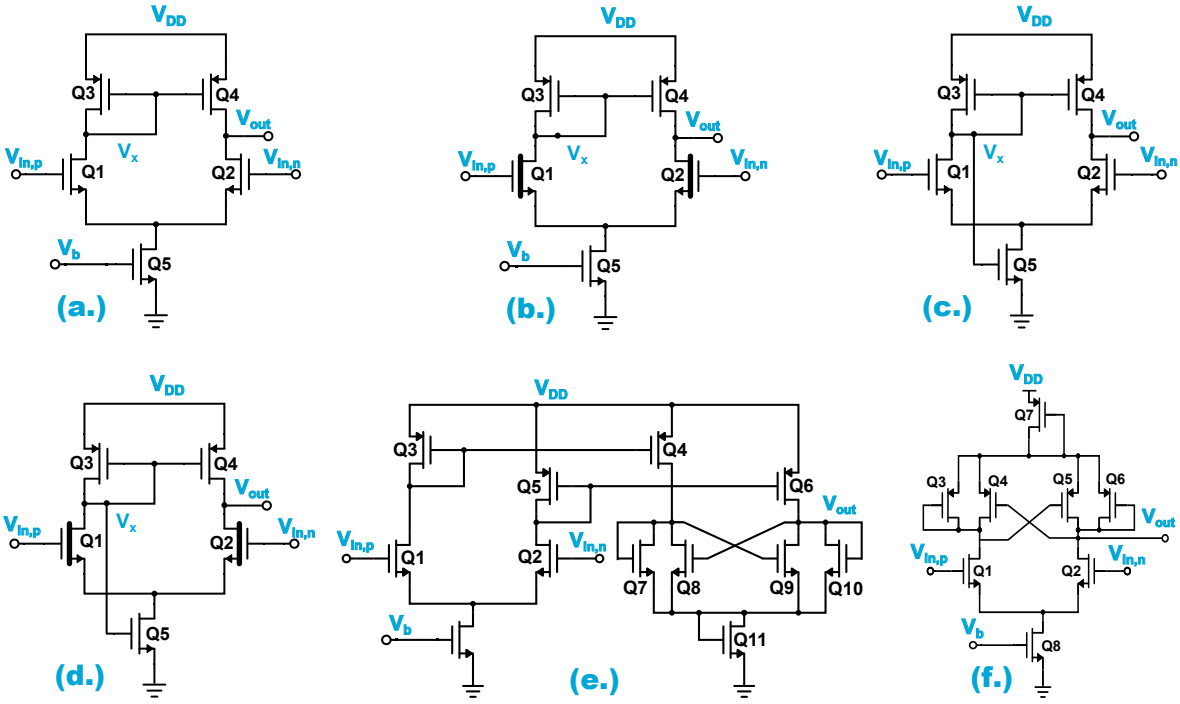

Figure 3.2: (a.) differential amplifier with external bias, (b.) differential amplifier with NMOS thick oxide and external bias, (c.) differential amplifier with self bias. (d.) Differential amplifier with thick oxide NMOS and self bias, (e.) high speed positive feedback comparator architecture, (f.) high speed positive feedback comparator architecture with alternate design technique

3.2 Pixel Architecture

As shown in figure 3.3, we have a cascode of quenching and recharge, with transistors Q1 and Q2 as proposed by [29]. The transistor Q2 acts as a passive quenching and recharge transistor, as a voltage applied to the transistor renders it as a high-impedence element. Both Q1 and Q2 are thick oxide transistors. The source of Q1, or input of the comparator, achieves a maximum voltage of $V_{clamp} - V_{th,Q1}$, where we have chosen V_{clamp} as 1.8 V with an added CM voltage. As the Anode of the pixel is biased at the CM voltage, the comparator input is also effectively biased at the CM.

3.2.1 Common Mode biasing of the in-pixel comparator

In order to optimize the fill-factor, the common mode of the comparator is input from the SPAD circuitry by directly connecting the input of the comparator with the anode of the SPAD. The cathode and the anode of the pixel are biased by taking the input CM of comparator into consideration. Let us consider a voltage smaller than breakdown

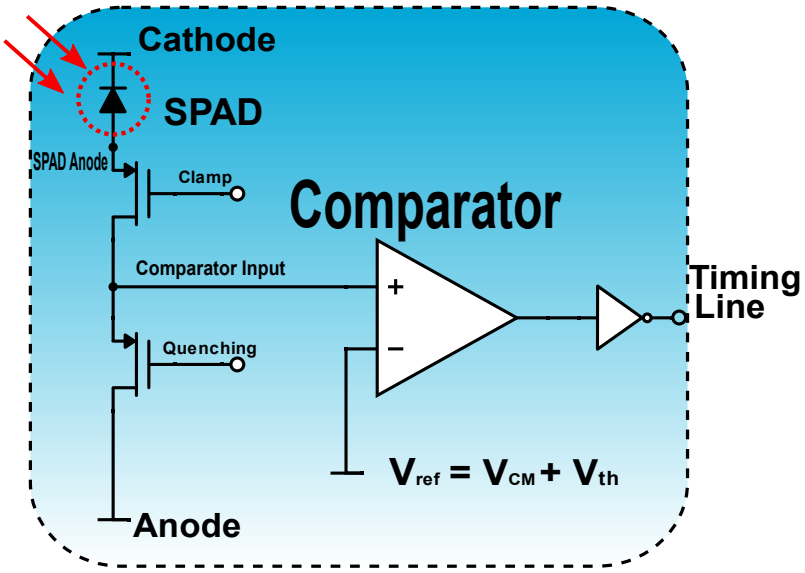

Figure 3.3: Pixel Architecture

voltage called V_{bd}. The SPAD is biased at an operating voltage (V_{op}), in which its excess bias voltage (V_e) is higher than the V_{bd}. At cathode, we have that

$$V_{cathode} = V_e + V_{bd} + V_{CM}, \tag{3.1}$$

$$V_{anode} = V_{CM}. \tag{3.2}$$

Overall, we have that

$$V_{cathode} - V_{anode} = V_e + V_{bd}. \tag{3.3}$$

The comparator's common mode can easily be biased by directly connecting it to the anode voltage of the SPAD. In this case, the input is connected after the pulse from SPAD anode is clamped using a clamping transistor Q1 as shown in figure 3.3.

As seen from figure 3.4, the SPAD anode and clamp, both are referred to a CM of 750 mV instead of ground. Also, we can observe that the SPAD anode has a maximum voltage of 4 V, which is very close to the breakdown voltage of a gate of a thin oxide NMOS transistor. The input of the comparator is directly connected to the gate of NMOS transistor (see figure 3.2. As a result we connected a clamping transistor, which decreases the maximum voltage to $V_{clamp} - V_{th,Q1}$. Henceforth, we mitigate the concern regarding crossing the breakdown voltage of a transistor and permanent damage of the comparator circuitry.

3.2.2 SPAD Architecture implementation

In this design, we use two different diameter circular SPAD with active area diameter of 12 µm and 40 µm in order to study the timing performance for different diameters. The SPAD has been designed by C. Veerappan et. al. [9]. The cross section area of such a design can be seen in figure 3.5.

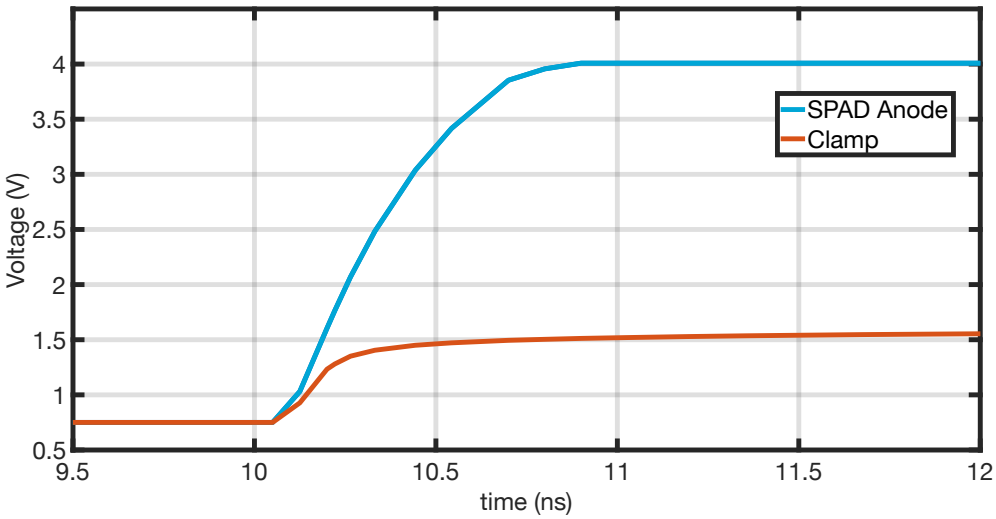

Figure 3.4: SPAD Model pulses transient response

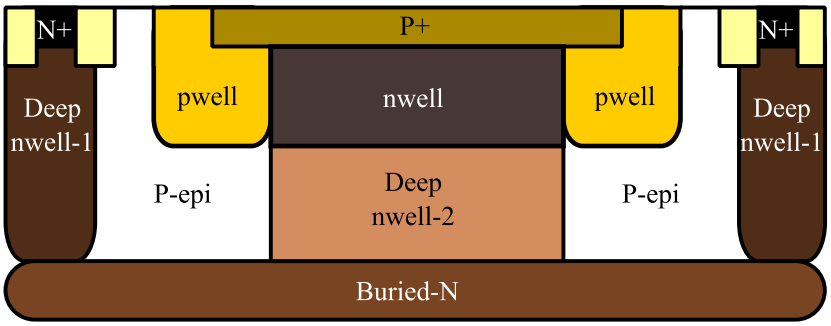

Figure 3.5: SPAD cross section view[9]

3.3 Design of a differential Amplifier

For the design of an asynchronous comparator, we begin with the design of a differential amplifier with active load, which has a current source of NMOS biased in the saturation region, as shown in figure 3.6. This circuit amplifies the difference of its two inputs and provides an output, which are single ended in our design. In this section, first, the large-signal analysis is explained. Next, the frequency analysis focused on high-speed design based on gain bandwidth product (GBW) calculations is introduced. The following analysis has been inspired from [30] and [31].

3.3.1 Large Signal Analysis

First, we assume that $V_{in,p}$ is much lower than $V_{in,n}$ which implies that Q1 is turned off (see figure 3.6). As a consequence, no current can flow through Q3 and Q1. As the current mirror is effectively turned off in such situation Q4 remains switched off as well.

