A Monolithic Radiation-Hard Testbed for Timing Characterization of Charge-Sensitive Particle Detector Front-Ends in 28 nm CMOS

Thesis

Presented in Partial Fulfillment of the Requirements for the Degree Master of Science in the Graduate School of The Ohio State University

By

Kennedy B. Caisley, B.S.E.E.

Graduate Program in Electrical and Computer Engineering

The Ohio State University

2021

Thesis Committee:

Dr. Wladimiro Villarroel, Advisor

Dr. Maurice Garcia-Sciveres

Dr. Ayman Fayed

Dr. Tawfiq Musah

Abstract

Next-generation hybrid pixel detectors aim to achieve timing resolutions on the order of 100 ps. Of primary concern is the analog front-end, composed of preamplifier and discriminator, which introduce significant timing uncertainty to the sensor charge signal they transduce. This work presents an on-chip test circuit capable of characterizing the jitter of pixel detector analog front-ends constructed in 28 nm bulk CMOS. The test system injects an artificial sensor charge pulse at the input of the device-under-test and then measures the output timing variation with a time-to-digital converter. The measurement circuit can inject charge quantities up to 24,000 electrons, with a timing precision of 10.1 ps RMS, a maximum differential non-linearity of 0.25 LSB, and a dynamic range of 64 ns.

Vita

May 2015High School Diploma, Moscow High SchoolAugust 2015 - May 2019B.S. Electrical Engineering, University of IdahoAugust 2019 - April 2021Graduate Researcher, Ohio State UniversityJune 2021 - December 2021ASIC Design Intern, Lawrence Berkeley National LaboratoryMajor Field: Electrical Engineering

Contents

A	bstra	let	i
V	ita		ii
\mathbf{Li}	st of	Figures	iv
\mathbf{Li}	st of	Tables	\mathbf{vi}
1	Intr	roduction	1
2	Bac	kground	2
	2.1	Pixel Detectors	2
	2.2	Time-of-Arrival Measurement	4
	2.3	28 nm AFE Timing Performance	7
	2.4	Proposed Testbed for 28 nm AFE Characterization	10
3	Cha	arge Injection Subcircuit	12
	3.1	Principle of Charge Injection	12
	3.2	Circuit Requirements	13
	3.3	Fine Resolution Tunable Charge Injection Circuit	14
	3.4	Coarse Resolution Charge Injection DAC	18
	3.5	Precision Timing and Charge Quantity	20
	3.6	Injection Circuit Top Level and Layout	21
4	Tin	ne Digitization Subcircuit	24
	4.1	Time-to-Digital Converter Requirements	24
	4.2	Unit Delay Cell Design	26
	4.3	Delay Line Gain Error and Calibration	28

	4.4	Top-Level TDC Architecture	29
	4.5	Top-level TDC Layout	32
	4.6	Overall TDC Performance	33
5	Syst	tem Integration and Testing	36
	5.1	System Level Integration	36
	5.2	System Testing	37
	5.3	Conclusion	38

List of Figures

2.1	3D assembly view of the ATLAS detector, with humans for scale	3
2.2	Cross section of a single hybrid pixel, with sensor, bump-bond, and CMOS electronics	3
2.3	Signal chain of a traditional hybrid pixel detector.	4
2.4	(a) Generic element with mean propagation delay t_{pd} . (b) Timing diagram of many trials with	
	varying Δt_{pd} . (c) Gaussian distribution of propagation delays with deviation σ_{pd}	6
2.5	Basic phenomenon of "time walk", where crossing time "t" depends on pulse amplitude	6
2.6	Time-walk of the AFE under test, vs input injection charge	8
2.7	TOT of the AFE under test, vs input injection charge	9
2.8	Jitter in measurement electronics introduces error to measurement of DUT jitter	10
2.9	Basic test setup, with only external test equipment.	10
2.10	Improved test setup, with integrated test fixtures	11
3.1	Basic principle of a charge injection circuit.	13
3.2	Unit circuit for charge injection, with tunable injection voltage	16
3.3	Fine resolution current transients vs time, at multiple injection voltages	17
3.4	Fine resolution slice charge injection vs injection voltage supply.	17
3.5	Binary weighted array of 15 injection circuits (CDAC), providing coarse resolution	18
3.6	Coarse resolution current transients vs time, at multiple CDAC codes	19
3.7	Fully range of charge injection vs CDAC code.	20
3.8	Charge injection precision and jitter (200 runs) $\ldots \ldots \ldots$	21
3.9	Top level schematic of the charge injection circuit	22
3.10	Top level layout of the charge injection circuit	23
4.1	Unit delay element with power supply cutoffs, cross-coupled inverters for de-skew/evaluation,	
	and output buffer.	28
4.2	Diagram of the dual stop edge generation circuit.	30

4.3	Top block-level diagram of the dual TDC architecture	30
4.4	Circuit-level diagram of one-half of the TDC, featuring the fine and coarse resolution circuits,	
	and PISO	32
4.5	Layout featuring edge selection circuit driving dual TDCs with coarse and fine resolution	
	sub-circuits	33
4.6	DNL vs input time of aggregate TDC (worst case of 50 runs via Monte-Carlo)	33
4.7	INL vs input time of aggregate TDC (worst case of 50 runs via Monte-Carlo) $\ldots \ldots \ldots$.	34
4.8	Single shot precision and mean error	35
4.9	TDC single-shot precision at end of range	35
5.1	Floorplan of the BigRock testbed ASIC.	37

List of Tables

3.1	Charge injection circuit design specifications	15
3.2	Expected performance of charge injection circuit vs design specifications	23
4.1	Time-to-digital converter circuit design specifications	26
4.2	Simulated time-to-digital converter circuit performance, versus design specifications	35

Chapter 1: Introduction

The objective of this project is to design a monolithic testbed for automated characterization of nextgeneration charge sensitive analog front-ends (AFEs) in 28nm CMOS. The test application specific integrated circuit (ASIC), termed 'BigRock' is capable of measuring the timing characteristics of AFEs, including jitter, delay, time-walk, and time-over-threshold at a wide range of charge injection levels. The core focus of this research was the construction of an integrated configurable charge injection circuit and time-to-digital converter, along with supporting digital controls and readout. The aggregate system has a timing precision below 20 ps RMS, and has been designed to operate across a wide range of temperatures, and in high-radiation environments up to a lifetime total ionizing dosage of 1000 Mrad. Up to 30 AFEs can be tested per test-bed chip.

Chapter 2 details the relevant background information on the applications and construction hybrid pixel detectors. The characteristics of capacitive pixel sensors and their corresponding AFEs are described, and a proposed architecture for the BigRock test bed is introduced. Chapter 3 and 4 detail the design of the charge injection circuit and time digitization sub-circuits, respectively. Chapter 5 concludes with future plans for top-level integration, layout, and post-fabrication testing for the BigRock ASIC.

Chapter 2: Background

2.1 Pixel Detectors

Particle detectors are a type of instrumentation used in high-energy physics (HEP) to sense massive relativistic particles and photons at accelerator interaction points. The data collected by these detectors helps experimental physics research programs better understand the particles that constitute matter and mediate forces. ATLAS and CMS at the Large Hadron Colliders (LHC) are two high-profile examples of such detector-based experiments. To understand the construction of a large scale detector system, it helpful to examine a specific example. A 3D cross section of the ATLAS detector is shown in Figure 2.1 for illustration. Like other detectors in its class, ATLAS is composed of multiple specialized detector subsystems wrapped concentrically in layers around the collision point to record the trajectory, momentum, and energy of particles, allowing them to be individually identified and measured. Of primary interest to this work is the inner-most layer of the detector, referred to as a pixel detector. Pixel detectors are a variety of detector composed of vast two-dimensional arrays of micrometer-scale solid-state sensors with coupled readout electronics. As a group, pixel detectors are the standard for measuring the energy and trajectory of particle tracks with high spatial resolution. For a sense of magnitude: in the current ATLAS detector, the pixels have a 50 µm pitch with a channel count on the order of 90 million (at the time of this document's writing), but examples of detectors with lower channel counts also exist [1].

Modern pixel detectors can be broadly categorized as either *monolithic* or *hybrid*, depending on system integration [1]. In hybrid architectures the sensor is fabricated separately and then coupled to a complimentary metal-oxide-semiconductor (CMOS) read-out integrated circuit (ROIC) via bump-bonding. This approach allows the ROIC design to leverage modern deep sub-micron technology while the sensor is built with maximum fill-factor in an optimized high-resistivity process [2]. A simplified cross section of one channel in a hybrid pixel detector is shown in Figure 2.2.



Figure 2.1: 3D assembly view of the ATLAS detector, with humans for scale.

The majority of ionizing particles detected by pixel sensors are pions, i.e. stable sub-atomic relativistic particles. High energy photons (x-ray and gamma radiation) are not generally detected, as they are not absorbed by the geometry of the capacitive pixel sensors. The capacitive sensor can be essentially considered as a photodiode. When an incident particle track passes through the solid-state sensor, mobile charge carriers are freed in the semiconductor lattice, which subsequently drift under the externally applied bias voltage, producing a fixed-rise time transient charge pulse with peak amplitude corresponding to particle energy. The charge pulse typically has a rise and fall time on the order of 100 ps, and the area of the pulse corresponds to the momentum of the original incident particle. Particles on average produce $80 e^-$ worth of carriers per micrometer thickness of the sensor, and sensor thickness is on the order of 250 µm, so typical charge pulses are in the range of 10 ke^- to 20 ke^- [3]. The noise floor is typically around $100 e^-$ RMS.



