

SERCuit

Single Electron Readout Circuit

by

Matthew Al Disi

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Tuesday August 28, 2020 at 11:30 AM.

Student number: 4770935
Project duration: September 1, 2019 – August 28, 2020
Thesis committee: Prof. S. Nihtianov, TU Delft, supervisor
Prof. Q. Fan, TU Delft, supervisor
Prof. F. Sebastiano, TU Delft

This thesis is confidential and cannot be made public until December 31, 2022.

An electronic version of this thesis is available at <http://repository.tudelft.nl/>.

Acknowledgments

I want to express my deep and sincere gratitude to my supervisors Prof. Qinwen Fan and Prof. Stoyan Nihtianov. They taught me how to critically think, manage my time, communicate my ideas to academic and industry professionals, and work within a team. They also were patient with me whenever I made mistakes and supported me through stressful times, and for that, I came to consider them not only as supervisors but also as friends.

I am thankful for Shoubhik Karmakar for answering my questions and helping me demystify analog circuits and Cadence, and Jan Angevare for passing down his knowledge of the 65nm technology and giving me one-on-one guidance throughout my taping-out process.

I could not have survived my master studies without the continuous love and support of my sisters: Reem, Sara, Rana, and Dana and my close friends: Afonso, Mariano, Caitlin, Rik, Shivani, and Luc. I am forever indebted.

Finally, I extend my thanks for my lab mates at EI: Nuriel, Shrinidhi, Angqi, Nandor, Robert, Arthur, Roger, Huajun, Efraim, Matheus, and Shardul for making my masters experience memorable.

*Matthew Al Disi
Delft, August 2020*

Abstract

Particle detection circuits are used for a wide range of applications from experimental physics to material testing and medical imaging. State-of-the-art imaging systems demand the detection of small amounts of charge with small time-resolution and limited power consumption, creating an implementation dead-end for the typical readout topology. In this thesis, a particle detection readout based on an intersymbol interference cancellation scheme is introduced to address this issue. Evaluated in post-layout simulations, the proposed architecture can detect generated charge as small as 160 aC with 97.8% certainty. The readout can operate with event-rates up to 400 MEvent/s while only consuming 2.85 mW of power.

Contents

1	Background	1
1.1	Introduction	1
1.2	Feasibility	3
1.3	The Two-Threshold Comparator	6
1.4	Thesis Organization	9
2	Architecture	10
2.1	Feedback TIA	10
2.2	Gain Stage	11
2.3	Comparator	13
2.4	Thresholds	13
3	Design & Implementation	16
3.1	Feedback TIA	16
3.2	Gain Stage	20
3.3	Comparator	21
3.4	Buffers	23
4	Results	25
5	Design for Test (DfT)	28
5.1	DfT Blocks	29
5.2	Programmability	32
5.3	Measurement Setup	33
6	Discussion, Future Work, and Conclusions	35
6.1	Discussion	35
6.1.1	Decision Feedback Equalization	37
6.2	Future Work.	39
6.3	Conclusions.	39
	Bibliography	40

Background

1.1. Introduction

Semiconductor particle sensors have made possible the customizability and dense integration required for many applications in experimental physics, astronomy, medicine, and material testing [1]. They work by absorbing energy deposited by a particle and generating electron-hole pairs that are swept to the sensor's electrodes by an electric-field [1, 2]. Pairs are then collected by a readout circuit for further processing. Figure 1.1 shows the block diagram of a typical readout circuit which consists, in its simplest form, of a preamplifier, gain stage and a digitizer [2]. The preamplifier converts current pulses produced by the sensor into voltage signals. Unfortunately, it also smears the signal with noise and limits its rate of change; bottlenecking the system's resolution and bandwidth [1, 2]. The Gain stage boosts the signal voltage levels making it less prone to digitization errors. Meanwhile, the digitizer can come in many forms depending on the information of interest to the end-user. They can be (i) time-to-digital converters (TDCs); (ii) analog-to-digital converters (ADCs); or (iii) a binary discriminator (i.e. comparator or slicer) [2].

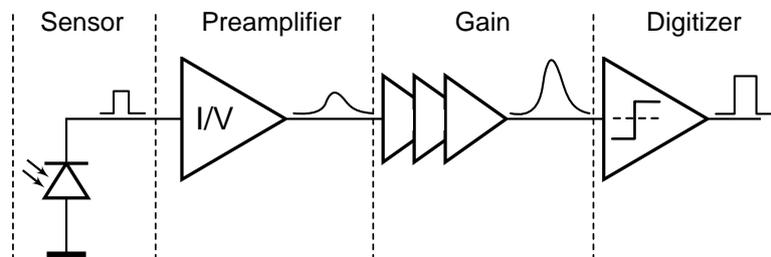


Figure 1.1: Block diagram of a particle readout. Current pulses generated by the sensor are converted to voltage, amplified, and fed to a comparator for discrimination.

An electron scanning microscope (SEM) employs particle detectors to image microscopic specimen [3]. Images are created by counting the number of electrons backscattered from a specimen bombarded with an electron beam [3, 4]. And ever since its inception, it has been known that low-energy electron beams are the key to image nanometer specimens; requiring particle readouts with a fine resolution [2, 5]. Furthermore, when sensitive or non-conductive samples are imaged, the beam must quickly scan the samples to avoid radiation damage and charging artifacts, hence, the readout must operate at high speeds [3]. Finally, the sensor area is segmented into many pixels to improve electronic noise and count-rate capabilities [1, 2, 4]. Because each pixel (i.e. channel) requires its own readout, the readout must consume limited power to keep the total power consumption acceptable. Particle detectors designed to imaging sensitive non-conductive nanometer specimen (e.g. a 10 nm MOSFET gate oxide) pose a challenge in terms of their readouts as they must be:

1. Low noise to detect low-energy particles.

2. Wideband to support high detecton-rates.
3. Low power to allow for a high amount of pixelization.

There already exist semiconductor sensors capable of detecting low-energy particles with remarkable efficiency and speed [4, 5]. Yet, no investigation has been performed on the feasibility of a low power readout solution sensitive to small charge portions while a having high event-rates. In this thesis, our objective is to create a readout circuit capable of detecting charge signals generated by a single pixel realized as a PIN diode, as a result of external electron hitting the surface of the detector at random times. The charge signals are as small as 160 aC, while the maximum allowed detection error is 5 % (not more than 5 false or missed detections per 100 generated charge signals). The detected charges have to be assigned to time intervals of 2.5 ns. In other words, the readout must have an event-rate of 400 MEvent/s where an event is either an electron hitting the pixel (a HIGH state or a logical '1') or no-electron hitting the pixel (a LOW state or a logical '0'). Furthermore, two assumptions are made regarding events:

1. A pixel is hit with no-more than 1-electron every 2.5 ns.
2. No-more than 3-electron can hit a pixel consecutively.

The above assumptions are based on the fact that the detector is heavily pixelized and the number of electrons backscattered from the specimen is low (≈ 20) [4]. Therefore, it is unlikely than any single pixel receive many electrons within a short period of time. The readout's target specifications and the PIN diode properties are summarized in Table 1.1 and Table 1.2 respectively.

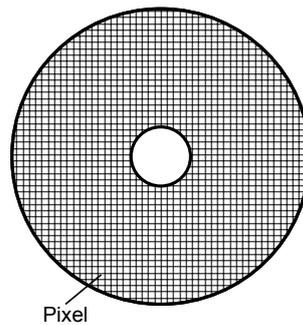


Figure 1.2: Bottom view of a highly pixelized detector.

Table 1.1: Readout target specifications

Specification	Target
Error-rate	<5 %
Event-rate	400 MEvent/s
Area	110 mm ² /(Nu. of pixels)
Power consumption	2.5 W/(Nu. of pixels)

Table 1.2: PIN diode propertise[‡]

Property	Value
Generated charge	160 aC
Capacitance	120 pF/(Nu. of pixels)
Area	110 mm ² /(Nu. of pixels)

[‡] PIN diode propertise are based on [5], electron energy of 6 keV, epi-thickness of 80 μ m, and bias voltage of 30 V.

Kleczek *et al.* created a counting charge detector for x-ray imaging which is 90 % accurate up to *count-rates* of 12 MHz—consuming only 100 μ W of power [6]. The readout relies on a charge sensitive amplifier (CSA) with an active feedback block known as Krummenacher feedback [6, 7].¹ The Krummenacher circuit emulates an inductance and a resistance which are tasked with compensating the sensor's leakage current and discharging the CSA respectively [7]. Unfortunately, the active feedback introduces multiple poles and thus has stability issues when the bandwidth is scaled up [8]. The counting implementation in [6] is done using a ripple-counter which, unlike a dynamic comparator, requires the preamplifier to completely discharge its previous input before subsequent charges can

¹This can also be viewed as a compensated transimpedance amplifier.

be counted. This is not the case for the readout architecture shown in Figure 1.1 which can typically supports event-rates up to $1.42\times$ the overall circuit bandwidth [9].² The circuit in [6] has a bandwidth of 37 MHz ³, and hence, it can theoretically achieve event-rates up to 52 MEvent/s if implemented in the form of the aforementioned architecture. Kleczek's work shows the feasibility of low-power highly pixelized readouts, however, the event-rate that can be achieved is much lower than 400 MEvent/s . Moreover, the circuit works with higher input signal energies ($\approx 350\text{ aC}$), does not require electro-static discharge (ESD) protection, and employs single-ended amplifier—resulting in much-improved signal-to-noise ratio (SNR).

Wider-band particle detectors are typically used in time-of-flight measurements. The authors in [10] and [11] created readout channels with up to 410 MHz of bandwidth, theoretically allowing the detection of 585 MEvent/s . However, the readout exhibits a minimum noise levels of 180 aC and consumes at least 17 mW of power [10, 11].

The above highlights the fundamental noise-power-bandwidth trade-off in analog circuits. Current particle readouts either meet low-power low-noise requirements as in [6] or are wideband as in [10] posing the question: are the target specifications attainable?

1.2. Feasibility

A feasibility study of the readout can be performed by analyzing the preamplifier because it determines the readout's SNR and bandwidth—the two parameters affecting error-rate—and typically dominates the readout's power consumption. The effects of SNR on particle detection accuracy are straightforward and are shown in Figure 1.3a. As the SNR degrades the signal is lost in the noise and the error-rate approaches 50% [9]. On the other hand, bandwidth and noise are inherently interlinked. Designing a circuit with an excessively large bandwidth means an increase of the integrated noise and hence the error-rate. Meanwhile, an excessively narrow bandwidth gives rise to inter-symbol interference (ISI). ISI distorts the analog '1' and '0' levels, bringing them closer to each other as drawn in Figure 1.3b. Consequently, it becomes harder to discriminate between them and the error-rate increases.

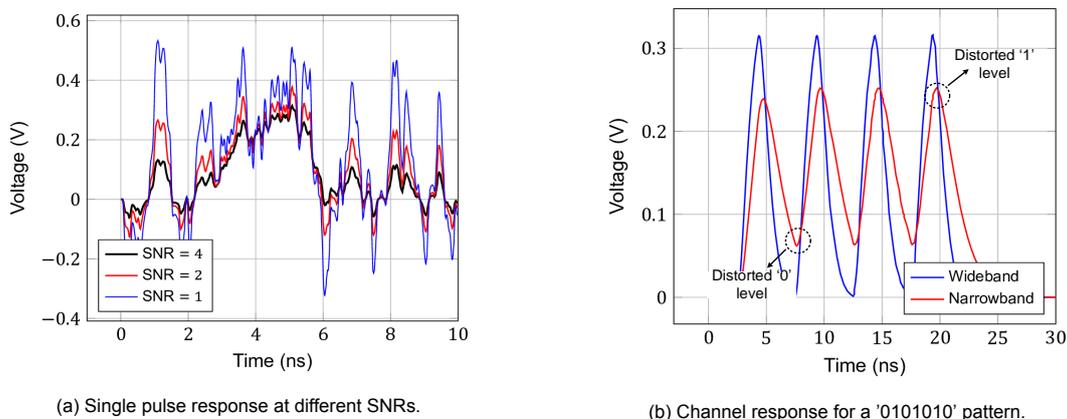


Figure 1.3: Effects of noise and bandwidth limitations.

The aforementioned interplay results in the graph illustrated in Figure 1.4 and an optimal bandwidth that balances between noise and ISI-induced errors [9]. It can be shown that detecting pulses in white noise with an error-rate $< 5\%$ requires an SNR of 3 to 4 [12], while a bandwidth of $1.42\times$ less than the desired event-rate (i.e. 280 MHz) is chosen as a starting point for the investigation. The question of feasibility then translates into whether a preamplifier can be built with the aforementioned qualities and an acceptable power consumption. In this analysis and throughout this work, the PIN diode is modeled as a current source with a parallel capacitance (e.g. Figure 1.5) where a current pulse is generated per electron hitting the surface of the diode. The duration of the pulse is determined by the diode's charge collection time t_c which, in this work, is 1.8 ns . The amplitude of the pulses is then set to 90 nA such that the charge generated is equivalent to $i_s \times t_c = 160\text{ aC}$.

²Approximation based on the optimal bandwidth of an optical receiver with non-return-to-zero (NRZ) digital signals [9].

³Calculated based on the reported discharging time-constant of 4 ns [6]

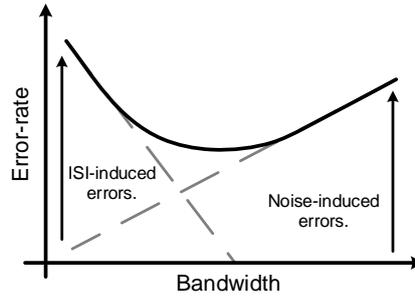


Figure 1.4: Conceptual plot of the trade-off between bandwidth, noise, ISI and the error-rate.

When considering preamplifier topologies, the feedback transimpedance amplifier (TIA) quickly emerges as the most suitable option. The CSA (Figure 1.5) amplifies incoming charges by $1/C_F$ but the circuit requires a resetting mechanism to discharge the feedback capacitor in preparation for the next incoming charge. This makes the CSA ill-suited for randomly generated charges as they may be generated, and consequently get lost, during the reset-phase. Furthermore, the resetting mechanism gives rise to noise-folding and charge injection, both of which degrade the performance of the circuit. Another option are open-loop TIAs based on common-gate amplifiers (Figure 1.6) which are typically used for wideband transimpedance implementations. Unfortunately however, they can be shown to be always noisier than the feedback TIA for the same bandwidth and power consumption due to the noise of the biasing device M2 appearing directly at its input [9]. Furthermore, common-gate TIAs have limited design freedom because the bandwidth provided by M1 and the noise contributed by M2 are affected by the same current but scale in the opposite direction. On the other hand, the feedback TIA parameters such as gain, noise, and bandwidth can be orthogonally designed. Additionally, feedback TIAs contribute less noise and can be built with a large bandwidth—making them the preferred preamplifier topology.

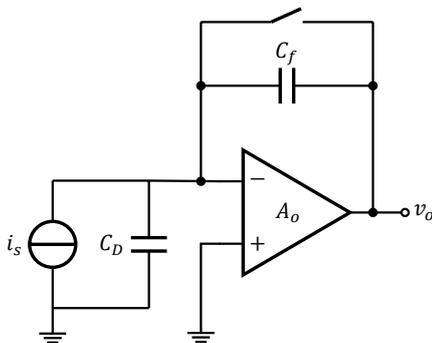


Figure 1.5: Block diagram of a charge sensitive amplifier.

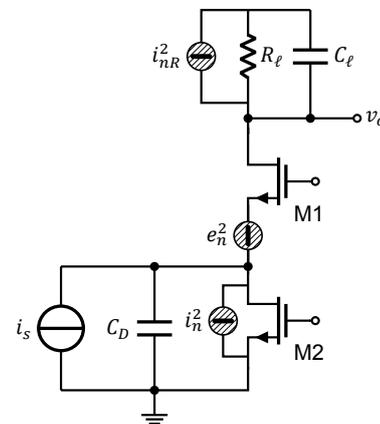


Figure 1.6: Block diagram of a common-gate TIA and its noise sources.

