

© Copyright by Jeremy Todd Townsend 2019

All Rights Reserved

FPGA-based Data Acquisition System for SiPM Detectors

A Thesis

Presented to

the Faculty of the Department of Electrical and Computer Engineering

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

in Electrical Engineering

by

Jeremy Todd Townsend

May 2019

FPGA-based Data Acquisition System for SiPM Detectors

Jeremy Todd Townsend

Approved:

Chair of the Committee
Dr. Jinghong Chen, Associate Professor,
Electrical and Computer Engineering

Committee Members:

Dr. Xin Fu, Associate Professor,
Electrical and Computer Engineering

Dr. Jiming Peng, Associate Professor,
Industrial Engineering

Dr. Suresh K. Khator, Associate Dean,
Cullen College of Engineering

Dr. Badri Roysam, Professor and Chair
Electrical and Computer Engineering

Acknowledgments

First and foremost, I would like to thank the Lord and Savior Jesus Christ for His strength and support throughout this journey. He gave me a dream to become the best engineer I can be, and His plan is always greater than I can imagine. To my wife Jennifer, I love you more than you will ever know, and I thank you for your constant push and belief in me. Thank you for being a Super Mom through this time which I will always strive to repay. To my children, Hailey, Tyler, and Jacob, thank you for your understanding while Dad couldn't spend as much time with you. I promise to help you more with your homework, read you more bedtime stories, and play video games and play catch as much as you want from now on. To my parents Jerry and Debbie, thank you for instilling in me the will to succeed and an example to follow. Your prayers were felt and always with me. To my graduate advisor, Dr. Jinghong Chen, words cannot express my gratitude for your continued support and advice since 2015. You have always given me such encouragement and helped me with the whole process. I pray you will have a constant flow of ideas and research opportunities for many years to come. Thank you, Yuxuan Tang, for your help in explaining SiPM detector technology and for developing an ASIC to use as inspiration for my design. Thank you Dr. Xin Fu and Dr. Jiming Peng for your support and consideration in my thesis review. Thank you, Dr. Steven Morris, for pushing me to obtain my graduate degree and thank you and Chaitanya Vempati for support as my managers and understanding of my time away from work. Thank you to the Xilinx University Program for providing an evaluation board free of charge and last, but not least, thank you to the University of Houston and to the Electrical and Computer Engineering department for giving me this opportunity.

FPGA-based Data Acquisition System for SiPM Detectors

An Abstract

of a

Thesis

Presented to

the Faculty of the Department of Electrical and Computer Engineering

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

in Electrical Engineering

by

Jeremy Todd Townsend

May 2019

Abstract

Positron Emission Tomography is a growing field with an increasing influence in nuclear medicine. Recent Time-of-Flight advancements have increased the cross-sectional image clarity and progression towards a miniature, silicon-based, photomultiplier has increased the resolution. As the photomultiplier size reduces, channel-count increases. Use of ASIC based Time-to-Digital converters with remote event processing is no longer viable due to the increasing cost and complexity. Instead, multichannel, FPGA based measurements with localized event processing is a must. Within this work, an FPGA based SiPM DAQ is presented with emphasis regarding the application of PET imaging. An introduction concerning PET concepts is included with a brief description of measurement techniques followed by a review of previous work with a detailed explanation of relevant methods. Next, an analysis of the problem covers specific implementation details and difficulties followed by a comparison of results to previous work. Use of a 28nm, Kintex 7 FPGA based DAQ is examined with implementation of a 32 channel, multichain averaged TDC with 15ps RMS resolution, 11ps average bin size, and less than 10ps of integral nonlinearity.

Table of Contents

Acknowledgments	v
Abstract	vii
Table of Contents.....	viii
List of Figures.....	x
List of Tables	xiii
Nomenclature	xiv
1. Introduction	1
1.1 Positron Emission Tomography.....	3
1.1.1 Theoretical Background.....	3
1.1.2 Measurement and Processing.....	9
1.2 Scope and Objectives	14
2. Review of Previous Work.....	15
2.1 Introduction.....	15
2.2 Time to Digital Converter	15
2.2.1 Analog TDC.....	16
2.2.2 Digital TDC	17
2.2.3 Performance Characterization.....	20
2.2.4 FPGA-based Digital TDC.....	25
2.3 Pulse Energy Measurement.....	28
2.4 Inter-DAQ and DAQ to PC Communications Link.....	29

3. Analysis of the Problem	34
3.1 Arrival Time Measurement.....	35
3.1.1 FPGA Selection	36
3.1.2 Maximum System Clock Frequency.....	38
3.1.3 Average Delay Time	39
3.1.4 Thermometer to Binary Encoder	43
3.1.5 Pulse Detection and Latching	48
3.1.6 Linearity Improvement	49
3.1.7 Calibration.....	54
3.1.8 Multichannel Implementation.....	55
3.2 Pulse Energy Measurement.....	56
3.3 High Speed Transfer of Events.....	58
4. Results and Discussion	61
5. Summary and Conclusions	64
5.1 Summary	64
5.2 Conclusion	65
References	66

List of Figures

Figure 1:1 – Example PET scans: (a) radiotracer imaging of tumors and (b) imaging of glucose metabolism within the brain [1].	2
Figure 1:2 – Targeting individual receptors with specific radiopharmaceuticals [1].	3
Figure 1:3 – Positron emission and annihilation [3].	4
Figure 1:4 – 3D gantry array and LOR.	5
Figure 1:5 – Coincidence detection [3].	6
Figure 1:6 – Time-of-Flight [6].	7
Figure 1:7 – Gain in sensitivity as defined by thickness $D/\Delta x$ [8].	8
Figure 1:8 – Multichannel SiPM FPGA-based DAQ.	9
Figure 1:9 – PMT structure [8].	11
Figure 1:10 – SiPM structure [5].	12
Figure 1:11 – SiPM array [5].	12
Figure 2:1 – Basic analog TDC (block and signal diagram) [10].	16
Figure 2:2 – Pure TDL-based TDC with buffer delay elements [10].	17
Figure 2:3 – Dual channel TDL-based TDC with inverter delay elements [10].	18
Figure 2:4 – Vernier TDL-based TDC.	18
Figure 2:5 – TDC coarse counter with start and stop error [10].	19
Figure 2:6 – Example measured vs. expected plot.	20
Figure 2:7 – Code density test for single channel.	21
Figure 2:8 – Code density example distribution.	22
Figure 2:9 – Example differential nonlinearity plot.	23
Figure 2:10 – Example integral nonlinearity plot.	24

Figure 2:11 – RMS measurement with fixed cable delay.	25
Figure 2:12 – CARRY4 (4 delay tap) Xilinx Series 7 [27].	27
Figure 2:13 – Energy measurement (post-sample digital integration).....	29
Figure 2:14 – Energy measurement (pre-sample analog integration).....	29
Figure 2:15 – Tree data path DAQ structure.	33
Figure 2:16 - Daisy-chain path DAQ structure.....	34
Figure 2:17 - Bus path DAQ structure.	34
Figure 3:1 – Xilinx KC705 evaluation board [31].....	38
Figure 3:2 – 7 Series SLICEL with registered tapped delay line.	40
Figure 3:3 – Initial single channel TDC.	40
Figure 3:4 – Initial TDC bin count (bin latch at state count = 5).....	41
Figure 3:5 – TDC bin count (bin latch at state count = 12).	42
Figure 3:6 – The Kintex 7 XC7K325T architecture with 14 clock regions.....	43
Figure 3:7 – Bubble error.....	44
Figure 3:8 – Clock skew in Kintex 7 SLICEL.....	45
Figure 3:9 – Post implementation timing simulation for average CARRY4 bins.	46
Figure 3:10 – TDC after bin realignment.	47
Figure 3:11 – Zoomed in view of TDC after bin realignment.....	47
Figure 3:12 – Typical pulse detector.	48
Figure 3:13 – Fast acting pulse detector with filtered output pulse.....	49
Figure 3:14 – 500MHz initial TDC with fast-acting pulse detector.	50
Figure 3:15 – 1 Chain, 2 Chain, and 4 Chain multichain histogram.	51
Figure 3:16 – 1 Chain, 2 Chain, and 4 Chain multichain DNL.	52

Figure 3:17 – 1 Chain, 2 Chain, and 4 Chain multichain INL.	53
Figure 3:18 – 1 Chain and 4 Chain RMS comparison.	54
Figure 3:19 – 2 Chain RMS result.	54
Figure 3:20 – Enlarged view of two channels with registers, latches, and encoders.	56
Figure 3:21 – UH designed SiPM detector.	57
Figure 3:22 – Latched ADC measurement.	58
Figure 4:1 – Final design.	62
Figure 4:2 – Post-calibrated 4 Chain DNL.	63
Figure 4:3 – Post-calibrated 4 Chain INL.	63

List of Tables

Table 2:1 – Previous TDC work.	26
Table 2:2 – Typical PET system specifications [4].	30
Table 2:3 – Typical key parameters of a PET DAQ system [4].	32
Table 3:1 – Kintex 7 and Virtex 7 F_{MAX}	39
Table 3:2 – Pre-calibrated 1 Chain, 2 Chain, and 4 Chain DNL/INL.	52
Table 3:3 – Proposed PET requirements.	59
Table 3:4 – PET single event packet.....	59
Table 3:5 – PET coincidence event packet.	60
Table 3:6 – Required bus speeds for single and coincidence transmission.	60
Table 3:7 – Common bus protocols and reference with KC705 availability.....	61
Table 4:1 – Comparison to previous work.....	64

Nomenclature

SiPM – Silicon Photomultiplier

PET – Positron Emission Tomography

TOF – Time-of-Flight

RMS – Root Mean Square

ADC – Analog to Digital Converter

TDC – Time-to-Digital Converter

DAQ – Data Acquisition

APD – Avalanche Photodiode

SNR – Signal to Noise Ratio

LOR – Line of Response

FWHM – Full Width Half Maximum

PVT – Process Voltage Temperature

LSB – Least Significant Bit

TDL – Tapped Delay Line

1. Introduction

An insurgence of silicon-based photomultipliers (SiPM) has driven the necessity of multichannel data acquisition (DAQ) systems with compact local processing capability. As a replacement to traditional photomultiplier tubes (PMT) in fields requiring smaller packaging and high-density arrays, SiPM technology has proven to be equivalent in performance including gain and noise. One field of interest is that of Positron Emission Tomography (PET) due to its rising influence in nuclear medicine and greater capabilities with advancements in Time-of-Flight (TOF) technology. Since the 1920s, nuclear medicine has been advancing with early research involving radionuclides and gamma ray counters. In 1953, Gordon Brownell from MIT constructed the first detector device to exploit positron-electron annihilation as a part of an imaging tool. This device was the precursor to modern PET scanners. In 1974, the first PET camera, PETT III, built for human studies was designed to use advanced algorithms for computing three-dimensional images. By the late 1980s, PET scanners had become a mainstay within hospitals and universities. PET scanners can image radiotracers traveling to tumor sites, patterns of glucose metabolism related to different mental tasks, areas of reduced blood flow contributing to atherosclerosis and cardiovascular disease, and even provide early diagnosis of Alzheimer's disease [1]. See Figure 1:1.

Present day PET scanners include 3D sensor gantry arrays and dynamic time sampling for 4D imaging. In order to accurately measure positron annihilation events spatially throughout any patient size, TOF measurements are required. Time-to-Digital Converters (TDC) are utilized for TOF measurement where resolution and accuracy are directly related to PET image clarity. TDCs can be implemented by costly application-specific integrated

circuit (ASIC) designs or more recently by cheaper field programmable gate array (FPGA) methods. The following research focuses on FPGA data acquisition for use with SiPM detector-based PET imaging systems. Organization is as follows:

- Chapter 1: Introduction and overview of PET imaging
- Chapter 2: A review of previous work including scope and objectives
- Chapter 3: Analysis of the problem
- Chapter 4: Results and discussion
- Chapter 5: Summary and conclusion

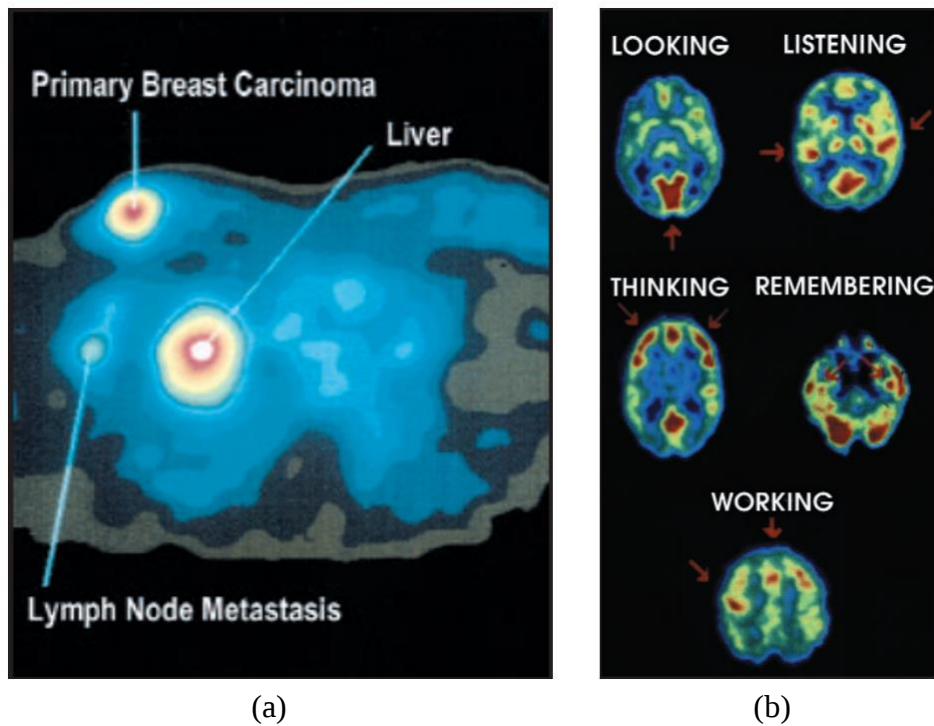


Figure 1:1 – Example PET scans: (a) radiotracer imaging of tumors and (b) imaging of glucose metabolism within the brain [1].

1.1 Positron Emission Tomography

1.1.1 Theoretical Background

PET is an imaging technique which captures pairs of gamma rays emitted from the interaction of a radionuclide tracer with biological tissue. Typically, PET scans are performed in vivo with live tissue. Different types of radionuclides exist with varying half-lives and absorption rates which prove useful for targeted applications and tissues. See Figure 1:2. Radionuclide tracers are designed to be biologically active molecules which act like compounds normally consumed by the body like water, glucose, or ammonia. Fluorine-18 fluorodeoxyglucose (F-18 FDG), a form of glucose sugar, is utilized in 90% of all PET scanning in the United States due to its long half-life of 110 minutes [2]. F-18 FDG can be commercially manufactured offsite and shipped to location for immediate use [2].

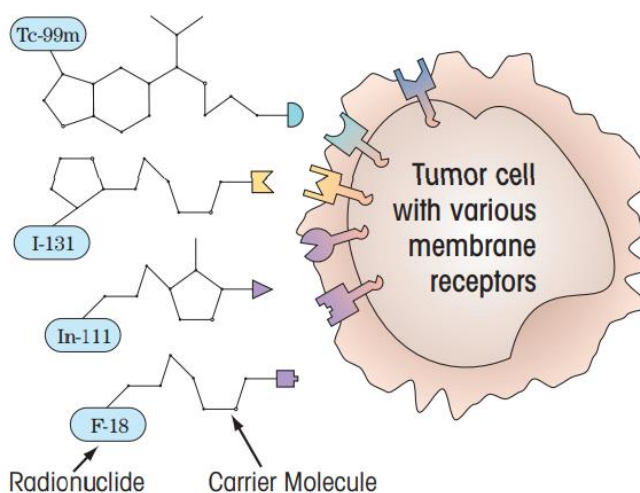


Figure 1:2 – Targeting individual receptors with specific radiopharmaceuticals [1].

1.1.1.1 Annihilation Event

Proton-rich radionuclides decay by positron emission. This process involves the decay of a proton within the nucleus to a neutron, positron, and a neutrino as seen in

$$p \rightarrow n + e^+ + \nu_e, \quad (1.1)$$

and with the decay written as

$$(Z, A) \rightarrow (Z - 1, A) + e^+ + \nu_e. \quad (1.2)$$

The released positron travels through live tissue approximately 0.5 to 1mm before giving up its kinetic energy when interacting with electrons. This sudden release in kinetic energy is called annihilation. Annihilation produces two 511 keV gamma photons travelling 180° apart in an anti-parallel path [3]. See Figure 1:3. It is possible for one or both released photons to scatter which changes the trajectory angle. The longer a gamma ray travels through tissue, the higher the likelihood of scattering. This phenomenon is called Compton scattering and it is the highest contributing factor to reduced spatial resolution in PET imaging [3].

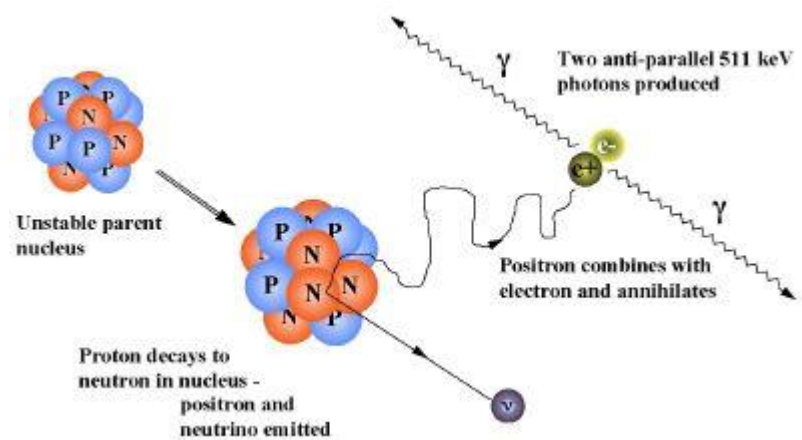


Figure 1:3 – Positron emission and annihilation [3].

1.1.1.2 Multichannel 3D Gantry Array

A 3D array of detector/amplifier pairs will capture annihilation events over a volume equal to $Length \cdot \pi \cdot radius^2$. As the axial length of the PET chamber is increased, more of the annihilation events are captured, increasing sensitivity. It is desirable to capture as many events as possible to increase confidence in event location. At the moment of annihilation, the anti-parallel gamma rays travel along a line of response (LOR). See Figure 1:4. The time between detection of a matching pair of annihilation gamma rays is very small and typically less than 6 ns [4]. This feeds into the coarse time requirement of a TOF TDC.