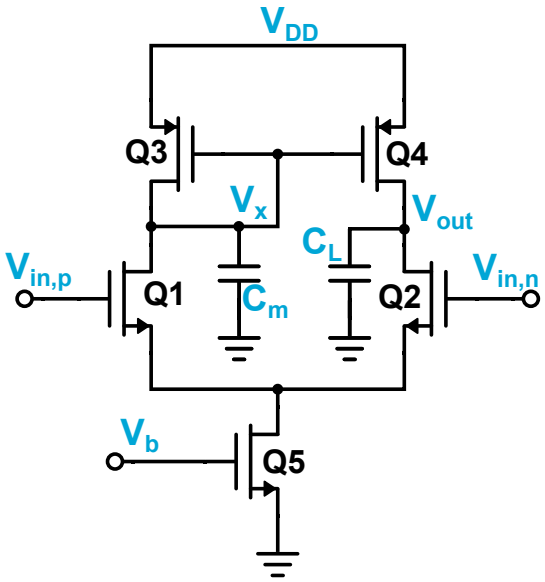

Figure 3.6: Differential pair amplifier architecture with parasitic capacitance C_m and load capacitance C_L

In such a case, Q2 and Q5 remain in deep triode region, as for Q2 and Q5,

$$V_{GS} - V_{th} > V_{DS}, \tag{3.4}$$
$$V_{in,n} > V_x + V_{th}. \tag{3.5}$$

As the $V_{in,p}$ starts approaching $V_{in,n}$, the Q1 turns on and current mirror activates. This is the linear region of operation of the differential amplifier, where the transistors are in saturation region, and the small voltage difference at the input directly affects the output voltage. This region of operation is high gain region.

As $V_{in,p}$ increases, and crosses $V_{in,n}$ and becomes larger, the current through Q1 increases and as a consequence the current in Q3 and Q4 increases. As a result, the current in Q2 starts to decrease, which causes Q4 to go into the triode region as eventually Q2 turns off. In this case, V_{out} is effectively equal to V_{DD}.

3.3.2 Small signal DC analysis and frequency response of Differential Amplifier

To reach to the frequency response of the differential amplifier, we first look at its DC gain. In this analysis, we consider that the differential amplifier is working in the linear region, or in the small signal region. Simplifying the architecture, we can replace Q4 in figure 3.7 with its output impedance r_{o4}. Similarly, we perform the same replacement Q2. The total output resistance, hence, can be summarized as,

$$R_{out} \approx r_{o2} || r_{o4}. \tag{3.6}$$

In the steady state, where all transistors are in saturation region, Q1 and Q2 are same in size, and so is Q3 and Q4, we have equal current flowing in two branches. The current

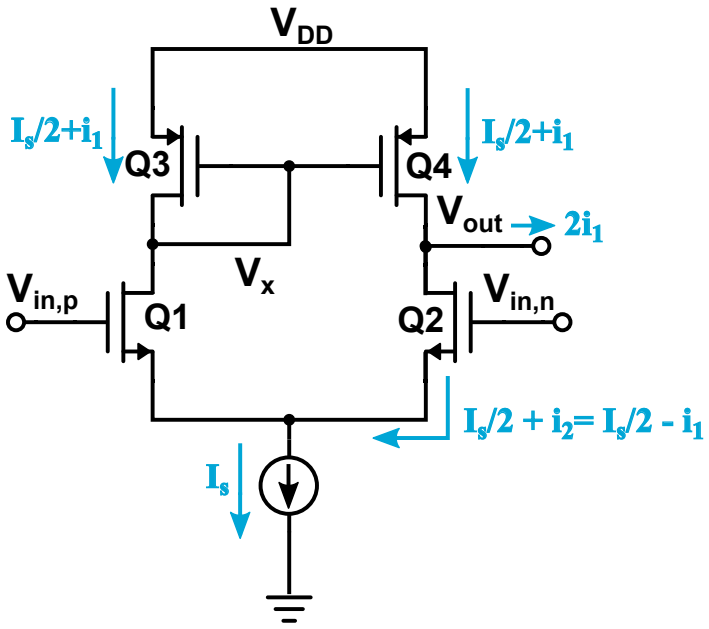

Figure 3.7: Differential amplifier DC current distribution

in this case is $I_s/2$ (see figure 3.7). Now, a small signal current i_1 is flowing through Q3 due to a differential change in voltage at the inputs. The same current is mirrored in Q4. The current flowing from Q2 is, say, i_2, which will be the difference of i_1 and current flowing through the output resistance (R_{out}). Hence, the small signal DC output voltage is

$$v_{out} = (i_1 - i_2) \cdot (r_{o2}||r_{o4}). \tag{3.7}$$

Due to the presence of current source at the tail of the differential amplifier, we have, $i_1 = -i_2 = i$. Hence,

$$v_{out} = 2 \cdot i \cdot (r_{o2}||r_{o4}). \tag{3.8}$$

The differential DC gain is, thus,

$$A_d = \frac{v_{out}}{v_{in,p} - v_{in,n}} = \frac{i_1 - i_2}{v_{in,p} - v_{in,n}} \cdot (r_{o2}||r_{o4}) = g_{m1,2} \cdot (r_{o2}||r_{o4}), \tag{3.9}$$

where $v_{in,p}$ and $v_{in,n}$ are small signal inputs in the linear region of the transistors.

By adding C_L, from the gain in equation 3.9, we can derive the transfer function expression of differential amplifier in the frequency domain. Such a transfer function helps us in estimating the noise frequency at output of the comparator for a fixed threshold. As there are two paths to the output, namely, Q1 $\rightarrow$ Q2 and Q1 $\rightarrow$ Q3 $\rightarrow$ Q4. The path Q1 $\rightarrow$ Q2 to the output contains one pole due to the presence of load capacitance, C_L. The value of such a pole is,

$$\omega_{p1} \approx \frac{1}{(r_{o2}||r_{o4}) \cdot C_L}. \tag{3.10}$$

At node V_x we can see a presence of a capacitance, C_m (see figure 3.6). Such capacitance contributes at high frequencies in the form of a pole and a zero. These pole and zero occur at the respective frequencies of,

$$\omega_{p2} \approx \frac{g_{mp}}{C_m}, \tag{3.11}$$

$$\omega_z \approx \frac{2g_{mp}}{C_m}. \tag{3.12}$$

The frequency response can be expressed as,

$$\frac{V_{out}}{V_{in}} = \frac{A_d(2 + s/\omega_{p2})}{(1 + s/\omega_{p1})(1 + s/\omega_{p2})}. \tag{3.13}$$

3.4 Calculations of the Load Capacitance of the Comparator

The comparator drives an inverter before input a timing line of the D-SiPM. In the utilized CMOS technology, we estimated the load capacitance of the comparator by calculating the input capacitance of a minimum size inverter. In the CMOS 180 nm technology used for this design, the minimum sized inverter has a width of 280 nm for NMOS. As PMOS provides about half the mobility it requires almost double width for similar functionality. The calculations for the input capacitance are implemented using [10].

The propagation delay can be divided into two types, which are t_{pHL} and t_{pLH}. t_{pHL} is time between 50 % of the input rising pulse and the 50 % of the output falling pulse. Similarly, t_{pLH} is the time between 50 % of the input falling pulse and the 50 % of the output rising pulse. The propagation delay results using 2 fF and 4 fF load at the inverter output and 10 ps rise time and fall time input square pulse (see figure 3.8(a.)) are summarized in the table 3.1.

Table 3.1: Propagation delay for minimum sized inverter for 2 fF and 4 fF load

t_p	$C_L = 2fF$	$C_L = 4fF$
t_{pHL}	40.392 ps	58.883 ps
t_{pLH}	61.2 ps	91.1 ps

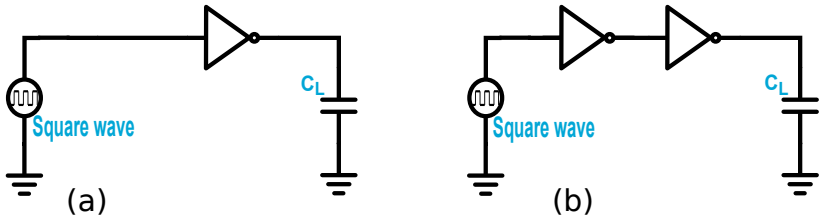

Figure 3.8: (a.) Inverter with load of C_L, (b.) Inverter with load connected to second inverter

Approximating $R_{on,nmos}$ and $R_{on,pmos}$ equal to R_{eq}, we can estimate the propagation

delay as

$$t_p = 0.69 \cdot R_{eq} \cdot C_{eq}. \tag{3.14}$$

Hence, from table 3.1, the average of t_{pHL} and t_{pLH} was calculated to find the propagation delay. Furthermore, the output equivalent capacitance is simplified as the sum of the drain capacitances of the MOSFETs (C_d) and the load capacitance. Therefore,

$$\frac{40.392\,\text{ps} + 61.2\,\text{ps}}{2} = 0.69 \cdot R_{eq} \cdot (2\,\text{fF} + C_d), \tag{3.15}$$

$$\frac{61.2\,\text{ps} + 91.1\,\text{ps}}{2} = 0.69 \cdot R_{eq} \cdot (2\,\text{fF} + C_d), \tag{3.16}$$

which results in a R_{eq} of 17.533 kΩ and a C_d of 2.2 fF. Furthermore, when the same minimum sized inverter is connected to another minimum sized inverter as the load, as shown in fig. 3.8(b), the output capacitance (C_{eq}) of the first inverter is,

$$C_{eq} = C_d + C_g. \tag{3.17}$$

Where C_g is the input capacitance of the second inverter. The simulation results shows,

$$t_{pHL} = 32.404\,\text{ps}, \tag{3.18}$$

$$t_{pLH} = 48\,\text{ps}. \tag{3.19}$$

Henceforth,

$$\frac{32.404\,\text{ps} + 48\,\text{ps}}{2} = 0.69 \cdot 17.533\,\text{k}\Omega \cdot (C_g + 2.2\,\text{fF}), \tag{3.20}$$

$$C_g = 1.12\,\text{fF}. \tag{3.21}$$

The performance of the comparator is critical, as it should work for even a stringent specification. Due to such considerations, we consider the load of the comparator to be 10 fF, higher than 1.12 fF input capacitance of the inverter, which serves as the load of the pixel comparator .

3.5 Calculations Involving Accurate Output Capacitance

From the previous section, we have defined a load capacitance (C_L) of 10 fF. Next, we consider extra parasitic capacitances of the MOSFETs. The motivation behind determining the parasitic capacitances at the output is because it helps to size the Q1 and Q2 accurately based on the gain bandwidth product (GBW) defined in the design (see figure 3.6). As shown in figure 3.9, we have following capacitances, which add to the effective capacitance in a MOSFET,

$$C_{gd} = C_{ovd} = C_{GDO} + C_{GCD}, \tag{3.22}$$

where C_{gd} is effective capacitance from gate to drain. The presence of C_{gd} is due to an overlap capacitance (C_{GDO}) and a channel capacitance (C_{GCD}) between gate to drain.

