Quantitative Measurement and Development of Back-end Processing System-on-Chip for the Pixelated CdZnTe Detector

by

Zhuo Chen

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Nuclear Engineering and Radiological Science)
in the University of Michigan
2022

Doctoral Committee:

Professor Zhong He, Chair
Professor Michael Flynn
Professor Igor Jovanovic
Associate Research Scientist Yuefeng Zhu

Zhuo Chen
chzhuo@umich.edu
ORCID iD: 0000-0001-5401-6006

Dedication

To my family.
To my friends.
To a better world.

Acknowledgments

My deepest gratitude is to my committee chair, my advisor, Professor Zhong He. His high-standard to research motivates me to seek excellence all the time. Every time we discuss, I am amazed by his insight to technology and attention to details. Throughout my Ph.D. career, he was a constant source of inspiration. Without his support, this work can never become possible. He is not only my academic advisor, outside the research he's also my life model.

I want to express my gratitude to my committee members, Professor Michael Flynn, Professor Igor Jovanovic and Associate Research Scientist Yuefeng Zhu, for serving on my dissertation committee. Professor Jovanovic taught me NERS515, where I gained the fundamental knowledge of this research field. Professor Flynn taught me so much about circuit design. His knowledge and expertise in the integrated circuit design amazes me. Dr. Yuefeng Zhu is one of the most hard-working people I've known. His engineering skills and knowledge of the radiation detection are something I kept learning from him in the past few years. Their feedback and insight are invaluable to this work.

Without the technological development and innovation from the members of the Orion Measurements Group, this work will never become possible. Their solid work and brave exploration in the past three decades laid the solid foundation and I feel lucky and honored to work around these smart and hard-working colleagues. The collaboration with current members was essential for my work and the quality of this thesis work were greatly improved from their critics. Not only are we colleagues, but we enjoy the leisure times as friends. I want to thank James Berry, Visiting Professor Gianluigi De Geronimo, Dr.Jiyang Chu, Dr.Sean O'Neal, Dr.David Goodman, Dr.Benette Williams, Dr.Charles Leak, Dr. Niral Shah, Dr.Valerie Nwadeyi, John Kuchta, Sara Abraham, Damon Anderson, Alexander Rice, Prabhjot Kaur and Peter Hotvedt. Especially, I want to thank Dr. Jiawei Xia, Dr. Daniel Shy, Matthew Petryk and Erik Hall. They helped me so much, both academically and non-academically.

My gratitude goes to my family, especially my father Zhizhan Chen and my mother Liping Wang. I truly own them so much. Their financial and emotional support lend me the courage to tackle any difficulty that's in front of me. We planned to visit each other at least once a year when I decided to start this journey, but due to various reasons out of our control, we were only able to unite twice in the past five and half years.

I thank all my friends. Together we visit the arboretum in the spring, float on the river in the summer, collect colorful leaves in the fall and walk in knee-deep snow in the winter. We encourage and comfort each other with long talks (sometimes late-night) about the uncertain future in this fast-changing world, the challenges we face both in research and life. We sincerely admire each other's achievement and share the joys. Star-gazing, grilling, skiing, cooking, gaming, etc. all make the life at Ann Arbor the golden age of my life. The support from the friends who I am not able to visit physically during these years are of equal importance. Though we may live in different time zones, we are always one phone call away.

At last, I want to thank myself, for having the determination to finish the journey. I vaguely remember that chilling early-spring night at the Tsinghua Garden, when the acceptance offer from Professor He reached my E-mail. At that time, the younger me could never imagine how fruitful this journey would become, nor how challenging it actually turns out to be. Five and half years later, in the beautiful campus of University of Michigan, surrounded by the fall foliage, the journey is about to end. Looking back, it's totally worth it.

Disclaimber: this material is based on work supported by the Defense Threat Reduction Agency under contract #HDTRA 1-18-C-0073. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Defense Threat Reduction Agency.

Table of Contents

List of Figures

List of Tables

List of Acronyms

ADC Analog-to-Digital Converter.

ASIC Application Specific Integrated Circuits.

BNL Brookhaven National Laboratory.

CMOS Complementary Metal Oxide Semiconductor.

CPU Central Processing Unit.

CZT CdZnTe.

DAC Digital-to-Analog Converter.

DRC Design Rule Check.

DSA Domain Specific Accelerator.

DSP Digital Signal Processor.

FEM Finite Element Method.

FIFO First-in First-out.

FPGA Field Programmable Gate Array.

FWHM Full-width at Half-maximum.

Ge Germanium.

GPIO General Purpose Input Output.

GPU Graphic Processing Unit.

GUI Graphical User Interface.

HDL Hardware Description Language.

HPGe High-purity Germanium.

IC Integrated Circuits.

IDEAS Integrated Detector Electronics AS.

IoT Internet-of-Things.

IP Intellectual Property.

LVS Layout-Schematic Comparison.

MOSFET Metal–Oxide–Semiconductor Field-Effect Transistors.

PEX Parasitic Extraction.

PMT Photo-Multiplier Tube.

RLC Resistance-Inductance-Conductance.

RTL Register Transfer Level.

SAR Successive Approximation.

Si Silicon.

SNM Special Nuclear Material.

SoC System-on-Chip.

SPI Serial Peripheral Interface.

TEI Time-encoded Imaging.

TSMC Taiwan Semiconductor Manufacturing Company.

UM University of Michigan.

VLSI Very-large Scale Integrated Circuits.

Abstract

The pixelated CdZnTe (CZT) detector is a promising alternative to the High-purity Germanium (HPGe) detectors in gamma-ray spectroscopy. Our research group at the University of Michigan has demonstrated that CZT detector can achieve close to HPGe energy resolution at room temperature [37]. Aside from the energy resolution, the detection efficiency is also of great importance. The efficiency characteristics of CZT detector using Integrated Detector Electronics AS (IDEAS) VAS/TAT system and Brookhaven National Laboratory (BNL)'s analog ASIC have been investigated previously [36], and the first half of this work focused on the efficiency characteristics of the latest digital ASIC-based systems.

The intrinsic detection efficiency of a 20mm×20mm×15mm CZT detector from 59 keV to 2614 keV is measured and compared with simulation results. A detailed simulation package modelling the physical processes, digital readout electronics and event reconstruction has been developed to explain the mismatch between measurements and simulation, in order to diagnose the cause of the efficiency deficit. Several efficiency loss mechanisms are revealed. The efficiency loss due to the guard ring and anode side inaccurate reconstruction are the most important factors. Other miscellaneous efficiency loss mechanisms are discussed too. This study presents the most accurate efficiency benchmark experiment on the digital ASIC-based system. It is the first time the details of the efficiency response of CZT crystal are understood.

The hardware design of current detectors are based on the concept of "system on board". The discrete electrical components are integrated on the board-level. The dimension of the assembled systems and the overall power consumption are usually not optimal. With careful engineering, the disadvantages of "system-on-board" can be minimized, but after all, these hardware are not tailored to the specific application. The room for further improvement and optimization are limited. Similar to how any modern technologies evolve, the constant desire to increase the performance, decrease the

physical dimension and reduce the power consumption requires deeper-level integration. The SoC, where components are integrated on a single silicon die, becomes the most suitable solution for the next-generation radiation detector.

A back-end data processing SoC for CZT was designed at University of Michigan (UM) and named as the DSP/ADC ASIC. The SoC methodology provides the convenience to combine different hardware Intellectual Property (IP)s together on a single piece of silicon, therefore provides extra room for power-performance-area (PPA) improvement. The area of the final silicon die is 4.32mm×4.32mm and it is packaged into a 60-pin QFN package. An ASIC design workflow and the script system were built around various design software programs. The workflow covers the full life cycle of the ASIC design and can be used in the future projects. Two hardware platforms and a data acquisition software are developed to test the fabricated DSP/ADC ASIC. Several problems are identified during the test. Their cause are understood with the help of simulation and will be fixed in the next design iteration.

1 Introduction

1.1 Radiation Detection

1.1.1 Physics of Radiation

Radiation, in a more general sense in Physics, is the emission or transmission of energy through space in the form of either waves or particles. Detecting the emission or transmission of radiation allows one to characterize the property of source or that of the transmission medium.

The most common "radiation detector" is our eyes. They detect the visible light, which is a form of electromagnetic radiation. Although the underlying physics and biology mechanisms are not yet fully clear, they are sensitive to electromagnetic waves whose wavelength range from about 380nm to about 750 nm. Measurement of the incident direction and energy (color) of light allows sighted human beings to discern the world spatially.

However, visible light is only a narrow fraction of radiation. Radiation in the form of waves includes radio waves, microwaves, infrared, visible light, ultraviolet, x-rays, and gamma radiation. Radiation in the form of particles includes alpha radiation (α), beta radiation (β), proton radiation, and neutron radiation. Most of these radiations do not trigger the sensory system of human body, and therefore is invisible to us. For example, the energy of x-rays and gamma-rays are several order of magnitude larger than these visible photons (Figure.1.1). They are important, but they are impossible to be sensed by humans directly. On the one hand, it is desired to utilize the physics of these radiations in order to build tools around them. On the other hand, it's of significant importance to know where the radiation is and avoid its potential hazards to human body. Therefore, to aid humans to detect and measure radiation events, various

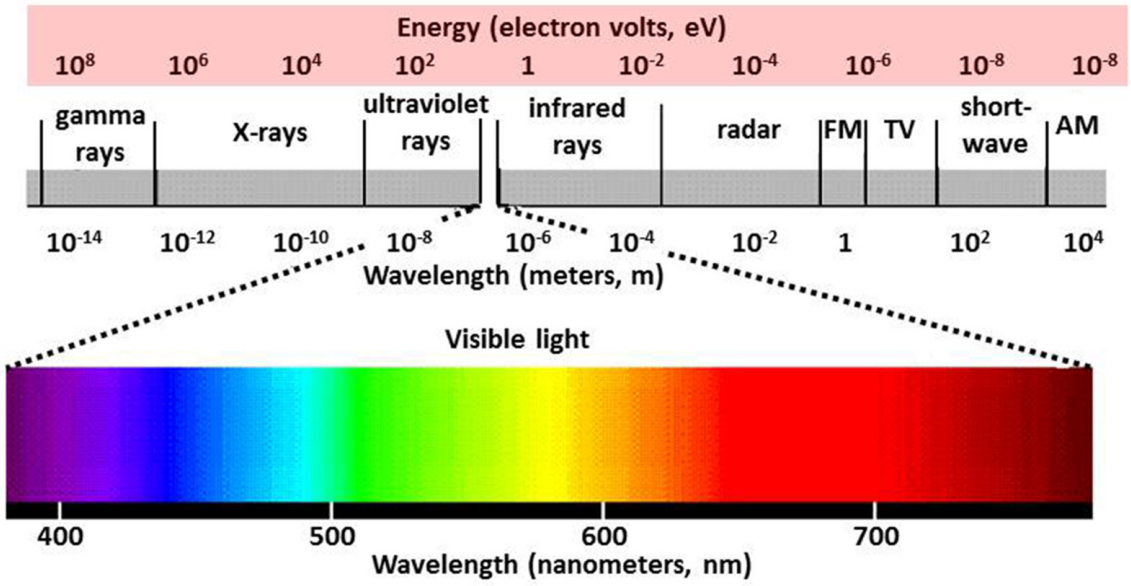

Figure 1.1: Gamma rays and X-rays are both electromagnetic radiation. The electromagnetic spectrum shows the overlap of frequency between X-rays and gamma rays

instruments and algorithms have been developed in the past century.

The radiation of primary concern in the scope of the nuclear detection and measurement field originate in the atomic or nuclear processes. They can be categorized into four general types[14].

1. Fast electrons include beta particles (positive or negative) emitted in nuclear decay, as well as energetic electrons produced by any other process.

2. Heavy charged particles denote a category that encompasses all energetic ions with mass of one atomic mass unit or greater, such as alpha particles, protons, fission products, or the products of many nuclear reactions.

3. The electromagnetic radiation of interest includes X-rays emitted in the rearrangement of electron shells of atoms, and gamma rays that originate from transitions within the nucleus itself.

4. Neutrons generated in various nuclear processes constitute the final major category, which is often further divided into slow neutron and fast neutron subcategories.

The main topic in this study will revolve around gamma rays. In 1905, Einstein was the first to propose that energy quantization was a property of electromagnetic radiation itself and therefore can be regarded as flux of point-like quanta, a.k.a. photons (later known as the wave-particle duality). In this sense, a gamma ray can also be abstracted as a high-energy photon particle.

1.1.2 Photon Interactions with Matter

Photons have zero rest mass, have no charge, and travel at the speed of light. Only when they interact with matter can we detect and further reveal their information. Such properties and their notorious negative effects to the human body usually scares the general public.

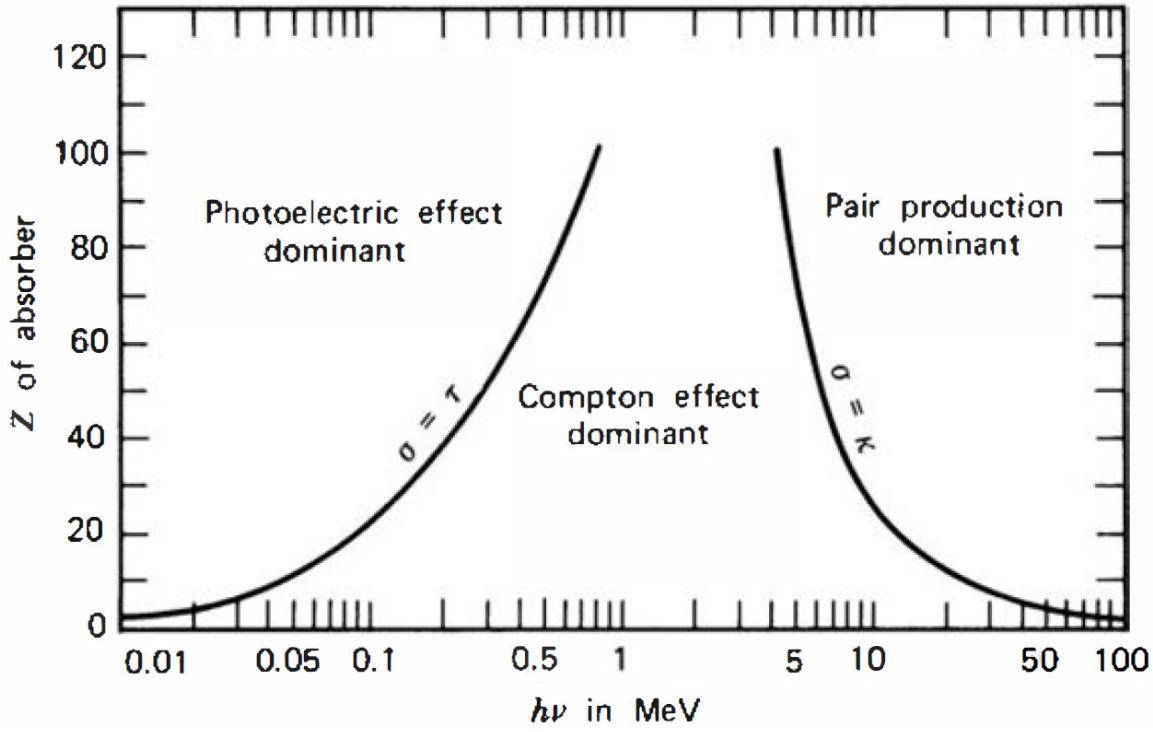

Figure 1.2: The relative importance of the three major types of gamma-ray interaction. The lines show the values of Z and hv for which the two neighboring effects are just equal [14].

Photons interact with matter and deposits energy through three major mechanisms,

1. Photoelectric Absorption, in which a photon interacts with an atom of the absorbing material, and the photon completely disappears; its energy is transferred to one of the orbital electrons of the atom.

2. Compton Scattering, in which the photon scattering off a stationary charged particle (usually an electron), results in a decrease in energy of the photon.

3. Pair Production, in which the photon energy exceeds twice the rest-mass energy of an electron (1.02 MeV), the photon disappears and is replaced by an electron-positron pair.

The relative importance of the three major types of gamma-ray interaction are well depicted in Figure.1.2[14]. No matter which interaction the photon undergoes, part (or all) of its energy is transferred to a charged particle.

1.1.3 Charge Induction

If the detection medium is gas, the travelling energetic charged particle will ionize the gas particles along the track, leaving electrons and positively-charged nuclei (hole) behind. The shape of the electron-hole clouds would resemble the particle trajectory in the famous Wilson Cloud Chamber, which was invented in 1930s by the Scottish physicist Charles Wilson. In the Wilson Chamber, the ionized particles act as condensation centers, around which the supersaturated vapour condenses. In later gas-based detectors, the electron-hole clouds can be sensed electrically (e.g. proportional counter).

If the detection mediums is semi-conductor, like Silicon (Si), Germanium (Ge), or CZT, the physics picture is similar. Recall in semiconductors, each atom shares its valence electrons with its four neighbors in a paired configuration called a covalent bond. Quantum Mechanics (QM) predicts that the allowed energy levels of electrons in a solid are grouped into two bands, called the valence band (VB) and the conduction band (CB). When the charged particle moves through the detection medium, it transfers energy to electrons through electromagnetic force, exciting some electrons to jump from the valence band to the conduction band. This leads to a free electron and a positively-charged vacancy (hole), therefore, the net effect is the generation of electron-hole pairs. Clusters of these electron-hole pairs are called electron-hole clouds.

If the clouds stay where they were generated, these opposite charges will eventually recombine and disappear under the influence of static Coulomb Force. To separate and sense them, an electric field must be applied. Under the influence of the applied electric field, the electron-hole pairs are separated and drift inside the detection medium at a constant speed $v = \mu E$ after a quick speed saturation. During their motion, the charge induced on the electrode can be measured, and then used to reconstruct the original gamma interactions.

The whole charge induction process can be understood by considering an example, where a point charge q is placed some distance away from a perfectly conducting metal plate. The electric field on the surface of the plate can be analytically solved. A surface integration over the plate would yield the total induced charge on the plate, which is proportional to the point charge. By sensing how much charge resides on the metal plate, we can tell how much charge the point charge carries. Therefore, even if we lack a method to directly measure the original charge, we can still reconstruct it by observing the induced charge on the electrode indirectly.

To summarize, the principle of detecting photons (either x-ray or gamma-ray) is the fact that the photons need to deposit energy through one of the three interaction mechanisms within the medium, generating charges q. Under the applied electric field, the charge q moves and the induced charge Q on the electrode changes. It is the measurement of this induced charge Q on the electrode that the detection system senses and processes.

1.2 The Shockley-Ramo Theorem

In detectors whose electrode configuration are complicated (co-planar grid or pixellated), even when the material is ideal, it's hard to solve for the induced charge on each electrode by brute-force surface integration. If the non-uniformity of material or electrode geometry are considered, it's fair to claim it's impossible to solve it analytically. The Shockley-Ramo Theorem provides a simplistic method to calculate the induced signal on an electrode due to the motion of the charge carriers [9].

The Shockley-Ramo Theorem states that, for a moving charge, q, the charged induced on the electrode, Q, can be calculated as,

$$\Delta Q = -q\phi_0(\vec{x}) \tag{1.1}$$

where $\phi_0(\vec{x})$ is the unit-less weighting potential profile of the negatively induced charge by the moving charge carrier. Taking first order derivative again time t, the same theorem can be applied to the instantaneous induced current i on the electrode.

$$i = -q\vec{v} \cdot \vec{E}_0(x) \tag{1.2}$$

where $\vec{v}$ is the instantaneous velocity and $\vec{E}_0(\vec{x})$ is the weighting field potential. The weighting potential and weighting field are not real potential or field, but only theoretically exist to ease the calculation of induced charge on any electrode.

The weighting potential ϕ_0 and weighting field $\vec{E}_0$ are the electric potential (field) when only the selected electrode of interest is at unit potential (1 Volt) while all others are at zero potential (0 Volt), and all space charges are removed. Mathematically, they are the solution to the following Poisson equation groups with a special set of boundary conditions,

$$\begin{cases} \nabla^2\phi(x) = 0 & (1.3) \\ \phi|_{S_i} = 0V,\ i \neq k & (1.4) \\ \phi|_{S_i} = 1V,\ i = k & (1.5) \end{cases}$$

where electrode k is our electrode of interest and S_i denote the surface of the i-th electrode. Therefore, the change of the induced charge ΔQ_L on an electrode L due to the movement of charge q from position $\vec{x_i}$ to position $\vec{x_f}$ is,

$$\Delta Q_L = \int_{x_i}^{x_j} q\vec{E_0}(\vec{x})dx = -q[\phi_0(x_f) - \phi_0(x_i)] \tag{1.6}$$

where $\vec{E}_0$ is the weighting electric field, and ϕ_0 is the weighting potential.

Most modern detection systems record and base their processing algorithms on $Q_L(t)$, not ΔQ_L. Because the change of the induced charge ΔQ_L is independent from the actual electric field and the space charges. It is only determined by the starting position x_i and stopping position x_j of the charge carriers relative to the electrode L. The waveform $Q_L(t)$ carries a lot more information. Since its profile depends on the trajectory the charge clouds went through, the trapping inside the material, its thermal diffusion

or material non-uniformality etc. By examining $Q_L(t)$ closely, information about the detection medium, the charge cloud distribution, etc. can be revealed. To examine $Q_L(t)$ (how Q_L changes with time t), the drift trajectory $\vec{x(t)}$ must be solved.

The drift trajectory of the charge $\vec{x(t)}$ from point x_i to point x_j is and only is determined by the actual operating electric field and the space charges. The operating field can be calculated from the same Poisson equation groups but with the boundary conditions that describes the actual electrode configuration.

To summarize, if the weighting potential (or field) and the operating field are solved, the induced charge can be calculated on any specific electrode at any time easily, even if the electrode configuration is complicated. This greatly simplifies the detector system design and analysis.

1.3 Motivation and Outline of This Work

This Ph.D. thesis work covers two major topics.

The first topic is the study on the efficiency performance of the radiation detection system built from 20-mm CZT crystals. The intrinsic photopeak efficiency of a single 20mm×20mm×15mm CZT detector from 59 keV to 2614 keV are measured in an experiment. However, compared to the simulated ideal intrinsic photopeak efficiency, the measured efficiency showed deficit. A detailed simulation model including gamma-ray interaction physics, readout electronics as well as the signal reconstruction process is built to explain this observed efficiency deficit. The simulated efficiencies of the new model match with experimental results well after taking these factors into account. Efficiency loss due to the guard ring and anode side inaccurate reconstruction are revealed. Other miscellaneous efficiency loss mechanisms are discussed too. This is the most detailed study on the efficiency of CZT detectors as of the writting of this work and could be the foundation of future quantitative detection study.

Intrinsic photopeak efficiency of pixelated CZT detectors is an important property for quantitative measurements and analysis, especially for detection and characterization of Special Nuclear Material (SNM). The recent advancement in TEI enabled 3-D localization of radiation sources with depth refocusing. A quantification demo experiment was performed to showcase the quantitatively measurement capability of the Orion detection system.

The second topic is the development of a back-end readout SoC. The discrete electric components are integrated into a single silicon die with small form factor and an order of magnitude decrease in power consumption. SoC is an ASIC that integrates the entire functional system. It is becoming increasingly popular with the growth of Internet of Things (IoT) and mobile computing due to various advantages. As we essentially size down what is normally a multi-chip design onto a single silicon substrate, the space of the whole system can be reduced when compared to their multi-chip counterparts. Besides, since theSoC is designed and optimized to finish a very specific task, it's generally faster and less power-hungry than the general-purpose circuits. What's more, since the communication among modules is on-chip, the bandwidth of data transfer and processing can be improved too.

The architecture of the DSP/ADC ASIC and the design work flow are elaborated. Later, the test sytem and the current test results are presented. Several problems are identified in the first tape-out iteration. They are presented and discussed in this work too.

2 Overview of the Current Technology

This thesis work aims to advance the detection technology based on CZT. It's important to review the fundamental knowledge in the related field so that the following work can be better understood. This chapter will illustrate the current detection system and the modern IC development methodology on which this work based.

2.1 The Detection System

Detectors are the bridge between the unknown particle and the humans in the particle physics research field. With detectors, scientists are able to observe and record the physical interaction between particles and matter, which can't be sensed by humans directly. The evolution of radiation detection has come a long way and the frontier of the research field is still advancing. The constant desire to achieve better energy resolution, to reduce the form factor and to decrease the power consumption of the system drive the development of crystal, the ASIC and the processing algorithms. This section will review the detection system[1] at the UM.

2.1.1 CdZnTe Crystal

CdZnTe (CZT) is a semiconductor material with a band-gap of about 1.6 eV. The band-gap of CZT is larger than that of HPGe detectors, which implies that in CZT, the energy required to excite electrons to the conduction band is greater compared to that of the

[1] In the context of this work, a detection system refers to a combination of the detection medium (the semiconductor material), the readout electronics (ASIC, its peripheral electronics) and the compatible reconstruction algorithms.

HPGe (built with Ge). The probability of thermal excitation in CZT is much lower due to this, allowing CZT systems to operate at room temperature.

The energy deposition in the CZT detector will excite a number of electrons into the conduction band. The number of electron and hole pairs is approximately proportional to the deposited energy. If a strong electric field is applied in the crystal, the excited electrons and the positively charged holes will be separated and they drift towards the opposite electrodes. Their motion induce signals on the anode and cathode electrodes. As described in the Chapter I, it is possible to recover the position and the energy information of this interaction from the signals.

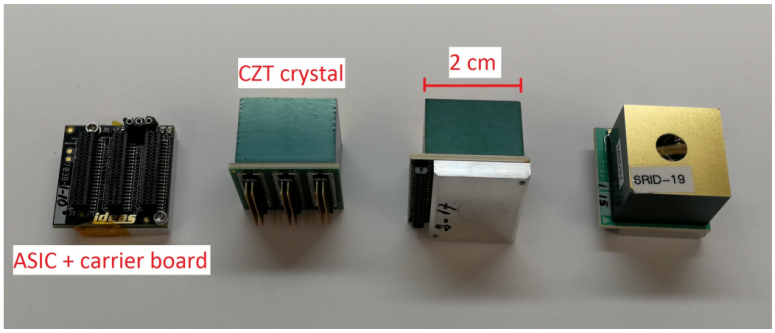

Figure 2.1: From left to right, a VAD_UM v2.2 ASIC mounted on the carrier board, a 20mm×20mm×15mm CZT detector, a direct-attachment CZT and ASIC module (side view) and a direct attachment detector module (top view).

The CZT crystals used in this thesis are cuboids of 20mm×20mm×15mm (Figure.2.1). The square side is covered with a planar cathode electrode, and the opposite side is attached to an 11×11 array of anode electrodes (Figure.2.2). The anode electrodes form pixels with a pixel pitch of 1.72 mm. To improve charge collection, many detectors have a negatively-biased grid electrode between the anode electrodes. Each of the 123 electrodes in one crystal is read out through a single VAD_UM v2.2 ASIC. A detector system may contain multiple crystals to form a detector array.

The development of CZT has been going on for more than three decades. In the 1990s, 10mm×10mm×10mm crystals were successfully grown and were used to build radiation detectors. In the early 2000s, 15mm×15mm×10mm CZT crystals became available and were studied. In the late 2000s, the crystal growth technique continued to improve. Later the detector size was standardized to 20mm×20mm×15mm. Over 300 of the

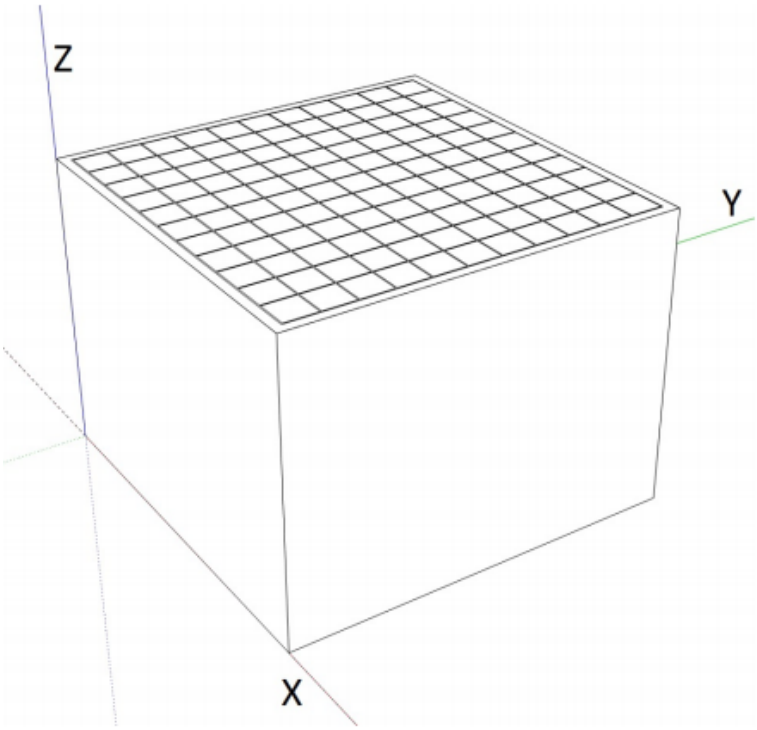

Figure 2.2: Simplified scheme of a pixelated CZT detector. The size of the crystal is 20mm×20mm×15mm. One side of the crystal has 121 pixelated anodes[29]. The other side is a planar electrode. The coordiante axis are valid throughout this study.

standard-sized detectors were delivered by Redlen Technologies and eV-Products to UM between 2009 and 2022. Beginning in 2019, large crystals of size 40mm×40mm×15mm have been fabricated and tested (Figure.2.3). As of 2022, the best energy resolution performances was reported to be near 0.50% FWHM at 662keV in a batch of crystals that were tested in early 2020.

Though CZT has excellent performance, it is hard to grow in large volume and contains rare-earth element. Both factors make CZT expensive. The search for the next generation radiation detection medium has been going on for decades at the same time as the development of CZT-based systems [15][18]. Recently, due to the improvement of crystal growth technique, perovskites can already produce about 1% FWHM energy resolution at 662keV, which seems to be a promising candidate in the future.

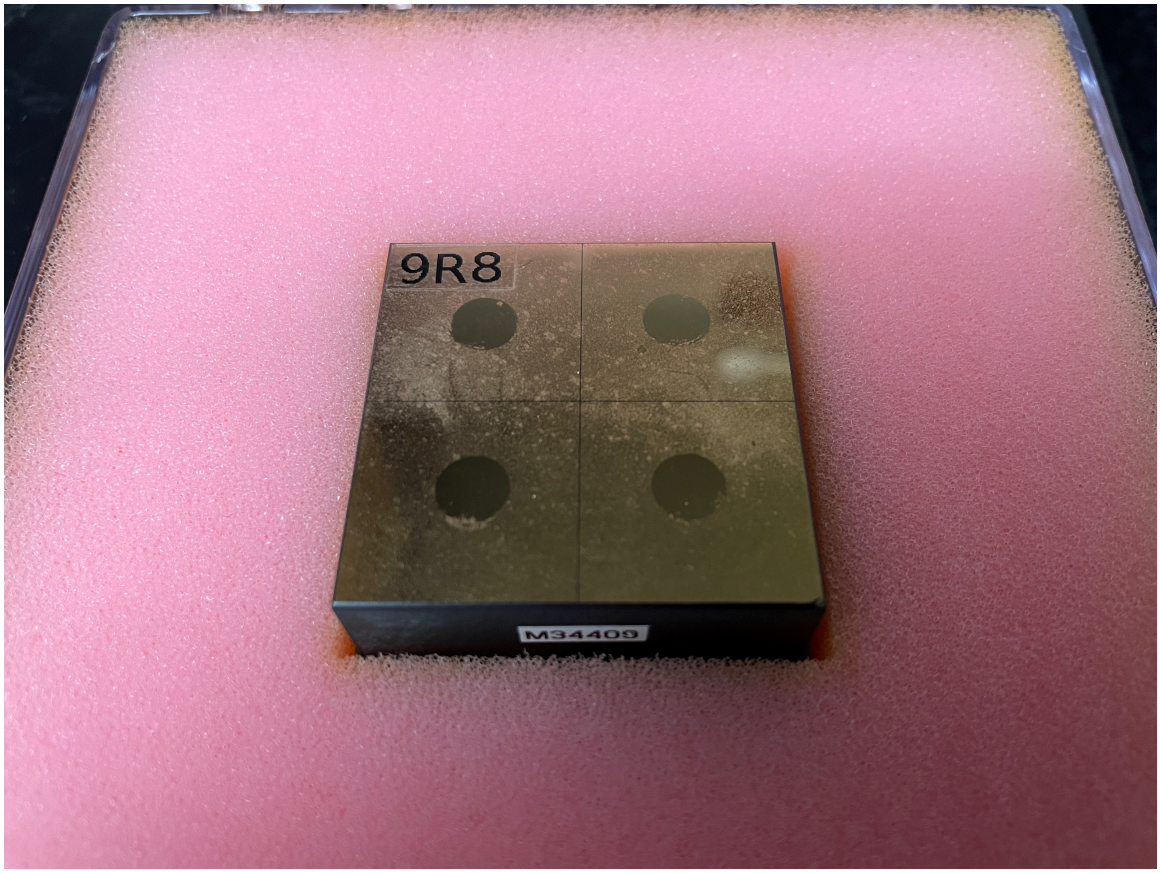

Figure 2.3: One 40mm×40mm×15mm CZT (ID No.9R5) received from Relden in May 2022.

2.1.2 The Read-out ASICs

In earlier models of radiation detectors, the number of signal channels was limited to only a few, therefore, the corresponding readout electronics are fairly simple. However, as radiation detectors evolved into the pixelated style, the number of channels that the readout electronics needed to handle increased to more than one hundred. Each anode requires one electronic channel to read out its signal. Ideally, the anode signal can be routed onto the circuit board and then sensed by electronics, however, engineering experiences have shown the longer trace length induce extra noise to the signal and is undesirable [38]. Therefore, there's motivation to design the high-performance circuits in a dense area as close to the electrodes as possible. The ASIC, due to its fast speed, compact size and low-power, is the most suitable solution to this challenge.

Over the past two and half decades, several family of ASICs have been developed by UM and the collaborators. These ASICs generally are capable of reading out the signals from 123 channels (121 anode pixels channels, 1 cathode channel, and 1 guard ring channel).

The ASIC in earlier years was analog ASIC, who uses on-chip passive components (resistors and capacitors) to achieve waveform filtering. Only limited information (e.g. amplitude and timing) can be readout and processed. In 1998, the first ASIC readout system for CZT was delivered (VA1 from IDEAS). The energy range of interest is from a few keV to 1 MeV. 1.75% FWHM energy resolution for single-pixel events at 662 keV was achieved on a 10mm×10mm×10mm cubic CZT crystal [16]. In 2004, another ASIC generation was designed by UM and IDEAS (VAS2/TAT2). Each chip has 32 channels and four chips are needed for each 121-pixel CZT detector. 1.11% FWHM energy resolution for single-pixel events at 662 keV was reported on a 15mm×15mm×10mm CZT crystal [35]. One year later, the next version further improved the energy resolution performance to 0.76% [34].

Around the same time, UM worked together with the BNL to develop another analog ASIC. Systems with this ASIC have successfully demonstrated sub-0.5% FWHM resolution at 662 keV for single-pixel events on a 20mm×20mm×15mm CZT crystals[36].

The next family of ASIC are the VAD_UM ASICs, which marks the start of the development of the digital readout ASICs. The VAD_UM v1.0 ASIC was delivered in 2010. It had excessive electronic noise and several design flaws. The re-designed VAD_UM v1.2 ASIC was delivered in 2011 and demonstrated 0.41% FWHM energy resolution for single-pixel events at 662keV [39]. The VAD_UM v2.0, which was delivered in 2015, had several dynamic ranges that can be selected by users based on their need. The most matured version, VAD_UM v2.2 ASIC, was delivered in 2016. This version has showed about 0.4% FWHM resolution at 662 keV for single-pixel events [37].

After the VAD_UM v2.2, the development of the next-generation high-performance readout ASIC continued. Starting from late 2017, the design of the new H3DD_UM ASIC started. As of the writting of this work, four generations have been fabricated and tested. The H3DD_UM v1.0 was delivered and tested in 2018 but unfortunately was not working properly. The H3DD_UM v2.0 attempted to fix the problems in its predecessor and was fabricated in 2019. The noise performance was worse than expected due to the CMOS gate leakage and clock distribution. The H3DD_UM v3.0 was initially fabricated

in June 2020, but the thick-oxide transistor in the design wasn't fabricated correctly due to a foundry mistake. The correctly fabricated H3DD_UM v3.1 was delivered in March 2021. It had excessive noise, leading to poorer energy resolution than expected. Other problems like the long readout settling time and the 16-sample spiking problem in the output waveform triggered another round of revision and tape-out. The test of the most current H3DD_UM v4.0 shows that it can achieve 0.40% FWHM energy resolution at 3MeV HF (high-gain) and 800keV LF (low-gain) dynamic range.

Looking back at the history of the readout ASIC, it's clear that each generation of high-performance ASIC can take years to design and mature. Although complicated to develop, once matured, the performance of the whole system is almost guaranteed to take a leap forward.

2.1.3 The Current System

Two detection systems, used throughout this work, are called the VAD_UM and the Orion-α system, respectively. The VAD_UM system was manufactured by IDEAS, and the Orion-α is developed by our research group [29]. They both can house a 3×3 detector array and are based on the VAD_UM v2.2 digital ASIC [29]. The dimensions of CZT crystals used in this study are 20mm×20mm×15mm (Fig.2.1)[37]. The anode side has 11 by 11 anode pixels with a guard ring anode surrounding them (Fig.2.2). The pixel pitch is 1.72 mm between neighboring pixel anodes (Fig.2.4). A more detailed description to these two system can be found in reference [29].

When in operation, the cathode is biased to -3000V, while the anodes are grounded. The high voltage on the cathode is provided by a high voltage generation circuit board covering the whole cathode surface. The guard ring can be biased at a small negative voltage in current system, so that events under it would be steered into the edge pixels, but this usually degrades the energy resolution of edge pixels. In this study, the guard ring was grounded to achieve best performance in energy resolution. Under this configuration, the guard ring is equi-potential with the anode pixels during operation.

The signals on the anodes and the cathode are read out by the VAD_UM v2.2 digital ASIC (Fig.2.1). The digital ASIC is directly attached to the CZT crystal through a carrier board by Redlen. Direct attachment yields superior noise performance as the electronic noise can be reduced to about 1.5 keV (FWHM). The energy resolution of

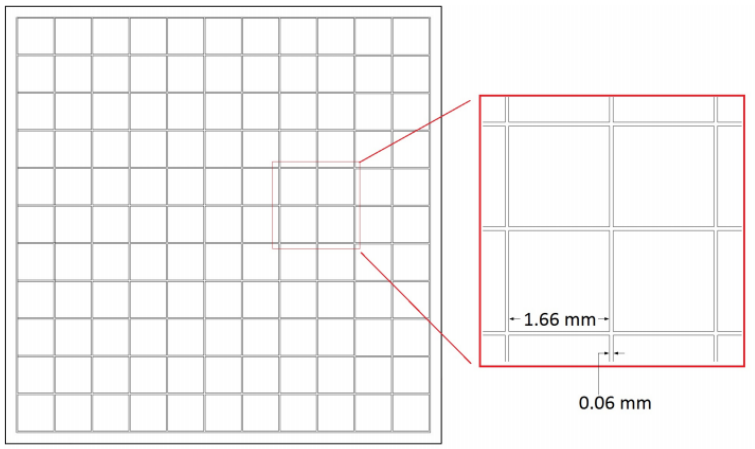

Figure 2.4: Top view of the anode side of a pixelated CZT detector. Each pixel pitch is 1.66×1.66mm and the width of gap between two neighboring pixel is 0.06mm[29].

single pixel events can reach better than 0.4% FWHM energy resolution at 661.7 keV. The ASIC samples the preamplifier output signal on electrode with a 40 MHz sampling frequency. The sampling window on each readout channel for a single-pixel interaction is 4 μs and 160 samples are recorded.

2.2 The Integrated Circuits (IC)

An integrated circuit (also referred to as a chip, or a microchip) is a set of electronic circuits built on one small piece of semiconductor material (usually Si). Large numbers of tiny Metal–Oxide–Semiconductor Field-Effect Transistors (MOSFET) are integrated into a small chip. The integrated circuits are orders of magnitude smaller, faster, and less expensive (when produced in large volume) than those constructed of discrete electronic components. The IC's mass production capability and reliability has ensured the rapid adoption of standardized ICs in place of designs using discrete transistors.

After decades of development, ICs have shaped the landscape of modern world. ICs are now used in virtually all electronic equipment nowadays and have revolutionized the world of electronics. Assembled together, silicon chips constitute hardware platform (personal computers, mobile phones, cloud servers, and other home appliances) on which various services ran (Figure.2.5). They are all made possible by the small size and low

cost of IC such as modern computer processors and micro-controllers. ICs have become inextricable parts of the structure of modern societies and the backbone of the constantly evolving modern technology and economy.

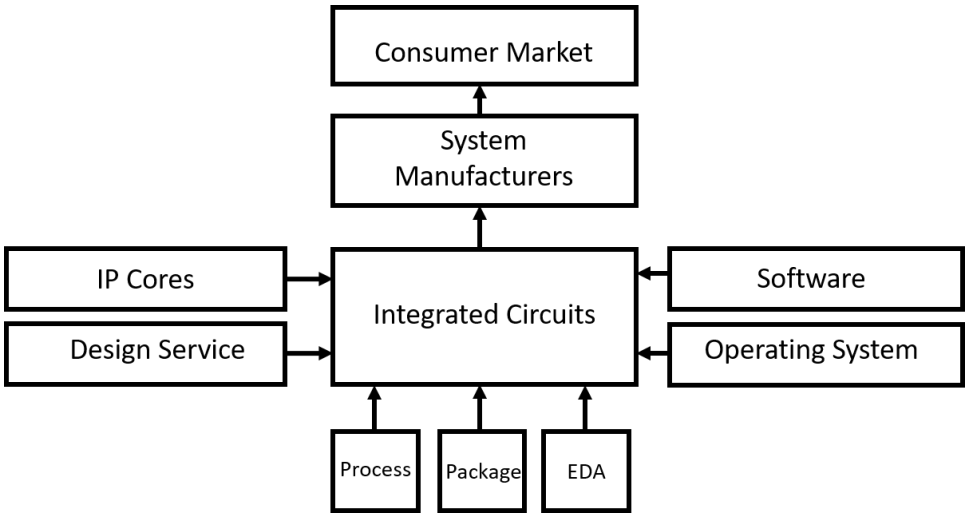

Figure 2.5: The eco-system of integrated circuits. The process foundry, the packaging company and the EDA company provide the design tools and fabricate the IC. The design companies design the IC and other companies provide the compatible software and operating system. The system manufactures will assembled various ICs together and produce the hardware platform on which services run. The final product are sold in the consumer market.

