Study of High-Performance Detectors for a Charged-Lepton-Flavor-Violating Muon Conversion Search

by

Yongyi Wu

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Physics)
in the University of Michigan
2022

Doctoral Committee:

Professor Myron Campbell, Chair
Professor Igor Jovanovic
Professor James T. Liu
Associate Professor Joshua Spitz
Professor Junjie Zhu

Yongyi Wu

wuyongyi@umich.edu

ORCID iD: 0000-0002-2543-2462

© Yongyi Wu 2022

Acknowledgments

I would like to thank my parents for their unconditional love and support over the years, especially during my years of studying overseas. I would also like to express my wholehearted gratitude to my advisor, Dr. Myron Campbell, for his continuous guidance, support, and encouragements through my graduate studies. His keen insights and immense knowledge are always an inspiration to me. I am extremely grateful to our Research Scientist, Dr. Monica Tecchio for her great patience and assistance in guiding me into the world of firmware programming. I am greatly indebted to my Mu2e Tracker team colleagues, Dr. Vadim Rusu and Dr. Richard Bonventre. From them I learned a great deal about the Mu2e Tracker. I also want to extend my appreciation to other current and former Tracker team members, especially Dr. Robert Tschirhart, Dr. Yujing Sun, Dr. Aseet Mukherjee, Dr. David N. Brown, and Dr. Mete Yucel. It was a great pleasure to be a collaborator with them. Thanks should also go to Fermilab Scientists Dr. Paul Derwent and Dr. Kiyomi Seiya for their support and help on the Beam Extinction Monitor R&D work. Many thanks to Michigan Physics Department electronics engineer Jon Ameel for his help in PCB board designs. Special thanks to Dr. Brian Beckford, our former postdoc and current DOE staff for the great career advice. I would also like to thank my committee members for their time and patience.

Table of Contents

Acknowledgments ii
List of Figures vi
List of Tables ix
List of Appendices x
List of Abbreviations xi
Abstract xiii
Chapter 1 Charged Lepton Flavor Violation 1
1.1 Lepton Flavor Violation in the Standard Model 2
1.1.1 Standard Model with Massless Neutrinos 2
1.1.2 Neutrino Oscillation and Charged Lepton Flavor Violation 4
1.2 Charged Lepton Flavor Violation Beyond the Standard Model 7
1.2.1 Supersymmetric Theories 7
1.2.2 Additional Higgs Doublets 8
1.2.3 Leptoquark Models 8
1.2.4 Additional Neutral Gauge Boson 10
1.3 Charged Lepton Flavor Violation Searches and Muon Rare Decays 11
Chapter 2 The Mu2e Experiment 18
2.1 Experiment Concept 18
2.2 Backgrounds and Mitigations 20
2.2.1 Intrinsic Backgrounds 20
2.2.2 Prompt Processes 23
2.2.3 Delayed Processes from Antiprotons 25
2.2.4 Cosmic-Ray-Induced Backgrounds 26
2.2.5 Accidental Activities 26
2.3 Experiment Setup 28
2.3.1 Accelerator Systems and Pulsed Proton Beam 28
2.3.2 Solenoid Systems 32
2.3.3 Muon Beamline 34
2.3.4 Straw Tube Tracker 37

2.3.5 Electromagnetic Calorimeter . . . 39
2.3.6 Cosmic Ray Veto . . . 40
2.4 Background and Sensitivity . . . 42
2.5 Lepton Number Violation Search at Mu2e . . . 45
Chapter 3 Mu2e Tracker Panels . . . 48
3.1 Drift Tubes . . . 48
3.1.1 Gas Ionization . . . 49
3.1.2 Drift of Ions and Electrons . . . 50
3.1.3 Avalanche Multiplication . . . 52
3.1.4 Signal Creation and Propagation . . . 56
3.1.5 Example Detector Responses . . . 58
3.2 Tracker Panel Details . . . 64
3.3 Tracker Front-End Electronics . . . 67
3.3.1 Pre-Amplifiers . . . 68
3.3.2 The Digitizer Readout and Assembler Controller . . . 69
3.4 Requirements on Tracker Performance . . . 71
Chapter 4 Firmware-Based Data Acquisition and Tracker Operation . . . 74
4.1 Experiment and Tracker Data Acquisition . . . 74
4.2 Alternate Panel Control and Operation System . . . 76
4.2.1 Tracker Firmware . . . 77
4.2.2 Micro-Controller Command Set . . . 80
4.2.3 Software Control Console . . . 82
4.3 Realization of Panel Operations . . . 83
Chapter 5 Tracker Vertical Slice Test . . . 88
5.1 Vertical Slice Test and Setup . . . 88
5.2 Source Scan Studies with Iron-55 . . . 89
5.3 Scan Analyses and Results . . . 92
5.3.1 Reconstruction of Source Locations . . . 92
5.3.2 Time Division and Effective Propagation Speed . . . 95
5.3.3 Straw Longitudinal Resolution . . . 100
5.4 Longitudinal Resolution Asymmetry between Straw Ends . . . 103
5.4.1 Unbalanced Channel Thresholds and Longitudinal Resolution . . . 104
5.4.2 Examination of Channel Thresholds . . . 106
Chapter 6 Research and Design: Beam Extinction Monitor . . . 114
6.1 Proton Improvement Plan-II and Mu2e-II . . . 114
6.2 Beam Extinction Measurement of Negative Hydrogen Ion Beam . . . 116

6.3 Selection of Electron Detector 119
6.3.1 Secondary Electron Multipliers 119
6.3.2 Solid State Devices 122
6.4 Proof-of-Principle Measurement and Discussions 125
6.5 Development Plans 127
Chapter 7 Conclusion 130
Appendices 132

List of Figures

Figure 1.1 Elementary particles in the Standard Model of particle physics 1
Figure 1.2 Leading order Feynman diagrams of some CLFV processes in the SM 6
Figure 1.3 SUSY contribution to $\mu \to e\gamma$ 7
Figure 1.4 Constraints on flavor-violating Yukawa couplings to the $125\ \mathrm{GeV}$ Higgs 9
Figure 1.5 Leptoquark interactions contribute to some CLFV processes at the tree level . . 10
Figure 1.6 History and outlook of branching ratio limits in muon rare decay modes 12
Figure 1.7 Toy CLFV Lagrangian parameter space coverage by past and upcoming muon rare decay experiments 14
Figure 2.1 An illustration of the Mu2e experiment concept 18
Figure 2.2 Michel spectrum of muon decay-in-orbit 21
Figure 2.3 A schematic view of the Mu2e experiment 28
Figure 2.4 Pulsed proton beam delivery 29
Figure 2.5 The pulsed proton beam timeline with the live search window 30
Figure 2.6 The beam Extinction Monitor 31
Figure 2.7 Solenoid module and superconducting cable 32
Figure 2.8 A closer look at the Mu2e muon beamline 34
Figure 2.9 Dependence of muonic atom properties on the atomic number 36
Figure 2.10 The Mu2e straw tube Tracker 38
Figure 2.11 Annular design of Mu2e detectors 39
Figure 2.12 The Mu2e electromagnetic Calorimeter 40
Figure 2.13 The Mu2e Cosmic Ray Veto 41
Figure 2.14 Simulated conversion electron and backgrounds 43
Figure 2.15 Discovery reach of the Mu2e experiment 43
Figure 2.16 Confidence belt using Feldman-Cousins ordering at 90% CL for background $b = 0.41$ 44
Figure 3.1 A schematic diagram of a drift tube 48
Figure 3.2 Electron drift velocity versus electric field in Ar-$\mathrm{CO_2}$ mixtures of different proportions 51

Figure 3.3 A display of electron density near an anode wire in an avalanche originated from a single electron . . . 53
Figure 3.4 First Townsend ionization coefficient of some noble gases as a function of reduced electric field . . . 53
Figure 3.5 Dependence of particle gain on applied voltage in gaseous ionization detectors 55
Figure 3.6 Photon absorption length versus photon energy in argon under normal temperature and pressure . . . 59
Figure 3.7 Example $^{55}\mathrm{Fe}$ energy spectrum from a straw tube measurement . . . 59
Figure 3.8 Example cosmic ray energy spectrum from a straw tube measurement . . . 62
Figure 3.9 MIP interactions in a straw tube . . . 62
Figure 3.10 Electron stopping power in the 80%/20% Ar-$\mathrm{CO_2}$ mixture . . . 64
Figure 3.11 A picture of the Mu2e straw tube comparing with a pencil . . . 65
Figure 3.12 Straws fully assembled in the Tracker panel . . . 65
Figure 3.13 Straw arrangement in a tracker panel . . . 66
Figure 3.14 An overview of the Tracker FEE . . . 67
Figure 3.15 A picture of the Mu2e Tracker preamp . . . 69
Figure 3.16 A picture of the DRAC board . . . 69
Figure 3.17 A chart of the signal flow in the Tracker FEE . . . 70
Figure 3.18 Example digitized signal from a Tracker straw channel . . . 70
Figure 4.1 Architecture of the Mu2e DAQ system . . . 75
Figure 4.2 A block diagram of the alternate panel control and operation system . . . 76
Figure 4.3 A block diagram of the firmware on the DIGI FPGA . . . 78
Figure 4.4 Bit alignment and bit slippage in the SerDes interface . . . 78
Figure 4.5 A block diagram of the firmware on the ROC . . . 79
Figure 5.1 Pictures of VST setup . . . 89
Figure 5.2 Signal dispersion and effective speed . . . 90
Figure 5.3 Source positions in the two collected scan datasets . . . 91
Figure 5.4 Example channel occupancy plot and fittings . . . 93
Figure 5.5 Channel occupancy plot of a source position close to panel edge . . . 94
Figure 5.6 Two example plots of source location reconstruction . . . 95
Figure 5.7 Spectrum and time distribution of hits at a fixed source location . . . 96
Figure 5.8 2-D histograms of hit time division and amplitude for various source locations along a straw . . . 97
Figure 5.9 Source coordinates, time divisions, and local effective speeds . . . 98
Figure 5.10 Maps of local half effective propagation speeds across the entire panel . . . 99
Figure 5.11 Longitudinal resolutions at various straw locations . . . 100