The overlap capacitance is due to extra diffusion of channel from the gate to the drain or source region. This extra distance extended by the gate oxide into drain or source region is called lateral diffusion, x_d, and it contributes to the overlap capacitance. The overlap capacitance is formulated as

$$C_{GDO} = C_{GSO} = C_{OL}W = C_{ox}x_dW. \tag{3.23}$$

Where, C_{ox} is equal to $\epsilon_o \cdot \epsilon_r / t_{ox}$, and it is the oxide capacitance per unit area. W is the width of the MOSFET and C_{OL} is the overlap capacitance per unit transistor width.

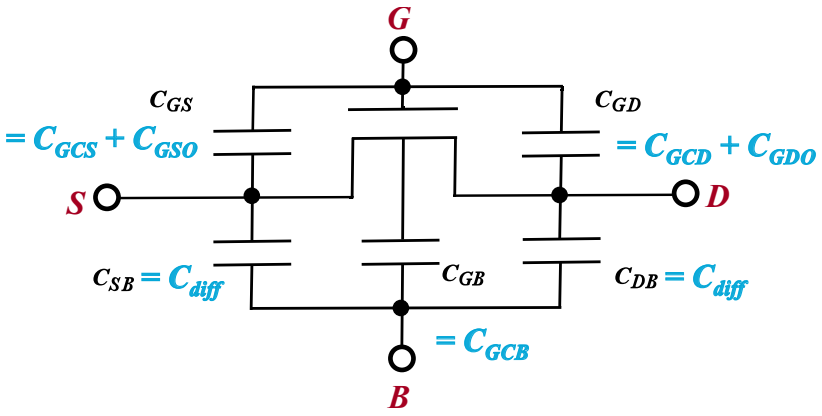

Figure 3.9: MOSFET capacitances[10]

The second capacitance is the channel capacitance between gate and drain. The channel capacitance is dependent on device's region of operation. This is due to the fact that in different region of operation, the channel has different properties. As for our design, the MOSFET will operate in saturation region, where the $C_{GCD} = 0$, therefore,

$$C_{gd} = C_{ovd} = C_{GDO} = C_{OL}W = C_{ox}x_dW. \tag{3.24}$$

In the saturation region of operation the channel capacitance between gate and source, C_{GCS}, is,

$$C_{GCS} = \frac{2}{3} \cdot W \cdot L \cdot C_{ox}. \tag{3.25}$$

Hence, the capacitance between gate to source becomes,

$$G_{gs} = C_{GCS} + C_{GSO} = C_{ox}.x_d.W + \frac{2}{3}.W.L.C_{ox}. \tag{3.26}$$

Moving forward with the consequential parasitic capacitances, we looked into the drain to bulk capacitance, which occurs predominantly due to the presence of junction capacitance, also know as diffusion capacitance (C_{diff}). The junction capacitance occurs due to the presence of reverse bias pn junction between bulk to drain or bulk to source. In our calculations for output parasitic capacitance, we are only interested in the bulk to drain junction capacitance. The junction capacitance is divided in following two components, namely,

- The bottom plate junction capacitance, which results from the bottom plate area of the drain junction with the substrate and it is expressed as,

$$C_{bottom} = C_j.Area = C_j.W.L_d, \tag{3.27}$$

were, C_j is junction capacitance per unit area and L_d is the length of drain junction.

- The side wall junction capacitance, which is due to the drain region and p^+ channel stop implant, is summarized as,

$$C_{sw} = C_{jsw} \cdot Perimeter = C_{jsw} \cdot x_j \cdot (2L_d + W). \tag{3.28}$$

Where C_{jsw} and x_j are capacitance per unit perimeter and junction depth, respectively.

The whole junction capacitance can be summarized as,

$$C_{db} = C_{diff} = C_j.W.L_d + C_{jsw}.x_j.(2L_d + W) \approx \alpha_1.W. \tag{3.29}$$

We simulated the DC analysis of the differential amplifier (see figure 3.6) with different values of $W_{1,2}$ and $W_{3,4}$ and L of 1 µm. The simlulation results provided different parasitic capacitances, from which C_{OL} and α_1 were constructed using equation 3.24 and 3.29(see table 3.5). The results of the DC simulations are given in table 3.5.

$\boldsymbol{W_{Q1,Q2}}$[µm]	$\boldsymbol{C_{gd}}$[fF]	$\boldsymbol{C_{db}}$[fF]	$\boldsymbol{C_{OL} = C_{gd}/W_{Q1,Q2}}$[fF µm^{-1}]	$\boldsymbol{\alpha_1 = C_{db}/W_{Q1,Q2}}$[fF µm^{-1}]
3.00	5.32	1.50	1.78	0.50
4.00	6.68	1.89	1.67	0.47
5.00	7.97	2.25	1.59	0.45

Table 3.2: NMOS C_{OL} and α_1 simulation results

From the simulation results we concluded that,

$$C_{OL} \approx 1.65\,\text{fF}\,\mu\text{m}^{-1}, \tag{3.30}$$
$$\alpha_1 \approx 0.47\,\text{fF}\,\mu\text{m}^{-1}. \tag{3.31}$$

Finally, from the knowledge about the parasitic capacitances, we constructed the output capacitance and also the mirror capacitance at the nodes V_{out} and V_x, respectively. The accurate calculation results are

$$C_{out} = C_L + C_{db,Q4} + C_{db,Q2} + C_{gd,Q4} + C_{gd,Q2}, \tag{3.32}$$
$$C_{out} = 10\,\text{fF} + C_{db,Q4} + \alpha_1.W_{Q2} + C_{gd,Q4} + C_{OL}.W_{Q2}, \tag{3.33}$$
$$C_{out} = 10\,\text{fF} + C_{db,Q4} + (0.47).W_{Q2} + C_{gd,Q4} + (1.65).W_{Q2}, \tag{3.34}$$
$$C_{out} = 10\,\text{fF} + C_{db,Q4} + C_{gd,Q4} + (2.1).W_{Q2}, \tag{3.35}$$

$$C_m = C_{gd,Q1} + C_{db,Q1} + C_{db,Q3} + C_{gs,Q3} + C_{gs,Q4}, \tag{3.36}$$
$$C_m = C_{OL}.W_{Q2} + \alpha_1.W_{Q2} + C_{db,Q3} + C_{gs,Q3} + C_{gs,Q4}, \tag{3.37}$$
$$C_m = (2.1).W_{Q2} + C_{db,Q3} + C_{gs,Q3} + C_{gs,Q4}. \tag{3.38}$$

In the above analysis, we have not derived PMOS constants as we can directly find parasitic capacitance in this case from simulations once the dimensions of the PMOS are known, which we elaborate in the next section.

3.6 Design Steps of Differential Amplifier Comparator based on Specifications

Based on the specifications derived in chapter , we begin with finding sizes of all of the transistors in the design of the differential amplifier is performed in the following procedure:

Step 1: Current I_o from Slew Rate(SR)

Step 2: M_3 and M_1 sizes from Input common mode range maximum (ICMR+)

Step 3: M_1 and M_2 sizes from Gain Bandwidth Product (GBW) requirement

Step 4: The tail transistor sizing (M_5) from Input common mode range minimum (ICMR-)

3.6.1 Current calculations from Slew Rate (SR)

The power limitation, as presented in chapter 3, essentially limits our slew rate (SR), which is $3.0\,\mathrm{V\,ns^{-1}}$. Henceforth, calculated the maximum current that can be drawn by a single pixel comparator as,

$$SR = \frac{dV}{dt} = \frac{I_s}{C_L}, \tag{3.39}$$

$$I_s = C_L \cdot SR, \tag{3.40}$$

$$I_s = 10\,\mathrm{fF} \cdot 3.0\,\mathrm{V\,ns^{-1}} = 30\,\mu\mathrm{A}. \tag{3.41}$$

3.6.2 Sizing of Active load transistors Q3 and Q4 based on the input common mode maximum (ICMR+)

As defined in chapter 1, we have ICMR+ of 900 mV. From figure 3.6, we can conclude that in order for Q1 and Q2 to remain in saturation region, we must have,

$$V_{DS} \geq V_{GS} - V_{th,n}, \tag{3.42}$$

$$V_D \geq V_G - V_{th,n}. \tag{3.43}$$

For Q1 and Q2, the NMOS must stay in saturation for the whole input common mode range. The worst case input in this case is ICMR+. Henceforth, for Q1 we can formulate equation 3.42 as,

$$V_x \geq V_{in,p} - V_{th,n}, \tag{3.44}$$

$$V_x \geq 0.9\,\mathrm{V} - 0.4155\,\mathrm{V} = 0.4845\,\mathrm{V}. \tag{3.45}$$

From equation (3.45), we can assume V_x is approximately equal 0.5 V. Therefore,

$$V_{SD(Q3)} = V_{SG(Q3)} = V_{DD} - V_x = 1.8\,\mathrm{V} - 0.5\,\mathrm{V} = 1.3\,\mathrm{V}, \tag{3.46}$$

which results in the I_{DS} vs. V_{GS} equation to find the sizing of Q3 and Q4,

$$I_{SD,Q3,Q4} = \frac{\mu_p C_{ox}}{2}\frac{W}{L}(V_{SG,Q3} - |V_{th,p}|)^2, \tag{3.47}$$

$$\frac{27}{2} = \frac{79.2}{2}\mu\mathrm{A\,V^{-2}}\frac{W}{L}(1.3\,\mathrm{V} - 0.468\,\mathrm{V})^2, \tag{3.48}$$

$$(\frac{W}{L})_{Q3,Q4} = 0.49 \equiv 1. \tag{3.49}$$

As now the sizes of Q3 and Q4 are defined, using DC simulation on figure 3.6 we extracted the values of capacitance as $C_{db,Q3,Q4}$ to 0.75 fF, $C_{gd,Q3,Q4}$ to 2.2 fF, and $C_{gs,Q3,Q4}$ to 3 fF. From these values we can refined equation (3.35) and (3.38) as,

$$C_{out} = 11.362\,\mathrm{fF} + (2.1).W_{Q2}, \tag{3.50}$$

$$C_m = 6.75\,\mathrm{fF} + (2.1).W_{Q2}. \tag{3.51}$$

3.6.3 Sizing of the gain NMOS transistors Q1 and Q2 using gain bandwidth product (GBW)

As $C_{out} > C_m$, we approximated from equation (3.10) and (3.11) that ω_{p1} occurs at smaller frequency than ω_{p2}. Henceforth, we concluded that our -3 dB bandwidth is at ω_{p1}. Constructing the gain bandwidth product (GBW) from the bandwidth and the gain, as presented in equation (3.9) we obtained,

$$GBW = A_d BW = A_d \omega_{p1}, \tag{3.52}$$

$$GBW = \frac{g_m}{2\pi C_{out}}, \tag{3.53}$$

$$GBW = \frac{g_m}{2\pi[11.362\,\mathrm{fF} + (2.1)W_{Q2}]}. \tag{3.54}$$

Then, we extrapolated g_m from I_d vs. V_{GS} as the NMOS is in saturation region under strong inversion.