2.2.1 A Historical Perspective

The development of integrated circuits has come a long way, and new devices and new mythologies are still under active study.

In the first half of 20th century, the electronic circuits are built with large, expensive, power-hungry and un-reliable vacuum tubes. The first computer, ENIAC, was built between 1943 and 1945. ENIAC contained 18,000 vacuum tubes, 7,200 crystal diodes, 1,500 relays, 70,000 resistors, 10,000 capacitors, and approximately 5,000,000 hand-soldered joints[5]. It weighed more than 27 tons and was roughly 2m×1m×30m in size and occupied $170m^2$ and consumed 150kW of electricity.

As early as 1952, British electronics engineer Geoffrey Dummer became the first person to conceptualise an integrated circuit. Meanwhile, several other scientists from

around the world, such as Yasuro Tarui and Sidney Darlington, had also suggested the design of a single board that contains several transistors.

Later in 1958, the first practical IC was demonstrated by Jack Kilby, who at the time worked at Texas Instruments. His team was working on a solution for the "tyranny of numbers". The problem that the need to integrate a large number of discrete modules increases the complexity of electronic circuits. He realised that the solution to such an issue was to integrate all the electronic components into a single piece of semiconductor material. Jack Kilby constructed the first IC flip-flop with two transistors from a Ge slice and gold wire at Texas Instrument. While it could only be produced in limited quantities, the invention of transistors earned him and his supervisor William Shockley the 2000 Nobel Prize in Physics.

In 1959, Robert Noyce, a physicist working at Fairchild Semiconductors, designed what came to be known as the first monolithic IC. Rather than using wires, Noyce connected all the components on a silicon substrate with copper traces. This design revolutionised semiconductor manufacturing in terms of cost, performance, and size, ultimately allowing manufacturers to integrate increasing numbers of components on a single chip. Notably, surface and planar passivation enabled the production of true monolithic IC.

In 1960, MOSFET began to enter production. The unique advantage of such device is that it consumes almost zero current when idle, which is ideal for reducing the power consumption and heat dissipation.

In 1963, Frank Wanlass at Fairchild described the first logic gate using MOSFET, using both nMOS and pMOS (earning the name Complementary Metal Oxide Semiconductor (CMOS)).

In 1965, Gordon Moore found transistor count doubling every 18 months. This so-called Moore's law become a self-fulfilling prophecy in the next several decades and predicts the development of ICs very nicely (Figure.2.2.1).

The innovation race in the IC industry led to unprecedented progress in the digital technology since the late 1940s. As the early ENIAC was roughly 2m×1m×30m in size and occupied $170m^2$, the Intel Itanium Microprocessor was smaller than a credit card but contained more than 2 billion transistors and 16Gb flash memory (more than 4 billion transistors) in 2008. Nowadays in 2022, Apple's M2 chip contains more than 20 billion transistors but is orders of magnitude faster.

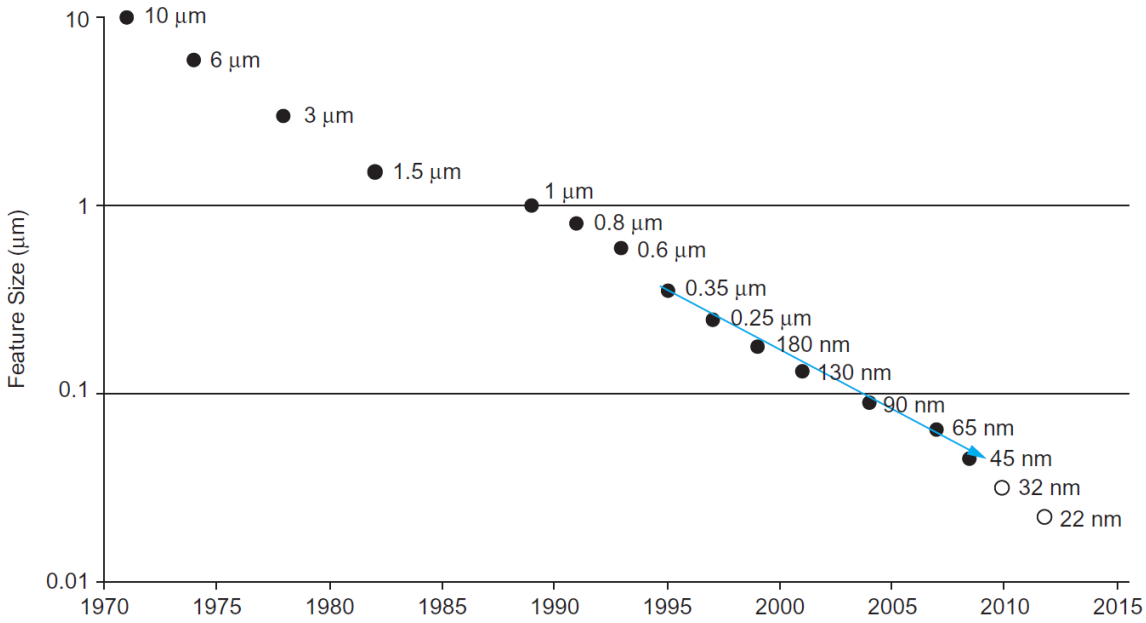

Figure 2.6: The process generation of integrated circuits [27]

Nowadays, ICs and its related products have become the backbone of the economy. All aspect of the IC and the industries built upon it are responsible for almost 10% of the gross domestic product of the US.

Loosely speaking, IC chips can be categorized based on the number of transistors it has. Small IC contains only 10 to 100 transistors. Medium chips usually contain on the order of 100 to 1000 transistors. Large chips have about 1000 to 100000 transistors, while Very-large ICs have up to 1 million transistors. With the advancement of technology, ultra-large chips nowadays can even have millions to billions of transistors whose feature size is as small as 5 nm. As of the writting of this work, the most advanced IC process node under test production is 3 nm by Taiwan Semiconductor Manufacturing Company (TSMC).

2.2.2 The System-on-Chip (SoC)

Electronic design based on IC technology can take various forms. To better describe and understand them, a spectrum of architecture are shown in Figure.2.7.

On the right end of the architecture spectrum is the general purpose computing IC. Such a system intends to run a fully general set of applications. The end-user has

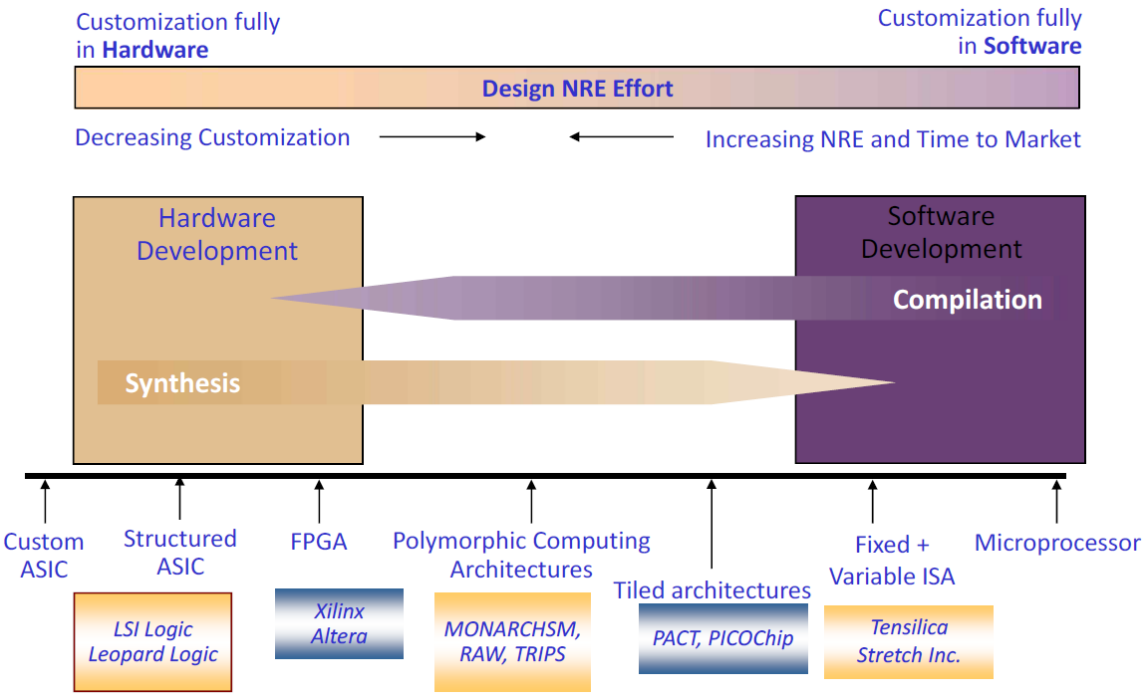

Figure 2.7: The Spectrum of the IC Architecture

the ability to program it at the software level with the help with compilers. Since the task is unknown during its design stage, they are designed to be faster at general arithmetic operations. Typical examples includes the Central Processing Unit (CPU) or the Graphic Processing Unit (GPU).

On the other end of the spectrum is the customized system, which runs a few applications often known at the design time. They are designed and optimized to be fast at these given applications and usually are not end-user programmable. Due to this nature, they are called Domain Specific Accelerator (DSA) in some literature. Examples include the H3DD_UM front-end ASIC, the off-the-shelf ADC ASIC or various Power Management Units (PMU).

Residing between them are some ICs that try to combine the benefit of the two by mixing the design (Hybrid Computing Architecture). A typical example is the FPGA. It sometimes provides users with programmable arrays of gates (the fabric), together with general purpose microprocessors. Users can benefit from the advancement in general purpose processors without losing the flexibility to customize the system on the hardware level.

SoC is a special group of ASIC (left side of the spectrum), that integrates part of or the entire functional system into a single silicon chip, instead of integrating them on the printed circuit board (Figure.2.8).

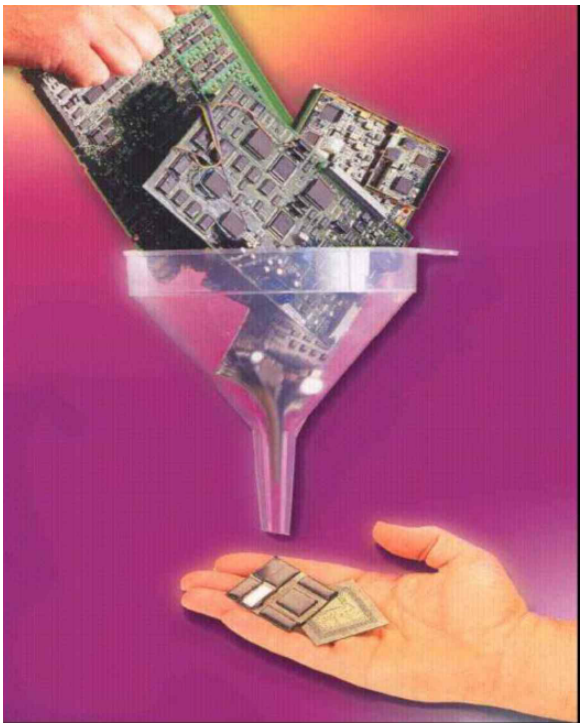

Figure 2.8: SoC replaces the board-based system by integrating multiple functional units into a single silicon die.

The functional units (sometimes called IP core) being integrated can be either digital or analog. By stitching together multiple stand-alone Very-large Scale Integrated Circuits (VLSI) designs, SoC provides most of or even the full functionality of an application. It composed of pre-designed models of complex functions known as cores that serve a variety of applications. It is a system on an IC that integrates software and hardware intellectual property using more than one design methodology for the purpose of defining the functionality and behavior of the proposed system.

The SoC became increasingly popular with the growth of Internet-of-Things (IoT) and mobile computing due to various advantages.

1. Synthesizing a core from Register Transfer Level (RTL) description allows for full control over it. This brings additional degrees of freedom in the quality of imple-

mentation (pay only for what you use).

2. As we essentially size down what is normally a multi-chip design onto a single silicon substrate, the space of the whole system can be reduced when compared to their multi-chip counterparts.

3. Since the SoC is designed and optimized to finish a very specific task, it's generally faster and less power hungry than the general-purpose circuits.

4. Since the communication among modules is on-chip, the bandwidth of data transfer and the processing can be sped up too.

2.2.3 The Modern Design Methodology

The methodology where one person or one group cover the design and fabrication in house gradually becomes impossible. With the growth of number of transistors on die, breaking down the logic into basic logic gates becomes tedious and can't be done manually in a timely manner. Besides, the decrease in the size of each transistor increases the density of transistors in the final layout. Manually placing and routing each individual devices is unrealistic due to its complexity.

Two factors revolutionised the IC design workflow. Nowadays, using advanced design and manufacturing techniques, several billions of transistors can be fitted into a silicon chip roughly the size of a human fingernail. The first factor is the Computer Aided Design (CAD). IC development helped making smaller, yet more powerful computers. The computers, in turn, free humans from the tedious work of working on logic functions, placing and routing the netlists. Nowadays, these tasks can be performed by computers or servers within a short period of time. The second is the separation of tasks in the workflow. Each engineer (or team) can focus on a small portion of the whole flow, therefore reducing possibility of making errors and improving the delivery quality. The workflow is illustrated in Figure.2.9 and will be explained shortly after.

Front-end Design

The front-end design (left part of Figure.2.9) covers everything before the physical layout. Common tasks include,

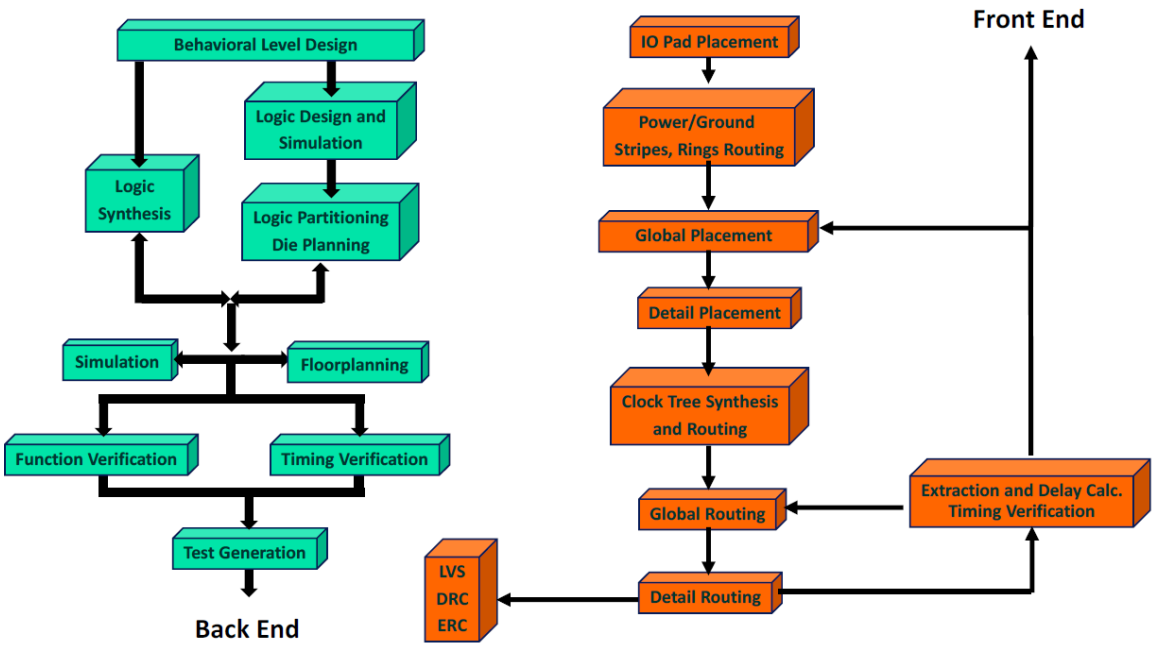

Figure 2.9: The front-end (left) and back-end(right) of the IC design workflow

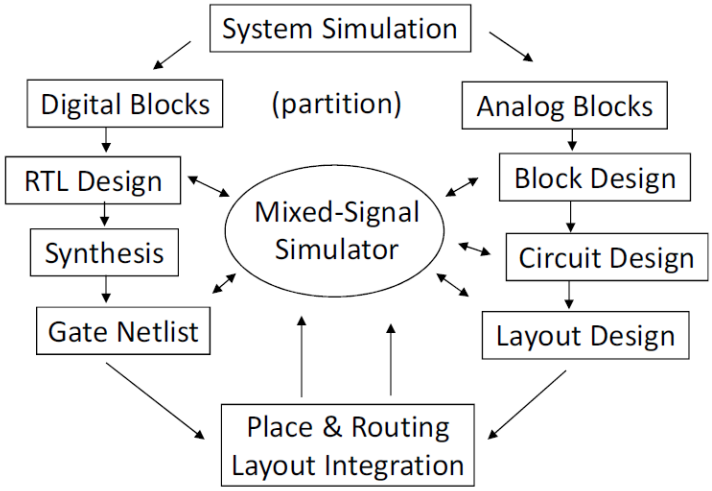

Figure 2.10: The mixed-signal simulation flow

1. System Partition: come up with the system design specification. Clarify what the system is supposed to do. Partition the design into major sub-components. Define the state of each components at the high level.

2. Behavior Simulation (Behavior Domain): Write Hardware Description Language (HDL) of each module. Build testbenches to simulate the behavior of each components. Occasionally, if mixed-signal behavior is of interest, mixed-signal simulation engine is used to perform co-simulation (Figure.2.10).

3. Logic Synthesis: Synthesize modules into a netlist of basic logic gates. The basic logic gates are usually standard cells provided by the foundry nowadays.

4. Gate-level Simulation (Structure Domain): Verify the correctness of the synthesized netlist with the testbenches. Check and solve any timing violations inside the netlist.

Back-end Design

The front-end design (right part of Figure.2.9) covers everything that's about the physical layout. Common tasks include,

1. Layout Design: set the geometry of individual circuit elements. Insert the power net and clock trees. Check for timing violations in the layout due to extra parasitic Resistance-Inductance-Conductance (RLC) networks.

2. Design Rule Check (DRC): Check for any design rule violations so that the layout can be fabricated by the foundry.

3. Layout-Schematic Comparison (LVS): compare the layout and the schematic, so that we are sure the layout realises what we intended to design.

4. Parasitic Extraction (PEX): extract the parasitic RLC network from the physical layout. The RLC network affects the delay of the signal and may cause extra timing violation, therefore, they must be taken into account during the design.

5. Post-layout Verification (Physical Domain): verify the correctness of the layout by running the previous simulation again with the same testbench. Check for any timing violation since the delay from the parasitics are now taken into account.

6. Silicon Test: after the fabrication, the silicon chip will be tested extensively to verify it can achieve our design goal and meet the expected specification.

2.3 Summary

In this chapter, we reviewed the technologies on which this thesis work based. First, we reviewed the history of radiation detection system in the Orion group at UM. The most recent detection systems at UM are described. Second, we took a look at the history of IC technology and the modern IC design methodology. The team at UM followed this methodology to develop the ASIC. Having the knowledge of both of them, the adventure of this thesis work can now begin.

3 Efficiency Study on 20mm×20mm CdZnTe Crystal

3.1 Introduction

The energy resolution is probably the most important characteristic of any radiation detectors developed for gamma-ray spectroscopy. Our research group at the University of Michigan has demonstrated that CZT detector can achieve close to HPGe energy resolution at room temperature [37]. Detection efficiency is also of great importance as well as energy resolution. Intrinsic photopeak efficiency of pixelated CZT detectors is an important property for quantitative measurements and analysis, especially for detection and characterization of SNM. The efficiency characteristics of CZT detector using IDEAS VAS/TAT system and BNL's analog ASIC have been investigated previously [36], and the work presented in this chapter focused on the efficiency characteristics of the latest digital ASIC-based systems.

First, the intrinsic photopeak efficiency of a single 20mm×20mm×15mm CZT detector from 59keV to 2614keV are measured in experiment. This is the most accurate efficiency benchmark experiment on the digital ASIC-based system.

Second, it was noticed compared to the predicated ideal intrinsic photopeak efficiency from physics, the measured efficiency showed deficit. A detailed simulation model including gamma-ray interaction physics, readout electronic as well as signal reconstruction process is built to explain this observed efficiency deficit. The simulated efficiencies from the new model match with experimental results well after taking these factors into account.

Third, several efficiency loss mechanisms are revealed. Among them, the efficiency loss due to the guard ring and anode side inaccurate reconstruction are the most important

factors. Other miscellaneous efficiency loss mechanisms are discussed too. This study is the first time the details of the efficiency response of any CZT crystal are understood. It provides useful guidance in future radiation detector design.

At last, a demonstration experiment to showcase the quantitative measurement capability of the CZT detectors were presented. With the 3-D imaging capability of TEI, the location of the point source can be determined within an uncertainty of 0.5cm. The activity of the source can be estimated from the energy spectrum within an uncertainty of 10% up to 1,274keV.

3.2 Measurement of Intrinsic Photopeak Efficiency

3.2.1 Theory

The intrinsic photopeak efficiency of a detector primarily depends on the detector material, the radiation energy and the physical dimensions of detector [14]. The intrinsic photopeak efficiency is defined as the ratio between the number of recorded photopeak counts N_{pp} and the number of radiation quanta incident on the detector surface $N_{incident}$ [14]. The efficiency can be calculated using the following equation,

$$\epsilon_{int,pp} = \frac{N_{pp}}{N_{incident}} = \frac{N}{A \cdot B \cdot T \cdot (1-f) \cdot \Omega} \tag{3.1}$$

where, N is the number of counts in the photopeak region, A is the activity of the source, Ω is the solid angle that the source sees. In far field measurement when the source is not too close to the detector, Ω can be calculated by considering the 20mm×20mm front area of a CZT crystal,

$$\Omega = \frac{S}{4\pi d^2} = \frac{20mm \times 20mm}{4\pi d^2} \tag{3.2}$$

in which d is the distance between the center of the source and the upper surface of the CZT crystal. The hidden assumptions are that the source is an ideal point source and emits gammas isotropically. B is the branching ratio of the photopeak emission (e.g. 85.10% for 661.7 keV of ^{137}Cs) and can be found in published nuclear database. T is the total measurement time. f is the dead time fraction of the system, and can be obtained from the firmware in our data acquisition computer program.

3.2.2 Experiment Uncertainty Analysis

The uncertainty of final efficiency result can be calculated by Eq.(3.3). Since T and B are known values in experiment, their uncertainty should be negligible. The elapsed time of experiment is usually long enough such that the uncertainty of N and f are very small too and therefore can be omitted in uncertainty estimation.

$$(\frac{\Delta\epsilon_{int,pp}}{\epsilon_{int,pp}})^2 = (\frac{\Delta A}{A})^2 + (\frac{\Delta\Omega}{\Omega})^2 \tag{3.3}$$

In Eq.(3.3), the uncertainty of the geometry factor Ω can be calculated from Eq.(3.4).

$$(\frac{\Delta\Omega}{\Omega})^2 = (2 \cdot \frac{\Delta d}{d})^2 \tag{3.4}$$

According to the vendor, the labelled activity of the disk source can have an uncertainty as large as 5%. The activity of several sources are calibrated using a well-calibrated high-purity Germanium detector and the uncertainty was measured to be 2.5%. Besides, the disk source is 5mm thick, and introduces a location uncertainty of $\Delta d = \pm 2.5$mm.

In Eq.(3.1), the source is assumed to be an ideal point source and the emission of gammas is isotropic. However, the radioactive material in the disk source can be a little extended. Besides, the emission could also be anisotropic due to self-shielding. These factors are hard to model quantitatively and could cause systematic bias on the measured intrinsic photopeak efficiency.

3.2.3 Experimental Setup

The setup of efficiency measurement is shown in Figure 3.1. The source irradiates the crystal from the cathode side. It is placed 41 cm away from the detector outer case to minimize the uncertainty of source location. Even though the systems can house up to nine crystals, the intrinsic photopeak efficiency is measured separately on each individual crystals.

3.2.4 Definition of Photopeak Region

In experiment, the measured efficiency is highly dependent on how the photopeak region is defined. Figure 3.2 shows the photopeak of ^{137}Cs (all collected events combined).

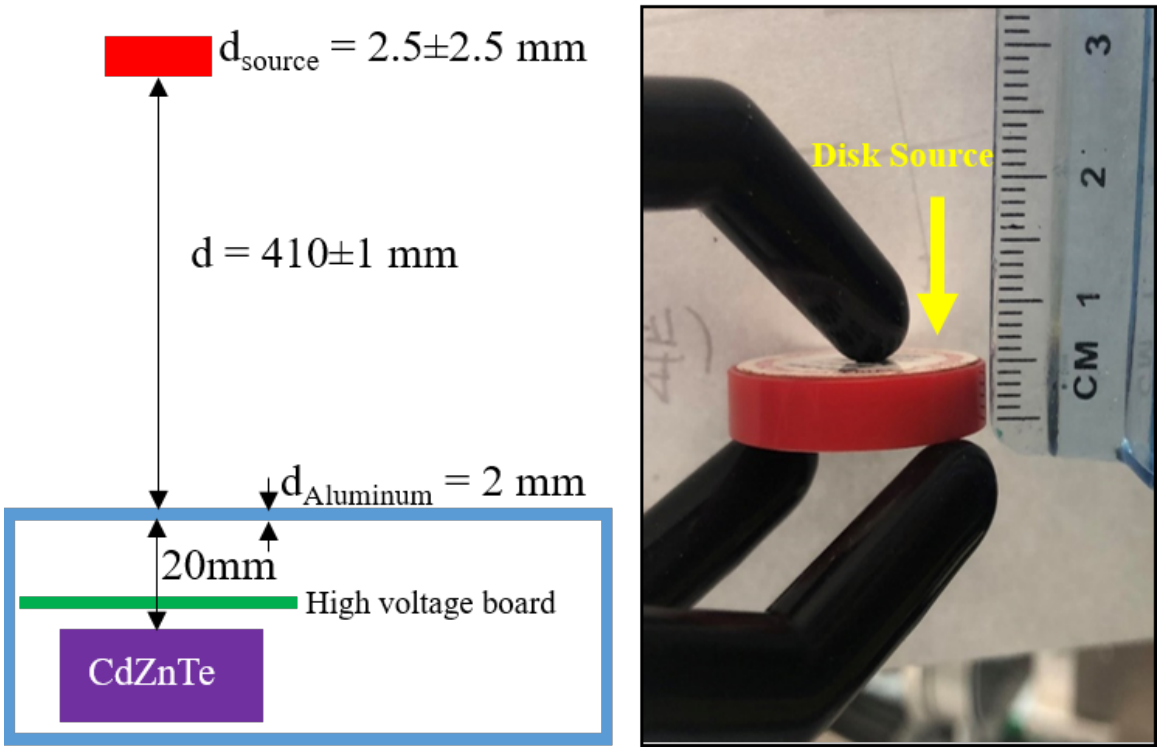

Figure 3.1: (Left)The setup of the efficiency measurement experiment. The cathode side of the crystal is facing up. (Right) The sealed disk source that is used in the experiment. The thickness is 5 mm.

The simulated efficiency will be discussed later in this chapter. Even though the energy resolution of the photopeak is superior (0.66% for all events), it's challenging to set a clear boundary to define the photopeak region. The significant low energy tail of the photopeak results from multiple factors, including charge leak, charge sharing and weighting potential cross talk (WPCT).

When a narrow energy window is used, the measured efficiency is lower since some of the full-energy deposition events in the shoulders are lost. A wider photopeak region can increase photopeak efficiency, but will degrade resolving power of gamma-ray lines in real-world applications, especially when multiple characteristic gamma-ray peaks are close to each other. The photopeak region used when measuring photopeak intrinsic efficiency at different energies are shown in Table 3.1.

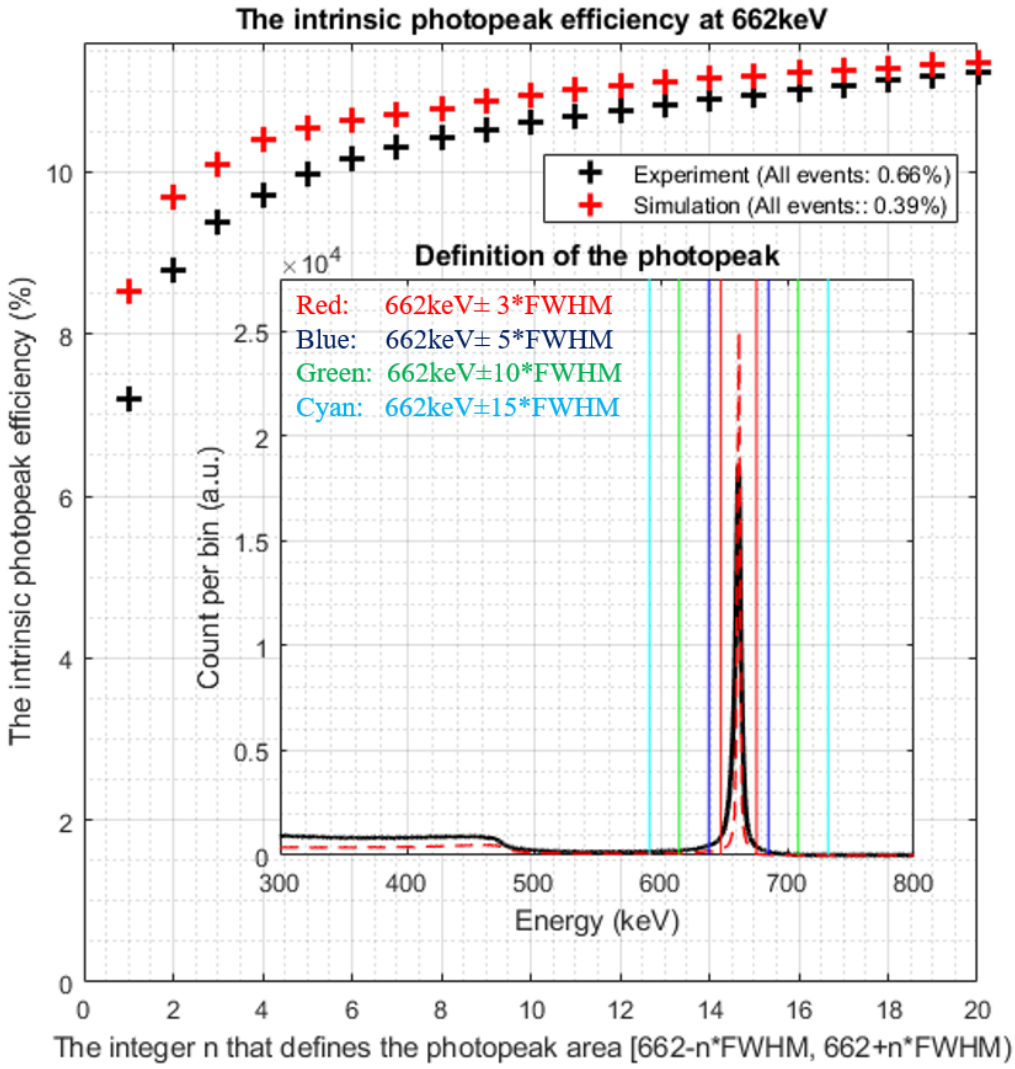

Figure 3.2: (Big) Photopeak efficiency as a function of photopeak region in simulation and experiment. (Small) Comparison between the measured all-event spectra of a ^{137}Cs source in simulation and experiment. Four energy windows to define the photopeak region are visualized (Red:$\pm$3$\cdot$ FWHM;Blue:$\pm$5$\cdot$ FWHM;Green:$\pm$10$\cdot$ FWHM;Cyan:$\pm$20$\cdot$ FWHM)

3.2.5 Results

The intrinsic photopeak efficiencies of a single CZT crystal (ID number 6RID-23) whose size is 20mm$\times$20mm$\times$15mm are measured on Orion-α system with ^{241}Am, ^{152}Eu, ^{137}Cs,

^{22}Na and ^{228}Th. When measuring the intrinsic photopeak efficiency at low energy, the high voltage board (Figure 3.1) was removed to minimize gamma-ray attenuation. The measured intrinsic photopeak efficiency at different energies are summarized in Table 3.1.

Table 3.1: Measured, Ideal and simulated Intrinsic Photopeak Efficiency of a single crystal at Different Energies

Source	Photopeak Range	ϵ_{exp}	ϵ_{ideal}	ϵ_{sim}
^{241}Am (59keV)	55-70keV	79.98%	90.97%	81.96%
^{152}Eu (121keV)	115-135keV	75.74%	90.66%	80.68%
^{137}Cs (662keV)	600-700keV	10.81%	14.01%	11.46%
^{22}Na (1274keV)	1190-1350keV	4.18%	5.68%	4.35%
^{228}Th (2614keV)	2550-2800keV	1.35%	2.12%	1.41%

In Table 3.1, ϵ_{exp} represents the measured intrinsic photopeak efficiency of a single 20mm×20mm×15mm CZT crystal. ϵ_{ideal} represents the intrinsic photopeak efficiency under ideal condition, where only the physics of gamma interaction with CZT was considered. The signal induction, collection are perfect and algorithms reconstruct every event without any error. At low energy range, around 12% discrepancy between ideal efficiency ϵ_{ideal} and experiment ϵ_{exp} is observed. At 662keV, such discrepancy becomes 23% and at 2614keV, it's almost 40%. As energy goes higher, the discrepancy becomes more significant. This is known as the efficiency deficit problem. ϵ_{sim} represents the intrinsic photopeak efficiency after taking the actual physical processes, readout electronic and event reconstruction into account. ϵ_{sim} will be elaborated in the following section.

Table 3.2: Measured Intrinsic Photopeak Efficiency of a single crystal at 661.7keV on two systems.

Triggering Mode	VAD_UM	Orion-α
Trigger Only	$10.69 \pm 0.43\%$	$10.79 \pm 0.31\%$
Trigger + 4	$10.70 \pm 0.43\%$	$10.82 \pm 0.32\%$
Trigger + 8	$10.64 \pm 0.43\%$	$10.81 \pm 0.32\%$

^{137}Cs source is used to benchmark the simulation program, so the intrinsic photopeak efficiency at 661.7 keV are measured on two systems with different crystals and different triggering modes. Introduction on different triggering modes can be found in the reference [29]. The efficiency of a crystal (ID number 6RID-56) was measured on VAD_UM system. The efficiency of a crystal (ID number 6RID-23) inside Orion-α was measured. The measured intrinsic photopeak efficiency performance at 661.7 keV on this two different systems are summarized in Table.3.2.

3.3 Modelling of the Detection System

To explain the efficiency deficit, actual signal induction, charge collection and signal processing are included in the system modelling. A custom simulation software was developed to model the detection system.

3.3.1 Electron-Hole Pair Generation

When a gamma particle incidents with the CZT crystal, it may undergo Compton scattering, photoelectric absorption or pair production and transfers energy to the electrons [14]. The primary electrons produced in these processes will travel through the crystal and generate secondary electrons through ionization and excitation. As the secondary electron leaves their original site, a vacancy is also created (called a hole). Such physic process are simulated using GEANT4 10.04 package [1]. The program tracks the electron ionization process and energy deposition along the track. The electronic PCB, detector chamber and the system housing are modelled in the simulation. The total signal induced by the whole charge clouds on electrodes can be simulated through a superposition of the signals induced by all carriers in the charge cloud [13].

3.3.2 Signal Induction on Electrodes

The signal induction on electrodes is calculated using the Shockley-Ramo Theorem [9]. The Shockley-Ramo Theorem states that in a charge-sensitive device, the current i induced on an electrode from the movement of a point charge q can be calculated by

$$i = q\vec{v} \cdot \vec{E}_0(\vec{x}) \tag{3.5}$$

in which $\vec{v}$ is the instantaneous velocity of the point charge. $\vec{E_0}(\vec{x})$ is the weighting electric field at the charge location $\vec{x}$. The integral form of Eq.(3.5) is,

$$Q = -q \cdot \Delta\phi_0 = -q \cdot (\phi_0(\vec{x_1}) - \phi_0(\vec{x_0})) \tag{3.6}$$

in which $\phi_0(\vec{x})$ is the weighting potential when the point charge is at location $\vec{x}$. Given the drift trajectory of the charge carrier from the initial position $\vec{x_0}(t)$ to the finial position $\vec{x_1}(t)$, the induced signal $Q(t)$ on any electrode can be simulated.

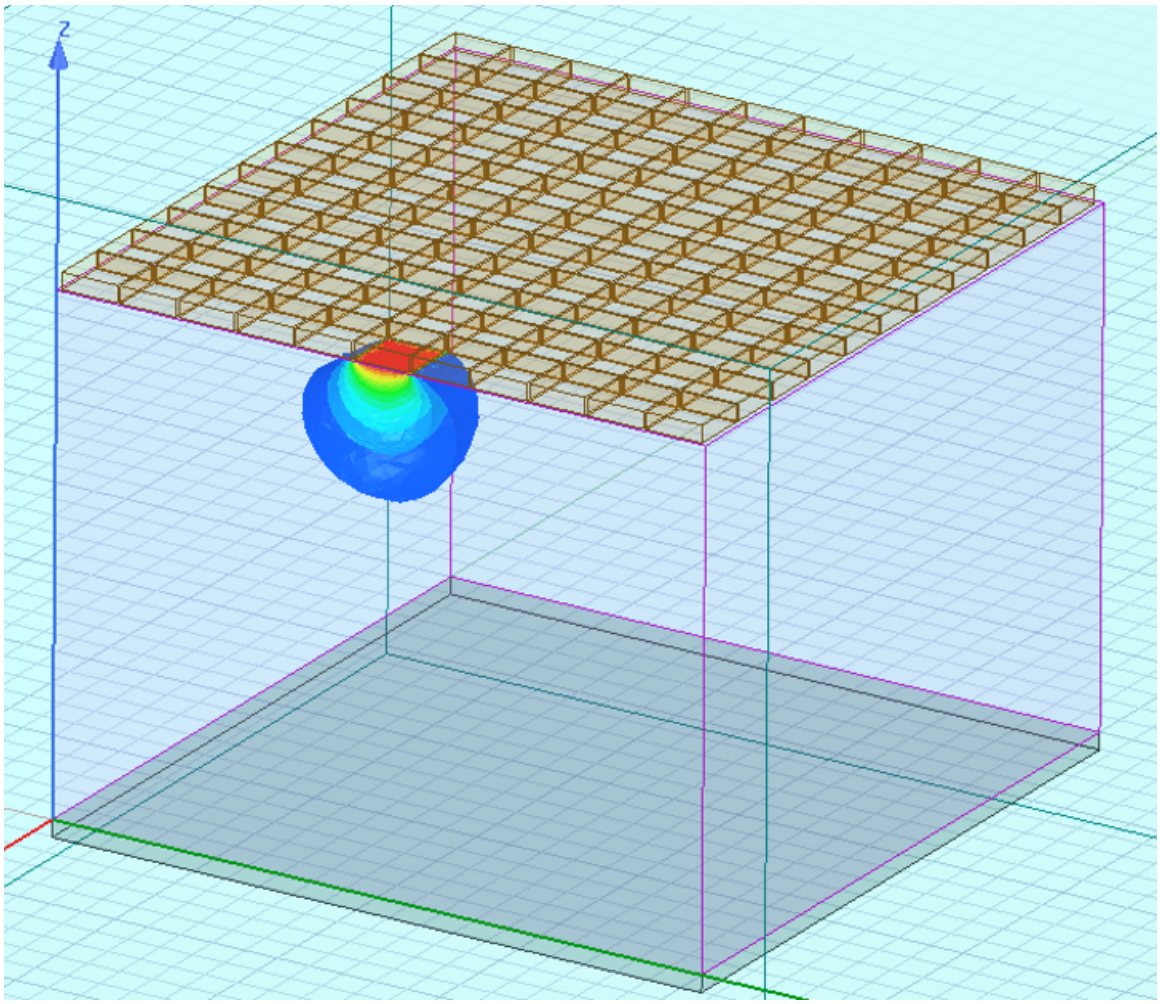

Figure 3.3: The weighting potential in a 15 mm thick CZT crystal was simulated in ANSYS Maxwell. The cathode is at 0 mm and anode is at 15 mm. This figure visualizes the weighting potential of an edge pixel, where only its potential is set to 1V and the rest are all at 0V.

The weighting potential can be calculated by solving the Poisson equations when only the electrode of interest is set be unit potential and all other to be zero potential, and with all spatial charge removed. Numerically, the weighting potential can be solved using Finite Element Method (FEM) with ANSYS Maxwell 18.1 software. A 20mm×20mm×15mm CZT crystal with pixelated anodes is built in software and simulated (Figure 3.3). The weighting potential of one anode pixel is sampled with a spatial resolution of 20μm×20μm×50μm. The simulated weighting potential of different pixels and at different sub-pixel locations are shown in Figure 3.4 and Figure 3.5 . As shown, the weighting potential starts to show significant sub-pixel variation when it's around 1.5 mm away from the anode.

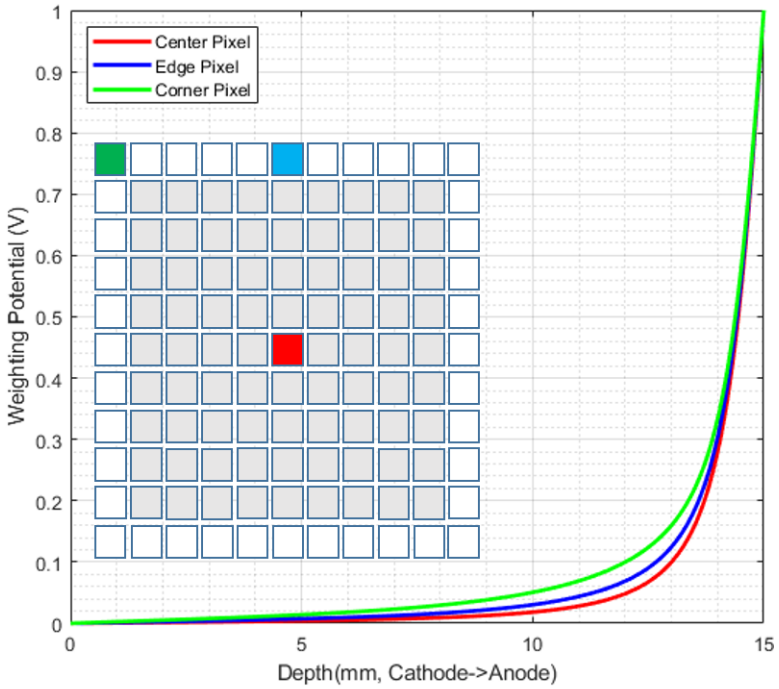

Figure 3.4: The simulated weighting potential in a 15 mm thick CZT crystal. In the figure the cathode is at 0 mm and anode is at 15 mm. For pixels at different locations, the weighting potential under the pixel center show variations.