Figure 5.12 Longitudinal resolutions across panels 101
Figure 5.13 Reported cosmic-ray track fitting and resolutions 102
Figure 5.14 Longitudinal resolutions with different threshold settings 103
Figure 5.15 Longitudinal resolution asymmetry between straw ends 104
Figure 5.16 Unbalanced threshold settings and longitudinal resolution asymmetry 105
Figure 5.17 An illustration demonstrating the longitudinal resolution changes after a single-ended threshold change 105
Figure 5.18 Setup for measurements to compare actual thresholds and threshold readings . 107
Figure 5.19 Threshold and activation curve 108
Figure 5.20 Threshold examination data cleaning and normalization 108
Figure 5.21 Extraction of thresholds from activation curves 109
Figure 5.22 An example of a channel with relatively large gain correction 110
Figure 5.23 Threshold comparison using transition positions in activation curves 111
Figure 5.24 Threshold deviations from the readings 112
Figure 6.1 A schematic illustration of a possible bunch sequence for Mu2e-II 115
Figure 6.2 A schematic drawing of a Bunch Shape Monitor 117
Figure 6.3 A schematic drawing of the envisioned Beam Extinction Monitor 117
Figure 6.4 One-electron photoejection cross-section of H^- ions 118
Figure 6.5 An example of Beam Extinction Monitor mirror system optimization 118
Figure 6.6 An example electron multiplier in a photomultiplier tube 120
Figure 6.7 A schematic illustration of a microchannel plate 120
Figure 6.8 An example of secondary electron emission coefficients for different primary electron energies 121
Figure 6.9 Geiger-mode Avalanche Photodiode and Multi-Pixel Photon Counter 123
Figure 6.10 Low energy electron detection in VUV-MPPC from literature 124
Figure 6.11 The test circuit that biases the VUV-MPPC and amplifies the output 126
Figure 6.12 VUV-MPPC responses to electrons 126
Figure A.1 Simulated conversion positron signals and radiative muon capture (RMC) backgrounds in SINDRUM II 133
Figure C.1 PCB for VUV-MPPC efficiency measurements 141
Figure C.2 A diagram of the upgraded MPPC biasing and signal amplification circuit design 141

List of Tables

Table 1.1 Branching ratio limits of selected CLFV processes 11
Table 2.1 Summary of backgrounds and mitigations of muon-to-electron conversion experiments 27
Table 2.2 A summary of estimated background events from various sources 44
Table 4.1 Functions realized in the alternate panel control and operation system 84

List of Appendices

Appendix A Lepton Number Violation Search at Mu2e 132
Appendix B Functions of Alternate Tracker Control and Operation System 135
Appendix C VUV-MPPC Measurement PCB Design 140

List of Abbreviations

0$\nu\beta\beta$ neutrinoless double beta
2HDM two Higgs doublet model
ADC analog-to-digital converter
AMB Analog Motherboard
APD Avalanche Photodiode
BEM Beam Extinction Monitor
BR branching ratio
BSM Beyond-Standard-Model
BSM Bunch Shape Monitor
CFO Command Fan-Out
CL confidence level
CLFV Charged Lepton Flavor Violation
CRV Cosmic Ray Veto
CSDA continuous slowing down approximation
CW continuous wave
DAC Digital-to-Analog Converter
DAQ data acquisition
DCS Detector Control System
DIO decay-in-orbit
DMB Digital Motherboard
DOCA distance of closest approach
DRAC Digitizer Readout and Assembler Controller
DS Detector Solenoid
DTC Data Transfer Controller
ESS Electrostatic Septum
EM electromagnetic
EWM Event Window Marker
FEE front-end electronics
FIFO First-in, First-Out
FPGA Field-Programmable Gate Array
FWHM full width at half maximum
GAPD Geiger-mode Avalanche Photodiode
HPGe high purity germanium
I^{2}C Inter-Integrated Circuit
I/O Input/Output
IP Intellectual Property
IPA Inner Proton Absorber
LBNF/DUNE Long Baseline Neutrino Facility/Deep Underground Neutrino Experiment
LFV Lepton Flavor Violation
LNV Lepton Number Violation
LQ leptoquark
MCP microchannel plate
MIP minimum ionizing particle
MBS Muon Beam Stop
MPPC Multi-Pixel Photon Counter
μ-DIF muon decay-in-flight
NTP normal temperature and pressure
OPA Outer Proton Absorber
op-amp operational amplifier
PCB printed circuit board
π-DIF pion decay-in-flight
PIP-II Proton Improvement Plan-II
PMT photomultiplier tube
POT protons-on-target
preamp pre-amplifier

PS Production Solenoid
RF radio frequency
RFQ radio frequency quadrupole
RMC radiative muon capture
ROC Readout Controller
RPC radiative pion capture
SerDes Serializer/Deserializer
SES single-event sensitivity
SiPM Silicon Photomultiplier
SM Standard Model
sNVM secure Non-Volatile Memory
SPI Serial Peripheral Interface
STM Stopping Target Monitor
SUSY Supersymmetry
TDC time-to-digital converter
TOT time-over-threshold
TS Transport Solenoid
UART Universal Asynchronous Receiver-Transmitter
UV ultraviolet
VHDL Very High-Speed Integrated Circuit Hardware Description Language
VST Vertical Slice Test
VUV vacuum-ultraviolet

Abstract

In search of physics beyond the Standard Model (SM), the SM symmetries are put to scrutiny through seeking violations to the corresponding conservation laws. In particular, Charged Lepton Flavor Violation (CLFV) searches provide model-independent probes of new physics into high effective mass scales. The Mu2e experiment at Fermilab looks for the CLFV process of neutrino-less muon-to-electron conversion in the presence of aluminum nuclei. By applying careful background mitigation to a sub-event level and using high-performance detectors to detect the conversion signature of a single monoenergetic 105-MeV electron, the experiment aims to reach a single-event-sensitivity of 3×10^{-17}, and to set an upper limit for the muon conversion-to-capture ratio of $R_{\mu e} < 8 \times 10^{-17}$ at 90% confidence level. This improves the current experimental limit by four orders of magnitude. The Mu2e Tracker tracks the conversion electrons and provides momentum measurements. It consists of panels containing arrays of straw tubes and is required to measure the conversion electrons with a momentum resolution better than 180 keV/c. When translating to spatial resolutions on a single panel, the conversion electrons need to be measured with a longitudinal resolution of ~ 40 mm and a transverse resolution of ~ 200 μm. The Tracker uses a firmware-based system for data acquisition and operation. The system was developed with great robustness to fulfill various needs during testing, commissioning, and data taking. A Tracker Vertical Slice Test (VST) was performed using 1/36 of the Tracker, to verify detector performances and exercise the full readout chain. During the VST, panel scan studies using a radioactive ^{55}Fe source were performed to map effective signal propagation speeds as well as longitudinal resolutions across the entire panel. The Tracker longitudinal resolution was measured to be around 30 mm across the Tracker panels, consistent with the cosmic studies and satisfying the resolution requirement. The scan studies also provided more insights into straw channel threshold settings.

In addition to the work on the Mu2e Tracker, the progress in a Research and Design project to design a Beam Extinction Monitor (BEM) for the PIP-II Linac and Mu2e-II is also discussed. A system was designed to measure the extinction level of the H^- beam using a laser system. VUV-MPPCs were investigated for the electron detector in the BEM, and a proof-of-principle measurement was made to confirm its single-electron detection capability.

Chapter 1 Charged Lepton Flavor Violation

The Standard Model (SM) of particle physics provides the most successful model so far to describe our universe on the fundamental level. This $SU(3) \times SU(2) \times U(1)$ gauge theory [1, 2] explains the interactions among the elementary particles, and unifies three out of four fundamental forces of nature (Figure 1.1). Three generations of leptons and quarks (the fermions), together with their antiparticles, comprise the known matter of the universe. The gauge bosons, including photons (γ), gluons (g), and $W^{\pm}$ and Z bosons, are force carriers; they respectively mediate the electromagnetic interaction, the strong interaction, and the weak interaction. The Higgs boson, which is the only known scalar boson to date, gives rise to particle masses through the Higgs mechanism. The SM provided many successful predictions and has been verified through a wide range of experiments. For instance, it predicted the existence of the top quark and Higgs boson before their experimental discoveries [3–10]. Another example is the agreement between the SM prediction and the measurement of the electron magnetic dipole moment to within one part per trillion [11, 12].

Despite the great success, the Standard Model is not a "theory of everything". Many questions are still left unanswered. To name a few, the Standard Model does not include a theory of quantum gravity, or a candidate for dark matter. The absence of antimatter (a.k.a. matter-antimatter asymmetry) in the universe cannot be sufficiently explained by the CP violation predicted in the

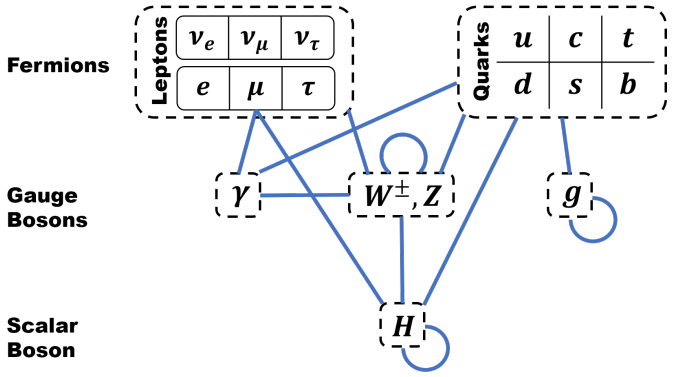


Figure 1.1: Elementary particles in the Standard Model of particle physics. Connections indicate the interactions between the particles.