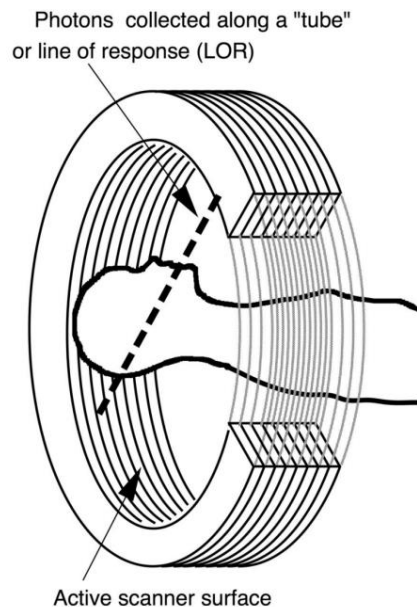


Figure 1:4 – 3D gantry array and LOR [5].

1.1.1.3 Coincidence

As annihilation events occur, each channel of a PET detects event pulses. If two independent, 180° opposing, channels detect pulses within a coincidence window (~6ns), a coincidence is confirmed and an LOR is recorded. Figure 1:5 illustrates this concept.

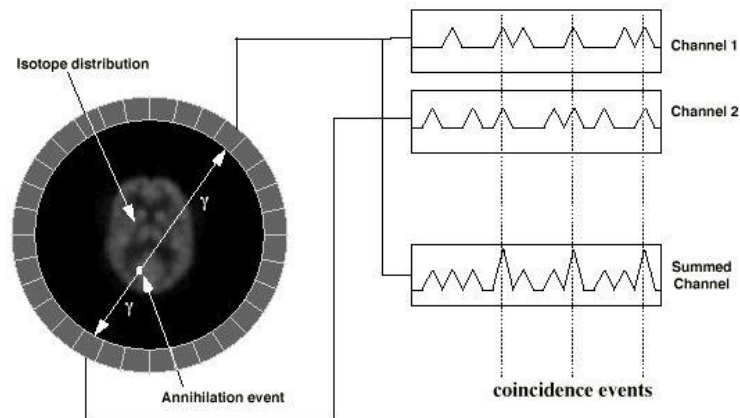


Figure 1:5 – Coincidence detection [3].

In addition to standard LOR detection, coincidence may be incorrectly determined due to random events falling within a coincident window or because of Compton scattering. Incorrect detection acts as a noise floor to true coincidence detection. Unfortunately, Compton scattering increases as test subjects become larger. This means that PET image clarity and spatial resolution is inversely related to patient size. In the late 1980s, researchers developed the first PET system to combat the effects of Compton scattering. This new PET system involved the use of precise time measuring devices to quantify the Time-of-Flight for annihilation events [6].

1.1.1.4 Time of Flight

For increased spatial resolution throughout an entire patient cross-section, measurement of the time difference between coincident events is utilized within TOF PET. Along a line of response, the corresponding annihilation event can be located relative to the midpoint of the LOR. Figure 1:6 shows the LOR and an uncertainty distribution centered at the location of annihilation.

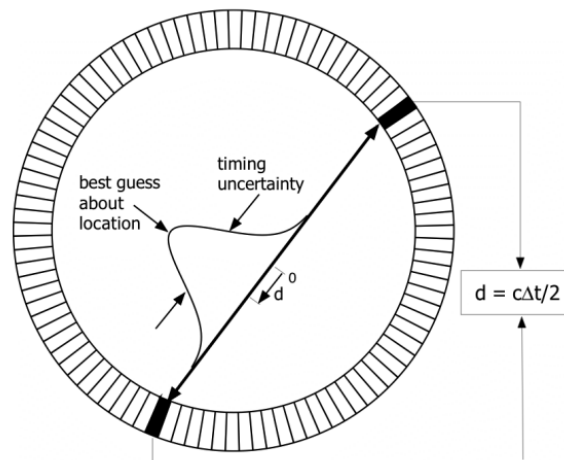


Figure 1:6 – Time-of-Flight [5].

A formula for the distance can be seen here

$$d = \frac{c\Delta t}{2}, \quad (1.3)$$

where the distance d can be calculated with c being the speed of light in a vacuum and Δt , the difference in time measured between matched coincident events. Statistically, as more events are detected from the same location, a better guess can be determined for the location. In addition, the energy of the event is averaged to produce a better approximation for the pixel intensity for the PET image. Traditional PET scanners use complex matrix

algorithms with each LOR contributing to the overall image. Location uncertainty builds as Compton scattering increases towards the center of large test subjects. TOF measurements do not eliminate the need for these algorithms, but instead feed them with additional spatial information which keeps the resolution constant throughout the cross-section [6]. Figure 1:7 illustrates the increase in sensitivity as the overall timing resolution decreases. Notice as the test sample thickness D increases, timing resolution has the highest impact on sensitivity gain maximizing around 100ps [7]. One requirement for TOF PET is for a fast time response of the scintillating material. In the late 1980s, the only options with fast time response were limited in gain. Eventually, slow, high-gain materials with traditional PET edged out TOF PET with better images and tumor detection. With recent advancements in scintillating materials, fast, high-gain options are now available. Because of this, there has been a resurgence in TOF PET research and commercial development. TOF PET is now again the best option for maximum clarity with short scan times for all patient sizes [7].

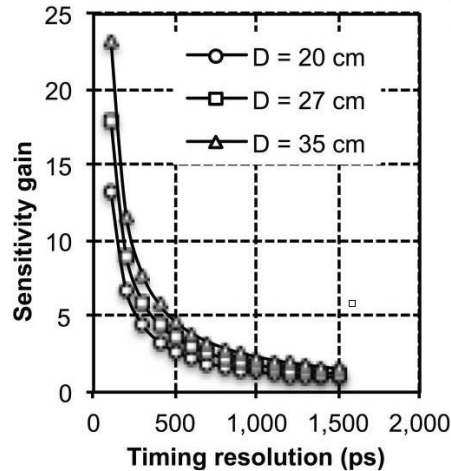


Figure 1:7 – Gain in sensitivity as defined by thickness $D/\Delta x$ [7].

Traditional TOF measurement is accomplished by ASIC TDC converters. As PET imaging becomes more complex with higher and higher quantities of SiPM detectors, ASIC TDCs can be cost prohibitive and burdensome from a pcb layout perspective. As fabrication technology becomes more advanced with smaller cell size and power requirements, FPGA based TDCs are now a more economical and convenient solution. As many as 16 to 32 TDCs are possible within the fabric of one FPGA along with room for local coincidence processing.

1.1.2 Measurement and Processing

1.1.2.1 Data Acquisition

A data acquisition (DAQ) unit is required for each channel of a PET scanner and each DAQ supports three primary functions: Detection, Measurement, and Coincidence Processing. See Figure 1:8.

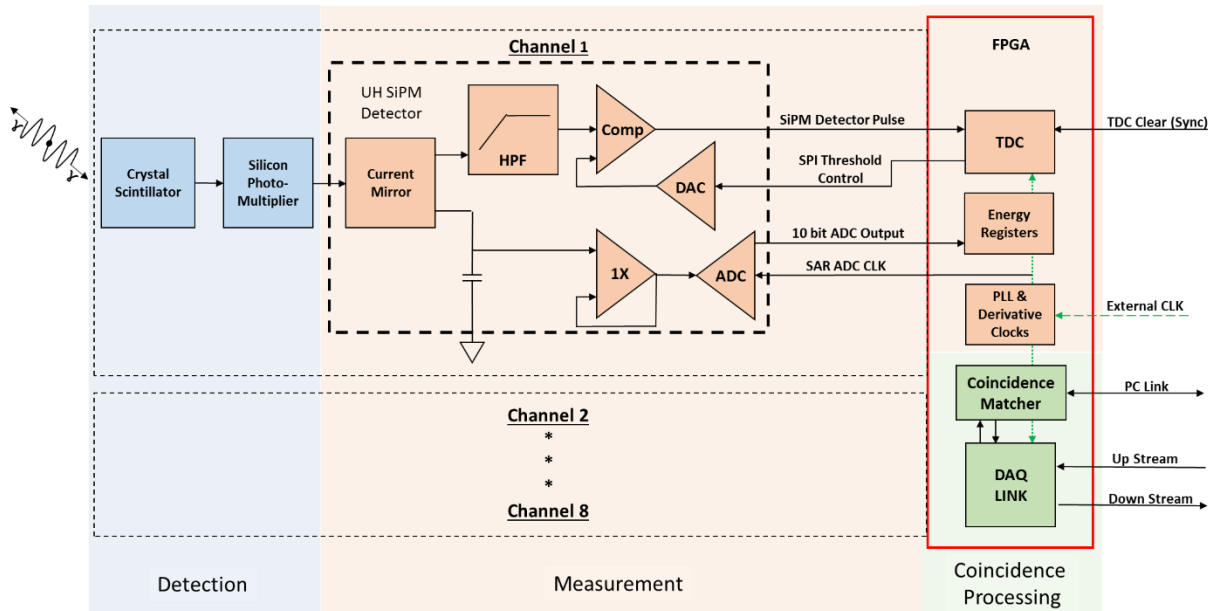


Figure 1:8 – Multichannel SiPM FPGA-based DAQ.

1.1.2.2 *Detection*

Scintillation

Detection of the 511 keV photons produced during annihilation is accomplished by scintillation and photo-multiplication. When nuclear radiation hits scintillating material, scintillation occurs and emits light within the ultra violet to infrared range (100 – 800 nm wavelength). The most common scintillating material for PET application is $\text{Bi}_4\text{Ge}_3\text{O}_{12}$ or Bismuth Germanate due to its high gain and high stopping power which increases efficiency and cost effectiveness. Newer Time-of-Flight PET scanners utilize BaF_2 (Barium Fluoride) which has a faster time response that maximizes TOF accuracy for finer resolution imaging. BaF_2 has both fast and slow light components. A description of the light emission and the number of photons emitted can be represented here with

$$N(t) = A\exp\left(\frac{-t}{\tau_f}\right) + B\exp\left(\frac{-t}{\tau_s}\right), \quad (1.4)$$

where N is the number of photons emitted in the visible spectrum at time t and τ_f and τ_s are the fast and slow time constants with A and B representing the total number of photons emitted per respective fast or slow time constant. Scintillation effectively converts the energy from annihilation photons to light which can be amplified and measured.

Amplification

Amplification is performed by photomultiplication either by a photomultiplier tube (PMT) or by a silicon photomultiplier (SiPM). Figure 1:9 illustrates operation of a PMT. First, as the 511 keV gamma ray passes through the scintillation material, photons of light are produced which are then converted to electrons via a photo-cathode and the

photoelectric effect. A high-voltage power supply charges a cross pattern of dynodes with increasing voltage from cathode to anode. The impact of the first electron with the first dynode releases a higher number of secondary electrons. These secondary electrons are then accelerated and amplified at each subsequent dynode stage until impact with the PMT anode.

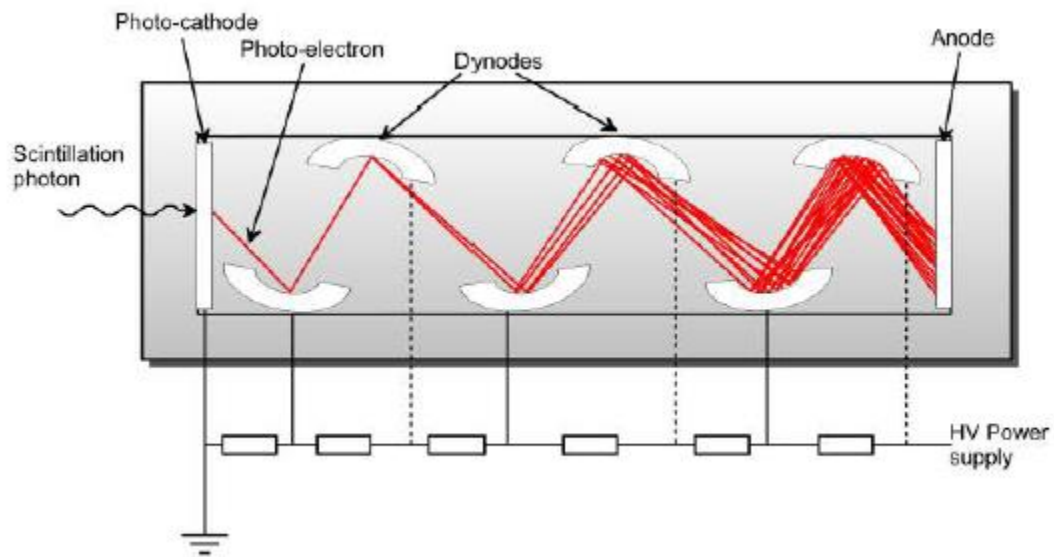


Figure 1:9 – PMT structure [8].

SiPM amplification offers a smaller, less complex alternative with comparable gain and noise. Operation is illustrated in Figure 1:10 and Figure 1:11. Conversion of light energy is handled by a special high-gain avalanche photodiode (APD) instead of photo-cathodes. SiPM diodes operate with a low bias voltage set above the diode breakdown voltage. When a single photon is detected at the cathode, avalanche breakdown is triggered [9]. An array of SiPM microcells are connected in parallel to reduce the complex impedance seen by the current mirror with an active area typically from $1.3 \times 1.3 \text{ mm}^2$ to $6.5 \times 6.5 \text{ mm}^2$ [8], [9]. As the array size increases, the SiPM is more susceptible to dark

current noise. Several distinct advantages exist for SiPM over PMT. The SiPM is much smaller in size due to the fact it does not need a vacuum tube for operation and is immune to both electric and magnetic field influence. Magnetic field immunity makes SiPM particularly useful for hybrid MRI-PET scanning applications. Before SiPMs, PMTs had to be placed outside of the magnetic field which reduced signal strength. For this reason, SiPMs offer an increased signal to noise ratio (SNR) vs PMT amplification even with the same gain [8].

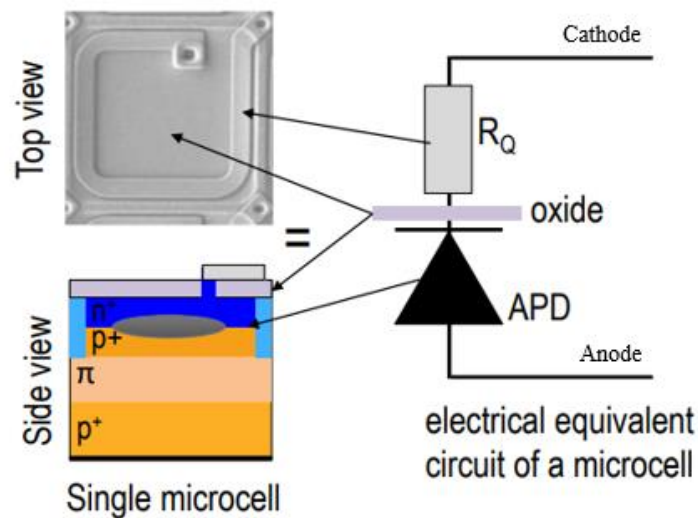


Figure 1:10 – SiPM structure [10].

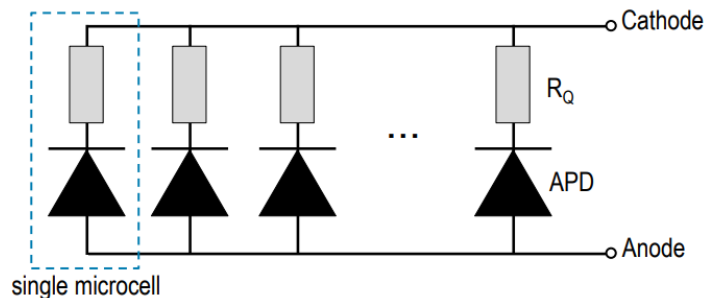


Figure 1:11 – SiPM array [10].

With SiPM amplifiers, PET scanners are now able to pack more and more individual detector/amplifier pairs in the same volume allowing for increased spatial resolution. As more detectors are employed, centralized processing becomes increasingly difficult. FPGAs offer localized processing and reduced bandwidth for connection to an imaging PC.

1.1.2.3 *Measurement*

Analog Front End

The current pulse arriving at the anode of the amplifier is fed through a current mirror and distributed with two taps. The current first travels to a high-pass filter and second to a precise, temperature stable, capacitor. The high-pass filter cuts out any DC offset before comparing the output to a variable threshold for low-energy pulse filtering. Silicon photomultipliers operate with a high electric field which generates random electrons and holes within the diode depletion region. This noise is called dark noise and can trigger a false detection if the comparator threshold isn't set high enough [9]. A second tap feeds a capacitor which integrates the current pulse and provides a measurable voltage proportional to the energy of photons arriving at the cathode.

Analog Back End

If the energy of a pulse is high enough, an analog-to-digital converter (ADC) samples the voltage and feeds it back to a processing unit. A precise TDC measures and records a coarse and fine time stamp for each event. The required timing precision of a TDC must be less than the scintillator time constant. The Full Width Half Maximum (FWHM) of a TOF PET is dependent on both the scintillating material and the TDC.

1.1.2.4 Coincidence Processing

Coincidence Matcher

Once a coincidence is detected and the pulse energy and timing have been captured, the timing must be compared to coincidences from the opposite side of the scanner array. A matched pair is discovered if both coincidences fall within the same coincident window (TDC coarse time) and are geometrically 180° apart. This process can be handled remotely by an image reconstruction processor or locally by a master DAQ FPGA.

DAQ Communications Link

Each DAQ must pass coincidence information to the image reconstruction processor. This can be accomplished by linking each DAQ directly to the processor or by linking each DAQ to the next in a daisy chain and using one master DAQ FPGA to pass only matched coincidence pairs to the processor.

1.2 Scope and Objectives

The primary purpose for this research is to review all relevant work regarding FPGA DAQ systems for use with SiPM detectors with emphasis on the PET imaging application. The full scope consists of items relating to a multichannel FPGA based DAQ with TDC, coincidence processing, external ADC, and DAQ communications link.

There were four primary objectives:

1. Determine any relevant constraints and requirements
2. Review all relevant time measurement methods and topologies and compare each implementation quantitatively against each DAQ requirement.

3. Design an FPGA based DAQ system which is customized to the unique requirements of SiPM detectors. Each design requirement will be measured and compared to the requirements of other TDC designs published.
4. Investigate a master/slave communication bridge between multiple FPGAs which will illustrate the need for bandwidth reduction using coincident event filtering before transmission to an imaging PC.

2. Review of Previous Work

2.1 Introduction

Three primary topics were examined to provide a wholistic view of a PET SiPM DAQ system with implementation in a convenient FPGA. As can be seen in Figure 1:8, an FPGA can fulfill the role of TDC, ADC pulse energy capture, coincidence processor, and high-speed DAQ link. Use of an FPGA based TDC is of particular interest and will be examined fully in regard to implementation methods.