Previous simulation study on analog-ASIC-based readout system assumed the weighting potential under one pixel is identical [36]. The weighting potential profile under the center of the anode pixel was selected to represent the weighting potential profile of all

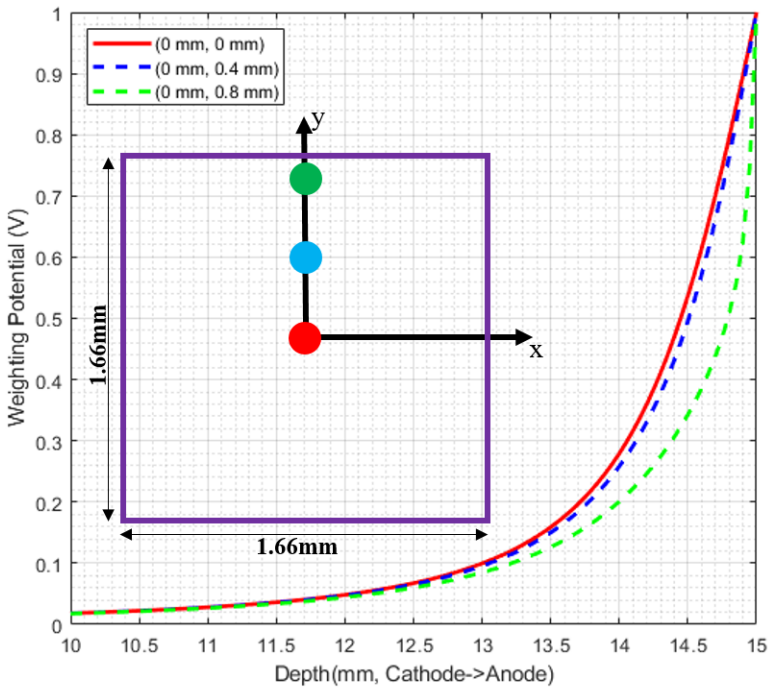

Figure 3.5: The simulated weighting potential in a 15 mm thick CZT crystal. In the figure, the cathode is at 0 mm and anode is at 15 mm. For one pixel, the weighting potential at different sub-pixel location show variations. To better visualize the difference, only the weighting potential near the anode are shown.

pixels. However, the weighting potential of central pixels, edge pixels and corner pixels do show some variations, as shown in Figure 3.4. Besides, the weighting potential under one pixel could also have slight lateral variations, as shown in Figure 3.5. For the latest system based on the digital ASIC (VAD_UM v2.2), the collected signal pulse waveform carries richer information. Therefore, the weighting potential variation among different pixels and at different sub-pixel positions should be taken into consideration to better model the detector response. Besides, the weighting potential of the non-collecting pixels are also simulated and taken into consideration to model the transient feature of the waveform more precisely (Figure 3.6).

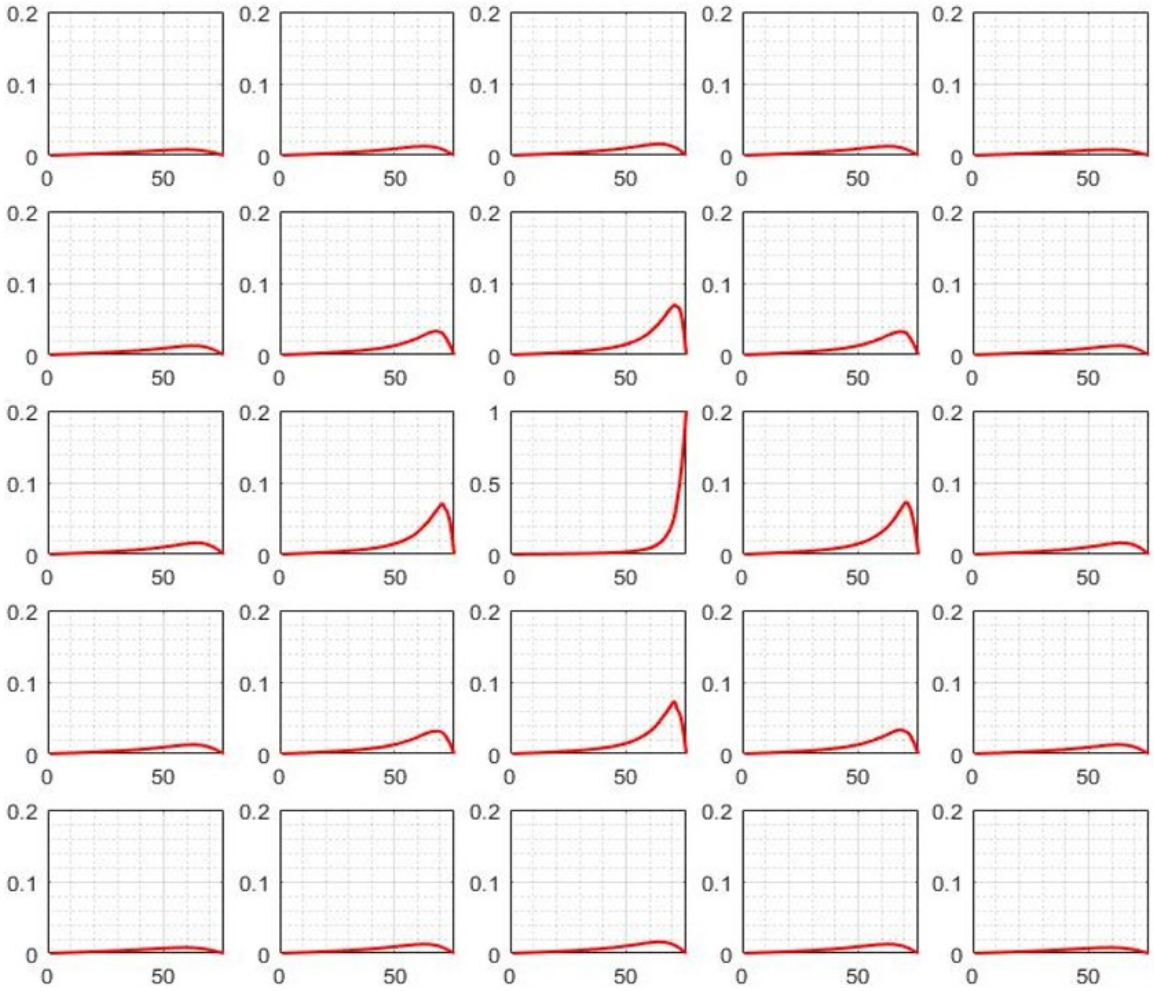

Figure 3.6: the weighting potential of the non-collecting pixels are also simulated and taken into consideration to model the transient feature of the waveform more precisely. The center picture is the weighting potential of the collecting pixel. The surrounding figures are the weighting potential of the pixels that are located at the relative location.

3.3.3 Drift Trajectory of Electron-Hole Pairs

To simulate the transient signal on electrodes, the drift trajectory $\vec{x}(t)$ also needs to be determined. Assuming the drift of electron-hole pairs only depends on their mobility and local electric field, and the material is ideal, the drift velocity can be expressed as,

$$\begin{cases} \frac{dx}{dt} = -\mu_e \cdot E_x(x,y) = -\mu_e \cdot \frac{\partial V(x,y)}{\partial x} \\ \frac{dy}{dt} = -\mu_e \cdot E_y(x,y) = -\mu_e \cdot \frac{\partial V(x,y)}{\partial y} \end{cases} \tag{3.7}$$

Eq.(3.7) can be re-written into the finite difference form, which is Eq.(3.8),

$$\begin{cases} \frac{x^{n+1}-x^n}{\Delta t} = -\mu_e \cdot \frac{V(x^n+\Delta x, y^n)-V(x^n,y^n)}{\Delta x} \\ \frac{y^{n+1}-y^n}{\Delta t} = -\mu_e \cdot \frac{V(x^n, y^n+\Delta y)-V(x^n,y^n)}{\Delta y} \end{cases} \tag{3.8}$$

in which $V(x,y)$ is the operating electric potential when the detector is biased. From Eq.(3.8), the location of electron (x^{n+1}, y^{n+1}) at time step n+1 can be updated from previous location (x^n, y^n),

$$\begin{cases} x^{n+1} = x^n - \mu_e \cdot \Delta t \cdot \frac{V(x^n+\Delta x, y^n)-V(x^n,y^n)}{\Delta x} \\ y^{n+1} = y^n - \mu_e \cdot \Delta t \cdot \frac{V(x^n, y^n+\Delta y)-V(x^n,y^n)}{\Delta y} \end{cases} \tag{3.9}$$

The actual operating field $V(x,y)$ is simulated with ANSYS Maxwell 18.1 software with actual electrode configurations (Figure 3.7).

In operation, the electric field in most volume of the detector is uniform. For those electrons and holes that are right under one pixel, they will drift perpendicularly to their corresponding electrodes. However, for those electrons and holes that are generated under the gaps between pixels, the drift behavior is different. The drift electric field becomes non-uniform under the gap in near anode region. As shown in Figure 3.7 and Figure 3.8, when the charged particle is more than 0.05mm away from anodes, the electric field is still quite uniform. But when it's less than 0.05mm away from anodes, the electric field is no longer perpendicular and has some tangential components.

With $\Delta t = 0.02$ ns, $\mu_e = 75\ mm^2/V \cdot ms$ and $\Delta x = 1\ \mu m$, the drift trajectory of an electron is solved and shown in Figure 3.8.

Two phenomenons are observed from the drift trajectory simulation. First, the non-uniform electric field under the gap will bend the drift trajectory only when the electron is very close to the anode ($<50\mu m$). In most of the detector volume, electrons and holes do drift perpendicularly to their corresponding electrodes. Second, the electrons under the gap will eventually be collected by its closest anode and induce waveforms (Figure

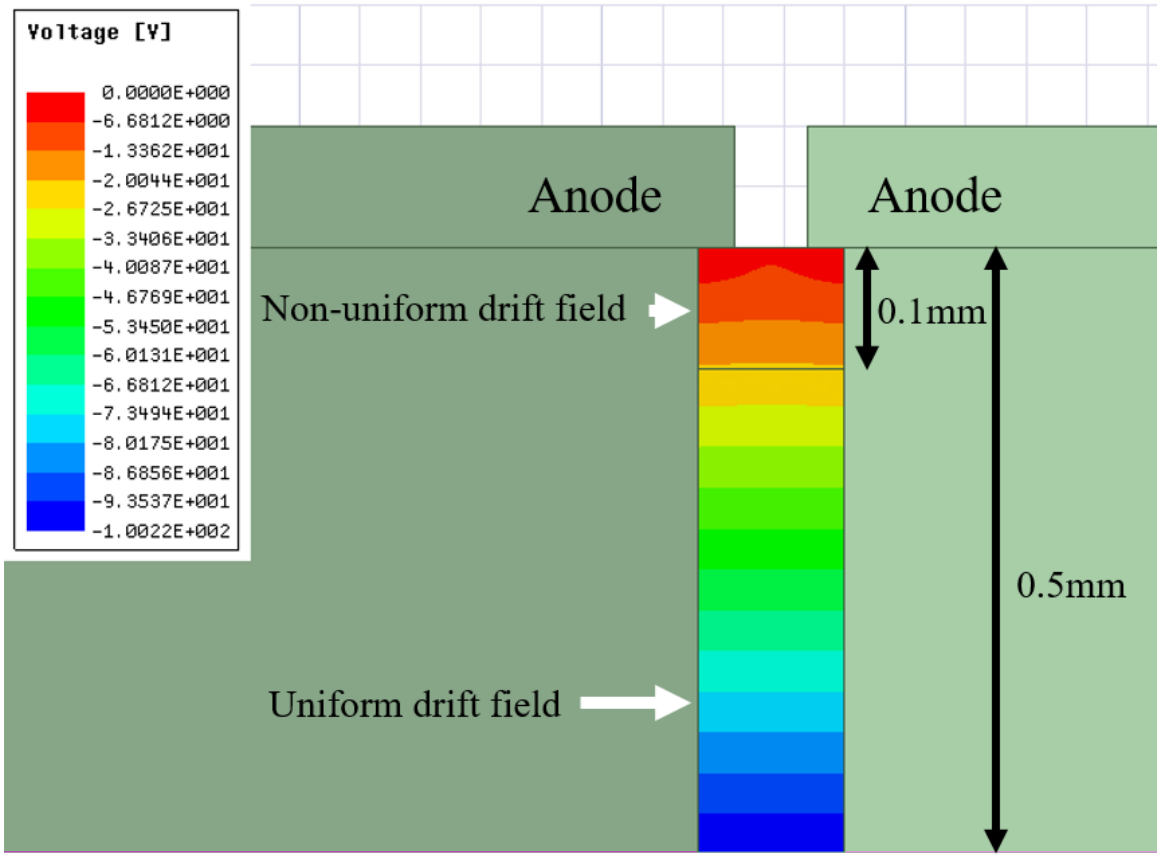

Figure 3.7: The drift electric field between pixels becomes non-uniform in near-anode region. Since in deeper depth, the electric field is very uniform, only the field near the anode side is shown.

3.9). These are the drift behaviors of a charge carrier under ideal condition. In actual CZT crystal, the drift behavior can be more complicated and is influenced by several other effects (charge trapping, pixel-jumping effects, non-uniform electric field inside crystal, etc.). The simulation should be viewed as a close approximation to the actual case.

With the simulated drift trajectory of charge carrier and weighting potential on all electrodes, the induced transient pulse waveform on every electrode can be simulated and recorded for event reconstruction. In simulation computer program, calculating the track of each individual electron and hole is computation-intensive and unnecessary. When the electrons or holes are more than 0.05 mm away from the anodes, they drift

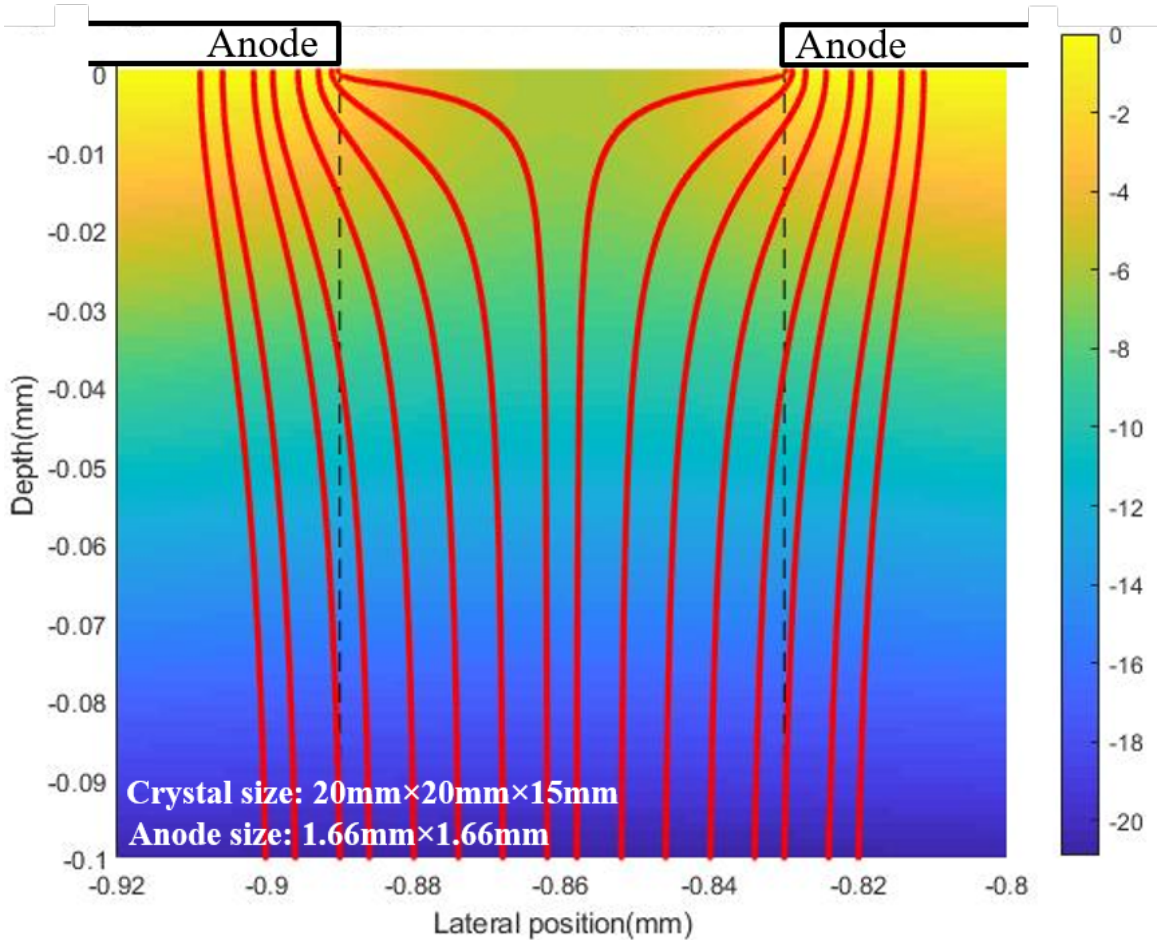

Figure 3.8: The background shows the operating electric field near anode side. The red curves show the drift trajectories of several electrons that are generated at different locations under the gap. Since in deeper depth, the trajectory is straight due to uniform electric field, only the trajectory near the anode side is shown. The color bar on the right indicates the operating electric potential (Unit: Volt)

perpendicularly to the electrode. They will be steered into the closest anode when they are 0.05 mm away from the anodes, and the tail amplitude is then determined.

3.3.4 Other Factors

Some other factors are also included in the model. (1) The diffusion of electron clouds are taken into account [31]. (2) The drift of holes are included in the model too. The

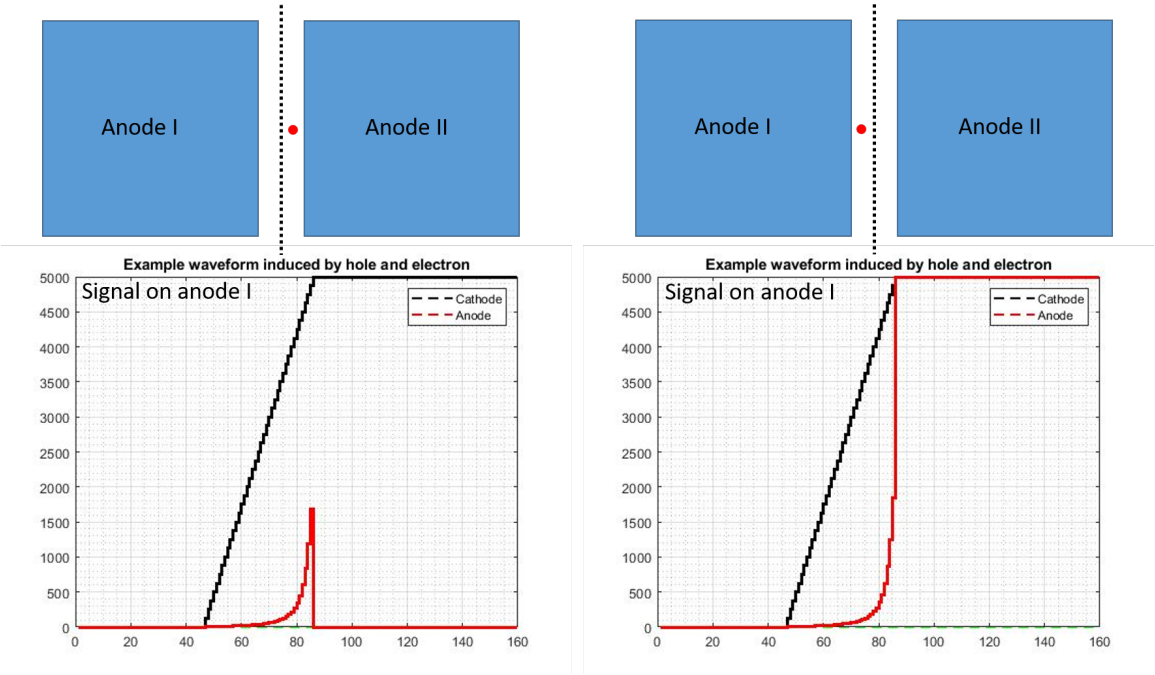

Figure 3.9: The simulated waveform of two events that happened under tha gap of the two anodes. Based on their location, they drift to their closest anode and induced different waveforms on the Anode I.

mobility of holes are assumed to be 100 times smaller than electrons [17]. (3) A 2keV equivalent white noise was added into the simulated waveform. (4) A 18keV software threshold was added to determine whether a pixel has real signal or just noise. This same threshold is used in experiment. (5) The surrounding materials are included in the GEANT4 simulation program.

Electron and hole trapping are not included. Since their impact on detection efficiency is negligible [36].

3.3.5 Event Reconstruction

The simulated waveform are processed in the same way as experimental waveform. The position and energy of a radiation event is reconstructed by our actual data processing algorithms. Detailed elaboration of event reconstruction algorithms can be found in reference [33, 38]. The simulated energy resolution are very close to the actual energy resolution of the detector (Figure 3.10). The comparison between the measured all-

event spectra in simulation and the experiment are shown in Figure.3.2. The simulated efficiency is always higher than the experimental one, since some factors are hard to quantitatively model. With a narrow photopeak window, the discrepancy is larger due to the different FWHM. As the window widens the discrepancy becomes smaller, since the window is wide enough to cover all full-energy deposition events, no matter the energy resolution.

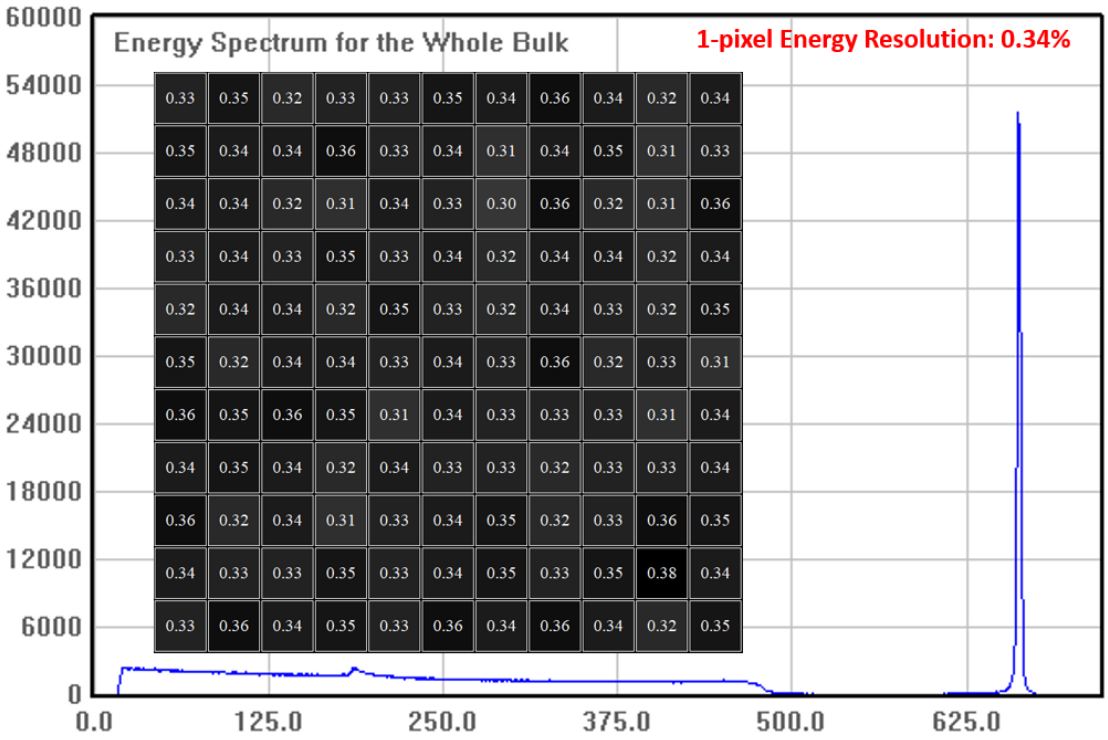

Figure 3.10: The energy spectrum of the single-pixel from the simulation. The simualted resolution was 0.34%, which matches the experiment results very well. The photopeak FWHM resolution distribution of single pixel events are shown in the inner figure.

3.4 Efficiency Loss Mechanisms

After considering the signal induction and event reconstruction, the simulated intrinsic photopeak efficiency ϵ_{sim} at different energies are summarized in Table 3.1. In simulation, the photopeak region is defined in the same way as in experiment (Table 3.1).

As shown, they agree very well with the measured intrinsic photopeak efficiency in experiment.

At 661.7keV, the simulated intrinsic photopeak efficiency is 11.46±0.02% while the measured one is 10.80±0.31%. Considering the uncertainty from experiment, there's still slight difference. The remaining discrepancy could come from the non-isotropic gamma emission from the source itself. Or the radioactive material inside the disk source cannot be simplified as a point source, and self-shielding can make the actual activity of source less than its labelled one. Such systematic error is hard to model quantitatively.

From simulation, the position distribution of the lost full-energy event can be extracted. An event is defined as "lost" if this event deposited all of its initial energy in the detector but the reconstructed energy is out of the photopeak region. The position distribution of the lost single pixel full-energy deposition events in ^{137}Cs simulation are shown as density map. Figure 3.11 (left) shows the position distribution in X-Y plane. Figure 3.11 (right) shows the position distribution in X-Z plane. Since any full-energy deposition that happened under the gap is a charge sharing event and will trigger two pixels, they are not included in Figure 3.11. This results in dark strips between pixels in the density map.

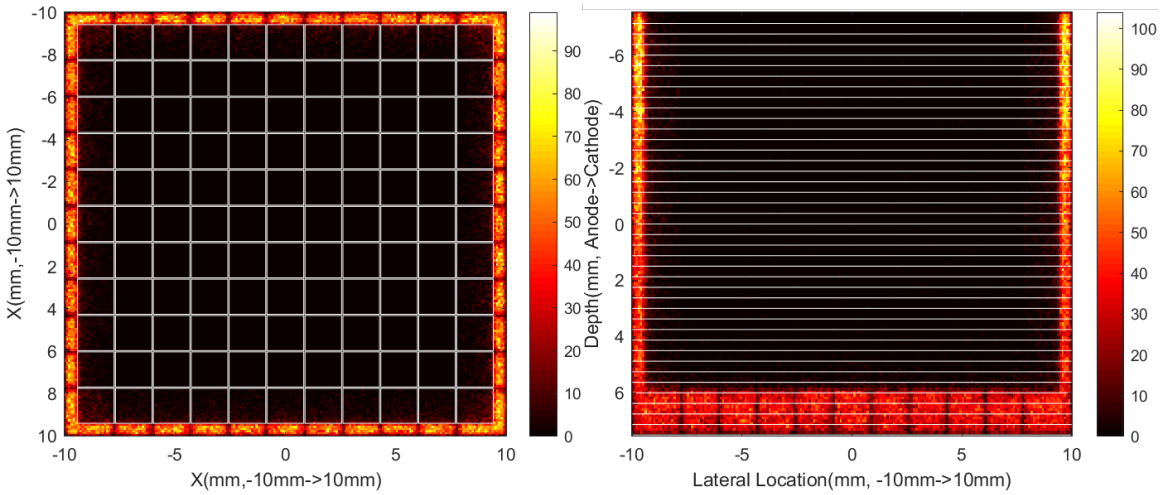

Figure 3.11: The location distribution of the lost single-pixel full-energy events from simulation (661.7keV of ^{137}Cs). (Left) the lateral position distribution. The pixel grids are marked in white for reference. (Right) a cross sectional view of the position distribution. The white lines mark the 40 depth bins.

Several mechanisms that can cause efficiency deficit are revealed and are elaborated in the following sections.

3.4.1 The Guard Ring Loss

For a single CZT crystal in Orion-α system, the guard ring, covering around 11.1% of the total 20mm×20mm front area, is equi-potential with pixelated anodes during operation to achieve better energy resolution. Electron-hole pairs generated under the guard ring region will not be steered towards the edge anode pixels, but will end up with being collected by the guard ring. Since this region is effectively not contributing to photopeak events detection, intrinsic efficiency deficit is observed under it, as shown in Figure 3.11(left). The guard ring loss is the dominant mechanism in low energy region. Most low-energy interactions happen near the cathode side and only triggers one anode pixel. The guard ring causes around 12% efficiency loss was observed at 59keV (^{241}Am), as shown in Table 3.1.

A single large crystal is more appealing in terms of efficiency if compared to several smaller crystals. For a single large crystal, there's only one large guard ring (assuming identical width) surrounding all pixels and it effectively covers less fraction of the total front area. The guard ring covers around 5% of the total front area for a larger 40mm×40mm×15mm crystal. The guard ring loss will therefore be reduced by half and the efficiency performance should be better.

3.4.2 The Anode Side Loss

Simulation and experiment also reveal the efficiency loss near the anodes. As shown in Figure 3.11 (right), full-energy events are lost near the anode side.

In current event reconstruction algorithm, the material under an anode is uniformly partitioned into 40 artificial bins. Depth bin No.1 is the anode side and depth bin No.40 is the cathode side. All events in the same voxel are reconstructed with the same set of calibration parameters [33]. The weighting potential shows significant variations on both lateral and depth directions in depth bins near the anode. On the one hand, since the weighting potential changes drastically in depth direction within one depth bin, a full-energy events that happens at different depth within same depth bin can induce a signal with very different amplitude on the collecting anode. The lateral variation, on

the other hand, results in different signal amplitude even though two interactions with same energy deposition happen at the same depth. Such variation in weighting potential near anode side can cause two problems in detector calibration and reconstruction, which could lead to a failed reconstruction for full-energy deposition event.

First, in the first several depth bins near anodes, the recorded photopeak is so broad that the photopeak centroid can't be identified. Therefore, the depth-correction in calibration stage will fail since the gain at this depth is unavailable from calibration data. In Figure 3.11 (right), most of the lost full-energy event near anode side are within the first four depth bins (roughly 1.5mm thick).

Second, in the next few depth bins, even though the peak centroid can be identified from the broad photopeak in calibration stage, the reconstruction always yields broad photopeaks that extend below 600keV. Even though the photopeak region is defined to be very wide, some full-energy events could still be lost. This explains the few scattered lost full-energy events beyond depth bin #5. Since the corner pixels have a different weighting potential profile, slightly more loss are observed under them.

It should be noted that such loss doesn't mean a 1.5mm dead layer exists near anode. Even if one event happens in this region, it can still trigger the readout electronic and can still have a chance being correctly reconstructed into the photopeak region.

Another effect that could cause efficiency loss near anode side is the dead layer on anode side. If one event happens very shallow in the crystal on the anode side, the induced signal on anode may be too small to trigger the readout electronics. Due to the superior noise performance in direct-attachment configuration, such dead layer is estimated to be very thin and therefore is just a minor efficiency loss mechanism.

Thicker crystal is preferable in terms of detection efficiency, not only because thicker crystal is better at absorbing incident gammas, but also because in thicker crystal, the small pixel effect is more significant. The steep change of weighting potential will only happen in depth bins that are very close to the anodes. Therefore, such bad reconstruction layer on anode side will be relatively thinner in thicker crystals. The loss due to anode side bad reconstruction should be less in thicker crystals.

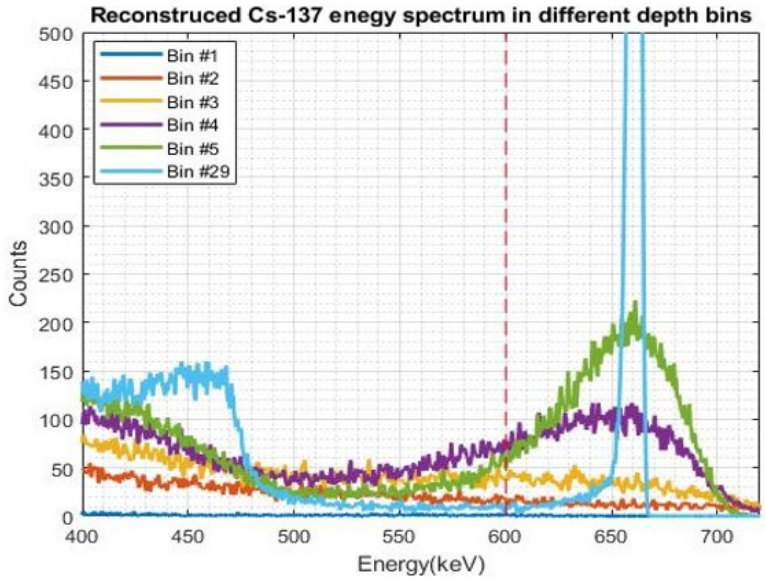

Figure 3.12: The broad photopeak near anode side after calibration (experiment). Bin #1 to Bin #5 is close to anode. Bin #29 is far away from anode and is shown here for comparison.

3.4.3 Miscellaneous Loss

At 662keV, only around 0.75% of full energy deposition will trigger five or more pixels so in typical experiments, the waveform of these events are not recorded nor processed. However, at higher energy, many more events will trigger five or more pixels during readout. The Orion-α system does have the capability to record and process these multi-pixel events. In this study, to keep consistency, such feature was not enabled. If detection efficiency is a critical design consideration in higher energy, the system should be designed to record and reconstruct these events, otherwise significant fraction of full-energy events would be lost.

Attenuation from surrounding materials can also degrade the intrinsic photopeak efficiency. This is negligible at 661.7keV or higher energy since the photons in this energy range are very penetrating. But the attenuation can cause significant efficiency loss at low energy range. For the 59keV peak from ^{241}Am, the intrinsic photopeak efficiency on a single 20mm×20mm×15mm CZT detector is 79.98% without the high-voltage PCB board (Figure 3.1), and reduces to 60.15% with the PCB covering the cathode. If the

intrinsic photopeak efficiency at low energy region is a critical design consideration, the attenuation material between the source and crystal should be minimized.

3.5 Efficiency Deficit Analysis

3.5.1 Efficiency Deficit Breakdown

In lower energy range, most interaction are cathode side single-pixel events. Therefore, the guard ring loss is the dominant mechanism. Since the guard ring covers 11.11% of total front area, around 12% loss was observed between experimental and ideal efficiency in Table 3.1. In higher energy range, multiple mechanisms can cause efficiency loss. The significance of each loss mechanisms at higher energies are summarized in Figure 3.13, Figure 3.14 and Figure 3.15.

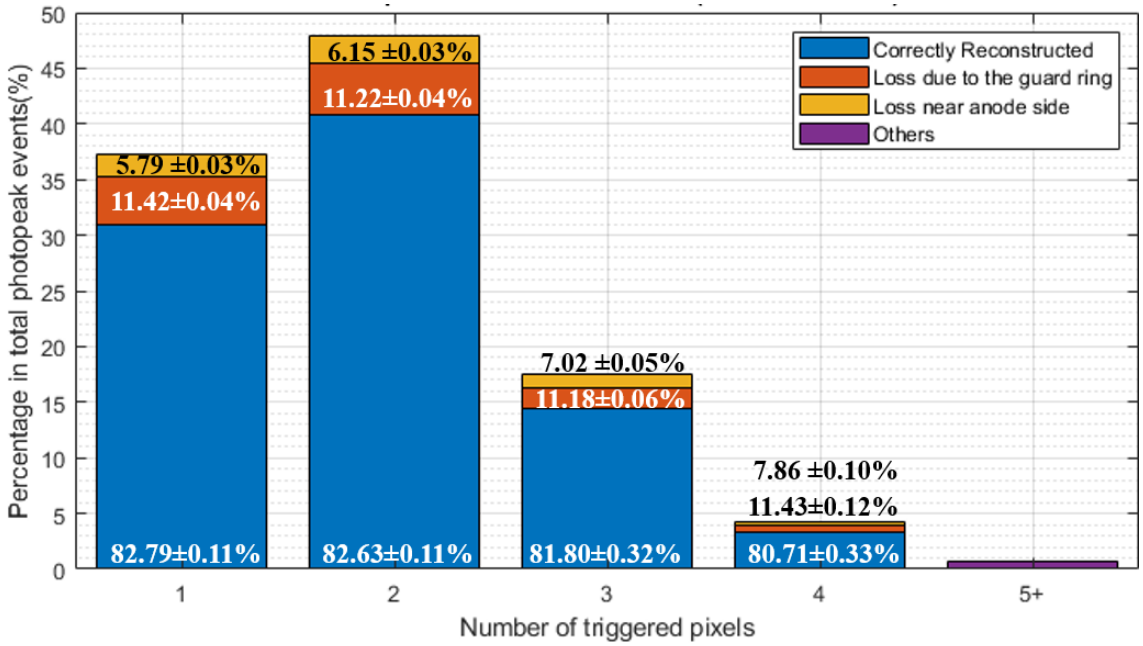

Figure 3.13: The significance of loss mechanisms for different types of events at 661.7 keV (from ^{137}Cs).

As the incident energy increases, more events would trigger five or more pixels (the rightmost bar in Figure 3.13, Figure 3.14 and Figure 3.15), which are discarded during data acquisition. Therefore, this mechanism becomes more significant.

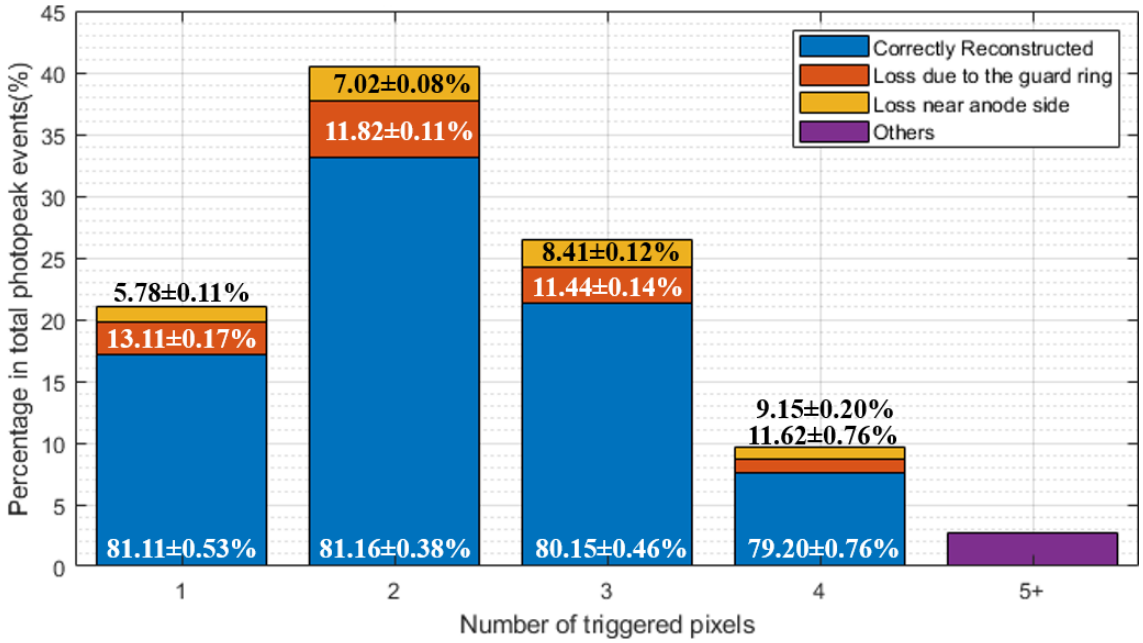

Figure 3.14: The significance of loss mechanisms for different types of events at 1274 keV (from ^{22}Na).

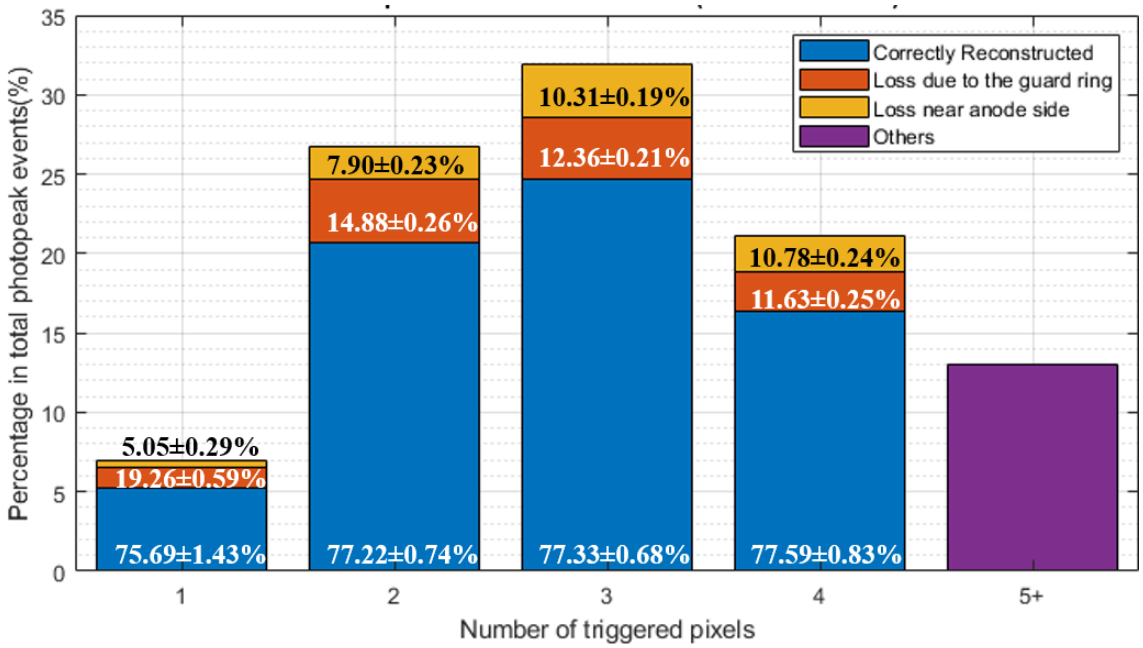

Figure 3.15: The significance of loss mechanisms for different types of events at 2614 keV (from ^{228}Th).

3.5.2 Loss Analysis on Single-pixel Events

The location distribution of the lost single-pixel full-energy events from simulation are shown in Figure 3.11 and Figure 3.16.