The block diagram of the feedback TIA is shown in Figure 1.7. As current flows into R_F , the amplifier forces its virtual ground to be constant causing the voltage $v_o = I_S R_F$ to be established. The amplifier gain suppresses the input resistance seen by the resistor, leading to the bandwidth of the TIA f_{3dB} being expressed by:

$$f_{3dB} \approx \frac{A_0}{2\pi R_F C_T} \tag{1.1}$$

Where A_0 is the amplifier's DC gain and C_T is the total capacitance seen at the input. Equation 1.1

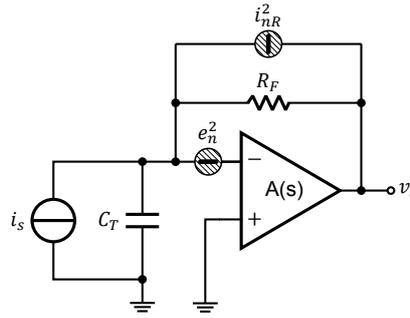


Figure 1.7: Block diagram of a feedback TIA and its noise sources.

holds only if the amplifier has its pole f_A at:

$$f_A = \frac{2A_0}{2\pi R_F C_T} \quad (1.2)$$

Feedback TIAs made in CMOS have two noise sources: (1) current noise due to the feedback resistor i_{nR}^2 and (2) voltage noise e_n^2 due to the channel noise of the MOS devices used in the amplifier. The input referred noise spectrum is:

$$i_n^2 = \frac{8kT}{R_F} + \frac{8kT\Gamma}{g_m} (2\pi f C_T)^2 \quad (1.3)$$

Where k is Boltzman's constant, T is the absolute temperature, Γ is Ogawa's noise factor, and g_m is the transconductance of the amplifier's input pair. Equations 1.1, 1.2, and 1.3 show that for an optimal and power efficient preamplifier one must:

- **Maximize R_F :** Increases the transimpedance gain and reduces the noise contribution of R_F . High transimpedance gain also reduces the subsequent gain required (see Figure 1.1) and the noise contribution of the following circuitry.
- **Minimize C_T :** Assuming the detector capacitance dominates the total input capacitance (i.e. $C_T \approx C_D$) and the TIA dominates the power consumption of the readout, then the sensor must be heavily segmented to reduce C_T . Minimizing C_T optimizes the readout power consumption because e_n^2 noise contribution scales with C_T^2 while g_m scales linearly with power. Reducing C_T also means that for the same R_F and f_{3dB} , the amplifier specifications are relaxed.

Numerical Example 1: the sensor segmented into 2400 pixels such that C_D is 50 fF and the readout's target power consumption is 1 mW. The amplifier is designed to match the detector capacitance for optimal noise performance [2] and to have a high g_m/I_D . It achieves a g_m of 5.5 mS while consuming 600 μ W of power.⁴ The feedback resistance is selected to be 200 k Ω , requiring an amplifier with A_0 of 27 and an $A_0 f_A$ of 10 GHz. Then, the TIA achieves a bandwidth of 280 MHz and an SNR of 4—concluding that the design is feasible.

Unfortunately, many CMOS non-idealities make the first-order approximation far from the truth:

1. Polysilicon resistors have crippling parasitics, a 200 k Ω poly-resistor introduces a pole at 140 MHz causing instability. R_F must be made smaller, reducing the transimpedance gain and increasing its noise and the noise contributed by subsequent circuitry.
2. Wideband circuits are power consuming, and the TIA no longer dominates the power consumption. Consequently, increasing the number of channels to very large values no longer reduces overall power consumption.
3. The application covered in this work requires ESD protection at its input. The ESD circuit, as drawn in Figure 1.8, adds a 100 fF capacitance and a 200 Ω resistor, limiting the minimum input capacitance that can be achieved and adding the thermal noise of R_{ESD} .

⁴The amplifier's g_m and power consumption values are extracted from an differential telescopic cascode design in 65nm CMOS.

4. Since all pixels must be read simultaneously, a large number of channels (i.e. 2400) leads to scaling problems with the digital circuitry and memory following the readouts.

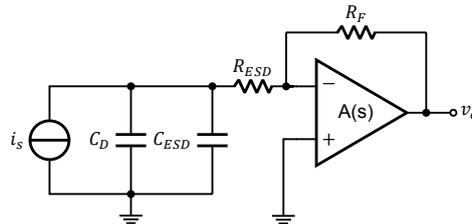


Figure 1.8: TIA circuit with ESD protection.

Numerical Example 2: Taking the above into account, the amplifier is adjusted to have 100 fF parasitic capacitance to match the ESD circuit, a g_m of 11 mS, and consume 1.2 mW. The detector capacitance is matched to the sum of the parasitics at the input (including the ESD and the amplifier), leading to 600 channels, a C_D of 200 fF, and a power consumption target of 4.2 mW per readout. Based on simulations, a bandwidth of 280 MHz requires an R_F of 35 k Ω to guarantee stability. The resulting SNR of this readout is less than 1.

Opportunities to improve SNR are limited. Increasing the TIA's power consumption while keeping the parasitic capacitance constant is inefficient (i.e. reduces g_m/I_D) and is limited by the power consumption specifications and the noise of R_{ESD} . Meanwhile, removing the ESD protection or cooling down the readout improves SNR but is impractical. On the other hand, reducing bandwidth can be very effective. The voltage noise contribution to the total noise scales with f^2 while R_F can be increased for narrower-band TIAs without destabilizing the circuit. To investigate how effective the bandwidth reduction is, the TIA is paired with an ideal gain stage and a comparator and the error-rate is computed as the bandwidth is swept. The results are plotted in Figure 1.9. The error-rate has a downward trend till ISI-induced errors start to dominate and the error-rate increases. Unfortunately, the system reaches its optimum error-rate at 15% concluding that the most straightforward readout topology, shown in Figure 1.1, fails to meet the targeted specifications.

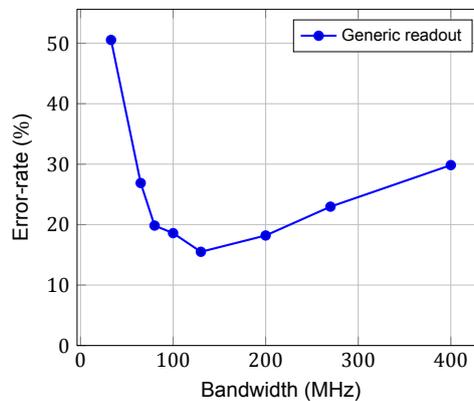


Figure 1.9: Bandwidth versus error-rate for a non-ideal TIA with ESD.

1.3. The Two-Threshold Comparator

The previous chapter concludes that the specifications set in Table 1.1 cannot be met by a standard solution. The input signal is too weak and the power budget is low, and hence, an acceptable wide-band SNR cannot be realized. Noise is a fundamental and nondeterministic source of errors; readout architectural changes cannot correct for them. The system bandwidth must be reduced to improve SNR. Unfortunately, at lower bandwidths, ISI-induced errors limit the minimum achievable error-rate. Luckily however, ISI is a deterministic source of errors. It occurs because, under limited bandwidth, the

pulse-response of a generated charges spreads out in time and interferes with the subsequent arriving electrons. And as mentioned earlier, it causes the distortion of '1' and '0' by bringing them closer to each other. But since ISI is predictable (i.e. deterministic), architectural changes can correct for some of its effects which is the main idea behind the two-threshold comparator.

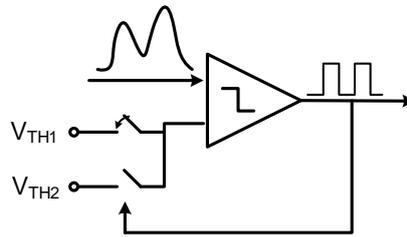


Figure 1.10: Block diagram of the two-threshold comparator.

A block diagram of the two-threshold comparator is shown in Figure 1.10. The two-threshold comparator works based on the observation that, under specific bandwidths, the 'pile-up' caused by ISI is limited. Pile-up refers to the build-up in voltage that occurs at the output of a slow preamplifier when inputs arrive in close proximity in time. But as seen in Figure 1.11, where an 80 MHz TIA is injected with a sequence of pulses, the pile-up saturates after the second successive pulse (i.e. no significant voltage build-up occurs between the second the third consecutive pulse.). This allows us to define two states in which arrival events can be classified:

1. Single pulse with no pile-up.
2. Successive pulses with limited pile-up.

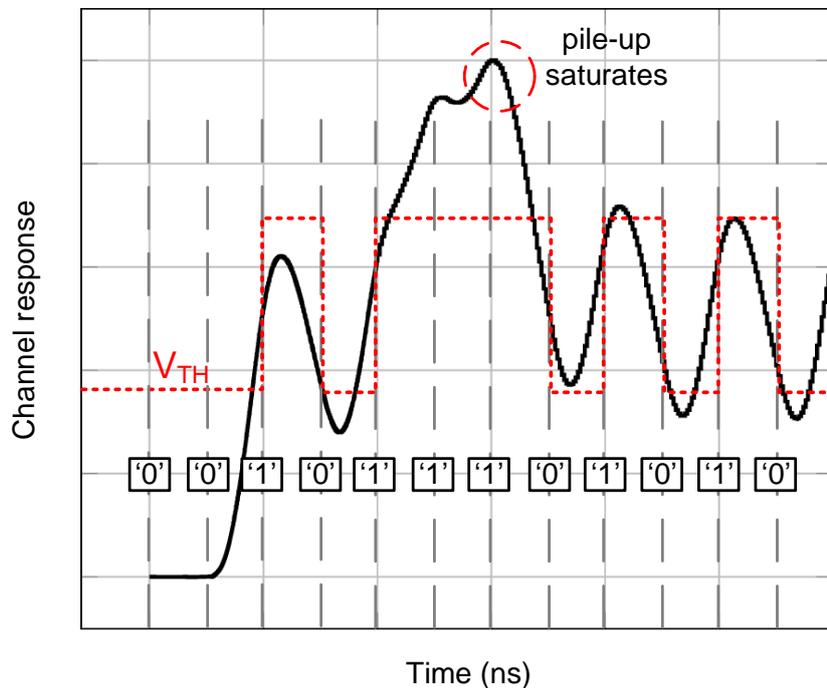


Figure 1.11: An example of the two-threshold comparator's operation. The grid represents the sampling moments.

For each state above, a threshold is defined to maximize correct detections. The threshold for the present discrimination is selected based on the previous decision as can be seen in Figure 1.11. This

assists the comparator to discriminate between arrival ('1') and non-arrival ('0') states. When a pulse has been detected the value of the threshold is increased, and hence, the comparator can more easily reject a subsequent '0' state while still being able to detect a subsequent '1' state thanks to pile-up.

Based on the aforementioned, the readout is designed to minimize nondeterministic noise-induced errors by reduction of bandwidth; intentionally inducing deterministic ISI. The negative effects of ISI are later removed using the two-threshold comparator. When plugged into the model created in the previous chapter, and swept over the TIA bandwidth (Figure 1.12), a significant improvement at low-bandwidths is observed. Of course, at extremely low bandwidths ISI becomes more severe and the error-rate rises again. Still, at its optimum, the two-threshold comparator improves the error-rate by over 12 %—allowing the readout to reach its target event-rate and accuracy with little added complexity and power consumption.

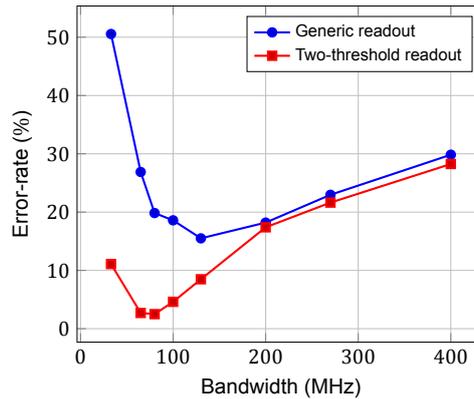


Figure 1.12: Bandwidth versus error-rate for the generic and the two-threshold readouts.

1.4. Thesis Organization

The remainder of the thesis is focused on designing a readout based on the two-threshold comparator in CMOS to meet the specifications highlighted in Table 1.1. Chapter 2 presents the detailed architecture and sets out the required performance of each block. The blocks are made in CMOS and their performance is verified in Chapter 3. System-level results based on post-layout simulations are summarized in Chapter 4. The designed chip and measurement setup used to evaluate the architecture are described in Chapter 5. Finally, Chapter 6 concludes and gives recommendations for future work.

2

Architecture

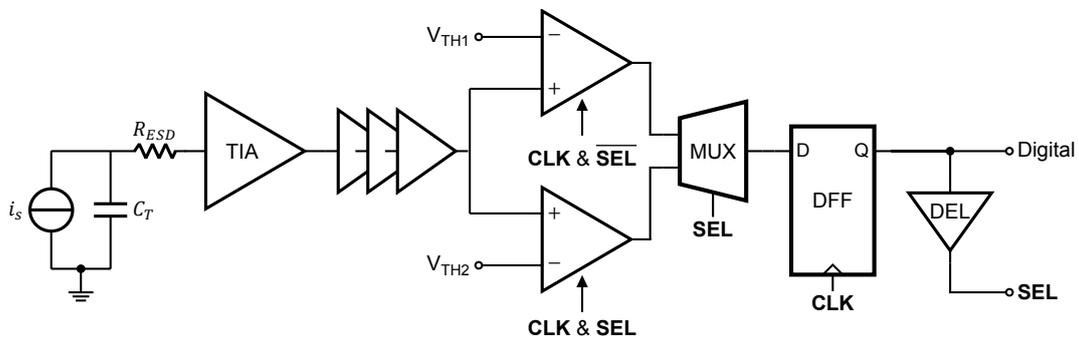


Figure 2.1: Block diagram of the two-threshold readout.

The beauty of the two-threshold technique is the significant reduction in error-rate for small amount of added complexity, the system block diagram presented in Figure 2.1 attests to that. The TIA is followed by a gain stage as in the generic readout introduced in Chapter 1. Then, one of two dynamic and multiplexed comparators is selected to sample the signal, representing the switch between the two thresholds. The selection signal is a delayed version of the output digital waveform which guarantees digital feedback loop stability. Since only one comparator is active at a sampling moment and the power consumption of digital circuitry is negligible, the two threshold architecture adds little to no additional power consumption. In this chapter, we set out to find the specifications of each block show in Figure 2.1, and later in Chapter 3, we design the blocks to meet them.

2.1. Feedback TIA

Being the system's bottleneck, the TIA must be carefully designed to optimize the performance of the readout. SNR and bandwidth, as emphasized in Chapter 1, are the most influential design parameters and are the focus of this section.

Minimizing the bandwidth is, theoretically, always beneficial because voltage noise contribution to the total noise scales with f^2 . Furthermore, the feedback resistor can be made large for lower-band TIAs. However, excessively slowing down the system induces excessive ISI which the two-threshold comparator cannot correct for—increasing the error-rate. The relationship between the error-rate and bandwidth was plotted earlier in Figure 1.12 showing that the error-rate reaches its optimum at around 70 MHz. Designing a TIA with a fixed bandwidth across process, voltage, and temperature (PVT) variations can be challenging. And it is especially the case for the feedback TIA circuit shown in Figure 1.7 due to the bandwidth being linearly proportional to A_0 which can vary with 1 dB to 4 dB across PVT. Adding a feedback capacitor C_F can help control the circuit's bandwidth such that:

$$f_{3dB} = \frac{1}{2\pi R_F C_F} \quad (2.1)$$

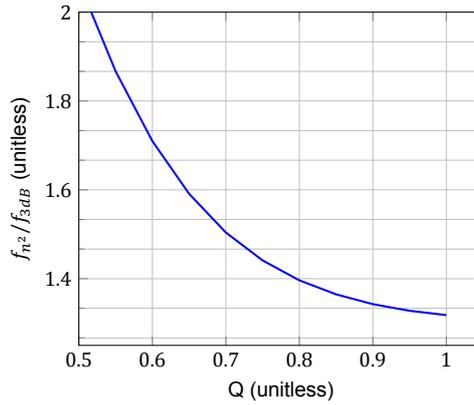


Figure 2.2: The quality-factor (Q) versus f_{n^2} normalized by the TIA bandwidth f_{3dB} .

Another benefit of C_F is that the zero it implements in the feedback network restores some phase lost to R_F 's parasitic poles.