$$I_{d,Q1,Q2} = \frac{\mu_n C_{ox}}{2}\frac{W}{L}(V_{GS,Q1,Q2} - V_{thn})^2. \tag{3.55}$$

Differentiating with respect to V_{GS}, we calculated

$$\frac{dI_d}{dV_{GS}} = g_m = \mu_n C_{ox}\frac{W}{L}(V_{GS} - V_{thn}). \tag{3.56}$$

$$\tag{3.57}$$

Elaborating equation (3.54) using equation (3.56), we get,

$$GBW = \frac{\mu_n C_{ox}\frac{W}{L}(V_{GS} - V_{thn})}{2\pi\{11.362\,\mathrm{fF} + (2.1)W\}}. \tag{3.58}$$

Since, we know I_d is 27/2µA, we further simplified by squaring both sides as

$$\{GBW \cdot 2\pi \cdot (11.362\,\text{fF} + (2.1)W)\}^2 = \mu_n C_{ox} \frac{W}{L} \cdot 2I_d, \tag{3.59}$$

$$\{10^{10} \cdot 2\pi \cdot (11.362\,\text{fF} + (2.1)W)\}^2 = 287.3\,\mu\text{A}\,\text{V}^{-2} \cdot \frac{W}{L} \cdot 27\,\mu\text{A}. \tag{3.60}$$

Solving equation (3.60) for W, where we equated L equals 1 µm, we obtained imaginary results, which indicates, for such specifications it is be possible to achieve such a high GBW. Since, we cannot increase our current, we redefine our GBW as 1 GHz. Solving equation (3.59) for 1 GHz GBW, we get real roots as,

$$W_{root1} \approx 2\,\mu\text{m}, \tag{3.61}$$

$$W_{root2} \approx 32\,\mu\text{m}. \tag{3.62}$$

As we have area constraint, we chose lower root which is $W_{root1} \approx 2\,\mu$m. Furthermore, using a ratio (w/L) of two does not leave much headroom for the tail transistor to be in saturation. Threfore, we increased the (W/L) to three, to have relaxation on current source sizing.

$$\{\frac{W}{L}\}_{Q1,Q2} = 3. \tag{3.63}$$

3.6.4 Tail transistor Q5 sizing from input common mode range minimum

For thin oxide NMOS, we have ICMR- of 700 mV. The size of the tail current source (Q5) was obtained by considering our input at ICMR- (see figure 3.6). Using saturation region equations we obtained,

$$I_{d,Q5} = 27\,\mu\text{A} = \frac{\mu_n C_{ox}}{2} \frac{W}{L} (V_{DSat,Q5})^2. \tag{3.64}$$

As V_{in} must be greater than $V_{gs,Q2} + V_{DSat,Q5}$, we calculated the maximum $V_{DSat,Q5}$ in order to keep Q2 and Q5 in saturation region. Foremost, we calculated $V_{gs,Q2}$ from,

$$I_{d,Q1,Q2} = \frac{\mu_n C_{ox}}{2} \frac{W}{L} ((V_{GS} - V_{thn}))^2, \tag{3.65}$$

$$\frac{27}{2}\mu A = \frac{287.3}{2}\mu\text{A}\,\text{V}^{-2} \cdot 3 \cdot ((V_{GS} - 0.4155\,\text{V}))^2, \tag{3.66}$$

$$V_{GS} = 0.59\,\text{V}. \tag{3.67}$$

From $V_{GS,Q1}$, we found the $V_{DSat,Q5}$ by using the following

$$V_{in} \geq V_{gs,Q2} + V_{DSat,Q5}, \tag{3.68}$$

$$0.73\,\text{V} \geq 0.59\,\text{V} + V_{DSat,Q5}, \tag{3.69}$$

$$V_{DSat,Q5} \leq 0.140\,\text{V}. \tag{3.70}$$

Hence, we can use the I_d vs. V_{GS} equation in saturation region, as shown in equation (3.64), to find the sizing of the current source as

$$27\,\mu\text{A} = \frac{287.3}{2}\mu\text{A}\,\text{V}^{-2}\frac{W}{L}(0.140)^2, \tag{3.71}$$

$$\{\frac{W}{L}\}_{Q5} = 9.58 \equiv 10. \tag{3.72}$$

3.7 Regenerative Feedback architecture Design

In this section, we discuss the design of the structure five and structure six of the SPAD farm design as shown in figure 3.2. Let us begin with understanding the design of the structure five as shown in figure 3.10. This design is modified architecture as presented in [31]. [31].

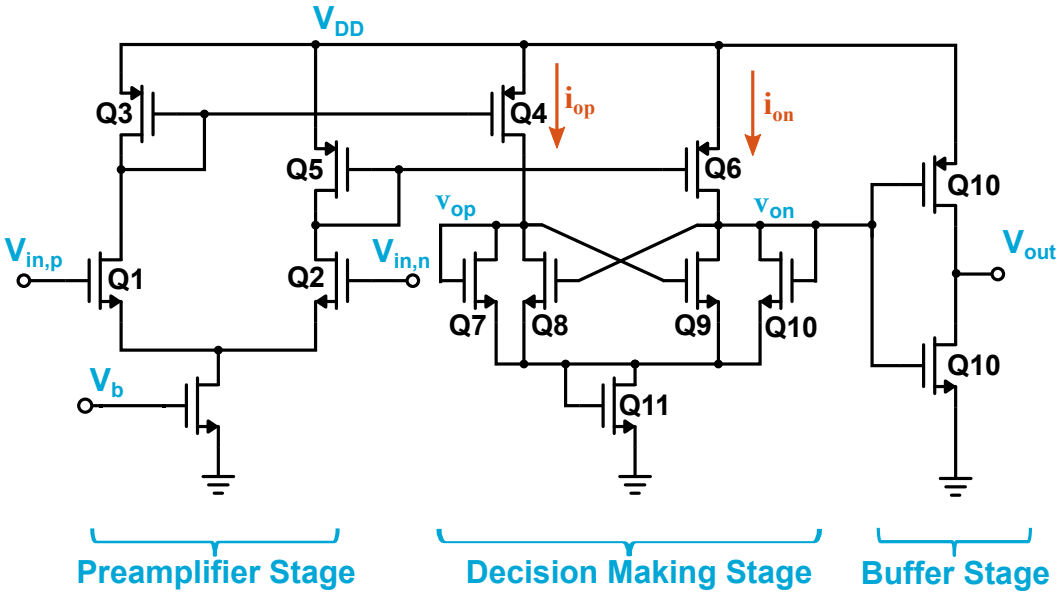

Figure 3.10: High speed regenerative latch design

This high speed comparator is divided into three individual stages, which are, input preamplification stage, positive feedback latch or decision making stage, and output buffer stage. The input preamplification stage amplifies the input from the SPAD and helps to improve the sensitivity of the input difference for the next stage of decision making. The decision stage inputs the large current difference at the nodes v_{op} and v_{on} as shown in figure 3.10, to provide a positive feedback to make the difference between v_{op} and v_{on} high. As we do not have synchronous circuitry, we implement diode transistors Q7 and Q10, to clamp the latch voltages to a level that can be reset using a pulse at the input which is opposite in polarity. The final stage is the output buffer stage, which is a minimum sized inverter. The inverter provides rail-to-rail swing of the output produced at v_{on} in our case.

We began the preamplifier design using the previous designs of the section 3.3. This preamplifier design requires low input capacitance, high speed. Henceforth, the Q1

and Q2 are sized keeping the differential amplifier transconductance, g_m and input capacitance into consideration. A large gain is critical as the latch in the second stage provides enough gain. The current gain from Q1 and Q2 is expressed as,

$$i_{op} = \frac{g_m \cdot (v_{in,p} - v_{in,n})}{2} + \frac{I_s}{2}. \tag{3.73}$$

The second stage, which is regenerative latch, has a minimum sized inverter as the load which has input capacitance of 1.12 fF as per the calculations from section 3.5. This input capacitance is the load capacitance of the diode clamped latch made from Q7 till Q10 in figure 3.10. We analyzed the small signal model of the latch made of Q8 and Q9 and reached to the conclusion that the voltage change at the output of the latch Q8 and Q9 can be expressed as,

$$\Delta V_{out}(t) = e^{\frac{t}{\tau_L}} \Delta V_i. \tag{3.74}$$

Where $e^{\frac{t}{\tau_L}}$ is the gain of the latch in time and τ_L is the time constant of the latch. In order to have high a gain from the latch we needed to lower τ_L. From the analysis of the small signal model of the NMOS based latch, if the gate-source capacitance at the output is considered as dominant, we reach to the following equation [22],

$$\tau_L = \frac{C}{g_m} = 0.67\sqrt{\frac{C_{ox} W L^3}{2\mu_n I}}. \tag{3.75}$$

Henceforth, to increase the sensitivity, or the gain from the latch, we needed to use minimum width and length allowed by the technology. The minimum length used by the technology is 180 nm. However, as we decrease the sizes of the transistor more mismatch occurs which appears in the form of offset. As a trade-off, we used W_{Q8}, W_{Q9}, L_{Q8}, and L_{Q9} equal to 300 nm. Furthermore, the diode Q7 and Q10 are sized equally to Q8 and Q9 in order to avoid hysteresis. This can be understood from the following analysis in [31].

As seen in figure 3.10, we started with the assumption that $i_{op} >> i_{on}$. In such situation Q7 and Q9 are on due to $v_{op} > V_{th,n}$, and Q8 and Q10 are off due to $v_{on} < V_{th,n}$. Also, we assumed that,

$$\beta_{Q7} = \beta_{Q8} = \beta_{op} = \mu_n C_{ox} \frac{W_{op}}{L_{op}}, \tag{3.76}$$

$$\beta_{Q9} = \beta_{Q10} = \beta_{on} = \mu_n C_{ox} \frac{W_{on}}{L_{on}}. \tag{3.77}$$

We started to decrease i_{op} and increase i_{on} simultaneously. As a result, v_{op} starts to decrease and v_{on} starts to increase until v_{on} is equal to $V_{th,n}$. As v_{on} increases further, the current i_{op} starts to split between Q7 and Q8. The latch operation starts to form a positive feedback loop which finally makes $v_{op} < V_{th,n}$. In this state the switching concludes, such that Q8 and Q10 are turned on and Q7 and Q9 turned off. At the stage when v_{on} has almost reached $V_{th,n}$, but the current has not started to flow in Q8 and

Q10, we have all of the currents flowing in Q7 and Q9. This stage is described by the following equations,

$$i_{op} = \frac{\beta_{op}}{2}(v_{op} - V_{th,n})^2, \tag{3.78}$$

$$i_{on} = \frac{\beta_{on}}{2}(v_{op} - V_{th,n})^2. \tag{3.79}$$

Hence, we concluded that,

$$i_{op} = \frac{\beta_{op}}{\beta_{on}} i_{om}. \tag{3.80}$$

The β_{op} were β_{on} made equal so that currents are same and we would not see the effect of hysteresis.