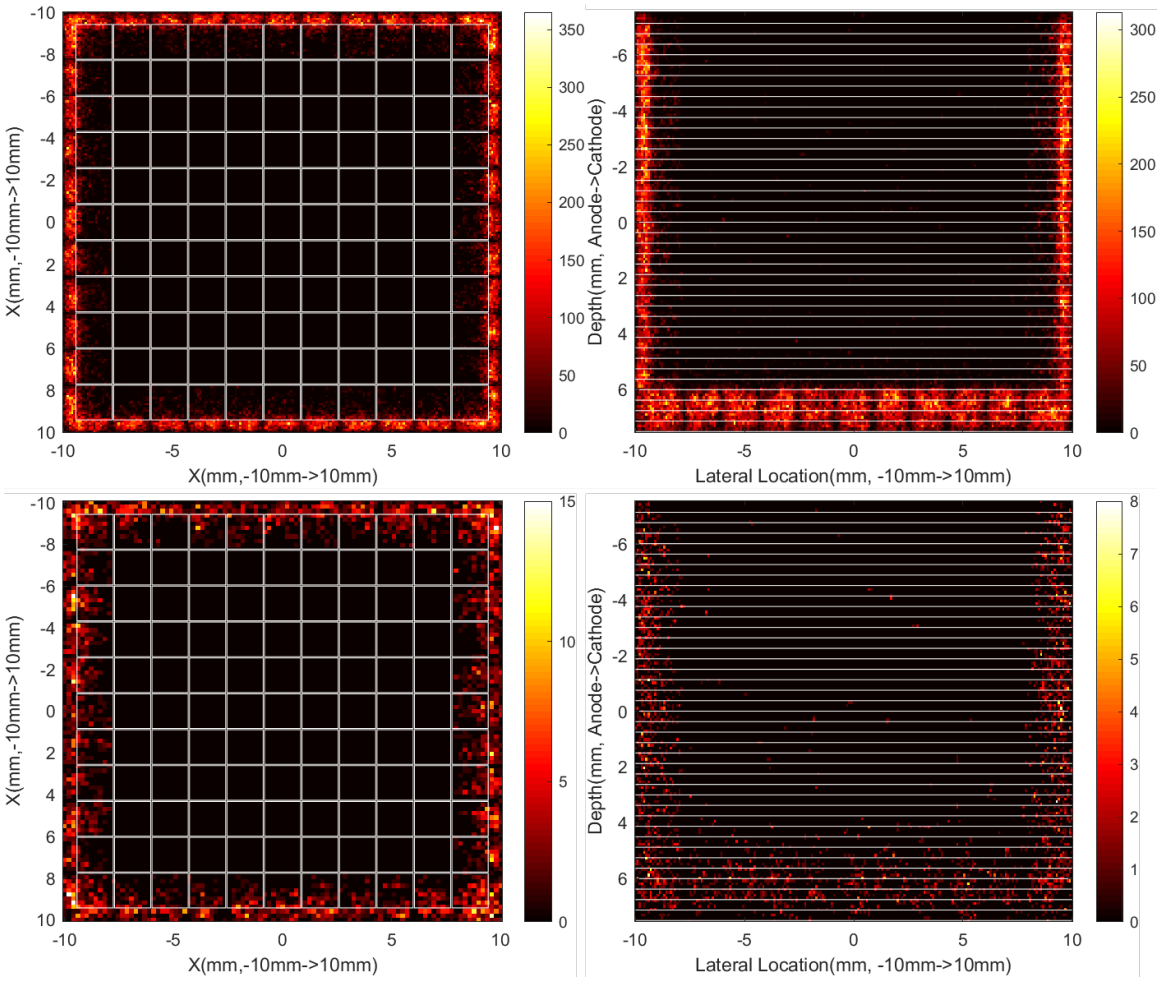

Figure 3.16: The location distribution of the lost single-pixel full-energy events from simulation. Top row is at 1274 keV. Bottom row is at 2614 keV. In each row, the lateral position distribution is shown on the left. The pixel grids are marked in white for reference. The cross sectional view of the position distribution is shown on the right. The white lines mark the 40 depth bins.

For single pixel events at different energies, the loss due to the guard ring shows obvious increase. At 661.7 keV, it's 11.42% and it increases to 19.29% at 2614 keV. Figure 3.11 and Figure 3.16 show that, as the energy goes higher, the effective width of

the guard ring "dead" region is expanding due to the increased size of electron clouds. At 2614 keV, the electron cloud size can be about 1.5 mm [33]. At 2614 keV, even if one photoelectric absorption event happens under the center of the edge pixel, it can still be influenced by the "dead" region since some electrons in its electron cloud are under the guard ring and therefore are lost during signal induction.

For single pixel events at different energies, the loss near the anode surface increases as the energy deposition increases. This trend also results from the increased electron cloud size. But such loss is less rapid if compared to the loss under the guard ring.

3.5.3 Loss Analysis on Multi-pixel Events

The analysis on multi-pixel events are based on the analysis of two-pixel events. The location distribution of the lost two-pixel full-energy events from simulation are shown in Figure 3.17. If a two-pixel event is outside the photopeak region, the locations of both interactions are shown in Figure 3.17 as the scattered events inside the crystal. The loss is from the same mechanisms discussed before.

For two-pixel events at different energies, the loss due to the guard ring and near anode side both increase as energy increases. The major reason is that the effective dead region is expanding due to the increased size of the electron cloud.

For three-pixel and four-pixel events, the observation and reasoning remain the same.

3.5.4 Loss Analysis at A Given Energy

At a given energy, the loss near the anode side becomes more significant as the number of triggered pixels increases. For example, at 661.7keV, the loss on anode side is 5.79%, 6.15%, 7.02% and 7.86% for 1-pixel, 2-pixel, 3-pixel and 4-pixel events, respectively. The more interactions one event has, the more likely one or more interactions may happen near the anode side.

Figure 3.18 illustrates how the guard ring loss changes as the number of triggered pixels increases. Such trend results from three factors.

First, the electron cloud size increases as energy goes higher. For single pixel events, all of its energy are deposited in one location so the cloud size is the largest. For multi-pixel events, the initial incident energy is deposited at several locations so that each individual cloud size is relatively smaller. Therefore, for multi-pixel events, the effective

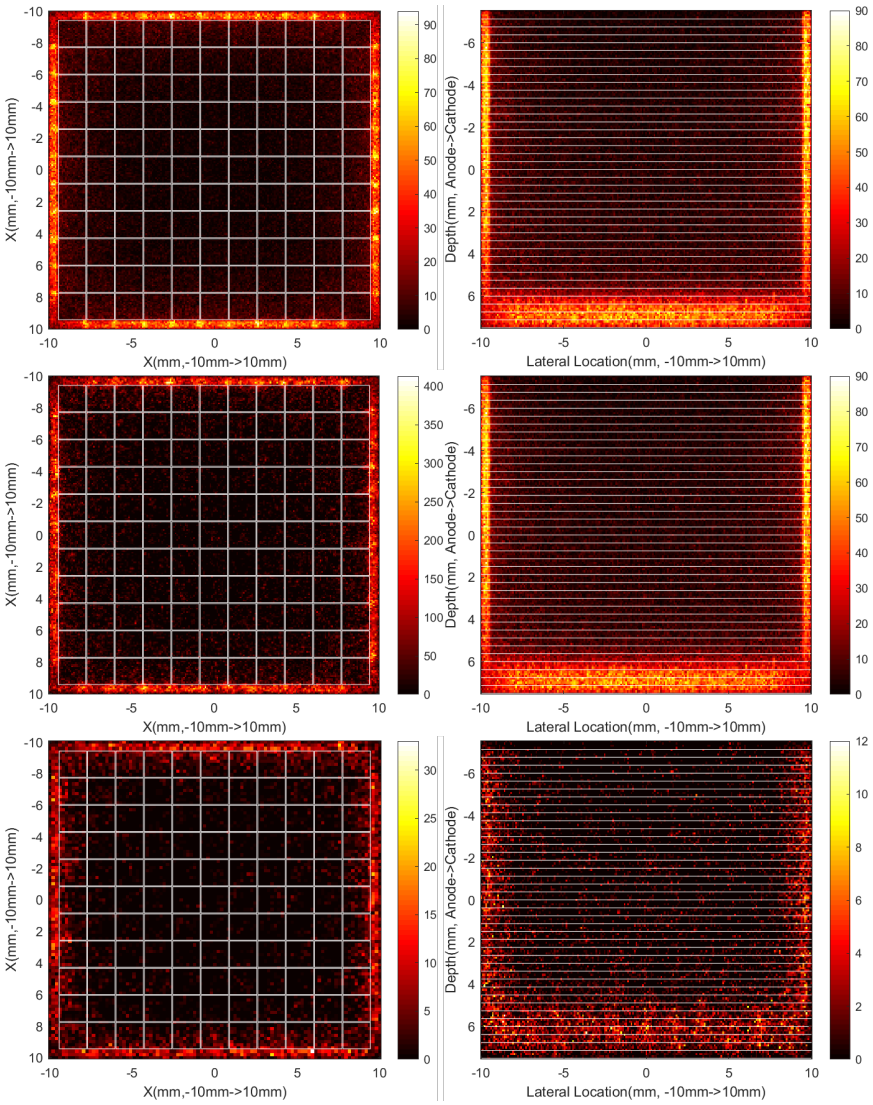

Figure 3.17: The location distribution of the lost two-pixel full-energy events from simulation. Top row is at 662keV. Middle row is at 1274keV. Bottom row is at 2614keV. In each row, the lateral position distribution is shown on the left. The pixel grids are marked in white for reference. The cross sectional view of the position distribution is shown on the right. The white lines mark the 40 depth bins.

width of guard ring dead region appears smaller. This factor will decrease the guard ring loss as the number of triggered pixels increases.

Second, the locations of multiple interactions in one radiation event are correlated [25]. At lower energy (e.g. 661.7 keV), the locations of two interactions tend to be close to

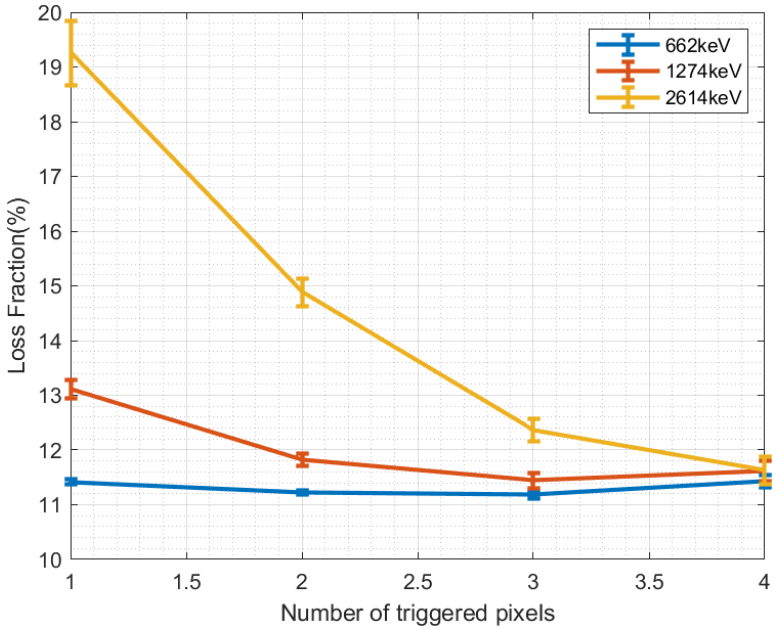

Figure 3.18: The significance of guard ring loss for different types of events at different energies

each other. However, at higher energy (e.g. 2614 keV), the mean distance between interactions in one event is larger, therefore, the two interactions appear more like two independent energy depositions and the loss probability is higher as the interactions spread out more. This factor will increase the guard ring loss as the number of triggered pixels increases.

Third, as number of triggered pixels increases, the initial incident energy are split between several interactions. In this case, even if one interaction falls into the "dead" region under the guard ring or the bad reconstruction layer near the anode, the whole event may still be reconstructed into the photopeak region. For example, if an event with 661.7keV initial energy deposited 621.7keV under one pixels and 40keV under the guard ring, even though the 40keV events are lost due to the dead region, this event can still be reconstructed into the photopeak region (600-700 keV). This factor will decrease the guard ring loss as the number of triggered pixels increases.

These three factors together caused the observed trend. At 661.7 keV, the second factor is the dominant one, so the loss fractions for different types of events are all around 11%. However, at higher energy like 1274 keV and 2614 keV, the first factor

starts to outweigh the other two, so the guard ring loss fraction decreases as the number of triggered pixels increases.

3.6 Application of the Efficiency Study

In SNM detection, it's critical to measure the enrichment of isotopes inside an object with minimum prior knowledge. Such precise quantitative measurement not only requires good understanding about the detection efficiency response of a radiation detector, but also requires accurate imaging capability.

The pixelated CZT detectors have already demonstrated its superior performance in both fields. On the one hand, the intrinsic photopeak efficiency of a single 20mm×20mm×15mm CZT detector from 59keV to 2614keV has been measured in experiment very accurately. The related results are presented in the previous sections. On the other hand, the positioning and imaging on the source can be achieved by either Compton imaging or TEI. Both have successfully imaged nuclear materials with high resolution before. Therefore, one can combine the superior imaging capability and the efficiency response of the pixelated CZT detector together aas a method to measure the isotope enrichment quantitatively with minimum knowledge about the object that is being measured.

In 2018, Dr. David Goodman imaged nuclear fuels with TEI [6]. And in 2016, Dr. Jiyang Chu reconstructed four point sources in a small room with Compton imaging[3]. In this part of study, the TEI is selected as a imaging method since it has better spatial resolution, and the scale of the field-of-view fits the problem better.

The TEI system consist of aCZT detector and a tungsten mask with specially designed patterns. During the measurement, the mask moves in a programmed trajectory. At any time, the recorded pattern on the detector is a convolution between the source and the mask pattern. After the experiment, a anti-mask is used to de-convolve the patterns. In this way, the source can be imaged. More elaboration of the TEI system can be found in reference [22].

Recently, the reconstruction algorithms of TEI was further improved by Dr. Daniel Shy [22]. With sub-pixel reconstruction included, the TEI can distinguish two sources with a separation of 1cm at the distance of 71 cm. Depth focusing technique can be used to estimate the distance of the source. When our guess deviates from the true value, the image resolution becomes worse. When the sharpest image (in terms of spatial

resolution) is observed, the guess to the distance between the source and the mask is the most probable location of the source.

3.6.1 Theory

The intrinsic photopeak efficiency at different energies have been measured and reported based on Equation (3.1). With known detection efficiency at a given energy, the activity of the source can be inversely calculated as

$$A = \frac{N}{\epsilon \cdot B \cdot T \cdot (1 - f_{dead}) \cdot \Omega} \tag{3.10}$$

In previous study, we have solved the problem of intrinsic photopeak efficiency ϵ. Estimating the source location d can be done by the depth refocusing technique of TEI. The geometry factor Ω is calculated from Equation.3.2.

If the absolute activity is not of concern, the enrichment [2] of any two radioactive isotopes M and N in an object can also be derived from Eq.3.10 as Eq.3.11.

$$\frac{A_M}{A_N} = \frac{N_M}{N_N} \cdot \frac{B_N}{B_M} \cdot \frac{\epsilon_N}{\epsilon_M} \tag{3.11}$$

Therefore, if the efficiency response of a detector is known, the enrichment of the two isotopes can be measured by recording the the photopeak counts in the characteristic peaks of each isotope. To use this equation, the dimension of the radioactive object being measured must be small enough so that they can be treated as point source. Under this condition, the common terms, such as the dead time fraction, geometry factor and measurement time will cancel out. The counts under each photopeak can be extracted from the measured energy spectrum.

3.6.2 Quantification with Prior Knowledge

Setup

A demonstration experiment has been performed in the laboratory with point check sources. The setup is shown in Figure 3.19. The sources are stacked together to create an artificial Uranium sample. Measured by a ruler, the precise location of each source is a

[2] The enrichment is the ratio between isotopes' absolute activities

prior knowledge at this stage. In this particular experiment, ^{22}Na and ^{57}Co were selected since their characteristic peaks (136keV of ^{57}Co and 1,274keV of ^{22}Na respectively) are close to those of ^{235}U and ^{238}U (185keV and 1,001keV respectively).

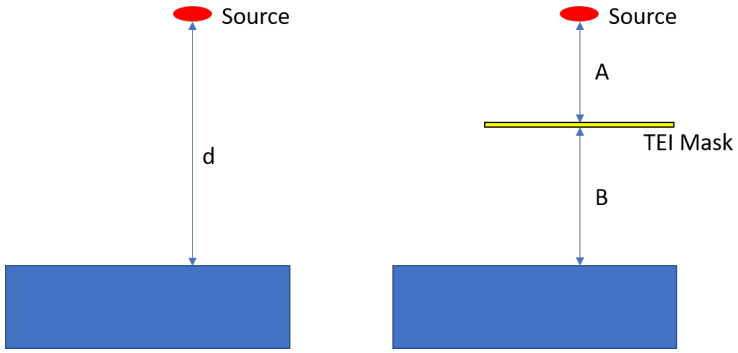

Figure 3.19: (Left) illustration of the experiment to benchmark the intrinsic photopeak efficiency without a mask at a given energy. (Right) illustration of the experiment to benchmark the intrinsic photopeak efficiency with a mask at a given energy and the experiment to measure an artificial Uranium sample.

In the experiment, the efficiency with and without the mask are both measured. The efficiency with the TEI mask was used to calculate the relative activity of the two sources.

Results

Figure 3.20 presents the collected spectra in the efficiency benchmark experiment (the lower two rows) and the enrichment measurement (the top row). In the enrichment measurement, the real activity of ^{22}Na and ^{57}Co were 29.00uCi and 39.39uCi (real enrichment is 74.26%). The measured enrichment according to Eq.3.11 is 74.28%, showing around 1% relative difference.

It should be noted that in this experiment, the sources are abstracted into point sources, which is rarely the case in any real-world circumstance. Self-attenuation is always a big factor to be taken into consideration.

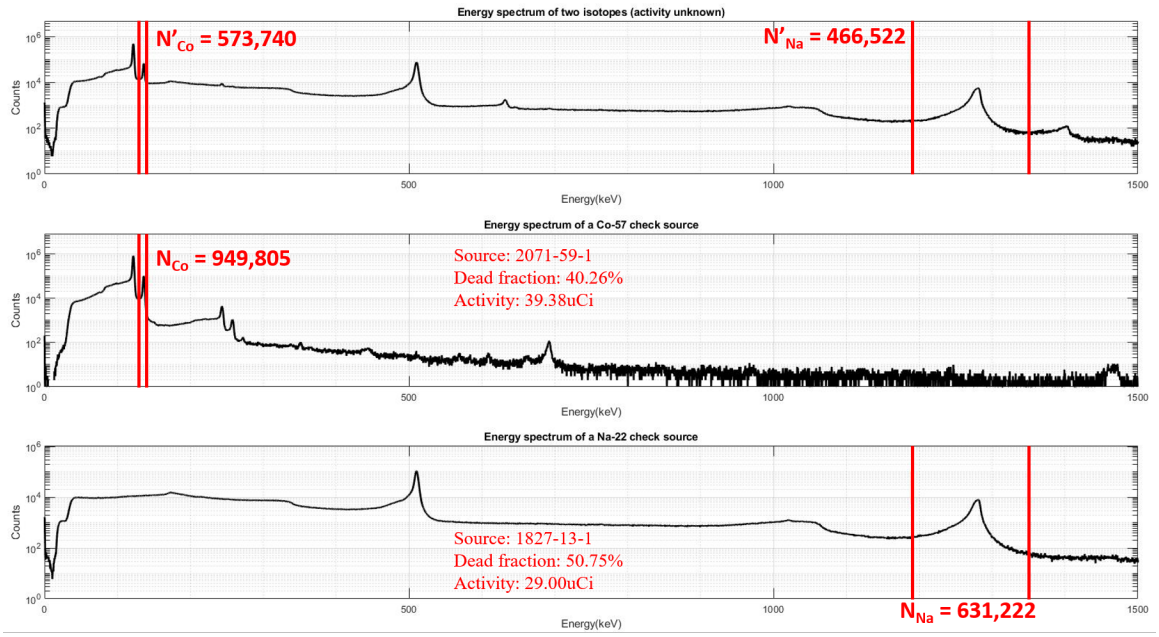

Figure 3.20: Measured spectra with (top) ^{22}Na and ^{57}Co combined, (middle) ^{57}Co only and (bottom) ^{22}Na only. The middle and bottom measurement is to benchmark the detection efficiency at those energies.

3.6.3 Quantification without Prior Knowledge

In applications where the source location is unknown before the field measurement, the source-to-mask distance A can be found by applying 3-D position estimation algorithm to the collected data set [22]. Such location estimation capability in 3-D space has been demonstrated before [22]. However, in previous study, this technique focused on low-energy gammas (below 300 keV).

To prove this technique is applicable to relatively higher energy gammas, an experiment with ^{137}Cs was set up to check whether TEI is capable of doing the 3-D imaging at 661.7kEv. The set up is the same as the setup shown in Figure 3.19 (right).

Figure 3.21 presents the reconstructed image of a ^{137}Cs source at 661.7keV with the estimated source-to-mask distance that gives the sharpest resolution. The estimated source-to-mask distance 56.9cm is close to the actual source-to-mask distance 55cm (less than 4% deviation). This result demonstrates,

1. The TEI system, though initially designed for imaging low-energy gamma sources, can work at higher energy with thick encoded mask and with enough statistics.

2. Even though due to lower statistics, the image quality is not as good as the image quality of the lower-energy peaks, the accuracy of the source location estimation is still good enough at higher energy.

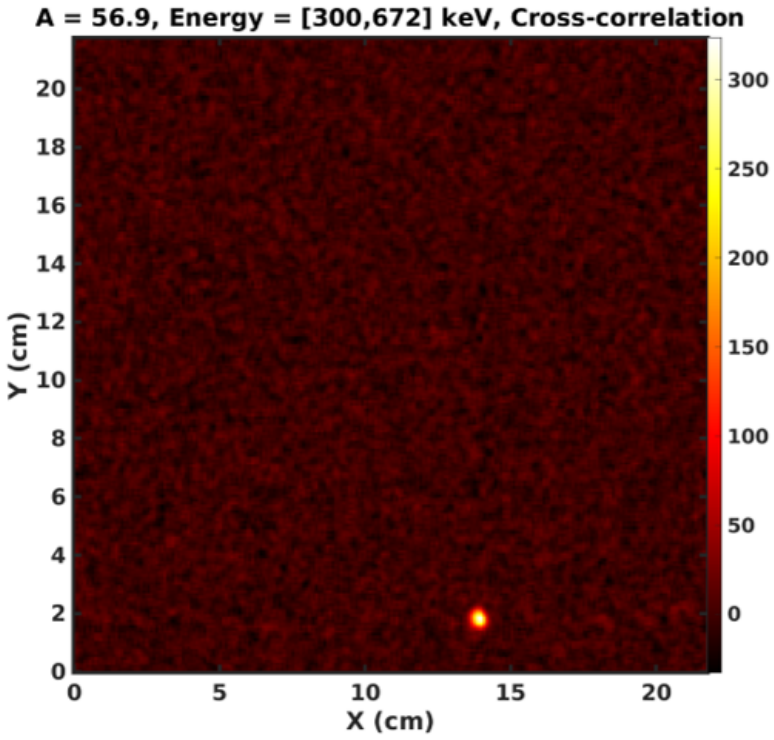

Figure 3.21: (Image reconstructed for a ^{137}Cs with the time-encoded imaging system. The source was 55 cm away from the TEI mask, and such distance is estimated as 56.9cm from TEI.)

Based on this finding, another quantification experiment where the location of the sources are unknown were conducted to further demonstrate the concept of quantification with TEI and the efficiency study results.

Setup

The experiment is set up as shown in Figure 3.22. ^{137}Cs, ^{22}Na and ^{152}Eu were placed at different locations in front of the mask. Their location will be estimated from the TEI instead of from the ruler.

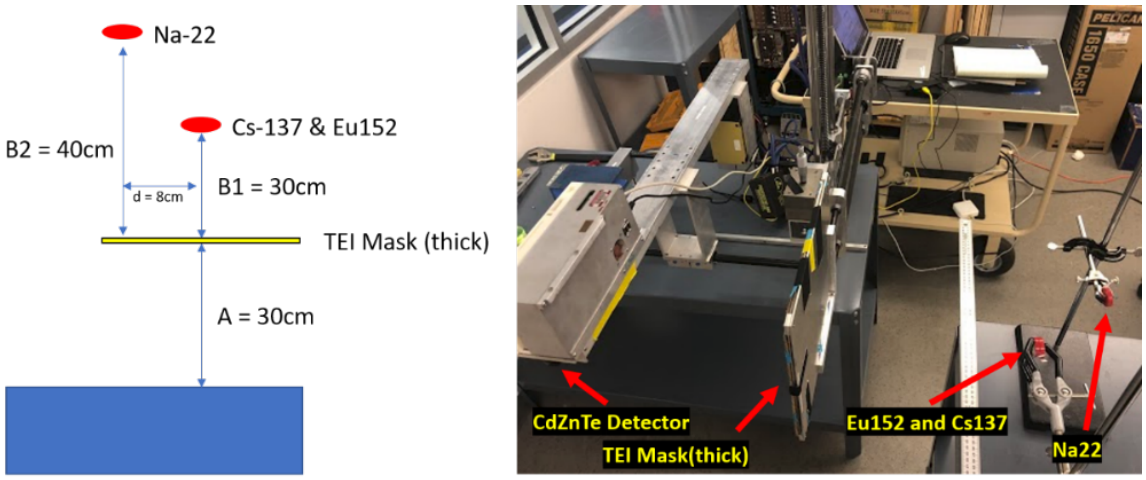

Figure 3.22: Experiment setup to verify that TEI can image above 1 MeV gammas and estimate the activity

Results

The imaging results on ^{152}Eu, ^{137}Cs and ^{22}Na are shown in Figure 3.23. As shown, although the signal-to-noise ratio is not ideal, the ^{22}Na and ^{137}Cs are still visible from the TEI reconstruction. Besides, compared with the actual location of those sources (30cm for the ^{152}Eu and the ^{137}Cs, 40cm for the ^{22}Na), the source-to-mask distance estimated from the depth-focusing technique had a maximum deviation of 1.5cm (Table.3.3).

From Eq. (3.10), the estimated activity of ^{152}Eu source is 14.61uCi, showing less than 1% difference. The estimated activity of ^{137}Cs is 22.44uCi, showing around 4% difference. The estimated activity of ^{22}Na is 22.32uCi, showing more than 10% difference. The excessive estimation bias is largely due to the poor signal-to-noise ratio in the TEI image. These experimental results laid a good foundation for future activity estimation algorithms.

3.6.4 Summary

The efficiency characteristics of the detector based on the latest digital ASIC have been studied [40]. With the recent advancement in the TEI technique, the detection system can locate the radiation source in 3-D space with good accuracy. When incorporated together, a system that can measure the enrichment of isotopes within an unknown object could be developed. In the presented demonstration experiment where the location of

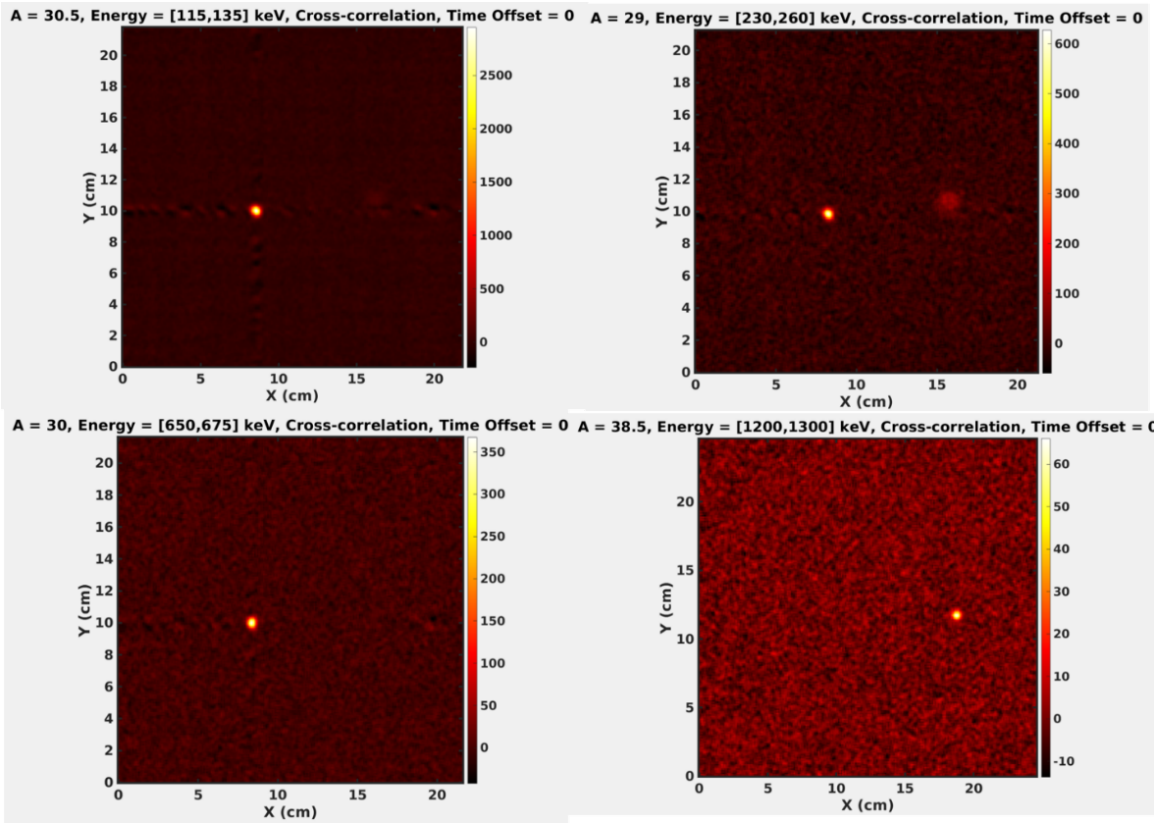

Figure 3.23: Images made from TEI reconstruction. (top left) Imaging result from the 121keV peak from the Eu-152. (top right) Imaging result from the 244keV peak from the Eu-152. The blurred point is the ghost artifact from the Na-22. (bottom left) imaging result from the 662keV from Cs-137. (bottom right) Imaging result from the 1,274keV peak from the Na-22.

the sources are prior knowledge to the user, the estimated enrichment match the actual enrichment of isotopes within 1% relative error.

In another demonstration experiment where the source location is unknown to the user, the source location can be estimated from TEI with less than 4% deviation. The estimated activity of the source match with their actual activity. Most of the deviation comes from the estimation to the source location.

The experiments also show that the TEI in 3-D space also works at higher energy. At energy as high as 1,273keV, even though the SNR decreased, the source can still be imaged and the 3-D lcation can be estimated.

Table 3.3: The summary of the quantification demonstration experiment with time-encoded imaging using CZT detector at room temperature.

Source	d_{actual}	d_{est}	A_{actual}	A_{est}
^{152}Eu (121keV)	30.00cm	30.50cm	14.72uCi	14.61uCi
^{152}Eu (244keV)	30.00cm	29.0cm	14.72uCi	12.48uCi
^{137}Cs (662keV)	30.00cm	30.00cm	25.33uCi	22.44uCi
^{22}Na (1,274keV)	40.00cm	38.50cm	25.10uCi	22.32uCi

3.7 Conclusion

The intrinsic photopeak efficiency of pixelated CZT system based on VAD_UM v2.2 digital readout ASIC is studied. The efficiency of a single 20mm×20mm×15mm CZT crystal is measured in experiment. The results are summarized in Table 3.1.

A detailed simulation program that includes the physical process, signal induction and event reconstruction was developed. Efficiency loss mechanisms are identified with the help of this computer program. The simulated and measured intrinsic photopeak efficiency match with each other considering the system uncertainties. Due to the low energy tail of the photopeak that originates from imperfect event reconstruction, the selection of the photopeak region can cause efficiency deficit. The region under the guard ring is basically dead in terms of photopeak efficiency response. On the anode side, due to the rapid variation of the weighting potential in one depth bin, the reconstruction of full-energy deposition event may fail and cause efficiency loss.

The efficiency measurement and simulation also provide useful guidance in future radiation detector design. First, to improve the efficiency performance at low energy region, the passive materials between the source and the detector crystal should be minimized. Second, a single larger thicker crystal is preferable, since the region covered by the guard ring and the layer in which weighting potential shows significant variation are relatively smaller. The efficiency loss will be less significant in this case.

Two demonstration experiments to showcase the quantitative measurement capability of the CZT detectors were presented. With the 3-D imaging capability of TEI, the location of the point source can be determined within an uncertainty of 0.5cm. The activity of the source can be estimated from the energy spectrum with an uncertainty

of 10%.

4 The Design of the DSP ASIC

4.1 From System-on-board to System-on-chip

The development of the electronics of the nuclear detection system has come a long way. The simplest nuclear detector is probably the Geiger Muller Counter. Its back-end electronics measures the current induced by the movement of radiated charge and reflect that on an analog gauge. A more advanced design will include a loudspeaker so that the radiation intensity can be identified from the frequency of the beeping sound. Another category of radiation detector is the scintillator. It convert the radiation into light (either visible or ultraviolet), and is further converted to electric signal and amplified by Photo-Multiplier Tube (PMT). The data acquisition and processing are usually on a computer.

As for the semiconductor-based radiation detectors, the golden standard in the detection field, the HPGe detector, requires large storage space and long preparation time due to the mandatory cooling. In the early 2000s, the development of radiation sensors based on CZT allows room-temperature operation. This eliminated the requirement of the Nitrogen cooling, therefore reducing the power and dimension significantly. It also increases its portability and quick field deployment is made possible. But to keep the detector performance stable, a constant ambient temperature still need to be maintained by power-consuming modules like peltier and cooling fans.

The aforementioned hardware design of all detectors are based on the concept of "system-on-board". General-purpose FPGA and CPU were employed to provide processing power to the system. Engineers and scientist integrate discrete electrical components on printed circuit board. The dimension of the assembled systems are usually big. Besides, to supply the power required by different components that are not always in use, the overall power consumption is not optimal.

As the way any modern technologies evolve, there's a constant desire to improve the performance, decrease the physical dimension and reduce the power consumption of the system. With careful engineering, all the disadvantages of "system-on-board" can be minimized. But after all, these hardware are not tailored to this specific application, eventually there's only a limited room for further improvement and optimization. The System-on-Chip (SoC) becomes the most suitable solution for next-generation radiation detector naturally.

The circuits to readout signals from a pixelated CZT detector can be categorized into two categories, the front-end circuits and the back-end circuits. **The front-end circuits** are typical closer to the detector and handle the signal sampling. Typical modules in this category include the low-noise charge amplifier, the sampling cell array (SCA) and the event trigger logic. The front-end circuits are mostly analog, while some digital blocks exist to assist the control or configuration. If a module is closer to the detector, the noise contribution from that circuits contributes more to the whole system. Therefore, to minimize the noise, these modules are typically designed and laid-out manually. Example ASICs of this category are the H3DD_UM ASIC, which has matured in the past few years. **The back-end circuits** are usually relatively far away from the detector and process the digitized waveform, which is the output from the front-end circuits. Typical modules includes the ADC, analog or digital filters, and the communication peripherals. The back-end circuits are mostly digital. They are responsible for processing the digitized analog waveform. Example ASICs of this category in the radiation detection field is rare. The trend in this research field is still to integrate commercial processing unit (either CPU or FPGA) into the system, instead of designing the digital processing chip in-house.

The first attempt to design the ASICs for radiation detectors started in the front-end category. The need to handle the large number of readout channels of the pixelated detectors strongly motivates such development. With the system-on-board style design, it's almost impossible to accommodate such large amount of components on a compact PCB. Besides, to achieve the lowest possible electrical noise, the front-end amplifiers must be tailored on the silicon level. Another advantage of customizing the front-end ASIC is to gain additional performance by adding features that are specific to the application (configurable dynamic range, hysteresis trigger etc.). The state-of-the-art front-end ASIC are the VAD_UM v2.2 from IDEAS and the H3DD_UM v4.0 ASIC

developed in the UM.

Though the front-end ASIC has been evolving for quite a long time, the interest in developing the back-end processing ASIC was not as strong. If compared to the development of the front-end ASIC, the development of the back-end ASIC has been very quiet in the radiation detection field. The state-of-the-art radiation detector are still built with separate general-purpose processing units.

The first reason is, due to the high-demand and fierce market competition, the commercial processors iterate fast. The processing power of micro-processor has improved by around four orders of magnitude in the past three decades [10]. Deep integration of processing cores with IC contributed the most to the smaller form factor and the lower power consumption. The second reason is, the prices of commercial processors remain reasonably low due to the large manufacturing volume. The detection community had the option to be the free-rider to the development of the consumer electronics, and lacked the motivation to devote money and time to develop back-end processing ASICs.

However, as the detector technology evolves, the further needs to shrink the size, reduce the power consumption and integrate more functionality into the system grow stronger. Based on the philosophy that "pay only what you need", the concept of SoC gradually outweigh the concept of "system-on-board". The basic concept and benefit of SoC have been elaborated in the Chapter II. Design the back-end processing ASIC in house is the new direction to explore in the radiation detection field.

Historically, ASICs are developed for radiation detectors of various architectures. In Chapter II, the history of the readout (front-end) ASIC at UM has been reviewed. In this chapter, Table 4.1 walks through the history of ASIC or SoC applications in the radiation detection field from early 1980s to most recent years by other research groups.

Table 4.1: A Literature Review of the SoC Development for Radiation Detectors

No.	Author	Title
1	You, Bihui, et al. [32]	A distributed readout network ASIC for high-density electrode array targeting at neutrinoless double-beta decay search in a Time Projection Chamber
2	Guo, Yuxiang, et al. [8]	Design of a Time-to-Digital Converter ASIC and a mini-DAQ system for the Phase-2 upgrade of the ATLAS Monitored Drift Tube detector
3	Wang, Jia, et al. (2021) [26]	Design of a low-noise, high-linearity, readout ASIC for CdZnTe-based gamma-ray spectrometers
4	Ahmad, S., et al. (2021) [2]	OMEGA SiPM readout ASICs
5	Prasad, K. Hari, et al. (2021) [20]	A versatile multi-hit, multi-channel Vernier time-to-digital converter ASIC
6	Kawamura, Tenyo, et al. (2020) [12]	Development of a low-noise front-end ASIC for CdTe detectors
7	Contino, G. I. O. V. A. N. N. I., et al. (2020) [4]	An ASIC front-end for fluorescence and Cherenkov light detection with SiPM for space and ground applications

Continued on next page

Table 4.1 – continued from previous page

No.	Author	Title
8	Hernandez, Hugo, et al. (2019)[11]	A Monolithic 32-Channel Front End and DSP ASIC for Gaseous Detectors
9	Vernon, Emerson, et al. (2019)[24]	Front-end ASIC for spectroscopic readout of virtual Frisch-grid CZT bar sensors.
10	Rivetti, A., et al. (2019) [21]	TIGER: A front-end ASIC for timing and energy measurements with radiation detectors
11	Ulyanov, Alexei, et al. (2017) [23]	Using the SIPHRA ASIC with an SiPM array and scintillators for gamma spectroscopy
12	Pasquali, G., et al. (2007) [19]	A DSP equipped digitizer for online analysis of nuclear detector signals
13	Yamamoto, et al. (2000) [30]	Development of a DSP-based real-time position calculation circuit for a beta camera
14	Wixted, R. L., et al. (1997) [28]	A 32-channel waveform sampling FASTBUS DSP processing

Most of the research focused on integrating the front-end readout electronics. Only a small portion of the back-end processing algorithms are transferred to the silicon die. Besides, all these chips used the old process node (above 130nm), which didn't utilize the recent advanced IC technology.

4.2 Architecture of the Chip

4.2.1 Hardware/Software Partitioning

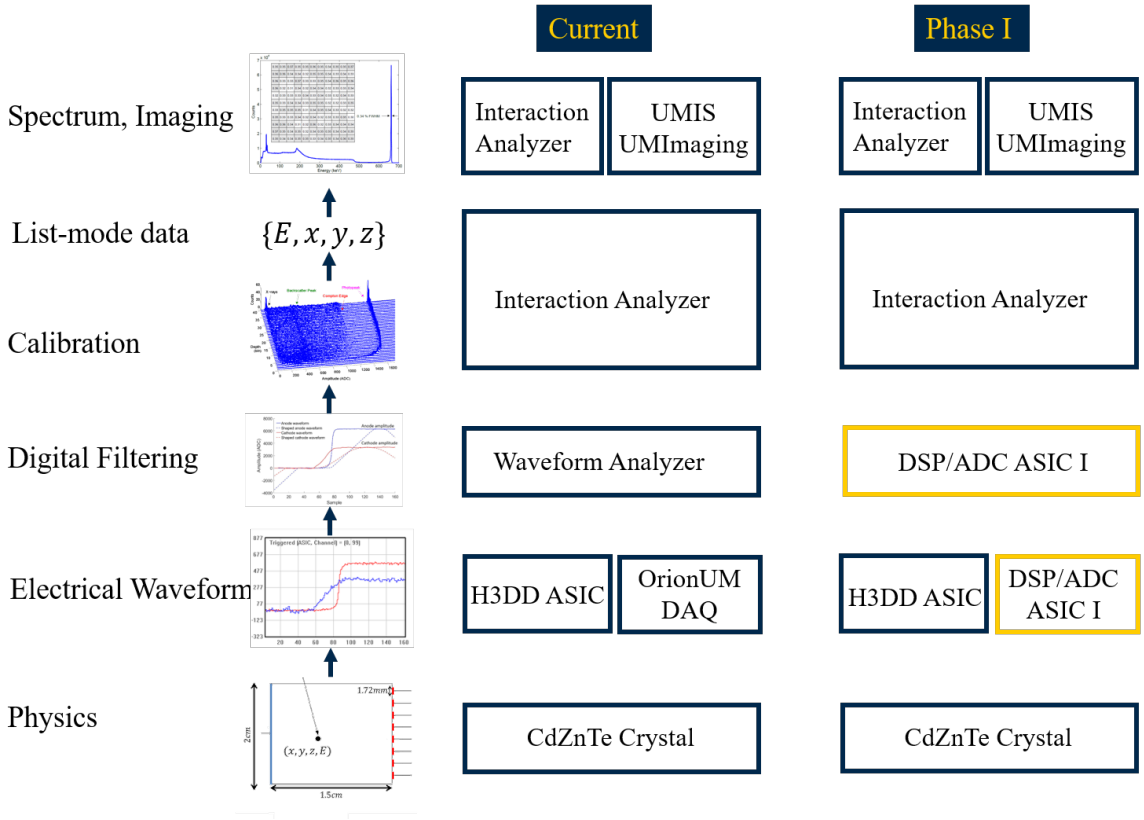

Figure 4.1: The system-level design of the whole radiation detection system

From the physical interactions between the gammas and the detection medium, to the energy spectrum with sharp peaks, the signal needs to be processed in multiples stages. Figure 4.1 summarizes the flow of information from the physics level to the energy spectrum level and imaging level.