Figure 2.2: Cross section of a single hybrid pixel, with sensor, bump-bond, and CMOS electronics.

The primary goal of the subsequent readout electronics is measurement of the pulse amplitude, as this serves as a direct proxy for the energy of the original incident particle. The polarity of the bias is generally set such that electrons, the carriers with higher mobility, are swept across the solder bump bond and collected by the read-out electronics. The signal processing chain typical of a state-of-the-art hybrid pixel is pictured in Figure 2.3 [4].

The initial charge pulse generated in the capacitive sensor is fed into a charge-sensitive amplifier (CSA) which produces a corresponding amplified voltage signal at its output. The CSA often has variable gain to accommodate input charge pulses of different magnitudes. The most obvious method for measuring the signal pulse amplitude would be to use a peak-detect and hold (PDH) circuit, followed by an analog-to-digital converter. This approach, while precise, is infeasible in modern pixel detectors where per-channel power consumption must be kept under 10 μ A and pixel area is on the order of 50 μ m x 50 μ m [4].

A more tractable alternative, shown following the preamp in Fig. 2.3, leverages a discriminator followed by a counter to estimate the pulse amplitude. The discriminator continuously compares the input signal to a threshold reference, and generates a corresponding *time-over-threshold* (TOT) signal at its output. The duration of the TOT signal, typically on the order of tens to hundreds nanoseconds, is then typically digitized using a simple synchronous counter. Clock frequencies approaching 1 GHz are feasibly distributed chip-wide, yielding counter resolutions on the order of 1 ns. For every clock period that the TOT signal is high the counter is incremented. Multiplying the digital counter value by the period of the reference clock yields the time the pulse signal was above threshold, which can be used to approximate peak amplitude [4]. In this manner, the time-over-threshold width of the discriminator output is used as a proxy to infer the integrated area of the charge signal produced in the detector. The tunable discriminator threshold also serves to reject false positive hits from noise. In hybrid pixel detectors the threshold it is typically set to an input-referred charge quantity on the order of $1 \, \text{ke}^-$ to $1.5 \, \text{ke}^-$.



Figure 2.3: Signal chain of a traditional hybrid pixel detector.

2.2 Time-of-Arrival Measurement

To advance the frontier of HEP experimental capabilities, the design of next-generation pixel detectors aims to incorporate functionality for measuring the arrival time of incident particles to a sub-nanosecond precision. This concept has been popularly referred to as the 4D pixel, due to the addition of the temporal domain to the three position dimensions traditionally measured [5]. The time-of-arrival (TOA) signal is measured by extending the circuit in Fig. 2.3, such that it also computes the time difference between the rising edge of the TOT signal and some fixed time reference.

To avoid obscuring the desired TOA signal with noise, the *timing precision* of the signal chain must be reduced below 100 ps RMS [2]. Timing resolution is a specification of the uncertainty expected in a time measurement, and is a function of the various sources of error in the system. Non-idealities in a pixel sensor, including charge deposit fluctuations, noise, and fluctuations of the signal shape due to weighting field variations [6], introduce an initial degree of uncertainty σ_{sensor} in the TOA signal¹. In recent years, R&D efforts have improved sensor timing performance to levels below 30 ps [7,8].

The charge signal is next processed by the readout circuit, where the timing resolution is further degraded by electronic *jitter*, which is defined as the random variation in phase or propagation delay of circuit elements. Jitter is typically predominantly stochastic in nature, and is thus quantified by its deviation from a mean. To illustrate this, consider the case of a buffer circuit with a nominal propagation delay t_{pd} as seen in Fig 2.4(a). Due to random non-idealities, however, such as electronic thermal noise, shot noise, and power supply noise, the actual measured delay differs from trial to trial by Δt_{pd} , shown in Fig. 2.4(b). With a sufficient number of samples n, we can express the jitter of the element, σ_{pd} , as a root mean square (RMS) variation of the samples [9].

$$\sigma_{pd} = \sqrt{\frac{\sum (\Delta t_{pd})^2}{n}} \tag{2.1}$$

These parameters indicate the element timing can be modeled by a Gaussian probability density function, where nominal delay t_{pd} is the mean value and jitter σ_{pd} is the standard deviation, as seen in Fig. 2.4(c) below. Jitter is present to varying degrees through the readout electronics, including the preamplifier, discriminator, and digitizer, depending on circuit architecture.

¹Similar to circuit jitter described in the next passage, the degree to which a sensor degrades timing resolution is quantified as an RMS quantity. Discussion of this, however, is beyond the scope of this work.



Figure 2.4: (a) Generic element with mean propagation delay t_{pd} . (b) Timing diagram of many trials with varying Δt_{pd} . (c) Gaussian distribution of propagation delays with deviation σ_{pd} .

In addition to jitter, another large source of timing uncertainty, termed 'time walk', is present in the discriminator. Discriminators thresholds are usually implemented with comparator thresholds, which convert the pulse to a time-over-threshold signal. Since the charge pulses from the capacitive sensor have a constant rise time, pulses with a large magnitude cross the threshold earlier. This amplitude dependent threshold crossing effect is pictured in Fig 2.5. Circuit techniques for eliminating time walk from discriminators do exist, including the constant fraction discriminator (CFD), but these are power and area intensive, and so are not used in compact pixel detectors with many channels.



Figure 2.5: Basic phenomenon of "time walk", where crossing time "t" depends on pulse amplitude.

The final stage of the read-out chain is digitization of the TOA signal. When a signal is quantized, error is introduced equal to the difference between the continuous true input value and the discrete output value it is mapped to. Similar to quantization noise in an ADC, if one assumes a continuous TOA signal is uniformly distributed from measurement to measurement, the expected RMS contribution of quantization error to the timing resolution can be estimated as

$$\sigma_{digitizer} \approx \frac{t_{LSB}}{\sqrt{12}} \tag{2.2}$$

where t_{LSB} is the nominal increase in the continuous time input that produces an increment in the *least* significant bit of the discrete output code [10]. As can be observed, this degradation of the timing resolution can be arbitrarily minimized if t_{LSB} can be reduced. With careful design, circuits providing this function can easily achieve time digitization precisions $\sigma_{digitizer}$ below 10 ps RMS [11–13].

For the complete readout signal chain to have an end-to-end timing resolution on the order 100 ps, the timing resolution of each element must be reduced such that when all terms are summed in quadrature, the sum is less than the link budget. Expressed mathematically, we have Equation 2.3, where σ_{preamp}^2 , $\sigma_{discrim}^2$, and $\sigma_{digitizer}^2$ are the degradation of timing resolution due to non-idealities in the preamplifier, discriminator, and digitizer, respectively [14]. As the terms are added in quadrature, the element with the poorest timing resolution dominate the performance of the entire link.

$$\sigma_{total} = \sqrt{\sigma_{sensor}^2 + \sigma_{preamp}^2 + \sigma_{discrim}^2 + \sigma_{digitizer}^2}$$
(2.3)

As current-generation readout electronics have a timing resolution worse that of state-of-the-art sensors [4], this motivates upgrades to ROICs to avoid severely bottlenecking sensor timing performance. Of particular importance are the jitter of the amplifier and the discriminator time walk. These two circuits together a typically referred to as the analog front-end (AFE) of the pixel detector. Improving the timing precision of the AFE is the key to achieving an overall performance in the channel below of 100 ps RMS.

2.3 28 nm AFE Timing Performance

Recent generations of hybrid pixel detector ROICs have been designed in 65 nm bulk CMOS processes [4], but these monolithic designs do not have the precision necessary to fit within the overall 100 ps budget. One promising avenue to build an AFE with acceptable timing performance is to migrate ROICs design a smaller process node, which have faster propagation delays and experience less jitter [2]. Additional benefits include reduced power consumption, better circuit density, and improvements to some types of radiation hardness [3, 14]. To this end 28 nm CMOS is being evaluated by the HEP community for use in nextgeneration hybrid pixel detectors. When migrating to a new process, however, all circuit blocks must be thoroughly tested, and some entirely redesigned to properly take advantage of smaller feature sizes. In the transition to 28 nm, thorough testing will be especially necessary for circuits with an analog architecture, such as the preamplifier and discriminator which together comprise the analog front-end (AFE) of a pixel detector, and represent the largest source of timing uncertainty [14] in the system. In light of this, it is critical that new 28 nm AFE designs be characterized in isolation, before being integrated into a larger hybrid pixel detector signal chain. The requirements for testing these AFEs can be informed from simulation data of predicted performance of AFE currently under development.

Figure 2.6 is produced from simulation data of a real 28 nm AFE currently under design; which will be the first to be fabricated and tested. This plot shows how the phase delay of the system varies with respect to input charge injection: a characterization of the time walk. The total variation in AFE propagation delay, $\Delta t_{pd,AFE}$ is approximately $2 \text{ ns} \rightarrow 6 \text{ ns}$.



Figure 2.6: Time-walk of the AFE under test, vs input injection charge.

As previously mentioned, the tunable AFE discriminator threshold Q_{th} is used to reject false positive hits, and in hybrid pixel detectors is typically on the order of 1 ke^- to 1.5 ke^- . It is defined experimentally: to precisely determine the AFE threshold, a fine incremental sweep of the charge injection magnitude is performed. Input injections significantly below the threshold have essentially a 0% hit probability, and those far greater than threshold the chance are detected with nearly 100% chance. Due to noise, however, immediately surrounding the true threshold, there is a region of uncertainty where the hit probability transitions from 0% to 100%. This produces an S-shaped detection curve (or 'S-curve') where the 50% detection point is defined as the discriminator threshold.