2.2 Time to Digital Converter

Many scientific and engineering fields over the past 30 years have relied on TDCs for precise time-interval measurement. Uses include particle and high-energy physics, TOF LiDAR and PET, time-over-threshold (TOT), high-resolution digital oscilloscopes, and logic analyzers [11], [12]. The most famous application involves convergence from analog phase-locked loops (PLL) to fully digital PLLs employing digital TDCs for ease of implementation and stability [11]. Low resolution only requires a precise crystal oscillator and a clock counter. Present day PLL technology can boost a 10 MHz crystal oscillator to 1GHz providing 1ns of resolution [11]. One nanosecond is too large for most applications

therefore sub-cycle interpolation has become a necessity. Both analog and digital methods exist for interpolation. Often, custom ASIC designs employ a combination of both to provide a maximum achievable RMS resolution on the order of 0.5ps [13]. This resolution is overkill for PET scanners which are limited by scintillating material decay rates near 300ps. Local processing via parallel computing methods has become a necessity which is best accomplished by FPGA. Design complexity and cost may be reduced by adding digital TDCs to the already present FPGA.

2.2.1 Analog TDC

Traditional TDC implementation converts a time interval into an analog voltage with an op amp current source integrator (often a precise series resistor and NPO shunt capacitor) and then digitizes the output with an ADC. See Figure 2:1.

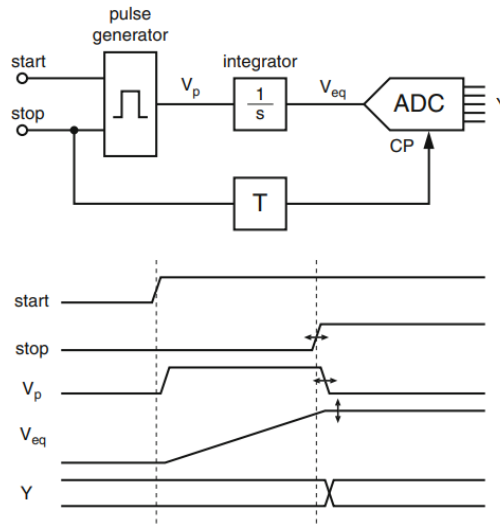


Figure 2:1 – Basic analog TDC (block and signal diagram) [11].

The dynamic range is restricted by the number of ADC quantization steps or bit count N ,

the smallest measurable time interval T_{LSB} , and maximum measurable time interval T_{MAX} .

Equations (2.1) and (2.2) can be used for dynamic range calculation with

$$T_{MAX} = 2^N T_{LSB} , \quad (2.1)$$

being the maximum time interval, and

$$DR = 20 \log_{10} \left(\frac{T_{MAX}}{T_{LSB}} \right) , \quad (2.2)$$

being the total dynamic range in dBs. Limitations exist with analog TDCs regarding variation with temperature and the lack of dynamic range which is dependent on bit count and op amp bandwidth or slew rate [11].

2.2.2 Digital TDC

Basic, fully digital TDC realization involves use of precision delay elements with latches or flip-flops capturing the count of bit flipped delay outputs following the rising or falling edge of a coarse time clock. Delay implementation is typically realized by buffers, flip-flops, or inverters [11]. Figure 2:2 represents a “pure” tapped delay line (TDL) structure with only one chain of delays. Figure 2:3 represents a dual channel TDL with inverter delays.

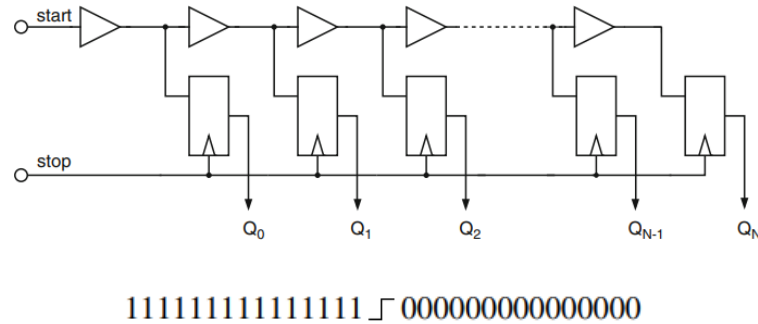


Figure 2:2 – Pure TDL-based TDC with buffer delay elements [11].

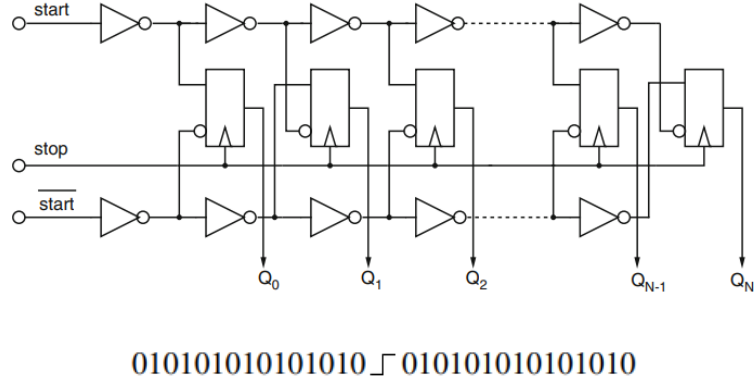


Figure 2:3 – Dual channel TDL-based TDC with inverter delay elements [11].

Figure 2:4 illustrates the Vernier TDL method utilized for sub-gate-delay resolution. Instead of using only one delay chain, two independent delay lines are used with delay $t_1 \neq t_2$ [14]. The delay on the stop path t_2 , is shorter than the delay of t_1 and the conversion is complete after the stop path outruns the start signal. The resolution is the difference between delays, $t_1 - t_2$ [14].

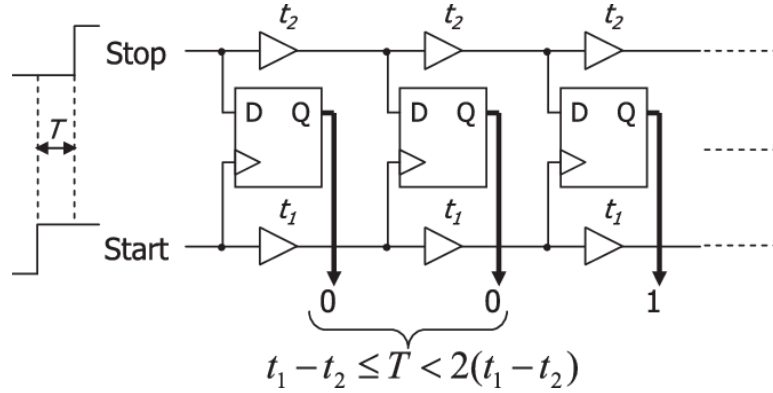


Figure 2:4 – Vernier TDL-based TDC.

For increased dynamic range, a coarse clock counter keeps track of clock cycles between the start and stop signal and one of the methods above acts as an interpolator quantifying ΔT_{start} and ΔT_{stop} with sub clock period precision. See Figure 2:5.

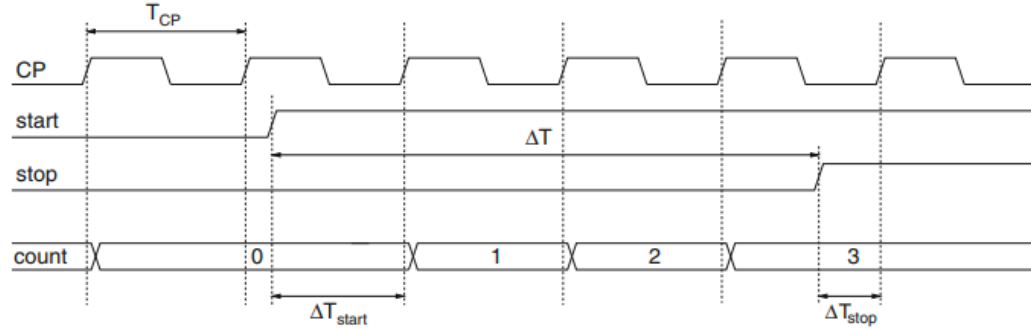


Figure 2:5 – TDC coarse counter with start and stop error [11].

Inverter delays offer nearly twice the resolution with the tradeoff of increased core area and non-linearity. Inverters are often not easily realized in FPGA design. A buffer offers a relatively consistent delay interval versus an inverter since the rising and falling transition of CMOS inverters are different [11]. The total measured time ΔT is represented here with

$$\Delta T = NT_{CP} + N_1 \frac{T_{CP}}{k} - \varepsilon_1 - N_2 \frac{T_{CP}}{k} + \varepsilon_2, \quad (2.3)$$

and total error

$$\varepsilon_T = \varepsilon_2 - \varepsilon_1 \in \left[-\frac{T_{CP}}{k}, \frac{T_{CP}}{k} \right], \quad (2.4)$$

with N equal to the number of coarse clock counts, T_{CP} the coarse clock period, N_1 and N_2 the fine delay count for start and stop, k the total number of delay cells within a clock

period, and ε_1 and ε_2 the start and stop error from an imperfect fine time measurement. Equation (2.4) shows the total possible error which is equivalent to $\pm T_{LSB}$ [11].

2.2.3 Performance Characterization

TDC performance is characterized by how close its measured value compares to the expected value and how fine the resolution is regarding each quantization step. Figure 2:6 shows a simulated TDC compared to the ideal measurement. Notice that the blue measured values are not perfectly placed on the ideal line.

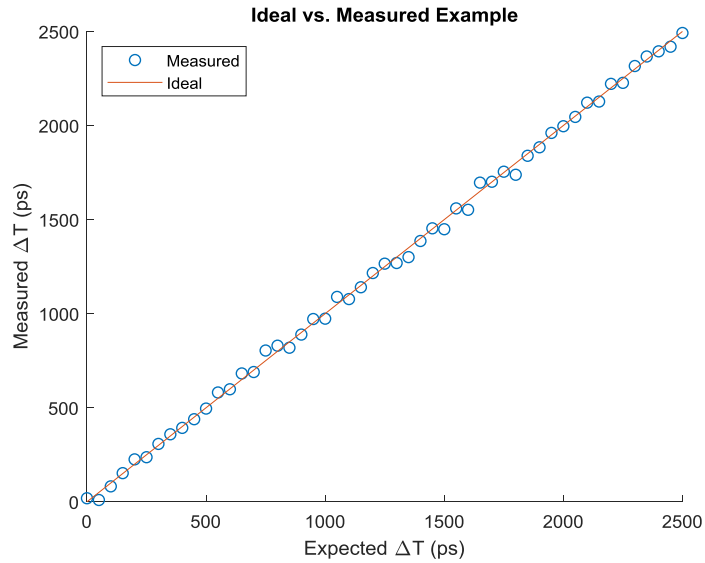


Figure 2:6 – Example measured vs. expected plot.

Within an ideal TDC, each individual delay cell is equal in length or “bin” size and distributed evenly over the entire coarse clock period T_{CP} . If asynchronous clock pulses are fed into a TDC, each bin should fill up equally because of the random distribution of ΔT_{START} and ΔT_{STOP} . An ideal TDC has a histogram of bins with equal amplitude. Any deviation from this ideal state is considered nonlinear behavior. This practice of using

random input pulses to produce a histogram is called code density testing and is also utilized in ADC and DAC characterization. Figure 2:7 shows how a signal generator should be connected for characterization of one TDC channel.

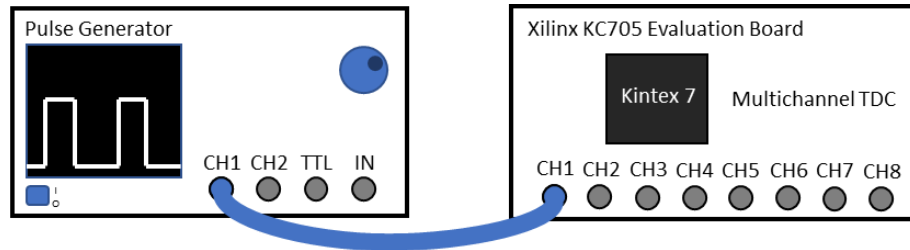


Figure 2:7 – Code density test for single channel.

Code density testing is the preferred method due to its simplicity and low equipment cost. An alternative approach is to produce pulses of very precise duration, at least five times better than the TDC itself. The pulses must be spread over the full input range for full coverage. The cost of equipment necessary for this type of precision is often prohibitive with little to no benefit versus the code density method. Figure 2:8 is an example of a TDC code density plot.

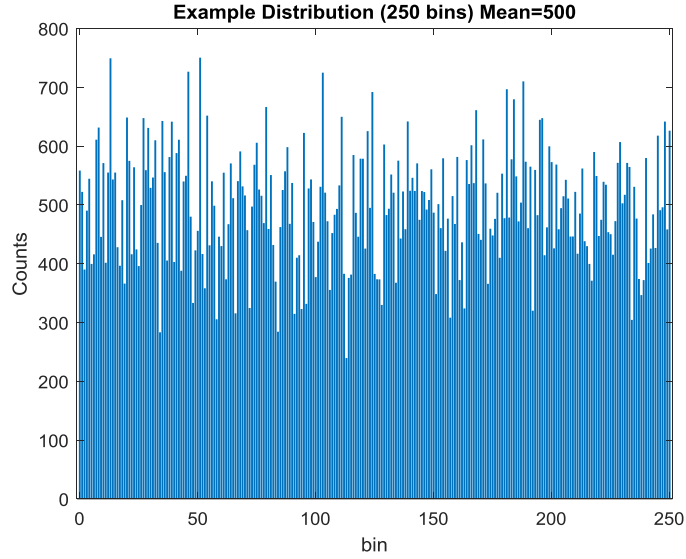


Figure 2:8 – Code density example distribution.

2.2.3.1 *Differential Nonlinearity*

Code density testing allows for TDC bin sizes to be measured. The best scenario would be for each bin to be equal length and this is the target or expected case. In reality, the bins will vary in size or duration which can be quantified with a differential nonlinearity plot. First, the average bin size $\bar{n}$ is calculated by

$$\bar{n} = \frac{N}{C}, \quad (2.5)$$

where the total number of test pulses N are spread across C number of bins [15]. In our example case shown in Figure 2:8, 125,000 pulses were utilized and spread over 250 bins giving an average bin size of 500. The differential nonlinearity for each bin, DNL_C , can now be calculated with

$$DNL_C = \frac{n_c}{\bar{n}} - 1, \quad (2.6)$$

where n_c is the number of counts at bin c [12]. Figure 2:9 is an example plot illustrating DNL. Notice the units are LSB which stands for Least Significant Bit. An LSB is the average smallest time interval which can be measured and is equal to $\bar{n}$. Use of LSB normalizes the plot. The DNL of an ideal TDC would be a horizontal line at LSB of zero.

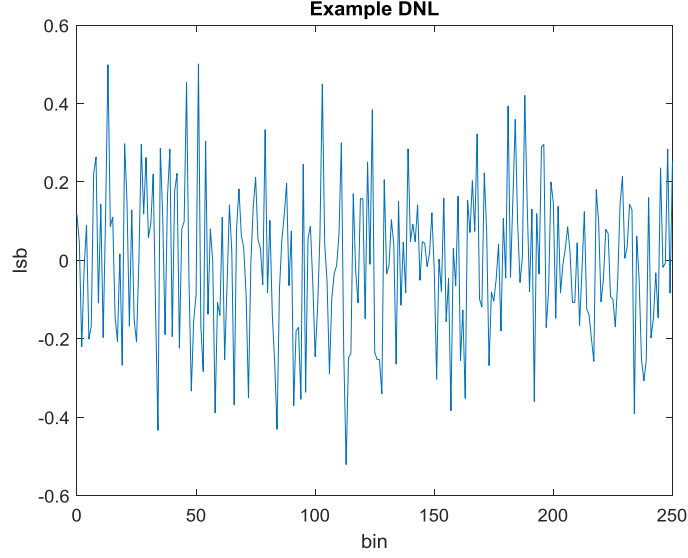


Figure 2:9 – Example differential nonlinearity plot.

2.2.3.2 *Integral Nonlinearity*

Since TDC measurements utilize more than one bin at a time, it is necessary to know the total error accumulated over the entire pulse length. If all DNL_c values are summed up through the current bin C , a cumulated sum of DNL_c yields the integral nonlinearity shown by

$$INL_C = \sum_{i=1}^C DNL_i, \quad (2.7)$$

with Figure 2:10 showing an example plot for INL [15]. Notice how the error is spread from -2.5 LSB to 2 LSB. For any given measurement, the total error can be off as much as the min or max INL.

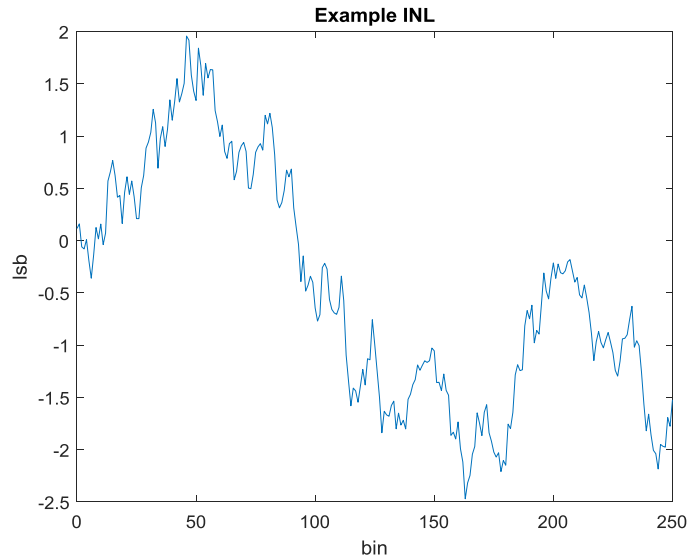


Figure 2:10 – Example integral nonlinearity plot.

2.2.3.3 *Root Mean Square Resolution*

Due to the variability of each bin size, a TDC will not measure a perfectly timed pulse exactly the same every measurement. Instead, a gaussian distribution of measurements will develop at a mean value approximately equal to the pulse length with a root mean square (RMS) resolution equal to one standard deviation from the mean. Measurement of the RMS resolution must be handled with care to provide a pulse with the exact same duration each cycle [15], [16], [17], [18]. One method which is most often employed, and which has the highest accuracy, is to feed a fixed pulse into two separate TDC channels at the same time with a known cable delay between both channels [15], [16], [17], [18].

Effectively, the start signal will be separated by the difference in cable length from a cable “T” to the input. See Figure 2:11 with cable B longer than cable A. A global clock will act as a stop pulse distributed to both TDC channels with a static delay between channels. As more and more pulses are included in the distribution, the measured RMS resolution will increase in precision. For this reason, it is common to have 100 thousand to 1 million pulses included.