Cabibbo-Kobayashi-Maskawa (CKM) mechanism [13, 14]. Moreover, tensions between theoretical predictions and experimental results emerged in recent high-precision measurements of the muon anomalous magnetic moment [15] and W boson mass [16]. And evidence of lepton-universality violations in recent b-quark decay measurements [17] have further shaken our basic understandings of the subject. The list goes on. We need to find physics beyond the SM to resolve these questions.

In the search of Beyond-Standard-Model (BSM) physics, it is important to identify and retain the fundamental symmetries in the current model. As Noether's theorem points out, each continuous symmetry of a local Lagrangian corresponds to a conservation law [1]. Thus testing conservation laws in the SM provides deeper insight into the underlying symmetries as well as guidance for the development of new theories. The symmetries in the SM include the discrete space-time symmetries,[1] as well as other global phase symmetries that correspond to conservation laws of certain quantum numbers. While some of these conservation laws have deeper roots (e.g., charge conservation is based from QED gauge symmetry, and fermion number conservation is rooted in Lorentz symmetry [18]), others seem to stem from accidental symmetries of the SM Lagrangian. These empirical conservation laws, once broken, can certainly indicate new interactions, and pilot the development of BSM physics.

One particular case that will be studied here is the conservation of the lepton family number. The process that violates this conservation law is generally referred to as Lepton Flavor Violation (LFV). This chapter will first introduce LFV in the SM. Discussions of Charged Lepton Flavor Violation (CLFV) in BSM theories will follow, and finally an overview of experimental searches will be provided, along with some discussions on the muon CLFV decay modes.

1.1 Lepton Flavor Violation in the Standard Model

1.1.1 Standard Model with Massless Neutrinos

In the "vanilla" version of the SM, there exists a single Higgs field Φ. The left-handed fermions are assigned as $SU(2)$ doublets, while the right-handed ones are singlets [1, 19]. In specific, the left-handed i-th generation quark doublets and lepton doublets are respectively denoted as $Q_{L\,i}$ and $L_{L\,i}$. Using u_i, d_i, ν_i, and e_i to indicate the i-th generation up-type quark, down-type quark, neutrino, and charged lepton, the transposes of the doublets are $Q^T_{L\,i} = (u_{L\,i} \;\; d_{L\,i})$ and $L^T_{L\,i} = (\nu_{L\,i} \;\; e_{L\,i})$. $u_{R\,j}$ and $d_{R\,j}$ are the right-handed quarks, while $e_{R\,j}$ is the right-handed (charged) lepton of the j-th generation. There is no right-handed neutrino. The Yukawa couplings of the fermion fields with the Higgs field $\mathcal{L}_Y$ contains a leptonic component $\mathcal{L}_e$ and a quark component

[1] The "discrete space-time symmetries" refer to charge conjugate C, parity P, time reversal T, as well as combined CP and CPT.

$\mathcal{L}_q$; $\mathcal{L}_Y = \mathcal{L}_e + \mathcal{L}_q$, and they are given by [19]:

$$-\mathcal{L}_e = (Y_e)_{ij}\overline{L}_{L\ i}e_{R\ j}\Phi + h.c. \tag{1.1}$$

$$-\mathcal{L}_q = (Y_u)_{ij}\overline{Q}_{L\ i}u_{R\ j}\widetilde{\Phi} + (Y_d)_{ij}\overline{Q}_{L\ i}d_{R\ j}\Phi + h.c. \tag{1.2}$$

Here $SU(2)_L$ indices are omitted, Y_f ($f = u, d, e$) are 3×3 Yukawa matrices containing coupling constants, and $\widetilde{\Phi} \equiv i\tau_2\Phi^*$ is the conjugate Higgs field (τ_2 is the second Pauli matrix). The fermions acquire their mass terms of $m_f\overline{f}_L f_R$ kind through the breaking of the $SU(2)_L \times U(1)$ symmetry by the vacuum expectation value of the Higgs field, which is given by [19]:

$$\langle\Phi\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ v\end{pmatrix},\ v \simeq 246\ \text{GeV}\ . \tag{1.3}$$

Then the fermion mass matrices can be obtained as:

$$(m_f)_{ij} = \frac{v}{\sqrt{2}}(Y_f)_{ij},\ f = u, d, e. \tag{1.4}$$

Clearly due to the non-existence of the right-handed neutrino singlet in this formulation, $\mathcal{L}_Y$ does not yield a neutrino mass term. Neutrinos are massless.

The Yukawa matrices can be diagonalized through unitary rotations. Writing the Yukawa matrices after diagonalization as $\hat{Y}_f$, that is:

$$Y_f = V_f\hat{Y}_f U_f^\dagger,\ f = u, d, e. \tag{1.5}$$

We denote the fermions in the rotated basis as f', then $f_L = V_f f'_L$ and $f_R = U_f f'_R$. The new basis is the physical basis, since it also diagonalizes the mass matrix. In other words, the couplings to the physical Higgs h takes the form:

$$-\mathcal{L}_{h\bar{f}f} = \frac{(\hat{m}_f)_{ij}}{v}\bar{f}'_{L\ i}f'_{R\ j}h + h.c.,\ f = u, d, e, \tag{1.6}$$

where $\hat{m}_f$ denotes the diagonalized mass matrices, and f' is used to emphasize the physical basis. Clearly, there are no flavor-violating terms, as $\hat{m}_f$ has no off-diagonal elements.

On the other hand, for neutral-current interactions (i.e., coupling with photon or Z boson), the unitary V_f and U_f in Equation 1.5 ensure that the neutral-current terms take the same form in either fermion basis, as in the formulation, $\bar{f}_L\gamma^\mu f_L = \bar{f}_L\gamma^\mu(V_f^\dagger V_f)f_L = \bar{f}'_L\gamma^\mu f'_L$ (similar for the right-handed components). There are no flavor-violating terms in neutral-current interactions either.

The gauge couplings with charged W boson currents are different. Here the interactions are:

$$\begin{aligned}\mathcal{L}_{CC} &= \frac{g}{\sqrt{2}}(\bar{u}_L\gamma^\mu d_L + \bar{\nu}_L\gamma^\mu e_L)W^+_\mu + h.c. \\ &= \frac{g}{\sqrt{2}}(\bar{u'}_L\gamma^\mu(V_u^\dagger V_d)d'_L + \bar{\nu'}_L\gamma^\mu(V_\nu^\dagger V_e)e'_L)W^+_\mu + h.c.\end{aligned} \tag{1.7}$$

In the quark sector, flavor violation arises as $V_u \neq V_d$. The flavor mixing is controlled by the matrix $V_{\rm CKM} \equiv V_u^\dagger V_d$, which is called the Cabibbo-Kobayashi-Maskawa (CKM) matrix [3, 20]. Meanwhile, as the lepton sector contains massless neutrinos, V_ν can be arbitrarily chosen as $V_\nu = V_e$. Once again, the lepton flavors are not mixed.

From the above discussions, there is no LFV of any kind in the SM with massless neutrinos. The Lagrangian $\mathcal{L}_Y$ remains invariant under three independent global $U(1)$ symmetries, which gives rise to three conserved lepton family numbers L_e, L_μ, and L_τ. Moreover, if the Yukawa coupling Lagrangian $\mathcal{L}_Y$ contains additional terms involving the lepton fields which uniquely determine some non-trivial basis through some extensions of the model, flavor violation will also occur in the lepton sector. Such addition can be, for example, a neutrino mass term or a second Higgs doublet [19].

1.1.2 Neutrino Oscillation and Charged Lepton Flavor Violation

First hints against the assumption of massless neutrinos arose with the solar neutrino problem. In the solar neutrino detection experiment carried out at Homestake in the 1960s, the observed number of solar neutrinos (neutrinos produced in the fission processes in the Sun) is significantly smaller than the expected value from the standard solar model (since the detector was only sensitive to ν_e) [18, 21]. Similar results were obtained in subsequent experiments using radiochemical detectors [18]. The following generation of experiments using Cherenkov detectors, including the Kamiokande experiment, the Super-Kamiokande experiment, and the SNO experiment, allowed real-time detection of the solar neutrinos. Through ν-e elastic scattering processes, these detectors were sensitive to neutrinos of all flavors. They established that there is a non-ν_e component in the solar neutrino flux [22, 23]. Combined with results from the KamLAND reactor neutrino experiment, the solar neutrino problem was resolved by the so-called large mixing angle solution [18]. The electron neutrinos produced in the Sun undergoes flavor transitions (a.k.a. neutrino oscillations) while propagating through the solar matter, as described by the Mihheev-Smirnov-Wolfenstein (MSW) effect [24, 25]. Three-generation neutrino oscillations have been established through many additional atmospheric, accelerator, and reactor neutrino experiments [18].