The time-over-threshold the same AFE $t_{tot,AFE}$ has also been characterized, the result of which is shown in Figure 2.7. This figure ranges from $0.1 \text{ ns} \rightarrow 18 \text{ ns}$, once again depending on input change quantity.



Figure 2.7: TOT of the AFE under test, vs input injection charge.

The importance of $\Delta t_{pd,AFE}$ and $t_{tot,AFE}$ will be addressed later in more detail, where it will be established that their sum determines the minimum temporal dynamic range requirement of the test system to be built. Similarly, the array of charge injections necessary to experimentally observe these timing characteristics determines the charge injection dynamic range.

Finally, jitter in this 28 nm AFE, $\sigma_{t_{pd,AFE}(jitter)}$, must also must be measured. AFE jitter is expected to be on the order of 50 ps RMS [5]. Jitter measurement is a classic statistical precision problem. To properly estimate with certain degree of confidence that a true population statistic is within a certain window, one must take a certain number of samples from that population, and ensure that the measurement itself does not have a larger degree of uncertainty that the magnitude of the population statistic being measured. In short, to be able to accurately observe the gaussian distribution of the jitter at the output of the AFE, the measurement system itself must exhibit timing uncertainty significantly less than that that less than the AFE itself. Assuming the AFE jitter is approximately 50 ps RMS, Figure 2.8 shows the measurement system precision that would be necessary to achieve a certain accuracy in the measurement of that AFE precision.



Figure 2.8: Jitter in measurement electronics introduces error to measurement of DUT jitter.

2.4 Proposed Testbed for 28 nm AFE Characterization

This work presents the design of a test system capable of characterizing the the time-walk, jitter, propagation delay, and TOT of AFEs constructed in 28 nm bulk CMOS. The system must first inject an artificial sensor charge pulse at the input of the AFE device-under-test (DUT) and then measures the output timing variation of the AFE TOT output. Unfortunately, accurate measurement of the jitter in a monolithic circuit is difficult with external lab equipment, as parasitics and other environmental conditions corrupt high speed analog signals. A basic test setup is shown in Figure 2.9 It has numerous issues, including external charge injection signals being distorted by wire bond impedances, drivers and limited wire bonds bandwidth adding considerable jitter to input/output analog waveforms, and the reliance on an external scope making automated multi-channel testing infeasible.



Figure 2.9: Basic test setup, with only external test equipment.

A better solution, as proposed by this work, is a monolithic testbed. The system, called 'BigRock' integrates all analog test fixtures on-chip. With the exception of a single external 1 GHz clock, all control

and readout interfaces are low speed digital signals. The device-under-test (DUT) is any variety of pixel detector AFE in 28 nm needing characterization. On the first rising edge of the 1 GHz clock, the circuit injects a test charge pulse into the DUT, meant to closely emulate those produced by pixel sensors. After some delay, the DUT will output its corresponding TOT signal. The integrated test circuit then measures the time delay between the input pulse (START) and this output TOT signal (STOP) using a fine-resolution TDC. The external FPGA controls the integrated test circuit and reads out measurements, which can then be used to characterize AFE jitter.



Figure 2.10: Improved test setup, with integrated test fixtures.

As a quick aside, before proceeding, some basic design principles must to be touched on. This monolithic test circuit is to be built on a 28 nm wafer with a post-shrink area of 2 mm x 2 mm, with the aim to integrate several dozen AFE test channels per chip. There is no hard power consumption cap, as long as the chip does not overheat due to thermal dissipation. The nominal supply voltage is to be 900 mV, and specified by the 28 nm process.

Standard cells were not available at the time of design and so all design was custom. Even if they were available, standard cells would likely not meet radiation hardness requirements that apply to this circuit. A total ionizing dose (TID) of 1000 Mrad of radiation hardness must be tolerated by the analog front-end designs, therefore so must the test circuitry integrated alongside it. This requirement can be met by paying special attention to device sizes. In particular minimum size devices should be avoided; device widths should be kept larger than 200 nm and length to 40 nm or more. Particularly long devices lengths of 1 µm or more should also be avoided [15–17].

The design should be kept as low-risk as possible, in order to increase the chances of a functional test on the first fabrication run. All work is to be verified by circuit simulation alone (no HDL usage), using Cadence Virtuoso IC6.1.7, Spectre APS, and Calibre DRC/LVS/PEX v2017.1.

Chapter 3: Charge Injection Subcircuit

3.1 Principle of Charge Injection

Charge injection is achieved by producing a transient current $i_{Qinj}(t)$ waveform at the node of interest. The magnitude of charge injected Q_{inj} is defined by integration

$$Q_{inj} \approx \frac{1}{1.602 \times 10^{-19}} \int i_{Qinj}(t) dt$$
 (3.1)

where, for unit conversion, the coefficient is equal to the number of elementary charges per coulomb [14]. A transient current with the desired integrated charge can be generated at a node by connecting one terminal of a capacitor to it and then applying a voltage step at the opposite terminal. The product of the voltage step size V_{step} and the series capacitance C_{inj} determines the area of the current transient, and thus the charge injected.

$$Q_{inj} \approx \left(\frac{1}{1.602 \times 10^{-19}}\right) (C_{inj}) (V_{step}) \tag{3.2}$$

The slew rate of the input voltage step, the impedance of the series injection capacitor, and the bandwidth of the output node determine the transient rise and fall time [4]. The polarity of the output signal depends on whether the nature of the voltage step is rising or falling. A simplified circuit showing this basic operation is pictured in Figure 3.1, where the example of a falling input step is shown, with a corresponding negativepolarity output transient. Z_{in} models the input impedance of the AFE receiving the injection plus the shunt pixel sensor capacitance C_{det} .

Initially (1), the pull-up network is enabled and the voltage $V_{step}(t)$ at the connected node is slowly charged to the injection voltage V_{inj} . Subsequently (2), the pull-up is disabled, pull-down enabled, and the voltage $V_{step}(t)$ falls to the ground supply voltage, producing a swing with magnitude $V_{inj} - V_{ss}$. This voltage swing causes the formation of the charge injection current waveform $i_{Qinj}(t)$ at the opposite terminal of the C_{inj} series injection capacitance.

It should be noted, for this charge injection to behave as described, two assumptions about Z_{in} must be true. First, the impedance at low frequencies must be quite high, on the order of M Ω . Low resistances would cause a large DC leakage which would add in superposition $i_{Qinj}(t)$. This is generally a nonissue, as the shunt detector C_{det} has high parasitic resistance and the AFE input node is composed of FET gate terminals and capacitor feedback, both high-impedance at AC. The second assumption is that the total shunt input capacitance must be significantly larger than that of the series injection capacitor C_{inj} . The input capacitance of Z_{in} , which can be denoted as $C_{in} = C_{det} + C_{AFE}$, forms a capacitance voltage divider with C_{inj} . If the AC voltage swing at output of the pull-up/pull-down network is denoted $v_{step}(t)$ and the swing at the input of the AFE is $v_{in}(t)$, we can express voltage drop across C_{inj} at high frequencies as

$$v_{step} - v_{in} = v_{step} - v_{step} \left| \left| \frac{Z_{in}}{Z_{inj} + Z_{in}} \right| = v_{step} \left(1 - \frac{C_{inj}}{C_{inj} + C_{in}} \right) \right|$$
(3.3)

For the approximation made by Equation 3.2 to be true, the difference between $V_{step}(t)$ and the voltage drop across C_{inj} , equal to $v_{step}(t) - v_{in}(t)$, must be negligible. Equation 3.3 indicates this is true when $C_{inj} \ll C_{in}$. Thus the series injection capacitor must be designed to be at least an order of magnitude smaller than the load capacitance for proper operation.



Figure 3.1: Basic principle of a charge injection circuit.

3.2 Circuit Requirements

To accurately characterize a charge-sensitive AFE the injection circuit must be able to emulate the diversity of signals produced by a typical capacitive pixel sensor, as described in Section 2.1. To measure the AFE time walk and time-over-threshold (TOT), an injection dynamic range stretching from the $100 e^-$ RMS

noise floor to more than $20 \,\mathrm{ke}^-$ is necessary.

Of particular importance are charge inputs close in magnitude to the AFE threshold, where the majority of timing variation occurs (time walk, etc). As mentioned in Section 2.2, the detection threshold is typically set in the range of 1 ke^- to 1.5 ke^- . To measure multiple points in the probability transition region of the AFE threshold S-curve, the injection circuit must be able to sweep or "scan" its output in steps smaller in magnitude than the noise floor. Too coarse a resolution would result in a scan which entirely misses the transition region of the S-curve, thus degrading precision of the threshold measurement. Thus the tunable resolution of the charge injection is therefore required to be much less than the 100 e^- RMS noise.

To approximate the transient characteristics of a real charge pulse, a rise and fall time on the order of 100 ps are desirable but not strictly specified. The bandwidth of the charge injection circuit does not impact the magnitude of charge injection (area of the current pulse), so the only requirement is that the dynamics of the injection pulse occur at a time scale much faster than the bandwidth of the feedback charge-sensitive amplifier that typically composes the AFE input stage. In addition, to meet the overall 23 ps RMS timing precision budget of the test system, it is critical that the timing of the charge injection pulses display minimal jitter. Reserving the majority of the budget for the time digitization circuit, 2 ps RMS is allotted to the charge injection circuit.