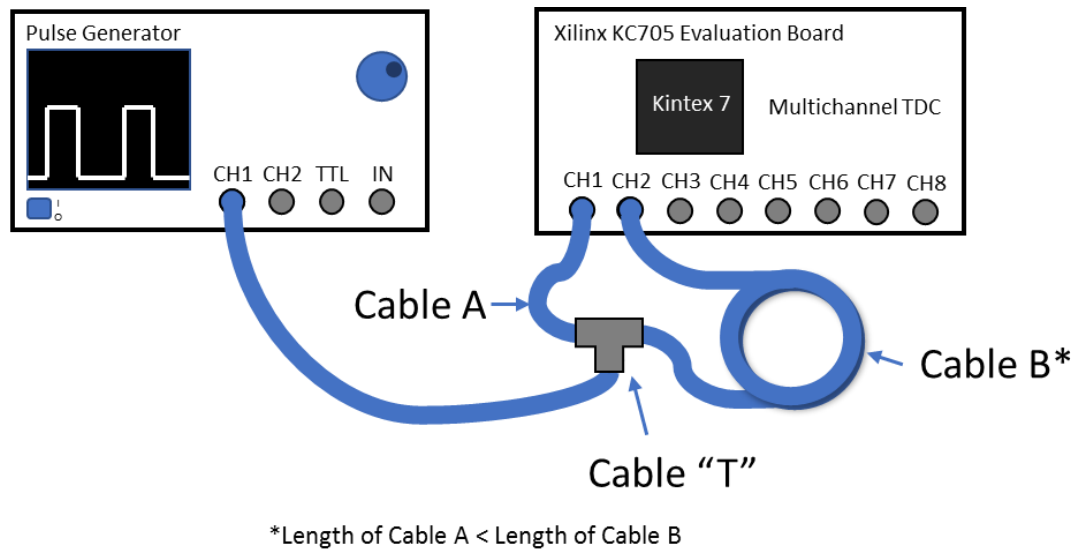


Figure 2:11 – RMS measurement with fixed cable delay.

2.2.4 FPGA-based Digital TDC

Early TDC development required use of ASICs to deliver fine resolution. ASIC TDCs are still in development today achieving 0.5ps to 30ps resolution with digital and analog methods respectively [13], [11]. For use with PET scanners, 0.5ps, is one to two hundred times smaller than is needed. Over the past three decades, many advancements in FPGA fabric technology and TDC techniques have reduced FPGA TDC resolution from 200ps and 129ps in 1997 to as low as 4.2ps in 2017 [19], [20], [21]. FPGAs now directly compete

with ASIC TDCs and often have an advantage in scalability and reduced development time. Many methods have been utilized, including Vernier TDL [19], [20], multi-phase quadrature clocks [22], pure carry chain TDL [12], [15], [23], multichain averaging [21], matrix of counters [24], wave union [25], [26], and ring oscillators [27], [28]. Each method has its own advantages and disadvantages regarding resolution, linearity, speed, dead-time, complexity, fabric slice count, and power utilization. Table 2:1 summarizes the three primary metrics for characterizing TDCs across 12 publications. As technology reduces in size, FPGA cells reduce in length and width effectively optimizing the achievable delay time. In addition, as time progresses, FPGA architecture improves and minimizes the number of cells needed for delay implementation.

Table 2:1 – Previous TDC work.

Ref.	Method	Year	RMS Resolution	Tech.	Integral Linearity (INL)	Non-Diff. Linearity (DNL)	Non-Diff. Linearity (DNL)
[19]	Vernier TDL	1997	200ps	650nm	<200ps	-94ps, +88ps	
[20]	Vernier TDL	1997	129ps	650nm	46ps	-144ps, +214ps	
[22]	Multi-phase clocks	2012	625ps	65nm	31.25ps	31.25ps	
[23]	Pure Carry TDL	-	50ps	90nm	-	300ps	
[15]	Pure Carry TDL	2009	17ps	65nm	-51ps, +43.86	-17ps, +60.35ps	
[12]	Pure Carry TDL	2013	15ps	65nm	±60ps	-15ps, +45ps	
[21]	Multichain averaging	2015	4.2ps	40nm	-28.7ps, +18.2ps	-2.9ps, +11.72ps	
[24]	Matrix of counters	2017	7.4ps	65nm	11.6ps	5.5ps	
[25]	Wave union	2008	10ps	90nm	-	-	
[26]	Wave union	2013	7.7ps	65nm	-	-	
[27]	Ring oscillators	2008	40ps	90nm	<40ps	<40ps	
[28]	Ring oscillators	2017	50ps	350nm	±65ps	±35.75ps	

At the core of most methods is interpolation which starts with a precise delay cell and is available within pre-designed FPGA fabric. FPGAs are not designed for TDC measurement specifically, but instead are meant to handle many concurrent tasks from high-speed communication to digital signal processing (DSP). Within a single cycle

multiply accumulate (MAC) function of a DSP, high-speed sub-cycle arithmetic is required. Implementation is carried out by use of a chain of fine buffer delays with the output of each fed to an adjacent D flip-flop [29]. The structure, or name of these carry chains is typically referred to as a carry tapped delay line or Carry TDL. ASICs often employ laser trimmed routing traces over low temperature drift substrates to create perfectly matched delays. Within an FPGA this is not possible, but routing constraints can be implemented in a similar fashion to construct nearly equivalent bins. The primary drawback with this method is the necessity for broad use of the FPGA fabric or usable area which often defeats the purpose of FPGA use for multichannel implementation.

Xilinx FPGAs contain configurable logic blocks (CLBs) which contain two slices with each slice containing carry chains and flip-flops [29]. Each carry chain consists of multiple CARRY4 primitives with one carry input and four carry outputs. The CARRY4 primitive is shown in Figure 2:12 [29].

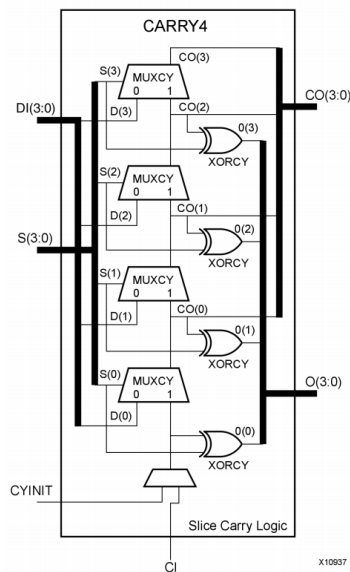


Figure 2:12 – CARRY4 (4 delay tap) Xilinx Series 7 [29].

Notice that multiplexers are used as delays with the first including an extra selector. Because of this nonuniformity, CARRY4 primitives have nonlinearity. Often, downsampling is utilized to include all four multiplexers and the initial selector in one tap. This increases linearity but reduces RMS resolution.

2.3 Pulse Energy Measurement

The pulse energy measured by PET DAQ can be accomplished by two primary methods. The traditional, post-sample digital integration, method detailed in Figure 2:13, first digitizes the wave form by ADC with sample rate between 65 and 100 MHz before digitally integrating the time domain samples within an FPGA [30], [31]. One high-speed, high-cost ADC is required per channel. The second approach, pre-sample analog integration, detailed within Figure 2:14, integrates first by use of a low-drift capacitor and then digitizes the result with a slow-speed (~ 1 MHz), low-cost ADC. The traditional method has an advantage regarding PVT drift since the ADC benefits from a very precise, temperature stable, band-gap voltage reference. The second approach has an advantage in lower system complexity and cost for high-channel count implementation. By integrating first, and holding the result, a single slow-speed, high dynamic range ADC is able to cycle across many channels. As many as 8 to 16 channels share an ADC with limitation being the allowable deadtime between pulse capture. This method also has a lower FPGA resource requirement since it only requires one data bus and one shared FIFO.

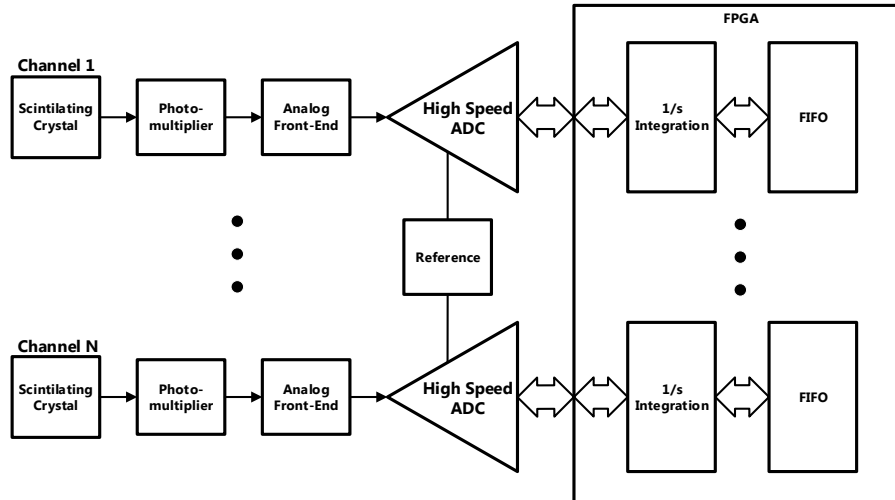


Figure 2:13 – Energy measurement (post-sample digital integration).

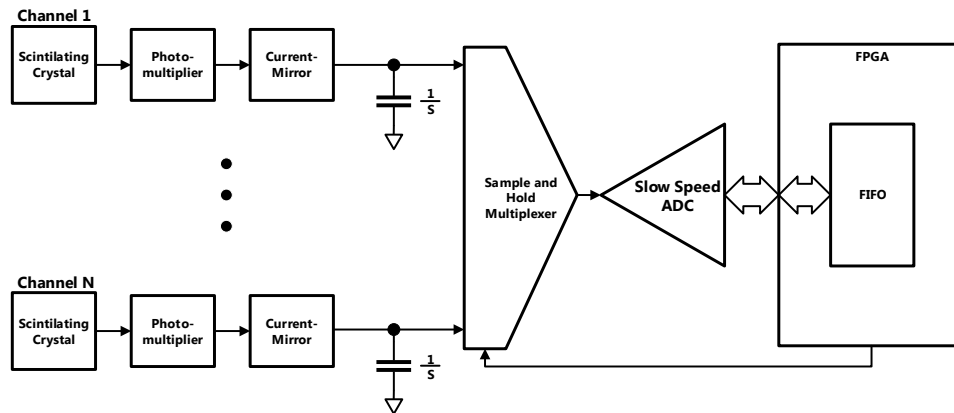


Figure 2:14 – Energy measurement (pre-sample analog integration).

2.4 Inter-DAQ and DAQ to PC Communications Link

A DAQ system not only measures the event arrival time and the energy of the detected photon, but also must be able to transfer single event and coincident event packets between DAQs and to the imaging PC. Before any PET DAQ system is designed, the requirements must be defined based on the size, resolution, and sensitivity needed. For each, the application must be considered. Common applications include: small animal PETs, breast

dedicated PETs, brain dedicated PETs, and whole-body PETs [4]. For adequate spatial resolution when imaging a small animal, each crystal must be the smallest possible at around 1.2~3 mm versus a whole-body PET at 4~5 mm [4]. For sensitivity, a Brain PET must detect a higher percentage of events at around 7% versus only needing 2% for a whole-body PET [4]. Ealgoo Kim et al. summarized typical PET specifications very concisely with an excerpt below in Table 2:2 [4]. By understanding the maximum possible number of single and coincident events and multiplying by the packet size, a maximum bus speed may be determined.

Table 2:2 – Typical PET system specifications [4].

Characteristic	Brain PET	Whole-Body PET
Bore diameter (cm)	30	60~80
Axial field-of-view (cm)	2~18	15~20
Detection area (cm ²)	~1400	~4000
Crystal pixel size (mm)	2.5~3.5	4~5
Number of crystals	10k~120k	10k~40k
Sensitivity (%)	<7%	<2%
Spatial resolution (mm)	3~4	5~6
Measurement time (min)	20~40	25~50

Each DAQ system must be designed to accommodate the number of single events received by detectors or crystals attached. As can be seen in Table 2:2, anywhere from 10 thousand to 120 thousand crystals may be needed within a PET gantry. Since each DAQ will most definitely be limited regarding the number of possible input channels, various multiplexing schemes are often employed to reduce the number of detectors needed per crystal count and to reduce the number of electronic channels needed per detector. The rate of the events received at each detector depends on the accumulated amount of radio-pharmaceutical inside the imaged subject, the coverage of the detectors, and the sensitivity of the PET imaging system [4], [32]. As single events are received, initial detector

comparators filter all low-energy events which fall well below the gamma photon energy of 511keV. Next, true, random, and scattered events are received and processed by the coincident detector to determine if each fall within a coarse time interval and match with a corresponding event 180° opposed. This allows for filtering of random events. Nearly 30% of all events which pass the initial energy filter are false random or scattered events [4]. Detector and TDC measurement timing resolution is key to minimize the coincidence (coarse) time interval and filter as many false events as possible to maximize the image quality during reconstruction. The photon traverse time across the bore of the PET system along with the timing resolution of the detector/TDC pair determines the smallest coincidence interval possible [4], [33]. Last, the measured energy after integration is utilized to precisely determine if an event is true or the result of Compton scattering which reduces the energy. Good energy resolution is required to correctly filter scattered events. As more and more crystals are multiplexed into a single energy measurement, a pile-up of multiple simultaneous events can increase the received energy. Sophisticated algorithms are required to remove pile-ups which need precise energy and timing measurements for correct filtering [4]. As filtering methods improve, the total requirement for DAQ communication speed reduces.

Ealgoo Kim et al. summarized key parameters for PET DAQ systems with an excerpt in Table 2:3 [4]. Notice that a whole-body scanner may have near 60 million single events in one second spread over 400 to 1,200 channels. This equates to 50 to 150 thousand events per DAQ channel in one second. Also note that at 1,200 channels, a single DAQ should accommodate as many channels as possible either through multiplexing and combining channels or through use of high-density FPGAs.

Table 2:3 – Typical key parameters of a PET DAQ system [4].

Requirement	Brain PET	Whole-Body PET
Number of channels	~4500	400~1200
Coincidence timing window (ns)	6	<6
Timing resolution (ns)	3	0.5~3
Energy resolution (%)	12~14	10~16
Single event input rate (count/s)	20M	<60M
Coincidence event throughput (counts/s)	1M	<1M
# of coincidences for an image (events)	200M	200M

As can be seen in Table 2:3, for the two most common PET applications, the known maximum single event rate per second is 60 million and the maximum coincidence throughput per second is one million. Depending on the packet size for each, careful design choices should be considered to keep the data transfer rate within an allowable range for typical bus sizes. Also, the data path structure should be evaluated to minimize unnecessary transfer bandwidth. The data path can be constructed in one of three possible configurations, all of which equally taking advantage of initial multiplexing and differ primarily in the interconnection between DAQs, coincidence detection modules, and the imaging PC. The traditional approach is a Tree Path structure with each DAQ separately transferring single event packets to a single coincidence detector before coincidence event packets are transferred to the PC. The amount of information sent to the PC is adjustable, but the progression from N channels per DAQ, to N x M channels per coincidence detection maintains the same with N dictated by how much bandwidth is possible between the DAQ and coincidence detector on a single bus. Figure 2:15 shows a typical Tree Path structure. The next approach is a Daisy-chain Path structure detailed in Figure 2:16 with each DAQ fulfilling a dual purpose of handling both single event receipt and coincidence detection. Each DAQ transfers its single event packet to the next DAQ in a chain with the last DAQ

transferring to the first. Only one of the DAQs must transfer the coincidence event packets to the imaging PC. The big advantage with this method is the reduction in hardware with each DAQ spreading the coincidence detection load and eliminating a central processor. The last method is the Bus Path structure detailed in Figure 2:17 which provides a single interconnecting bus between DAQs and one central coincidence detector. This method requires the highest single event bus speed since one bus handles transfers to and from each DAQ and to the coincidence detector simultaneously.

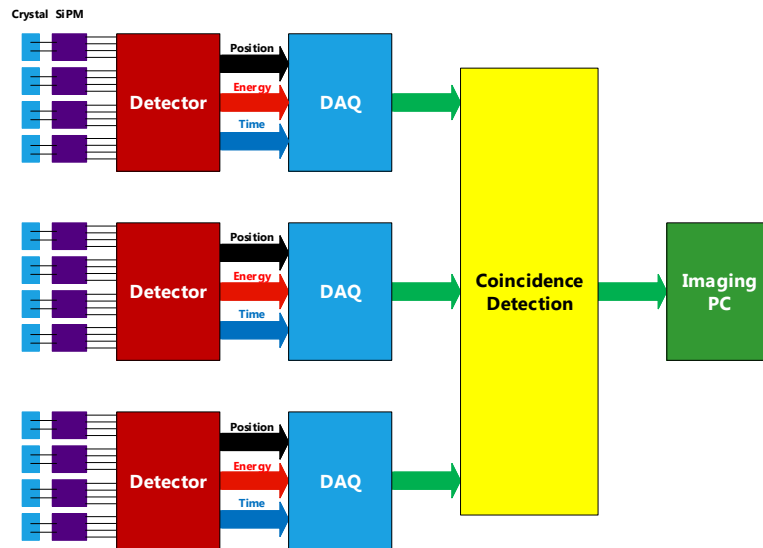


Figure 2:15 – Tree data path DAQ structure.

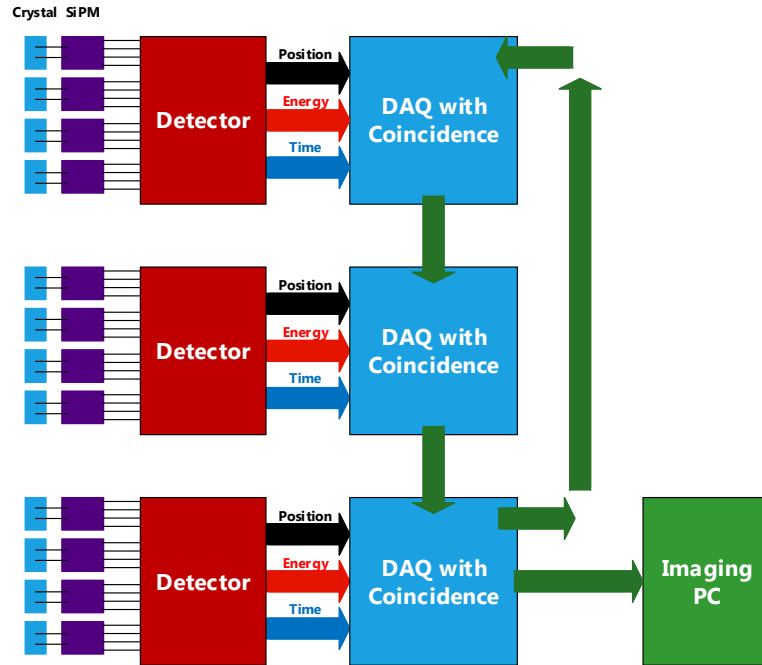


Figure 2:16 - Daisy-chain path DAQ structure.