The lepton-flavor-violating neutrino oscillations showed that the global $U(1)$ symmetries associated with the lepton family numbers are not fundamental symmetries. An extension to the SM has to be made to include neutrino mass terms. In a minimally extended SM, one can either

introduce a right-handed neutrino singlet ν_R, or add some non-renormalizable operators [19]. The former introduces an addition to the Yukawa coupling Lagrangian $\Delta\mathcal{L}_D$ given by:

$$-\Delta\mathcal{L}_D = (Y_\nu)_{ij}\overline{L}_{L\ i}\nu_{R\ j}\widetilde{\Phi} + h.c. \tag{1.8}$$

Analogous to the other fermions, a Dirac mass term of $m_\nu\overline{\nu}_L\nu_R$ type is generated through symmetry breaking:

$$(m_\nu^D)_{ij} = \frac{v}{\sqrt{2}}(Y_\nu)_{ij} \tag{1.9}$$

On the other hand, adding non-renormalizable operators can introduce Majorana masses for left-handed neutrinos alone. The corresponding $\Delta\mathcal{L}_M$ can be written as [26, 27]:

$$-\Delta\mathcal{L}_M = \frac{1}{2}(m_\nu^M)_{ij}\overline{\nu_{L\ i}^C}\nu_{L\ j} + h.c., \tag{1.10}$$

where $\overline{\nu_L^C}$ represents charge-conjugated fields. This process violates lepton and baryon numbers, and requires an operator of dimension $d = 5$ [28] to be consistent with SM symmetries; a minimal effective Lagrangian is given by [27]:

$$-\Delta\mathcal{L}_{M\ \text{eff}} = \frac{\mathcal{C}_{ij}}{\Lambda}(\overline{L_{L\ i}^C}\tau_2\Phi)(\Phi^T\tau_2 L_{L\ j}) + h.c. \tag{1.11}$$

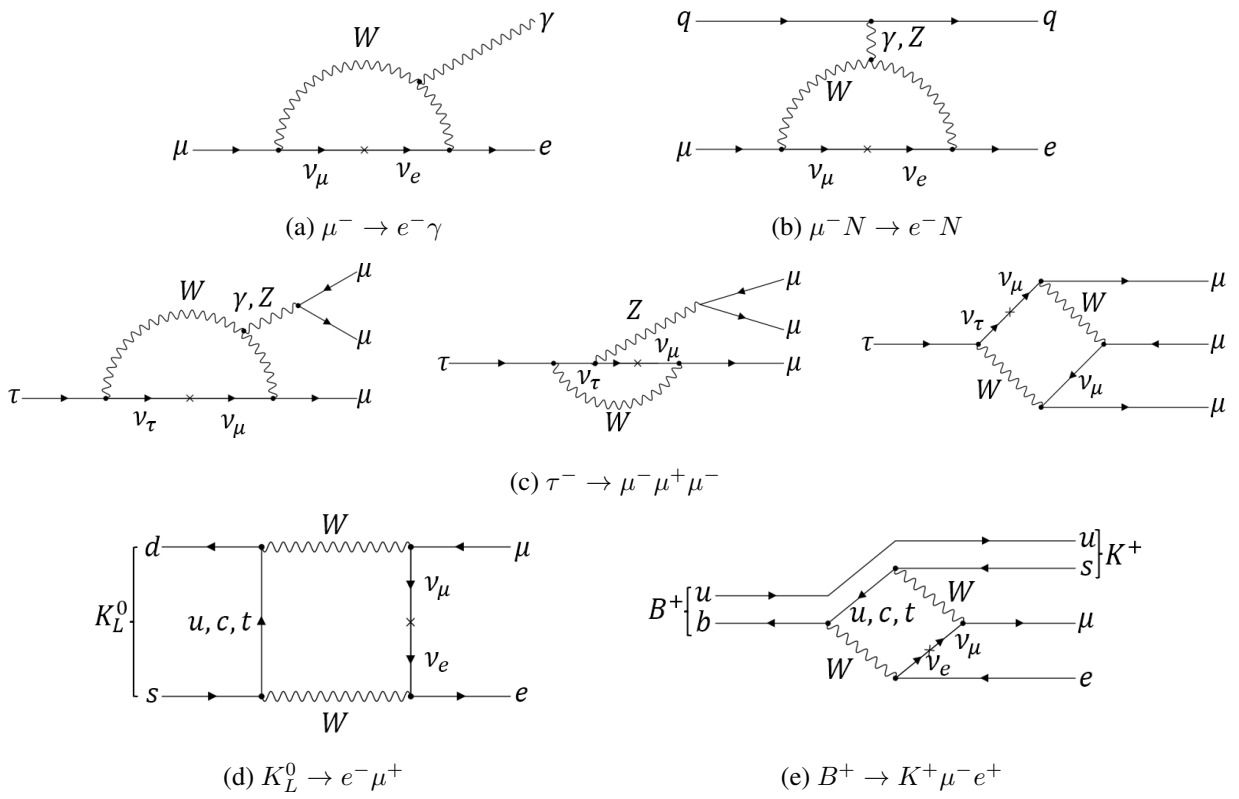
The Λ here is a mass scale characteristic of the extra degrees of freedom. $\mathcal{C}_{ij}$ is an antisymmetric charge conjugation matrix. The corresponding Majorana mass term is:

$$(m_\nu^M)_{ij} = \frac{\mathcal{C}_{ij}v^2}{\Lambda} \tag{1.12}$$

In this formulation, the smallness of the neutrino masses can be explained if $\Lambda \gg v$. Compared to the addition of ν_R singlet, where the same phenomenon requires a tiny value of $(Y_\nu)_{ij} \leq 10^{-12}$ [19], this may appear more natural.

Regardless of the exact form of the additional neutrino mass term, a physical basis diagonalizing the mass matrix is determined, so that in Equation 1.7, there is $V_\nu \neq V_e$ in general. Analogous to the CKM matrix, $V_{\text{PMNS}} \equiv V_\nu^\dagger V_e$ describes lepton mixing. It is normally referred to as the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix [29, 30]. As it is in the basis diagonalizing charged lepton masses, and it diagonalizes the neutrino mass matrix, V_{PMNS} also describes the mixing between neutrino flavor eigenstates ν_α and mass eigenstates ν_i:

$$\nu_\alpha = \sum_{i=1,2,3}(V_{\text{PMNS}})_{li}\nu_i,\ l = e, \mu, \tau \tag{1.13}$$



(a) $\mu^- \to e^-\gamma$ (b) $\mu^- N \to e^- N$

(c) $\tau^- \to \mu^-\mu^+\mu^-$

(d) $K_L^0 \to e^-\mu^+$ (e) $B^+ \to K^+\mu^- e^+$

Figure 1.2: Leading order Feynman diagrams of some CLFV processes in the SM. Examples involving muon, tau, kaon, and B meson are shown here. The SM rates of these processes are severely suppressed.

Apart from the neutral LFV seen in neutrino oscillations, in principle the mixing described by $V_{\rm PMNS}$ can also lead to Charged Lepton Flavor Violation (CLFV) processes, i.e., LFV involving charged leptons. In the discussed minimal extension to the SM (without adding new particles or couplings), leading order Feynman diagrams of CLFV processes contain loops and box diagrams with neutrinos and W bosons. Some examples are shown in Figure 1.2 [19, 31–33].

All these processes are severely suppressed in the SM [34, 35]. Taking $\mu \to e\gamma$ in Figure 1.2(a) as an example, the branching ratio (BR) of this process is calculated as [34, 36, 37]:[2]

$$\begin{aligned} BR(\mu \to e\gamma) &= \frac{3\alpha}{32\pi}\left|\sum_{i=2,3} V^*_{\mu i}V_{ei}\frac{\Delta m^2_{1i}}{M^2_W}\right|^2 \\ &= \frac{3\alpha}{32\pi}\left(\frac{1}{4}\right)\sin^2 2\theta_{13}\sin^2\theta_{23}\left|\frac{\Delta m^2_{13}}{M^2_W}\right|^2 , \end{aligned} \tag{1.14}$$

where α is the fine structure constant, $V_{\mu i}$ and V_{ei} are corresponding elements in the PMNS matrix, Δm^2_{1i} indicates the neutrino squared mass differences, M_W is the W boson mass, and θ_{13} and

[2]Here the branching ratio is normalized to partial width $\Gamma(\mu \to e\nu\bar{\nu}) = 1$.

θ_{23} are rotating angles in PMNS matrix parametrization [18]. The expression is evaluated to be $BR(\mu \to e\gamma) \sim \mathcal{O}(10^{-54})$ [36]; the huge mass difference between the neutrinos and the W boson determines that $|\Delta m^2_{13}/M^2_W|^2$ is forbiddingly small. Similar suppression terms are also present for other CLFV processes due to the similar topology of the Feynman diagrams. The SM rates are way too small for any experiment to measure within the foreseeable future.

On the other hand, many BSM theories include mechanisms that significantly enhance the rates of CLFV, as will be discussed in the next section. The tiny magnitude of the SM CLFV rates means detection of any CLFV processes in the experiments will be unambiguous evidence of new physics beyond the SM. Such approach is model-independent, and as will be shown later, allows us to probe into high mass scales. CLFV processes are great tools to search for BSM physics.

1.2 Charged Lepton Flavor Violation Beyond the Standard Model

Many BSM theories predict mechanisms contributing to CLFV processes, some of which can give rise to rates possibly detectable in experiments. A few representative examples of BSM theories with CLFV contributions are outlined here. This is by no means an exhaustive list; additional reviews can be found in [19, 38–42].

1.2.1 Supersymmetric Theories

Theories of Supersymmetry (SUSY) were developed as frameworks to address the naturalness problem,[3] and to unify quantum field theory and general relativity at the Planck scale [43, 44]. While details may vary for different versions of SUSY models, the basic idea is to introduce SUSY partners for SM particles. Squarks ($\tilde{q}$) and sleptons ($\tilde{l}$) correspond to quarks and leptons, Higgs boson matches with higgsino ($\tilde{H}$), while gauge bosons partner with gauge fermions (gluino ($\tilde{G}$), wino ($\tilde{W}$) and bino ($\tilde{B}$)). Electroweak symmetry breaking mixes $\tilde{H}$, $\tilde{W}$, and $\tilde{B}$ to form charginos ($\tilde{\chi}^{\pm}_i$) and neutralinos $\tilde{\chi}^0_i$ as mass eigenstates.