As specified by the 28 nm process, the nominal supply voltage is 0.9 V. The majority of gates in the injection circuit will need to be operated from this supply level, with the exception of those receiving a variable external reference voltage used to tune charge injection quantity. Both the main 0.9 V supply and tunable injection supply will exhibit noise originating from the external supply hardware, causing imprecision in the magnitude and timing of charge injection between runs. Therefore, the aforementioned injection precision and jitter must meet specification with at least 1 mV RMS supply noise.

The previously discussed radiation hardness, operating temperature, power, and area constraints apply at the sub-circuit level. Silicon area should be kept to a maximum of 60 µm x 60 µm as the time-digitization circuits are area intensive, and several dozen channels need to be integrated on-chip. A summary of prescribed design requirements is provided in Table 3.1.

3.3 Fine Resolution Tunable Charge Injection Circuit

Building on the principles introduced in Section 3.1, a unit charge injection circuit was designed, as shown in Figure 3.2. Design of this unit element circuit is critically influenced by the integrated devices able to be used for C_{inj} , the series injection capacitor. In the 28 nm CMOS process both n-type MOSCAPs and standard MOM capacitors are available. MOSCAPs posses good capacitance density per area, but are heavily

Figure of Merit	Requirement
Injection Dynamic Range	$100\mathrm{e^-}$ to $20\mathrm{ke^-}$
Injection Precision	$<<100\mathrm{e^-}\ \mathrm{RMS}$
Injection Rise Time	$\approx 100\mathrm{ps}$
Injection Fall Time	$\approx 100\mathrm{ps}$
Injection Jitter	$< 2 \mathrm{ps} \mathrm{RMS}$
Supply Voltage	$0.9\mathrm{V}$
Tolerable Supply Noise	$1\mathrm{mV}\ \mathrm{RMS}$
Radiation Hardness	$1000\mathrm{Mrad}$
Operating Temperature	$-10^{\rm o}{\rm C}$ to $50^{\rm o}{\rm C}$
Silicon Area	$< 60\mu{\rm m}$ x $60\mu{\rm m}$
Power Consumption	No Limit

Table 3.1: Charge injection circuit design specifications

dependent on bias voltage, due to its modulation of the effective oxide thickness transitioning the interface between accumulation and depletion. This rules out the use of MOSCAPs for the C_{inj} capacitor, as its value directly determines the quantity of charge. MOM capacitors are therefore selected; their reduced capacitance density is not problematic as the test circuit has a generous silicon area allowance.

Equation 3.2 illustrates that the charge injection quantity can be tuned by both voltage step and capacitor size. In the unit circuit (Figure 3.2), the voltage step is generated by a static pull-up/pull-down network. The pull-up network is formed by the PMOS/NMOS pair M1/M2, which can be considered as an analog transmission gate with M1 driven by the inverse of the signal applied to M2. This architecture is necessary because the positive rail V_{inj} is variable and can be as low as 50 mV. The gate control signals are full swing (0 mV - 900 mV), and so a PMOS-only pull-up would operate in cutoff for V_{inj} below the PMOS device threshold voltage. The addition of an NMOS device remedies this. The pull-down network, by contrast is formed by a single NMOS device M3, as the low-side supply rail is fixed at V_{ss} , or 0 mV. Therefore, the injection voltage step magnitude V_{step} is set by the value of V_{inj} .

For the unit circuit, a scan covering the typical AFE threshold from at least 1 ke^- to 1.5 ke^- is desired. The minimum size MOM capacitor that can be fabricated in the 28 nm process is 1.46 fF with an area of 0.4 µm by 1.6 µm. To provide a scan with margin above and below the AFE threshold, given a V_{step} ranging from 50 mV to 900 mV, five minimum size capacitors are combined in series for $C_{inj} = 0.29 \text{ fF}$. Extracted simulations of this MOM capacitor network indicate a more-than-acceptable parasitic resistance on the order of tens of m Ω .



Figure 3.2: Unit circuit for charge injection, with tunable injection voltage.

To incorporate functionality for enabling/disabling the circuit, a static CMOS NOR gate is added to the input. This logic allows the second input of the NOR gate to act as an active-high enable signal, disabling the injection when low. The inverting characteristics of the NOR gate conveniently eliminates the need for a second CMOS inverter to generate the complimentary control signals for the analog transmission gate pull-up.

The rise and fall time of the injection current pulse $i_{Qinj}(t)$ is determined by the bandwidth of the driver input and output nodes. In 28 nm, static gates exhibit rail-to-rail transitions less than 10 ps, and so the slew rate must be decreased to properly emulate signals from a real capacitive pixel sensor. To maximize on-resistance of the pull-up/pull-down network, near-minimum sizes for M1, M2, and M3 are used, with W = 500 nm and L = 50 nm. Dimensions any smaller would violate radiation hardness requirements. To further reduce speed, additional loading capacitance is added with $C_{slew} = 31$ pF and $C_{slew2} = 13$ pF.

Simulations for this architecture were performed with full parasitic extraction and a C_{det} load capacitance of 50 pF. Figure 3.3 displays the transient current $i_{Qinj}(t)$ vs time, at 50 mV increments of V_{inj} . The current waveforms peak in the μ A and have rise and fall times on order of 100 ps, as desired. Integration of these curves, via Equation 3.1, yields the charge injection quantities shown in Figure 3.4. The injection quantity continuously ranges from below 100 e⁻, for $V_{inj} = 50 \text{ mV}$, to maximum of 1.51 ke⁻, for $V_{inj} = 900 \text{ mV}$.



Figure 3.3: Fine resolution current transients vs time, at multiple injection voltages.



Figure 3.4: Fine resolution slice charge injection vs injection voltage supply.

A single instantiation of this unit-element circuit serves as the tunable "fine resolution" for our charge injection circuit. The enable input of this "slice" is tied to V_{dd} such that the element is always enabled, but tunable.

3.4 Coarse Resolution Charge Injection DAC

A single injection circuit slice can produce finely-tunable injections up to $1500 e^-$, but a total dynamic range greater than $20\ 000 e^-$ is required. To extend the dynamic range, a binary weighted array of 15 additional unit injection circuits are added, as shown in Figure 3.5. (These 15 slices are in addition to the 1 discussed in the previous chapter, for a total of 16.) The output of all slices is tied together in parallel, such that the current waveform from each adds in superposition. The 15 additional slices have their injection voltage fixed to V_{dd} (900 mV) and the binary-weighted grouping allows for simple 4-bit control via the coupled control signals. Unlike tuning the fine resolution V_{inj} voltage, incrementing the digital control code $B_{0:3}$ causes the charge output to increase in quantized steps. This circuit, which can be characterized as a charge digital-to-analog converter (CDAC), provides the coarse resolution of the complete injection circuit.



Figure 3.5: Binary weighted array of 15 injection circuits (CDAC), providing coarse resolution.

One issue of particular concern is how each of the parallel slices of the circuit impact the linearity of the circuit when enabled vs disabled. Enabled injection slices operate in parallel and behave as though their injection capacitance were summed. When disabled, an injection slice indefinitely holds the V_{step} node constant at $V_{inj} = 900 \text{ mV}$, or small signal ground at high frequencies. Therefore, the injection capacitance of disabled slices simply appears at the current summation output node in shunt with the input capacitance of the next stage, C_{in} . (Recall that C_{in} is composed of both C_{AFE} and C_{det} which typically dominates at around 50 fF.) At the minimum control input code $B_{0:3} = 0000$, only the tunable injection slice is enabled so $C_{inj} \approx 0.3$ fF, and the 15 other slices are disabled in shunt so $C_{in} \approx 50.5$ fF. In the case of the maximum injection code $B_{0:3} = 1111$, all slices are enabled so $C_{inj} = 4.8$ fF and $C_{in} \approx 50$ fF. Referring to Equation 3.3 and assuming no additional loading, an estimated 9 percent non-linear deviation is present at the maximum value ($B_{0:3} = 1111$), relative to the ideal linear slope established by the lowest input ($B_{0:3} = 0000$). In reality, the parasitic capacitance of the metal interconnects may increase the effective capacitance to ground in parallel with C_{det} , ameliorating the nonlinearity at the maximum input code.

Simulations with parasitic extraction and $C_{det} = 50 \text{ pF}$ were carried out with the coarse CDAC circuit and fine tunable injection circuit in conjunction. The transient current at the output node of the injection circuit is shown in Figure 3.6. The incremental steps between each current pulse correspond to a 1-bit increase of the 4-bit $B_{0:3}$ code. Across the dynamic range of the charge injection circuit, we can see a nearly constant rise time of 40 ps and a fall time of 100 ps, which is acceptable to approximate the charge injection pulses of a real capacitive pixel sensor.



Figure 3.6: Coarse resolution current transients vs time, at multiple CDAC codes.

The integrated charge under the curve of these pulses is displayed in Figure 3.7, in the form of the input-output coarse CDAC transfer function. As expected, for a digital control code of 0 ($B_{0:3} = 0000$), the charge injection quantity is 1.51 ke^- , as all the CDAC slices are disabled and the tunable unit is providing its maximum injection. When control code is set to 15 ($B_{0:3} = 1111$), all the slices are enabled. Considering

the unit capacitance C_{inj} of a single slice is 0.29 fF, and the total in parallel is $(16)(C_{inj}) = 4.67$ fF, Equation 3.2 predicts a maximum charge injection of 26.2 ke^- . Simulation, however, reveals nominal charge injection quantity at the maximum input code to be 24.1 ke^- , corresponding to a 8 percent nonlinearity. This is inline with prediction, as C_{inj} approaches the same order of magnitude as C_{det} . While difficult to measure externally once fabricated, this nonlinearity is of little consequence, as high accuracy of the injection quantity is only needed at small injections around the AFE discriminator threshold.