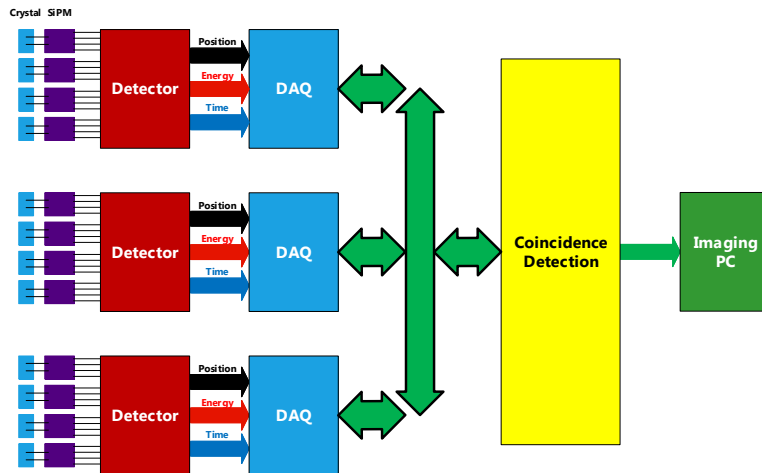


Figure 2:17 - Bus path DAQ structure.

3. Analysis of the Problem

An FPGA DAQ for PET SiPM detection has three primary objectives which all must be realized for successful implementation of a TOF PET scanner. Each objective carries

equal weight in significance but not in complexity. Listed below are the three objectives in order of complexity:

1. Measure the arrival time for annihilation events which feeds into TOF calculations
2. Record the integrated pulse energy from each annihilation event.
3. Forward event information to central PC for 3D spatial image recreation.

The first object must be met by use of a TDC capable of timing measurement two to five times smaller than the decay rate of the scintillation material/photomultiplier pair. At present, 300ps is the smallest scintillation material/photomultiplier rate achievable which dictates a TDC resolution of 60 to 150 ps [5]. The second objective must be capable of integrating/measuring the pulse energy and storing the result along with the timing information. Since an FPGA may be used for multiple PET channels, measurement must cover all channels within the same angular row. For the third objective, each FPGA DAQ must be capable of sending as many measured events to a PC as possible to avoid loss of pulses which would lengthen the required scan time. Careful consideration of how much information must be passed to the PC and a transmission rate must be selected which may be achievable by modern FPGA technology. This section will provide an analysis of each objective with a progression towards achievable implementation.

3.1 Arrival Time Measurement

Time-to-Digital Converters are required for PET arrival time measurement since simple, counter based designs are not capable of the needed sub-nanosecond resolution. Traditional TDC design has been from custom integrated circuits with complete control over architecture and only limited by existing fabrication technology. As discussed, delay

cells are required which should be equal in length for maximum linearity. FPGAs are pre-designed ICs meant for flexible, rapid development covering broad application and not inherently well suited for repeatable, precise time measurement. A review of past FPGA TDC design reveals a repeated pattern of selecting an FPGA with fabric best suited for chains of tapped delay lines and careful planning to manually instantiate and place primitives to achieve design targets. Analysis progresses in the following way:

1. Select an FPGA based on historical review
2. The maximum register-to-register clock speed is determined
3. A single channel TDC with a lengthy chain of delays and a basic thermometer to binary encoder is implemented to determine the average delay cell time
4. A review of bubble error is considered to select the best encoder type going forward
5. A review of linearity is considered to select methods for improvement depending on the application specific requirements and aggressiveness
6. A calibration method is chosen and implemented
7. The single channel design is replicated based on the number of channels needed and/or fabric size
8. Adjustments are made iteratively for optimization

3.1.1 FPGA Selection

A review of previous work indicates that FPGA based TDC design began in the late 1990s and has improved proportionally along with technology improvements in size, power, and logic block granularity [19]. Just as ASICs progressed to CMOS technology, FPGAs followed, due to its low manufacturing cost and low power consumption. Early

FPGAs utilized an amorphous antifuse structure with pASIC logic blocks for one-time programmable designs [19], [20]. The development cost of a design based on antifuse technology was lower than an equivalent ASIC but still much more expensive than present day flash-based designs. More recent designs utilize carry chains as delay lines with Xilinx brand as the preferred choice due to its best-in-class fabric size and IP selection. Most notably is the Virtex family optimized for highest system performance and capacity. Virtex based TDCs cover 80% of all FPGA based TDCs discovered. Due to its high cost, a PET scanner built with Virtex series FPGA may be less cost efficient and higher power than required. The Kintex family is optimized with best price-performance in mind. Since greater than 100 FPGAs are required for a basic TOF PET scanner, cost optimization is of primary concern after the primary objectives are met. For this reason, a Kintex 7 FPGA was chosen for initial experimentation. The KC705 evaluation board, Figure 3:1, at first glance, has all required features necessary for 16 to 32 channels (FMC headers), built in A to D conversion (XADC header), and high-speed communication (gigabit GTX transceivers) [34].

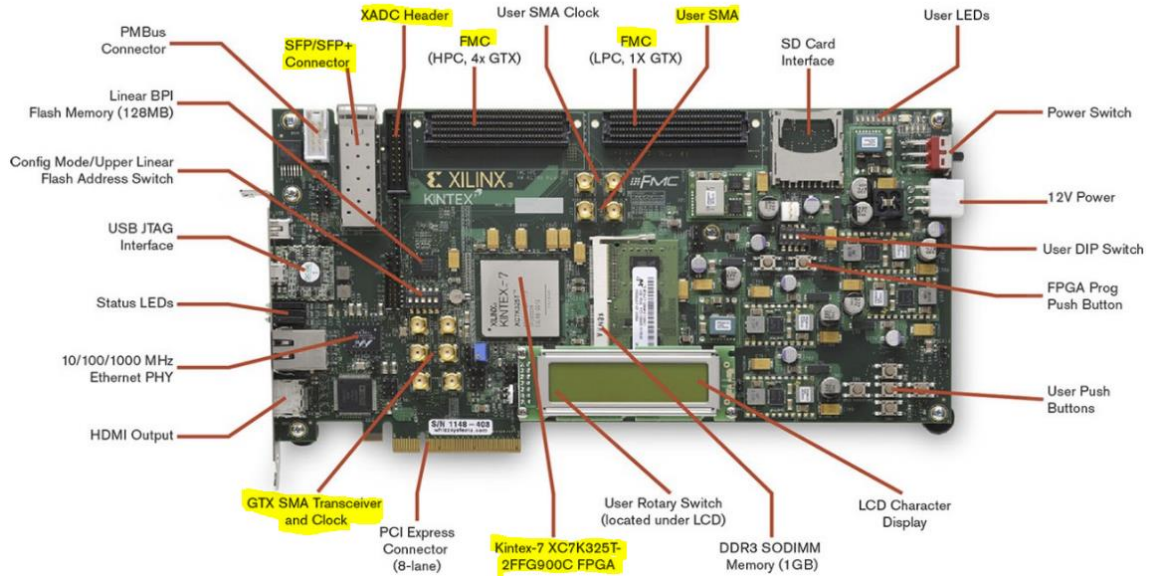


Figure 3:1 – Xilinx KC705 evaluation board [34].

3.1.2 Maximum System Clock Frequency

The system clock frequency of an FPGA must be high enough to coincide with the time interval for coincident detection of annihilation events. As the coincident interval decreases, unpaired events can be filtered more easily, thus increasing the signal-to-noise ratio. Based on review of previous work, a requirement of 2.5ns or less, is needed which indicates 400MHz or higher system clock. Both Virtex and Kintex 7 series FPGAs meet this requirement based on fabric and block ram F_{MAX} with a healthy margin for reliability. The KC705 evaluation board uses a -2 speed grade Kintex XC7K325T. Table 3:1 indicates 544MHz as the max -2 Kintex frequency. Although 544MHz is possible, 400MHz will be utilized at first to minimize implementation optimization.

Table 3:1 – Kintex 7 and Virtex 7 F_{MAX}.

Kintex®-7 FPGAs

Speed grade	-1	-2	-3
F _{MAX} [MHz]	464	550	741
Max GMAC/s	1,781	2,112	2,845

Virtex®-7 FPGAs

Speed grade	-1	-2	-3
F _{MAX} [MHz]	547	650	741
Max GMAC/s	2,756	3,276	3,734

Kintex-7 and Virtex-7 FPGAs

Speed grade	-1	-2	-3
True dual-port Block RAM F _{MAX} [MHz]	458	544	601

3.1.3 Average Delay Time

An initial TDC design was developed based on CARRY4 tapped delay lines within the Xilinx 7 series SLICEL with one register stage and one latched stage running at a system clock speed of 400MHz. A total of 75 CARRY4 primitives were utilized with 300 discrete tapped delays. Figure 3:2 illustrates in red the carry delay path and in blue the registered path leading to a latch stage. Figure 3:3 details the full single channel design with first stage registers in blue and the latch stage in green. Notice that the latch is enabled by a state machine counter which counts clock pulses occurring after the pulse rising edge. At state count equal to one, the latch is enabled and at state count equal to five, the binary result from the thermo-to-binary encoder is latched and transferred to block ram.

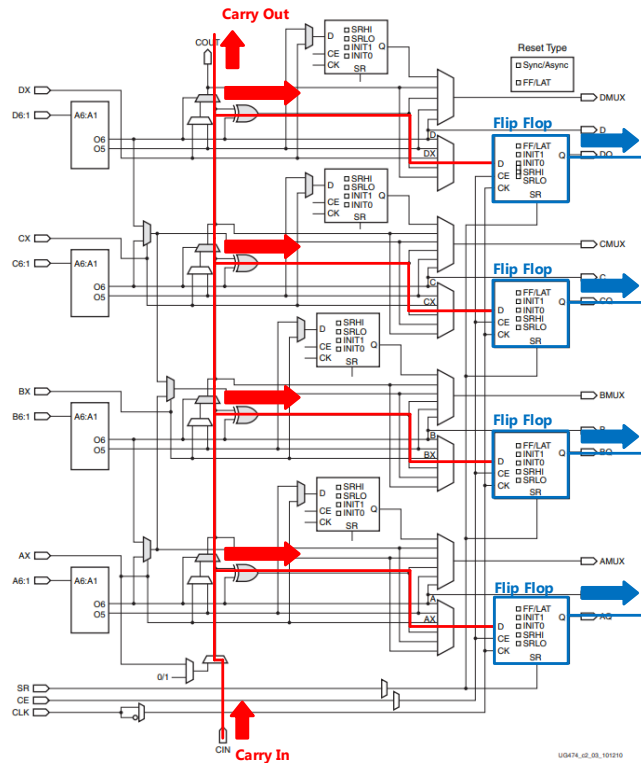


Figure 3:2 – 7 Series SLICEL with registered tapped delay line.

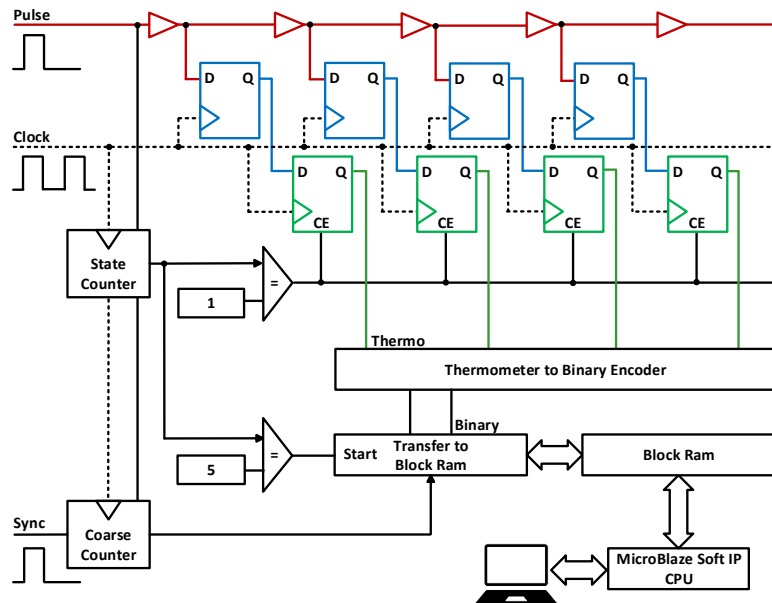


Figure 3:3 – Initial single channel TDC.

Utilizing the code density method, a histogram of bins is produced. Initial results reveal many zero count bins with the first bin starting at #27 and ending at #255. The full 2.5ns system clock divided by the bin spread equals an average bin size of 10.9ps. Since the number of missing bins is higher than expected, the design is investigated by reviewing the post implementation timing report. A higher than expected path time between the thermometer to binary encoder indicates a minimum of 6 clock cycles is needed before latching on the binary output. To be safe, the encoded output is changed to a count of 12.

Figure 3:4 and

Figure 3:5 represent the TDC Bin count with binary latch at State Count equal to 5 and 12 respectively.

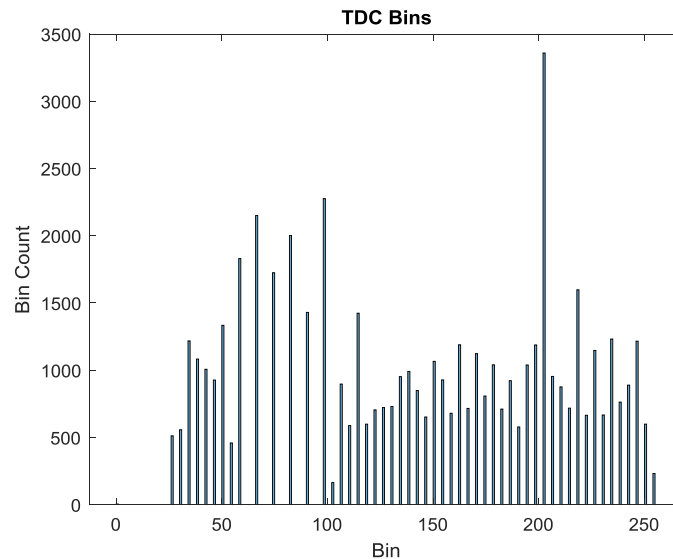


Figure 3:4 – Initial TDC bin count (bin latch at state count = 5).

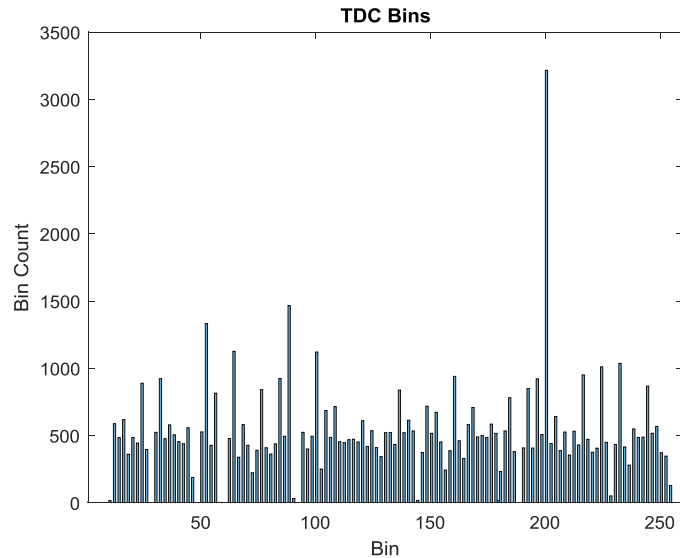


Figure 3:5 – TDC bin count (bin latch at state count = 12).

With the binary latch at State Count = 12, many more bins have filled in and the linearity has improved. The bin distribution is now from #10 to #255 with an average bin size of 10.2ps. A single bin stands out as the single highest bin size at a count of 3216 versus the non-zero average of 500. A review of the architecture of the Kintex 7 XC7K325T reveals 14 separate clock regions with a maximum carry chain of 200 delay cells per region. See Figure 3:6. As the carry chain crosses the clock region boundary, a significant delay is added. In order to maximize linearity, the goal is to increase the system clock frequency until the clock period can be covered by the delay chain interpolator in less than 200 delay cells with margin for process, voltage, and temperature fluctuations. Using a 10.2ps to 10.9ps average bin length spread across 200 bins would equate to 459 to 490 MHz. To be safe, 500MHz is chosen for the next attempt.

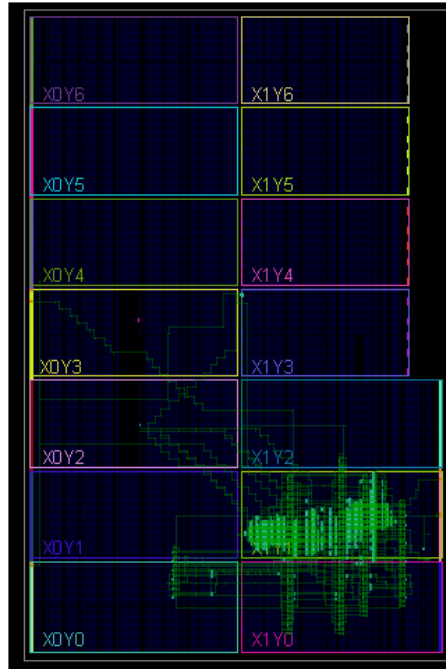


Figure 3:6 – The Kintex 7 XC7K325T architecture with 14 clock regions.

3.1.4 Thermometer to Binary Encoder

As the bin counts are inspected more closely in Figure 3:5, it is noticed that most even count bins remain zero in size. Now that the encoder timing is no longer a concern, the system bubble error must be examined and adjustments to the encoder must be considered. In order to look at the thermometer code, the Integrated Logic Analyzer IP (ILA) from Xilinx is employed. As expected, the last two thermometer bits were occasionally flipped from the correct state. In Figure 3:7, notice the rising edge of the clock pulse coinciding with the delay chain. Once the clock rising edge occurs, the binary state from each delay cell is latched into the first stage registers. Ideally, a continuous stream of 1s would exist followed by a continuous stream of 0s. Just as one might read the current temperature of a mercury thermometer, only the last transition from opaque (1) to clear (0) matters.

However, at the moment of latch, not all register setup time requirements have been fulfilled. This causes some bits to be incorrectly flipped.