The TIA, as a second-order system, has an associated quality-factor Q which turns out to be an important design parameter for the two-threshold readout. As seen in Figure 2.2, higher Q values lead to a reduction of equivalent noise bandwidth of the voltage noise f_{n^2} and consequently improves of the readout's SNR (Figure 2.2). However, higher Q values affect the readout time response in an unfriendly way. Figure 2.3a plots the pulse response of a narrowband TIAs with different quality-factors. A Higher Q widens the pulse causing it to interfere more severely with subsequent inputs and introduces an undershoot causing a 'pile-down' effect in addition to the 'pile-up' discussed in Chapter 1. This can be observed in the TIA's response to the '0101010' input pulse sequence shown in Figure 2.3b, where higher Q value leads to further distortion in the '1' and '0' levels. The effect of Q on the error-rate for different bandwidths and SNRs is plotted in Figure 2.4. Although higher Q s improve SNR, the error-rate increases due to ISI, and hence, the TIA is designed for $Q = 0.5$. Figure 2.4b also shows that for an SNR of 3.5 the readout can meet the target error-rate while leaving some margin for other sources of error.

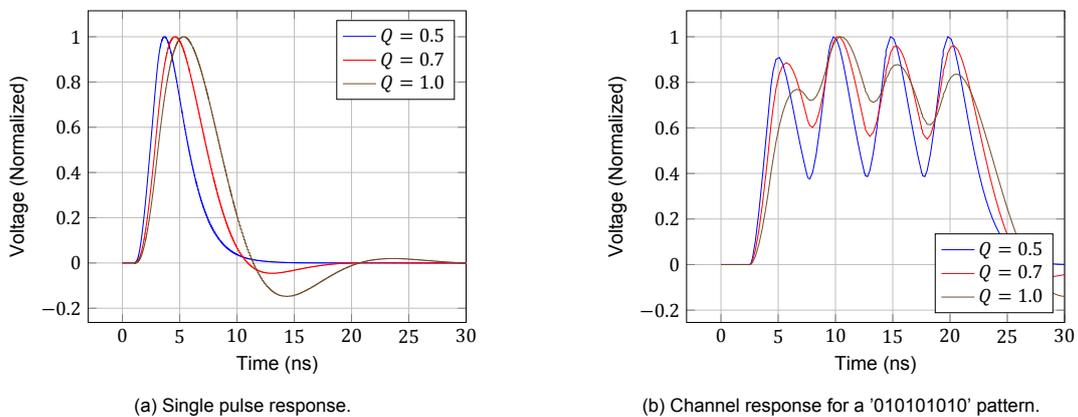


Figure 2.3: Effects of the quality factor Q on the readout time response and ISI.

Other TIA parameters such as R_F and C_F along with the specifications of the voltage amplifier are left to the implementation chapter as it requires consideration of technological limits and the complex models of poly-silicon resistors. The rule of thumb followed in the implementation is to maximize R_F to minimize noise and maximize transimpedance gain.

2.2. Gain Stage

The gain stage is responsible for boosting the voltage levels of the signal produced by the preamplifier to protect it from comparator non-idealities such as offset and kickback. The primary design parameters

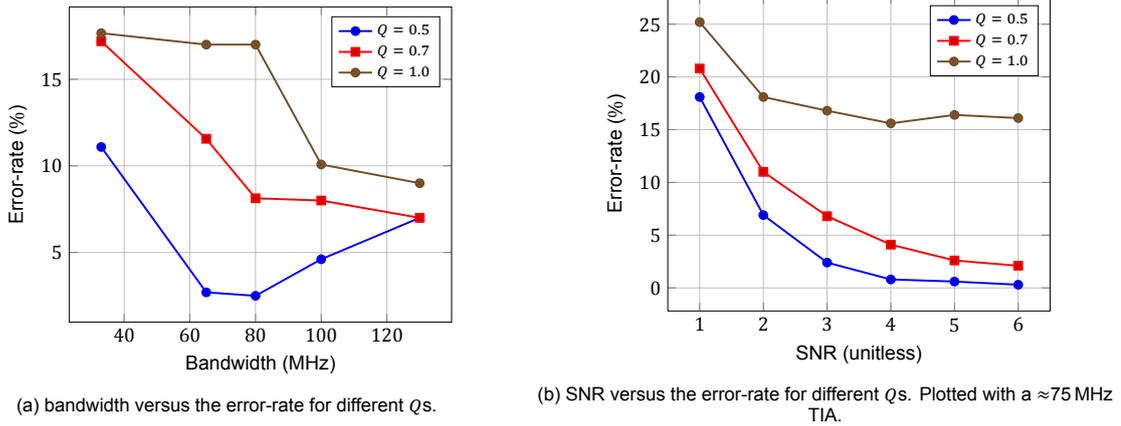


Figure 2.4: Effects of the bandwidth, SNR, and quality factor Q on the error-rate.

for the gain stage are its gain and bandwidth.

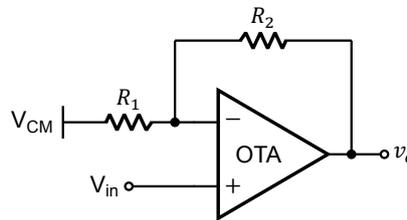


Figure 2.5: Block diagram of a feedback amplifier.

A closed-loop feedback amplifier (Figure 2.5) is selected as it can provide accurate gain with moderately large bandwidths. To determine the bandwidth of the amplifier, we note that the total bandwidth of the readout is a combination of the bandwidth of each block and is approximated by:

$$\frac{1}{f_{3dB,TOT}^2} \approx \frac{1}{f_{3dB1}^2} + \frac{1}{f_{3dB2}^2} + \dots \quad (2.2)$$

For example, if the gain stage has twice the bandwidth of the TIA (i.e. 140 MHz), the circuit bandwidth falls below 60 MHz, inducing ISI and increasing the error-rate. Since the TIA is designed to obtain the desired readout bandwidth, the gain stage must be designed with a significantly higher speeds to maintain it. Figure 2.6 plots the effect of the gain stage's bandwidth on the error-rate. An amplification bandwidth of ≈ 500 MHz ($7\times$ the TIA bandwidth) guarantees no degradation in error-rate and hence is set as the target specification.

The amount of gain required depends on the transimpedance gain (i.e. R_F) and the saturation voltage at the output of the TIA after the pile-up effect takes place—it can range from 20 V/V to 30 V/V. Generally, it is not possible to achieve such high gain values while maintaining a wideband. Hence, the gain stage must be split into two separate amplifiers. Splitting the gain stage reduces the gain required of each amplifier to $\sqrt{\text{Gain}_{TOT}}$. While if they amplifiers have the same bandwidth, the total bandwidth of the gain stage is reduced to $0.67f_{3dB,AMP}$. Consequently, each amplifier bandwidth must be increased by $1/0.67\times$ (≈ 750 MHz) to compensate for the bandwidth lost. Note however that splitting the gain stage reduces the gain-bandwidth required for each amplifier and hence improves the feasibility of the implementation. As for the gain accuracy, simulations show that for 10% gain-error the error-rate increases by 2%. To keep the error-rate contribution of the gain stage small, each amplifier must achieve their gain with $>1\%$ inaccuracy.

Offset can significantly affect the accuracy of the readout as will be shown in the next two sections. Unfortunately, a fast amplifier means small area and high offset. To address the total offset appearing from the channel (i.e. the TIA and gain stage), the voltage $V_{CM,AMP2}$ connected to the second amplifier in the gain stage is adjusted by an external voltage source to cancel the channel's offset. This emulates a servo-loop, the implementation of which is out of the scope of this work.

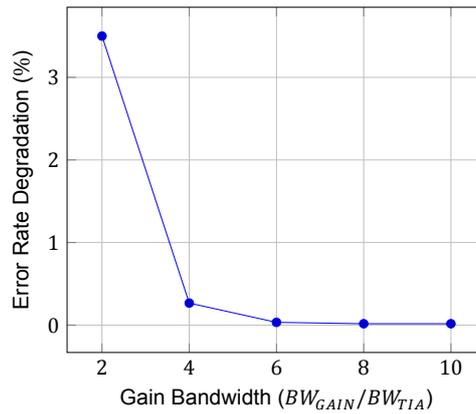


Figure 2.6: Effects of gain stage's bandwidth on the error-rate.

2.3. Comparator

One of the advantages of switching between comparators rather than switching between two-thresholds is the relaxation of the decision time specification of the comparator as no time needs to be reserved for threshold values to settle. Therefore, assuming 50% of the clock is assigned to the comparator's sampling phase, the comparators must make their decision in under 1 ns, reserving time for the storage device (DFF) to store the decision. The critical design parameter for the comparator is its offset voltage. Swept in Figure 2.7, the offset to error-rate relation shows that for ± 20 mV of offset the error-rate degrades by 3%. Subsequently, we limit the total offset seen at the input of the comparator to 10 mV.

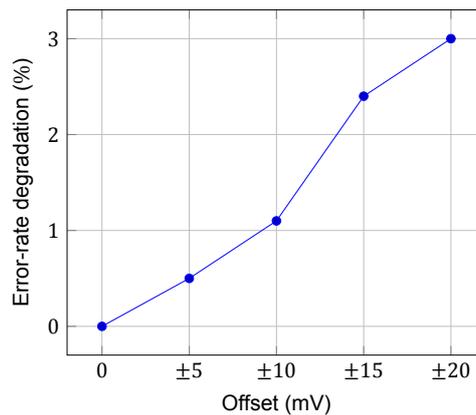


Figure 2.7: Error rate degradation versus offset seen at the input of the comparator.

2.4. Thresholds

Selecting adequate thresholds is critical to the readout's performance. And although formulation exists for optimal threshold values of digital signals arriving at a receiver at fixed times, none exists for electrons hitting a detector at random times. The way in which the random arrival of an electron affects its detection is illustrated in Figure 2.8. At particular times, it is possible that the comparator samples the signal when it's near the threshold value—giving noise the upper-hand in making the decision. Additionally, note how for a high V_{TH1} , an electron arriving at such compromised times can be missed by both sampling moments—giving rise to missed detections (i.e. false negatives). On the flip side, if V_{TH1} is too low, noise can be detected as generated charge—giving rise to false detections (i.e. false positives).

Deciding on the threshold values is done by an exhaustive search, optimizing their values to equate the rate of false positives and false negatives. The test signals used to search for the threshold values are shown in Figure 2.9. In Figure 2.9a, a single pulse is injected into the readout with a variable delay in steps Δt of 100 ps. For each delay, 100-samples of the readout's decision are recorded,

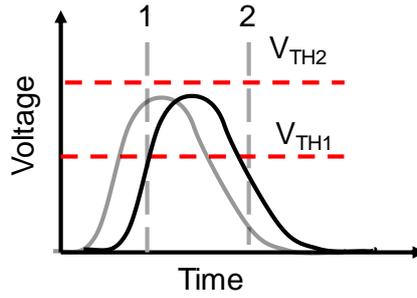


Figure 2.8: Example of a randomly arriving signal sampled at sensitive spots.

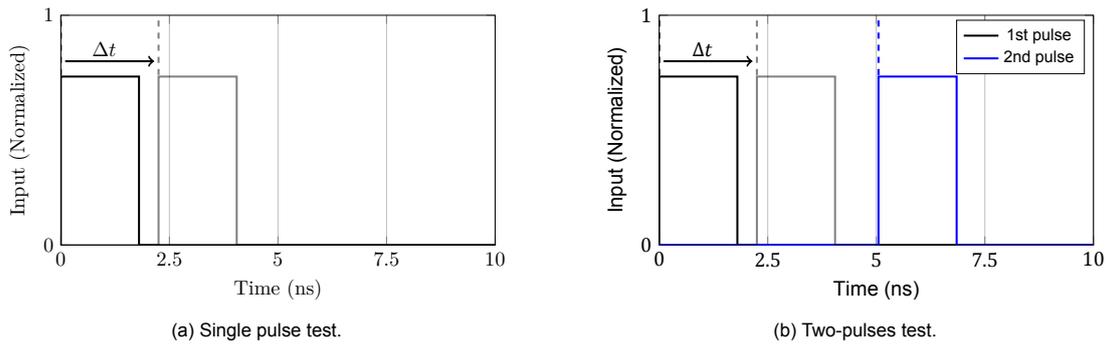
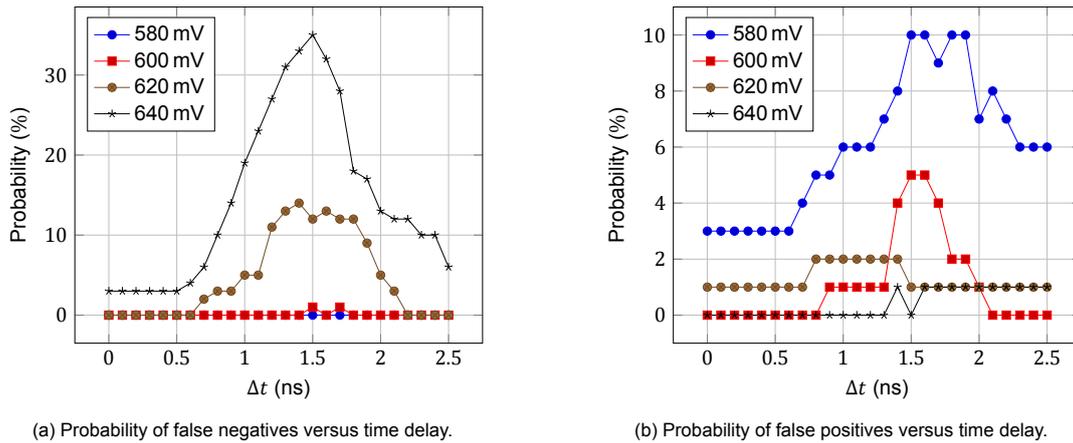


Figure 2.9: Tests to search for the optimal threshold values.

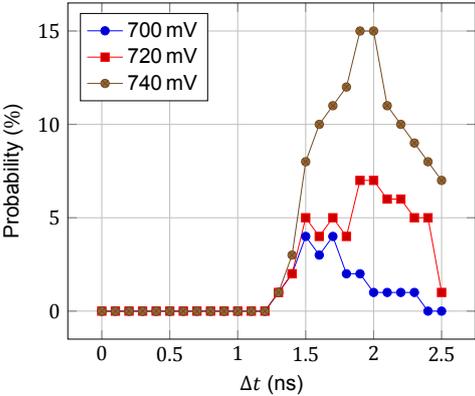
and subsequently, probabilities of false negatives and positives are derived for different first threshold values. For the second threshold, the swept pulse is followed by another pulse generated at a fixed time, and similar probabilities are derived for different threshold values. The results of the search are plotted in Figure 2.10 and 2.11. The selected values of the first and second thresholds are 610 mV and 720 mV respectively as they balance between false positives and false negatives. The reader can also observe that dramatic change in error probabilities for little change in the threshold values, emphasizing the effects of offset on the error-rate as discussed in Section 2.3.



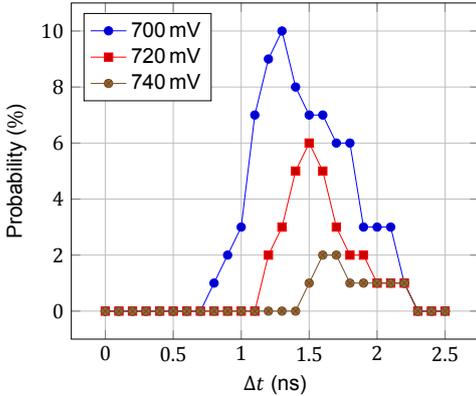
(a) Probability of false negatives versus time delay.

(b) Probability of false positives versus time delay.

Figure 2.10: Effects of the first threshold value V_{TH1} on the probability of false positives and false negatives.



(a) Probability of false negatives versus time delay.



(b) Probability of false positives versus time delay.

Figure 2.11: Effects of the second threshold value V_{TH2} on the probability of false positives and false negatives.

3

Design & Implementation

This chapter summarizes the design effort of the readout's functional blocks and verifies each block performance against the specifications purposed in Chapter 2. The readout is implemented in TSMC's General Purpose 65nm CMOS with a core voltage of 1 V, 2 fF MIM-cap density, and a 9-metal stack.

3.1. Feedback TIA

The TIA specifications are summarized in Table 3.1 next to a block diagram of the compensated feedback drawn in Figure 3.1. Beginning with the unknowns R_F and C_F , and noting that the bandwidth is proportional to $1/R_F C_F$, maximizing R_F means minimizing C_F . Unfortunately, C_F cannot be implemented arbitrarily small as it needs to be trimmed as will be shown below. Moreover, small C_F values are more affected by parasitics and PVT variations, and hence have reduced reliability. Consequently, we C_F set to be at least 15 fF. Following the bandwidth equation of the compensated TIA, for a bandwidth of 70 MHz, R_F must be set to 130 k Ω .

Table 3.1: TIA target specifications.

Specification	Target
Bandwidth	70 MHz
Q	0.5
SNR	3 to 4
R_F	Maximize.