The ionization and charge drift are governed by physics, as described in the Chapter I of the thesis work. Controlled by the data acquisition software, the electrical signals are sensed by the front-end ASIC (e.g. H3DD_UM front-end ASIC) and are sampled with high precision and low noise. After this stage, the signal is discrete in the time domain, but still continuous in the amplitude domain. The output from the front-end ASIC are quantized by the on-board ADCs.

After becoming discrete in both time and amplitude domain, the electrical waveform are shaped by digital filters to achieve the best signal-to-noise ratio [38]. In the current system at UM, the digital filtering are implemented at the software level in the WaveformAnalyzer. Different filter kernels are used for different purposes. More details about the digital signal processing algorithms can be found in reference [38] and reference [29].

Calibrations are made on per-voxel basis in the InteractionAnalyzer. The detailed calibration algorithms are described in reference [37] and reference [29]. Eventually, the processed data containing the calibration energy and 3-D position of the event is calculated based on the calibration data. In the end, the energy spectrum is obtained after collecting millions of events. If needed, the list-mode data and the energy spectra can be further processed by the imaging softwares (either UMIS or UMImaging) to create images showing the distribution of the radiation [3].

In the whole flow, the front-end ASIC, the ADC, the FPGA, the CPU and the processing software are responsible for a portion of the processing workflow and are integrated on the board level. Although after years of integration, the system is already very compact, like the Orion-α system described in the previous chapter. Further reduction in size and power consumption can be achieved by higher-level integration by integrating them onto an ASIC.

Designing a back-end processing ASIC for the radiation detection system doesn't mean integrating everything on the silicon. Some processing like calibration, sub-pixel reconstruction and imaging algorithms are computation-intensive, and are still more economical to be implemented on the software level or on the FPGA fabrics.

The design project is separated into two design phases. In the first phase, the goal is to design an ASIC that integrates several functional units onto a single silicon die. Each DSP/ADC ASIC will be connected with one front-end H3DD_UM ASIC and one CZT crystal as a detection unit. The DSP/ADC ASIC should be able to,

1. control the front-end H3DD_UM ASIC. Previously the data acquisition was controlled by a digital module, named as H3DD_Core, which was implemented on the FPGA fabric. This module is a relatively matured IP, but modifications are required to transfer it onto the silicon.

2. digitize the sampled waveform with the on-chip ADC. The front-end ASIC is only responsible for sampling the charge on the electrodes with high precision. Different

from the previous system-on-board design, an high-resolution ADC will be used for digitizing the data into digital codes.

3. apply digital filters to the digitized waveforms and extract the timing and amplitude information. Although the digital filters have been well optimized in the past decade, it's still important that the filter coefficients are field-programmable since this ASIC might be used to process the data from other detectors.

4. communicate with an external device so that the data can be sent out for future processing. It's desired that the ASIC can be configured through this communication interface as well.

In the next design phase, the problems identified in the first design phase will be fixed. More processing power like event calibration or event processing will be integrated in the future.

4.2.2 Design Specification

The design specification of the DSP/ADC ASIC are summarized in Table 4.3

Table 4.2: The mode of the signal monitor

Power	Die Area	Package	Speed	Vdd
$\leq 1W$	$\leq$5mm×5mm	60-pin QFN	100MHz	3.3V/1.2V

4.2.3 The Architecture

Figure 4.2 presents the architecture of the DSP/ADC ASIC in the first design phase. In the figure, the DSP/ADC ASIC is marked by the red rectangle.

Functional units are shown as the black boxes within the red rectangle. Their connections are marked by the arrows. The flow of information are shown by the blue arrows (control flow) and the green arrows (data flow). In the figure,

1. H3DD is the front-end ASIC, by which the electrical signals on the electrodes are sampled with high precision and low noise.

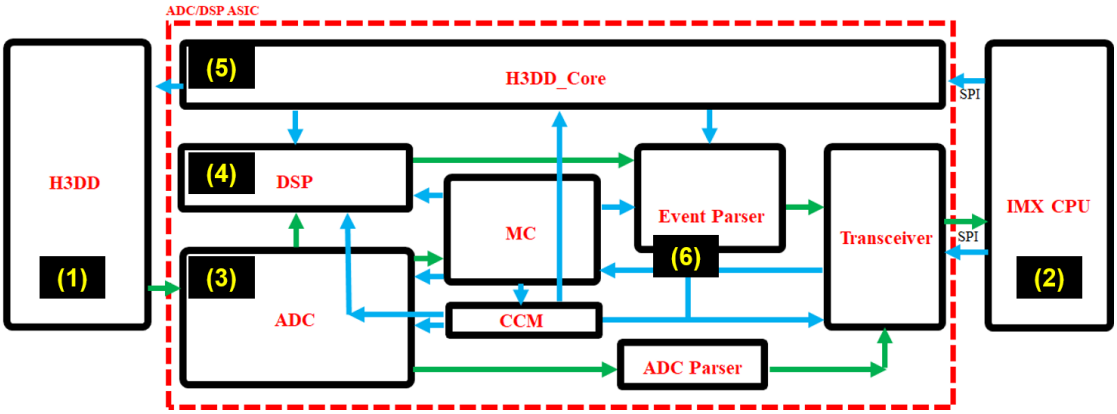

Figure 4.2: The archetecture of the DSP/ADC ASIC. The DSP/ADC ASIC contains the functional units enclosed by the red square. The function units and the connection among them inside the DSP/ADC ASIC.

2. A upper-stream device (usually a CPU) is responsible for acquiring the processed data from the DSP/ADC ASIC. The output of the DSP/ADC ASIC is further processed by it to obtain the calibrated energy spectrum. The DSP/ADC ASIC also receives command from it so that proper configuration can be made.

3. The ADC inside the DSP/ADC ASIC is for the analog-to-digital conversion.

4. The DSP inside the DSP/ADC ASIC applies different digital filters to the digitized waveform.

5. The H3DD_Core inside the DSP/ADC ASIC controls the operation of the front-end ASIC and coordinates the data processing.

6. various modules in the supporting circuits provide peripheral functions (control, clock, data parsing, communication).

The control flow is marked by the blue arrows in the figure. The upper-stream device (usually a CPU) sends instructions to the ASIC through the SPI. Since the SPI and the on-chip main controller don't necessarily work synchronously, the transceiver acts as a intermediate module to handle the clock domain crossing. After the main controller receives the command, it decodes the instruction and distributes the instruction to relevant modules accordingly.

The data flow is marked by the green arrows in the figure. The sampled waveform is sent from the front-end ASIC to the ADC. The digitized value is transferred to the DSP for digital filtering and the timing and amplitude information is extracted. Then the data package are transferred to the event parser to be packaged. In the end, they are transferred to the CPU via the transceiver. The transceiver still acts as a intermediate module to cross the clock domains.

4.2.4 Functional Units

ADC

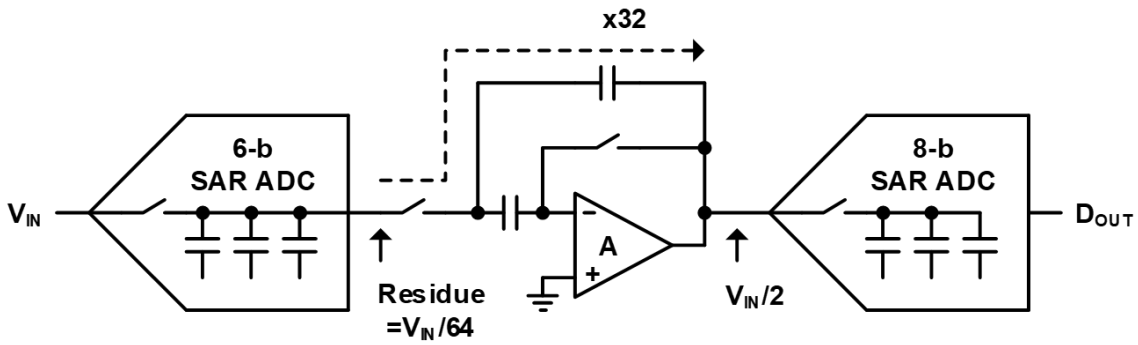

Figure 4.3: The architecture of the ADC. The first stage digitize the input with 6-bit precision and the residual voltage gets amplified by the residual amplifier. The second stage furthre digtitize the amplified residual. The output 13-bit code is assembled from the output of the two stages with one bit overlap.

The ADC is designed by Seungheun Song in Professor Michael Flynn's group. Only the basic design will be described in this thesis work.

The block diagram of the ADC is shown in Figure 4.3. The ADC is a two-stage Successive Approximation (SAR) ADC. The input analog voltage is digitized by a 6-bit coarse SAR ADC. The residual voltage is amplified by a ring amplifier and then fed to the second stage. The second stage is a 8-bit fine SAR ADC. Eventually, the digitized codes are combined. The resolution of the ADC is 13 bits and the target effective number of bits (ENOB) is larger than 11.5 bits. The target sampling rate is 12.5 MSPS. The layout of the ADC is shown in Figure 4.4 and its dimensions are 355um×195um. The operation timing diagram of the ADC are shown in Figure.4.5.

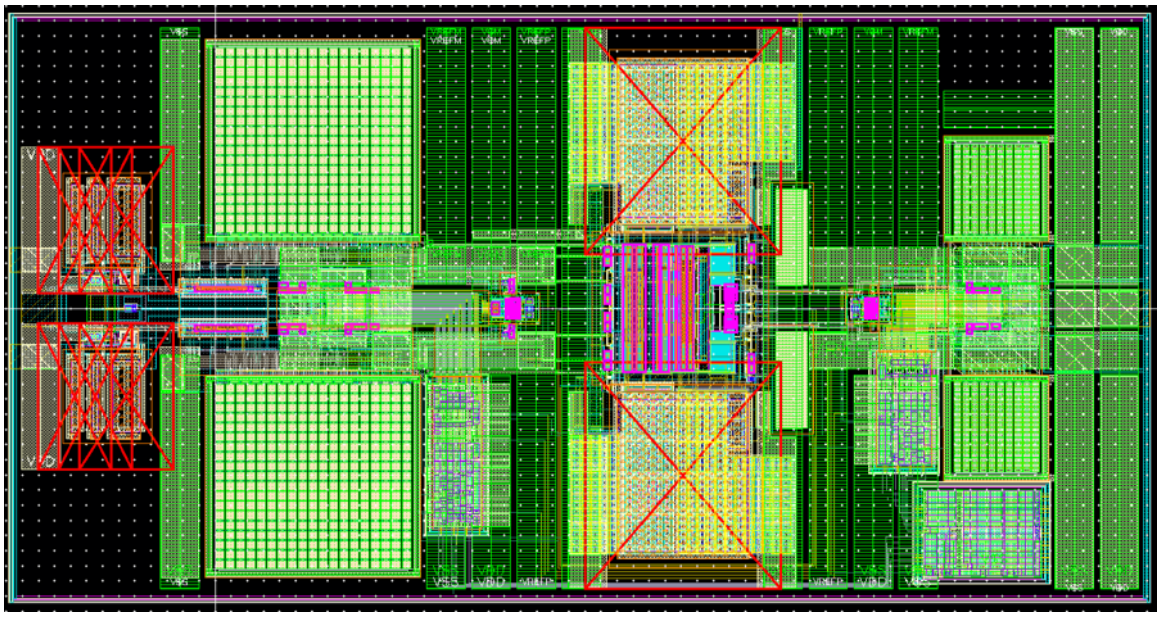

Figure 4.4: The physical layout of the ADC. The overall dimensions are 355um×195um.

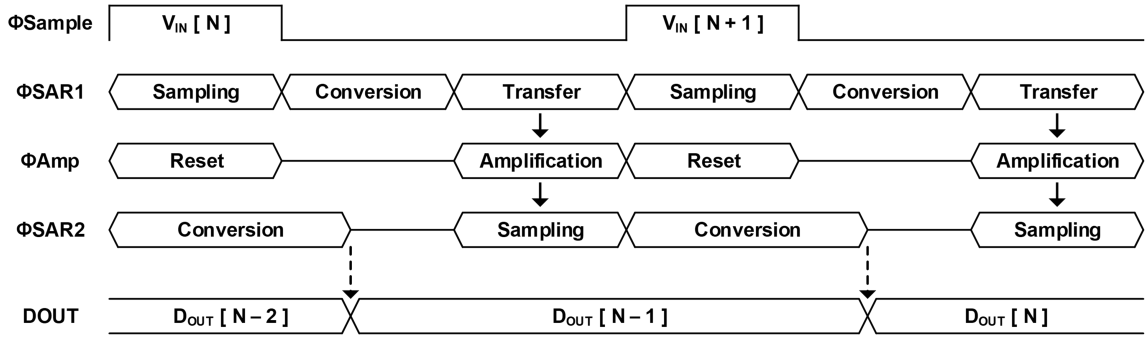

Figure 4.5: The timing diagram of the ADC. This ADC requires a clock with 25% duty cycle.

DSP

The DSP is the largest and the core module on the ASIC. It is designed by Matthew Petryk in the Orion Research Group. Only the basic design will be described in this thesis work.

The DSP has three stages. The first stage finds the baseline of each channel by averaging a programmable number of samples. Once the baseline is detected, baseline subtraction and front-end pre-amplifier decay correction will be performed on the in-

coming waveform. The second stage consists of three pairs of digital filters. Each pair has a trapezoidal filter to calculate the amplitude of the waveform , and an CR-$(RC)^4$ filter to calculate the timing of the waveform. Since the collecting anode, the non-collecting anode and the cathode require different filter coefficients, three filter pairs are designed to process the data. The third stage takes the output of the second stage and pick up the timing and amplitude information from the filtered waveform. Its output is sent to the event parser (described later) and transmitted to the computer for more processing. The layout of the DSP is shown in Figure 4.21. The overall dimensions are 2900um×2900um.

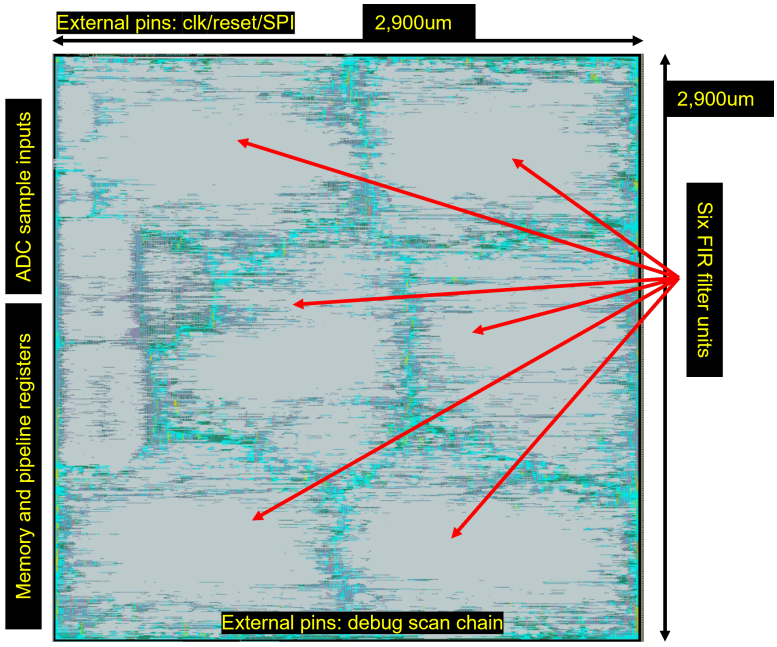

Figure 4.6: The layout of the DSP. The overall dimensions are 2900um×2900um.

Since the filter coefficients of the DSP need to be programmable, an interface is designed to configure the DSP. The timing diagram to configure the DSP is shown in Figure 4.7.

After digitized by the on-chip SAR ADC, the electrical waveform are sent to the DSP for processing. Meanwhile, information like the channel index, timestamps are collected from the H3DD_Core. Based on the property of the channel index (collecting pixel, neighboring pixel), different filter kernels are applied. The timing diagram of the DSP when it takes the input are shown in Figure 4.8.

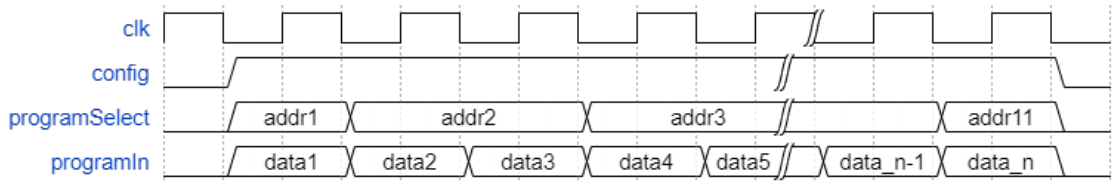

Figure 4.7: The DSP timing diagram during its configuration

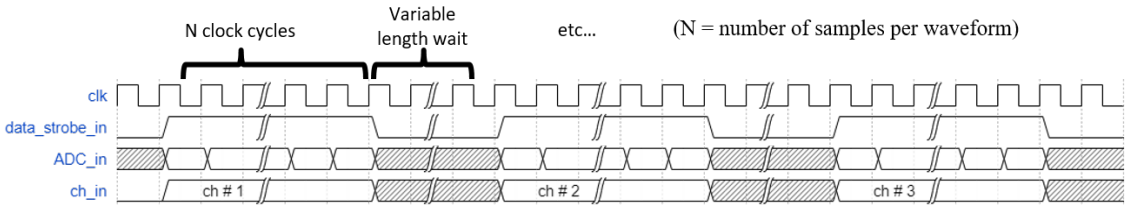

Figure 4.8: The DSP timing diagram during its operation (input)

After the data processing, the amplitude and the timing information of the waveform are presented at the output ports. The event parser will collect them and parse them into packet that can be transmitted by the transceiver. The timing diagram of the DSP when it outputs the processed data are shown in Figure 4.9.

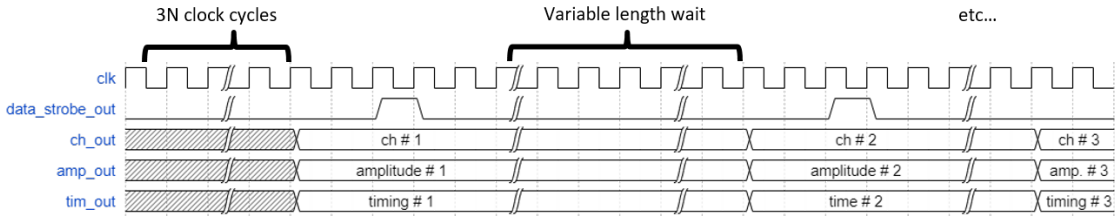

Figure 4.9: The DSP timing diagram during its operation (output)

H3DD_Core

The H3DD_Core is a module that the DSP/ADC ASIC inherited from the previous FPGA-based detection system. Not only does it control the front-end H3DD_UM ASIC, but it also controls the digitization of the ADC and the data processing sequence of the DSP. Previously, this module was implemented on FPGA platform and was written in VHDL. To integrate this module into the DSP/ADC ASIC, modifications were made to the legacy code.

The interface with the upper-stream device is to receive instructions. To de-couple this module from the rest of the digital modules, the H3DD_Core receives commands from external device through a dedicated SPI. The interface to the H3DD_UM front-end ASIC are six differential pairs (details can be found in its user manual). Altogether, these signals control the configuration, readout sequence of the detection system. The details of the interface can be found in the technical user manual of the H3DD front-end ASIC. The layout of the H3DD_Core module is shown in Figure 4.10. The layout dimension is 1000um×500um.

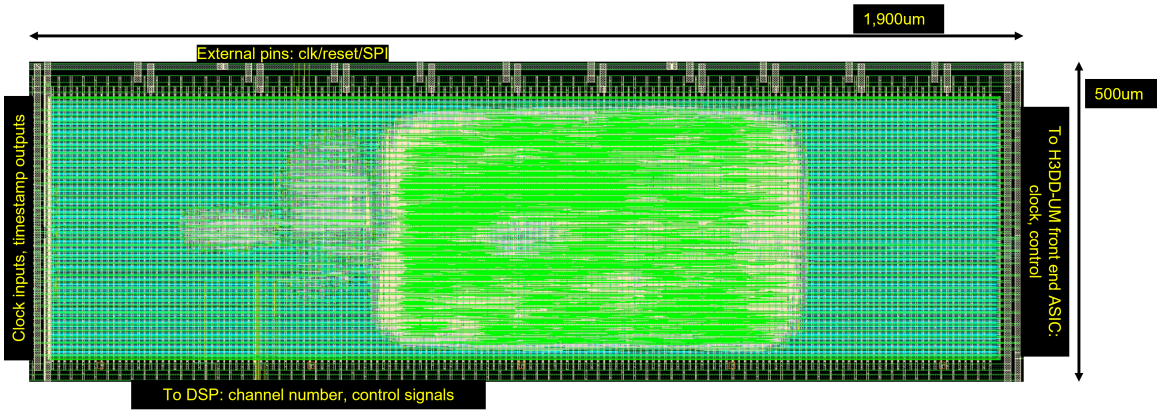

Figure 4.10: The layout of the H3DD_Core.

Main Controller

Although the data processing algorithms are implemented in the silicon, it is still desirable to reserve some flexibility to control and tune the parameters. The filter coefficients and some parameters are made programmable through the SPI. It is the main controller (MC) that is responsible for receiving the external instructions and controlling all functional units. The functions of the main controller include,

1. talking with the SPI receiver to receive instructions. The main controller is responsible for recognizing and decoding the instruction based on the instruction address (Figure 4.13).

2. issuing commands to the clock control module (CCM) to adjust the operation clock of all sub-modules. This includes the frequency, duty cycle and phase delay of each clock.

3. altering the value of registers that configures the SAR ADC.

4. issuing command to the DSP based on the SPI instruction, such that the internal filter coefficients can be programmed.

The block diagram of the main controller is presented in Figure.4.11.

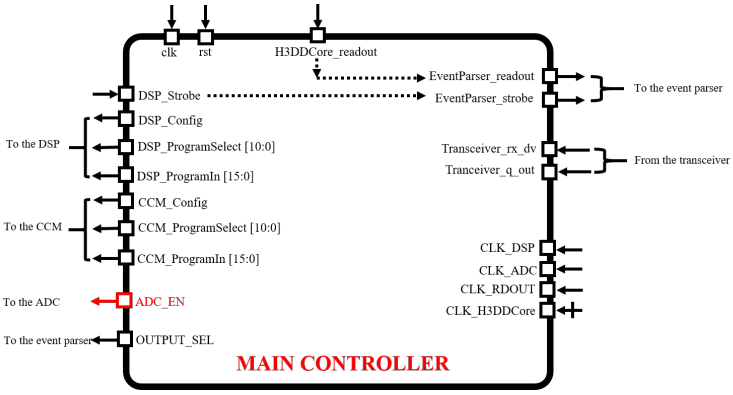

Figure 4.11: The block diagram of the main controller

To aid the design of the main controller, especially to simplify the design of the SPI in the trans-receiver, a customized instruction set is designed. Figure 4.12 summarizes the instruction set of the DSP/ADC ASIC. The instructions are all 32-bit wide, so that each single instruction can fit in the 32-bit FIFO in the SPI receiver. The instructions are decoded prior to being distributed to each functional units. The decoding logic is illustrated in Figure 4.13.

The main controller decodes the command based on the first 16 bits of the 32-bit instruction (the address of the instruction). The last 16 bits of the 32-bit instruction carry the value that we intend to send (the payload of the instruction). The 32-bit instruction is passed from the SPI receiver to the decoding logic in the main controller. The decoding logic translates the instruction based on the instruction set and output the correct logic sequence, so that the DSP, the ADC, the clock control module (CCM) etc. can react accordingly.

Clock Control Module

The Clock Control Module (CCM) is to provide the on-chip modules with their required clocks. As presented before, multiple modules exist in the ASIC and they need their own

Type	Command[31:0]	Length (bits)	Note	Module - 4 bits				FLG	Program Select -- 11 bits										Programm In -- 16 bits														
				31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
D	Threshold Compare -- Threshold Settings	16		x	x	x	x	1	0	0	0	0	0	0	0	0	0	1		int 16															
D	Threshold Compare -- Filter Coefficients	256*16	needs 256 cycles	x	x	x	x	1	0	0	0	0	0	0	0	0	1	0		int 16															
D	Neighboring Pixel -- Filter Coefficients	256*16	needs 256 cycles	x	x	x	x	1	0	0	0	0	0	0	0	1	0	0		int 16															
D	Collecting Pixel -- Timing pickoff start	8		x	x	x	x	1	0	0	0	0	0	0	0	1	0	0		0	0	0	0	0	0	0	0	int 8							
D	Collecting Pixel -- Fast filter coefficients	256*16	needs 256 cycles	x	x	x	x	1	0	0	0	0	0	1	0	0	0	0		int 16															
D	Collecting Pixel -- Slow filter coefficients	256*16	needs 256 cycles	x	x	x	x	1	0	0	0	0	1	0	0	0	0	0		int 16															
D	Cathode -- Timing pickoff start	8		x	x	x	x	1	0	0	0	1	0	0	0	0	0	0		0	0	0	0	0	0	0	0	int 8							
D	Cathode -- Fast filter coefficients	256*16	needs 256 cycles	x	x	x	x	1	0	0	1	0	0	0	0	0	0	0		int 16															
D	Cathode -- Slow filter coefficients	256*16	needs 256 cycles	x	x	x	x	1	0	0	1	0	0	0	0	0	0	0		int 16															
D	Sample Skip Setting	7		x	x	x	x	1	0	1	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	int 7						
D	Baseline length Setting	4		x	x	x	x	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	int 4				0	0	0	0	0	0	0
D	Rolling average enable/disable	1		x	x	x	x	1	0	1	0	0	0	0	0	0	0	0	int 1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	Front-end decay parameter	130*10	needs 130 cycles	x	x	x	x	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	int 10									
S	Freeze	--		x	x	x	x	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
S	Software-level reset command	--		x	x	x	x	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
S	Poweroff ADC	1		x	x	x	x	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	int 1
S	Select output (wf or filtered data)	1		x	x	x	x	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	int 1
S	Program the header (top 16-bit)	16	(reserved)	x	x	x	x	0	0	0	0	0	0	0	1	0	0	0	0	int 16															
S	Program the header (bottom 16-bit)	16	(reserved)	x	x	x	x	0	0	0	0	0	0	1	0	0	0	0	0	int 16															
S	Program the tail (top 16-bit)	16	(reserved)	x	x	x	x	0	0	0	0	0	1	0	0	0	0	0	0	int 16															
S	Program the tail (bottom 16-bit)	16	(reserved)	x	x	x	x	0	0	0	0	1	0	0	0	0	0	0	0	int 16															
C	ADC clock frequency	16		x	x	x	x	0	0	0	1	0	0	0	0	0	0	0	0	EN	DELAY [5:0]						Duty Cycle [4:0]					Frequency[4:0]				
C	DSP clock frequency	16		x	x	x	x	0	0	1	0	0	0	0	0	0	0	0	0	EN	DELAY [5:0]						Duty Cycle [4:0]					Frequency[4:0]				
C	Readout Module clock frequency	16		x	x	x	x	0	0	1	1	0	0	0	0	0	0	0	0	EN	DELAY [5:0]						Duty Cycle [4:0]					Frequency[4:0]				
C	H3DD_Core clock 100MHz	16		x	x	x	x	0	1	0	0	0	0	0	0	0	0	0	0	EN	DELAY [5:0]						Duty Cycle [4:0]					Frequency[4:0]				
C	H3DD_Core clock 50MHz	16		x	x	x	x	0	1	0	1	0	0	0	0	0	0	0	0	EN	DELAY [5:0]						Duty Cycle [4:0]					Frequency[4:0]				
C	H3DD_Core clock 25MHz	16		x	x	x	x	0	1	1	0	0	0	0	0	0	0	0	0	EN	DELAY [5:0]						Duty Cycle [4:0]					Frequency[4:0]				
C	H3DD_Core clock 6.25MHz	16		x	x	x	x	0	1	1	1	0	0	0	0	0	0	0	0	EN	DELAY [5:0]						Duty Cycle [4:0]					Frequency[4:0]				
R	Readout	---	(special command)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

Figure 4.12: The instrution set of the DSP/ADC ASIC

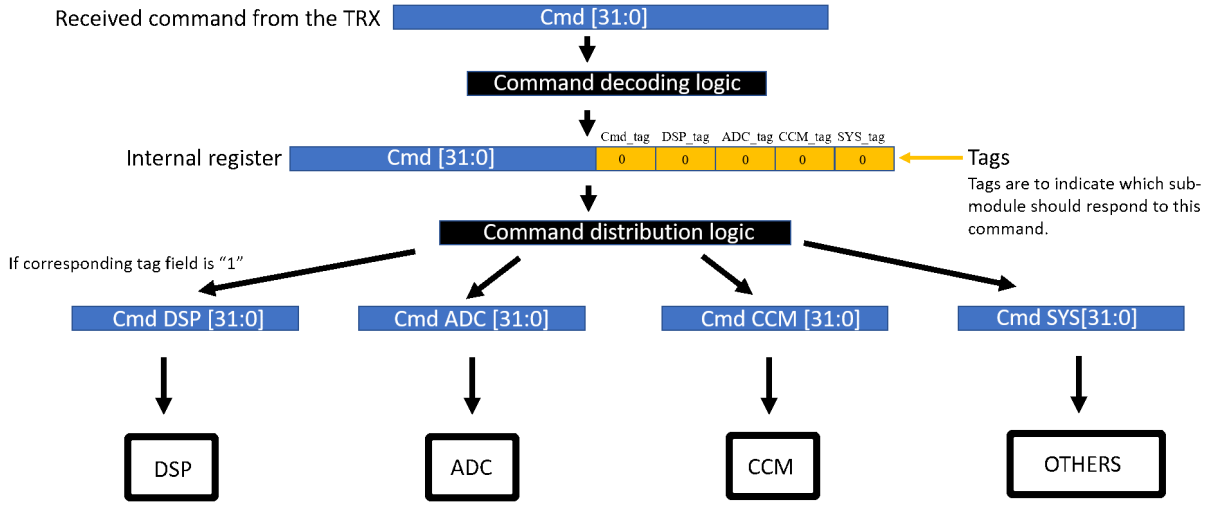

Figure 4.13: The instruction decoding logic of the ASIC

clock (frequency, duty cycle etc.). They also need to work synchronously in coordination. To ease the debugging in the future, the CCM should provide each functional units with a tunable clock. The desired output from the CCM under the normal operation condition are shown in Figure 4.14.

To program the CCM, an instruction must be issued from the upper-stream device to the main controller. The timing diagram that the main controller requires to program the clock control module is illustrated in Figure 4.15.

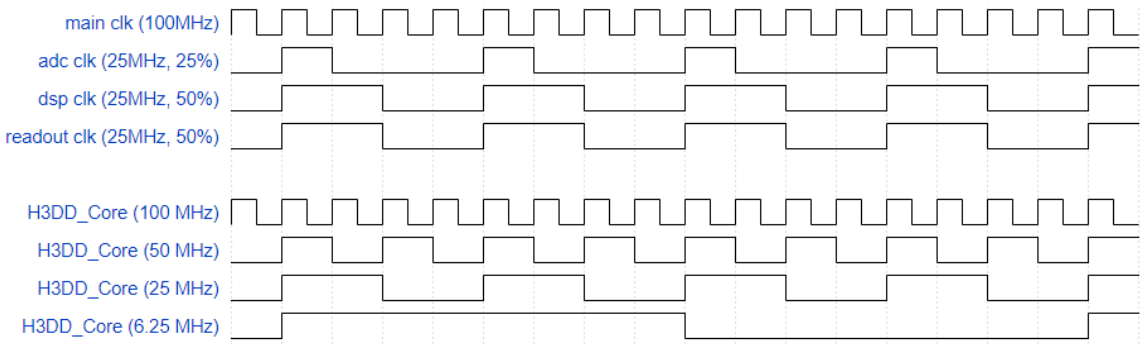

Figure 4.14: The target behavior of the CCM. The CCM is expected to provide each functional units with a tunable clock.

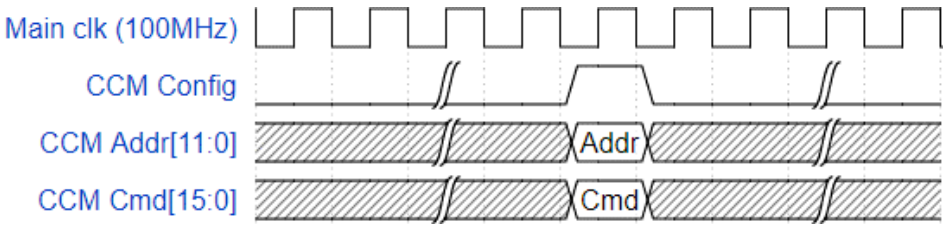

Figure 4.15: The timing diagram to program the CCM

Event Parser

The major functionality of the event parser is parsing the data and packaging the data into 32-bit packages. The timing, the amplitude, the module index, the pixel index and the timestamps are not inherently aligned into 32-bit FIFO width. Besides, the length of data stream is random since it depends on how many channels are triggered. What's more, the arrival time of the events can vary due to the DSP processing speed. Transmitting these information as raw binary stream will not fully utilize the communication bandwidth of the SPI and will require a very large FIFO in the transceiver to handle all the randomness. Considering these, an event parser module is designed to package the data into segment of 32-bit words, so that it can be transmitted by the 32-bit transceiver. The event parser are driven and controlled by the signals from the DSP to achieve its task. The block diagram of the event parser is illustrated in the following Figure 4.16. The timing diagram of the event parser is shown in Figure 4.17.

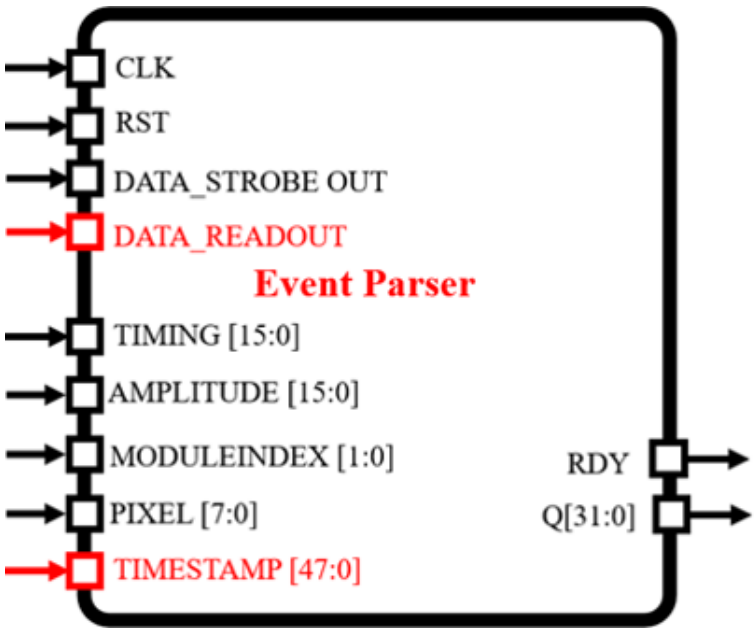

Figure 4.16: The block diagram of the data parser

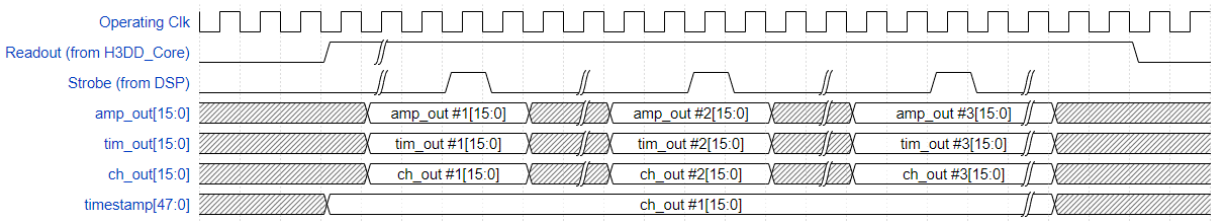

Figure 4.17: The timing diagram of the event parser.

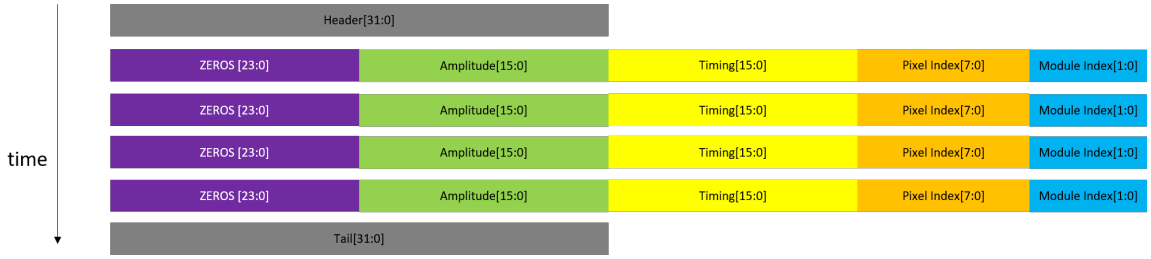

Figure 4.18: An example of the DSP readout package. The readout of a 4-pixel event is shown here.

Transceiver

The transceiver is the bridge between the internal silicon and the external devices. It communicates with the CPU to transmit the processed data and receive external

instruction to configure the operation of all on-chip modules. One design difficulty is that this module needs to work in two clock domains and the domain crossing must be handled carefully. To avoid meta-stability, FIFO buffer queue are used to cross the clock domains.

Figure 4.19 shows the block diagram of the transceiver. In the figure, the TX FIFO and the RX FIFO represent the output and input FIFO queues respectively. The former one is to buffer the incoming instructions, while the latter one is to buffer the outgoing processed data stream. The read and write of the data entry are independent, so that we can avoid meta-stability when crossing the clock domains. In a transfer test with iMX8MM CPU, the SPI receiver can operate at 100 MHz clock frequency and achieved 21 MBit/s transfer rate.

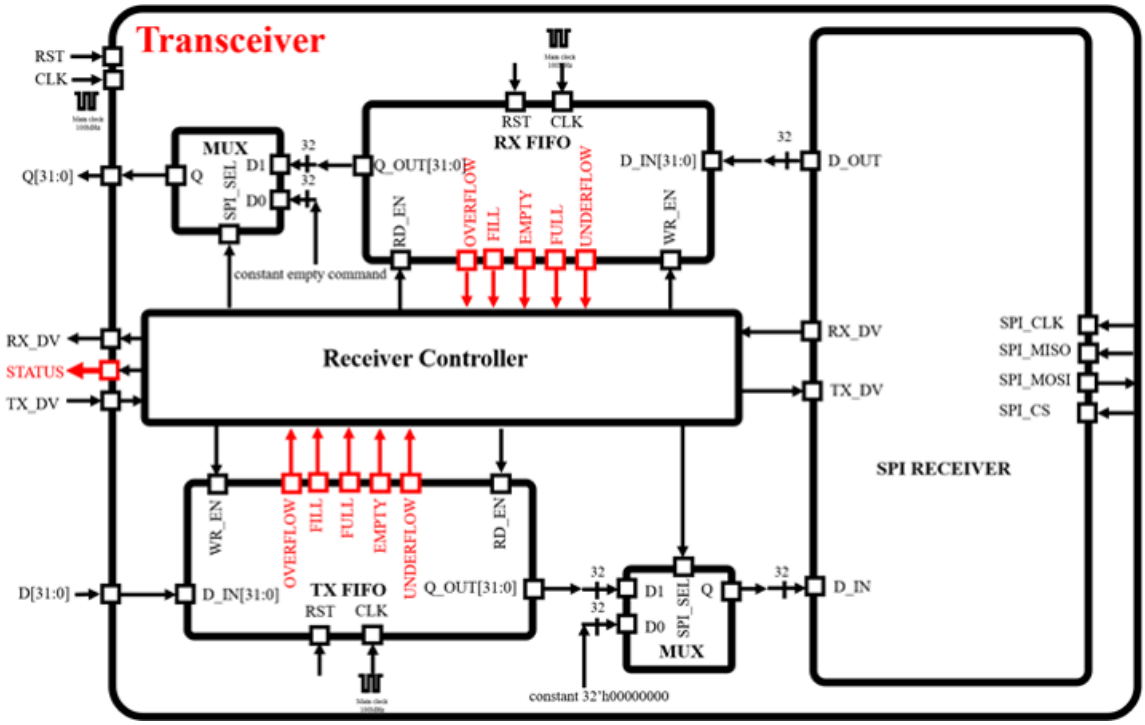

Figure 4.19: The diagram of the transceiver

The transceiver and the event parser are designed such that the data is packaged between a pre-determined header and tail. The CPU always monitors the SPI of the ASIC, but it only records the data between the pre-determined header and tail (Figure 4.18). After receiving the data package, the CPU will save the data on the disk for

future advanced processing.

These units who provide various peripheral functions are packaged together during the design. The layout of them are shown in Figure 4.20. The overall size is 1000um×500um. As shown in the figure, the IO pins that go to the DSP are placed on the bottom side, the IO pins to the H3DD_Core are on the right side, the control signals to the ADC are on the left side. The IO pins that needs to be routed out (SPI, main clock and the reset) are placed on the top so that the connection wires to the IO pads are shortened as much as possible. The eventual area of the modules in the supporting circuits are summarized in Table 4.3.