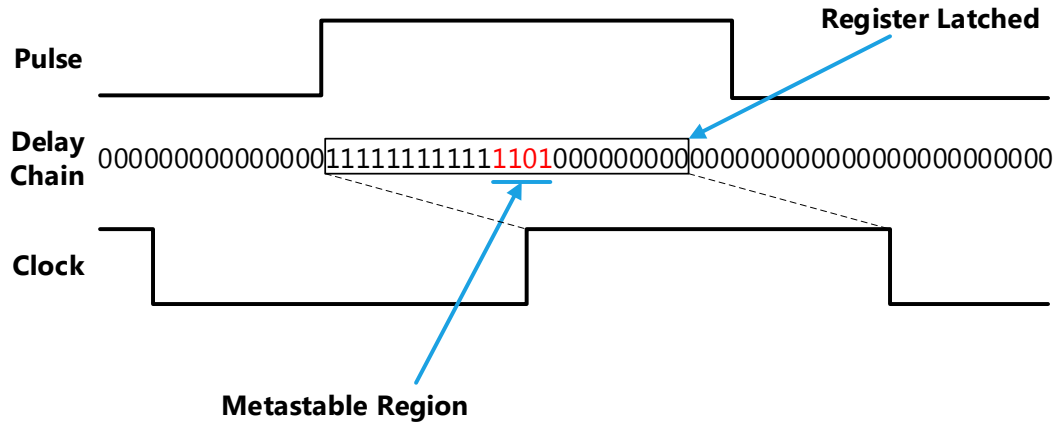


Figure 3:7 – Bubble error.

Since there is only one noticeable bit flip, most bubble error correcting encoders would be capable of filtering the bubble. The encoder utilized is recommended by Zbigniew Jaworski in review of common encoders with first and second order bubble correction [35]. In Jaworski's review, he compared "The Highest '1' (native)," "The Highest '1' with BEC," "One-Hot Encoding," "Sum of Bits with OR," "Direct Truth Table," and "Priority Encoder" [35]. The simplest, smallest in size and power, and best at all types of bubble error correction was "The Highest '1' (native)" [35]. Essentially, as soon as the thermometer input changes, combinational logic cycles through checking each input bit and finds the highest bit which is 1. Regardless of what any other lower bit value is, the highest 1 is found and resolved to a binary output. There is however, one assumption within this paper which is flawed for FPGAs with different clock skew between registers. Jaworski assumed that any value of 1 is true. Within the Kintex 7 series, some registers

within a SLICE see the clock edge sooner than others even though they should be later in the chain. See Figure 3:8 which highlights the clock path feeding into the third register before branching to two and four with the first receiving the clock last. This phenomenon has become more common with higher speed FPGAs since the delay cell time is small compared to the clock skew. Older, cheaper FPGAs have longer delay cell times allowing for each bin to fill with minimal zero width's [36]. The solution is bin realignment.

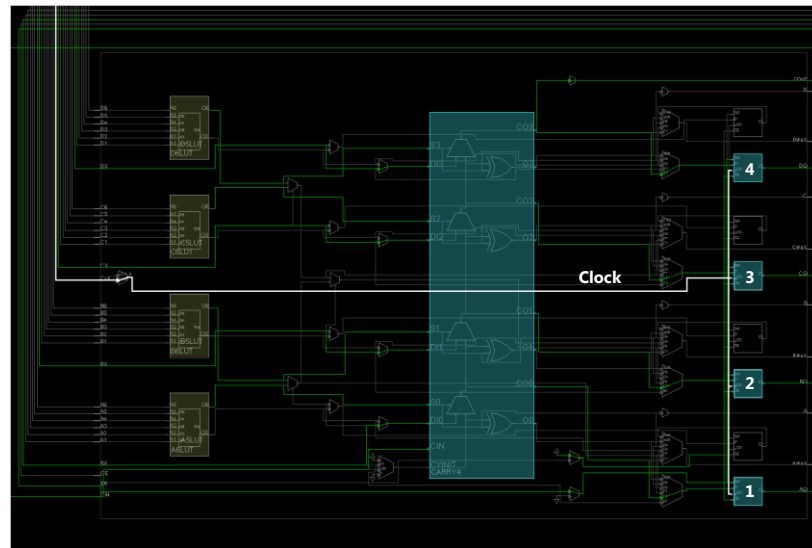


Figure 3:8 – Clock skew in Kintex 7 SLICEL.

Bin realignment, proposed by Liu and Wang, first uses the code density test to determine bin size and then by taking the results and utilizing statistical methods, a translation table is generated between the latched delay chain and the thermo-to-binary encoder [36], [37]. This method was very iterative and time consuming. G. Cao et al. proposed to perform a one-time realignment based purely on a single code density test [38]. Code density testing was utilized to determine that every other bin is very small in duration and swapping every even tap with odd before feeding the encoder was chosen. Post

implementation timing simulation seen here in Figure 3:9 confirmed that every odd bin is small in comparison to the neighboring even bin.

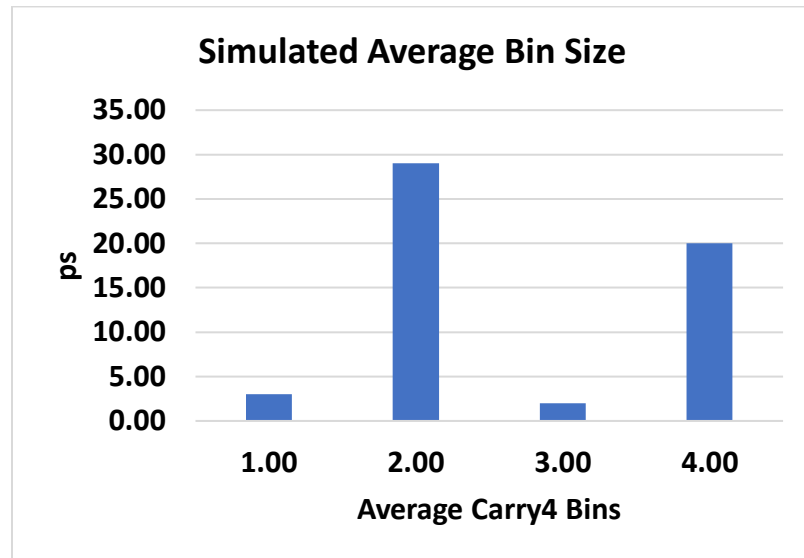


Figure 3:9 – Post implementation timing simulation for average CARRY4 bins.

After bin realignment, the code density test reveals almost all bins being filled except for the first 64. See Figure 3:10, and Figure 3:11. An attempt was made to reduce the front-end missing bins by placement and routing adjustments. Each movement of the carry chain shifted the front-end missing bins left and right with no definitive pattern. Good results occurred after buffering the input pulse signal with a global BUFG buffer. This, however is not a reasonable solution for multichannel implementation given the small quantity of on-board global buffers. Instead, after simulation, it was noticed that latching the carry chain when the State Coarse count is equal to one, causes a small percentage of pulses to be latched one clock cycle too late. This is due to the complex and slow process of

comparing a multibit count to a static condition. Instead, a faster acting pulse detection is needed.

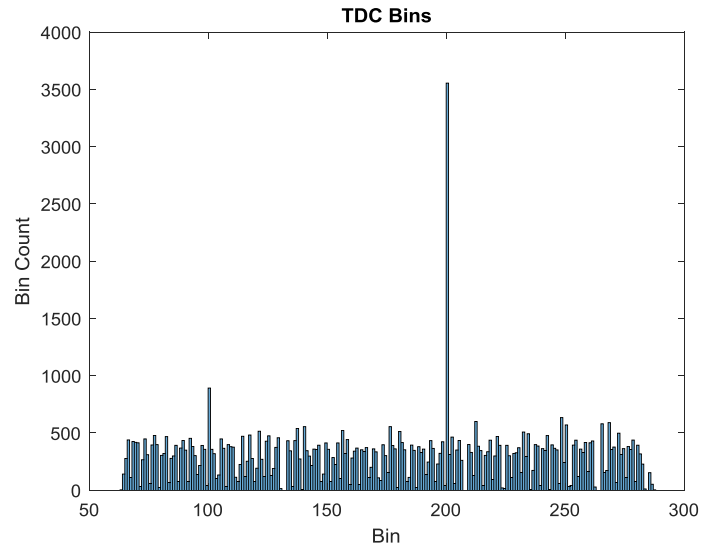


Figure 3:10 – TDC after bin realignment.

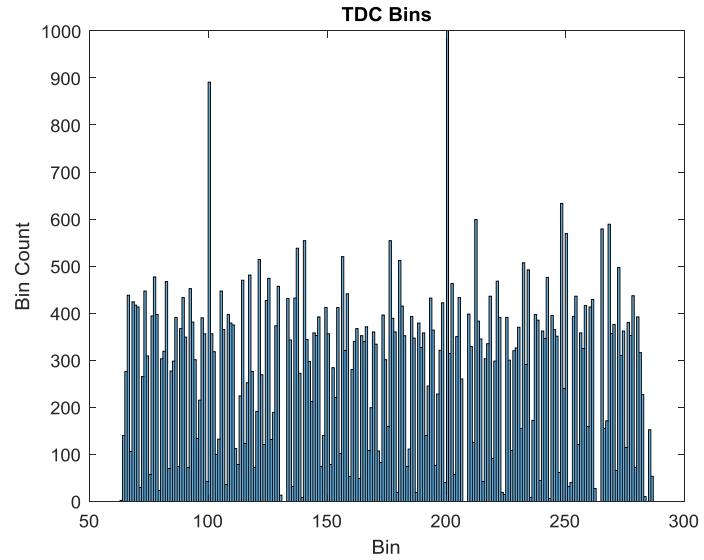


Figure 3:11 – Zoomed in view of TDC after bin realignment.

3.1.5 Pulse Detection and Latching

Little is mentioned within previous literature regarding a method for enabling the latching stage of registers. A suitable pulse or “hit” detector is required to enable the latch consistently with enough setup time for low metastability. Zhao et al. did reveal use of an AND gate with one inverted input fed by the first register from stages one and two [12]. This is a typical rising edge detector and may be the likely approach for most designs. Figure 3:12 details this method with the first two register stages followed by a latch stage in purple. This method was attempted with no benefit noticed. In fact, the frontend zero bin count increased substantially, and simulation revealed that the Stage Count = 1 method in Figure 3:3 produced a latch enable sooner for better results.

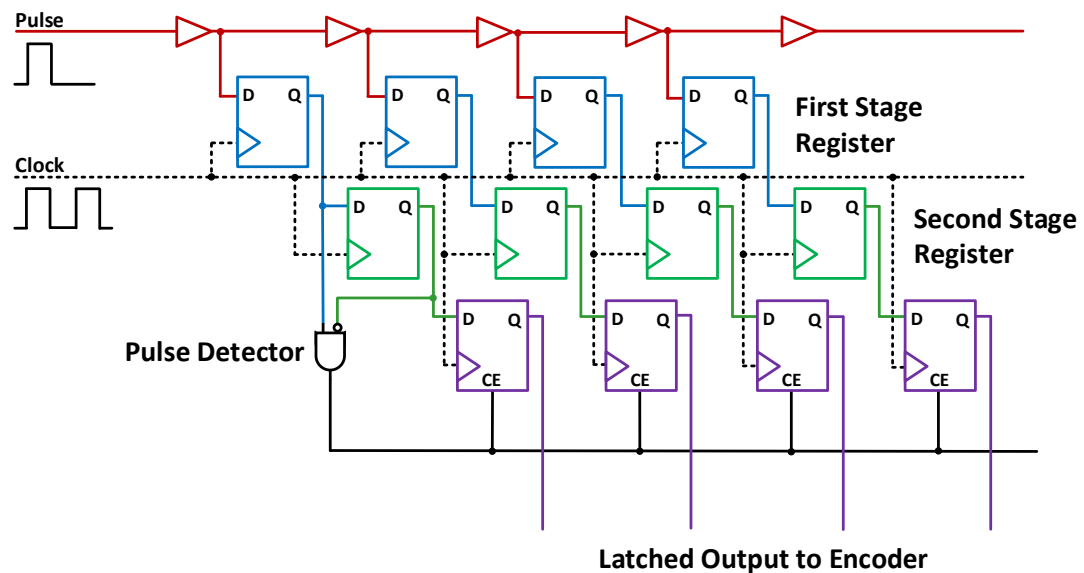


Figure 3:12 – Typical pulse detector.

Most references do, however, explain the necessity for keeping the system clock below F_{MAX} which allows for a healthy margin with timing analysis for register-to-register setup and hold times. Usually, the maximum clock speed realized is between 100 and 300

MHz. An attempt was made to reduce the clock frequency from 400 to 200 MHz. This resulted in marked improvement, but with the drawback of an increase in the delay chain length and size of the encoder. Before giving up on a better detector option, a concerted search effort resulted in a suitable option detailed by Harald Homulle et al. [39]. Figure 3:13 details the design which converts a pulse of variable length to one cycle in duration and produces a coinciding latch enable “Valid” signal.

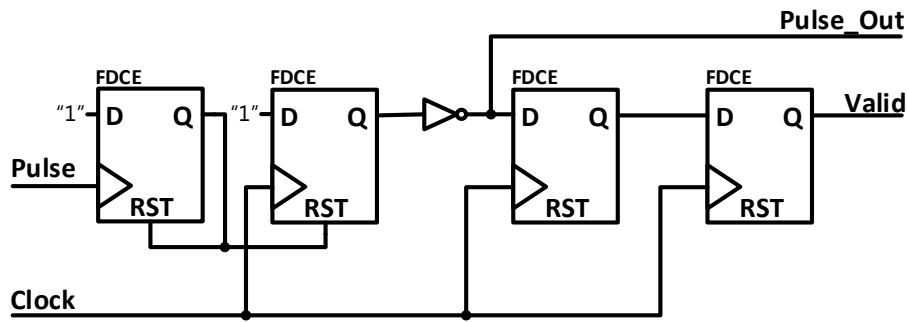


Figure 3:13 – Fast-acting pulse detector with filtered output pulse.

This method was implemented with two initial register stages for a reduction in metastability and a third latching stage enabled by the “Valid” signal. Simulation indicates that this approach provides an enable signal earliest of all tested methods and testing revealed little to no front-end zero count bins.

3.1.6 Linearity Improvement

With pulse detection improvements, the dependence on system clock frequency relaxes. For maximum linearity, the carry chain must be limited to a single clock region which eliminates the very high bin length when crossing the boundary. Each clock region maxes at 50 CARRY4 primitives in a vertical chain which equates to 200 delays. At an average delay time of ~11ps, 500MHz is a healthy choice giving 18 additional delay cells

for PVT changes. Figure 3:14 is the result of the first attempt with the fast-acting pulse detector.

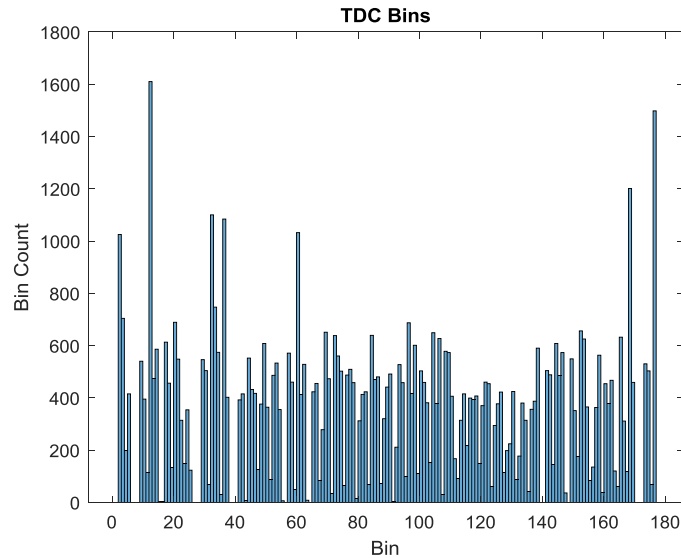


Figure 3:14 – 500MHz initial TDC with fast-acting pulse detector.

Additional improvement is possible by trying placement in different regions of the FPGA. Slight variation exists due to the inhomogeneous fabric structure and routing deviations. Place and route constraints are able to help with consistency but not perfectly. Instead, multichain averaging proposed by Qi Shen et al. has a much more significant impact [21]. As more parallel chains of delays are measured simultaneously, the average captured binary output reduces the bin size and the DNL and INL. Figure 3:15 reveals the results from increasing the number of parallel chains from one to four.

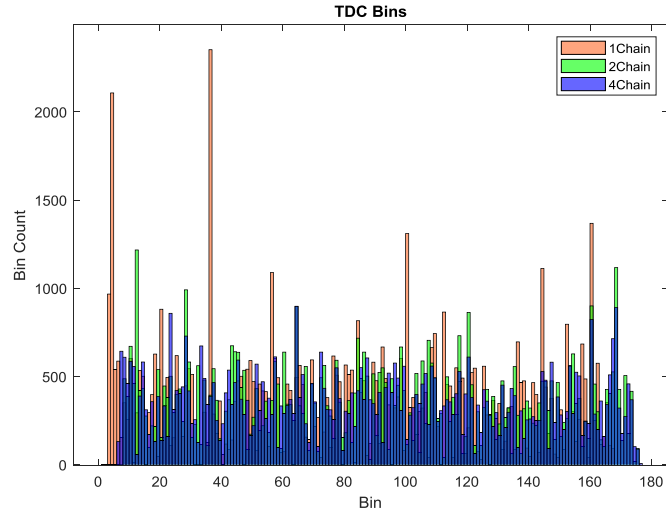


Figure 3:15 – 1 Chain, 2 Chain, and 4 Chain multichain histogram.

As more paths are added, the very large bins decrease in size and the overall difference from bin-to-bin reduces. A noticeable shift also occurs horizontally due to clock skew differences which exist between each path. Linearity improvement is most evident by viewing the differential nonlinearity plot in Figure 3:16. The DNL plot highlights how much difference exists for each individual bin and the average bin size. For a single chain, the min, max DNL is -1 to +5 LSB respectively. The two chain is -1 to +2.2 and the four chain is -1 to +1.4 LSB. Each LSB is 11ps.

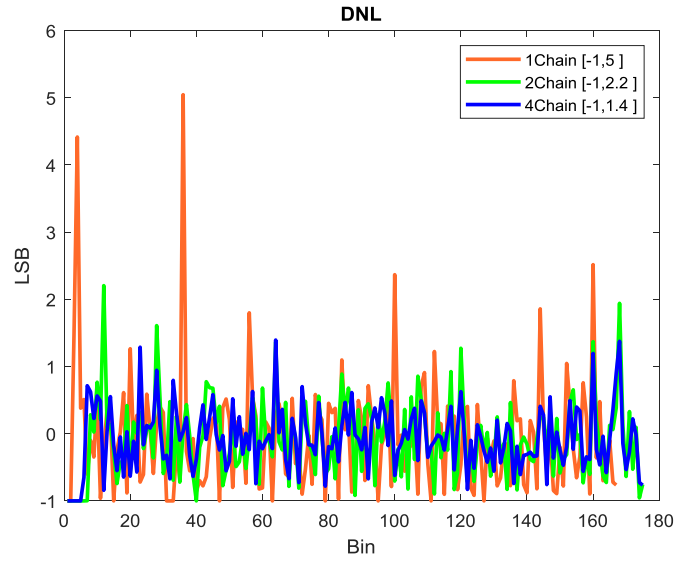


Figure 3:16 – 1 Chain, 2 Chain, and 4 Chain multichain DNL.