Figure 3.7: Fully range of charge injection vs CDAC code.

3.5 Precision Timing and Charge Quantity

Transient variation in the supply voltage modulates the propagation delay of the injection circuit. The magnitude of this effect, typically referred to as power supply induced jitter (PSIJ), determines the timing precision of the circuit alongside inherent device noise. To verify the injection circuit conforms to the allotted jitter budget, transient noise simulations were run with white supply noise $\sigma_{V_{dd}} = 0.5 \text{ mV}$ and device thermal noise with a 40 GHz bandwidth. The result (in Figure 3.8) of 200 runs, with $B_{0:3} = 0000$ and tunable slice $V_{inj} = 900 \text{ mV}$, reveal a jitter of $\sigma_{t_{pd}} = 685 \text{ fs}$, comfortably within the allotted 2 ps budget.

Power supply and device noise also impact the quantity of charge injected, and the data in Figure 3.8

can be used to examine the charge injection precision. While the coarse charge resolution of the CDAC is $Q_{LSB} = 1.51 \text{ ke}^-$, the continuously tunable nature of the fine injection circuit means the overall precision of the complete circuit is actually determined by voltage noise. Fortunately, as charge is an integrated quantity, uniformly distributed noise in the pulse period is largely eliminated. The total injection noise is measured by simulation to be a mere $\sigma_{Q_{in}} = 12 \text{ e}^-$, far below the approximate 100 e^- noise floor of a typical capacitive pixel sensor coupled to an AFE with discriminator. This precision, in conjunction with the high accuracy in the region of discriminator thresholds, means the injection circuit is expected to characterize the AFE threshold (and, predicated on the performance of the time digitization circuit, timing) with excellent fidelity.



Figure 3.8: Charge injection precision and jitter (200 runs)

3.6 Injection Circuit Top Level and Layout

A top-level block diagram of the circuit is shown in Figure 3.9. The always-enabled single injection slice with tunable V_{inj} for fine resolution is shown in parallel with the 15 slices of the coarse CDAC with V_{inj} fixed. The current i_{Qinj} from each slice is summed at the output node with the AFE Z_{in} modeled. As previously discussed, a 50 fF capacitor C_{det} is included at the output, in shunt with Z_{in} , to model the sensor capacitance present in real pixel detector configurations. To provide enough current to drive the trigger and enable inputs, a fanout buffer is constructed from inverter stages of increasing size, following a "fanout-of-four" (FO4) scheme.

Connected to the fanout buffer (left side of Figure 3.9) is a circuit which synchronizes the trigger signal for the injection slices with the first rising edge of the periodic start signal sent to the time-digitization circuit. A simple converter is included to drive the single-ended CMOS input of the flip-flop from the pseudo-differential clock. A fully-differential analog switch also buffers the clock signal, to create the start signal for the time-digitization circuit.

The behavior of this circuit is as follows: Throughout operation, the clock signal is constantly switching at rate of 1 GHz. Initially, the Q output of the flip-flop is low, and the differential analog buffer driving the start signal is disabled, so high impedance static pull-ups/pull-downs set the start output to low. To start the measurement sequence, the prime signal is raised, and on the next rising edge of the clock signal, the output of the flip-flop raises high, which both triggers the injection circuits and makes the differential analog buffer transparent, generating the periodic start signal from the clock. This trigger signal rises approximately 100 ps after the clock rising edge. To avoid this causing an offset in the timing of the first rising edge of the start signal, a delay element is added inside the analog buffer, greater than the sum of the differential-tosingle-ended clock buffer and the flip-flop clock-to-Q delay.

At the end of measurement, the system is reconfigured with the asynchronous active-high reset input of the flip-flop. When the flip-flop Q output falls upon reset, the trigger circuits restore their internal V_{step} node back to the voltage V_{inj} and the analog buffer for the start signal once again becomes opaque.



Figure 3.9: Top level schematic of the charge injection circuit

The layout of the charge injection circuit is shown in Figure 3.10. The various cells are organized in order of signal flow from left to right, starting first with the input trigger synchronization and the fanout buffers. In the center, the 16 injection slices can be seen, where the majority of the area is consumed by the C_{inj} , C_{skew} , and C_{skew2} capacitors. Along the right hand side of the cell is the distributed 50 pF C_{det} capacitor. The total area of the circuit is 30 µm by 50 µm.

The expected performance of the charge injection circuit is summarized in Table 3.2. All design parameters are verified via simulation to meet or exceed requirements.

Figure of Merit	Requirement	Simulated Performance
Injection Dynamic Range	$100\mathrm{e^-}$ to $20\mathrm{ke^-}$	$50 \mathrm{e^-}$ to $24.1 \mathrm{ke^-}$
Injection Precision	$<< 100 \mathrm{e^-}$ RMS	$12 \mathrm{e^-}\ \mathrm{RMS}$
Injection Rise Time	$pprox 100\mathrm{ps}$	$40\mathrm{ps}$
Injection Fall Time	$pprox 100\mathrm{ps}$	$100\mathrm{ps}$
Injection Jitter	$< 2\mathrm{ps}\ \mathrm{RMS}$	$0.69\mathrm{ps}\;\mathrm{RMS}$
Supply Voltage	$0.9\mathrm{V}$	$0.9\mathrm{V}$
Tolerable Supply Noise	$1\mathrm{mV}\ \mathrm{RMS}$	$1\mathrm{mV}\ \mathrm{RMS}$
Radiation Hardness	$1000\mathrm{Mrad}$	$1000\mathrm{Mrad}$
Operating Temperature	$-10^{\rm o}{\rm C}$ to $50^{\rm o}{\rm C}$	$-10^{\rm o}{\rm C}$ to $50^{\rm o}{\rm C}$
Silicon Area	$< 60\mu\mathrm{m} \ge 60\mu\mathrm{m}$	$50\mu\mathrm{m} \ge 30\mu\mathrm{m}$
Power Consumption	No Limit	$6.5\mathrm{mA}$ peak





Figure 3.10: Top level layout of the charge injection circuit

Chapter 4: Time Digitization Subcircuit

4.1 Time-to-Digital Converter Requirements

To characterize the timing properties of the AFE under test, a circuit capable of precise time interval measurement is required, typically referred to as a 'time-to-digital converter'. Time-to-digital converters (TDCs) are a broad class of circuits categorized by their ability to receive a continuous time difference signal as an input and derive a corresponding quantized digital value at their output. Similar the analogto-digital converters [18], TDC performance can be described by the metrics of dynamic range, nominal resolution/precision, effective resolution/precision, jitter, nonlinearity, offset error, and gain error.

The dynamic range of a TDC is the longest time period that can be measured without exceeding temporal limits of the counting and/or interpolation circuits that make up the TDC. In this work, the temporal dynamic range of the TDC must be large enough to capture the complete behavior of the TOT signal at the AFE output. To account for both the phase delay (as long as 6 ns with time walk) and the full range of possible time-over-threshold durations (up to 15 ns), a dynamic range of greater than 21 ns is required. To measure the AFE jitter with at least 90% accuracy, a measurement system with timing precision better than 23 ps RMS is required. As the charge injection circuit (detailed in Chapter 3) has a jitter less than 1 ps RMS, essentially all the timing precision budget can be allocated to TDC.

The precision and resolution of a TDC are two equivalent ways to describe the uncertainty of its timing measurements. Single-shot timing precision is the standard deviation of random timing variation observed for an input time, without repetition, and is measured in units of RMS time [10]. Resolution expresses this same uncertainty, but in terms of the hypothetical bit-count an ideal TDC, free of jitter and nonlinearity, that would achieve the same performance. In either case, nominal resolution and precision are set by the quantization error alone, whereas effective precision and precision account for the additional variation due to timing noise (jitter) and nonlinearity [10]. It the application space of HEP instrumentation, it is more common to work with RMS uncertainties, and so nominal and effective resolution, a.k.a. effective number of bits (ENOB), will be largely neglected in favor of using precision as a metric when describing timing uncertainty.

The nominal timing precision of a TDC is determined by quantization noise, which can be approximated via Equation 4.1, where t_{LSB} is the quantization interval or temporal 'least-significant-bit' (LSB) of the TDC [10].

$$\sigma_{TDC(qn)} \approx \frac{t_{LSB}}{\sqrt{12}} \tag{4.1}$$

 $\sigma_{TDC(qn)}$ represents the ideal, best case (minimum) uncertainty for a converter with a given nominal t_{LSB} . The approximation of Equation 4.1 assumes that the single-shot time interval inputs experience enough variation as to be uniformly distributed across the quantization intervals in which they fall. In most real converters, where the magnitude of electronic noise is comparable to the quantization noise, this assumption is appropriate [11].

Offset error describes the constant measurement inaccuracy caused by the aggregate propagation delay of all TDC components not directly leveraged for time quantization. It causes as a horizontal shift in the transfer function of a This type of error is fortunately easy to characterize and calibrate. The simplest method to determine the offset is measure a zero time interval, and then examine what value is reported by the TDC. In all subsequent measurements, this value can then be subtracted from the result to remove the offset.