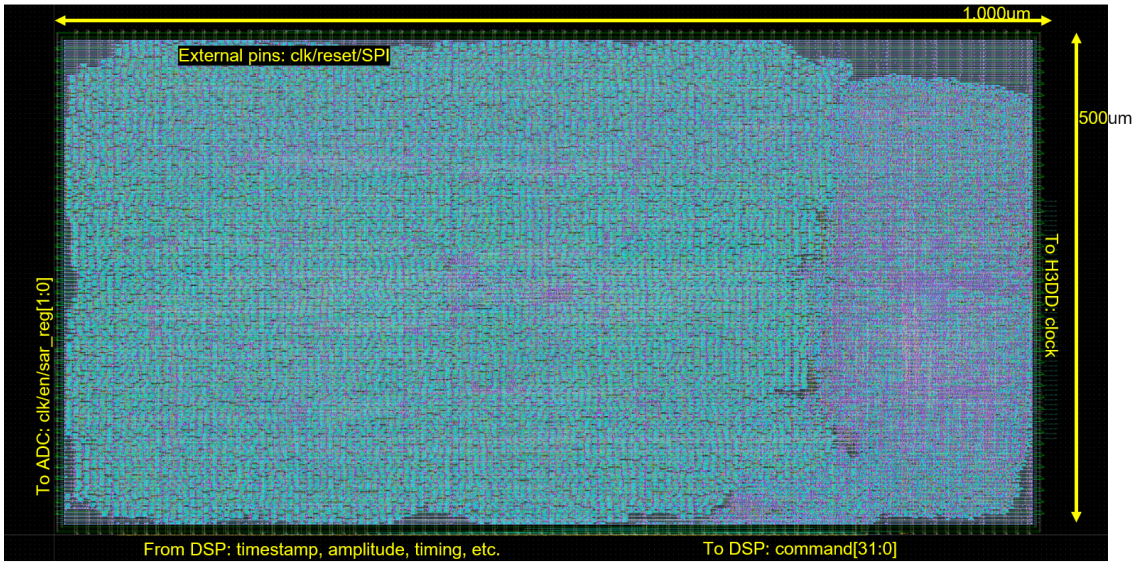

Figure 4.20: The layout of the supporting circuits

4.2.5 Design for Test

Very different from building a software, where breakpoints can be set freely in the code to diagnose the problems, once the ASIC layout is fabricated, probing the internal signals would be very difficult. The terminals of the transistors are buried under several layers of metals. The signal of interest usually travels in lower-level metals, whose width is too thin (at most several micro-meters) to probe with the tip of an oscilloscope (at least hundreds of micro-meters).

Table 4.3: The area estimation of the DSP/ADC ASIC before the actual layout.

Module Name	Area(μm^2)	Percentage
CCM	3481	0.10%
Event Parser	155643	37.20%
Main Controller	1598	0.40%
Transceiver	256879	61.60%
Monitor	230	0.10%
Total	418054	100%

To aid the silicon tests after the fabrication, a serial port and other interfaces should be implemented at the design stage. These ports or interfaces work independently from the normal operation of the chip and the internal signal can be probed to help diagnose the problem. An debug interface can be as complex as a JTAG, or can be as simple as several wires that route the important signals to the upper-level metal. But being able to reading from and writing to the registers on the chip is the minimum expectation. The compatible PCB should provide interfaces for a logic analyzer and oscilloscope. These can be implemented on the PCB in the form of multi-pinned headers. Two utilities are implemented on the DSP/ADC ASIC to help debugging the problems.

Scan Chain

Scan chain is a common technique used in design for test. The objective is to make testing easier by providing a simple way to set and observe the state of every register of an IC. The basic structure of the scan chain are three signals, scan_en, scan_in and scan_out. This set of signals together control and observe the scan mechanism. In a full scan mode, each scan_in input usually drives only one scan chain and the scan_out observe one as well.

The scan_en is a special signal that is enables the operation of the scan chain. When this signal is asserted, every flip-flop in the design is connected serially and form a long shift register.

The scan_in is the input port of the scan chain. Synchronized to the main clock of the chip, the bit being presented at scan_in port will be shifted into the chip at the

clock edges and enter the chain of registers. Since the internal registers are connected serially as a long shift register when scan_en is asserted, the value of the rest of the registers will be shifted respectively. Therefore, with enough clock cycle, the internal registers can be programmed to any arbitrary pattern.

The scan_out, on the other hand, is the output of the scan chain. When the scan_en is asserted, the value at the end of the shift register chain is output at the scan_out port. By capturing the bit-stream and reassemble them, the state of every internal register can be examined.

Signal Monitor

The scan chain is capable of probing the registers when the chip stops working. It is also crucial to examine the dynamic behavior of several control signals at the same time. To debug the chip dynamically, a signal monitor is implemented on the silicon.

The signal monitor is a big multiplexer, whose input are several groups of key signals that can indicate the operation of the circuits (Table 4.4). Based on the selection bus (Sel[2:0]), each group of signals are presented at the output bus (Out[3:0]), so that they can be captured by the logic analyzer or the oscilloscope. In this design iteration, the selection bus is 3-bit wide and the output bus is 4-bit wide.

Table 4.4: The operation mode of the signal monitor

Sel[2:0]	**Out[3]**	**Out[2]**	**Out[1]**	**Out[0]**
default	DSP Clk	Readout Clk	H3DDCore Clk	H3DDCore Clk Half
000	DSP Clk	Readout Clk	H3DDCore Clk	H3DDCore Clk Half
001	DSP Clk	Readout Clk	H3DDCore Readout	EvtParser Readout
010	ADC Clk	ADC En	ADC REG[1]	ADC REG[0]
011	Rd Trg	TRX DV	TRX Q[1]	TRX Q[0]
Continued on next page				

Table 4.4 – continued from previous page

Sel[2:0]	Out[3]	Out[2]	Out[1]	Out[0]
100	Rdout Clk	SPI Clk	TX Full	TX Empty
101	Rdout Clk	SPI Clk	RX Full	RX Empty
110	EvtParser DV	EvtParser Q[2]	EvtParser Q[1]	EvtParser Q[0]
111	AdcParser DV	AdcParser Q[2]	AdcParser Q[1]	AdcParser Q[0]

4.2.6 The Overall Layout

The overall layout of the DSP/ADC ASIC is shown in Figure 4.21. The total die area is 4320um×4320um.

The DSP is the largest and the core module on the ASIC. It is placed in the middle right (the blue color are the top-level metals). The supporting circuits and the H3DD_Core are placed north of it, in order to minimize the trace length of the interconnects. The ADC is a mixed signal circuits and requires dedicated power pads. To avoid cross talk from the digital domain to the analog domain, the ADC is placed away from the digital modules on the west side of the ASIC.

The placement of the pads have been shown in Figure 4.23. The IO pads on the north, east and south side of the chip are for the digital domain. The communication signals (two SPIs), the ASIC clock and the ASIC reset are placed on the north side. The signals related to the front-end ASIC are grouped together and placed on the east side. This is to minimize the metal trace on chip to guarantee a clean output signal to the front-end ASIC. The pads that are related to the verification utilities (the scan chain and the signal monitor) are grouped and placed on the south side. These signals are slow and less critical, a longer but wider metal trace is acceptable. The IO pads on the west side are for the ADC and are designed separately. The power pads are inserted periodically between the regular signal pads so that the IR drop across the power net is

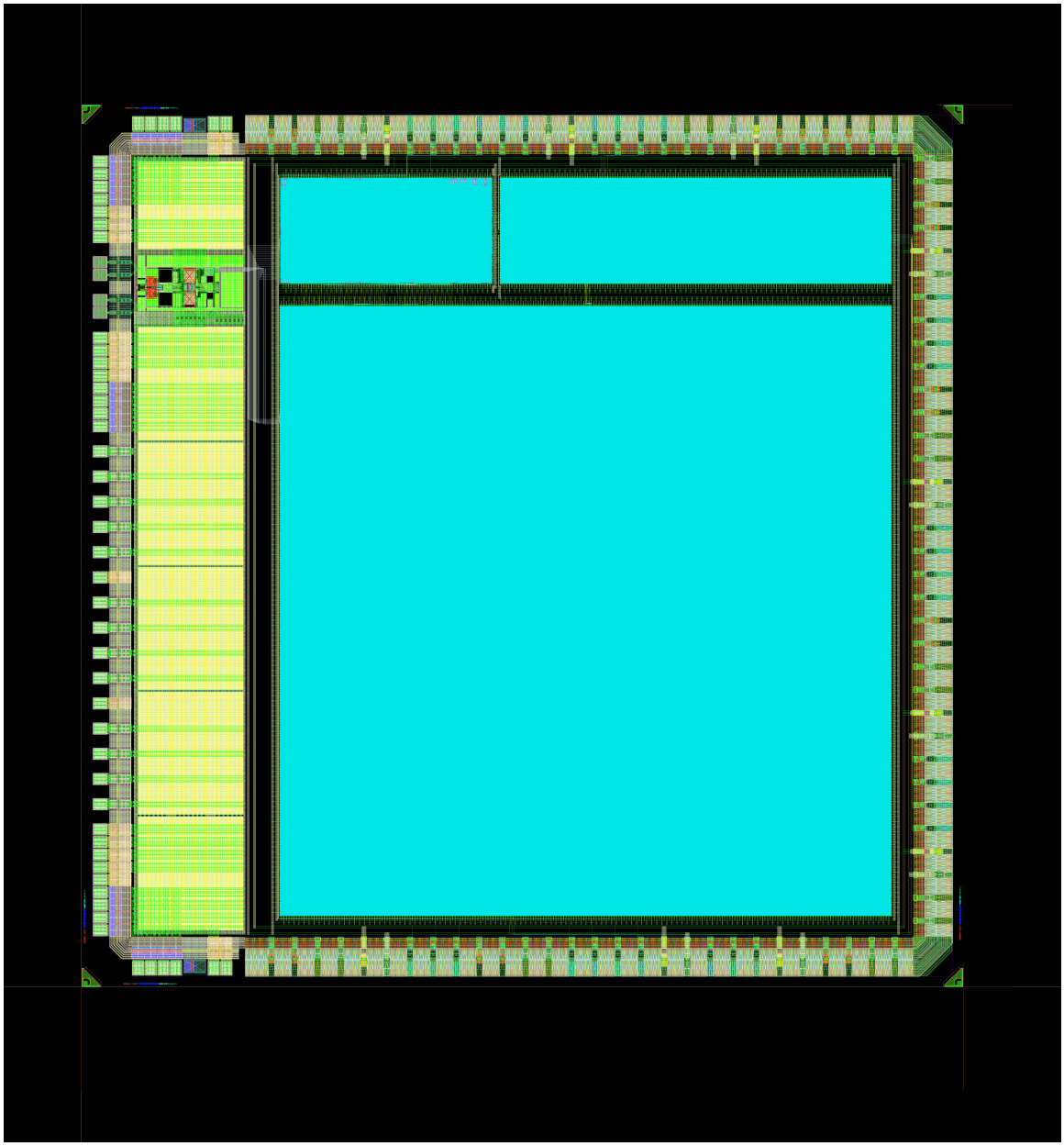

Figure 4.21: the layout of the DSP/ADC ASIC

minimized.

4.3 Design Workflow

Prior to this project, the Orion research group at UM had little experience designing the ASIC within the group. Together with this project, the chip design infrastructure is built. This includes the design software programs, automated design flow and its supporting scripts system.

4.3.1 Process Design Kit (PDK)

The PDK is a set of files used within the semiconductor industry to model a fabrication process for the design tools used to design an integrated circuit. The PDK we selected was TSMC 65nm GP from TSMC.

There're three major reasons to use this process node.

1. For purely digital ASIC, the smaller the technology node, the better the performance we would expect (speed, power consumption, area etc.). But the DSP/ADC ASIC is a mixed-signal chip, the process node is mostly dictated by the analog part. Since the analog behavior of the 65nm PDK is well understood, it's better to design the whole chip at 65nm.

2. Chips fabricated at smaller technology node can run at faster frequency. The most advanced technology node can have the chip run at several GHz. But the DSP/ADC ASIC needs to work in coordination with the front-end ASIC, who runs in the sub-GHz frequency. There's no need to use such advanced technology to fabricate the DSP/ADC ASIC.

3. The most advanced nodes is suitable for consumer electronics with large market volume. The non-recurring cost (development cost and mask production cost) is averaged into millions of chips, therefore reducing the cost of each individual chip. The DSP/ADC ASIC is a small volume research prototype and using government funding, so it's better to start with 65nm PDK. If in the future, the chip needs to be fabricated in more advanced nodes, the current design can be scaled down with the current netlist but with new standard cells.

4.3.2 Design Software

This section will describe the design and verification workflow of the DSP/ADC ASIC. Figure 4.22 shows the different stages of the chip design. Different software programs are used in each stage.

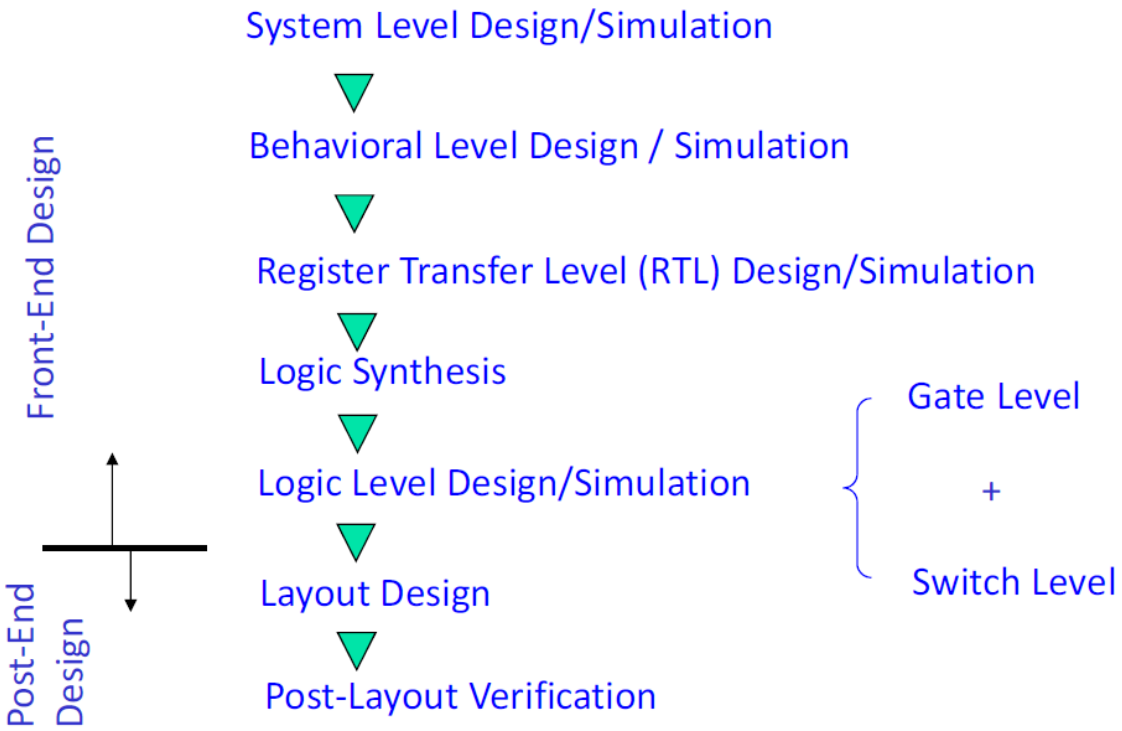

Figure 4.22: The design methodology of the DSP/ADC ASIC

System-Level Design/Behavior Simulation

All chip design starts from the top-level abstraction and design. The major task on the system level is to define the system functionality. Besides, the input and output ports of the readout module and the processing module must be defined. The function partition of the DSP/ADC ASIC has been described in the previous section.

After the specification and the IO timing of each module is defined, its flow control, the arithmetic operations and the complex delays are programmed in HDL language, SystemVerilog,

SystemVerilog is a HDL that is based on Verilog and has some extensions. It is commonly used in the semiconductor and electronic design industry and has become the de facto standard in the IC design industry. There are several advantages in using SystemVerilog for hardware design.

1. Expanded modern data structures like class, ports allows higher level of abstractions to the design. The design can be implemented in a more concise way, therefore reducing the workload to debug and adjust.

2. More advanced syntax specifically designed for easier verification. Examples like Assertions, interface to C and Randomization allows for efficient test pattern generation and mismatch capture. These features typically can not be implemented in actual hardware (not synthesizable). Instead, they assist in the creation of flexible test benches.

The module that controls the front-end ASIC, H3DD_Core, is inherited from the FPGA-based system and was written in another HDL language VHDL. This part was converted to SystemVerilog to fit in the design workflow.

Behavioral-level simulations are performed with VCS from Synopsys. To debug the problems and view the waveform, the Discovery Visual Environment(DVE) from Synopsys and another open-source tool GTKWave was used.

Logic Synthesis and Simulation

The HDL models of the functional units describe the desired behavior of the circuits. The design need to be implemented into a netlist of standardized logic gates. Examples of the standard cells include the common logic gates (AND, OR, NAND), where the timing of these modules are already carefully characterised. When the functional unit is as simple as an adder, such logic synthesis process can be done manually on paper. However, manual logic synthesis doesn't guarantee the optimal circuit performance. Besides, with the exponential increase in module complexity, the correctness of manual synthesis is prone to human errors. Tasks of this kind nowadays are taken care of by the logic synthesis tools.

The synthesis tool takes the upper-stream behavioral level HDL design as the input. User specifies the timing, speed requirement and the driving capability of each IO.

Based on these information, the original design is parsed and optimized into a netlist of standard cells that satisfies these requirements. Since the behavior of the standard cells are well-characterised, the performance of the circuits can be accurately characterized and implemented in the most optimal way.

The synthesis tool used in the DSP/ADC ASIC project is the Design Compiler from Synopsys.

Layout Design and Simulation

By the end of the front-end design, the behavioral design has been converted to a netlist of foundry-provided cells. The design then move to the back-end, where the physical layout of transistors and metal wires are built.

IC layout is the physical representation of an integrated circuit. A set of masks with planar geometric shapes correspond to the patterns of metal, oxide layers that make up the components of the integrated circuit.

The physical characteristic of the design are taken into consideration at this stage. The power delivery net and the clock distribution net are implemented. Each functional modules are routed together with metals at different layers. Eventually, the IO pads surrounding the silicon die are implemented. After the layout design, the RLC parasitic are extracted from the layout. The induced delay from them are considered in the post-APR simulation. If any new timing violation of function violation occurs, the design will iterate back to eliminated them.

The software used in the layout design is the Innovus from Cadence. The simulation tool is the same as before, VCS and DVE from Synopsys.

Post-layout Compilation and Verification

The DSP/ADC ASIC is a mixed-signal chip. The ADC (mostly analog) and the rest of the circuits (pure digital) are integrated in the layout stage manually.

The DRC is performed in this stage. The DRC aims to check the design with a set of rules that are given by the foundry, to make sure the designed mask satisfies the foundry requirement and can be manufactured. It is an essential part of the physical design flow and ensures the design meets manufacturing requirements and will not result in a chip failure. The process technology rules are provided by process engineers or the

fabrication foundry.

The LVS extracts the netlist from the layout and compare it to the original schematic netlist, in order to determine whether they are the same. The comparison check is considered clean if all the devices and nets of the schematic match the devices and the nets of the layout. Optionally, the device properties can also be compared to determine if they match within a certain tolerance. When properties are compared, all the properties must match as well to achieve a clean comparison.

The DRC and LVS tool used in this project is the Calibre from Mentor Graphics, which is built into the Virtuoso from Cadence. The spice netlist of the digital modules are extracted and parsed by v2lvs.

4.3.3 Automated Script System

Most of the front-end design software in the design flow are based on the command line terminal. Setting up the design context by typing the command line by line manually would be tedious and time-consuming. To avoid the repetitive work, an automated script system was developed to reduce the human workload when compiling, verifying, placing, routing and fixing DRC errors in the circuits. For example, with the script system, the variables (module name, clock speed, IO definition, design corners) are passed into the design system and the design context can be switched very quickly. Variables are not mandatory in every design stage, but when used, it simplifies the amount of code to be written and is easier to reuse the code.

The backbone of the automated script system is the Makefile, where it contains a list of customized command-line instructions that controls the aforementioned design softwares. Makefile rules are able to track the last time files were updated and only updates those files which are required (the ones containing changes) to keep the source file up-to-date. This feature comes handy when a revision is made in a file, the automated script system will only compiles this specific file and the one that depends on it, while the rest of the irrelevant files are untouched, Since the compilation is progressive, it can avoid compiling the whole chip from the scratch due to a small revision, the design iteration time is shortened.

The script language the design software use are Tcl. Multiple Tcl scripts are built to automated the synthesis, place and routing and the initial DRC.

The behavioral, post-synthesis and post-APR simulation results are compared automatically by the bash script. In small-scale simulation, where the complexity of the module under test is small and the number of simulation clock cycles are small, bugs can be identified by visually reviewing the output waveform. This method doesn't work when the scale of the verification problem grows. With the automated mismatch reporting functionality, any mismatch in a large simulation test benches can be captured quickly without manually inspecting the waveform cycle-by-cycle.

Aside from these, Pythons scripts are occasionally used to glue the workflow together, due to its simple syntax and extensive matured libraries.

4.3.4 The Design Repository

The design of the DSP/ADC ASIC is hosted on GitHub in the online git repository. The content of each folder are explained as follows,

The **rtl** folder holds the rtl-level design of the DSP and H3DD_Core. The **sim** folder holds the simulation testbench of the DSP and H3DD_Core. The **syn** folder holds the synthesis command and output of the DSP and H3DD_Core. The **apr** folder holds the APR command and output of the DSP and H3DD_Core. The **adc** folder holds physical design mask and the high-level abstraction of the ADC. The **cosim** folder holds mixed-signal simulation workflow and test benches. The simulation tool is still VCS as mentioned before. But the analog behavior of the circuit can be simulated based on its SPICE model. The **supporting** folder holds the supporting circuitry. In each sub-folder, the design has its own synthesis (syn) and apr folder, which isolate it from other design. The assembled supporting circuits resides in 00_TopLevel. The **integration** folder is where all digital modules are stitched together. The IO pads are placed in the physical layout and the internal signals are routed to their respective pads. The power net and the clock distribution nets of each functional units are connected so that the global power and clock nets are formed. All the digital modules are stitched together by connecting their interfaces. The post-APR netlists, in which the parasitic delays are included, are generated. Any timing violations that occur due to excessive parasitic RLC loads are fixed by adjusting the floor-planning and including extra buffer cells.

4.3.5 Packaging

The ASIC is packaged in a 7mm×7mm 60-pin QFN package. The IO map is shown in Figure 4.25. The bonding diagram together with the ASIC layout is shown in Figure 4.23.

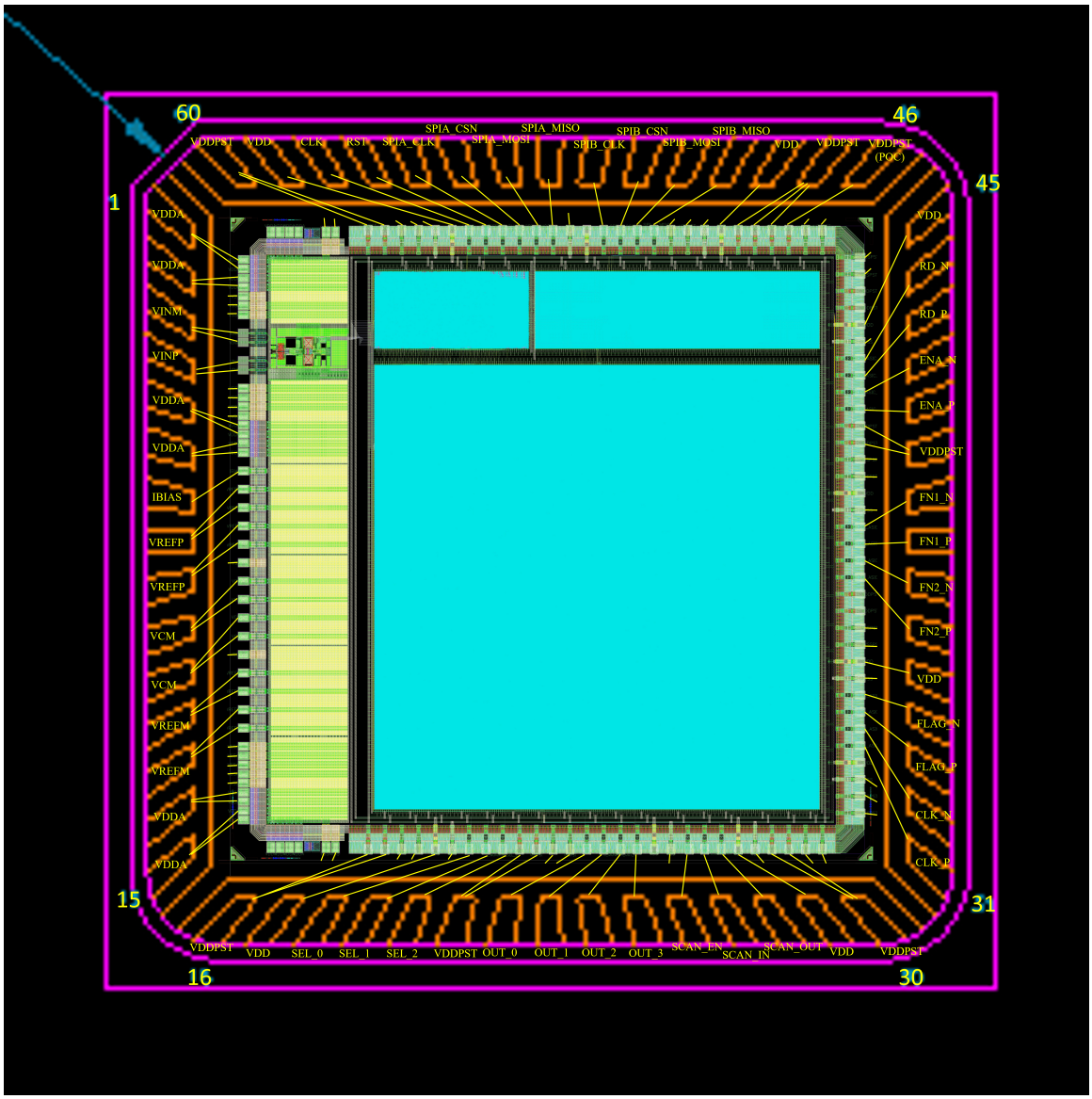

Figure 4.23: The wire bonding diagram of the DSP/ADC ASIC.

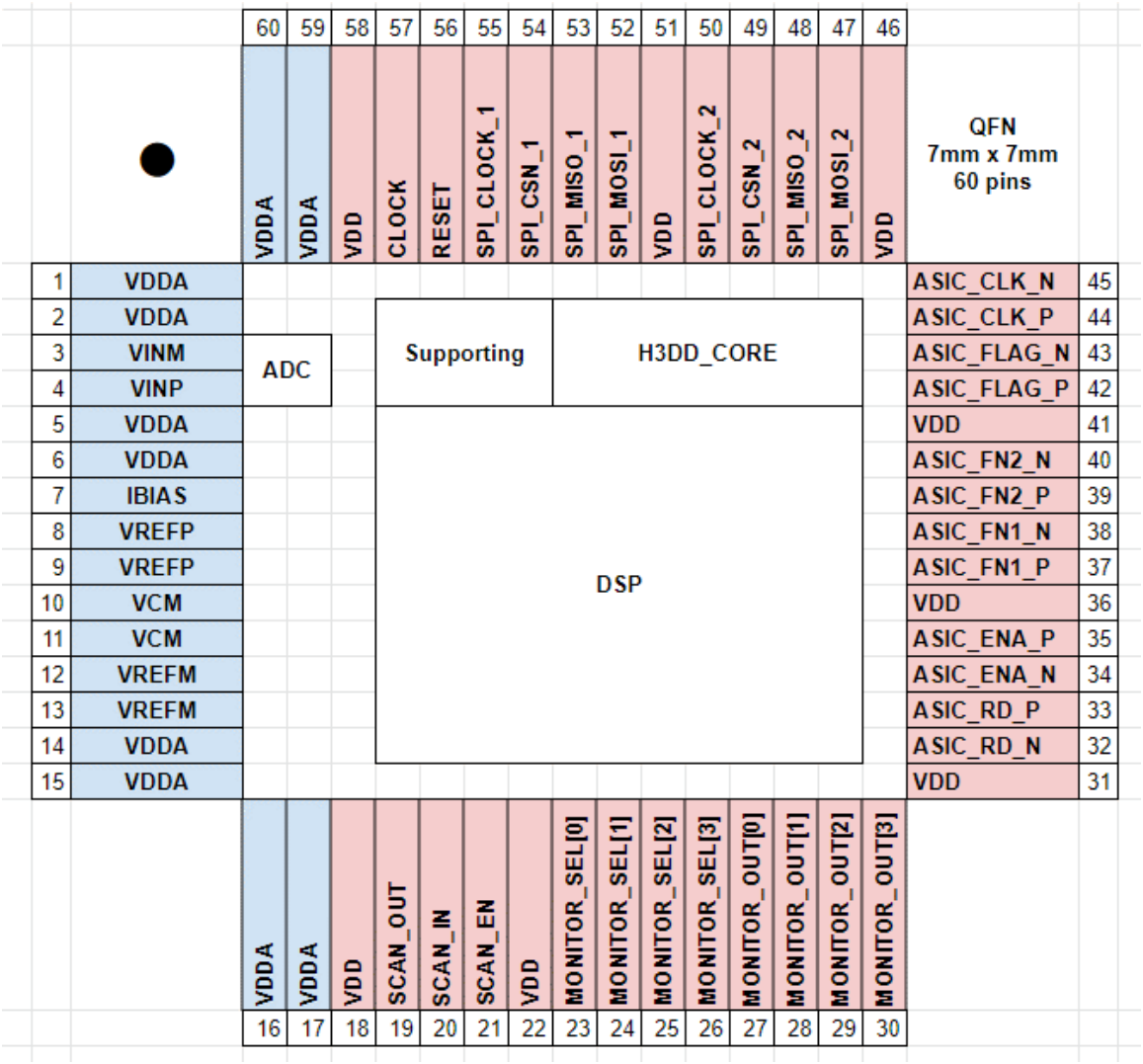

Figure 4.24: the 60-pin IO map of the DSP/ADC ASIC

4.4 The Fabricated Chip

The chip layout was designed by the Orion group and send to TSMC for fabrication. A total of 100 silicon dies are fabricated and half of them are packaged in 7mm×7mm 60-pin QFN package. The 50 packaged chips are shown Figure 4.25. The size compassion among a coin, the packaged ASIC and a bare silicon die are shown in Figure 4.25.

Upon receiving the chips, the silicon die are examined under a microscope. Figure 4.27 is the microscopic picture of the fabricated ASIC. The size of the silicon die is $4,320\mu m \times$

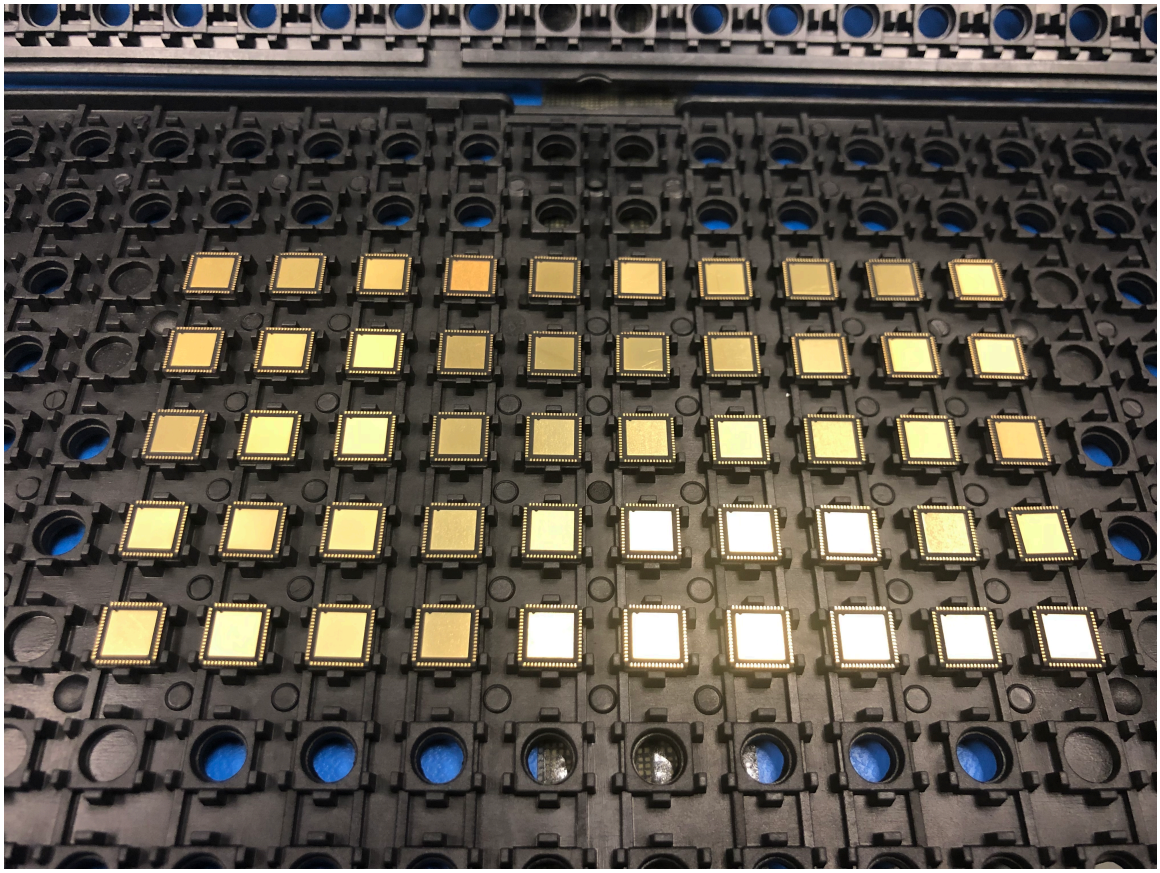

Figure 4.25: 100 ASICs have been fabricated and delivered. This figure shows 50 packaged ASICs.

$4,320\mu m$. The pattern in the highest-level metal layer is visible, especially the bonding pads near the edge in the Metal-9 layer. On the left side, the metal patterns corresponds to the ADC and its surrounding dummy cells. Several thin stripes can be seen on the right side of the figure. They corresponds to the power ring (on the Metal-7 and Metal-8 layer). The dark dots in this figure are dust fallen onto the silicon die when the chip was placed under the microscope in the lab.

4.5 Conclusion

This chapter summarized the design of the DSP/ADC ASIC. The motivation to develop a back-end processing ASIC it to overcome the limitation of the traditional system-on-

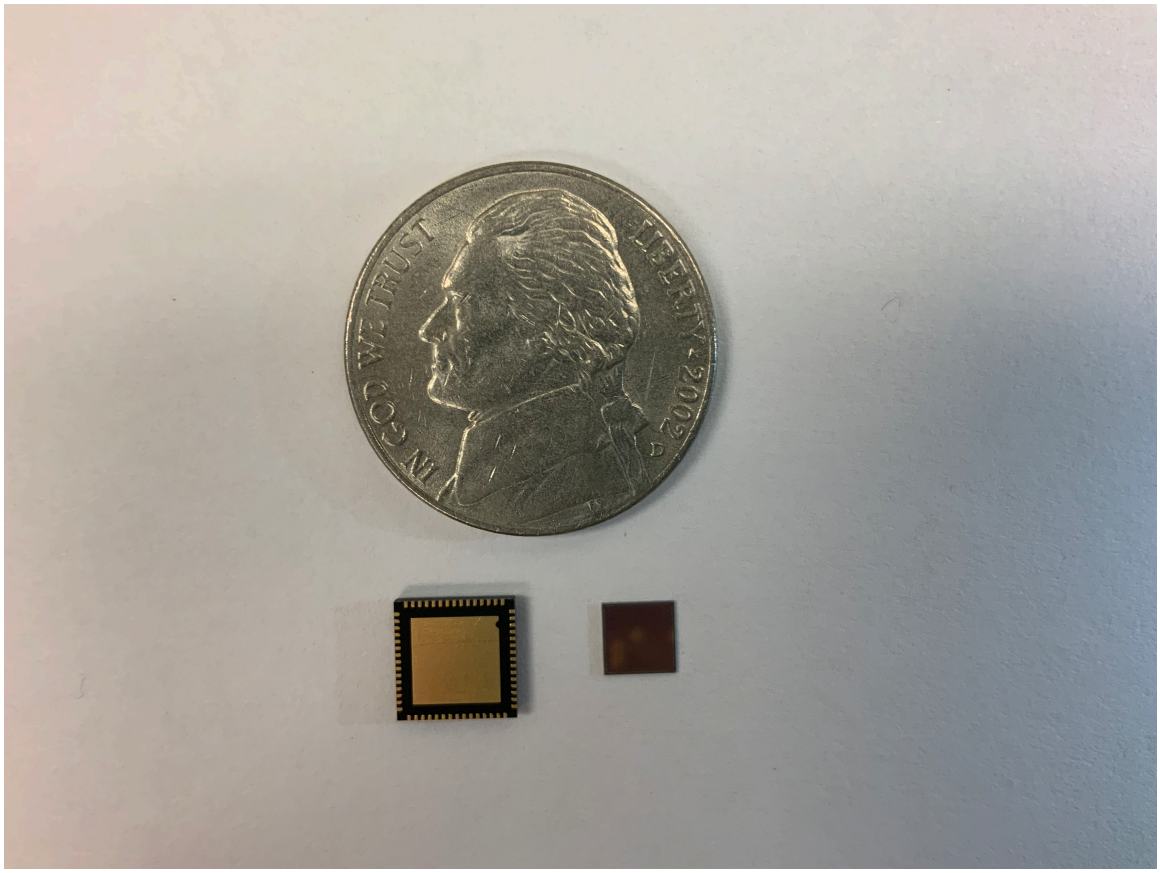

Figure 4.26: The size comparison among the packaged ASIC (bottom left), the bare die (bottom right) and a coin (top).

board design methodology. The SoC methodology provides us with the convenience to combine different hardware IPs together on a single piece of silicon, therefore provides extra room for power-performance-area (PPA) improvement. The chip was designed within our research group, with collaboration with EECS department at UM. The eventual layout was 4320um×4320um, as shown in Figure 4.27, During the chip design, the IC design workflow and a script system were built around various design software programs, which cover the full life cycle of the RTL design, synthesis, layout and verification. The chip was fabricated by TSMC in 2021 and packaged by QuickPak (now QP Technology) into a 60-pin QFN package. An automated chip design workflow involving several design software are built. The test system and the latest test results will be discussed in the next chapter.

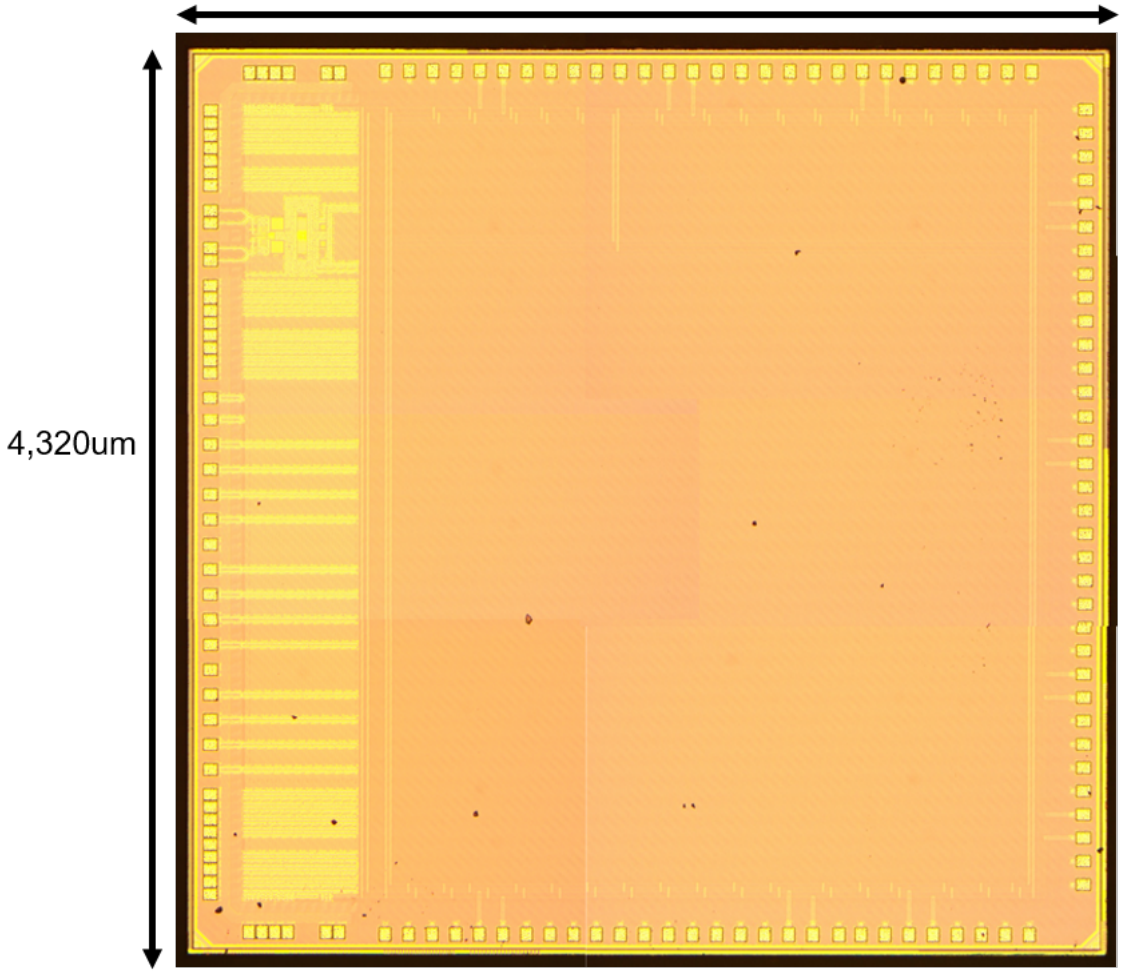

Figure 4.27: The microscopic picture of the fabricated ASIC (bare die). The dark dots in this picture are dust.

5 The Test of the DSP/ADC ASIC

5.1 Introduction

The development of an ASIC doesn't stop after the fabrication. The fabricated ASIC must undergo extensive test to verify it achieves the functions that it was designed to do. If it doesn't achieve the designed functions, only when the problems are identified and understood can the next tape-out start. Very often, the test of a chip can take as long time as designing it. In some cases, testing accounts for even more effort than the design [27]. In this chapter, we will summarize the test systems and the results of the DSP/ADC ASIC after its fabrication in middle 2021.

Before the design was submitted to TSMC for fabrication, it has undergone functionality tests or logic verification in the simulation environment with the foundry-provided data, in order to verify that the chip performs its intended function under various operating conditions. However, it's unrealistic to cover every possible input pattern in the pre-tapeout simulation.