The integral nonlinearity plot also shows improvement from one chain to four although there is a prolonged period of shorter than average bins near 120 which skew the results. See Figure 3:17. The overall results are summarized within Table 3:2

Table 3:2 – Pre-calibrated 1 Chain, 2 Chain, and 4 Chain DNL/INL.

Method	DNL (LSB)	DNL (ps)	INL (LSB)	INL (ps)
1 Chain	-1, +5	-11, +55	-9.4, +4.5	-103.4, +49.5
2 Chain	-1, +2.2	-11, +22	-5.6, +3.9	-61.6, +42.9
4 Chain	-1, +1.4	-11, +15.4	-4.2, +2.9	-46.2, +31.9

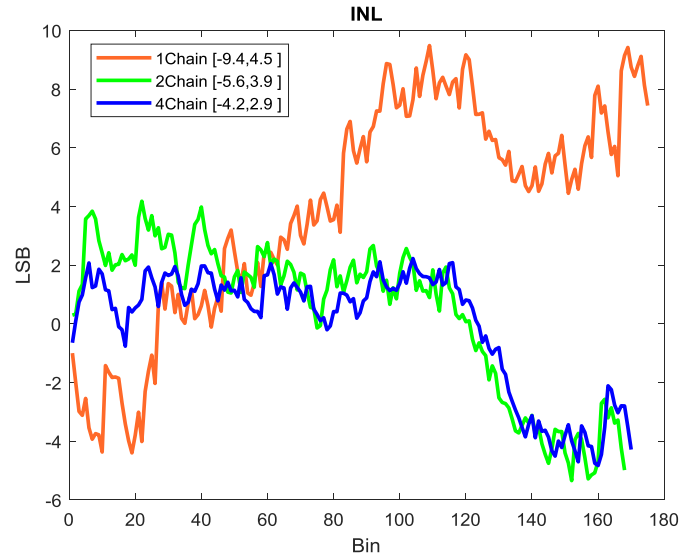


Figure 3:17 – 1 Chain, 2 Chain, and 4 Chain multichain INL.

The RMS TDC value is a measure of the system noise. Even if the average delay size is very small, the RMS value can be high. Also, it is possible to have a very low noise system with a relatively large average delay. Using a 12-inch cable and the RMS measurement method illustrated in Figure 2:11, the difference between a one, two, and four chain averaged TDC was determined. As more chains are added, the uncalibrated RMS value reduces from 24.64ps to 15.17ps as seen in Figure 3:18. It was noticed that with only two chains, the RMS value resulted in 15.26ps which is already very close to the four-chain result. See Figure 3:19.

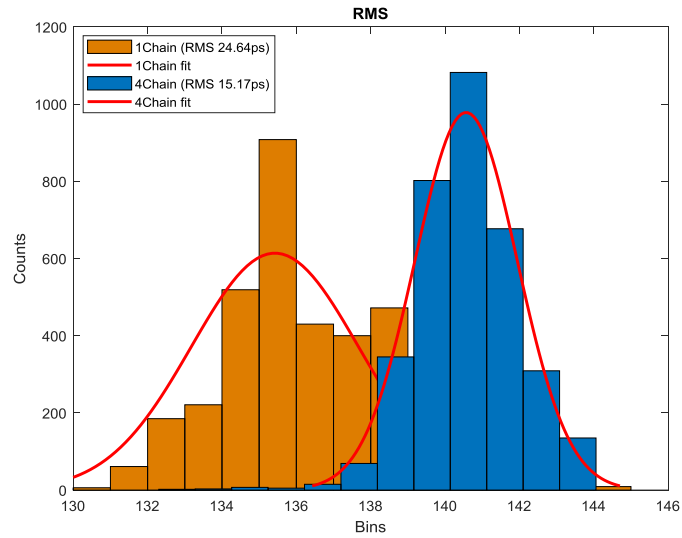


Figure 3:18 – 1 Chain and 4 Chain RMS comparison.

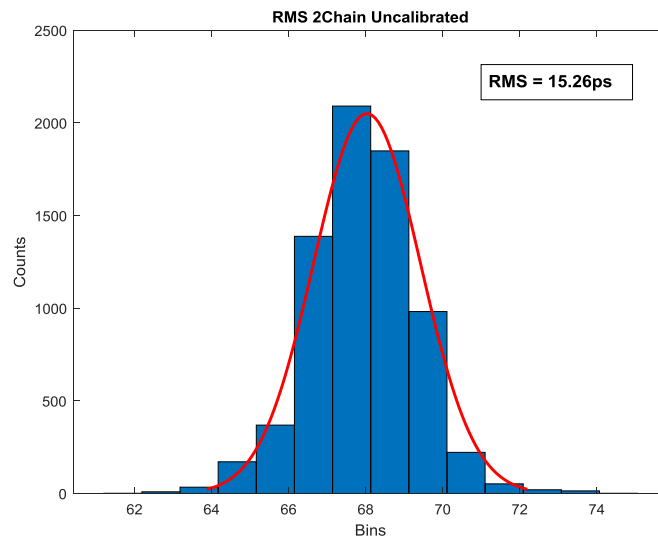


Figure 3:19 – 2 Chain RMS result.

3.1.7 Calibration

Use of the code density test to predetermine bin lengths allows for calibration of the arrival time measurement. Within an FPGA, even with replicated placement constraints

for each channel, slight inhomogeneous fabric differences exist causing nonlinearity. Once the bin sizes are known, each measurement can be converted from a bin count to a true time measurement. For the calibrated DNL and INL plots, instead of using an expected value of the mean bin size, the expected value becomes the code density measured bin length. Implementation involved use of a MATLAB application to first calculate bin sizes for each leg of the chain and then to combine into one averaged calibration result. Due to slight skew differences between legs, occasional overflow existed for one leg resulting in mismatch. An example would be 172, 184, 181, 3. A filter was set to remove these instances because the average would be skewed since the fourth leg had overflowed and passed its max bin count.

3.1.8 Multichannel Implementation

The method for multichannel implementation which resulted in the best linearity was to separate each leg of the chain by a minimum of 5 to 10 horizontal SLICES. This provided enough room between legs (and channels) for the place and route constraints to be met during implementation. A total of 32 channels were implemented with each channel consisting of four carry chains at a length of 200 delays. Only two channels were connected externally due to IO limitations on the KC705 evaluation board. Figure 3:20 is an enlarged view of the two tested channels.

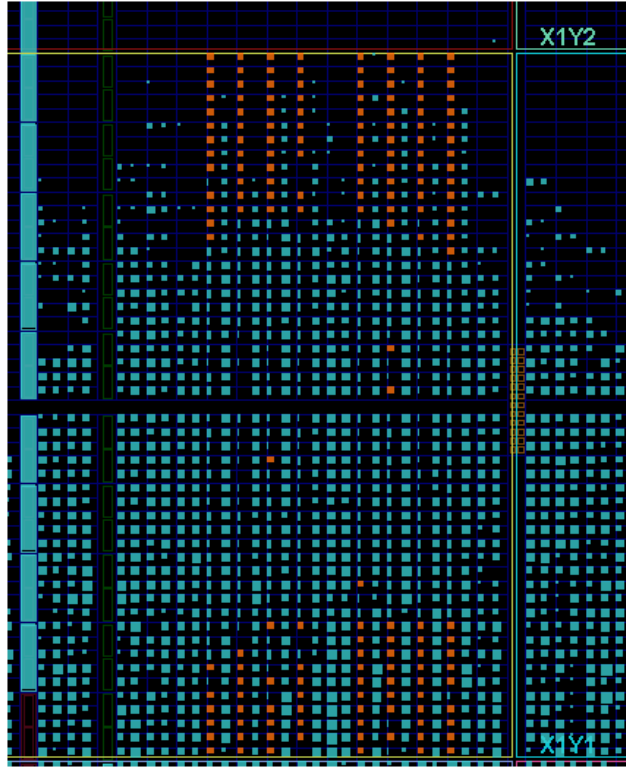


Figure 3:20 – Enlarged view of two channels with registers, latches, and encoders.

3.2 Pulse Energy Measurement

Implementation of pulse energy measurement involved use of the post-sampled analog integrator approach involving an ASIC designed by one of Dr. Jinghong Chen's PhD students, Yuxuan Tang from the University of Houston. A single channel tapeout was fabricated and connected to the KC705 evaluation board through a parallel bus on the Low Pin Count header. The tapeout includes one input connected to the output of a Silicon Photo-Multiplier. A current mirror feeds both a high-pass filter and an integrating capacitor. In the event that a pulse reaches the filter threshold, a comparator signals the FPGA TDC to take a fine measurement. The threshold is set to filter events below the 511keV energy of a true annihilation photon. When the ADC has finished conversion, the

TDC pulse de-asserts causing the FPGA to latch onto the ADC result and store along with the corresponding TDC channel coarse and fine measured time. The pulse length is around 400ns in duration. Figure 3:21 illustrates the ASIC detector design and Figure 3:22 shows the progression as the SiPM feeds the detector and the pulse is integrated and digitized before the FPGA latches on the result.

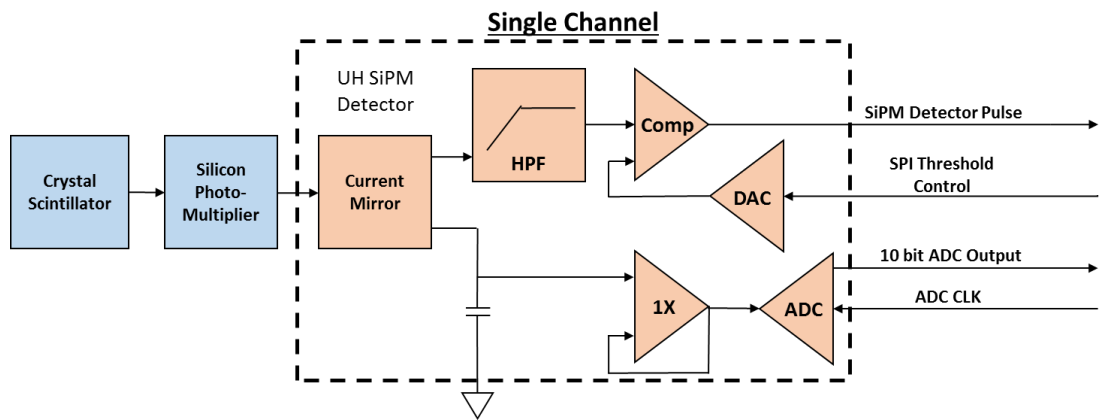


Figure 3:21 – UH designed SiPM detector.

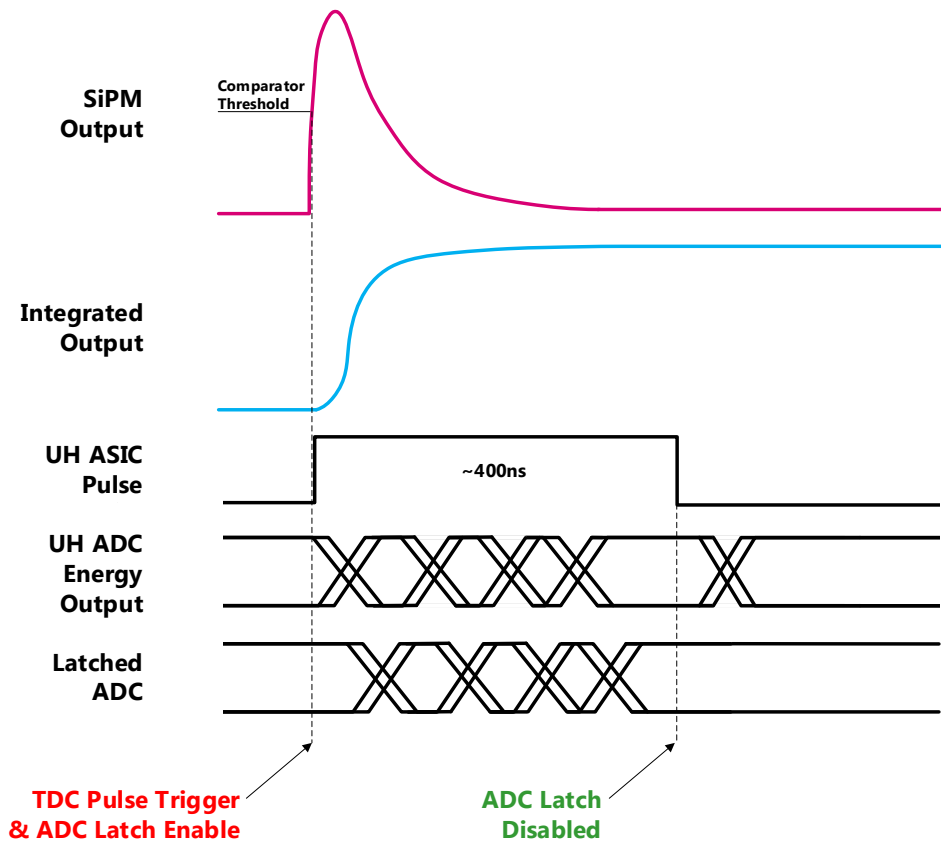


Figure 3:22 – Latched ADC measurement.

3.3 High Speed Transfer of Events

A comprehensive assessment regarding PET system requirements, available bus types, and path structure is necessary to successfully implement a high-speed DAQ communication system. The worst-case application for communication would be a whole-body PET since it requires transfer of the highest number of single and coincidence events per second. For this reason, the whole-body application is chosen for evaluation with a list of requirements captured in Table 3:3.

Table 3:3 – Proposed PET requirements.

Requirements	Whole-Body PET
Bore diameter (cm)	80
Measurement time (min)	50
Number of channels	1200
Coincidence timing window (ns)	<6
Timing resolution (ps)	100
Energy resolution (%)	10~16
Single event input rate (count/s)	<60M
Coincidence event throughput (counts/s)	<1M
# of coincidences for an image (events)	200M

Since the coincidence timing window is dependent upon the bore diameter, a quick calculation using equation 1.3 results in 5.33ns so 6ns is chosen. This will feed into the number of bits necessary in a packet. Before choosing a path structure, an assessment is needed regarding each packet size and the amount of bandwidth required. Table 3:4 and Figure 3:5 detail the single and coincidence packet structure respectively with the number of binary bits necessary for each element's representation. All elements are considered unsigned integers except where noted.

Table 3:4 – PET single event packet.

Single Packet Elements	Bits
Coincident ID (50min/6ns)	39
Channel ID (1,200 channels)	11
Event energy (UH design)	10
Fine time t (2ns coarse, 200 bin fine)	10
Clear signal fine time (200 bin)	8
CRC	16
Overhead	8
8b/10b encoding (20%)	21
Total	123

Table 3:5 – PET coincidence event packet.

Coincident Packet Elements	Bits
$\Delta t (t1 - t2 @ 10\text{bits each} + \text{sign})$	11
Event energy $\times 2$ (UH design)	20
Line of response (CH ID $\times 2$)	22
CRC	16
Overhead	8
8b/10b encoding (20%)	16
Total	92

Once the packet structure and bit count are determined, a calculation is necessary for the total bandwidth needed for each bus with a total of 32 channels per DAQ board as a given. Table 3:6 details the inputs to the calculation with the required bus speed of 198MHz for single and 92Mhz for coincidence. Now, a review of available bus types may be cross referenced along with those available to the Kintex 7 KC705 evaluation board to see if this speed is possible. Table 3:7 details many common bus types with those available to the KC705 noted and the available speed bolded. The maximum needed transfer rate of 198 Mbps can easily be achieved by the available gigabit ethernet or several GTX equipped interfaces like SFP+, common SMA coax, PCIe, or FMC connector.

Table 3:6 – Required bus speeds for single and coincidence transmission.

Packet Type	Bits	Total Events/s	Channels/ Bus	%Events /Bus	Min. Bus Speed
Single	123	60M	32	2.67	198Mbps
Coincidence	92	1M	1200	100	92Mbps

Table 3:7 – Common bus protocols and reference with KC705 availability.

Name	Speed	Available on KC705?
Ethernet (UDP)	100 Mbps, 1 Gbps , 10 Gbps	Yes, SGMII, Max 1Gbps
SFP+	4 Gbps	Yes, 1 lane GTX
PCIe	4 Gbps, 8 Gbps, 16 Gbps, 32 Gbps	Yes, gen 2, 8 lane GTX
SMA Coax	4 Gbps	Yes, 1 lane GTX
FMC HPC	16 Gbps	Yes, 4 lane GTX
FMC LPC	4 Gbps	Yes, 1 lane GTX
USB	12 Mbps, 480 Mbps, 5 Gbps	No, only slow speed UART
Thunderbolt	20 Gbps	No
SATA	1.5 Gbps, 3 Gbps, 6 Gbps	No

High end FPGAs with GTX transceivers are easily able to handle the maximum possible transmission rate necessary for PET DAQ systems. Any of the available data path structures discussed in section 2.4 are possible. The best choice for this application would likely be the daisy-chain method since it eliminates one board and is possible due to the high-density of Xilinx 7 series FPGAs.

4. Results and Discussion

A final DAQ block diagram is provided below within Figure 4:1. Use of a 4x multichain averaging technique with calibration implementation produced very favorable results. After calibration, the differential nonlinearity, Figure 4:2, resulted in less than +/- 4 picoseconds of error and the integral nonlinearity, Figure 4:3, resulted in less than 10ps of overall cumulative error. In comparison, no previous work reviewed had less DNL or INL error. Also, the final RMS resolution was 15ps which is much lower than the 100ps target and significantly lower than the 300ps resolution of most TOF PET scanners today. Table 4:1 provides a comparison between this work and 12 previous works.