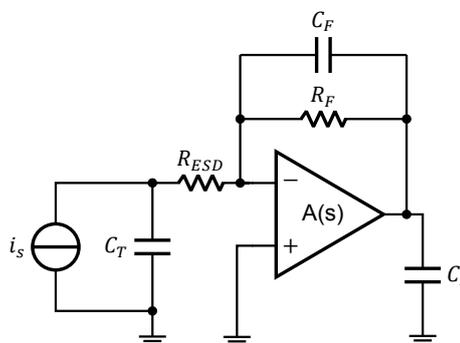
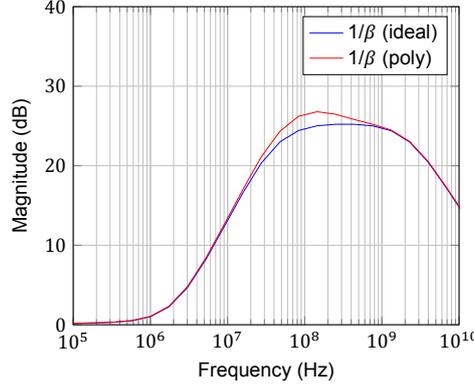


Figure 3.1: Compensated transimpedance amplifier.

Fortunately (and also, unfortunately) poly-resistors have parasitics and show inductive behavior at high-frequencies. This behavior can be utilized to tune-out some of C_F (i.e. inductive peaking) and increase the value of R_F . Due to the complexity of the resistor model and its non-lumped nature, this is done iteratively until a suitable (and stable) $R_F C_F$ frequency is achieved. The resultant is an R_F of 220 k Ω and a C_F of 18 fF. In theory and as shown in Figure 3.2, this should achieve an $R_F C_F$ frequency of 40 MHz. But thanks to R_F 's inductive behavior, the combination results in a corner frequency of ≈ 75 MHz. In equivalence, the feedback capacitor appears to be 10 fF instead of an 18 fF.

Now that R_F and C_F are selected, the open-loop gain and speed on the voltage amplifier can be defined. Equations derived for the compensated TIA do not do well in predicting the A_0 and bandwidth required to obtain the desired response, especially for a higher order system with zeros and a low-quality factor [9]. However, they can provide an initial guess that can be tweaked by simulations. Equation 3.1 shows that for $f_{3dB} = 1/2\pi R_F C_F$ to hold, $(A_0 + 1)C_F$ must be much greater than C_T . The total input capacitance is 400 fF, and since C_F is appears to be a 10 fF capacitor, A_0 must be much

Figure 3.2: $1/\beta$ plot of the feedback TIA with ideal and poly-resistors.

greater than 40 V/V and is therefore set to 400 V/V (52 dB). Meanwhile, the OTA time constant T_A is estimated by Equation 3.2 and is set to 11 ns (equivalent to an amplifier pole at 14 MHz). Simulations with an ideal amplifier (Figure 3.3 and 3.4) show that an amplifier with an open-loop gain of ≈ 55 dB and an dominant pole at 10 MHz is able to realize the desired bandwidth and pulse response (i.e. Q).

$$f_{3dB} = \frac{A_0 + 1}{2\pi R_F (C_T + (A_0 + 1)C_F)} \quad (3.1)$$

$$T_A \approx \frac{1}{4\pi^2} \frac{2A_0}{R_F (C_T + C_F) f_{3dB}^2} \quad (3.2)$$

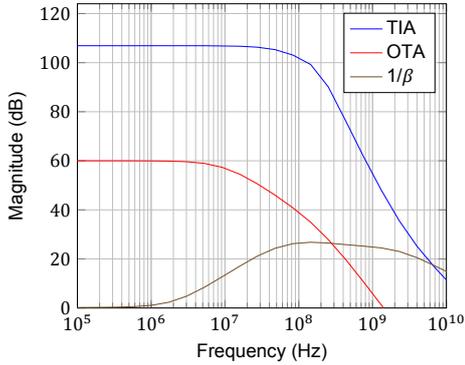
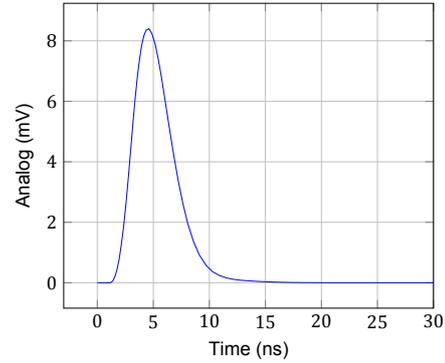
Figure 3.3: Frequency response of the TIA, OTA, and $1/\beta$ with real feedback elements and an ideal amplifier.

Figure 3.4: Transient response of the TIA with real feedback elements and an ideal amplifier.

Achieving such a high-gain bandwidth product (≈ 6 GHz) demands a two-stage design, especially in short-channel technologies. The schematic of the implemented amplifier is shown in Figure 3.5. The first stage is a short-channel, cascoded, and high-power stage that is only loaded by the second stage—resulting in a low input capacitance, relatively high-gain, a wideband, and low-noise. Furthermore, the resistor R_G is placed at its output to trade gain for bandwidth, moving its pole to a higher frequency to guarantee stability. On the contrary, the second stage is a long-channel low-power stage loaded with C_L , resulting in high-gain and a dominating pole. The transistor sizing strategy was as follows:

- **M₁, M₂**: Sized for an input capacitance of 100 fF.
- **M₅, M₆, M₉**: Sized to ensure the 1st stage's output and the gate of the diode connected device are approximately at $V_{DD}/2$.
- **Remaining**: Sized for proper operating conditions across corners.

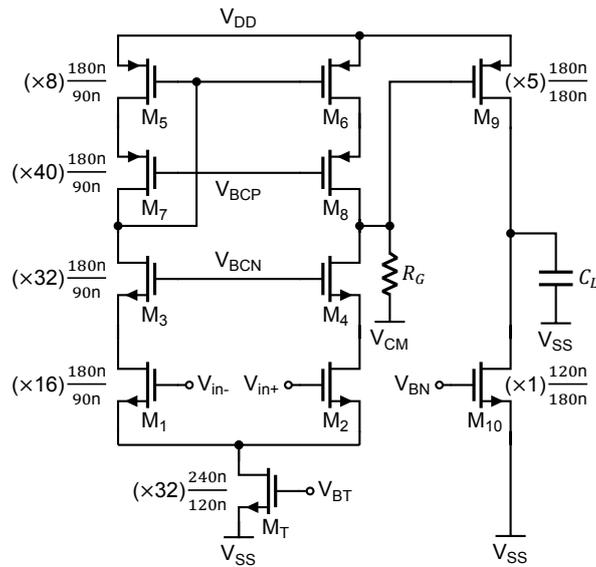


Figure 3.5: Schematic of the TIA's voltage amplifier.

As for noise, the current drawn by the first stage is selected such that the TIA circuit has an SNR of 3.5 and consequently set to 1.25 mA. The noise contribution breakdown is summarized in Figure 3.8 showing that voltage noise sources (the OTA and R_{ESD}) dominate the total noise as expected from a wideband low-power TIA. Finally, note that the values of R_F , C_F , and C_L directly affect the TIA's gain, bandwidth, and Q and must be trimmed to account for process variations. Figure 3.6 plots the error-rate degradation versus variation in the values of the aforementioned components. The feedback resistance plays the most critical role in the definition of the TIA parameters and causes severe error-rate degradation as it varies. And although the values C_F and C_L affect the readout less prominently, their values must be slightly adjusted across corners to meet the target specifications. Therefore, we choose to trim R_F to a $\approx 2.5\%$ tolerance, and C_F and C_L to a $\approx 5\%$ tolerance.

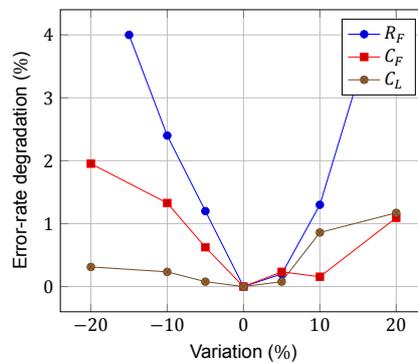


Figure 3.6: Effect of TIA's component variations on the error-rate.

The layout of the TIA including the feedback elements, load capacitance, and the input passives emulating the PIN diode's capacitance and ESD circuit is shown in Figure 5.1. Figure 3.10 shows the achieved TIA pulse response (encompassing f_{3dB} and Q) across the typical, fast, and slow corners while Figure 3.9 shows the TIA's frequency response. Table 3.2 summarizes the TIA performance.

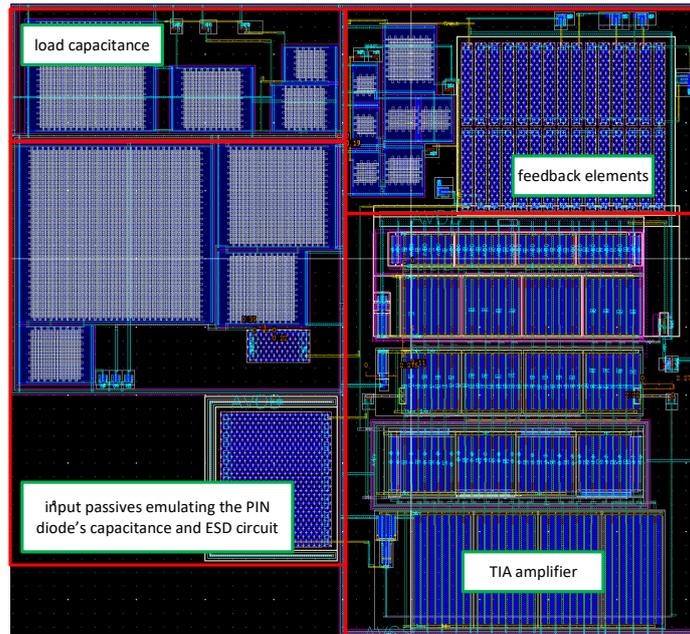


Figure 3.7: TIA circuit layout.

Table 3.2: Summary of the TIA's performance—post-layout simulations.

Specification	Min.	Max.
Transimpedance gain	216 k Ω	230 k Ω
Bandwidth	72 MHz	74 MHz
Q	0.55	0.6
SNR	3.4	3.5
Power	1.15 mW	1.3 mW

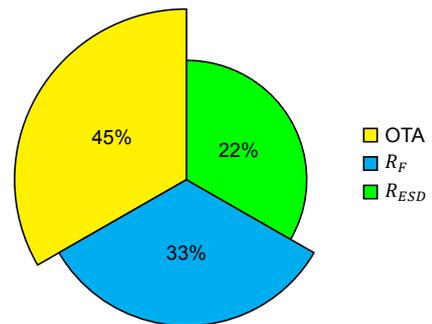


Figure 3.8: TIA circuit noise breakdown.

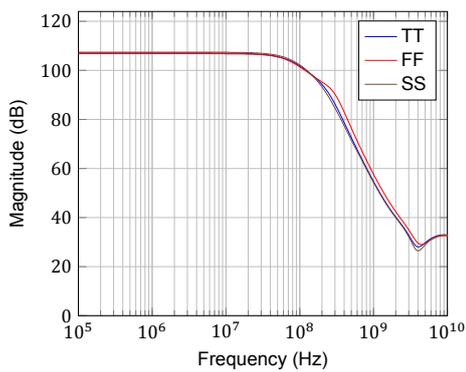


Figure 3.9: Frequency response of the implemented TIA across critical corners—post-layout simulations.

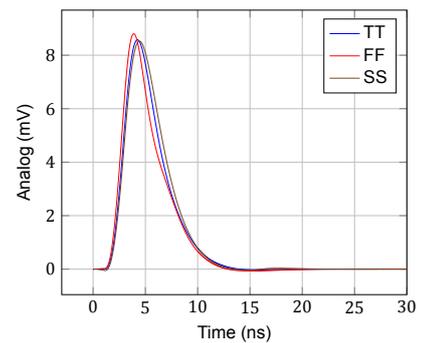


Figure 3.10: Transient response of the implemented TIA across critical corners—post-layout simulations.

3.2. Gain Stage

The pile-up of the TIA saturates at 15 mV after the detector generates multiple consecutive charges. Accounting for an amplifier clipping voltage of 400 mV and leaving a margin such that noise does not cause clipping (readout has a low SNR), each amplifier in the gain stage must provide a gain of 5. Table 3.3 summarizes the target specifications.

Table 3.3: Amplifier target specifications.

Specification	Target
Gain	5
Gain Error	<1 %
Bandwidth	800 MHz
Phase margin	>45°

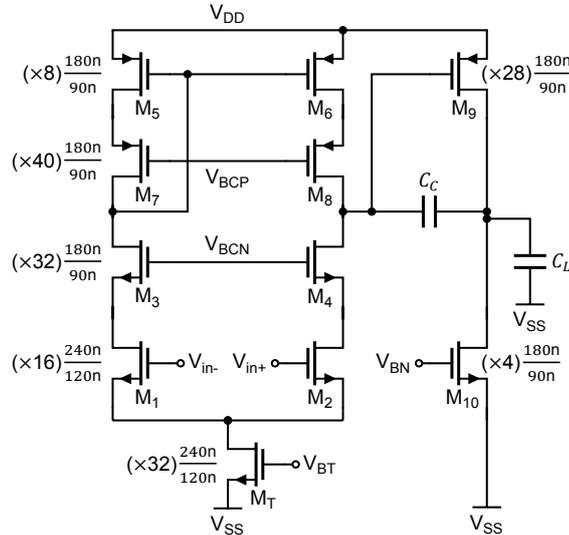


Figure 3.11: Schematic of the amplifier OTA.

Figure 3.11 illustrates the schematic of the amplifier. Besides replacing M1 and M2 with longer-channel devices to improve the open-loop gain, the input stage is identical to that of the TIA. A compensation capacitor C_C however must be added to split the poles of the amplifier located at:

$$\omega_{p1} \approx \frac{1}{r_{o1}(A_2 + 1)C_C} \quad (3.3)$$

$$\omega_{p2} \approx \frac{g_m}{C_L} \quad (3.4)$$

Where A_2 is the second stage DC gain. Short-channel devices are used for the second stage to reduce its output impedance and maximize its transconductance. The power consumed by both stages needs to be set while taking into account the values of R_1 and R_2 , noting that:

1. **Large** R_1 : The poly-resistor's parasitic pole degrades the phase margin for the same bandwidth.
2. **Large** R_1 : Input referred noise is significant due to low TIA gain.
3. **Small** R_1 : High power required for load driving.

An iterative optimization results in an R_1 of 8 k Ω while $R_2 = 4 \times R_1$. The first stage consumes 100 μ W to match its noise contribution to that of R_1 . With an assumed load capacitance of 25 fF, the second stage also consumes 100 μ W to drive its load and provide a good phase margin. Finally, the compensation capacitor is set to be 50 fF. The layout of a single amplifier is shown in Figure 3.12. Figure 3.13, Figure 3.14, and Table 3.4 show the achieved performance across critical corners.

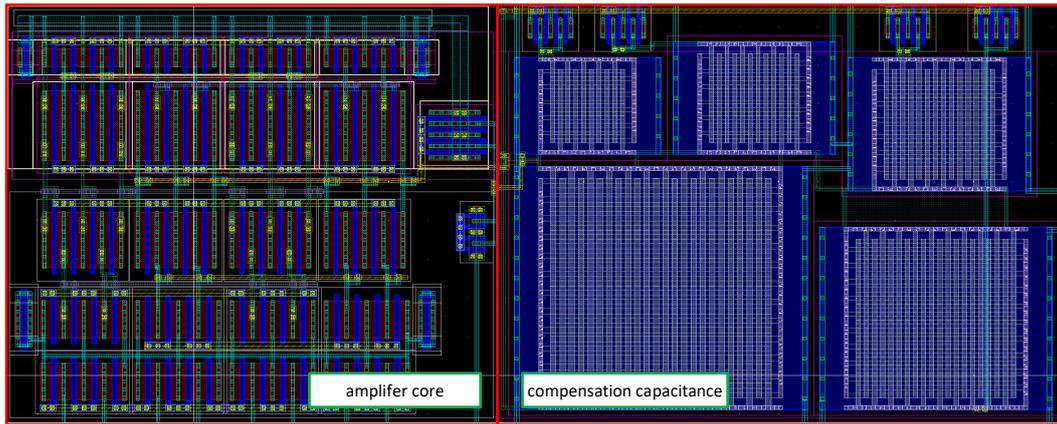


Figure 3.12: Layout of a single amplifier.