The sizing of the diode Q11 is calculated to shift-up the output v_{on} of the latch, in order to bias it in the middle of the inverter to get maximum gain and speed. The PMOS sizing on the other hand, represented by Q3 till Q6 in figure 3.10, was achieved considering that the transistor must stay in saturation (these calculations are similar to ones performed in section 3.6.2). To keep mismatch low, we have used W and L as 700 nm for all of the PMOS transistors.

3.7.1 Quiescent Current selection

As the maximum headroom allowed for the comparator design is 30 µA, we began by adjusting our design with 18 µA in total and performing DC and transient level simulations. This means that I_s of 9 µA is flowing in preamplifier and 9 µA in the second stage (the decision making stage). We ran the DC simulations by sweeping the $V_{in,n}$ from zero to V_{DD} and keeping $V_{in,p}$ equal to CM of 750 mV. We obtained the results as shown in graph 3.11 .

As we varied our input in figure 3.11 from 0.5 V till 1 V, we observed that the output of the comparator V_{out} goes from V_{DD} (1.8 V) to approximately zero volt. However, around 0.75 V we saw a state of transition for the output, V_{out1} to V_{out2}, and we require a difference of approximately 33 mV at the input to switch the output from V_{out1} to V_{out2}. Hence, we defined our threshold voltage for structure five as 33 mV. Furthermore, we observed that the uncertain state of the latch is between V_{on1} to V_{on2}. The observed current to switch the latch from V_{on1} to V_{on2} was defined as,

$$I_{on2} - I_{on1} = 4.656\,\mu A - 3.349\,\mu A \approx 1.3\,\mu A. \tag{3.81}$$

To ensure that we have enough quiescent current, total current in a single stage can be decreased up to 3 µA. However, in the transient response this results in decreased output rise time due to a minimum current consumption. From schematic transient simulations, we achieved a rise time of 150 ps degradation at the output if we use minimum current of 6 µA as the quisent current for structure five. The maximum current limit for our design was 30 µA. Henceforth, the total static current supplied to structure five was concluded as 18 µA.

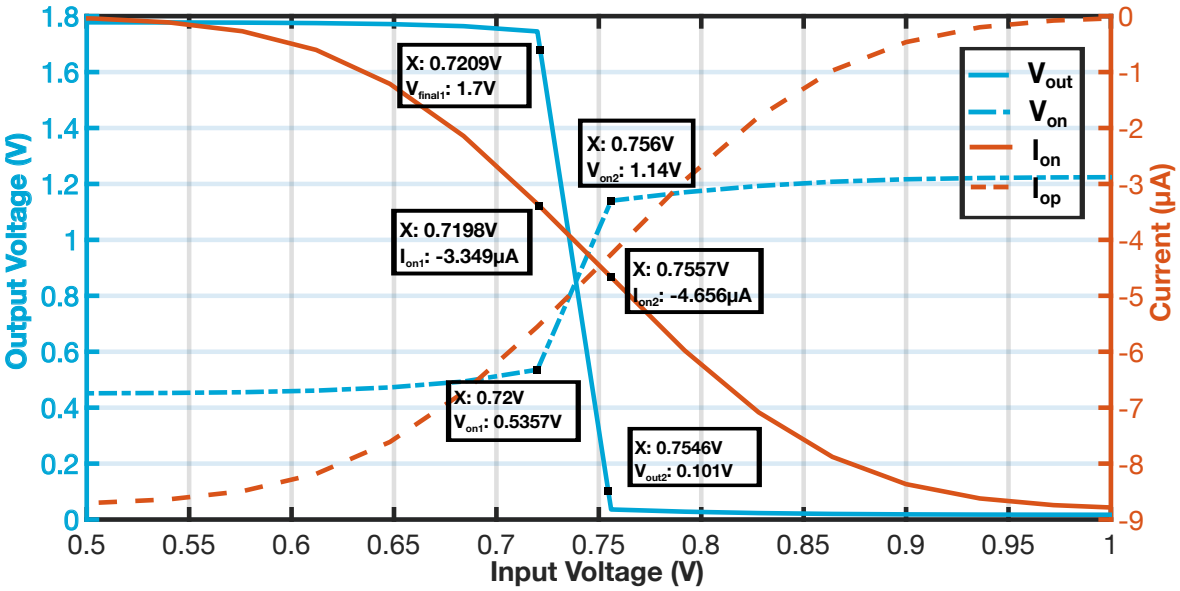

Figure 3.11: Structure five DC simulation for threshold and current analysis

3.8 Decoupling Capacitor

The comparator structures in the current design achieve less than the area requirement. The SPAD input rising pulse is fast in time, it was mandatory to add decoupling capacitors in each comparator. Such an integration allowed every pixel having its own decoupling capacitor. Since it is not practical that all SPAD fires at once, the adjacent decoupling capacitors also provide voltage against sudden pulse. In this design we used NMOS decoupling capacitors as shown in the fig. 3.12. The MOS decoupling capacitors

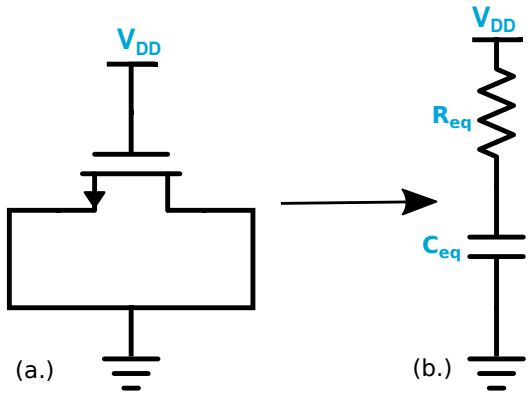

Figure 3.12: NMOS decoupling capacitor

provide high capacitance for low amount of area. The equivalent capacitance of nmos or pmos at low frequencies can be accurately formed as follows:[32]

$$C_{eq} = C_{ox}WL + 2C_{OL}W, \tag{3.82}$$

where C_{ox} is $\epsilon_o \cdot \epsilon_r / t_{ox}$, which is the gate oxide capacitance per unit area and C_{OL} stands for overlap & fringe capacitance per unit width. Also, there is a net effective resistance present in the form of R_{eq}, which was quantified for the low frequencies as

$$R_{eq} = \frac{L}{6\mu C_{ox} W (V_{gs} - V_{th})}. \tag{3.83}$$

As seen from equation 3.82, thick gate oxide devices have lower capacitance. In this regard, we have implemented thin oxide NMOS capacitors [33].

4 Post Layout Extracted Results

This chapter is divided in two parts. In the first part, we briefly discuss the design phase I. In the second part, we elaborate on the results achieved from in-pixel comparator for D-SiPMs.

4.1 Design Phase I

The design created for the A-SiPM, achieved results as provided in table 4.1.

Specification	**Result**
Preamplifier gain	24 dB
-3 dB **bandwidth**	1 GHz
Static power consumption	200 µW
Threshold without input preamplification	35 mV
Threshold with input preamplification	9.5 mV
Common mode rejection ratio (CMRR)	38 dB

Table 4.1: Summary of post-layout DC and AC simulation results of the design phase I

The most important specifications were the threshold sensitivity and speed of the comparator. The comparator displayed a -3 dB bandwidth of approximately 1 GHz in post-layout AC simulation. We ran the transient simulation in typical-typical corner with the post-layout extracted design (see figure 4.1). The rise time measured at the output of the comparator was 69 ps. The threshold voltage in typical-typical corner was kept at 150 mV for all of the simulations. The CM of 1.65 V was used in the DC and AC simulations of design I. Henceforth, we use the reference voltage of 1.5 V. It is important to notice that the power supply in this technology was 3.3 V.

We also ran four corner analysis of the preamplifier output and the final output pulses (see figure 4.2). The main observation from the corner analysis was that we required different reference voltages to be able to observe a pulse at the output. This is due to the presence of a high impedance node at the output of the preamplifier, which results in shifted DC levels at the comparator input (see figure 4.2(a.)). The summary of the rise time, slew rate, and required reference voltage of the five corners is summarized in table 4.2.

4.2 Design phase II

For the six comparator architectures (see figure 3.2), we ran post-layout DC and AC simulations. The summarized DC gain and BW are presented in table 4.3. It is important to notice that the power supply in this technology is 1.8 V. The maximum

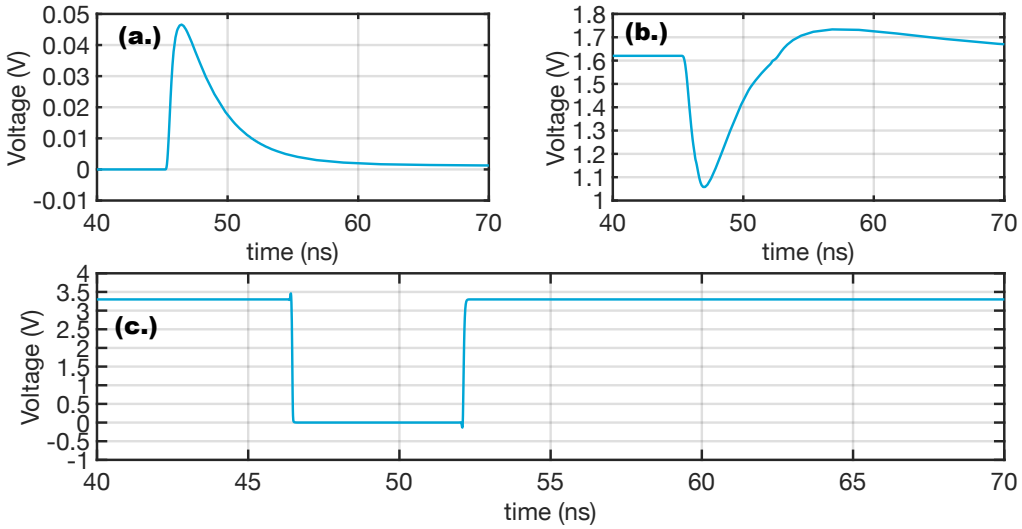

Figure 4.1: (a.) SiPM model pulse with 2.5 % microcells fired, (b.) preamplifier output pulse, (c.) comparator's final output post-layout transient response. The figures contain results from the design phase I

Corner	**Adjusted reference voltage** [V]	**Rise time** [ps]	**Slew rate** [V ns^{-1}]
Typical-Typical	1.50	69.04	48
Fast-Fast	1.50	51.08	65
Slow-Slow	1.50	94.00	35
Slow-Fast	1.75	53.50	62
Fast-Slow	1.15	94.06	35

Table 4.2: Summary of post-layout transient corner analysis of the design phase I

GBW of 2.61 GHz was achieved from the structure three. We also analyzed the noise frequency response of structure one. The integrated input referred noise in a bandwidth of 1 Hz untill 1 GHz was 1 $\mathrm{mV_{RMS}}$.