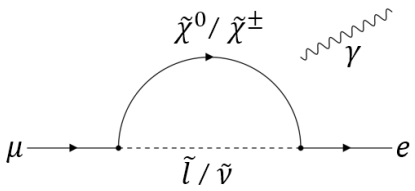


Figure 1.3: SUSY contribution to $\mu \to e\gamma$ [19]. The photon is attached to the charged fermion in the loop.

[3] Regarding the SM as a low-energy approximation of a more complete theory at a higher energy level (presumably at the Planck scale $\sim 10^{19}$ GeV), a precise fine-tuning between the radiative corrections and the bare Higgs scalar mass is necessary to keep the electroweak scale low ($\sim 10^2$ GeV). This is the so-called "naturalness problem", or "hierarchy problem".

Additional couplings are introduced along with the SUSY particles. These additional SUSY couplings lead to contributions to CLFV. Again using $\mu \to e\gamma$ as an example, a typical Feynman diagram of the SUSY contribution is shown in Figure 1.3 [19]. Despite the similar topology to that of the SM contribution, the typical SUSY mass is expected to be much higher than that of the neutrinos. Predictions of the BR of this process vary for different versions of SUSY models, dependent on the detailed mechanisms (involvement of different SUSY gauge fermions, handedness of sleptons, etc.) and particle masses. The rates can be significantly enhanced. For instance, in an $SU(5)$ SUSY grand unified theory, the calculated branching ratio can reach $\mathcal{O}(10^{-14})$ for a slepton mass of $\mathcal{O}(10^2\ \mathrm{GeV/c^2})$ [39], which is a value measurable in upcoming experiments.

1.2.2 Additional Higgs Doublets

While only a single Higgs is present in the SM, there is nothing preventing the existence of additional Higgs fields. The simplest example for a full theory containing multiple Higgs fields is the type III two Higgs doublet model (2HDM), where there are two Higgses; each couples to fermions and has a vacuum expectation value [41]. In general, the Lagrangian with extra Higgs fields after electroweak symmetry breaking can be expressed as [19, 41]:

$$-\mathcal{L} = m_i \overline{f}_{L\,i} f_{R\,i} + (Y^\alpha)_{ij} \overline{f}_{L\,i} f_{R\,j} h^\alpha + h.c. + \ldots \tag{1.15}$$

Here non-renormalizable terms of higher dimensions are omitted. m_i contains contributions from all vacuum expectation values. $(Y^\alpha)_{ij}$ represents couplings to a single scalar field, and the contributions from several Higgses are summed over. In general, the basis that diagonalizes the mass matrix does not diagonalize $(Y^\alpha)_{ij}$ matrices. The non-zero off-diagonal terms in $(Y^\alpha)_{ij}$ give rise to flavor-violating Yukawa couplings. This means CLFV processes such as $\mu \to eee$ and $\mu \to e$ conversion in the presence of a nucleus can proceed at the tree level through exchanging Higgs [19, 41].

Constraints on the off-diagonal coupling constants $Y_{l_1 l_2}$ of the 125 GeV Higgs boson can be obtained from various accelerator and precision experiments [41]. Figure 1.4 [41] shows the upper limits of the coupling constants. The colored regions have been excluded due to the non-observation of various CLFV processes. Meanwhile, for the 125 GeV Higgs boson, the upper limits of $H \to l_1 l_2$ branching ratios [18, 45, 46] are listed in Table 1.1 along with other CLFV processes in Section 1.3.

1.2.3 Leptoquark Models

Leptoquarks (LQs) are hypothetical particles first proposed in the Pati-Salam model [47]. Each LQ is associated with both baryon number (B) and lepton number (L). In various LQ models, the quark sector and the lepton sector are unified [42, 48–50], for example in groups $SU(5)$ and

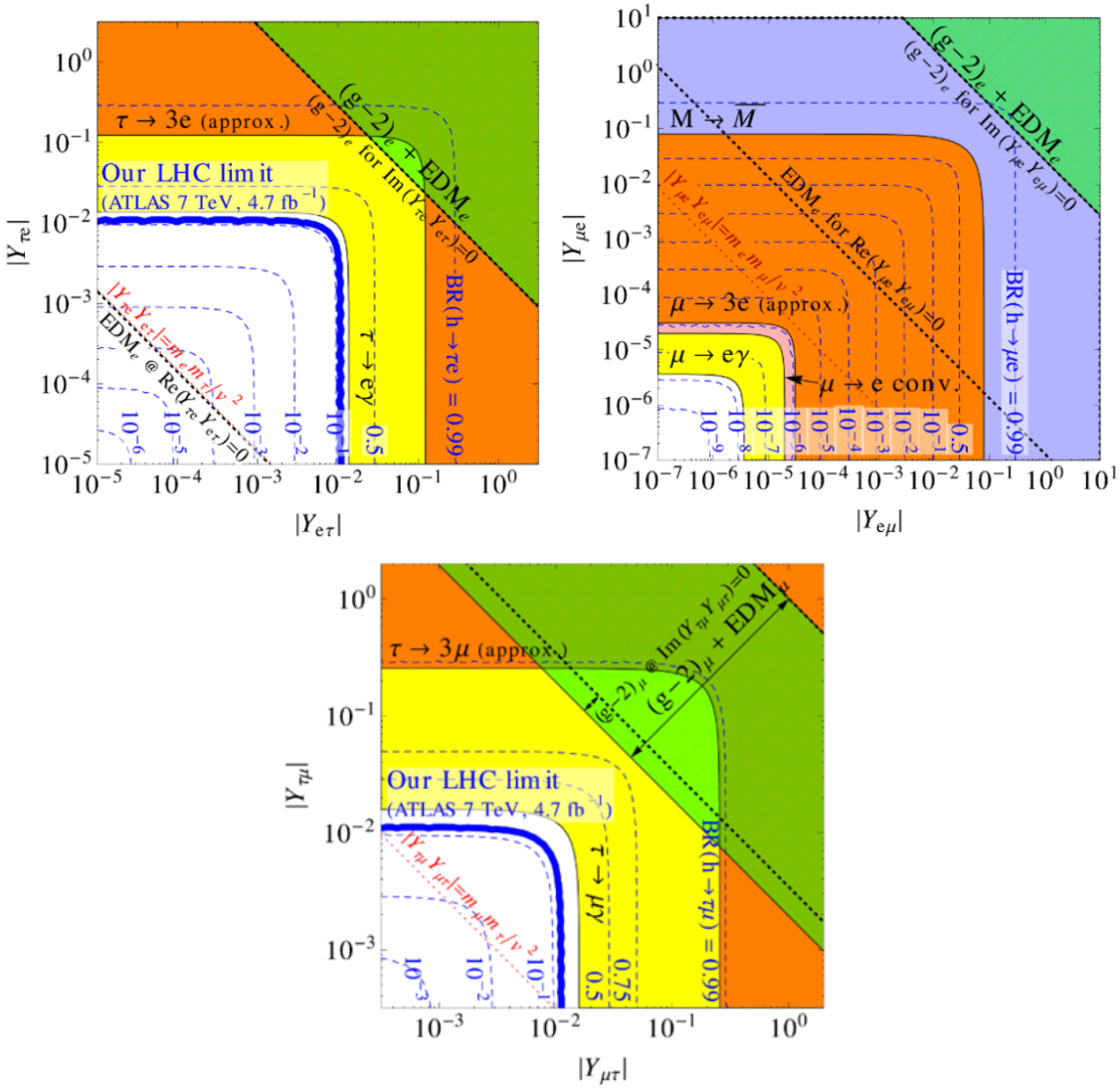
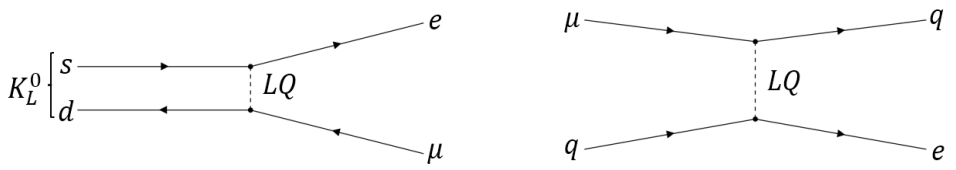


Figure 1.4: Constraints on flavor-violating Yukawa couplings to the 125 GeV Higgs. The three panels show couplings between $e\tau$ (top left), $e\mu$ (top right), and $\mu\tau$ (bottom) respectively. The colored regions have been excluded using the upper limits of various CLFV processes.

$SO(10)$.[4] Through exchange of LQs, the quarks are directly coupled with the leptons. As a result, certain CLFV processes such as $K_L^0 \to e\mu$ and $\mu \to e$ in nuclei receive tree-level contributions through an intermediate LQ, as shown in Figure 1.5.

Constraints on LQ models can be obtained both directly in collider experiments and indirectly through rare decay searches. The former approach has ruled out first and second generation scalar LQs under a mass of $\sim$ 1 TeV through searches at ATLAS and CMS [18]. The indirect searches, on the other hand, provide constraints on the mass-coupling plane, where the regions

[4]Certain grand unified theories based upon $SU(5)$ and $SO(10)$ include LQs. Some SUSY models also have scalar quarks with LQ-like Yukawa couplings [18].



(a) $K_L^0 \to e^- \mu^+$ (b) $\mu^- N \to e^- N$

Figure 1.5: Leptoquark interactions contribute to some CLFV processes at the tree level.

with higher couplings and lower LQ masses correspond to higher branching ratios. One example is shown in [51]. Additional constraints on LQ models are imposed by proton stability, as many models include LQs that can mediate proton decay. To suppress this process, the corresponding LQs have to be extremely heavy and / or the related couplings have to be extremely small [42].