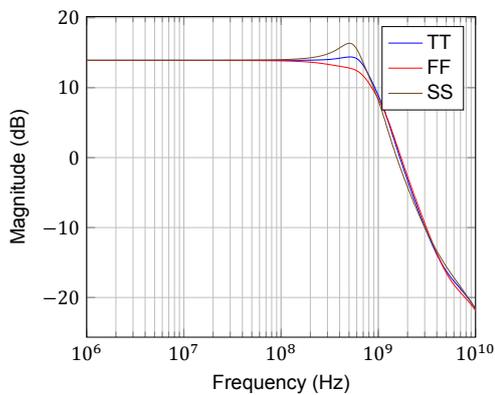


Figure 3.13: Frequency response of the implemented amplifier across critical corners—post-layout simulations.

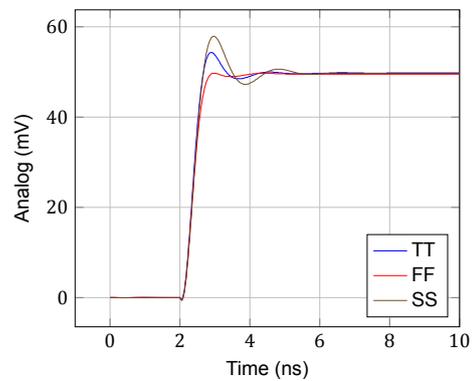


Figure 3.14: Transient response of the implemented amplifier across critical corners—post-layout simulations.

Table 3.4: Summary of the amplifier's performance—post-layout simulations.

Specification	Min.	Max.
Gain	4.95	4.98
Bandwidth	780 MHz	860 MHz
Phase Margin	46°	63°
Power	240 μ W	260 μ W

3.3. Comparator

Shown in Figure 3.15 is a double-tail comparator which offers a well-rounded choice for high-speed low-supply designs [13]. Furthermore, it requires no DC-current and hence achieves low-power consumption. When the clock is 'HIGH', the first stage amplifies the difference between the inputs via an integration on nodes V_{DIN} and V_{DIP} [13]. The second stage then latches on the difference and gives a logical '0' or '1' output, holding it steady until the clock is 'LOW'. Sizing the transistor has been mainly done to meet the offset requirements without giving rise to excessive kickback noise. The sizing trade-offs are:

- M_1, M_2 : Offset proportional to $1/\sqrt{WL}$ and kickback is proportional to the area.
- M_3, M_4 and M_9, M_{10} : Contributes to offset and determines decision time.
- M_{TN}, M_{TP} : Decision time and kickback.

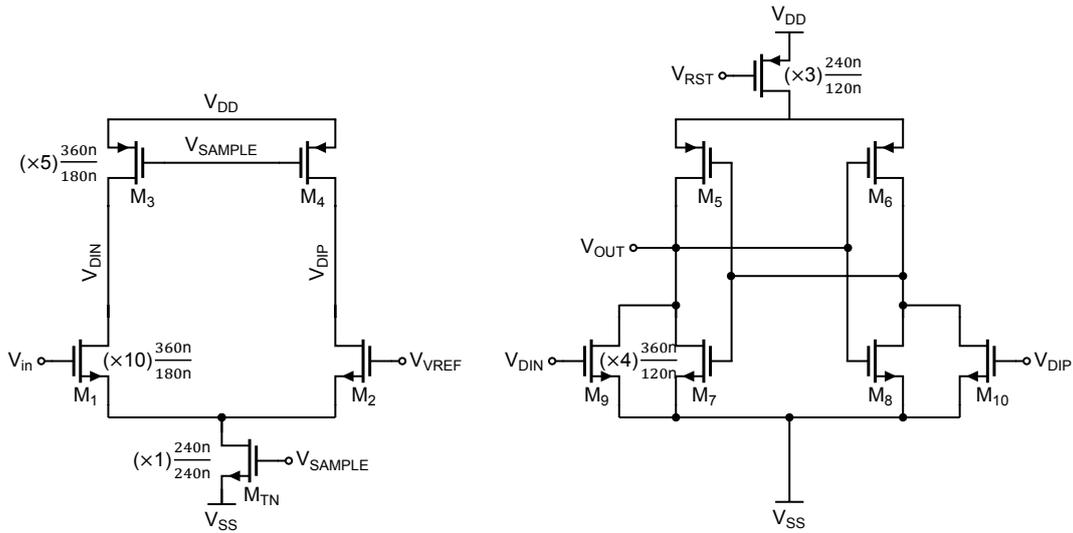


Figure 3.15: Schematic of the double-tail comparator.

The designed comparator has an offset σ of 3.5 mV and has a kickback of 3 mV ($2\sigma + \text{kickback} = 10 \text{ mV}$). It can make decisions within 500 ps for a 1 mV voltage difference at its inputs while only consuming 100 μW per decision. Figure 3.16 shows the layout of the comparator and Table 5.1 summarizes its performance characteristics across corners.

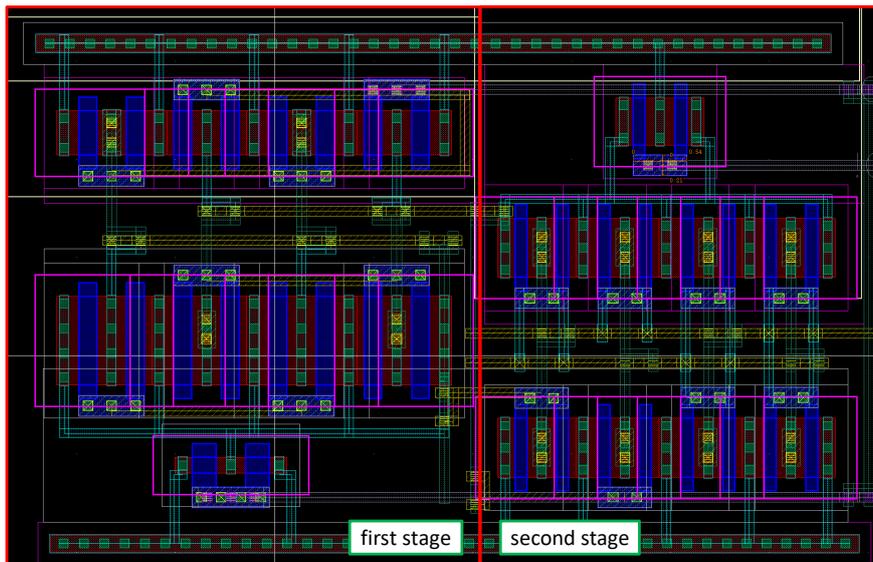


Figure 3.16: Layout of the comparator.

Table 3.5: Summary of the comparator's performance.

Specification	Min.	Max.
Decision time	300 ps	500 ps
Power	75 μW	100 μW
Offset	$\sigma = 3.5 \text{ mV}$	

3.4. Buffers

Although not mentioned in Chapter 2, wideband buffers are required for the proper operation of the circuit. They provide a low-impedance V_{CM} and $V_{TH1,2}$ connections to the amplifiers in the gain stage and comparators respectively. The buffer performance requirements is derived from the other blocks, and they must:

- Have a bandwidth higher than the amplifiers as not to limit the overall circuit bandwidth and to match their output impedance.
- An input range of 450 mV to 720 mV to accommodate for the adjustable offset canceling voltage (delivered to the second amplifier in the gain stage) and the second threshold value.
- An offset σ of <1 mV to limit its contribution to the total offset seen by the comparators.

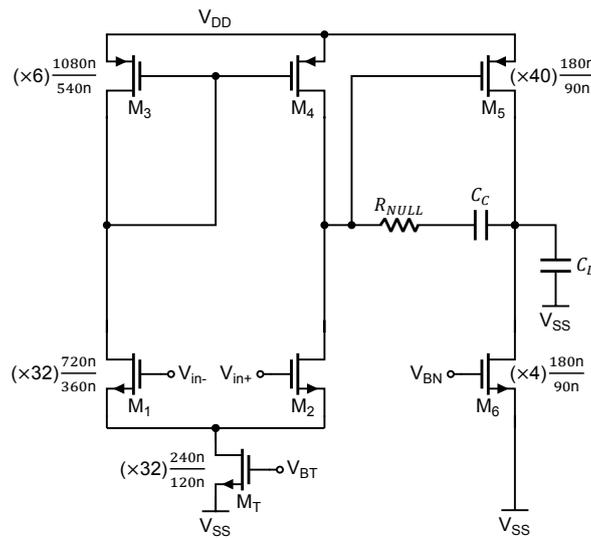


Figure 3.17: Schematic of the buffer OTA.

Figure 3.17 shows the block diagram and the schematic of the buffer. The cascodes are removed to improve input range and subsequently the length of the input transistor and current sources are increased to recover the lost open-loop gain. Additionally, larger input pairs and current source have lower offsets and help meet the buffer's σ specification. The second stage is identical to that of the amplifiers' to match their output impedance and minimize differential kickback. Differential kickback is the difference between the kickback seen at the output of the amplifier and the one seen at the output of the buffers connected to the comparator. Due to an increase in size of M1–M4 while the current drawn by the first stage remains low (≈ 50 μ A), the mirror pole introduced by M3 moves to lower frequencies and degrades the phase margin. To improve the phase margin, a nulling resistor is added to the compensation network, implementing a zero. C_C is chosen to 75 fF and R_{NULL} is 1.5 k Ω to set the zero location at ω_{p2} . The layout of the buffer is shown in Figure 3.18. Figure 3.19 shows the differential kickback between the buffer and the gain stage which reaches a maximum of 3 mV at the sampling moment (at ≈ 1.5 ns). The buffer's frequency and transient performance is shown in Figure 3.20 and Figure 3.21 respectively, and summarized in Table 3.6.

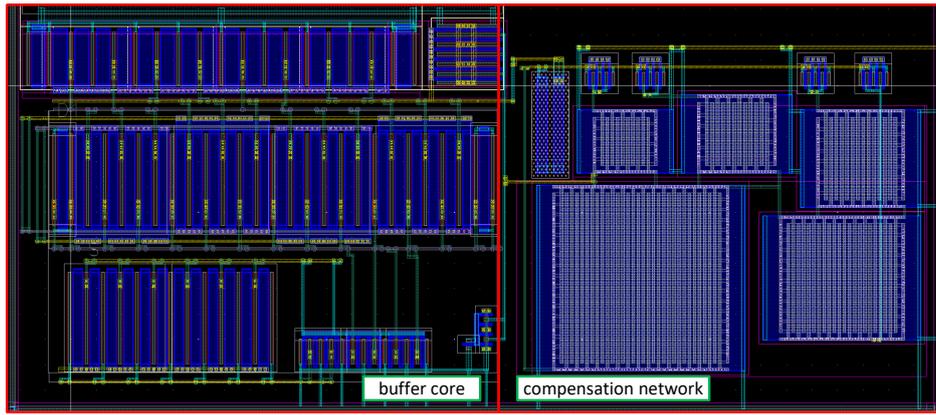


Figure 3.18: Layout of the buffer.

Table 3.6: Summary of the buffer's performance—post-layout simulations.

Specification	Min.	Max.
Bandwidth	800 MHz	1 GHz
Phase Margin	52°	56°
Power	185 μ W	205 μ W
Offset	$\sigma = 0.8$ mV	

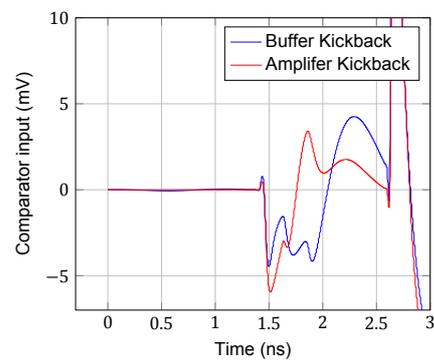


Figure 3.19: Comparison of the kickback at the amplifier and buffer outputs.

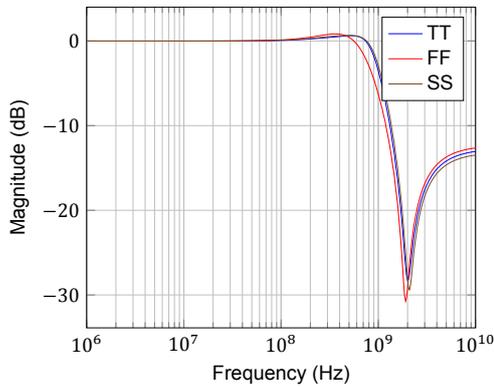


Figure 3.20: Frequency response of the implemented buffer across critical corners—post-layout simulations.

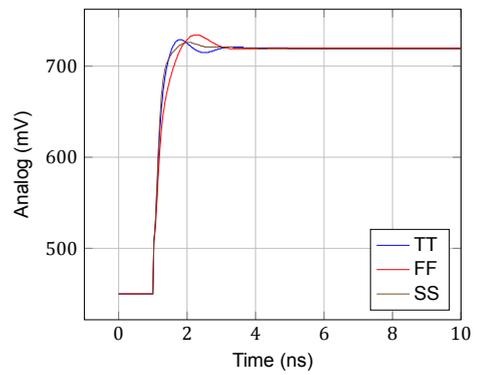


Figure 3.21: Transient response of the implemented buffer across critical corners—post-layout simulations.

4

Results

The two-threshold architecture shown in Figure 2.1, consisting of a TIA, two amplifiers, two comparators, and four buffers, has been laid-out to verify the architecture's effectiveness. The layout of the readout is shown in Figure 4.1. All data in this chapter are simulated under the following conditions:

- Post-layout RC extraction.
- Analog and digital supply voltages of 1 V.
- A Bondwire of 3 nH and 100 pF on-chip decoupling.

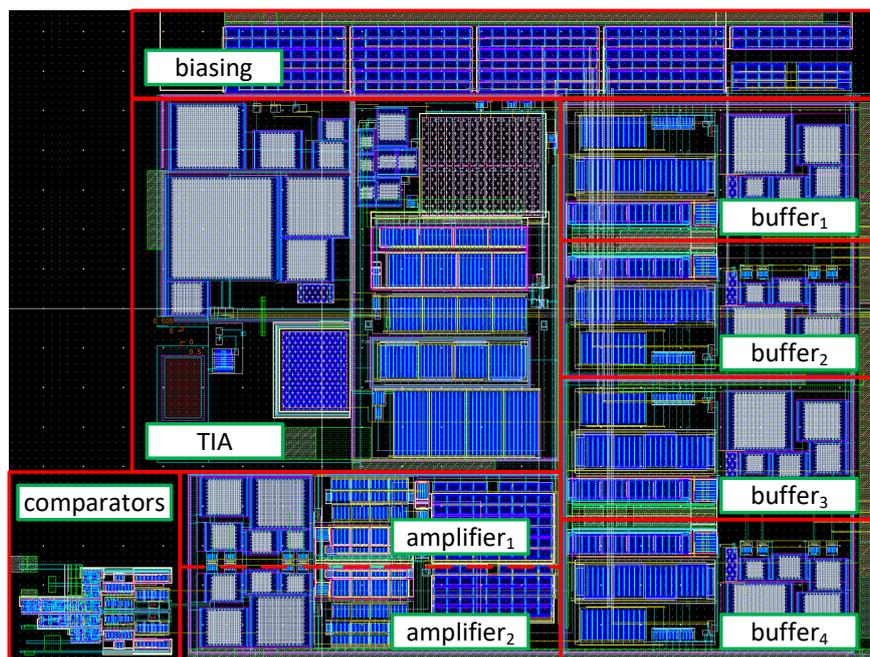


Figure 4.1: Layout of the readout including the TIA, amplifiers, comparators, and buffers.

The simulation test-bench used to evaluate the readout's error-rate is illustrated in Figure 4.2. The readout is first calibrated for the corner under-test which includes trimming the passive elements and eliminating the offset of the channel. Afterward, a random test signal generated in MATLAB is injected into the TIA and the digital output of the readout is recorded. The error-rate can then be readily calculated.