Structure	**DC gain** [dB]	**−3 dB Bandwidth** [MHz]	**Gain bandwidth product** [GHz]
Structure 1	41.21	20.68	2.38
Structure 2	44.55	12.13	2.05
Structure 3	42.62	19.31	2.61
Structure 4	45.53	10.25	1.94
Structure 5	32.64	20.24	0.87
Structure 6	26.85	20.06	0.44

Table 4.3: Post-layout DC and AC results for six structures of the design phase II

Furthermore, we performed a post-layout DC sweep of the input from 0 V to 1.8 V for the hysteresis behaviour. The reference voltage was set at the CM of the comparator.

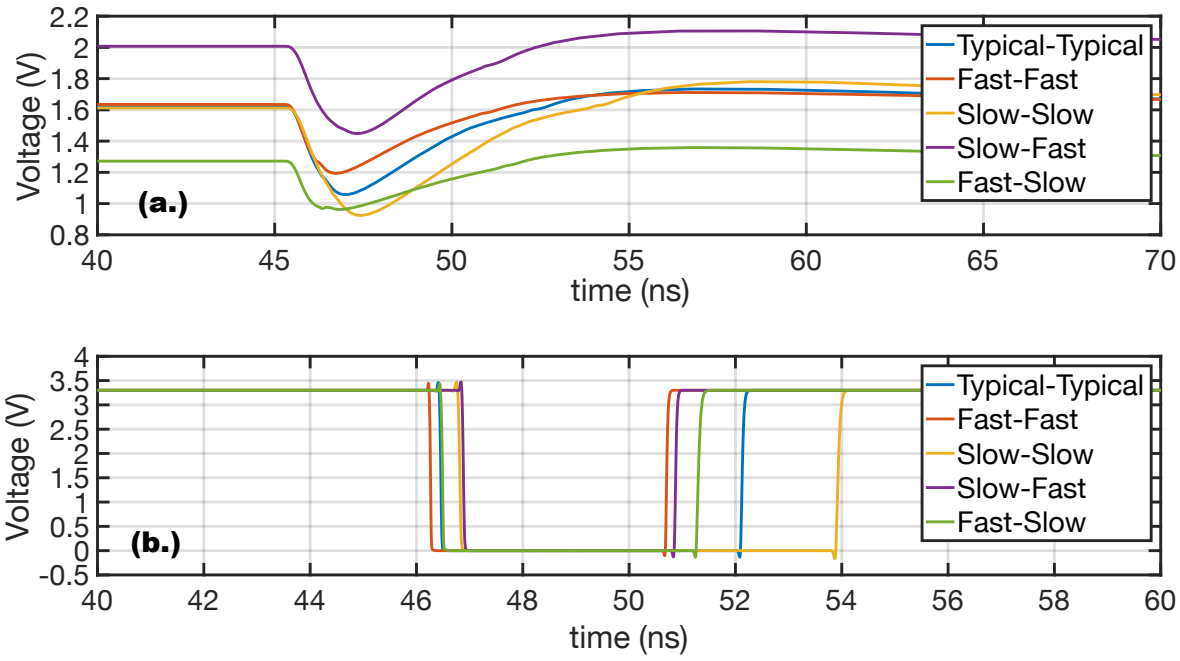

Figure 4.2: (a.) post-layout corner transient response of preamplifier output pulse, (b.) post-layout corner transient response of comparator's final output. The figures contain results from the design phase I

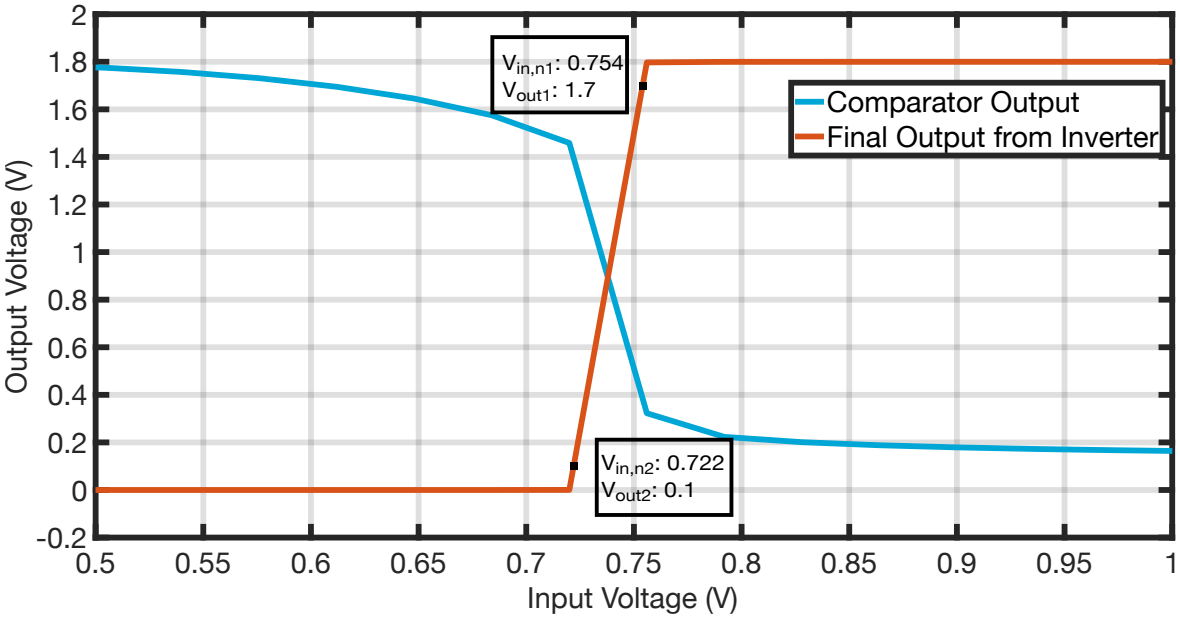

Figure 4.3: Post-layout DC simulation for structure one with input sweep from 0 V to 1.8 V of design phase II

As the input reached close to the reference voltage, the output switch from 1.8 V to 0.2 V for structure one (see figure 4.3). We measured the input range when the output of the

inverter switches from approximately 1.7 V to 0.1 V for all six comparator structures. We concluded that the comparator threshold resolution was limited by the one-sided hysteresis input range (see table 4.4). Additionally, we found the input referred noise of 1 mV_{RMS} to be significantly lower than the threshold resolution. We analyzed the hysteresis for the switching only in one direction as the relevant information for TOF-PET is the rise pulse of the input when SPAD fires. The minimum threshold of 32 mV was achieved with the first four structures. The static power consumption of all six architectures remain below 50 µW, minimum achieved power consumption of 16.2 µW corresponds to the structure six (see table 4.4).

Structure	**Minimum threshold achieved**[mV]	**Quiescent current**[µA]	**Static power consumption**[µW]
Structure 1	32	27	48.6
Structure 2	32	27	48.6
Structure 3	32	27	48.6
Structure 4	32	27	48.6
Structure 5	33.7	18	32.4
Structure 6	35	9	16.2

Table 4.4: Post-layout DC simulation results summary of the six comparator architectures of the design phase II

4.2.1 Transient analysis

We performed transient analysis in typical-typical corner for all six structures (see figure 4.4 for structure one). The post-layout extracted results of rise time and slew rate of all structures are summarized in the table 4.5. The comparator output swing is not from V_{DD} till ground; however, it reaches rail-to-rail after the addition of the timing line inverter stage (see figure 4.4). The highest achieved slew rate was 2.6 V ns^{-1} from structure three. This is due the presence of a self-biased differential amplifier, which is faster compared to the other structures.

Structure	**Comparator input rise time**[ns]	**Comparator output rise time** [ps]	**Inverter final output rise time (rail-to-rail)** [ps]	**slew rate** [V ns^{-1}]
Structure 1	6.65	534.4	295.8	2.4
Structure 2	7.99	542.9	295.4	2.4
Structure 3	7.23	485.3	317.4	2.6
Structure 4	9.10	518.3	317.8	2.5
Structure 5	2.41	648.3	400.9	1.2
Structure 6	5.92	1268.0	434.0	0.7

Table 4.5: Post-layout transient analysis result summary of the design phase II

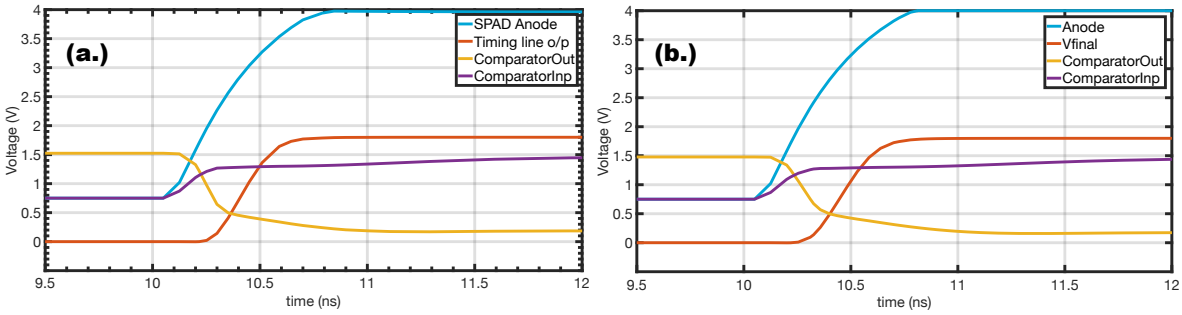

Figure 4.4: (a.)schematic transient simulation of structure one, (b.) post layout extracted transient simulation of structure one. The figures are results from the design phase II

4.2.2 Area Characterization

For the pixel farm of $1 \times 1\,\mathrm{mm}^2$, the layout of pixels containing two types of circular SPADs with structure 1 is shown in figure 4.5. The active region diameter of the bigger SPAD is 40 µm (see figure 4.5(a.)) and for the smaller SPAD is 12 µm (see figure 4.5(b.)).