Gain error is a function of how far the average t_{LSB} has shifted from its nominal value, which manifests as a systematic change in the slope of the TDC transfer function. In most cases gain error is caused by some combination of chip-wide deviations in process parameters, supply voltage, or ambient temperature [18, 19]. After removing offset error, a calibration factor can be determined for the linear gain error by measuring the difference between the actual and ideal final tripping point, and then dividing by the total number of tripping points. TDC measurements can then be simply scaled by this factor.

After offset and gain error have been accounted for, some residual static inaccuracy in the TDC transfer function remains. It can be described as random deviation in each tripping points from the ideal values, is referred to as nonlinearity. This error is difficult to calibrate, but can be minimized by paying attention to the potential for device mismatch in the design phase. The variation in t_{LSB} from tripping point to tripping point is defined as the TDC differential nonlinearity (DNL). Computing the running total of these incremental deviations along the converter transfer function yields the integral nonlinearity (INL). It should be noted, since calibration is never a perfect process, any offset or gain error that fails to be removed by prior calibration steps simply heightens the effective nonlinearity that must be tolerated in the TDC transfer function. As aforementioned, the baseline for single-shot precision is set by quantization error, but impairments including internal jitter, nonlinearity, and persistent offset/gain error further degrade the effective precision that is actually observed. [10] Thus a real TDC always has an effective precision worse than the nominal value. In light of this, to achieve the necessary effective timing precision of 23 ps, the nominal resolution of the TDC must be designed to a more conservative value, leaving headroom for non-ideal impairments.

A summary of the design requirements for the time-digitization subsystem, to be implemented by a TDC circuit, are shown in Table 4.1. The same requirements for supply voltage, operating temperature, and radiation hardness apply to the TDC as did apply to the charge injection circuit. Similarly, no power consumption requirement exists as the TDC will be used as part of a monolithic test fixture. Due to the fact that the TDC will likely operate from a supply with larger and more frequency switching transients (due to the complexity of a TDC itself), a higher supply noise of 3 mV RMS must be tolerated.

Figure of Merit	Requirement
Dynamic Range	$> 25\mathrm{ns}$
Effective Timing Precision	$< 23\mathrm{ps}\;\mathrm{RMS}$
Supply Voltage	$0.9\mathrm{V}$
Tolerable Supply Noise	$3\mathrm{mV}\ \mathrm{RMS}$
Temperature	$-10^{\rm o}{\rm C}$ to $50^{\rm o}{\rm C}$
Radiation Hardness	$1000\mathrm{Mrad}$
Silicon Area	$< 60\mu{\rm m}~{\rm x}~500\mu{\rm m}$
Power Consumption	No Limit

Table 4.1: Time-to-digital converter circuit design specifications

4.2 Unit Delay Cell Design

Time-to-digital converters are a broad class of circuit with diverse implementations, each with comparative advantages and deficiencies. In larger, older process nodes, TDCs are most often constructed by simply transducing the input time interval into a voltage, and then subsequently digitizing this voltage with a traditional ADC circuit [20]. This approach is advantageous when voltage supplies are on the order of several volts, as signals are readily processed in the amplitude domain.

However, as process technology nodes have shrunk, so has the available voltage headroom, thus compressing the range available for voltage signal processing. Fortunately, the speed and timing precision of gates has simultaneously improved with each new generation, which increasingly allows for processing of signals in the time domain [20]. Modern TDC architectures leveraging this trend can be divided into two categories: those with t_{LSB} set by the propagation delay of unit delay cell in the technology, and those with the ability to interpolate time differences at a sub-gate delay level [10]. The latter style of architectures is quite common in high-precision commercial TDCs, but require design-intensive internal calibration circuitry, as deviating interpolation circuits can easily cause non-monotonicity in the converter transfer function [21,22]. Fortunately, in 28 nm, the gate delay of static CMOS gates is on the order of 10 ps - 20 ps, meaning design requirements can be met without the need for sub-gate delay precision. In this designing the delay element, however, only a single CMOS stage can be used, as a two-gate solution would exceed the maximum 20 ps t_{LSB} allowable to stay within the 23 ps RMS timing precision budget.

Initially, simple single-ended static CMOS inverters were considered for TDC delay elements. Inverters have the advantage of being compact and energy efficiency, with minimal layout parasitics resulting in an excellent propagation delay on the order of 12 ps in 28 nm [23]. Several issues exist with this basic approach, however. First, as inverters change the polarity of the time signal, the t_{LSB} is alternatively determined by rising and falling propagation delay t_{pd} . Making these two equal is difficult considering process, voltage, and temperature variations, and impossible without calibration. Another issue is that the low device count of the inverter makes the cell delay very susceptible to mismatch; the propagation delay is determined by the size of a single transistor. Finally, as inverters can not independently provide a 1-bit quantization of their output voltage, a separate latching comparison circuit is needed. Since the output of the inverter is single-ended, a static mid-supply reference voltage would be needed for the second input of the comparison circuit.

To remedy the issues of toggling polarity and sensitivity to device mismatch, a pseudo-differential architecture was next considered, with two inverters acting as delay element, and two cross-coupled inverters across the output to correct skew. This design increases power and area, but this is a non-issue as discussed in the design requirements. The propagation delay increases to 15 ps, but this still fits within our timing precision budget. The mitigation of device mismatch reduces the propagation delay variation to around $\sigma_{t_{dcell(mismatch)}} = 250$ fs and the alternating delay effect no longer exists. To digitize the output of this circuit, and comparator built from a regenerative cross-coupled latch can be used. Since both inputs of the comparator are driven simultaneously, in opposite directions, the effective slew rate of timing signal being digitized is doubled. This means the time window during which the regenerative latch is susceptible to noise in evaluation is cut in half.

To further simplify the design of the unit delay cell, the final design, shown in Figure 4.1 was selected [24–26]. The power supply enable/disable transistors typically used within a regenerative latching comparator are moved to the feed-forward delay inverters, allowing the entire comparator circuit to be eliminated, as the delay element is now capable of self-evaluation. This design significantly reduces the transistor count needed without increasing susceptibility to mismatch or process variation; $\sigma_{t_{dcell(mismatch)}}$ remains at 250 fs and the jitter of the circuit $\sigma_{t_{dcell(jitter)}}$ is a mere 70 fs. The only cost paid is a marginal increase in propagation delay to $t_{cell} = 17.2$ ps, verified via parasitic extraction. The operation of this circuit is as follows: During normal operation (1), the power supply cutoff transistors are enabled, and the cell acts as a standard pseudo-

differential delay cell, with cross-coupled inverters at the output effectively removing skew. When the signal to stop and evaluate the delay cell arrives (2), the feedforward inverters have their power supply removed, forcing their outputs to a high impedance state. At this point, the cross-coupled inverters fully take over, and their regenerative feedback reinforces the current state until the complimentary output reaches full-swing logic levels. The pair of outputs are then finally converted to a single-ended digital signal by a buffer with low input capacitance.



Figure 4.1: Unit delay element with power supply cutoffs, cross-coupled inverters for de-skew/evaluation, and output buffer.

Using this delay cell, a TDC can be assembled in the form of either a simple delay line or ring-oscillator architecture. Ring-based topologies are very space efficient, but not appropriate for low-complexity testing infrastructure, as they are difficult to calibrate. Using more devices, in a long delay line, also improves TDC precision because the impact on linearity of a heavily mismatched element is relegated to a single occurrence, and the lack of wrap-around in the topology prevents jitter from accumulating. The unit cell in Figure 4.1 has a small layout area of only 3 µm by 4 µm. Given the silicon area allowance, inherently higher performance, and the need for design simplicity, a delay line architecture is chosen for the TDC.

4.3 Delay Line Gain Error and Calibration

Calibration of the delay line is necessary to remove gain error in the input-output transfer characteristic of the TDC. This gain error originates from two separate processes. The first is change in the nominal propagation delay of the cells in the delay line due to process, voltage and temperature (PVT) variation. Circuit simulation of the unit cell reveals the following potential ranges for t_{dcell} deviation: $t_{dcell(process)} = 16.2 \text{ ps} \rightarrow 19.8 \text{ ps}$ $t_{dcell(voltage)} = 16.6 \text{ ps} \rightarrow 19.4 \text{ ps}$ $t_{dcell(temp)} = 16.9 \text{ ps} \rightarrow 18.1 \text{ ps}$

The second source of gain error is the accumulation of non-mean-free non-linearity along the line, due to local device mismatch [18]. In most cases, by the end of a delay line, the INL should be approximately zero (via law of large numbers sampling from a mean-free gaussian) but this will not necessarily be the case for all fabricated instances of the TDC.

Any degree of gain error which fails to be removed via calibration contributes to the integral non-linearity of the TDC. Therefore, to avoid unnecessary degradation of the TDC precision, as much of this total gain error must be removed as possible. Delay lines are calibrated by ensuring that the total propagation delay through the line is kept constant across this PVT and aggregate mismatch variation. This is typically accomplished in real time using a delay-locked loop approach, with a frequency reference, phase detector, and voltage controlled delay line [27]. The inclusion of feedback, however, would run counter to a low risk design philosophy, so a different approach is needed.