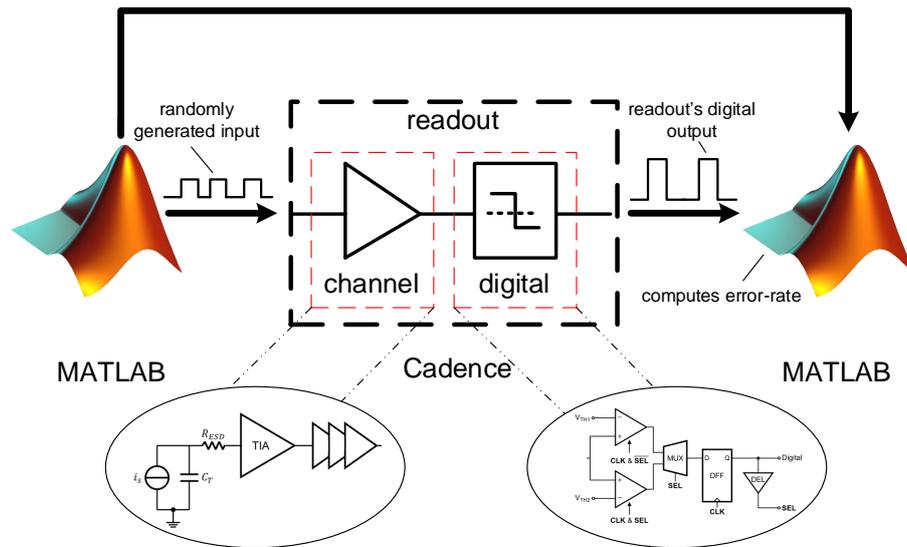


Figure 4.2: Simulation setup to evaluate the readout's error-rate.

Twenty-five Monte-Carlo samples have been collected for each critical corner, the statistical results of which are plotted in Figure 4.3 and summarized in Table 4.1. Under typical conditions, the architecture has an average error-rate of 2.2%. And in its worst case, the readout achieves a 3σ error-rate of 4.8% and hence meets the targeted specification. Also plotted (Figure 4.4) is the error-rate's distribution when the two-threshold technique is disabled (i.e. only the first threshold is used), and on average, the error-rate increases by 20%. Figure 4.5 shows the error breakdown of each corner and gives insight on the potential reason for the degradation of the error-rate at the slow-corner. Both the typical and fast corners have approximately equal rates of false negatives and false positives, indicating optimized threshold values. Meanwhile, the slow corner has a dominant rate of false positives. Optimizing the threshold values for the slow corner could correct for some of the erroneous detections, however, an exhaustive search to find them has not been performed.

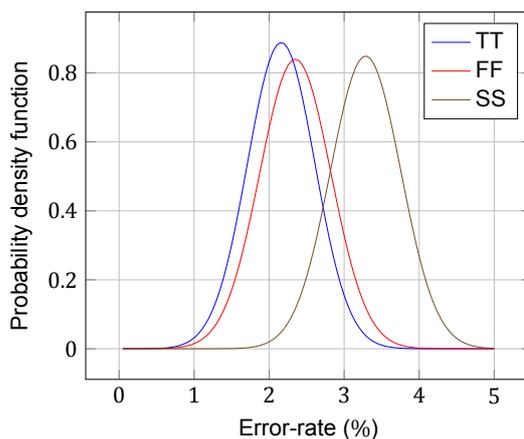


Figure 4.3: Error-rate distribution from a Monte Carlo simulation with $n = 25$.

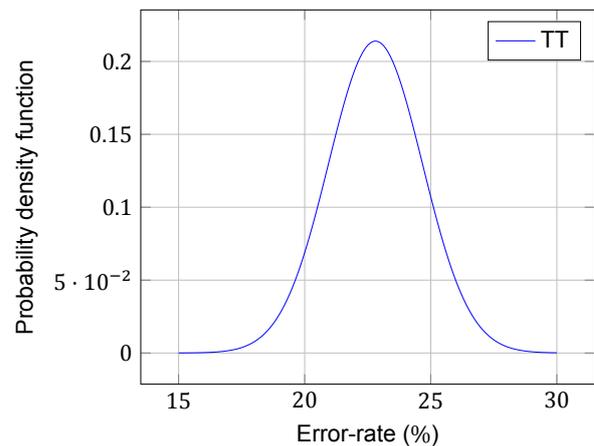


Figure 4.4: Error-rate distribution with the two-threshold technique disabled.

Table 4.1: Error rate performance in critical corners ($n = 25$).

Corner	Average error-rate	σ
TT	2.20 %	0.45 %
FF	2.35 %	0.48 %
SS	3.30 %	0.47 %

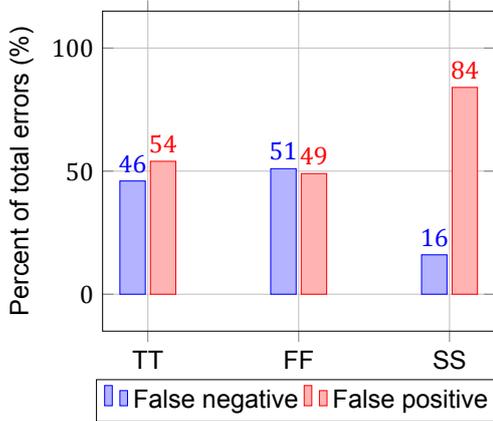


Figure 4.5: Breakdown of the errors into false positives and false negatives.

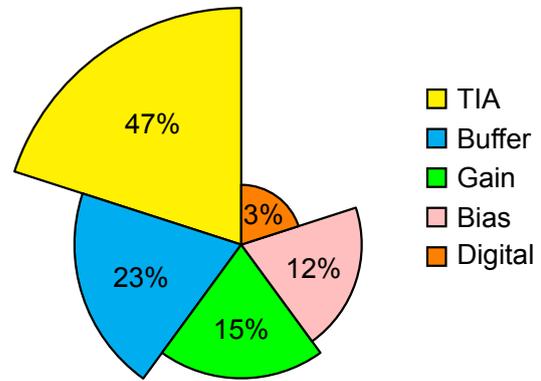


Figure 4.6: Power consumption breakdown.

The total power consumed by the readout is 2.85 mW (1.7 W for 600 pixels) and its component-by-component breakdown is charted in Figure 4.6. The TIA only accounts for 47 % of the total power consumed which attests to the point made during the feasibility study that excessive pixelization to reduce the detector's capacitance and improve the TIA's noise efficiency does not always bring the total power consumption down. The readout occupies an area of 100 μm by 80 μm , which is in-large the area of the TIA and the metal-oxide-metal (MOM) capacitors. Table 4.2 summarizes the achieved performance and compares it to the work done by Kleczek [6] and Ciaobanu [10].

Table 4.2: Readout performance summary and comparison with relevant works.

	This work	[6]	[10]
Technology	65 nm CMOS	40 nm CMOS	180 nm CMOS
Generated charge (fC)	0.16	0.35	1.00
Input capacitance (fF)	300 [†]	50 to 150	—
Event-rate (MEvent/s)	400	52 [‡]	585 [‡]
Error rate (%)	2.2	—	—
Power per channel (mW)	2.85	0.1	17
Area (μm^2)	100×80	—	—

[†] Including the ESD circuit.

[‡] Calculated as Event-rate = 1.42× achieved preamplifier bandwidth.

5

Design for Test (DfT)

This chapter details the integrated circuit (IC) designed to verify the performance of the proposed architecture. The IC includes two readout circuits and a plethora of additional blocks that aid in the testing of the chip. Section 5.1 briefly discusses the functionalities added, Section 5.2 summarizes the chip's built-in programmability, and Section 5.3 describes the measurement setup.

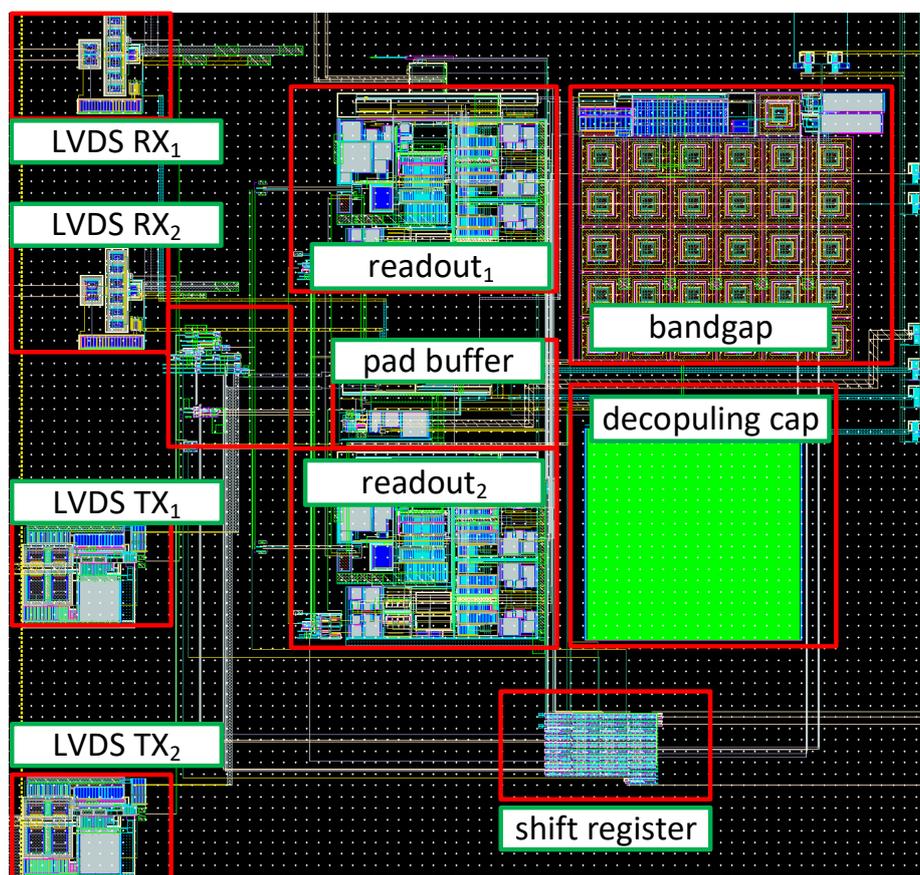


Figure 5.1: The layout of the test IC.

5.1. DfT Blocks

Functional blocks must be added to enable the monitoring, debugging, and troubleshooting of the IC. Moreover, the pixilated PIN diode is not available for testing and hence a circuit implementation is required to emulate the sensor's characteristics by generating small and fast current pulses.

Current-DAC: The steering current-DAC shown in Figure 5.2 is placed at the input of the readout to emulate the PIN diode. Two switches steer the current between the input of the TIA and an arbitrary node where the current is dumped. The switches are controlled by a 400 Mbit/s data line provided by an FPGA and delivered to the chip through high-speed low-voltage differential signaling (LVDS). Using a programming bit, the current of DAC can be measured by connecting the dump-node to an output pad. Additionally, the biasing current I_b is implemented using an external current source, allowing DAC's current to be calibrated. The current pulses produced by the DAC have a pulse-width of 2.5 ns (unlike the PIN diode's pulse-width of 1.8 ns) following the bit period of the control signal. Consequently, the amplitude of the pulses must be reduced to 65 nA to keep the equivalent charge generated to 160 aC. Figure 5.3 plots the response of the readout to pulses generated by an ideal current source (used in the previous chapters) and pulses generated by the DAC. The figure shows that the DAC can accurately emulate the input signal.

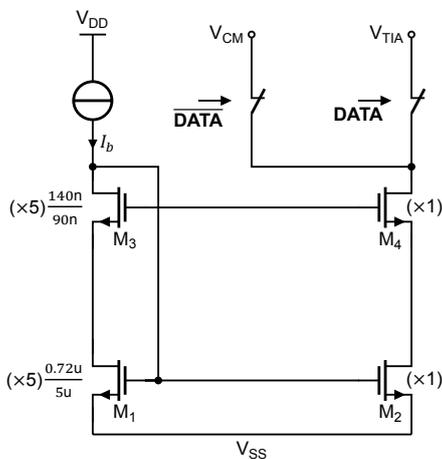


Figure 5.2: Schematic of the current-DAC.

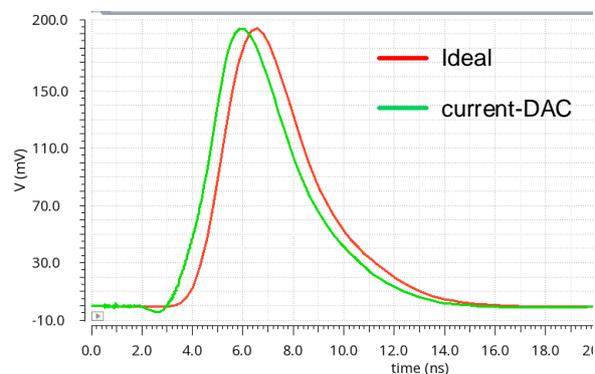


Figure 5.3: The pulse-response of the channel for an ideal current source and the current-DAC—post-layout simulation.

Unfortunately, the DAC is not capable of generating randomly-spaced pulses which have the consequence of making the error-rate dependent on the sampling moment as shown in Figure 5.1. This is because if a pulse is sampled when its voltage level is near the threshold, the probability of its detection error is higher (see Figure 2.10b). And since the DAC generates pulses that are equally-spaced by the sampling period, all subsequent pulses will also be sampled at this sensitive spot, and therefore, the error-rate increases rapidly. To account for this effect, a trimmable delay is applied to the DAC control signal to optimize the pulses' arrival time with respect to the sampling moment. Using programmable delays, the DAC can also be programmed to one of three test-modes (Figure 5.5):

- Single-pulse: A single-pulse arrives with a variable time-delay—implementing the single-pulse test.
- Two-pulses: Two-pulses arrive with variable inter-pulse time-spacing—implementing the two-pulses test.
- Large sequence of pulse: Pulses arrive with fixed inter-pulse time-spacing—used to evaluate the readout's error-rate.

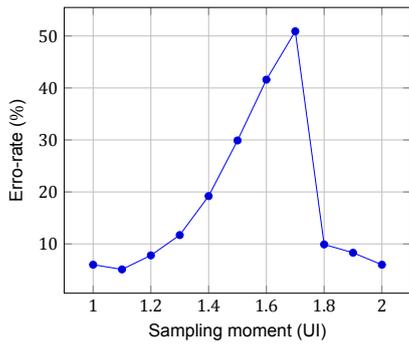


Figure 5.4: Effects of sampling moment on the error-rate for input signals generated by the DAC. The UI in this work is 2.5 ns.

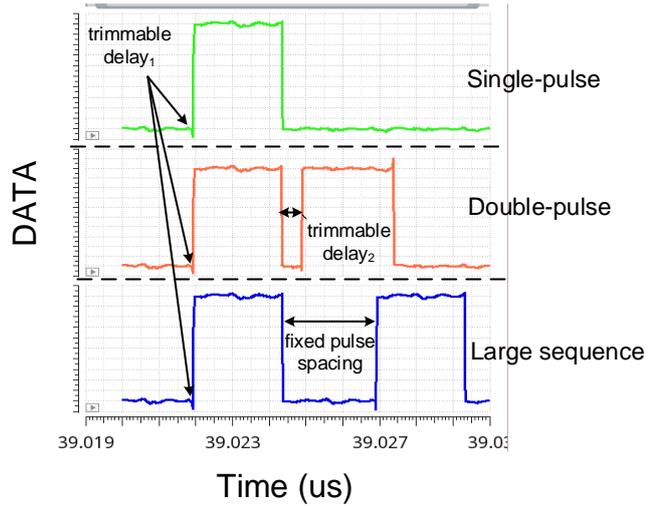


Figure 5.5: The operational modes of the DAC—post-layout simulation.

Bandgap Reference: The threshold values are generated on-chip by a bandgap reference, the schematic of which is drawn in Figure 5.6. The amplifier implemented by M1–M4 forces its inputs to the voltage V_{BE1} and consequently a ΔV_{BE} is established across the resistor R_3 —generating a proportional-to-absolute-temperature (PTAT) current [14]. On the other hand, the resistor R_2 draws a current proportional to V_{BE1} which has a negative temperature coefficient [14]. The total current flowing through M6 is then:

$$I_{M6} = \frac{1}{R_2} (V_{BE1} + \frac{R_2}{R_1} \times V_T \ln 30) \tag{5.1}$$

Which has a positive and negative temperature coefficient and is theoretically temperature insensitive. This current is copied by M7 to generate a V_{TH} value of $I_{M6} \times R_4$. To guarantee accurate threshold values across PVT variations, the resistor R_4 is made 3-bit trimmable. Moreover, both nodes V_{TH1} and V_{TH2} can be measured and calibrated or completely bypassed by an external voltage through two analog input/output (I/O) pads. The layout of the bandgap reference and its performance across corners and temperatures is shown in Figure 5.7 and Figure 5.8 respectively.