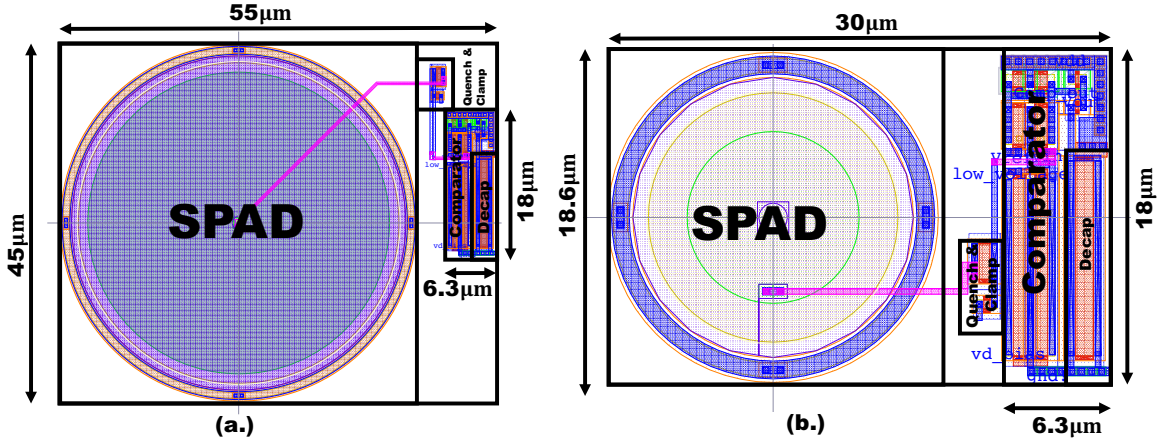

Figure 4.5: (a.)pixel with a SPAD of active region diameter of 40 µm and structure one, (b.)pixel with a SPAD of active region diameter of 12 µm and structure one. The figures are results from the design phase II

As the specified in-pixel area allocated for the comparator is a critical factor for the optimization of the fill-factor, we have summarized the area of all six comparator structures and the pixel area in table 4.6. For the implementation of the $3 \times 3\,\mathrm{mm}^2$ D-SiPM, we chose to use the bigger SPAD with active region diameter of 40 µm. We have also presented the percentage of area consumed by the comparator structure for all six designs (see table 4.6). The comparator area also includes the timing line inverter(inv.) and the decoupling capacitors (decap.) (see figure 4.6 and 4.7). The minimum area achieved was from structure three, which consumed merely 3.9 % area of the whole pixel (see figure 4.6(c.) for the layout of structure three). The entire chip

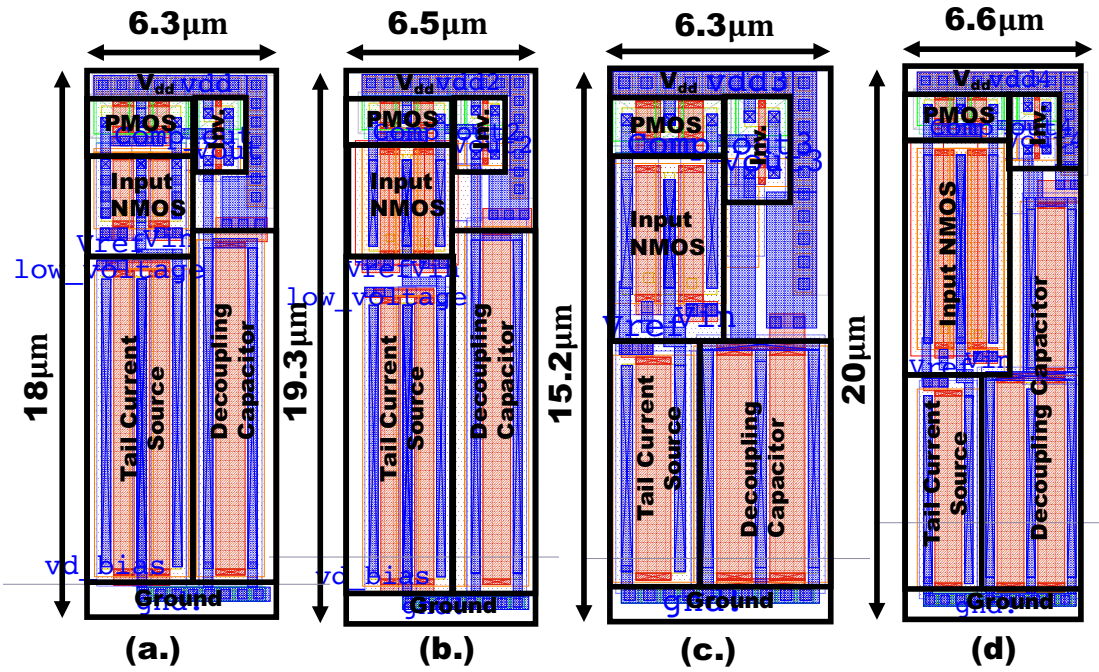

Figure 4.6: (a.)layout of structure one, (b.)layout of structure two, (c.)layout of structure three, (d.)layout of structure four. The figures are results from the design phase II.

area from the 12 pixel layouts resulted in $1 \times 1\,\mathrm{mm}^2$ (see figure 4.8).

Structure	Pixel Area for SPAD variant one[μm^2]	Pixel Area for SPAD variant two [μm^2]	Comparator + Decap. Area [μm^2]	Comparator + Decap. Area percentage with SPAD variant one [%]
Structure 1	2475	558	113.4	4.5
Structure 2	2475	549	125.4	5.1
Structure 3	2475	526	95.8	3.9
Structure 4	2475	570	132.0	5.3
Structure 5	2628	595	161.0	6.1
Structure 6	2628	597	143.0	5.4

Table 4.6: Area summary of the six structures and 12 pixels of the design phase II

4.2.3 High frequency behaviour of decoupling capacitors

The equivalent capacitance and resistance described in section 3.8 only works at low frequency. In order to formulate models that try to approximate the capacitance and resistance of decoupling capacitor at high frequency, the research done by [34] and [33] gives us following equations at high frequencies.

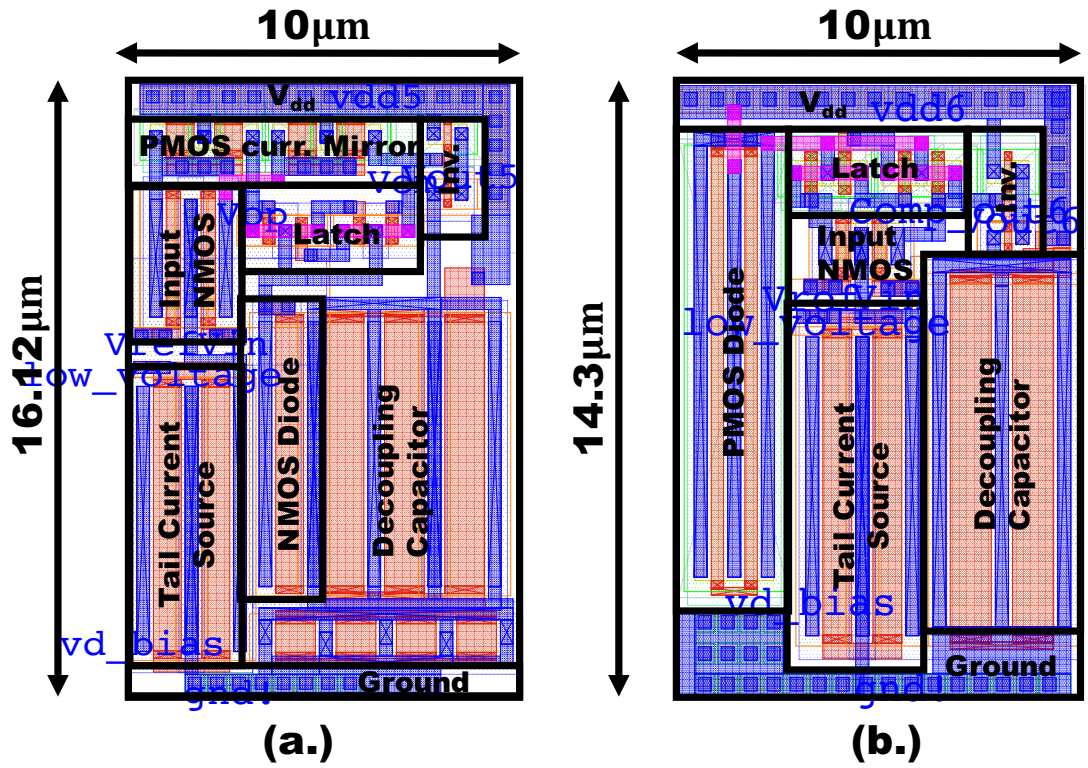

Figure 4.7: (a.)layout of structure five, (b.)layout of structure six. The figures are results from the design phase II

$$C_{eff} = \frac{C_{eq,o}}{1 + \frac{\omega}{\omega_o}}, \tag{4.1}$$

$$R_{eff} = \frac{R_{eq,o}}{1 + \frac{\omega}{\omega_o}}, \tag{4.2}$$

where $C_{eq,o}$ and $R_{eq,o}$ are given by equation 3.82 and 3.83 respectively. Here, the cuttoff frequency,

$$\omega_o = \frac{\mu(V_{GS} - V_{th})}{L^2} \tag{4.3}$$

At high frequencies, it was inferred from above that NMOS performs better than PMOS at high frequencies as mobility of NMOS (μ_n) is almost double compared to PMOS devices, and thus NMOS has higher cutoff frequency. As a result of such analysis, we have implemented NMOS decoupling capacitors.

The technology parameters, which are minimum length of 180 nm, μ_n of $0.0288\,\mathrm{m^2\,V^{-1}\,s^{-1}}$, supply of 1.8 V and $V_{th,n}$ of 0.415 V, were used to calculate the cuttoff frequency. We observed that the cutoff frequency for L of 1 µm is approximately

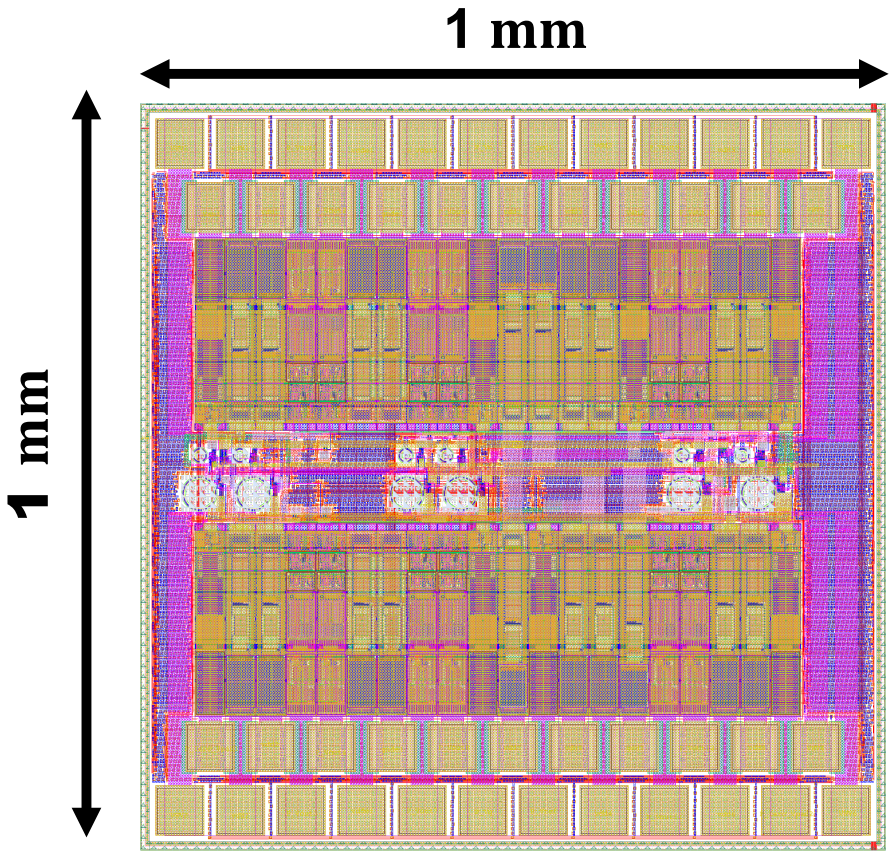

Figure 4.8: layout of the whole chip of $1 \times 1\,\mathrm{mm}^2$ pixel farm. The figure is a result of design phase II

40 GHz, which is much higher than the SPAD pulse frequency. Hence, we chose not to reduce the L further than 1 µm as it comes at the expense of reducing the equivalent capacitance. Hence, we use L of 1 µm.