1.2.4 Additional Neutral Gauge Boson

In the search for a more fundamental model, grand unified theories based on extended gauge groups (such as E_6 and $SO(10)$) are constructed [52, 53]. At low energy, the original extended gauge groups are thought to effectively break to the direct product of the SM gauge group $SU(3) \times SU(2) \times U(1)$ and an additional $U(1)$ factor [53]. The neutral gauge boson associated with this $U(1)$ group can mix with the original SM neutral gauge boson, which gives rise to two mass eigenstates Z and Z'. On the other hand, additional fermion fields are required in extended gauge theories (for cancelling anomaly free currents beyond those of $SU(5)$) [52]. These "new" fermions can mix with the known SM fermions with the same electric and color charges, and affect their couplings with the gauge bosons. Off-diagonal terms in neutral current couplings to fermions can appear, which lead towards flavor-changing couplings to Z and Z'. Certain CLFV processes, such as $\mu \to eee$ and μ-e conversion, again receive tree-level contributions through intermediate Z and Z' bosons [53].

Additional review on the phenomenology of the Z' boson can be found in [54]. The existence of Z' is searched for through channels such as $Z' \to f\overline{f}$ at hadron colliders. Mass lower limits of Z' from various specific models are listed in [18]; the values are mostly in the low TeV range. In particular, the mass lower limits reported in CLFV final states $e\mu$, $e\tau$, and $\mu\tau$ are between 3.5 TeV and 4.5 TeV [55, 56]. In terms of couplings to the normal Z boson, the upper limits of $Z \to l_1 l_2$ [18, 46, 57] are also listed in Table 1.1.

* * *

While all the BSM models remain theoretical, CLFV searches can provide more insights into these theories. Observation of any CLFV process in experiments will be a huge breakthrough

leading towards new physics. By combining results from various CLFV channels, we may eventually pinpoint the underlying model. Meanwhile, null results from these searches can help us constrain the models and put limits to their parameter spaces.

1.3 Charged Lepton Flavor Violation Searches and Muon Rare Decays

CLFV processes have been long searched for in experiments, and no observation has been made so far. Table 1.1 provides a list of current best (upper) limits on selected channels [18].[5] Apart from the direct search at collider experiments (e.g., the Z decays and the Higgs decays in Table 1.1), the rare decay experiments comprise an important complimentary approach in the CLFV search. While the former allows probing of different CLFV couplings at the same time including those involving the heavy τ, complexities in data selection and reconstruction, along with limitations on statistics, make it challenging to significantly improve the limits. This is when the rare decay

Table 1.1: Branching ratio limits of selected CLFV processes (confidence level $\mathrm{CL} = 90\%$ unless otherwise noted) [18]

Process	Branching Ratio Upper Limit
$Z \to e\mu$	$BR < 7.5 \times 10^{-7}$, $\mathrm{CL} = 95\%$
$Z \to e\tau$	$BR < 9.8 \times 10^{-6}$, $\mathrm{CL} = 95\%$
$Z \to \mu\tau$	$BR < 1.2 \times 10^{-5}$, $\mathrm{CL} = 95\%$
$H \to e\mu$	$BR < 6.1 \times 10^{-5}$, $\mathrm{CL} = 95\%$
$H \to e\tau$	$BR < 4.7 \times 10^{-3}$, $\mathrm{CL} = 95\%$
$H \to \mu\tau$	$BR < 2.5 \times 10^{-3}$, $\mathrm{CL} = 95\%$
$\mu^- \to e^-\gamma$	$BR < 4.2 \times 10^{-13}$
$\mu^- \to e^-e^+e^-$	$BR < 1.0 \times 10^{-12}$
$\mu^-\mathrm{Au} \to e^-\mathrm{Au}$	$R_{\mu e} < 7 \times 10^{-13}$
$\tau^- \to e^-\gamma$	$BR < 3.3 \times 10^{-8}$
$\tau^- \to e^-\pi^0$	$BR < 8.0 \times 10^{-8}$
$\tau^- \to \mu^-\eta$	$BR < 6.5 \times 10^{-8}$
$\tau^- \to e^-e^+e^-$	$BR < 2.7 \times 10^{-8}$
$\tau^- \to \mu^+e^-e^-$	$BR < 1.5 \times 10^{-8}$
$K_L^0 \to e\mu$	$BR < 4.7 \times 10^{-12}$
$K^+ \to \pi^+e^-\mu^+$	$BR < 1.3 \times 10^{-11}$
$B^0 \to e\mu$	$BR < 1.0 \times 10^{-9}$
$B^0 \to e\tau$	$BR < 2.8 \times 10^{-5}$
$B^+ \to \pi^+e\mu$	$BR < 1.7 \times 10^{-7}$
$B^+ \to K^+e^-\mu^+$	$BR < 6.4 \times 10^{-9}$

[5] $R_{\mu e}$ is the conversion-to-capture rate. More details are in the next chapter.

experiments following the intensity frontier approach become useful. In an experiment dedicated to a specific process (or a few similar processes), high statistics can be accumulated using an intense particle beam. With background signals carefully suppressed, the experiment sensitivity can be largely improved, hence providing a stringent limit on the CLFV process. Such an approach also allows probing into high mass scales, which will be shown in a later part of this section.

Among the rare decay experiments, the ones probing the muon rare decay channels $\mu \to e\gamma$, $\mu \to eee$, and $\mu N \to eN$ stand out in particular. As shown in Table 1.1, these channels provide the lowest branching ratio limits. These searches are very sensitive in the $e\mu$ CLFV mode. The small muon mass means a short list of viable decay modes, along with relatively simple search backgrounds compared to other heavier particles. However, it comes at a cost of not being able to probe couplings to τ. Over the past decades, the branching ratio limits of the muon rare decays have quickly improved as shown in Figure 1.6 [58, 59]. The next-generation experiments aim at further improvements by several orders of magnitude, reaching the levels predicted by some BSM models discussed in the previous section.

All next-generation experiments start with stopped muons on some stopping targets. The MEG-II at Paul Scherrer Institute (PSI) in Switzerland searches for the $\mu^+ \to e^+\gamma$ process [60]. Aiming at a branching ratio measurement to the level of 6×10^{-14}, it measures the two-body

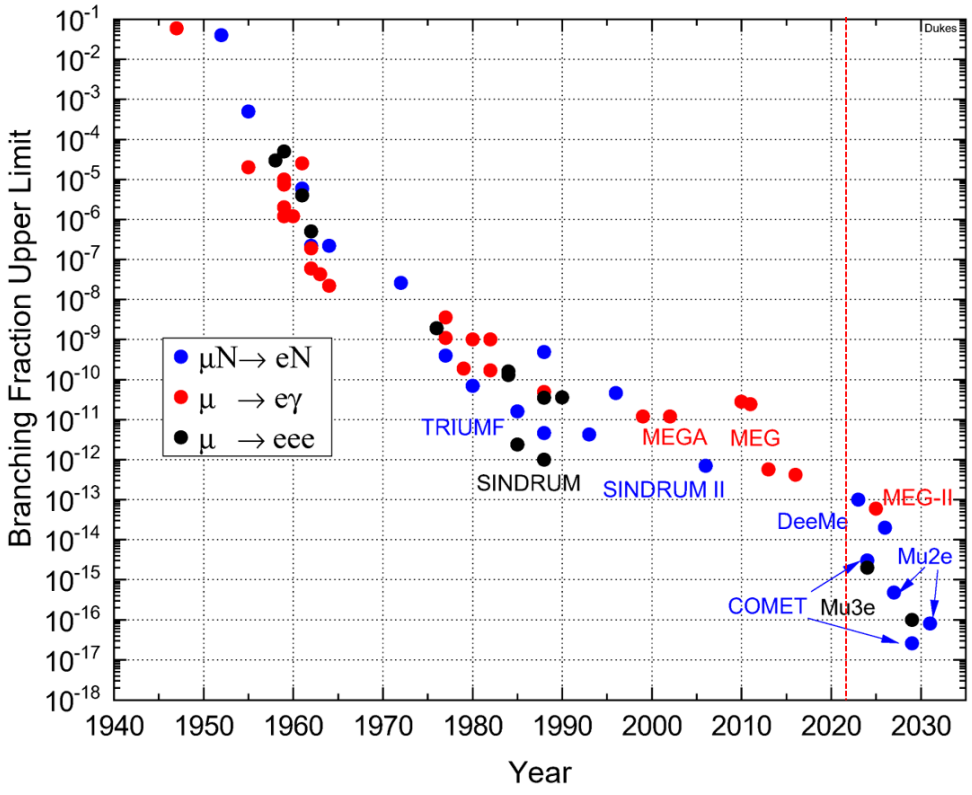


Figure 1.6: History and outlook of branching ratio limits in muon rare decay modes [59].

decay products from stopped muons. The positrons are tracked in a drift chamber, whose time are precisely measured using pixelated timing counters comprised of scintillation tiles, and the photons are detected in a liquid xenon detector through scintillation light in the vacuum-ultraviolet range. The Mu3e experiment, also located at PSI, searches for $\mu \to eee$ with an aimed branching fraction of 10^{-16} [61]. Using solid-state pixel tracking modules, scintillation fibers, and tile detectors, the electron and positron momenta and time are measured. And as for the $\mu N \to eN$ process, both the Mu2e experiment at Fermilab [62] and the COMET experiment at J-PARC (Japan) [63] are searching for the process. The two experiments have similar search goals on the 10^{-17} level and share resemblances in detector designs; the major difference is in the geometries of detector layouts. More about the Mu2e experiment will be introduced in the following chapters.