After the fabrication, the chip needs to be mounted on a test system to be tested at full speed. Such test, called silicon test, aims to confirm that the chip can operate as it was intended to be. If any discrepancy is found, the bug should be traced down and understood. Since we have an actual chip to test, the silicon test can be much more extensive than the logic verification tests in the pre-tapeout stage. One difficulty in the silicon test is it requires creativity to locate the cause of failures because the designer has much less visibility to the fabricated chip.

This chapter will cover two major topics. The first topic is the construction of the test systems. On the hardware level, two test boards were designed, with emphasis on different part of the test. The FPGA-assisted test system requires programming the firmware for the FPGA, which is introduced too. A unified data acquisition software

with Graphical User Interface (GUI) was programmed based on the Qt framework.

The second topic is the test result of the DSP/ADC ASIC. The initial ADC readout results are shown and discussed. Several problems were identified during the test and their causes are explored. At last, the revision plan to fix the problems of the DSP/ADC ASIC are proposed.

5.2 The Test System

Testing and debugging an ASIC in the industry environment and in the university environment can be vastly different. Industry environment are market-driven, and are usually more well-funded than the university environment. More advanced test tools are available to ensure product-grade test effort. The infrastructure in the university environment is not quite as affluent as that in the industry. Not only may some test tools are unavailable, building and debugging a test board can be an extra amount of work. Due to this, the team adapt to the environment and built the test system that suits the project need and utilize the equipment in lab to the fullest.

5.2.1 Hardware

When the chip returns from the fabrication, the first tests are always run on the test board in the lab environment. Therefore, a test board that are able to provide the following attributes should be constructed.

1. Provides the stable ground and power rail for the ASIC with the ability to vary the Vdd, and measure the power dissipation.

2. Provides the real-world analog and digital inputs to the chip as the input test pattern. The board should be able to capture the outputs too.

3. Provides clock inputs as required by the chip. A stable but variable external clock generator is preferred.

4. A digital interface to the PC for fast data exchange and instruction

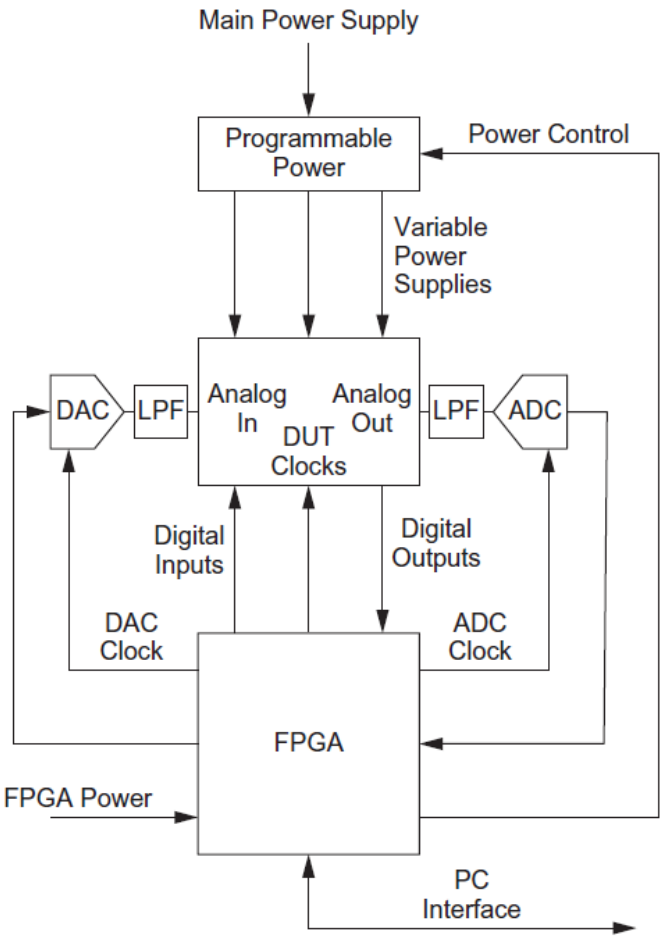

Figure 5.1: The concept of FPGA-assisted test system

FPGA-assisted Verification

In the silicon test, many approaches can be employed to achieve the aforementioned goal. Naively, one can use the most primitive method like a bread board and provide the power and the test pattern with several power supplies. Clearly this method is only applicable when testing small chips with limited number of IOs. Another method is to use a logic analyzer with pattern generation functionality. This power and test pattern delivery is more efficient. A specialized test fixture can be build around the logic analyzer. But with the development of electronics in the past decades, another FPGA-based test methodology is becoming more popular. The test board is designed around a FPGA, where it drives complex stimuli pattern to the chip under the control

of the data acquisition software. The response of the chip can be captured and stored either by the FPGA, or be captured by an external logic analyzer or oscilloscope. The architecture of a typical FPGA-assisted test system is shown in Figure.5.1.

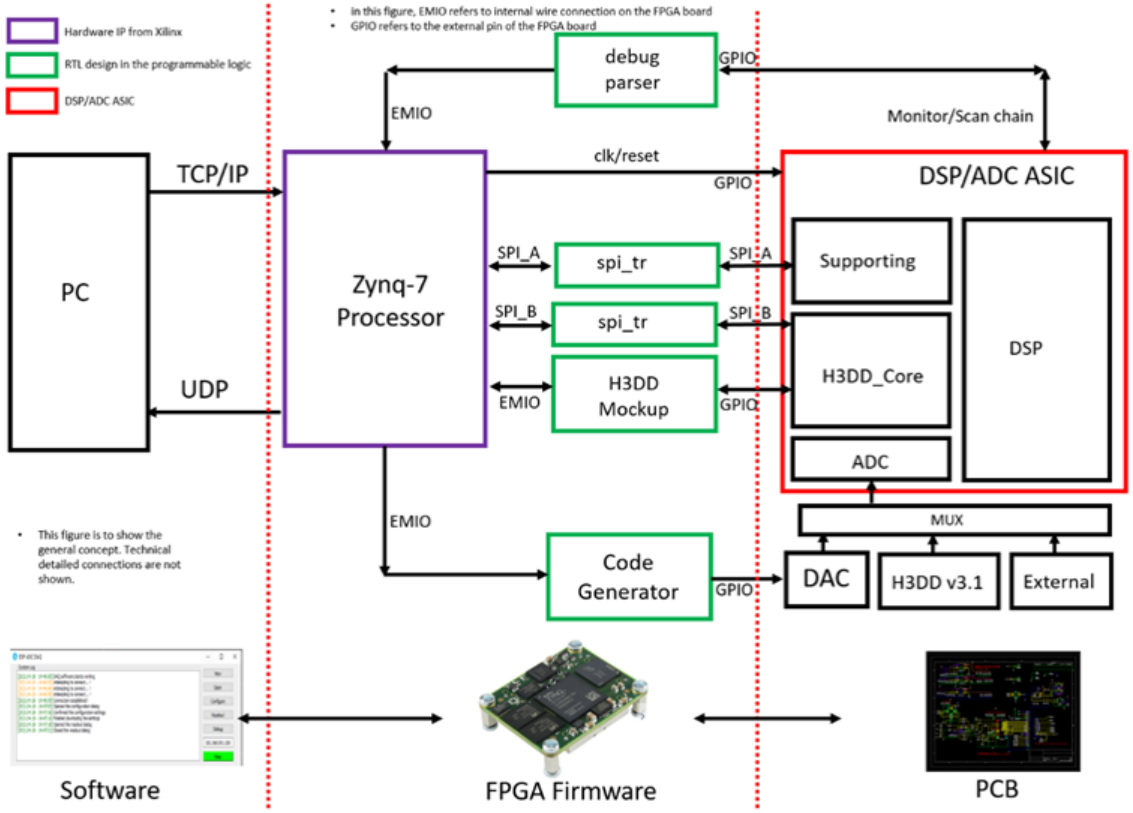

Figure 5.2: The architecture of the test system for the DSP/ADC ASIC. The left side is the software that runs on the computer. The rightmost side is the test board. In the middle, the FPGA serves as a bridge between the data acquisition software and the hardware.

The FPGA-assisted test system for the DSP/ADC ASIC was designed and shown in Figure 5.2. In Figure 5.2, the left side is the computer (or laptop), on which the data acquisition software runs. User can control the readout from the GUI on the software level. The user input will trigger subroutines written in C++. These subroutines will generate the corresponding instruction sequences and send them through the TCP/IP interface of the PC to the FPGA. The right side is the test board. The test board hosts the DSP/ADC ASIC, which is marked by the red rectangle. The input of the ADC

in the DSP/ADC ASIC can be connected to one of the three signal sources. The first signal source is from a Digital-to-Analog Converter (DAC), whose input digital codes are controlled by the FPGA. The second signal source is from the H3DD_UM ASIC, whose output is the actual waveform from the CZT. This is the most integrated test, since two ASICs are to be tested together like in a real production system. The third signal source is an external signal generator, which provides extra flexibility to the test. In the middle of Figure 5.2, The FPGA serves as a bridge between the data acquisition software program and the hardware test board.

The FPGA consists of two parts, the PS (processing system) side and the PL (programmable logic) side. The PS side is a ZYNQ-7 processor (represented by the purple rectangle). Various pre-built IPs can be conveniently implemented and programmed on the FPGA. The IPs that the FPGA firmware utilizes on the PS side are,

1. The TCP/IP Control. This module is responsible for communicating with the computer. On the one hand, the instructions from the data acquisition software is transmitted from the PC to the chip. On the other hand, the processed data can be transferred back to the PC through the UDP for event reconstruction and other data processing.

2. The SPI Module. As described in the previous chapter, the chip have two SPI ports that can be used independently to control the operation of the chip. The SPI modules on the FPGA firmware are to send the instructions to the ASIC. A programming interface is available in C so that the timing of the SPI can be adjusted so that it is compatible with the DSP/ADC ASIC.

3. Clock Wizard. This is responsible for providing the main clock to the DSP/ADC ASIC. The frequency, duty cycle and the delay of the clock can be controlled by the aforementioned TCP/IP interface.

4. General Purpose Input Output (GPIO). Multiple GPIOs are employed to control key input stimulus These include, the monitor selection bus, the scan chain, power-enable on the PCB, the hardware reset, and control signales to the on-chip H3DD_Mockup.

The PL part is a logic gate array that can be programmed. Several functions are implemented on this programmable fabric,

1. Code Generator. This digital code generator, can be programmed such that a arbitrary waveform can be sent to the test board. The digital code is converted by the DAC so that it behaves like there's a CZT and a H3DD_UM front-end ASIC.

2. H3DD_Mockup. Since testing the DSP/ADC ASIC together with the actual H3DD_UM front-end ASIC is too complex, a trimmed-down version of the front-end ASIC is implemented on the FPGA. This mock-up module only has the digital readout control logic of the front-end ASIC.

3. SPI Translator. Although the SPI IP from the Xilinx FPGA already suffice the timing requirement of the DSP/ADC ASIC, a special SPI translator module is developed so that the SPI transfer can be optionally synchronized to other external clocks. This is useful when debugging the ASIC.

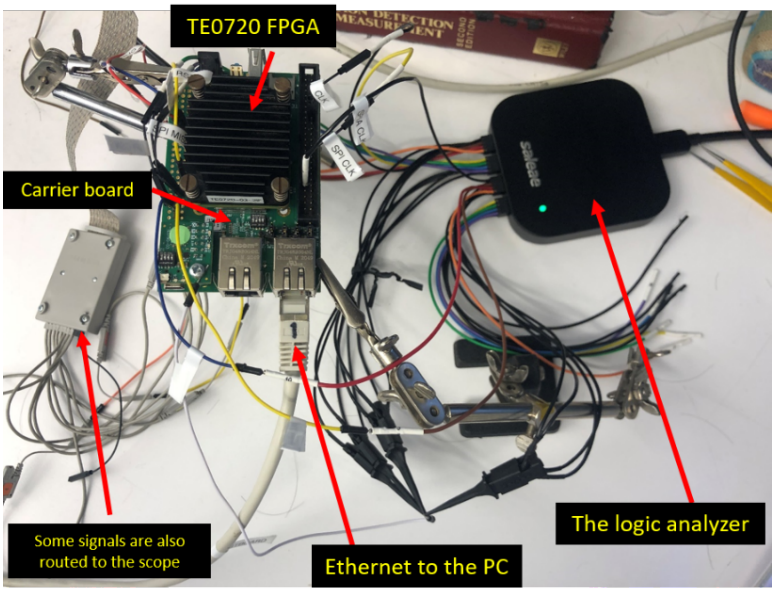

Figure 5.3: the test environment for firmware development verification. Individual components are marked in the figure. The TE0720 FPGA is mounted on its official carrier board. A digital oscilloscope and a logic analyzer are used to check the signals.

It should be noted that the goal of designing the DSP/ADC ASIC is to get rid of the FPGA in the final assembled detection system. Currently in the test phase, the FPGA still exists to help generating the test pattern and collect signals for debugging. When

the chip is fully verified, any personal computer or micro-processor with the compatible SPI can communicate with the chip directly.

Integrated Test System

The integrated test system is a compact and sophisticated test system for the DSP/ADC ASIC. This test system integrates the power delivery, input pattern delivery and the signal readout onto two compact boards with a dimension of 60mm×60mm. This integrated system is modified from the test board of the H3DD_UM front-end ASIC. Power regulators and signal readout circuitry for the DSP/ADC ASIC were incorporated on the board.

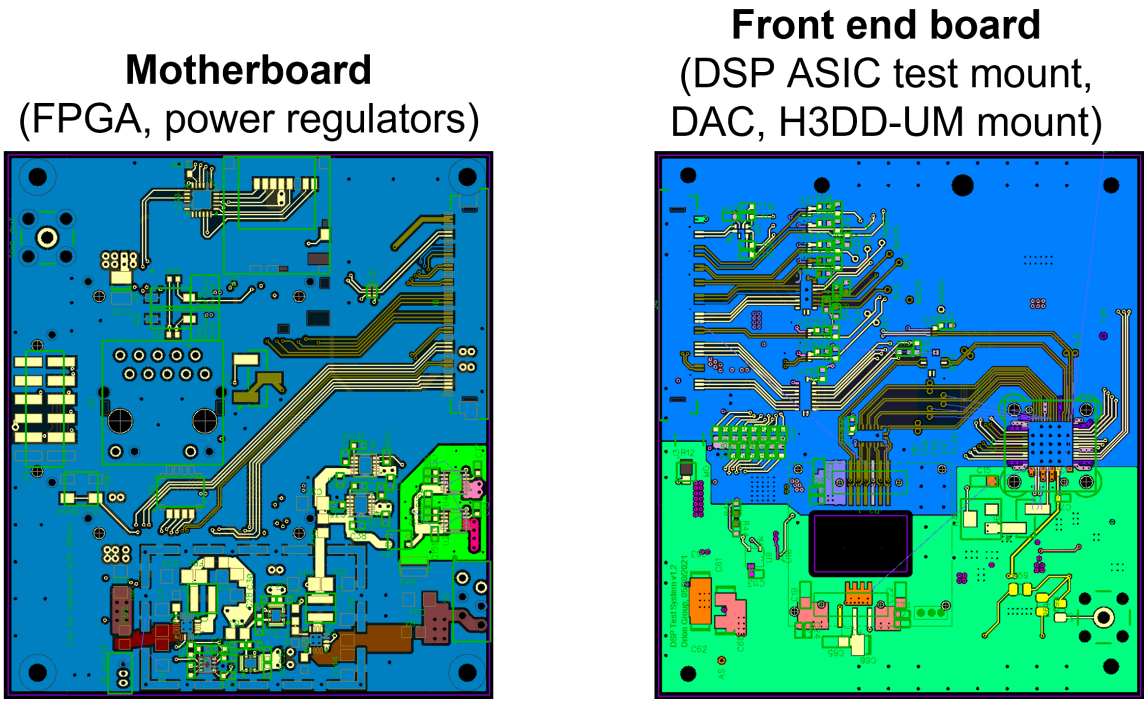

Figure 5.4: The layout of the integrated test system

The motherboard (Figure 5.4 (left)) hosts the power regulator and the FPGA. The design of the FPGA has been elaborated in the previous subsection. The power regulators provide 1.0V, 1.2V and 3.3V power rails for the front-end board. The power and signals are transmitted to the front-end board by a cable. The front-end board (Figure 5.4 (right)) hosts the DSP/ADC ASIC under test, one CZT crystal that is directly

attached to the front-end H3DD_UM ASIC. Two pairs of the integrated test systems were assembled and they are shown in Figure 5.5.

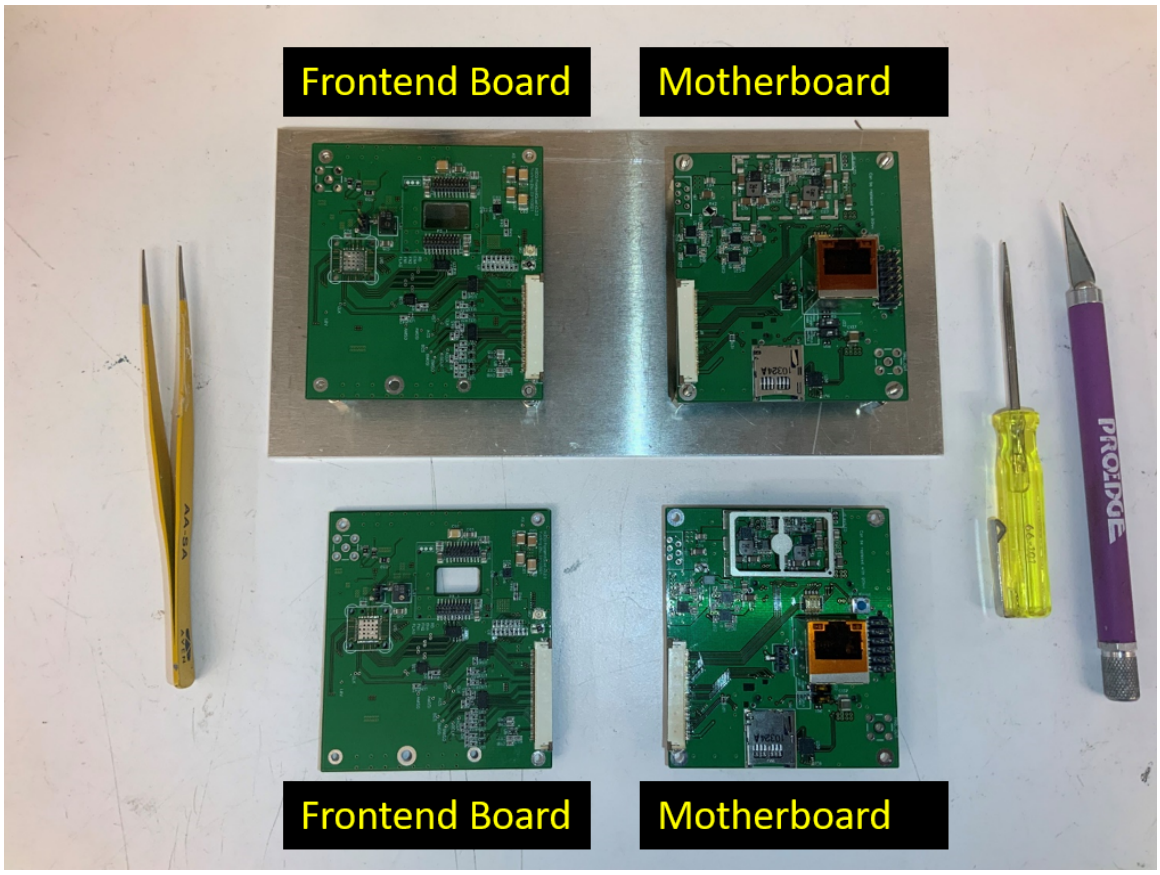

Figure 5.5: The assembled integrated test system

The most prominent advantage of the integrated system is its excellent signal integrity. The traces on the board are short and the cross talk among signals are minimized by careful routing. Due to the same reason, once the DSP/ADC ASIC is confirmed working, it can be converted to a research system with little revision.

However, several disadvantages of the integrated test system limit its use in the actual silicon test phase. First, since the bias voltages to the DSP/ADC ASIC are generated on board by ICs, it's less flexible to adjust the bias voltage and biasing current. When testing the ASIC under different conditions, it requires modifying the board on the hardware level, which is very slow and prone to error. Second, due to its small size, it's hard to assemble and debug the board itself. Besides, since the system lacks probing

headers, using logic analyzer and oscilloscope to check the signals of the scan chain and the signal monitor is very inconvenient.

Simplified Test System

Although the integrated test system is compact and sophisticated, it's hard to debug the board and probe key signals with the logic analyzer or the oscilloscope. Therefore, it was desired to built another simplified test board to compensate these inconvenience.

The simplified board aims to,

1. Provide more flexibility to the test. Adjusting the biasing voltage and the biasing current is easier.

2. Provide multi-pin headers so that it's easier to probe the ASIC signals with the oscilloscope and the logic analyzer.

3. Reduce the effort to debug the problems on the test board itself. The board is made as simple as possible.

4. Monitor the power consumption of the system more easily.

Figure 5.6 shows the layout of the simplified test system. The dimensions of the test boards are 120mm×120mm. The left figure in Figure 5.6 shows the layout of the simplified test board as viewed from the top. Near the left edges, six SMA connectors are used to provide the bias voltages to the analog domain on the ASIC. The bias current can be provided by the on-board IC, or through a two-pin connector from an external precise current source. The two SMA connectors on the top right provide the power to the digital domain of the ASIC. On the right side of the layout, two multi-pin connectors are used for easier access to the IO pads. The left figure in Figure 5.6 shows the layout of the simplified test board as viewed from the bottom. Several resistors and capacitors networks are implemented to terminate the transmission lines to avoid signal reflection.

Figure.5.7 presents the assembled system and the block diagram of the simplified test system. In the simplified test system, the FPGA is no longer mounted on the test board. Instead, it's mounted on the official carrier board from the FPGA vendor (Model TE0706, same board as shown in Figure 5.3). The FPGA receives commands from the PC through the Ethernet port and the FPGA firmware is the same as described in the

previous subsection. The DSP/ADC ASIC carrier board (the left green PCB in the figure) and the test board (the right green PCB in the figure) are connected by a ribbon cable. The test board receives external powers, therefore, the biasing voltage and the biasing current can be tuned conveniently. Monitoring the power consumption is also made easier. The probes of the oscilloscope and a 16-channel logic analyzer can be routed to the multi-pin connectors to capture the waveform.

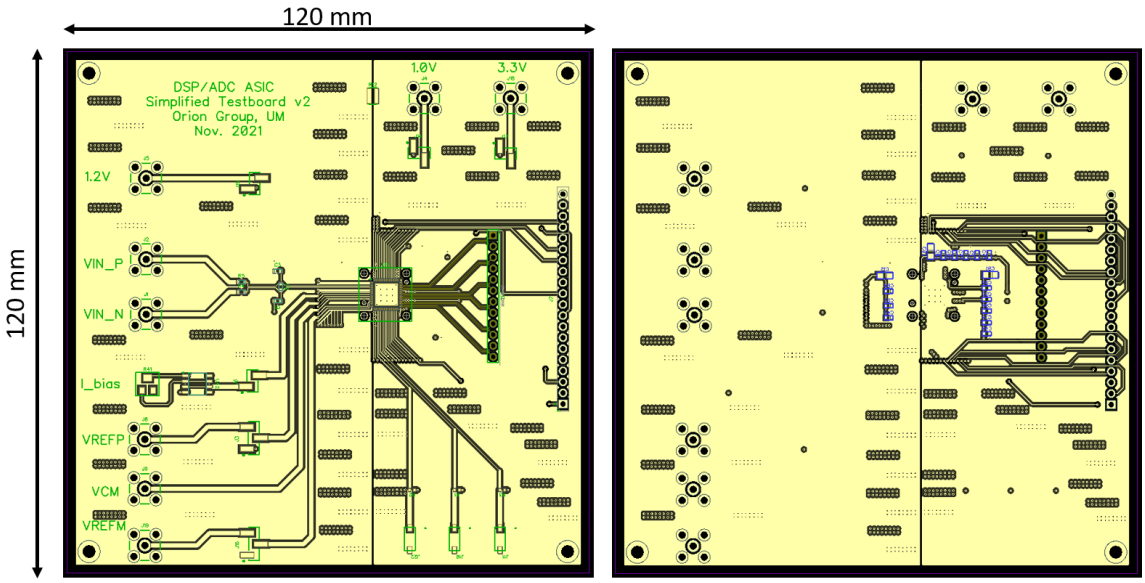

Figure 5.6: The layout of the simplified test board. (Left) the top view of the layout. (Right) the bottom view of the layout.

Equipment List

The following equipment is used in constructing the test system.

1. The FPGA used in both test systems is TE0720-03-2IF from Trenz Electronics.
2. The logic analyzer is Logic Pro 16 from Saleae.
3. Two source-meters from Keithley are used during the test (the model number is 2401). Since they are able to measure and track the current and the voltage while working as high-precision power supplies, both supplies are responsible for biasing the ADC.

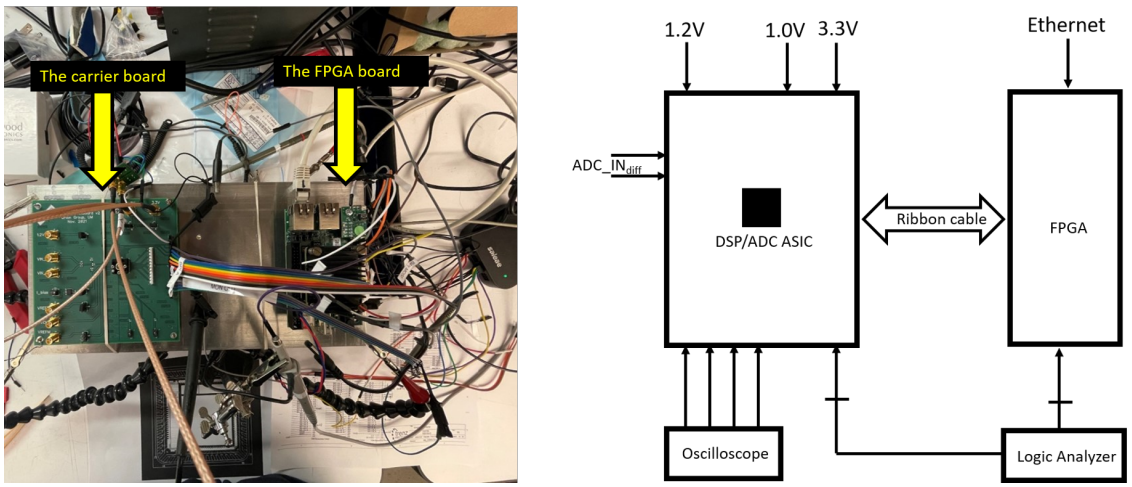

Figure 5.7: (left)The simplified test system of the DSP/ADC ASIC (right) the block diagram and connection.

4. RIGOL DP832A is used to provide the biasing voltage to the digital domain.

5. A signal generator (module numerb DG2102 from RIGOL) is used to provide the clocks signals to the ASIC. When required, it can also provide a customized input waveform to the ADC.

5.2.2 Software

To control the data readout and verification, a customized data acquisition software was programmed.

Before the test pattern is applied, the program has to set up the various attribute of the test environment. This includes,

1. set and apply the supply voltages

2. set and apply the clocks

3. set the input pattern, apply the digital stimulus and record responses.

4. collect the output signals and save onto the disk for future processing.

The data acquisition software is written in C++ under the Qt Framework. The lower-level TCP/IP interface is inherited from the legacy code in the Orion DAQ software.

The higher-level user interface are customized to suit the need for the DSP/ADC ASIC project.

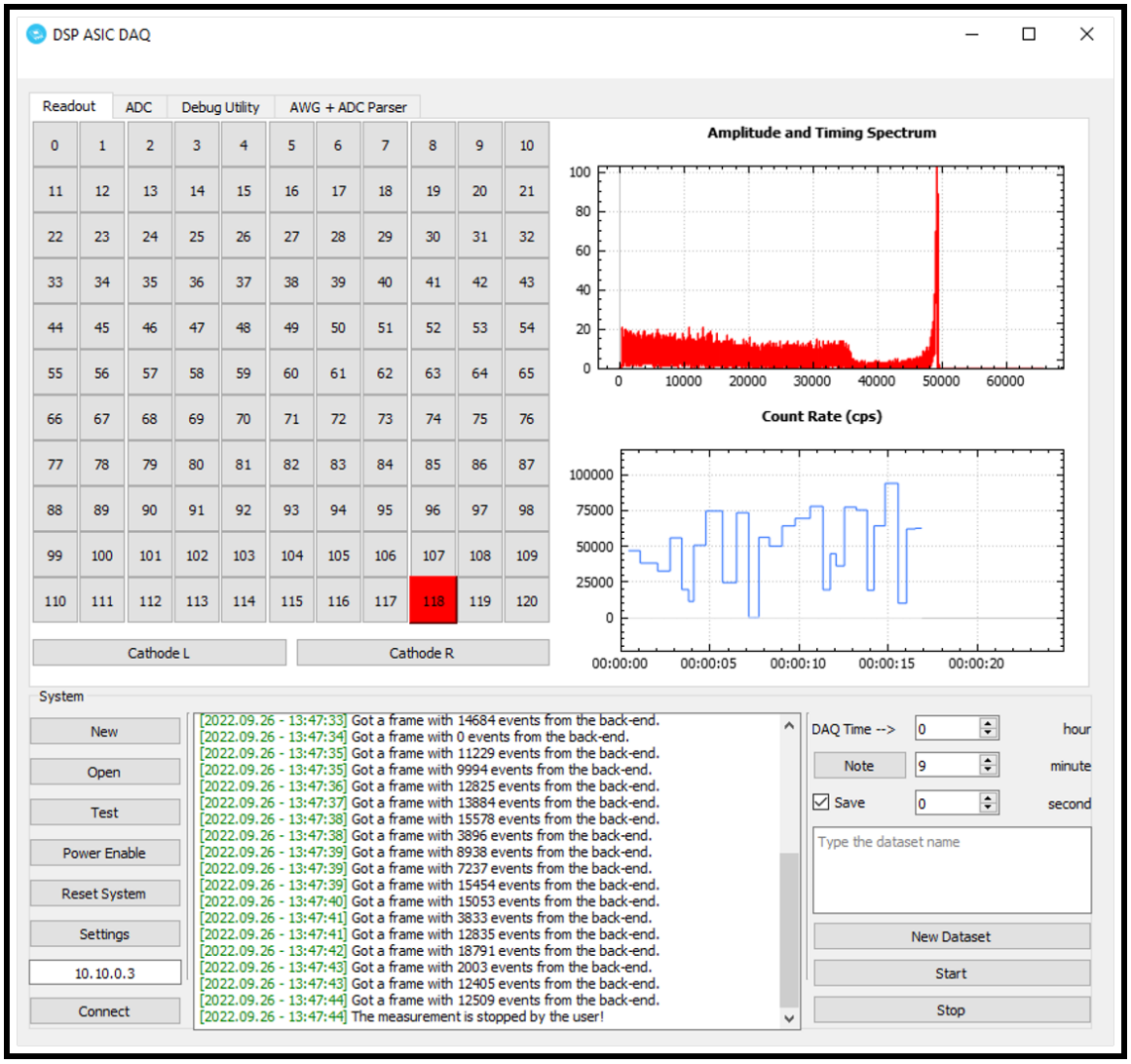

Figure 5.8: The main user interface of the data acquisition software. This figure shows a simulated readout with dataset collected on the Orion system

The readout interface of the data acquisition software are presented in Figure 5.8 (simulated readout with dataset collected on the Orion system). The top half of the software shows the triggered channel, the recorded raw spectrum and the count rate. The buttons in the left bottom corner are for TCP/IP connection, system settings, enabling the power (only applicable for the integrated test system) and global hardware

reset. The buttons in the right bottom corner are the readout control. A timer can be set up to control the readout time. The text in the middle bottom shows the log of the software, so that the user can confirm whether the instructions are executed successfully or not.

During the test of the ASIC, multiple tests need to be performed on each module. Some handy utilities are built to avoid the repetitive work and speed up the workflow. They are grouped together in different tags under the main user interface (Figure 5.9).

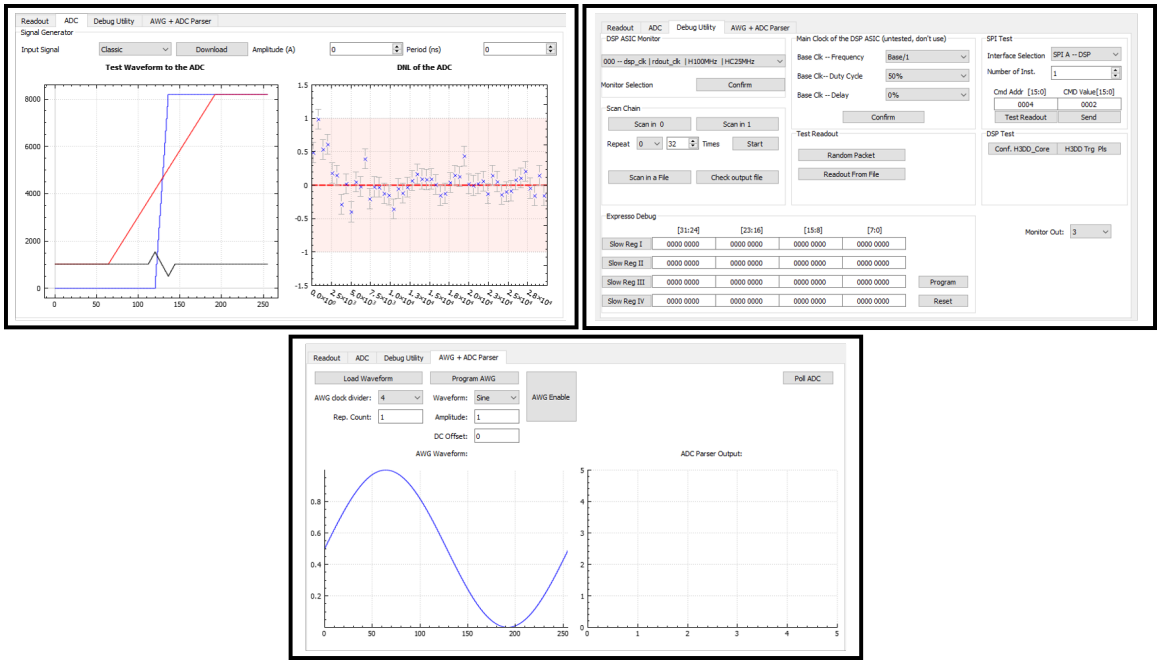

Figure 5.9: The main user interface of the other tabs. (Top left) some utilities to test and debug the ADC. The input waveform can be customized and sent to the digital coder on the FPGA. (Top right) the debug utilities controls the scan chain and the signal monitor. The SPI instruction sequences can be customized and sent to verify the operation of the SPI. (Bottom) another group of utilities that focus on testing the on-chip H3DD_Core.

Before the normal operation starts, the ASIC must be configured. The clock speed and duty cycles of each module must be programmed through the SPI. The on-chip DSP, the ADC and the H3DD_Core must be programmed too so that they can collect, digitize and process the data in the correct way. Eventually a front-end H3DD ASIC will be included in the final detection system once the DSP/ADC ASIC is ready, therefore, a separate

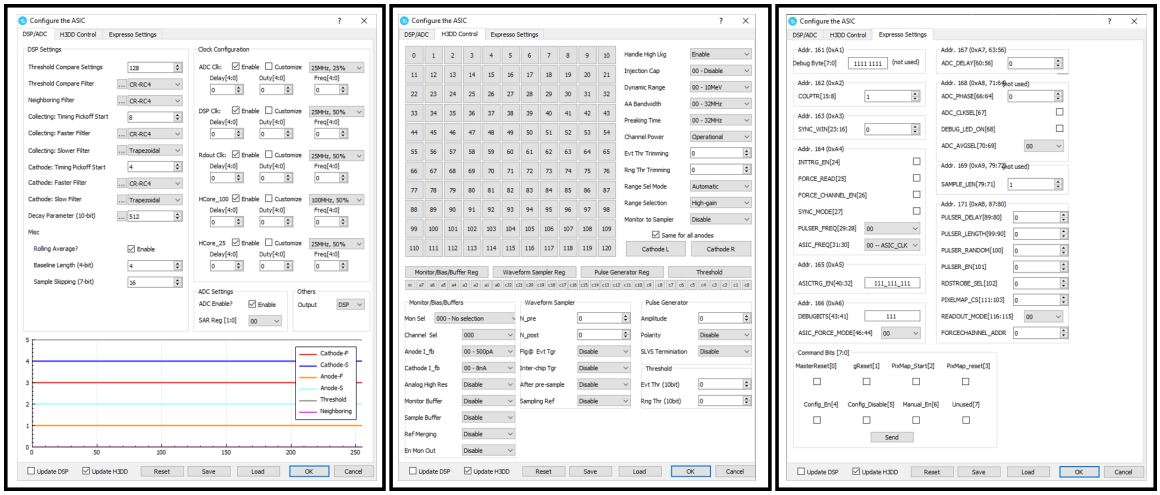

Figure 5.10: The settings page in the data acquisition software. (Left) the DSP, the ADC and the clock setup page. (Middle) the H3DD front-end ASIC setup page. (Right) the H3DD_Core setup page.

page to set up the front-end H3DD ASIC are programmed. The three configuration pages are presented in Figure 5.10.

5.3 Test Results, Known Issues and the Fix

The fabricated DSP/ADCASICs are tested extensively on the both test systems. The tests revealed several problems and they are described in the following section.

5.3.1 The Digital VDD Problem

Upon powering on the ASIC, the chip didn't respond to any instructions. The observed power consumption was close to zero watts.

First, an open-lid ASIC was used to understand the cause. The input and output signals are probed in the way as shown in Figure 5.11. The minimum probe header that's available in the lab is 0.5mm in diameter. If compared with the bonding pad (Figure 5.11 (left)), the probe pin is still slightly larger. It must be placed carefully onto the silicon die in order to capture the signal (Figure 5.11 (right)). Weirdly, all input signal, including the power and the clocks, can reach the bonding pad without any problem. The signal integrity was acceptable when observed from the oscilloscope.

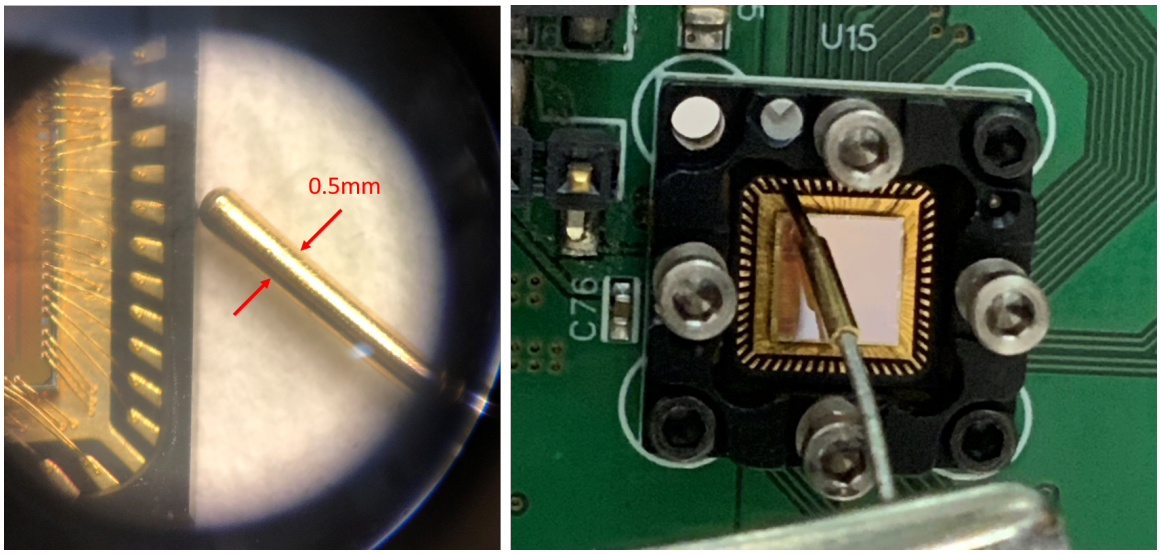

Figure 5.11: Use probes to check the input and output signal in an open lid chip

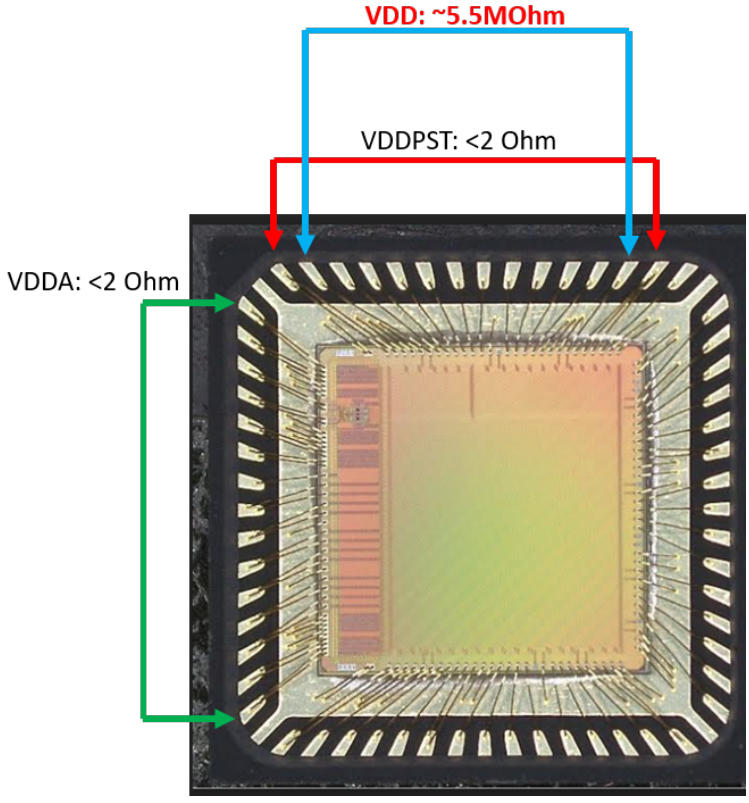

Figure 5.12: The abnormal 1.2V VDD connections

This indicates that the problem must be within the silicon die, instead of in the test hardware or the software.

Checking the connectivity between the power pads of the ASIC gives a clue about the possible cause of the problem. The resistance between the power pads in each power domain are expected to be small, since they are connected internally through the power nets built with metals. As shown in Figure 5.12, the resistance between the analog VDD (VDDA) and between the post-driver VDD (VDDPST) are both smaller than 2Ω. However, abnormally high resistance were found between the digital VDD pads. As shown in Figure 5.12, 5.5MΩ was measured between the digital VDD pads, indicating they are disconnected.