As the external 1 GHz clock has a known fixed frequency, with minimal jitter (below 2 ps RMS), it can be used as a timing reference for calibration. By simply including enough elements in the delay line such that the total line, over all PVT variation, is temporally long enough to contain at least two periods of the periodic START signal, each TDC measurement will have a calibration reference embedded. Rather than being using in real time, this information can simply be extracted from the output data and used to correct gain error with a high accuracy in post processing. Therefore, considering $t_{dcell} = 17.2$ ps the delay line is designed to have 128 elements to ensure it always continuously stores two full cycles of the start signal.

4.4 Top-Level TDC Architecture

The periodic start signal to the TDC is generated by the analog buffer inside the charge injection circuit (shown in Figure 3.9). The stop signal is derived from the TOT output signal from the AFE discriminator. Thus the interval between the start and stop signal is duration between the time when the injection circuit fires, and when the AFE responds. As we want to characterize both the rising and falling edge of the TOT signal, a circuit must be built to generate two corresponding time-aligned stop signals. The circuit in Figure 4.2 implements this function with a complimentary pair of positive and negative edge-triggered D flip-flops.



Figure 4.2: Diagram of the dual stop edge generation circuit.

Figure 4.3 is the top-level diagram of the complete TDC circuit which receives the pair of stop signals. It uses a symmetric dual architecture, with one half TDC for measuring the delay and jitter of the rising edge of the TOT signal, and the other identical half for measuring the falling edge, on each single-shot experiment.

To extend the dynamic range of the TDC, a coarse-fine topology is implemented. The aforementioned tapped delay line with picosecond precision will provide fine resolution necessary for measuring jitter, while the coarse resolution is provided by a up-counter triggered by the gated 1 GHz (1 ns period) start waveform. In both halves, the up-counter and delay line bits are loaded in parallel into a shift register which is then subsequently clocks to read out the data in a sequential manner. This allows all the measurement data to be extracted using a single clock signal, and two serial output data lines.



Figure 4.3: Top block-level diagram of the dual TDC architecture

Figure 4.4 shows the circuit-level implementation of the fine resolution delay line, coarse resolution delay line, and parallel-in serial-out (PISO) shift register. At the input of the circuit, the differential start signal coming from the injection circuit is fed into the delay line where is periodically propagates down the line, with the waveform at each subsequent node separated by the propagation delay of the cell. The delay line is constructed of 128 elements, such that the line has nominal time length of great than 2200 ps. Additionally, the start signal is converted to a single-ended waveform, which is used to increment the 6-bit coarse count every 1 ns, for a total dynamic range of 64 ns, allowing the TOT and time walk of the AFE to be captured regardless of input charge level.

When the rising edge of the stop signal arrives, it also branches, entering a large fanout buffer and disabling the gate on the coarse resolution counter, freezing it. The fanout buffer is a large array of inverters arranged with increasing drive strength such that the latch inputs of all 128 elements along the fine resolution delay can be switched at once. Each incremental drive strength stage is tied together to prevent mismatch from accumulating. At the output of the fanout buffer is a row of single-ended to pseudo-differential converters designed to ensure the complimentary latching inputs of the delay line can be driven with minimal skew. [28,29].

When the latch signal to the delay element is active, the delay line is frozen in its current state. Any taps along the delay line that were mid-supply, due to being in transition, are reinforced to the rails by the internal regenerative cross-coupled feedback, as detailed in Figure 4.1. After a waiting period to avoid metastability on the delay line taps, the system is ready for readout. With the PISO mode control set to 0, the serial clock is cycled once, to load all bits from the delay line and counter in parallel into the shift register. Small tap buffers are used for compatibility between the differential delay line and single-ended inputs of the CMOS flip-flops in the PISO.

After this operation, the PISO mode is switched to 1, and the serial clock is free to cycle 134 (128+6) times to sequentially read out all the data bits. The PISO should be clocked at a MHz rate to avoid issues with clock-to-Q delay. Finally, the whole measurement system can be reset by simply using the asynchronous reset pin on all the flip-flops to lower stop signal and erase the value stored in the coarse resolution counter and PISO.



Figure 4.4: Circuit-level diagram of one-half of the TDC, featuring the fine and coarse resolution circuits, and PISO.

4.5 Top-level TDC Layout

The nearly symmetric top level layout of the TDC is shown in Figure 4.5, with a total area of 45 µm by 475 µm. The delay lines with readout buffer and PISO dominate the majority of the TDC area, with the compact edge generation circuit and coarse counters found along the far left side of the cell layout. Not pictured in Figure 4.5 are the power supply decoupling capacitors which are needed to minimize power supply voltage fluctuations along the delay line length due to frequency switching transients. These capacitors are shared between adjacent TDCs.



Figure 4.5: Layout featuring edge selection circuit driving dual TDCs with coarse and fine resolution subcircuits.

4.6 Overall TDC Performance

The static non-linearity characteristics of the TDC were found via simple scans along the converter transfer function, with noise disabled. The simulation data is generated in the following manner: Monte Carlo simulation is used to first generate mismatches across the whole circuit of a hypothetical converter. For each of the 50 generated mismatched hypothetical TDCs, an incremental sweep along its transfer function is completed, and the tripping points for each quantization interval are recorded. Figure 4.6 shows the worse case observed from those runs, where the DNL max peaks at 0.23 LSB, or 3.95 ps. Integrating these, we find the worse case INL to be 0.15 LSB, or 2.58 ps. INL is pictured in Figure 4.7.



Figure 4.6: DNL vs input time of aggregate TDC (worst case of 50 runs via Monte-Carlo)



Figure 4.7: INL vs input time of aggregate TDC (worst case of 50 runs via Monte-Carlo)

To evaluate aggregate performance of the TDC, the single-shot precision of the circuit is first considered. Figure 4.8 displays the result of an in-depth simulation along the length of the converter transfer function, with device noise and non-linearity from mismatch incorporated. The whole range of the TDC's fine delay line is evaluated at 0.5 ps increments along the line, with 25 points per input time. The top plot is of static gain error, and the bottom is of single-shot precision, and both are the worst case observed across 50 separate runs. From this we can observe that the peak worst case precision we can expect along the converter is 10.1 ps.

Figure 4.9 is a plot of this same data, however zooming in time scale to the last dozen elements in the line. This allows us to observe some of the nuanced in the curves of precision and gain error. Since noise is enabled, jitter in the system causes some variation around the tripping points. This is because noise has the potential to add enough error to push a value into an adjacent quantization interval. The higher the noise, the more this dithering effect is observed. In the case of our TDC, even with 3 mV RMS supply noise enabled, it is clear that the system is quantization noise dominated, as there are input time intervals with no uncertainty as to how they will evaluate. Also interesting to not is the difference in peak quantization error from interval. This is a function of non-linearity randomly shifting each tripping point.



Figure 4.8: Single shot precision and mean error Figure 4.9: TDC single-shot precision at end of range

The results from the above simulations, the layout, and other work are summarized in Table 4.2 below where the expected performance of the TDC is compared against its design specifications. All requirements are met or exceeded.

Figure of Merit	Requirement	Simulated Performance
Dynamic Range	$> 25\mathrm{ns}$	$64\mathrm{ns}$
Effective Timing Precision	$< 23\mathrm{ps}\;\mathrm{RMS}$	$10.1\mathrm{ps}\;\mathrm{RMS}$
Supply Voltage	$0.9\mathrm{V}$	$0.9\mathrm{V}$
Tolerable Supply Noise	$3\mathrm{mV}\ \mathrm{RMS}$	$3\mathrm{mV}\ \mathrm{RMS}$
Temperature	$-10^{\rm o}{\rm C}$ to $50^{\rm o}{\rm C}$	$-10^{\circ}\mathrm{C}$ to $50^{\circ}\mathrm{C}$
Radiation Hardness	$1000\mathrm{Mrad}$	$1000\mathrm{Mrad}$
Silicon Area	$< 60\mu\mathrm{m~x~500\mu\mathrm{m}}$	$50\mu\mathrm{m} \ge 475\mu\mathrm{m}$
Power Consumption	No Limit	$34.6\mathrm{mA}$ peak

Table 4.2: Simulated time-to-digital converter circuit performance, versus design specifications

Chapter 5: System Integration and Testing

5.1 System Level Integration

The next stage of the project is integrate these two test fixture blocks as part of full chip system, alongside supporting circuitry and the AFE under test. A planned top-level layout for the chip is shown in Figure 5.1. The chip area, pre-shrink is designed to be a square 2.2 mm by 2.2 mm. During the fabrication stage, this design will be scaled smaller by a factor of 0.9, such that the final chip is 2.0 mm by 2.0 mm.

Each analog front end will be paired with a charge injection circuit and TDC. The heights of each are all approximately equal, and so vertical columns of channels can be constructed, with roughly 15 channels per side of the chip, for a total of 30 channels. Each channel will need a pad to support a buffered version of the TOT signal for debugging and additional testing purposes. The chip will have an estimated 68 wire bond pads available, leaving plenty of space for the control and outputs signals, including prime, clock, channel select, charge injection CDAC configuration, PISO select, PISO clock, and PISO output. Channel select and CDAC configuration inputs will likely all share the same serial interface. The remaining pins are used for power supply and bias lines.

The initial chip GDS^1 submission is January 12th, and the final submission is February 2nd, with an expected delivery in early April 2022.

¹GDS is a database file format which is the de facto industry standard for data exchange of integrated circuit layout artwork.



Figure 5.1: Floorplan of the BigRock testbed ASIC.