Apart from the high sensitivities, the muon rare decay experiments can provide theory differentiation power when combining the results from $\mu \to e\gamma$, $\mu \to eee$, and $\mu N \to eN$ channels. We can consider the following model-independent effective Lagrangian of the $e\mu$ CLFV mode [64]:

$$\begin{aligned}\mathcal{L}_{\text{CLFV}} =& \frac{m_\mu}{(1+\kappa)\Lambda^2}\overline{\mu}_R\sigma_{\mu\nu}e_L F^{\mu\nu} + \\ & \frac{\kappa}{(1+\kappa)\Lambda^2}\overline{\mu}_L\gamma_\mu e_L\left(\sum_{q=u,d}\overline{q}_L\gamma^\mu q_L\right) + h.c.\end{aligned} \tag{1.16}$$

Here m_μ is the muon mass, $F^{\mu\nu}$ is the electromagnetic field tensor. Two parameters are introduced in this toy Lagrangian. Λ represents the effective mass scale of the new degree of freedom, and κ determines the relative strengths of the two operators. The first term in the Lagrangian represents a magnetic moment type operator which directly mediates $\mu \to e\gamma$. Additional couplings are present when it mediates $\mu \to eee$ and $\mu N \to eN$, so these rates are much lower when this term dominates, which requires $\kappa \ll 1$. The second term in $\mathcal{L}_{\text{CLFV}}$ corresponds to a four-fermion operator, which mediates $\mu N \to eN$ at tree level and the other two processes at one-loop level. When $\kappa \gg 1$, the second term is dominant, and the rates of $\mu N \to eN$ will be much higher than the other two.

The corresponding parameter space (the Λ-κ plane) is shown in Figure 1.7 [64]. The colored regions in the figure have been ruled out by past experiments. The solid and dashed lines in the figure represent contours of branching ratios in a certain decay mode. Experiments measuring $\mu N \to eN$ are more sensitive when $\kappa \gg 1$; the same branching ratio limit allows probing into higher effective mass scales in this κ range. For instance, the Mu2e experiment can probe an effective mass scale up to $\mathcal{O}(10^4\ \text{TeV})$ with its designed sensitivity assuming $\kappa \gg 1$. This effective mass (energy) scale is much higher than energy levels of any foreseeable collider experiments. On the other hand, $\mu \to e\gamma$ experiments are more sensitive when $\kappa \ll 1$; the dominant magnetic moment type term determines the other two processes have lower rates in such a case. Were CLFV ever observed in muon rare decays, ratios between rates of different decay modes could be used to

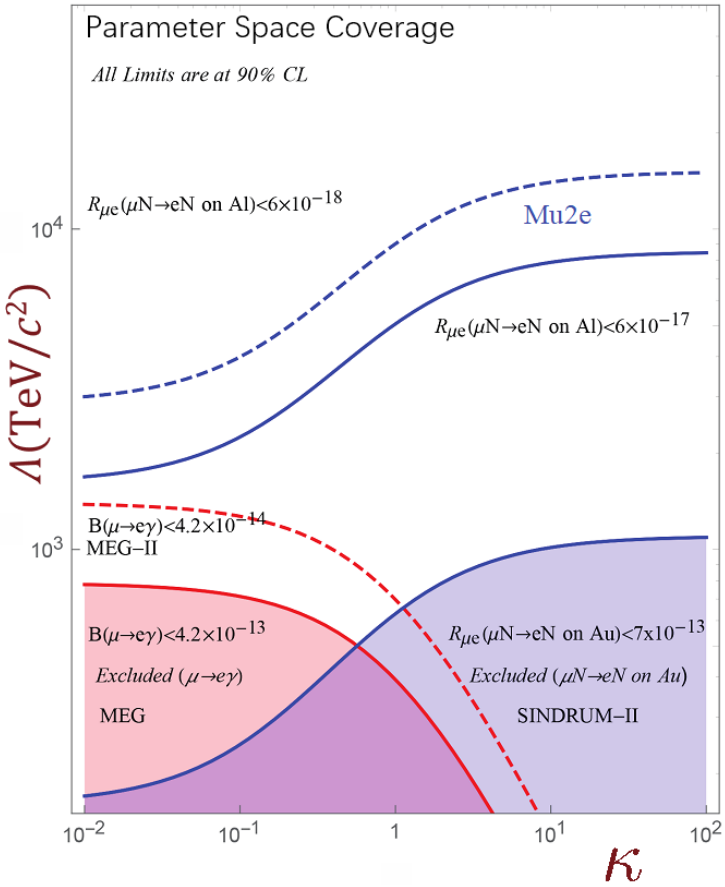


Figure 1.7: Toy CLFV Lagrangian parameter space coverage by past and upcoming muon rare decay experiments [64]. The solid and dashed lines correspond to certain branching ratios in respective decay modes.

differentiate the underlying theory [65].

* * *

Now we have established the motivations to search for CLFV processes and demonstrated the advantages of utilizing the muon rare decay modes for the search. Starting from the next chapter, we will focus on a particular CLFV search in the $\mu N \to eN$ channel, i.e., the Mu2e experiment.

References

[1] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory* (Westview Press, 1995).

[2] M. K. Gaillard, P. D. Grannis, and F. J. Sciulli, Rev. Mod. Phys. **71**, S96 (1999).

[3] M. Kobayashi and T. Maskawa, Prog. Theor. Exp. Phys **49**, 652 (1973).

[4] F. Abe, *et al.* (CDF Collaboration), Phys. Rev. Lett. **74**, 2626 (1995).

[5] S. Abachi, *et al.* (D0 Collaboration), Phys. Rev. Lett. **74**, 2632 (1995).

[6] P. W. Higgs, Phys. Rev. Lett. **13**, 508 (1964).

[7] F. Englert and R. Brout, Phys. Rev. Lett. **13**, 321 (1964).

[8] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, Phys. Rev. Lett. **13**, 585 (1964).

[9] G. Aad *et al.* (ATLAS Collaboration), Physics Letters B **716**, 1 (2012).

[10] S. Chatrchyan *et al.* (CMS Collaboration), Phys. Lett. B **716**, 30 (2012).

[11] D. Hanneke, S. Fogwell, and G. Gabrielse, Phys. Rev. Lett. **100**, 120801 (2008).

[12] T. Aoyama, *et al.*, Prog. Theor. Exp. Phys **2012**, 01A107 (2012).

[13] A. D. Sakharov, Pisma Zh. Eksp. Teor. Fiz. **5**, 32 (1967).

[14] A. Riotto, in *Proceedings, Summer School in High-energy Physics and Cosmology*, edited by E. Gava *et al.* (World Scientific, 1998) pp. 326–436.

[15] B. Abi, *et al.* (Muon g-2 Collaboration), Phys. Rev. Lett. **126**, 141801 (2021).

[16] T. Aaltonen, *et al.* (CDF Collaboration), Science **376**, 170 (2022).

[17] R. Aaij *et al.* (LHCb Collaboration), Nat. Phys. **18**, 277 (2022).

[18] P. Zyla *et al.* (Particle Data Group), Prog. Theor. Exp. Phys. **2020**, 083C01 (2020), and 2021 update.

[19] L. Calibbi and G. Signorelli, Riv. Nuovo Cim. **41**, 71 (2018).

[20] N. Cabibbo, Phys. Rev. Lett. **10**, 531 (1963).

[21] R. Davis, D. S. Harmer, and K. C. Hoffman, Phys. Rev. Lett. **20**, 1205 (1968).

[22] Q. R. Ahmad, *et al.* (SNO Collaboration), Phys. Rev. Lett. **89**, 011301 (2002).

[23] B. Aharmim, *et al.* (SNO Collaboration), Phys. Rev. C **72**, 055502 (2005).

[24] L. Wolfenstein, Phys. Rev. D **17**, 2369 (1978).

[25] S. P. Mikheyev and A. Y. Smirnov, Sov. J. Nucl. Phys. **42**, 913 (1985).

[26] S. M. Bilenky and S. T. Petcov, Rev. Mod. Phys. **59**, 671 (1987), [Erratum: Rev. Mod. Phys. 61, 169 (1989), Erratum: Rev. Mod. Phys. 60, 575–575 (1988)].

[27] S. Bilenky, C. Giunti, and W. Grimus, Prog. Part. Nucl. Phys. **43**, 1 (1999).

[28] S. Weinberg, Phys. Rev. Lett. **43**, 1566 (1979).

[29] B. Pontecorvo, Sov. Phys. JETP **6**, 429 (1957).

[30] Z. Maki, M. Nakagawa, and S. Sakata, Prog. Theor. Phys. **28**, 870 (1962).

[31] G. Hernández-Tomé, G. L. Castro, and P. Roig, Eur. Phys. J. C **79**, 84 (2018), [Erratum: Eur. Phys. J. C 80, 438 (2020)].

[32] A. Crivellin, *et al.*, Phys. Rev. D **93**, 074038 (2016).

[33] B. Aubert, *et al.* (BABAR Collaboration), Phys. Rev. D **73**, 092001 (2006).

[34] W. Marciano and A. Sanda, Phys. Lett. B **67**, 303 (1977).

[35] A. Rossi, Phys. Rev. D **66**, 075003 (2002).

[36] R. Bernstein and P. S. Cooper, Phys. Rep. **532**, 27 (2013).

[37] B. W. Lee and R. E. Shrock, Phys. Rev. D **16**, 1444 (1977).

[38] M. Raidal, *et al.*, Eur. Phys. J. C **57**, 13 (2008).

[39] Y. Kuno and Y. Okada, Rev. Mod. Phys. **73**, 151 (2001).

[40] W. Altmannshofer, *et al.*, Nucl. Phys. B **830**, 17 (2010).

[41] R. Harnik, J. Kopp, and J. Zupan, J. High Energy Phys. **2013** (3).

[42] I. Doršner, *et al.*, Phys. Rep. **641**, 1 (2016).

[43] H. Nilles, Phys. Rep. **110**, 1 (1984).