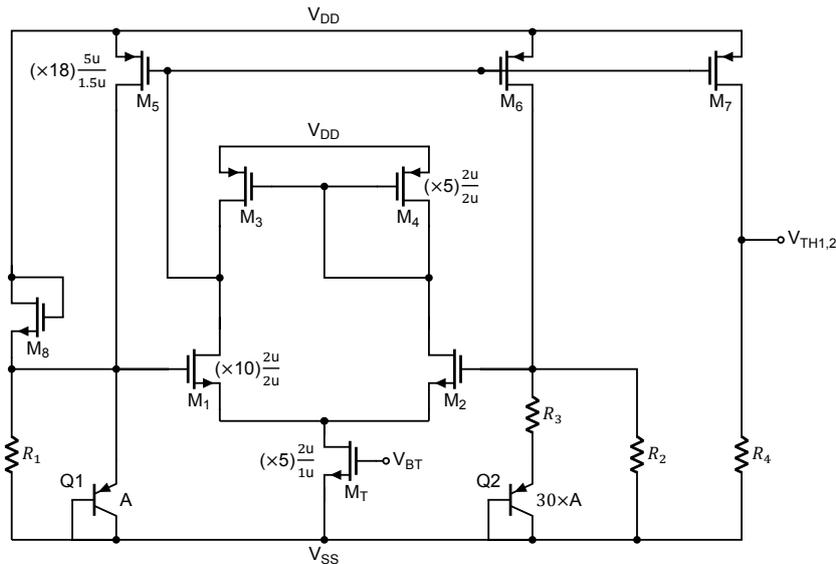


Figure 5.6: Schematic of the bandgap reference.

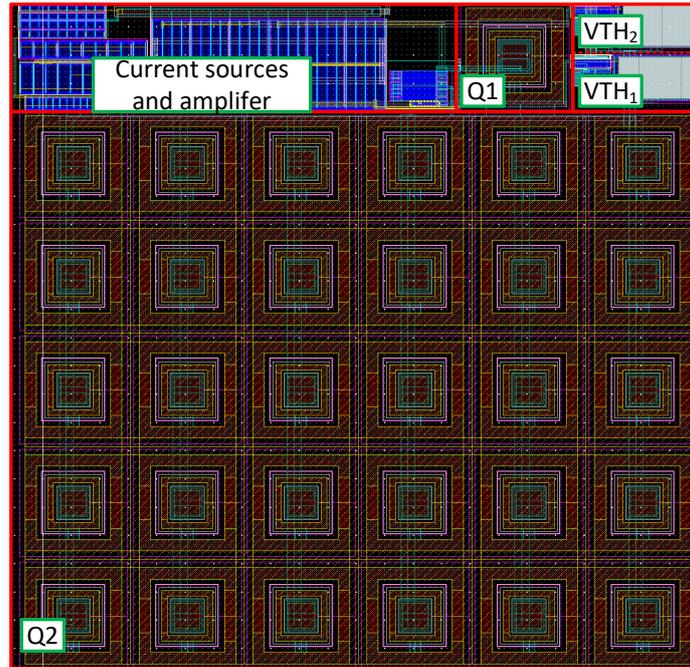


Figure 5.7: Layout of the bandgap reference.

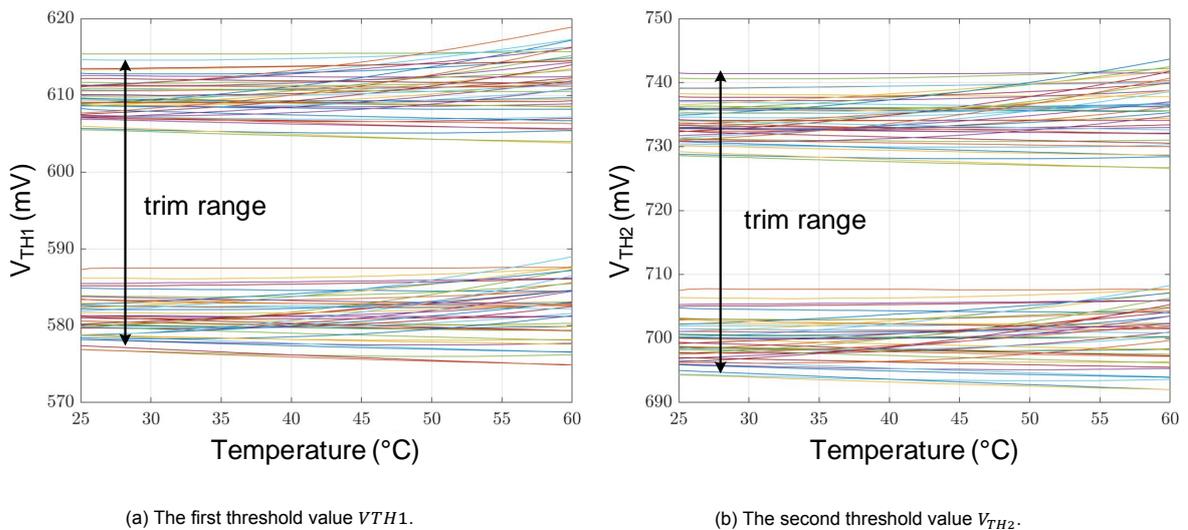


Figure 5.8: Performance of the bandgap across corners and temperature.

Pad Buffer: A wideband high-power buffer facilitates the monitoring of the readout's analog voltage (i.e. it probes the voltage at the input of the comparators). The pad buffer is designed to drive a 10 pF load (I/O pad and PCB parasitics) and provide 400 MHz of bandwidth to accurately represent the pulse-response of the readout. It has the same topology of the buffers designed for the readout (See Figure 3.17), however, the following changes were made:

- The input pairs are made shorter and narrower such that the buffer does not capacitively load the readout.
- The first stage's current-sources are shorter and narrower to move the mirror pole to higher frequencies.

- The second stage is sized up and draws 1 mA of current to drive its large load.

The buffer's input terminal can be programmed to probe one of the two readouts or an internal node fixed to $V_{DD}/2$ —allowing the measurement and calibration of the channel's offset. The layout of the pad buffer is shown in Figure 5.9. Table 5.1 summarizes its achieved performance.

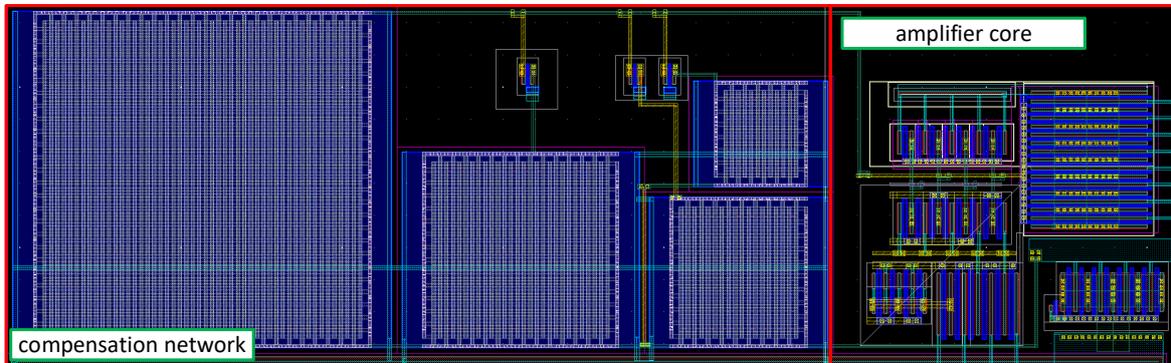


Figure 5.9: Layout of the pad buffer.

Table 5.1: Summary of pad buffer's performance—post-layout simulations.

Specification	Min.	Max.
Bandwidth	380 MHz	535 GHz
Phase margin	41°	52°
Power consumption	1.2 mW	1.3 mW
Load capacitance	10 pF	

Miscellaneous: The other blocks implemented to facilitate testing are:

- High-speed LVDS transceivers: receive high-speed input data and clock from an FPGA and transmit the readout's output data and clock to the FPGA.
- Shift register: holds bits that can be written and read by an external FPGA to trim, calibrate, monitor, and debug the chip.
- A replica of R_F : is connected to an IO pad to measure and calibrate the value of the TIA's feedback resistor.

5.2. Programmability

The IC is programmable through the shift register to select testing modes, calibrate components and nodes, and troubleshoot potential issues. Below is a summary of the chip's programmability:

- **Select a readout:** enables one of the two readouts. The enabled readout receives input data (i.e. DAC control signal) and its output data and output data clock are connected to the LVDS transmitters. The disabled readout is turned off.
- **Enable both:** enables both readouts. Both readouts receive input data, however, only the selected readout's output (analog and digital) can be monitored.
- **Set DAC mode:** programs the DAC to perform the single-pulse and two-pulses tests or to inject large sequences of data to evaluate the error-rate.
- **Cross-talk test:** injects current-pulses onto the selected readout while monitoring the non-selected, but enabled, readout.

- **Program pad buffer:** connects the pad buffer's input to an internal-node set to $V_{DD}/2$ or to the analog output of the selected readout.
- **Measure and calibrate DAC:** connect the DAC to an IO pad for measurement and calibration of the DAC's current.
- **Enable two-thresholds:** enables/disables the two-threshold technique.
- **Troubleshoot LVDS:** connects the LVDS receivers to the transmitters to verify the data injected into the readouts.
- **Trim:** bits are programmed to trim passive components (e.g. R_F , C_F , etc.), the threshold values, and the delay imposed on the input data.

5.3. Measurement Setup

The block diagram of the proposed measurement setup is illustrated in Figure 5.10. An FPGA sends and receives 400 Mbit/s data and clock signals to and from the IC—controlling the input DAC and reading the output of the selected readout. The FPGA also reads and writes bits into the shift register to program the operational mode of the IC. A Multimeter is used to measure the value of the feedback resistance R_F , the DAC's current, and the threshold values. Current and voltage sources bias the IC core components, calibrate the TIA's gain and the DAC's current and eliminate the channel's offset. Finally, an oscilloscope facilitates monitoring the analog output of the readout.

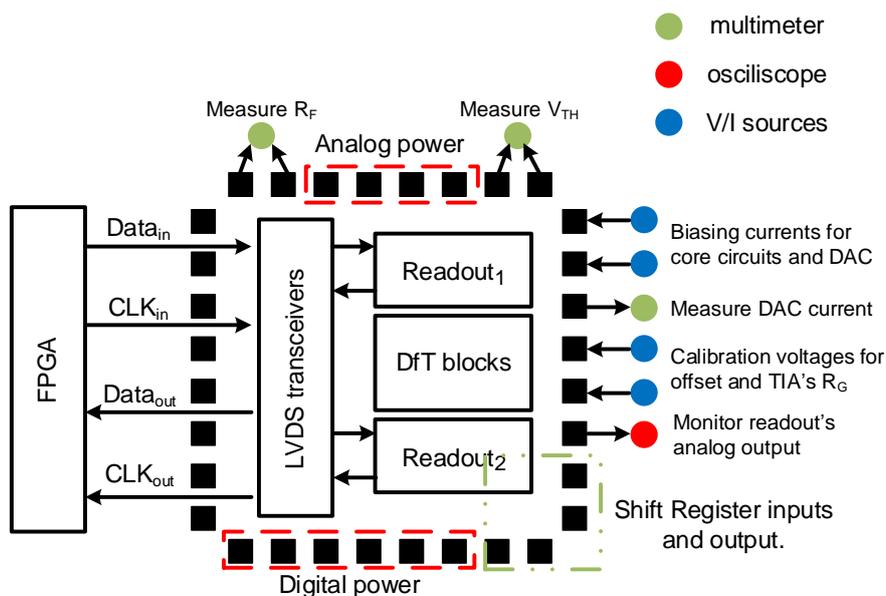


Figure 5.10: The IC's measurement setup.

After the IC parameters have been calibrated to achieve the desired bandwidth, pulse response, and threshold values, the following measurements are performed to evaluate its performance:

1. Monitor the pulse-response and measure the channel's discharge time-constant.
2. Measure the probability of error by performing the single-pulse and two-pulses tests.
3. Measure the error-rate by using a large-sequence of inputs.
4. Measure the effectiveness of the two-threshold technique by performing (2) and (3) with the technique disabled.

5. Measure sensitivity to cross-talk by performing (2) and (3) with the non-selected readout is enabled.
6. Measure the readout robustness by performing (2) and (3) using different threshold values, adding offsets, and varying the values of R_F , C_F , and C_L .
7. Measure the power consumption of the readouts.
8. Monitor cross-talk using the cross-talk mode.

6

Discussion, Future Work, and Conclusions

6.1. Discussion

On paper, the two-threshold comparator appears to be an effective (and perhaps a required) technique to allow for the fast detecting of small charge portions. But what's the catch? After all, designs seldom have advantages without drawbacks.

On randomly arriving particles: We first must discuss how a detection error was defined (and consequently how the error-rate was calculated) throughout this work. In digital systems where the typical readout structure is used, signals arrive in a binary format at fixed times and are sampled at fixed times as illustrated in Figure 6.1. For example, a '1' bit arrives at the 1st unit interval (UI) and is sampled at center of that UI (when the output peaks). Errors are then easily defined: at the detector's output, a bit must retain its state ('0' or '1') and the UI (time slot) it arrives in to be considered correct.

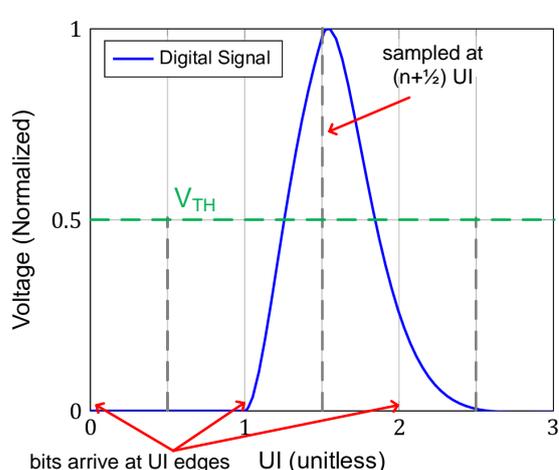


Figure 6.1: A digital system. Bits are assigned to the UI edge they arrive at.

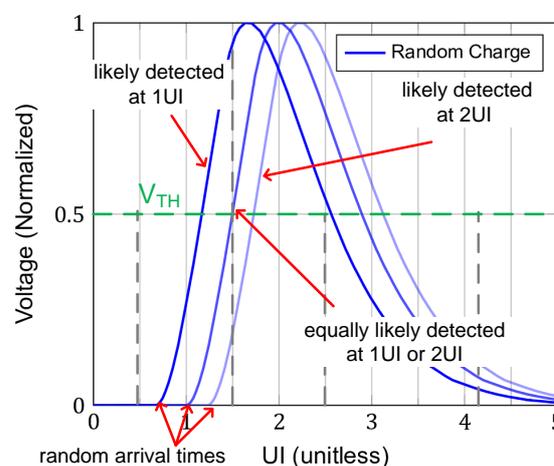
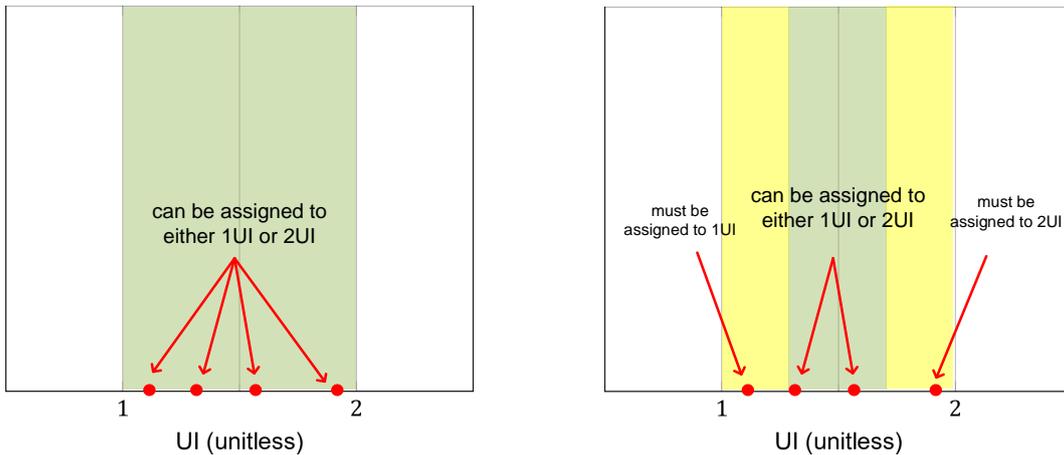


Figure 6.2: A narrowband particle detector. Charge arrive anywhere inducing uncertainty with respect to the UI they are assigned to.