5.2 System Testing

While waiting for chip delivery, efforts will switch to supporting hardware and hardware description language (HDL) development. An external FPGA module will be used to control the operation of the test integrated circuit. Before each measurement, a reset step will be performed, where the injection trigger and delay line flip-flop will be reset. After being primed, the FPGA will supply the 1 GHz reference signal to the input of the test chip. After a set wait time, it can be assumed that the TDC has performed a measurement, and that the result is available in the flip-flops connected to each tap of the delay line. The FPGA will then use digital controls to load this data in parallel into the PISO register; which can then be read out at an arbitrary speed by cycling the clock of the shift register. The first couple cycles of measurement are used for calibration only, where an external power supply is adjusted to tune the average delay of unit cells in the line. After the calibration phase, jitter measurements read out by the FPGA will be saved for later analysis. Design of the FPGA controller will be performed using SystemVerilog as the hardware description language of choice.

In addition to the FPGA controller, a printed circuit board will be designed to carry the test chip and interconnect it with the FPGA and other external equipment. A custom footprint will be designed to match the pad ring of the chip, allowing for wire bonding. The printed circuit board will likely also carry some supporting ICs, including level-shifters to convert the FPGA square wave and control signals for compatibility with the 0.9 V supply of the test chip, and linear voltage regulators for reduced power supply noise. Addition functionality will be likely be added as needs arise throughout the project.

Once fabricated chips are received, the first step will be to characterize the test circuits. This will be done by a couple channels on the chip that will not be populated with an AFE. With no AFE, the adjacent pin can be used to measure charge pulse signals from the injection circuit, or to input arbitrary time intervals to TDC circuits. Will be necessary to verify the accuracy and precision of these two converters match the values expected by simulation, before they can then be used in turn to characterize the AFEs themselves.

5.3 Conclusion

Referencing Figure 2.8, it is calculated that the excellent timing precision performance of the testbed (dominated by the TDC precision) will yield measurements of AFE timing characteristics with accuracy better than 95%. Dynamic range requirements for both timing and charge injection are also met. In all over 30+ unique base-level cells were constructed in 28nm CMOS in the process of designing charge injection and TDC circuits. This test circuit will allow for automated, high precision testing of AFEs in 28 nm, yielding vast improvements over older testing methodologies.

Bibliography

- L. Rossi, P. Fischer, T. Rohe, and N. Wermes, *Pixel Detectors: From Fundamentals to Applications*. Particle Acceleration and Detection, Berlin, Heidelberg: Springer, 2006.
- [2] N. Wermes, "Pixel detectors ... where do we stand?," Nucl. Instrum. Methods Phys. Res. A, vol. 924, pp. 44–50, Apr. 2019.
- [3] M. Garcia-Sciveres and N. Wermes, "A review of advances in pixel detectors for experiments with high rate and radiation," *Rep. Prog. Phys.*, vol. 81, p. 066101, May 2018.
- [4] "RD53B users guide," tech. rep., RD53 Collaboration, Dec. 2020.
- [5] N. Cartiglia, R. Arcidiacono, A. Bellora, F. Cenna, R. Cirio, S. Durando, M. Ferrero, P. Freeman, Z. Galloway, B. Gruey, M. Mashayekhi, M. Mandurrino, V. Monaco, R. Mulargia, M. Obertino, F. Ravera, R. Sacchi, H. F.-W. Sadrozinski, A. Seiden, V. Sola, N. Spencer, A. Staiano, M. Wilder, N. Woods, and A. Zatserklyaniy, "The 4D pixel challenge," J. Inst., vol. 11, pp. C12016–C12016, Dec. 2016.
- [6] W. Riegler and G. A. Rinella, "Time resolution of silicon pixel sensors," J. Inst., vol. 12, pp. P11017– P11017, Nov. 2017.
- [7] R. Mendicino, G. T. Forcolin, M. Boscardin, F. Ficorella, A. Lai, A. Loi, S. Ronchin, S. Vecchi, and G.-F. Dalla Betta, "3D trenched-electrode sensors for charged particle tracking and timing," *Nucl. Instrum. Methods Phys. Res. A*, vol. 927, pp. 24–30, May 2019.
- [8] L. Anderlini, M. Aresti, A. Bizzeti, M. Boscardin, A. Cardini, G.-F. D. Betta, M. Ferrero, G. Forcolin, M. Garau, A. Lai, A. Lampis, A. Loi, C. Lucarelli, R. Mendicino, R. Mulargia, M. Obertino, E. Robutti, S. Ronchin, M. Ruspa, and S. Vecchi, "Intrinsic time resolution of 3D-trench silicon pixels for charged particle detection," J. Inst., vol. 15, pp. P09029–P09029, Sept. 2020.
- [9] A. Zjajo, Stochastic Process Variation in Deep-Submicron CMOS. Springer Series in Advanced Microelectronics, Dordrecht: Springer Netherlands, 2014.

- [10] S. Henzler, Time-to-Digital Converters, vol. 29 of Springer Series in Advanced Microelectronics. Dordrecht: Springer Netherlands, 2010.
- [11] S. Henzler, S. Koeppe, W. Kamp, H. Mulatz, and D. Schmitt-Landsiedel, "90nm 4.7ps-Resolution 0.7-LSB Single-Shot Precision and 19pJ-per-Shot Local Passive Interpolation Time-to-Digital Converter with On-Chip Characterization," in 2008 IEEE International Solid-State Circuits Conference - Digest of Technical Papers, pp. 548–635, Feb. 2008.
- [12] H. Huang, A 0.1 Ps Resolution Coarse-Fine Time-to-Digital Converter with 2.21 Ps Single-Shot Precision. PhD thesis, May 2018.
- [13] M. Lee and A. A. Abidi, "A 9-bit, 1.25 ps Resolution Coarse–Fine Time-to-Digital Converter in 90 nm CMOS that Amplifies a Time Residue," *IEEE Journal of Solid-State Circuits*, vol. 43, pp. 769–777, Apr. 2008.
- [14] A. Rivetti, CMOS: Front-End Electronics for Radiation Sensors. Devices, Circuits, and Systems, CRC Press, 1st ed., 2017.
- [15] S. Mattiazzo, M. Bagatin, D. Bisello, S. Gerardin, A. Marchioro, A. Paccagnella, D. Pantano, A. Pezzotta, C.-M. Zhang, and A. Baschirotto, "Total Ionizing Dose effects on a 28 nm Hi-K metal-gate CMOS technology up to 1 Grad," J. Inst., vol. 12, pp. C02003–C02003, Feb. 2017.
- [16] C.-M. Zhang, F. Jazaeri, A. Pezzotta, C. Bruschini, G. Borghello, F. Faccio, S. Mattiazzo, A. Baschirotto, and C. Enz, "Characterization of GigaRad Total Ionizing Dose and Annealing Effects on 28-nm Bulk MOSFETs," *IEEE Transactions on Nuclear Science*, vol. 64, pp. 2639–2647, Oct. 2017.
- [17] F. Faccio, "Radiation effects in the electronics for CMS," 2017.
- [18] F. Maloberti, Data Converters. Dordrecht, Netherlands: Springer, 2007.
- B. Razavi, High-Speed, High-Resolution Analog-to-Digital Conversion in VLSI Technologies. PhD thesis, Stanford University, United States – California, 1991.
- [20] J. Christiansen, "Picosecond Stopwatches: The Evolution of Time-to-Digital Converters," IEEE Solid-State Circuits Magazine, vol. 4, no. 3, pp. 55–59, 2012.
- [21] N. U. Andersson and M. Vesterbacka, "A Vernier Time-to-Digital Converter With Delay Latch Chain Architecture," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 61, pp. 773–777, Oct. 2014.

- [22] P. Dudek, S. Szczepanski, and J. V. Hatfield, "A high-resolution CMOS time-to-digital converter utilizing a Vernier delay line," *IEEE Journal of Solid-State Circuits*, vol. 35, pp. 240–247, Feb. 2000.
- [23] C.-H. Wu, S.-Y. Huang, Y.-F. Chou, and D.-M. Kwai, "Time-to-Digital Converter Compiler for On-Chip Instrumentation," *IEEE Design Test*, vol. 37, pp. 101–107, Aug. 2020.
- [24] S. Hwang, J. Koo, K. Kim, H. Lee, and C. Kim, "A 0.008 mm2 500 uW 469 kS/s Frequency-to-Digital Converter Based CMOS Temperature Sensor With Process Variation Compensation," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 60, pp. 2241–2248, Sept. 2013.
- [25] X. Zhang, J. Acharya, and A. Basu, "A 0.11-0.38 pJ/cycle Differential Ring Oscillator in 65 nm CMOS for Robust Neurocomputing," Nov. 2020.
- [26] M. Moazedi, A. Abrishamifar, and A. M. Sodagar, "A highly-linear modified pseudo-differential current starved delay element with wide tuning range," in 2011 19th Iranian Conference on Electrical Engineering, pp. 1–4, May 2011.
- [27] B. Markovic, S. Tisa, F. A. Villa, A. Tosi, and F. Zappa, "A High-Linearity, 17 ps Precision Time-to-Digital Converter Based on a Single-Stage Vernier Delay Loop Fine Interpolation," *IEEE Transactions* on Circuits and Systems I: Regular Papers, vol. 60, pp. 557–569, Mar. 2013.
- [28] Y. Toh and J. A. McNeill, "Single-ended to differential converter for multiple-stage single-ended ring oscillators," *IEEE Journal of Solid-State Circuits*, vol. 38, pp. 141–145, Jan. 2003.
- [29] B. K. Davis, "Low-skew single-ended to differential converter," Oct. 2006.