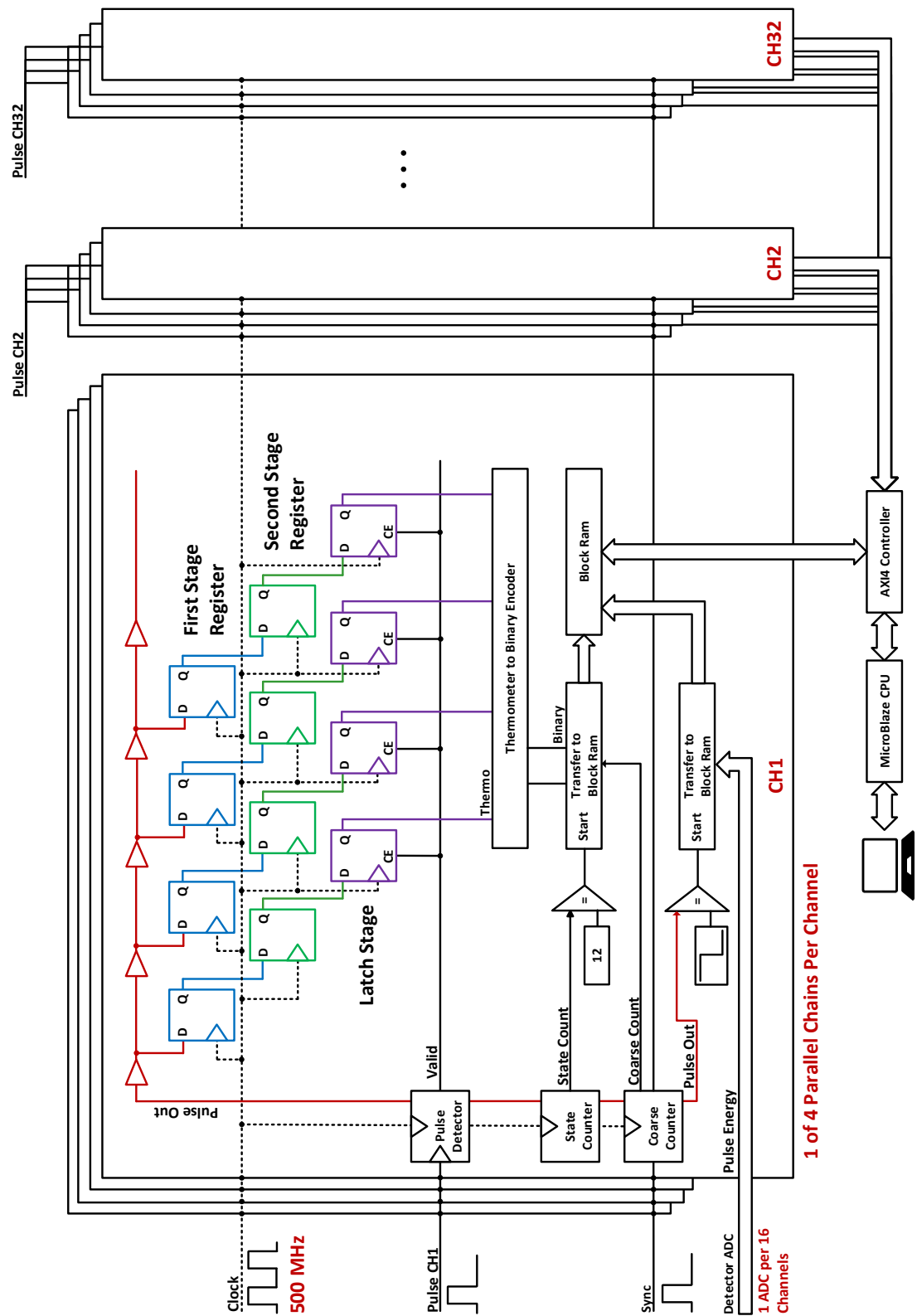


Figure 4:1 – Final design.

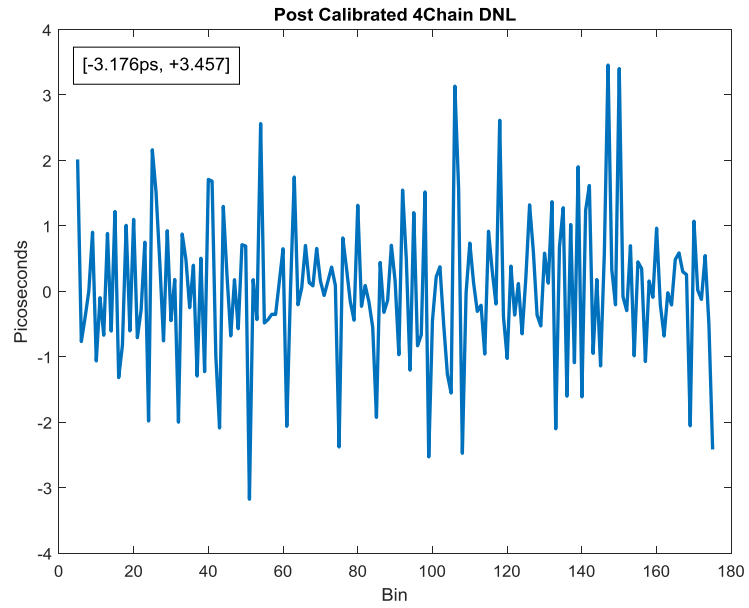


Figure 4:2 – Post-calibrated 4 Chain DNL.

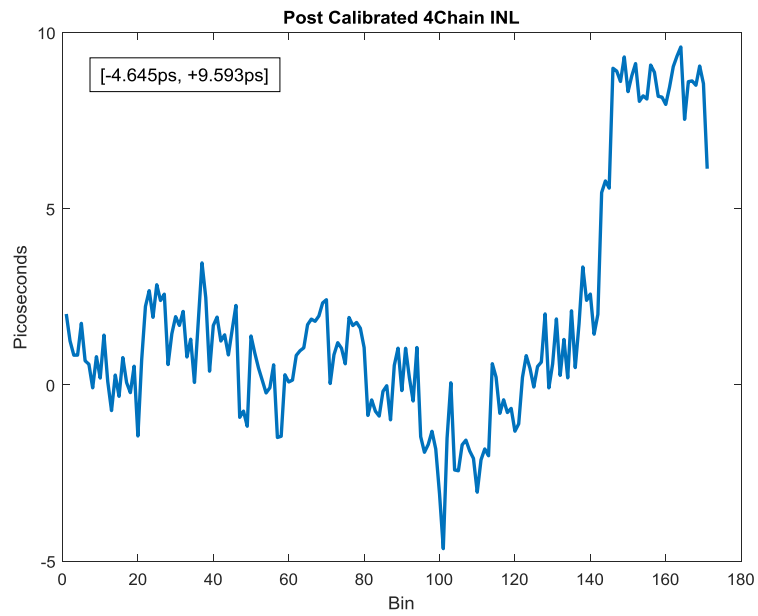


Figure 4:3 – Post-calibrated 4 Chain INL.

Table 4:1 – Comparison to previous work.

Ref.	Method	Year	RMS Resolution	Tech.	Integral Linearity (INL)	Non-Linear	Diff. Linearity (DNL)	Non-Linear
	This Work (Multichain averaging)	2019	15ps	28nm	-4.65ps, +9.59ps		3.18ps, +3.46ps	
[19]	Vernier TDL	1997	200ps	650nm	<200ps		-94ps, +88ps	
[20]	Vernier TDL	1997	129ps	650nm	46ps		-144ps, +214ps	
[22]	Multi-phase clocks	2012	625ps	65nm	31.25ps		31.25ps	
[23]	Pure Carry TDL	-	50ps	90nm	-		300ps	
[15]	Pure Carry TDL	2009	17ps	65nm	-51ps, +43.86		-17ps, +60.35ps	
[12]	Pure Carry TDL	2013	15ps	65nm	± 60 ps		-15ps, +45ps	
[21]	Multichain averaging	2015	4.2ps	40nm	-28.7ps, +18.2ps		-2.9ps, +11.72ps	
[24]	Matrix of counters	2017	7.4ps	65nm	11.6ps		5.5ps	
[25]	Wave union	2008	10ps	90nm	-		-	
[26]	Wave union	2013	7.7ps	65nm	-		-	
[27]	Ring oscillators	2008	40ps	90nm	<40ps		<40ps	
[28]	Ring oscillators	2017	50ps	350nm	± 65 ps		± 35.75 ps	

5. Summary and Conclusions

5.1 Summary

Positron Emission Tomography is a growing field with an increasing influence in nuclear medicine. Recent time-of-flight advancements have increased the cross-sectional image clarity and progression towards a miniature, silicon-based, photomultiplier has increased the resolution. As the photomultiplier size reduces, channel-count increases. Use of ASIC based Time-to-Digital converters with remote event processing is no longer viable due to the increasing cost and complexity. Instead, multichannel, FPGA based measurements with localized event processing is a must. Within this work, an FPGA based SiPM DAQ was presented with emphasis regarding the application of PET imaging. An introduction concerning PET concepts was first covered including a brief description of measurement techniques followed by a review of previous work with a detailed explanation of relevant methods. Next, an analysis of the problem covered specific implementation

details and difficulties followed by a comparison of results to previous work. Use of a 28nm, Kintex 7 FPGA based DAQ was examined with implementation of a 32 channel, multichain averaged TDC with 15ps RMS resolution, 11ps average bin size, and less than 10ps of integral nonlinearity.

5.2 Conclusion

Modern PET scanners no longer require costly ASIC based hardware for data acquisition. As progression moves towards higher and higher channel count with need for precise arrival time measurement, cheaper FPGA based DAQs are proving to be even better suited for the task. This research proves that modern FPGAs can handle at least 32 independent channels with enough timing precision for time-of-flight PET requirements and with built in communication hardware capable of transferring the maximum known load of single and coincidence event information.

References

- [1] L. E. Ketchum, M. V. Viola, P. C. Srivastava, P. T. Kirchner and L. S. James, "Nuclear Medicine, Converting Energy to Medical Progress," Medical Sciences Division of the Office of Biological and Environmental Research (BER) of the Office of Science of the U.S. Department of Energy (DOE), New York, New York, April 2001.
- [2] D. Fornell, "What is PET Imaging?," itn - Imaging Technology News, 26 July 2016. [Online]. Available: <https://www.itnonline.com/article/what-pet-imaging>. [Accessed 23rd February 2019].
- [3] R. Badawi, "Introduction to PET Physics," University of Washington, 12 January 1999. [Online]. Available: https://depts.washington.edu/nucmed/IRL/pet_intro/intro_src/section2.html. [Accessed 23 February 2019].
- [4] E. Kim, K. J. Hong, J. Y. Yeom, P. D. Olcott and C. S. Levin, "Trends of Data Path Topologies for Data Acquisition Systems in Positron Emission Tomography," *IEEE Transactions on Nuclear Science*, vol. 60, no. 5, pp. 3746–3757, October 2013.
- [5] R. E. Schmitz, A. M. Alessio and P. E. Kinahan, "The Physics of PET/CT scanners," [Online]. Available:

<http://depts.washington.edu/imreslab/education/Physics%20of%20PET.pdf>.

[Accessed 23 February 2019].

- [6] J. S. Karp, "Time-of-Flight PET," *Pet Center of Excellence Newsletter*, vol. 3, no. 4, pp. 1, 3-4, Fall 2006.
- [7] S. Surti, "Update on Time-of-Flight PET Imaging," *The Journal of Nuclear Medicine*, vol. 56, no. 1, pp. 98–105, January 2015.
- [8] A. A. Muntean, "Design of a Fully Digital Analog SiPM with Sub-50ps Time Conversion," Delft University of Technology, Delft, The Netherlands, 2017.
- [9] S. Dey, J. C. Rudell, T. K. Lewellen and R. S. Miyaoka, "A CMOS Front-End Interface ASIC for SiPM-based Positron Emission Tomography Imaging Systems," IEEE, Seattle, Washington, USA, 2017.
- [10] S. Piatek, "Silicon Photomultiplier - Operation, Performance & Possible Applications," [Online]. Available: https://www.hamamatsu.com/sp/hc/osh/sipm_webinar_1.10.pdf. [Accessed 23 February 2019].
- [11] S. Henzler, "Time-To-Digital Converter Basics," in *Time-to-Digital Converters*, Springer Science+Business Media B. V., 2010, pp. 5–18.

- [12] L. Zhao, X. Hu, S. Liu, J. Wang, Q. Shen, H. Fan and Q. An, "The Design of a 16-Channel 15 ps TDC Implemented in a 65 nm FPGA," *IEEE Transactions on Nuclear Science*, vol. 60, no. 5, pp. 3532–3536, 2013.
- [13] X. Liu, L. Ma, J. Xiang, NaYan, H. Xie and X. Cai, "A Low Power TDC with 0.5ps Resolution for ADPLL in 40nm CMOS," IEEE, Shanghai, 2015.
- [14] Y. Cao, P. Leroux and M. Steyaert, "Background on Time-to-Digital Converters," in *Radiation-Tolerant Delta-Sigma Time-to-Digital Converters*, Springer, 2015, pp. 15-23.
- [15] C. Favi and E. Charbon, "A 17ps Time-to-Digital Converter Implemented in 65nm FPGA Technology," *FPGA '09*, pp. 22–24, 2009.
- [16] J. Wang, S. Liu, Q. Shen, H. Li and Q. An, "A Fully Fledged TDC Implemented in Field-programmaable Gate Arrays," *IEEE Trans. Nucl. Science*, vol. 57, no. 2, pp. 446–450, Apr. 2010.
- [17] J. Doernberg, H. S. Lee and D. A. Hodges, "Full-speed Testing of A/D Converters," *IEEE J. Solid-State Circuits*, Vols. SSC-19, no. 6, pp. 820–827, Dec. 1984.
- [18] K. Cui, X. Li, Z. Liu and R. Zhu, "Towards Implementing Multi-Channels, Ring-Oscillator-Based, Vernier Time-to-Digital Converter in FPGAs: Key Design Points and Construction Method," pp. 1–9, 17 June 2017.

- [19] J. Kalisz, R. Szplet, J. Pasierbinski and A. Poniecki, "Field-Programmable-Gate-Array-Based Time-to-Digital Converter with 200-ps Resolution," *IEEE Trans. Instrum. Meas.*, vol. 46, no. 1, pp. 51–55, Feb. 1997.
- [20] J. Kalisz, R. Szplet, R. Pelka and A. Poniecki, "Single-chip Interpolating Time Counter with 200-ps Resolution and 43-s Range," *IEEE Trans. Instrum. Meas.*, vol. 46, no. 4, pp. 851–856, Aug. 1997.
- [21] Q. Shen, S. Liu, B. Qi, Q. An, S. Liao, P. Shang, C. Peng and W. Liu, "A 1.7 ps Equivalent Bin Size and 4.2 ps RMS FPGA TDC Based on Multichain Measurements Averaging Method," *IEEE Transactions on Nuclear Science*, vol. 62, no. 3, June 2015.
- [22] A. Balla, M. Beretta, P. Ciambrone, M. Gatta, F. Gonnella, L. Iafolla, M. Mascolo, R. Messi, D. Moricciani and D. Riondino, "Low Resource FPGA-based Time to Digital Converter," *To be submitted to IEEE Transaction on Instrumentation and Measurement*, pp. 1–7, 2012.
- [23] Y. Zhang, P. Huang and R. Zhu, "Upgrading of Integration of Time to Digit Converter on a Single FPGA," pp. 1–4.
- [24] M. Zhang, H. Wang and Y. Liu, "A 7.4 ps FPGA-Based TDC with a 1024-Unit Measurement Matrix," *Sensors* 2017, 17, 865, pp. 1–18, 2017.

- [25] J. Wu and Z. Shi, "The 10-ps Wave Union TDC: Improving FPGA TDC Resolution beyond Its Cell Delay," *IEEE Nuclear Science Symposium*, pp. 3440–3446, 19–25 Oct. 2008.
- [26] Q. Shen, L. Zhao, S. B. Liu, S. K. Liao, B. X. Qi, X. Y. Hu, C. Z. Peng and Q. An, "A Fast Improved Fat Tree Encoder for Wave Union," *Chinese Phys. C*, vol. 37, no. 10, p. 106102, Oct. 2013.
- [27] S. Junnarkar, P. O'Connor and R. Fontaine, "FPGA Based Self Calibrating 40 Picosend Resolution, Wide Range Time to Digital Converter," *2008 IEEE Nuclear Science Symposium Conference Record*, pp. 3434–3439, 2008.
- [28] A. A. Muntean, "Design of a fully digital analog SiPM with sub-50ps time conversion," *Delft University of Technology Masters Thesis*, pp. 1–67, 2017.
- [29] Xilinx Inc., "Vivado Design Suite 7 Series FPGA Libraries Guide (UG953)," 25 July 2012. [Online]. Available: https://www.xilinx.com/support/documentation/sw_manuals/xilinx2012_2/ug953-vivado-7series-libraries.pdf. [Accessed 2 Nov. 2018].
- [30] E. Kim, K. J. Hong, P. D. Olcott and C. S. Levin, "PET DAQ System for Compressed Sensing Detector Modules," *IEEE Nuclear Science Symposium and Medical Imaging Conference Record (NSS/MIC)*, pp. 2798–2801, 2012.
- [31] M. Nakazawa, J. Ohi, T. Furumiya, T. Tsuda, M. Furuta, M. Sato and K. Kitamura, "PET Data Acquisition (DAQ) System Having Scalability for the

Number of Detector," *IEEE Nuclear Science Symposium and Medical Imaging Conference Record (NSS/MIC)*, 2012.

- [32] M. E. Phelps, *PET Physics, Instrumentation, and Scanners*, New York, NY: Springer, 2006.
- [33] M. E. Casey and E. J. Hoffman, "Quantitation in Positron Emission Computed Tomography: 7. A Technique to Reduce Noise in Accidental Coincidence Measurements and Coincidence Efficiency Calibration," *Computer Assisted Tomography*, vol. 10, pp. 845–850, 1986.
- [34] "KC705 Evaluation Board for the Kintex-7 FPGA," 4 February 2019. [Online]. Available:
https://www.xilinx.com/support/documentation/boards_and_kits/kc705/ug810_KC705_Eval_Bd.pdf.
- [35] Z. Jaworski, "Verilog HDL Model Based Thermometer-to-Binary Encoder with Bubble Error Correction," *MIXDES*, pp. 249–254, 23–25 June 2016.
- [36] C. Liu and Y. Wang, "A 128-Channel, 710 M Samples/Second, and Less than 10 ps RMS Resolution Time-to-Digital Converter Implemented in a Kintex-7 FPGA," *IEEE Transactions on Nuclear Science*, vol. 62, no. 3, June 2015.

- [37] Y. Wang and C. Liu, "A 3.9 ps Time-Interval RMS Precision Time-to-Digital Converter Using a Dual-Sampling Method in an UltraScale FPGA," *IEEE Transactions on Nuclear Science*, vol. 63, no. 5, pp. 2617–2621, 2016.
- [38] G. Cao, H. Xia and N. Dong, "An 18-ps TDC Using Timing Adjustment and Bin Realignment Methods in a Cyclone-IV FPGA," *Rev. Sci. Instrum.* 89, 054707, 2018.
- [39] H. Homulle and E. Charbon, "Basic FPGA TDC Design," TUDelft, 2015.
[Online]. Available: https://cas.tudelft.nl/fpga_tdc/TDC_basic.html. [Accessed April 2019].
- [40] S. M. Visser, "A 1 GSa/s Deep Cryogenic, Reconfigurable Soft-core FPGA ADC for Quantum Computing Applications," Delft University of Technology, 2015.