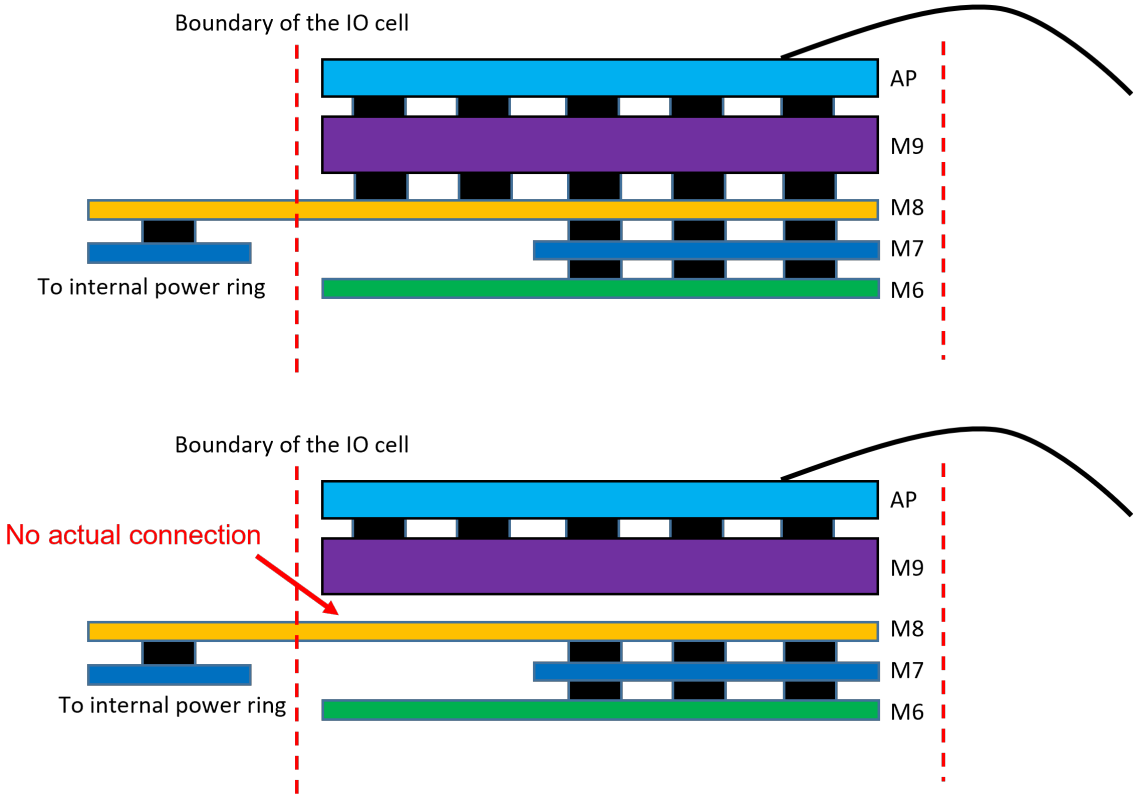

Figure 5.13: The missed connections in the via8 layer. The correct layers of masks are shown in the top figure and the actually fabricated masks are shown in the bottom figure.

After reviewing the layout, it was noticed that the revision in the Via8 mask is not correctly submitted to the foundry. Figure 5.13 shows the comparison between the

correct Via8 mask (top) and the actually fabricated Via8 mask (bottom). Since it is the old mask that is used, the Via8 layer connection of VDD and VSS nets is missing (marked by the red arrow in the figure). Therefore, although the VDD can reach the M9 layer, it never go further and reach the digital circuits. Since the VDD power supply is not available internally, the digital modules are not running,leading to zero power consumption of the ASIC. Due to the same reason, the power-on sequence is not correct and the output pads are disabled, leading to the observed high inpedence state.

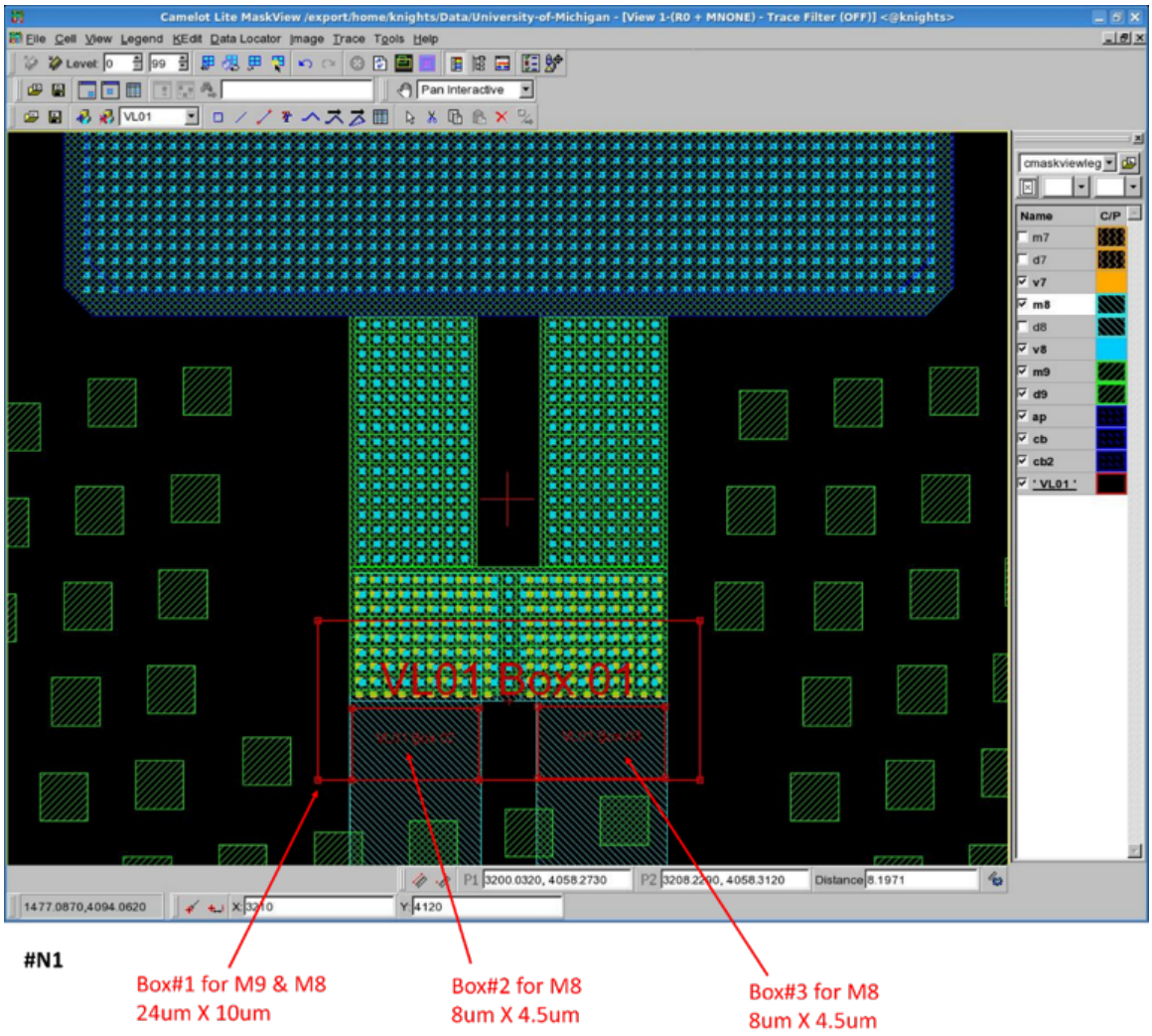

Figure 5.14: The FIB fix plan, as provided by the FIB company ABER.

Re-fabricating the ASIC was the initial plan, but the eventual solution is to use the

Focused Ion Beam (FIB) to manually build the connection between the M9 and M8 layer. The original FIB plan is to drill four 4um×4um contacts in the pad regions with the focused ion beam. However, after reviewing the layout, the FIB company, ABER, proposed a new plan, as shown in Figure 5.14. The new plan is electrically the same as the old plan but has two extra advantages. First, it's easier to work in this region since there's no need to remove the thick top aluminum layer (the AP layer) Second, the resistance and connection quality are supposed to be better, since the connection area in the Via8 layer is larger.

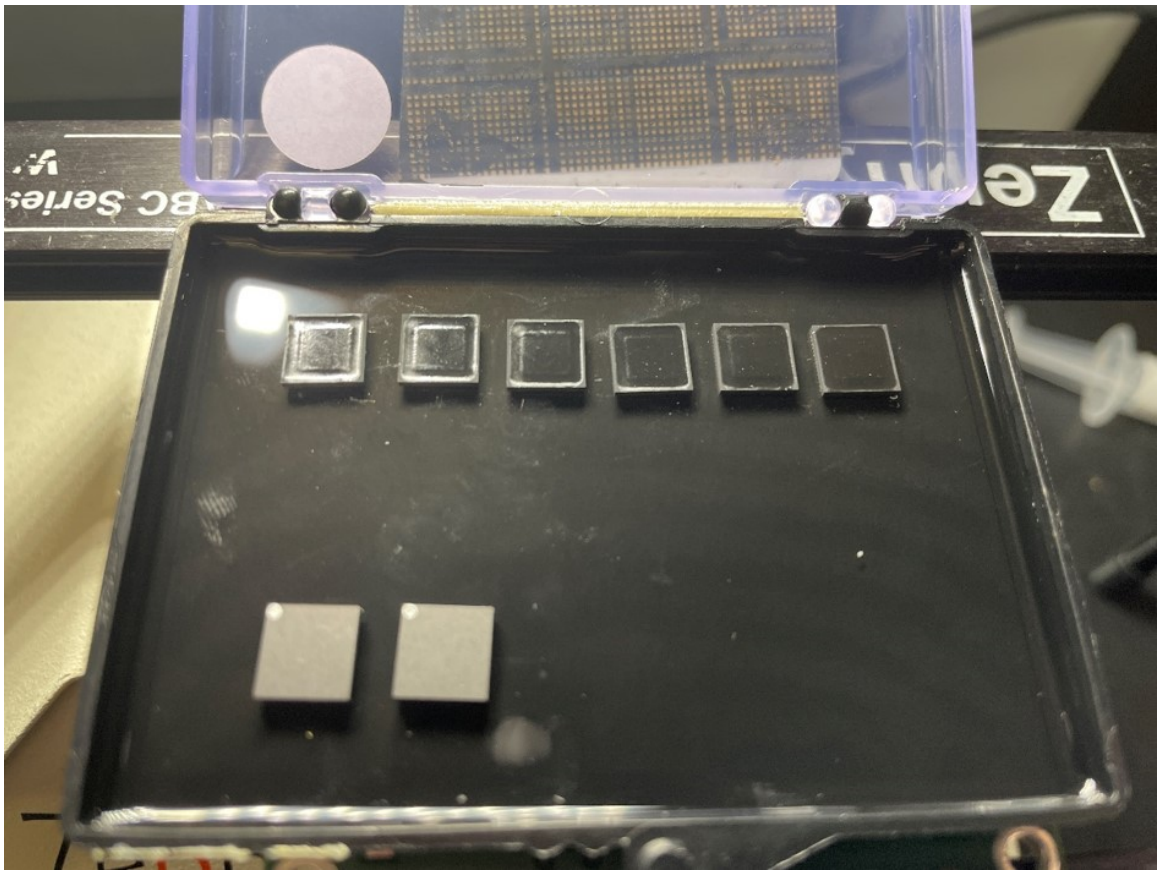

Figure 5.15: the package of the fib-fixed asic

Six DSP/ADC ASICs are FIB-fixed by ABER. The fixed and packaged ASIC are shown in Figure 5.15. Before the fix, the VDD nets were not connected inside the ASIC. Measured on the IO pads, the impedance between two VDD pads was 5.5 MΩ (Figure 5.12). Upon receiving the fixed ASIC, the impedence between the digital VDD

are measured again. After the fix, the impedance is reduced to less than 2 Ohm, confirming a good connection has been made. Now the digital power is able to reach the internal silicon.

5.3.2 The ADC Readout Test

The ADC testing requires real-time access to the digital output of the ADC. Providing parallel digital test ports by reassigning pins on the chip I/O is the most straight-forward way to facilitate this testing. However, the output of the ADC on the DSP/ADC ASIC is a 13-bit bus, the number of available I/O pins are not enough, thus, an alternative method was used. During the design phase, the ASIC are designed to have two readout modes. The first one is the normal readout mode, where the input is digitized by the ADC, then processed by the DSP. The results are then packaged by the Event Parser and transmitted by the transceiver (see Chapter IV for more details). Another readout mode is to by pass the DSP processing. The ADC output are sent directly through the parser into the transceiver.

Figure 5.16 shows an example ADC readout results, where the sin wave from an signal generator are fed to the input. As shown, the input sine wave can be successfully digitized. The spikes (zero codes) are to be discussed shortly after.

The linearity of the ADC is estimated for now by feeding a ramp function to the input. The readout results are shown in upper row in Figure 5.17. The estimated non-linearity is shown in the bottom row in Figure 5.17. It should be noted that this is not an rigorous non-linearity measurement due to the limitation in the current test system. More non-linearity studies will be performed for the next version.

Several problems were identified during the ADC readout test.

First, the ADC readout is working only when the digital power rail is raised to 1.3V, which is higher than its nominal value 1.0V. When the digital domain operates at 1.0V, the stability of the readout is quite poor. Very often, no data can be readout from the ASIC at all. Once the digital VDD is raised to 1.3V, it was noticed that the occurrence of this failure is significantly less. The reason is possibly the power drop due to the FIB fixed connection in the power net . It was also noticed that the dynamic range of the ADC is smaller than it should be when the analog power is at its nominal value 1.2V. Besides, some distortion are visually observable when the amplitude is near the floor

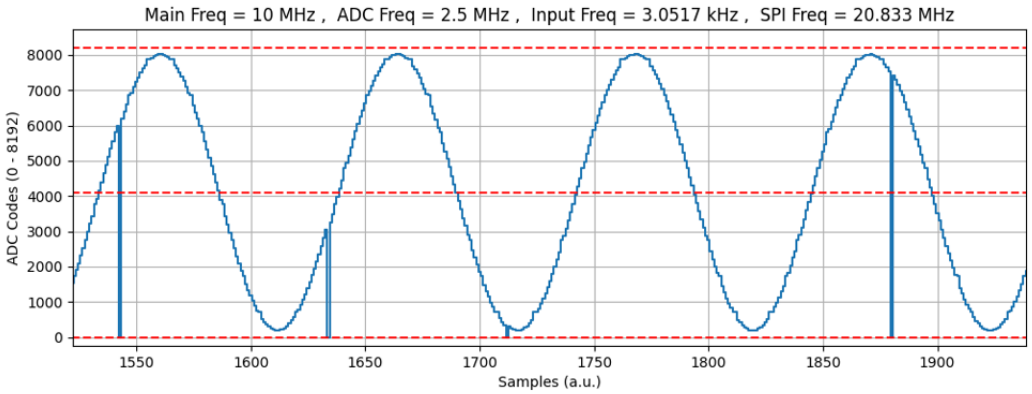

Figure 5.16: An example ADC readout results with a sine wave

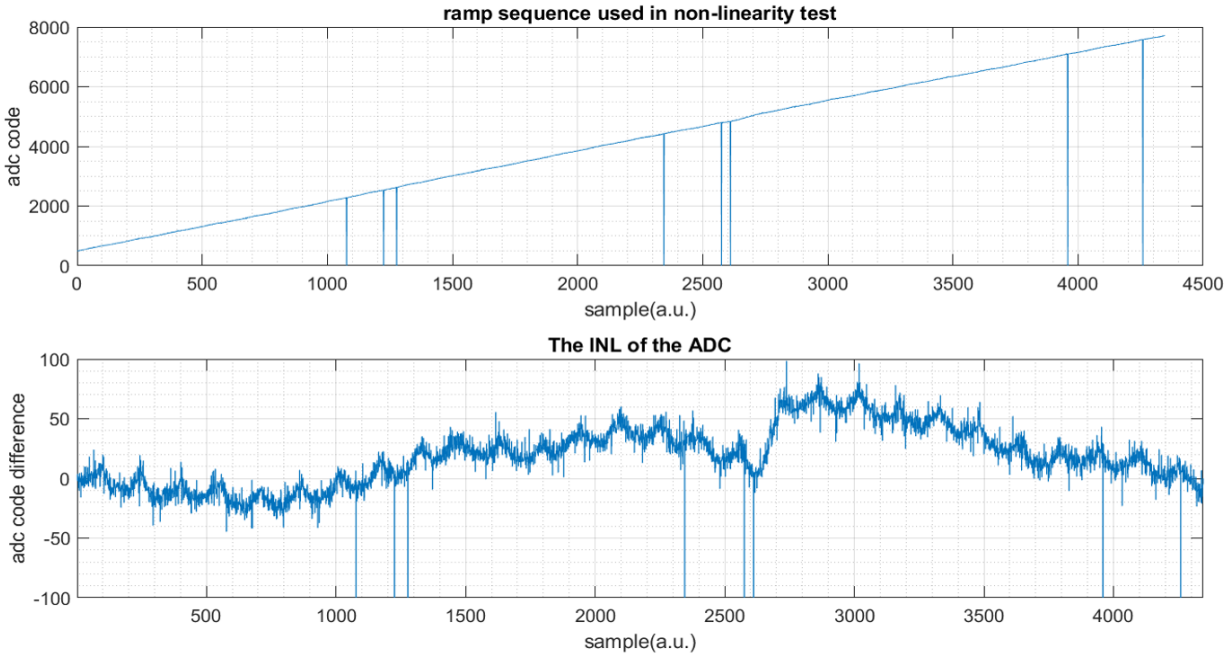

Figure 5.17: An example ADC readout results with a ramp function

and the ceiling of the dynamic range. It's suspected that the residual amplifier between the two stages of digitization (see Figure 4.3) doesn't amplify the residual voltage with an enough gain, resulting to the observed distortion. Raising the analog VDD from the nominal value to 1.3V recovers the dynamic range.

Second, as shown in the Figure 5.17 and Figure 5.16, zero codes occasionally appear in the output data stream. Simulation confirmed this is a glitch in the FIFO data pointer. In the readout test, this is less an issue since when such glitch happens, an zeros code is always observed. Such zeros code can be fairly easily identified and removed.

Third, dis-continuity is occasionally observed during the test. Such discontinuity is likely from a design flaw in the FIFO inside the transceiver. By design, the FIFO is supposed to stop the readout when its "full" flag is asserted. But due to a design flaw, the full flag is not always reliably asserted. Therefore, the newer data stream will overwrite the existing data that's already inside the FIFO buffer, leading to abrupt discontinuity in the readout waveform. When the data entries are taken out faster than they are written in, this is less a problem since the FIFO is never full. But in a real application, especially when the DSP readout is enabled, the FIFO can be full very often. This problem can't be ignored and will be fixed in the next design phase.

To summarize, the ADC readout test shows, the ADC can successfully digitize the input waveform. Although the exact linearity is not possible right now, its non-linearity is estimated from a ramp function readout test. Several issues in the readout modules (the event parser, transceiver) are identified and their underlying mechanisms are understood with the help from the simulation.

5.3.3 The DSP Readout Test

After the test on ADC confirms the performance of the ADC and the readout circuitry, the test of the DSP/ADC ASIC shirts to testing the DSP readout. Unfortunately, as of the writing of this thesis, the DSP readout still has major problems and doesn't operate reliably.

Many problems are related to the H3DD_Core. As described in the previous chapter, the H3DD_Core is a legacy module inherited from the FPGA-based detection system. This module controls the front-end ASIC for data sampling, and coordinate the ADC, the DSP and the readout modules for data processing. If the H3DD_Core is not functioning as expected, testing the rest of the ASIC is no possible. During the test, it's found the H3DD_Core has repeatability and observability problem.

The repeatability refers to the ability to produce the same outputs when given the same input pattern. When testing the H3DD_Core, it was found the response from the

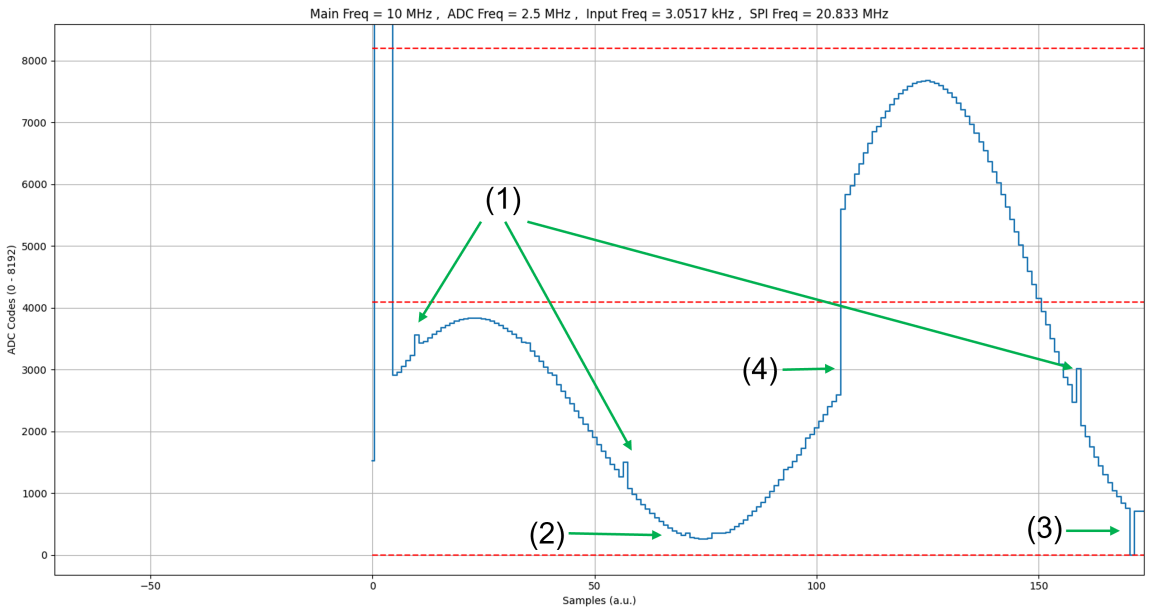

Figure 5.18: The summary of the problems observed in the ADC test readout. (1) The small spikes in the output code is likely from within the ADC. (2)Some minor distortions are from the test board. They disappear after tweaking the termination network and the power supply. (3) The zero codes are from the readout logic. Its occurrence is unpredictable, but since it always generates an empty code, this is a less concerning problem. (4) The discontinuity in the output stream is most likely from the FIFO as described before.

H3DD_Core is not repeatable from time to time.

First, the signal integrity of the test system is one factor. On the simplified test system, where most of the test on the H3DD_Core are done, the input signal integrity is not ideal. For example, the clock signals to the chip sometimes have small spikes, leading to false triggers in the digital logic. In the later stage of test, the chip is transferred to the integrated test system.

Second, the logic design style in the HDL code is problematic. The first design flaw that leads to this problem is a counter that controls the state machine. Sample the counter and changing the value of the counter happens at the same time, which is prone to racing condition, leading to unexpected behavior. The second design flaw is the unsafe clock domain crossing method. In general, digital circuits in the same clock domain tend to work in a deterministic way and are quite reliable. But the H3DD_Core actually

works in two asynchronous clock domain. Certain asynchronous sequential circuits are non-deterministic and prone to error during the operation. The internal registers are set up with the SPI and an intermediate register is sampled by another clock. If the sampling clock is significantly faster (at least four times) than the SPI clock, the instruction can reliably reach the H3DD_Core. But if this condition is not met, it was observed that the response would be chaotic and is impossible to be used for normal detection and measurement.

Third, the chip tends to overheat during the test. During the test, the chip tends to work fine initially after the power is applied. But after being tested for several hours, the chips would be noticeably less responsive to the incoming instructions. If cold spray is applied to the chip, the chip seems to become "alive" again. Such overheating tendency makes the ASIC more vulnerable for timing violations.

The observability refers to the degree to which you can observe behavior of an IC. Ideally, it's desired to be able to observe every gate output within an ASIC directly or indirectly.

Since the H3DD_Core is inherited from the previous FPGA-based system, it was believed that this module is matured enough. When integrating the H3DD_Core into the silicon, no debug ports were implemented for this module at all. The aforementioned scan chain and the signal monitor were mostly designed for the DSP and the readout modules. To avoid this problem in the next design phase, more signals in the H3DD_Core will be routed to an expanded signal monitor.

5.4 Conclusion

To summarize, this chapter covers the test of the DSP/ADC ASIC, which is probably the most important activity aside from the ASIC design. Usually it takes the longest time no matter what design methodology is used [27].

In this chapter, the test systems of the DSP/ADC ASIC project are elaborated. Two hardware systems are built. One system is the integrated test system and the other is the simplified test system. The former one is a compact and more sophisticated system, which integrates the power delivery, input pattern generation and the signal readout on two PCB boards. Although compact, it brings some inconvenience during the test. The latter one is designed as an complement to the integrated system. Probing the signals

are made easier since the logic analyzers and the oscilloscopes can be connected to the multi-pin headers. Adjusting the operating condition (speed, power etc) is alos made easier on the simplified testboard.

Both systems use an FPGA to assist the test. The architecture of the firmware on the FPGA are described. A unified data acquisition software with graphic user interface was programmed under the Qt Framework.

The test on the DSP/ADC ASIC revealed several problems in the design. First, the VDD net of the digital domain is not fabrication as designed, leading to a power delivery issue. It is fixed by creating connection in the power net with Focused Ion Beam. Second, The ADC test readout bypasses the DSP and is a good way to examine the performance of the ADC. Results shows that the ADC and its readout modules are mostly working fine. With the help of the simulation, the glitches in the readout waveform are confirmed to be from the FIFO inside the transceiver and will be addressed in the next tape-out iteration. Third, the problems identified in the H3DD_Core make the DSP readout test impossible at this stage. The repeatability issue must be improved by modifying the implementation of this module. The observability issue will be eliminated by routing more internal signals to an expanded signal monitor in the future.

6 Concluding Remarks

6.1 Summary of Accomplished Work

This thesis work discussed two topics. The first topic is the efficiency response of the radiation detection system based on the VAD_UM v2.2 digital ASIC. The intrinsic photopeak efficiency of a single 20mm×20mm×15mm CZT detector from 59 keV to 2,614 keV are measured in experiment. A detailed simulation model including gamma-ray interaction physics, readout electronic as well as signal reconstruction process is built to explain this observed efficiency deficit. The simulated efficiencies of the new model match with experimental results well after taking these factors into account. Efficiency loss due to the guard ring and the inaccurate reconstruction in the anode vicinity are revealed. Other miscellaneous efficiency loss mechanisms are discussed too. This is the most detailed study on the efficiency of CZT detectors as of the writing of this work and can be used in future quantitative detection study. The recent advancement in TEI enabled 3-D localization of radiation sources with depth refocusing. A quantification demo experiment was performed to showcase the quantitatively measurement capability of the Orion detection system.

The second topic is the development of a DSP/ADC ASIC for back-end data processing of the CZT detector. The previously discrete electric components are integrated into a single silicon chip (SoC), instead of integrating on the board-level. A smaller form factor and an order of magnitude decrease in power consumption are expected. The architecture of the DSP/ADC ASIC are elaborated. Since this is the first time the Orion research group design an ASIC in house, the design infrastructure and the work flow was constructed. Automated script systems were built to help reducing the human workload during the design. The ASIC was fabricated by TSMC in middle 2021. Two test systems for the DSP/ADC ASIC are constructed. The integrated test system has

good signal integrity, compact size and multiple functionalities, but lack flexibility and hard to test. The simplified test system provides easy access to the signal and provides more flexible test conditions, but occasionally suffers from signal integrity issues. They share the same test FPGA firmware and the data acquisition software, which are introduced in chapter IV. Several issues were identified in the first tape-out iteration and their mechanisms are discussed. At last, the revision plan to build the second version of the was proposed to fix the identified bugs.

6.2 Challenges and the Future Work

6.2.1 Quantitative Measurement with CZT

In the past several decades, various systems based on the CZT crystals have consistently demonstrated their superior performance as gamma-ray detector. The existence of a certain isotopes in the environment can be distinguished by sharp characteristic peaks. The imaging resolution of the advanced imaging technologies like Compton imaging and TEI keep getting improved and are still under active study. These techniques solved the problem "where is the source", but yet are unable to tell the end-user "how strong is the source". To measure a radiation source quantitatively, it requires understanding the details of the efficiency response of the detector and measuring the location of the source precisely in the 3D space. The study presented in Chapter III serves as an initial attempt to combine the two techniques. Under the lab condition, the sources are almost ideal point sources with isotropic emission and the effect of the surrounding environment shielding are minimum. About 10% deviation were observed in the estimation of the source activity at 1,273keV. The precision is better at lower energy range due to higher TEI image quality. However, in a real-world measurement, the scenarios would be much more complicated than that in the lab environment. Several factors limit its application in the real world.

First, the physical dimension of the source can induce errors in the estimation. For example, in a nuclear powerplant, the nuclear wastes are typically stored in barrels. Since nuclear materials are usually high-Z material, the object itself can attenuate some of the gammas that it emits, leading to significant self-attenuation. The gammas or neutrons that are from the material away from the detector can be attenuated by the

materials that are closer to the detector, degrading the accuracy of the estimation to the source activity. Besides, the efficiency response to extended source are still not well studied. This is also a good topic to explore for future researchers.

Second, the measurement environment will induce extra inaccuracy. The nuclear material might be shielded intentionally or unintentionally by the surrounding environment. Since a strong source with shielding and a weaker source without shielding essentially delivers similar counts to a detector, the nature of the shielding must be determined when quantitatively measuring radiation sources. Studies have been done by the Orion Group to estimate the intermediate shielding based of features in the spectrum [7]. The back scattering from the surroundings (wall, ground, lab table or air etc) can alter the shape of the spectrum. Any errors in the background and continuum subtraction will lead to inaccuracy in the later activity estimation.

Third, the current imaging technology is not robust enough. The limitations of the TEI are its speed and the energy range. The speed of TEI is not fast enough to handle a nuclear emergency. With TEI, the object under detection must be placed stationary in the field of view and enough counts must be collected before a high-quality image is reconstructed. The energy range that TEI can be applied is limited to low energy. If the characteristic gammas are under 300keV, the imaging efficiency of the TEI is satisfactory, since the mask can block the gammas. The pattern created on the detector array are well-defined, resulting to high reconstruction quality and accurate location estimation. However, some nuclear material can emit gammas that are above 1 MeV (^{238}U). In this case, even a thick TEI mask is not enough to guarantee a clear image. 3-D imaging with Compton Imaging requires the inspector to carry the detector in a given pose. The inspector must also move slowly and keep the position of the detector consistent relative to the IMU [22]. This requirement can be relaxed by placing another IMU on the detector and such system is under active development in the Orion research group.

6.2.2 The DSP/ADC ASIC

Previous review to the readout front-end ASIC have demonstrated the development of ASIC is a long-term work. It takes about half an decade for one generation of high-performance ASIC to mature and produce useful results. Developing ASIC in-house is

a big step the Orion group took to further improve the performance of the CZT-based detection system. In the first design iterations, several problems were identified and they are scheduled to be fixed in the next design iteration. The details of them can be found in Chapter IV of this thesis work.

Two future work are recommended for the DSP/ADC ASIC.

First, in the short-term, the problems in the first design iteration must be addressed. The power delivery issue in the digital VDD was a design procedural error. Although it's temporarily fixed with focused ion beam in this design phase, the second tape-out cycle should definitely get rid of this so that high-quality power can be delivered reliably to the internal silicon. The readout FIFO had design flaw in its overflow detection logic. When the FIFO is full, it didn't stop receiving the data from the DSP (or the ADC). The H3DD_Core module has timing issues in the first design phase. It was found that the implementation of the digital logic is problematic. Some data are sampled and changed at the same time, leading to unreliable data transfer during its operation. This module also suffered from lack of debugging ports. When problems occur, there's hardly any method to understand what happened internally, therefore, the width of the signal monitor are scheduled to be extended to help debugging in the next phase.

Second, in the longer future, the DSP/ADC ASIC can be further improved in two ways. On the one hand, more functions can be integrated on-chip. The current ASIC integrated part of the functions in the WaveformAnalyzer (digital filtering, timing and amplitude pickup). Future developers can consider integrating the calibration algorithms and the event reconstruction algorithms in the InteractionAnalyzer. It should be noted that the calibration algorithms and event reconstruction algorithms are highly computation-intensive. The designer might consider integrating a small general purpose processing unit on-chip. A good suggestion would be the open-source RISC-V processor. On the other hand, the width of the DSP/ADC ASIC can be expanded. As of now, the DSP/ADC ASIC can support one front-end H3DD_UM ASIC and therefore one crystal. In the future, it's advantageous to have it support four front-end H3DD_UM ASICs. Therefor, one DSP/ADC ASIC can handle the data from four 20mm×20mm×15mm CZT crystals. This can increase the detection efficiency and enable coincidence detection among different modules. Or alternatively, it can process the data from one large 40mm×40mm×15mm CZT crystal.

6.3 Concluding Remarks

The research and development on the CZT-based radiation detection system has been going on for almost three decades. It's fascinating that after this long time of extensive research, this field still have refreshing and exciting discoveries. Larger-volume crystals are grown, high-performance ASIC are built, and new data processing algorithms are constantly evolving. In the coming years, we expect to see more new applications in the space research field and medical field. We look forward to the future of the CZT detectors and hope future researcher in this field find the work presented in this thesis useful.

Bibliography

[1] S. Agostinelli, J. Allison, K. a. Amako, J. Apostolakis, H. Araujo, P. Arce, M. Asai, D. Axen, S. Banerjee, G. 2. Barrand, et al. "GEANT4—a simulation toolkit". In: *Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303.

[2] S. Ahmad, S. Blin, S. Callier, J. Cizel, S. Conforti, C. de La Taille, F. Dulucq, J. Fleury, G. Martin-Chassard, L. Raux, et al. "Omega sipm readout asics". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 986 (2021), p. 164628.

[3] J. Chu. "Advanced Imaging Algorithms with Position-Sensitive Gamma-Ray Detectors". PhD thesis. 2018.

[4] G. Contino, O. Catalano, G. Sottile, P. Sangiorgi, M. Capalbi, G. Osteria, V. Scotti, H. Miyamoto, C. Vigorito, M. Casolino, et al. "An ASIC front-end for fluorescence and Cherenkov light detection with SiPM for space and ground applications". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 980 (2020), p. 164510.

[5] H. H. Goldstine and A. Goldstine. "The electronic numerical integrator and computer (eniac)". In: *Mathematical Tables and Other Aids to Computation* 2.15 (1946), pp. 97–110.

[6] D. Goodman. "Passive Characterization of Unknown Spaces Using Large-Volume, Pixelated CdZnTe". PhD thesis. 2019.

[7] D. Goodman, M. Streicher, Y. Zhu, and Z. He. "Identification of intervening materials in gamma-ray spectroscopy measurements using angularly deconvolved spectra with multiple sources in the field of view". In: *IEEE Transactions on Nuclear Science* 65.3 (2018), pp. 924–931.

[8] Y. Guo, J. Wang, Y. Liang, X. Xiao, X. Hu, Q. An, J. Chapman, T. Dai, L. Zhao, Z. Zhao, et al. "Design of a Time-to-Digital Converter ASIC and a mini-DAQ system for the Phase-2 Upgrade of the ATLAS Monitored Drift Tube detector". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 988 (2021), p. 164896.

[9] Z. He. "Review of the Shockley–Ramo theorem and its application in semiconductor gamma-ray detectors". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 463.1-2 (2001), pp. 250–267.

[10] J. L. Hennessy and D. A. Patterson. *Computer architecture: a quantitative approach*. Elsevier, 2011.

[11] H. Hernandez, B. C. de Souza Sanches, D. Carvalho, M. Bregant, A. A. Pabon, R. W. da Silva, R. A. Hernandez, T. O. Weber, A. L. do Couto, A. Campos, et al. "A monolithic 32-channel front end and DSP ASIC for gaseous detectors". In: *IEEE Transactions on Instrumentation and Measurement* 69.6 (2019), pp. 2686–2697.

[12] T. Kawamura, T. Orita, S. Takeda, S. Watanabe, H. Ikeda, and T. Takahashi. "Development of a low-noise front-end ASIC for CdTe detectors". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 982 (2020), p. 164575.

[13] J. C. Kim, S. E. Anderson, W. Kaye, F. Zhang, Y. Zhu, S. J. Kaye, and Z. He. "Charge sharing in common-grid pixelated CdZnTe detectors". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 654.1 (2011), pp. 233–243.

[14] G. F. Knoll. *Radiation detection and measurement*. John Wiley & Sons, 2010.

[15] C. Leak. "Techniques for Pixelated Ambipolar-Sensitive Semiconductor Gamma-Ray Spectrometers". PhD thesis. 2021.

[16] W. Li, Z. He, G. Knoll, D. Wehe, and J. Berry. "A data acquisition and processing system for 3-D position sensitive CZT gamma-ray spectrometers". In: *1998 IEEE Nuclear Science Symposium Conference Record. 1998 IEEE Nuclear Science Symposium and Medical Imaging Conference (Cat. No. 98CH36255)*. Vol. 2. IEEE. 1998, pp. 885–890.

[17] J. Mann. "Improving Cadmium Zinc Telluride Spectrometer Performance and Capabilities." PhD thesis. 2016.

[18] S. O'Neal. "Advancements in TlBr for gamma-ray detection and imaging". PhD thesis. 2018.

[19] G. Pasquali, R. Ciaranfi, L. Bardelli, M. Bini, A. Boiano, F. Giannelli, A. Ordine, and G. Poggi. "A DSP equipped digitizer for online analysis of nuclear detector signals". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 570.1 (2007), pp. 126–132.

[20] K. H. Prasad, V. Chandratre, and M. Sukhwani. "A versatile multi-hit, multi-channel Vernier time-to-digital converter ASIC". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 990 (2021), p. 164997.

[21] A. Rivetti, M. Alexeev, R. Bugalho, F. Cossio, M. D. R. Rolo, A. Di Francesco, M. Greco, W. Cheng, M. Maggiora, S. Marcello, et al. "TIGER: A front-end ASIC for timing and energy measurements with radiation detectors". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 924 (2019), pp. 181–186.

[22] D. Shy. "Super-MeV Compton Imaging and 3D Gamma-Ray Imaging Using Pixelated CdZnTe". In: (2020).

[23] A. Ulyanov, D. Murphy, A. Fredriksen, J. Ackermann, D. Meier, N. Nelms, B. Shortt, S. McBreen, and L. Hanlon. "Using the SIPHRA ASIC with an SiPM array and scintillators for gamma spectroscopy". In: *2017 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*. IEEE. 2017, pp. 1–3.

[24] E. Vernon, G. De Geronimo, A. Bolotnikov, M. Stanacevic, J. Fried, L. O. Giraldo, G. Smith, K. Wolniewicz, K. Ackley, C. Salwen, et al. "Front-end ASIC for spectroscopic readout of virtual Frisch-grid CZT bar sensors". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 940 (2019), pp. 1–11.

[25] C. G. Wahl. "Imaging, Detection, and Identification Algorithms for Position-Sensitive Gamma-Ray Detectors." PhD thesis. 2011.

[26] J. Wang, X. Luo, X. Wei, R. Zheng, C. Feng, and Y. Hu. "Design of a low-noise, high-linearity, readout ASIC for CdZnTe-based gamma-ray spectrometers". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 988 (2021), p. 164927.

[27] N. H. Weste and D. Harris. *CMOS VLSI design: a circuits and systems perspective*. Pearson Education India, 2015.

[28] R. Wixted, S. Chidzik, D. Marlow, C. Mindas, S. Sekikawa, I. Arai, A. Suzuki, A. Watanabe, and Y. Kuno. "A 32-channel waveform sampling FASTBUS module with local DSP processing". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 386.2-3 (1997), pp. 483–486.

[29] J. Xia. "Interaction Reconstruction in Digital 3-D CdZnTe Under Various Circumstances". PhD thesis. 2019.

[30] S. Yamamoto, T. Matsuda, and I. Kanno. "Development of a DSP-based real-time position calculation circuit for a beta camera". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 441.3 (2000), pp. 359–365.

[31] H. Yang. "Applications of Digital ASIC Array System for Noise Analysis, Non-linearity Correction, Event Classification and Reconstruction." PhD thesis. 2013.

[32] B. You, L. Xiao, X. Sun, Y. Mei, and G. Huang. "A distributed readout network ASIC for high-density electrode array targeting at neutrinoless double-beta decay search in a Time Projection Chamber". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 988 (2021), p. 164871.

[33] F. Zhang. "Events reconstruction in 3-D position sensitive CdZnTe gamma-ray spectrometers". In: (2005).

[34] F. Zhang, Z. He, G. F. Knoll, D. K. Wehe, and J. E. Berry. "3-D position sensitive CdZnTe spectrometer performance using third generation VAS/TAT readout electronics". In: *IEEE Transactions on Nuclear Science* 52.5 (2005), pp. 2009–2016.

[35] F. Zhang, Z. He, D. Xu, G. F. Knoll, D. K. Wehe, and J. E. Berry. "Improved resolution for 3-D position sensitive CdZnTe spectrometers". In: *IEEE Transactions on Nuclear Science* 51.5 (2004), pp. 2427–2431.

[36] F. Zhang, C. Herman, Z. He, G. De Geronimo, E. Vernon, and J. Fried. "Characterization of the H3D ASIC Readout System and 6.0 cm^3 3-D Position Sensitive CdZnTe Detectors". In: *IEEE Transactions on Nuclear Science* 59.1 (2012), pp. 236–242.

[37] Y. Zhu, M. Streicher, J. Xia, and Z. He. *Energy resolution improvement and high-energy photon detection in CdZnTe spectrometers with second generation digital electronics and reduced preamplifier trace length*. 2018.

[38] Y. Zhu. "Digital Signal Processing Methods for Pixelated 3-D Position Sensitive Room-Temperature Semiconductor Detectors." PhD thesis. 2012.

[39] Y. Zhu and Z. He. "Performance of a 2-keV digitizer ASIC for 3-D position-sensitive pixellated semiconductor detectors". In: *2012 IEEE Nuclear Science Symposium and Medical Imaging Conference Record (NSS/MIC)*. IEEE. 2012, pp. 4109–4112.

[40] C. Zhuo. "Intrinsic Photopeak Efficiency Measurement and Simulation for Pixelated CdZnTe Detector". In: (2020).