Unfortunately, the scenario is ambiguous considering electrons randomly hitting the surface of a detector. Charge can be generated anywhere between two UIs, and subsequently, there is no longer a definitive time slot the incoming particle can be assigned to (Figure 6.2). To account for this randomness we redefine a 'correct' output as follows: charge arriving anywhere between the (n_{th}) UI and the ($n_{th} + 1$)

UI can be assigned to either. For example, a charge hitting the detector at 1.33UI can be assigned to 1UI or 2UI (Figure 6.3a) while detecting the charge in both UIs or neither is counted as an error. This definition of correct and erroneous detections might appear arbitrary, and it is. We could define stricter criteria: a particle must be assigned to the UI it arrives closer to, and can only be assigned to either UI when it arrives in the middle between the two (Figure 6.3b). Using the data in Chapter 4, the stricter definition results in an average error-rate of 10% due to a combination of random arrival and the two-threshold technique as we will see in the next section. Nevertheless, the best way to define correct and erroneous detections must be decided based on the application employing the readout. Unfortunately at the time of writing this thesis, the effects of this timing uncertainty on the imaging system are unknown.



(a) Definition of correct detection time slot used in This Work.

(b) An example of a strict definition of correct detection time slot.

Figure 6.3: Definition of the correct detection time slot for randomly arriving charge portions. Red dots represent charge arrival times.

On the propagation of uncertainty: The two-threshold technique uses past decisions to determine future ones, and consequently, any uncertainty (or error) made in past decisions propagates down to the future decisions. The two-pulses test performed in Chapter 2, where a pulse arrives at different spots in a UI followed by a second pulse arriving at a fixed time (Figure 6.4), demonstrates this phenomenon well. Figure 6.5 shows the probability of detecting the first pulse at the 1st UI and the second pulse at the 3rd UI. Logically, as the arrival time of the first pulse moves from 1UI to 2UI its less likely to be assigned to the 1st UI, reaching 50% approximately in the middle. The second pulse, although always arriving at the edge of the 3rd UI, has some chance of being detected at the 4th UI due to the propagation of uncertainty from the past decision. If the application at hand does not tolerate such timing uncertainty, the aforementioned propagation becomes a prohibiting drawback of the two-threshold technique.

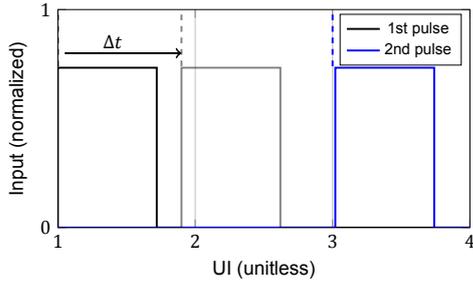


Figure 6.4: The two-pulses test: the arrival time of the first pulse is swept while a following pulse arrives at a fixed time.

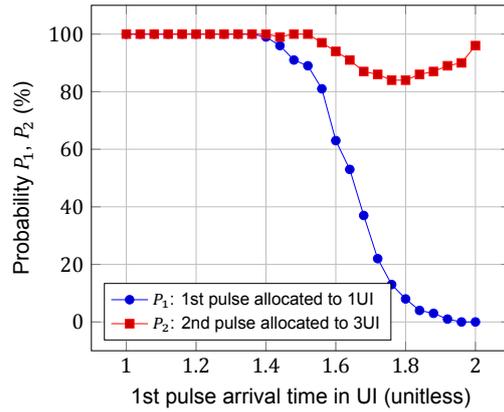


Figure 6.5: Probability of allocating the first pulse to 1UI and the second pulse to 3UI.

On controlling the channel response: In this work, we decided to use the two-threshold technique to cancel ISI then proceeded to optimize the channel response (i.e. bandwidth and quality factor) for minimum error-rate. Bandwidth is well controlled through a trimmable feedback resistor and capacitor, the use of which is typical for precision TIAs in the literature. However, controlling the quality-factor remains challenging. In this design, two additional trimmable components are implemented to perform this task: (1) a resistor adjusts the TIA OTA's non-dominant pole, and (2) a load capacitance adjusts TIA OTA's dominant pole. Unfortunately, many more parameters affect Q such as temperature, bond wire inductance, and supply voltage variations making the design less reliable and becoming one of its primary drawbacks. But why does the quality-factor affect the error-rate so severely? (see Figure 2.4.) A glance at decision feedback equalization can offer some insight.

6.1.1. Decision Feedback Equalization

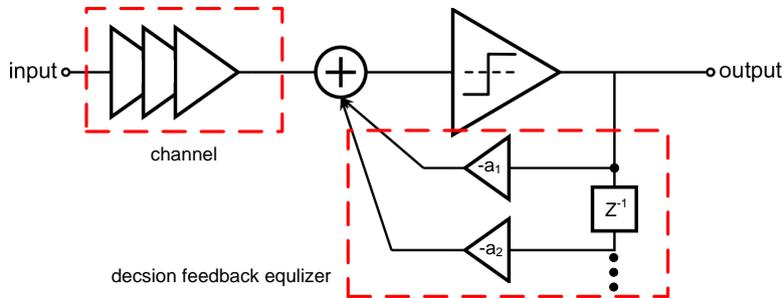


Figure 6.6: Block diagram of a decision feedback equalizer.

Equalization is the process of correcting imperfections in a channel's transfer by implementing the inverse transfer in the signal path. Decision feedback equalizers (DFEs¹) are used in gigabit optic fiber receivers to reconstruct the ISI from past decisions and subtract it from future decisions (Figure 6.6). An example of a channel response requiring DFE is illustrated in Figure 6.7. The comparator samples a pulse at time x_0 , and due to the channel's inadequate bandwidth some non-zero values of this pulse leak into the subsequent sampling moments x_1 , x_2 , and x_3 , and are commonly referred to as post-cursors. Since DFEs are used in digital systems, the input arrives and is sampled at fixed times, and thus, the values of the post-cursors are predictable and can be eliminated. To remove any 'pile-up-or-down' effects arising from the post-cursors, a 3-tap DFE with feedback coefficients a_1 , a_2 , and a_3 equal to their perspective post-cursor values can be implemented to subtract the pulse's residual voltages. Optical receivers employing DFE do not rely on a fixed channel response, but instead, the feedback

¹We came across DFEs at the ending phase of this project. This small section attempts to place the work done within the framework of feedback equalization and gives recommendations for future work based on insights from DFEs.

coefficients are calibrated on-chip using a pseudo-random test-signal to minimize the bit error-rate.

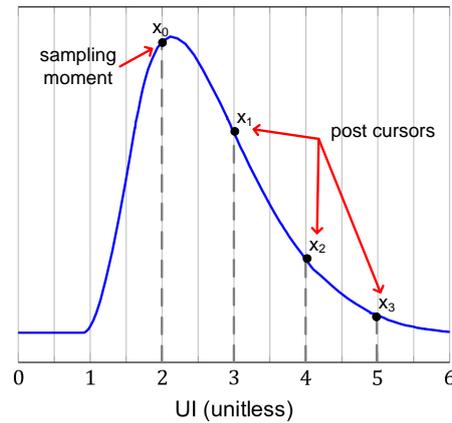


Figure 6.7: Single-bit response of a narrow band digital system.

The reader may have noticed that the two-threshold comparator is nothing but a 1-tap DFE.² Instead of automatically calibrating the threshold values, we optimized the channel to a 1-tap DFE which explains the error-rate's sensitivity to the quality-factor. A channel bandwidth of 70 MHz and Q of 0.5 has a pulse response with approximately a single non-zero post-cursor (Figure 6.8a) because the 2nd post-cursor x_2 approaches zero—and hence it can be optimally equalized by the 1-tap DFE. However as Q increases, the pulse response widens and has an undershoot as seen in Figure 6.8b. The first post-cursor moves to higher values and a fourth and fifth post-cursors are introduced, inducing additional ISI and making the 1-tap DFE less effective.

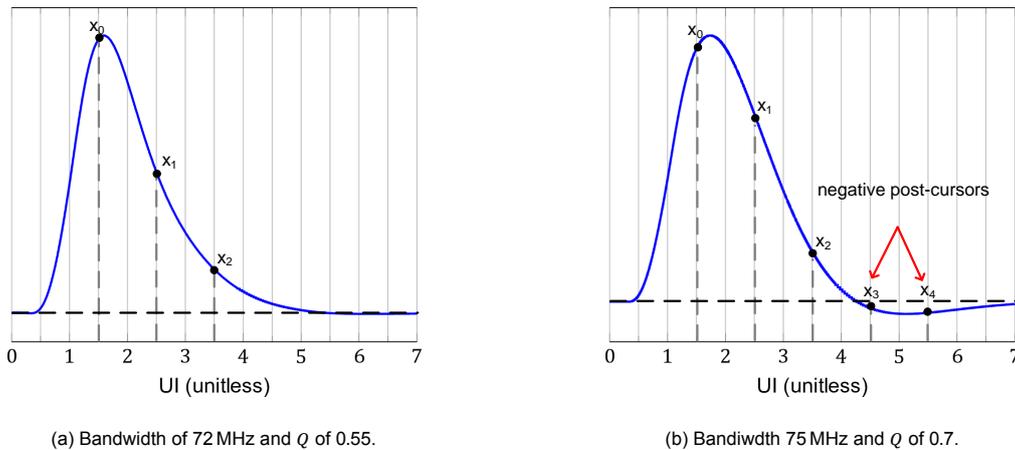


Figure 6.8: Pulse response of the implemented readout (a) and a readout with a higher Q .

Is it then possible to introduce additional thresholds (DFE taps) and calibrate them on silicon to make the electron readout less sensitive to the channel response or to further reduce the bandwidth of the channel? The short answer is that we do not know and future work must examine this question. Nevertheless, there are two key differences between DFE in digital systems and DFE employed in particle detectors that can limit the scalability of the technique:

1. **Random arrival of particles:** Pulses do not arrive and are not sampled at fixed times, and consequently, the post-cursor values are less predictable. Post cursors are now a range of values depending on the time the pulse is detected as illustrated in Figure 6.9.

²Increasing the threshold value from 600 mV to 720 mV implements a_1 of 120 mV.

2. **Automatic calibration requires digital signals:** Calibration algorithms use XOR gates to calculate the error between a known test signal and the system's output. This does not work for randomly arriving charges due to the aforementioned timing uncertainty.

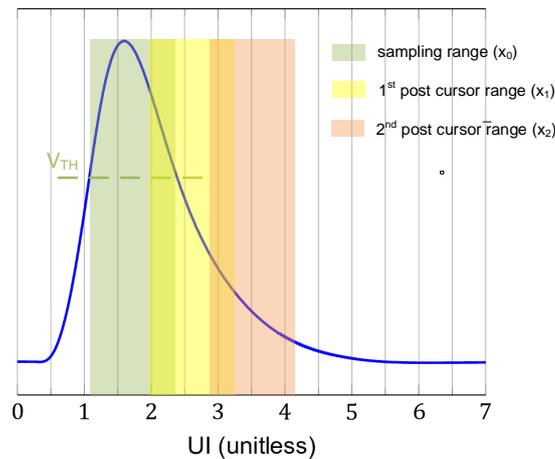


Figure 6.9: Random particle arrival causing post-cursors to be less-well defined.

6.2. Future Work

Recommendations for future work are to:

- **Extend:** How much slower can the readout be? How many thresholds (DFE taps) can be implemented while maintaining the technique's effectiveness? Exploring the limit of ISI cancellation using DFE or improving upon DFE to make it more suitable for randomly arriving charges.
- **Harden:** Improve the system reliability by having built-in robustness to process, voltage, and temperature variations such as automatically calibrating thresholds (DFE tap coefficients), automatic gain control, and offset cancellation schemes.
- **Optimize:** Functions such as gain, buffering, comparison, and decision feedback can be implemented in many ways in CMOS. Optimize the CMOS realization to reduce overhead³ power consumption or to increase the system's robustness.

6.3. Conclusions

State-of-the-art imaging systems require fast and accurate particle detectors that can deal with small charge portions and consume limited power. The conflicting requirements of a small input signal energies, large bandwidth, and low-power create an implementation dead-end considering typical methods of particle detection as a sufficient wideband SNR cannot be achieved. The two-threshold architecture is a simple addition with a strong effect. It permits the analog components of the particle readout to be designed with a lower bandwidth than required—improving their noise performance—while reducing the deterministic ISI-induced errors associated with narrowband circuit and hence maintain a low error-rate. The readout can detect 160 aC charge portions arriving randomly at 400 MEvent/s with error-rates below 5 % while consuming 2.85 mW and occupying 8000 μm^2 . Future work should focus on extending the technique to allow for further bandwidth reduction and introduce measures to guarantee its robustness across process variations.

³overhead refers to non-TIA components.

Bibliography

- [1] G. Lutz, *Semiconductor Radiation Detectors*. Springer Berlin Heidelberg. [Online]. Available: <http://link.springer.com/10.1007/978-3-540-71679-2>
- [2] H. Spieler, *Semiconductor Detector Systems*, ser. Series on Semiconductor Science and Technology. Oxford University Press, no. 12.
- [3] J. I. Goldstein, D. E. Newbury, J. R. Michael, N. W. Ritchie, J. H. J. Scott, and D. C. Joy, *Scanning Electron Microscopy and X-Ray Microanalysis*. Springer New York. [Online]. Available: <http://link.springer.com/10.1007/978-1-4939-6676-9>
- [4] Y. Wang, Z. Dong, R.-L. Lai, and K. Kanai, "SEMICONDUCTOR CHARGED PARTICLE DETECTOR FOR MICROSCOPY," patent 20 190 378 682, Decmeber 12 2019.
- [5] A. Sakic, G. van Veen, K. Kooijman, P. Vogelsang, T. L. M. Scholtes, W. B. de Boer, J. Derakhshandeh, W. H. A. Wien, S. Milosavljevic, and L. K. Nanver, "High-Efficiency Silicon Photodiode Detector for Sub-keV Electron Microscopy," vol. 59, no. 10, pp. 2707–2714.
- [6] R. Kleczek, P. Grybos, R. Szczygiel, and P. Maj, "Single Photon-Counting Pixel Readout Chip Operating Up to 1.2 Gcps/mm² for Digital X-Ray Imaging Systems," vol. 53, no. 9, pp. 2651–2662. [Online]. Available: <https://ieeexplore.ieee.org/document/8421204/>
- [7] F. Krummenacher, "Pixel detectors with local intelligence: An IC designer point of view," vol. 305, no. 3, pp. 527–532. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/016890029190152G>
- [8] R. Szczygiel, *Krummenacher feedback analysis for high-count-rate semiconductor pixel detector readout*. Proceedings of the 17th International Conference Mixed Design of Integrated Circuits and Systems. [Online]. Available: <http://ieeexplore.ieee.org/servlet/opac?punumber=5543946>
- [9] E. Säckinger, *Analysis and Design of Transimpedance Amplifiers for Optical Receivers*. Wiley.
- [10] M. Ciobanu, N. Herrmann, K. D. Hildenbrand, M. Kis, A. Schuttauf, H. Flemming, H. Deppe, S. Lochner, J. Fruhauf, I. Deppner, P. A. Loizeau, and M. Trager, "PADI, an Ultrafast Preamplifier - Discriminator ASIC for Time-of-Flight Measurements," vol. 61, no. 2, pp. 1015–1023. [Online]. Available: <http://ieeexplore.ieee.org/document/6786378/>
- [11] F. Anghinolfi, P. Jarron, F. Krummenacher, E. Usenko, and M. Williams, "NINO, an ultra-fast, low-power, front-end amplifier discriminator for the Time-Of-Flight detector in ALICE experiment," in *2003 IEEE Nuclear Science Symposium. Conference Record (IEEE Cat. No.03CH37515)*. IEEE, pp. 375–379 Vol.1. [Online]. Available: <http://ieeexplore.ieee.org/document/1352067/>
- [12] R. N. McDonough, A. D. Whalen, and A. D. Whalen, *Detection of Signals in Noise*, 2nd ed. Academic Press.
- [13] M. Pelgrom, *Analog-to-Digital Conversion*. Springer International Publishing. [Online]. Available: <http://link.springer.com/10.1007/978-3-319-44971-5>
- [14] B. Razavi, *Design of Analog CMOS Integrated Circuits*, second edition ed. McGraw-Hill Education.