Table 4.2.3 summarizes the zero frequency decoupling capacitance, reducing the resistance (from equation 3.82 and 3.83), and the respective values at 10 GHz, which is the maximum frequency that comparators require.

Structure	W[µm]	L[µm]	$C_{eq,o}$[fF]	$R_{eq,o}$[Ω]	C_{eff} **at 10GHz** [fF]
Structure 1	11.00	1	118	29	94
Structure 2	12.50	1	134	26	107
Structure 3	12.52	1	134	26	107
Structure 4	21.00	1	225	15	180
Structure 5	24.04	1	257	13	206
Structure 6	16.52	1	177	20	141

Table 4.7: Decoupling Capacitor detailed summary for six structures of the design phase II

4.2.4 Common Mode Rejection Ratio

The common mode rejection ratio(CMRR) is formulated as the ratio between differential gain (A_d) and the common mode gain (A_c), or,

$$CMRR = 20 \log \frac{A_d}{A_c}. \tag{4.4}$$

We ran a post-layout AC simulation for the structure one and calculated differential gain and the common mode gain. The CMRR for structure one resulted as 50.5 dB at low frequencies. The -3 dB BW where the CMRR starts to decrease was 12.6 MHz.

Conclusion and Future Research

5

We have investigated on the improvement of the timing resolution in the TOF-PET imaging. Firstly, we have realized an in-pixel ultra-low power comparator. For such realization we have implemented six independent comparator structures for improved timing jitter SPAD cell for D-SiPMs. All the six comparator structures consume a static power that is less than 50 µW, with a lowest value of 16.2 µW. The maximum achieved slew rate was 2.6 $\mathrm{V\,ns^{-1}}$, and a GBW of 2.61 GHz was procured. The future research is to investigate if the achieved slew rate, and hence achieved rise time, is sufficient for the required timing performance. The transient monte carlo analysis needs to be performed for such investigation, which considers the SNR at the output of the comparator. Using noise transient analysis in SPICE with different noise seeds, the timing jitter of the comparator can be extracted as

$$\sigma_t = \frac{\sigma_n}{\frac{dV_{in}}{dt}}. \tag{5.1}$$

Where σ_t is the standard deviation of the comparator flipping times, σ_n is the standard deviation of the noise source, and $\frac{dV_{in}}{dt}$ is the slew rate[23]. In addition, we allocated a comparator area that is smaller than the required specifications. The lowest area of 95.8 μm^2 was achieved with one of the comparator structures, with the total area percentage of 3.9 % of the whole pixel. This includes comparator architecture, timing line inverter, and an on-pixel decoupling capacitor. The maximum capacitance achieved in one of our designs was of 200 fF till 10 GHz frequency. We have investigated the CMRR of one structure, which resulted as 50 dB. The input referred noise achieved by one of the structures is estimated to be 1 $\mathrm{mV_{RMS}}$ till the frequency of 1 GHz.

Furthermore, we implemented a high PDE and low DCR A-SiPM with integrated readout electronics. A high speed comparator was successfully designed which achieved a BW of approximately 1 GHz.

The characterization of the $1 \times 1\,mm^2$ pixel farm silicon is the next immediate future work for this research, which includes the statistical analysis of the timing resolution of the pixel. After testing the chip, the next step is to implement a full photodetector in $3 \times 3\,mm^2$. Also, the offset of the comparator needs to be computed as a future research. Such computation can be achieved by monte carlo simulation of mismatch and process variations, which causes offset to appear. The offset is an important characterization for the $3 \times 3\,mm^2$ D-SiPM, where multiple comparator architectures are operated simultaneously. We conclude this dissertation with the successful implementation of a comparator design for integrated A-SiPMs and for a SPAD pixel which achieves high speed, low power and low threshold.

Bibliography

[1] S. Surti, "Update on time-of-flight pet imaging," *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*, vol. 56, pp. 98–105, 01 2015.

[2] S. Mandai, *Multichannel Digital Silicon Photomultipliers for Time-of-Flight PET*. PhD thesis, Delft University of Technology, The address of the publisher, 7 2014.

[3] M. Fishburn, *Fundamentals of CMOS Single-Photon Avalanche Diodes*. fishburn, 2012.

[4] S. Mandai and E. Charbon, "Multi-channel digital sipms: Concept, analysis and implementation," *2012 IEEE Nuclear Science Symposium and Medical Imaging Conference Record (NSS/MIC)*, pp. 1840–1844, 2012.

[5] M. Ghioni, A. Gulinatti, I. Rech, P. Maccagnani, and S. Cova, "Large-area low-jitter silicon single photon avalanche diodes," 02 2008.

[6] SensL, "An introduction to the silicon photomultiplier," 2017. https://www.sensl.com/downloads/ds/TN%20-%20Intro%20to%20SPM%20Tech.pdf, accessed Sept 2018.

[7] SensL, "J-series spice model release note," 2017. release date Dec 11, 2018.

[8] M. Bazes, "Two novel fully complementary self-biased cmos differential amplifiers," *IEEE Journal of Solid-State Circuits*, vol. 26, pp. 165–168, Feb 1991.

[9] C. Veerappan and E. Charbon, "A low dark count p-i-n diode based spad in cmos technology," *IEEE Transactions on Electron Devices*, vol. 63, pp. 65–71, Jan 2016.

[10] J. M. Rabaey, *Digital Integrated Circuits: A Design Perspective*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1996.

[11] M. E. Phelps, "chapter 1 - physics of positron emission and annihilation," in *PET: Physics, Instrumentation, and Scanners* (M. E. Phelps, ed.), pp. 19 – 30, Los Angeles: Springer, New York, NY, 2006.

[12] S. R. Cherry, J. A. Sorenson, and M. E. Phelps, "chapter 3 - modes of radioactive decay," in *Physics in Nuclear Medicine (Fourth Edition)* (S. R. Cherry, J. A. Sorenson, and M. E. Phelps, eds.), pp. 19 – 30, Philadelphia: W.B. Saunders, fourth edition ed., 2012.

[13] S. Brunner and D. Schaart, "Bgo as a hybrid scintillator/cherenkov radiator for cost-effective time-of-flight pet," *Physics in Medicine & Biology*, vol. 62, no. 11, p. 4421, 2017.

[14] W. Moses, "Time of flight in pet revisited," *Nuclear Science, IEEE Transactions on*, vol. 50, pp. 1325–1330, Oct 2003.

[15] S. Seifert, J. H. L. Steenbergen, H. T. van Dam, and D. R. Schaart, "Accurate measurement of the rise and decay times of fast scintillators with solid state photon counters," *Journal of Instrumentation*, vol. 7, no. 09, p. P09004, 2012.

[16] G. Ripamonti and S. Cova, "Carrier diffusion effects in the time-response of a fast photodiode," *Solid-State Electronics*, vol. 28, no. 9, pp. 925 – 931, 1985.

[17] G. Ranucci, "Time statistics of the photoelectron emission process in scintillation counters," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 335, no. 1, pp. 121 – 128, 1993.

[18] A. Lacaita and M. Mastrapasqua, "Strong dependence of time resolution on detector diameter in single photon avalanche diodes," *Electronics Letters*, vol. 26, pp. 2053–2054, Nov 1990.

[19] A. Spinelli and A. L. Lacaita, "Physics and numerical simulation of single photon avalanche diodes," *IEEE Transactions on Electron Devices*, vol. 44, pp. 1931–1943, Nov 1997.

[20] A. Gulinatti, P. Maccagnani, I. Rech, M. Ghioni, and S. Cova, "35 ps time resolution at room temperature with large area single photon avalanche diodes," *Electronics Letters*, vol. 41, pp. 272–274, March 2005.

[21] A. Muntean, "Design of a fully digital analog sipm with sub-50ps time conversion," 2017.

[22] P. Allen and D. Holberg, *CMOS Analog Circuit Design*. The Oxford Series in Electrical and Computer Engineering, OUP USA, 2011.

[23] A. Rivetti, *CMOS: Front-End Electronics for Radiation Sensors*. Devices, Circuits, and Systems, Taylor & Francis, 2015.

[24] S. O. Rice, "Mathematical analysis of random noise," *Bell System Technical Journal*, vol. 23, no. 3, pp. 282–332.

[25] S. O. Rice, "Mathematical analysis of random noise," *The Bell System Technical Journal*, vol. 24, pp. 46–156, Jan 1945.

[26] S. J. Kim, D. Kim, and M. Seok, "Comparative study and optimization of synchronous and asynchronous comparators at near-threshold voltages," in *2017 IEEE/ACM International Symposium on Low Power Electronics and Design (ISLPED)*, pp. 1–6, July 2017.

[27] B. A. Chappell, T. I. Chappell, S. E. Schuster, H. M. Segmuller, J. W. Allan, R. L. Franch, and P. J. Restle, "Fast cmos ecl receivers with 100-mv worst-case sensitivity," *IEEE Journal of Solid-State Circuits*, vol. 23, pp. 59–67, Feb 1988.

[28] V. Milovanović and H. Zimmermann, "Analyses of single-stage complementary self-biased cmos differential amplifiers," in *NORCHIP 2012*, pp. 1–4, Nov 2012.

[29] S. Lindner, S. Pellegrini, Y. Henrion, B. Rae, M. Wolf, and E. Charbon, "A high-pde, backside-illuminated spad in 65/40-nm 3d ic cmos pixel with cascoded passive quenching and active recharge," *IEEE Electron Device Letters*, vol. 38, pp. 1547–1550, Nov 2017.

[30] B. Razavi, *Design of Analog CMOS Integrated Circuits*. New York, NY, USA: McGraw-Hill, Inc., 1 ed., 2001.

[31] R. Baker, *CMOS: Circuit Design, Layout, and Simulation*. IEEE Press Series on Microelectronic Systems, Wiley, 2011.

[32] X. MENG, "Decoupling capacitor design issues in 90nm cmos," 10 2018.

[33] J. Chia, *Design, layout and placement of on-chip decoupling capacitors in IP blocks*. PhD thesis, 2004.

[34] W. Liu, "Mosfet models for spice simulation including bsim3v3 and bsim4," 01 2001.