[44] H. Haber and G. Kane, Phys. Rep. **117**, 75 (1985).

[45] A. M. Sirunyan *et al.* (CMS Collaboration), Phys. Rev. D **104**, 032013 (2021).

[46] T. Davidek and L. Fiorini, Front. Phys **8**, 149 (2020).

[47] J. C. Pati and A. Salam, Phys. Rev. D **10**, 275 (1974).

[48] H. Georgi and S. L. Glashow, Phys. Rev. Lett. **32**, 438 (1974).

[49] H. Georgi, AIP Conf. Proc **23**, 575 (1975).

[50] H. Fritzsch and P. Minkowski, Ann. Phys. (N. Y.) **93**, 193 (1975).

[51] J. M. Arnold, B. Fornal, and M. B. Wise, Phys. Rev. D **88**, 035009 (2013).

[52] E. Nardi, E. Roulet, and D. Tommasini, Phys. Rev. D **46**, 3040 (1992).

[53] J. Bernabéu, E. Nardi, and D. Tommasini, Nucl. Phys. B **409**, 69 (1993).

[54] A. Leike, Phys. Rep. **317**, 143 (1999).

[55] M. Aaboud *et al.* (ATLAS Collaboration), Phys. Rev. D **98**, 092008 (2018).

[56] A. M. Sirunyan *et al.* (CMS Collaboration), J. High Energy Phys **2018**, 73 (2018).

[57] G. Aad *et al.* (ATLAS Collaboration), Nat. Phys. **17**, 819 (2021).

[58] W. J. Marciano, T. Mori, and J. M. Roney, Annual Review of Nuclear and Particle Science **58**, 315 (2008).

[59] E. C. Dukes, "Mu2e-II: The Mu2e Experiment in the PIP-II Era," (NuFact 2021 2021-09-08), mu2e-doc-39318-v3.

[60] A. M. Baldini, *et al.*, Eur. Phys. J. C **78**, 10.1140/epjc/s10052-018-5845-6 (2018).

[61] K. Arndt, *et al.*, Nucl. Instrum. Methods Phys. Res. A: Accel. Spectrom. Detect. Assoc. Equip. **1014**, 165679 (2021).

[62] L. Bartoszek, *et al.* (Mu2e Collaboration), *Mu2e Technical Design Report*, Tech. Rep. (Fermilab, 2015) arXiv:1501.05241 [physics.ins-det] .

[63] R. Abramishvili, *et al.* (COMET Collaboration), Prog. Theor. Exp. Phys **2020** (2020), 033C01.

[64] A. de Gouvêa and P. Vogel, Prog. Part. Nucl. Phys. **71**, 75 (2013).

[65] J. L. Hewett *et al.*, in *Proceedings of the 2011 workshop on Fundamental Physics at the Intensity Frontier* (2012) arXiv:1205.2671 [hep-ex] .

Chapter 2 The Mu2e Experiment

The Mu2e experiment is located on the Fermilab muon campus. It searches for the neutrinoless muon-to-electron conversion in the presence of a nucleus ($\mu^- N \rightarrow e^- N$), a CLFV process as discussed in the previous chapter. The experiment expects to set an upper limit for the conversion-to-capture ratio of $R_{\mu e} < 8 \times 10^{-17}$ at 90% confidence level (CL). This is an improvement of $\sim 10^4$ compared to the current best limit set by the SINDRUM II experiment in 2006 [1].

In this chapter, the experiment concept of Mu2e will first be introduced. A discussion on various sources of backgrounds in the experiment will follow, along with the mitigations. The rest of the chapter will focus on the experiment setup. The systems of the Mu2e experiment will be introduced, detailing how the background mitigation techniques are implemented in the experiment, enabling it to reach the designed level of sensitivity.

2.1 Experiment Concept

Figure 2.1 illustrates the basic experiment concept of a muon to-electron conversion experiment, which Mu2e follows closely. The first step is to produce a beam of negative muons μ^-. This is achieved by impinging a proton beam on a production target. The process produces negative pions, among other mesons and hadrons. The pions are short-lived and decay in flight. Due to

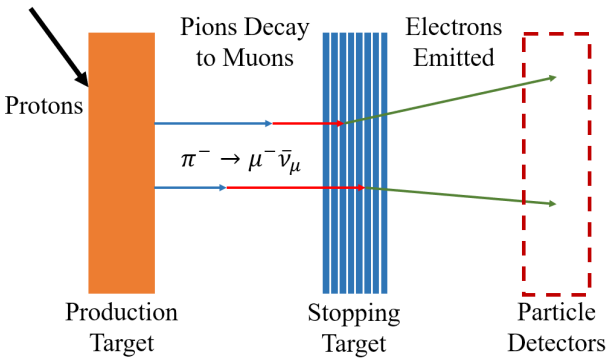


Figure 2.1: An illustration of the Mu2e experiment concept. Secondary muon beam is captured on a stopping target, and particle detectors measure the electrons emitted from the muonic atom decays.

helicity suppression [2], $\pi^- \to \mu^- \bar{\nu}_\mu$ occurs at a BR of more than 99.9% [3]. An intense muon beamline is required to accumulate enough statistics within realistic time.

Low momentum secondary muons are then captured on a stopping target to form muonic atoms. Within a timescale of $\sim 10^{-13}$ s, a muonic atom cascades down to the $1s$ state [4]. Given the fact that the cascading time is so short compared to the mean lifetime of a muon in a muonic atom, which is typically on the order of 10^{-6} s [5], it is very unlikely for muons to decay before reaching the ground state. Cases involving muons decaying in non-ground states are thus negligible [6]. The cascades emit x-ray photons, whose spectrum can be used to estimate the number of muons captured on the stopping target [6].

The muonic atoms decay after a certain lifetime that is dependent on the material of the stopping target. The Mu2e experiment uses aluminum stopping target, which corresponds to a lifetime of 864 ns [5]. The muons decay through two major processes, nuclear capture and muon decay-in-orbit (DIO). The ratio between the two processes is also dependent on the stopping target material. For aluminum, about 60.9% of the muonic atoms go through muon nuclear capture, i.e., $\mu^- N \to \nu_\mu N'^*$, where N'^* signifies an excited magnesium nucleus (^{27}Mg). The rest of the 39.1% undergoes muon DIO, i.e., the three-body decay with neutrinos $\mu^- N \to e^- N \nu_\mu \bar{\nu}_e$ [4].

The experiment, on the other hand, searches for a third decay mode, the neutrinoless muon-to-electron conversion $\mu^- N \to e^- N$. The detectors search for a signature of a single monoenergetic electron, which has an energy of

$$\begin{aligned} E_{\mu e} &= m_\mu c^2 - (\text{B.E.})_{1s} - E_{\text{recoil}} \\ &= 104.97 \text{ MeV} . \end{aligned} \tag{2.1}$$

Here m_μ is the rest mass of a muon, $(\text{B.E.})_{1s}$ is the muonic binding energy at $1s$ state, and E_{recoil} is the energy of the nuclear recoil [6]. Ultimately, the Mu2e experiment measures the ratio of neutrinoless muon-to-electron conversions to all muon captures [6], denoted as $R_{\mu e}$:[1]

$$R_{\mu e} = \frac{\Gamma[\mu^- + A(Z,N) \to e^- + A(Z,N)]}{\Gamma[\mu^- + A(Z,N) \to \nu_\mu + A(Z-1,N+1)]} . \tag{2.2}$$

In the case of Mu2e, $A(Z,N)$ is $^{27}_{13}$Al and $A(Z-1,N+1)$ is $^{27}_{12}$Mg. The previous best limit set by SINDRUM II, on the other hand, was obtained in muonic gold. At 90% confidence level, $R_{\mu e}(\mu^- \text{Au} \to e^- \text{Au}) < 7 \times 10^{-13}$ [1]. The Mu2e experiment expects to reach a limit

[1]While the equation gives a definition of $R_{\mu e}$, in the experiment the decay widths are not directly measured. What can be measured are the number of observed ~ 105 MeV electrons (backgrounds included) and the spectrum of x-ray photons emitted during muonic atom formations. The decay widths are then calculated based on thorough understandings of the physics processes and the detector performances (e.g., the detector acceptances and efficiencies). So it is crucial to know the detectors well, in order to get an accurate evaluation of $R_{\mu e}$.

of $R_{\mu e}(\mu^- \mathrm{Al} \to e^- \mathrm{Al}) < 8 \times 10^{-17}$ (90% CL), which improves the previous limit by a factor of $\sim 10^4$ [7].

2.2 Backgrounds and Mitigations

To achieve the four orders of magnitude improvement in sensitivity, it is useful to understand various experiment backgrounds in $\mu^- N \to e^- N$ conversion searches, and then discuss their mitigations in the Mu2e context. In general, anything that produces $105\ \mathrm{MeV}$ electrons can introduce a background. There are five categories of background sources [6], which are:

1. Intrinsic backgrounds;
2. Prompt processes from an incoming beam particle;
3. Delayed processes from slow-moving particles (antiprotons) down the muon beamline;
4. Backgrounds induced by cosmic rays;
5. Reconstruction errors due to accidental activities.

These background sources are discussed in detail below, along with their respective mitigation methods. The applications of these mitigations in the Mu2e experimental design will be introduced in the following section focusing on the experiment setup.

2.2.1 Intrinsic Backgrounds

The two physics processes that can contribute to the intrinsic backgrounds are muon DIO and RMC. The word "intrinsic" here means the rates of these backgrounds are proportional to the number of muons captured on the stopping target.

Muon Decay-in-Orbit

The energy of electrons coming from the 3-body decay $\mu^- \to e^- \nu_\mu \bar{\nu}_e$ of a free muon is described by the Michel spectrum. The differential decay rate is the following [8]:

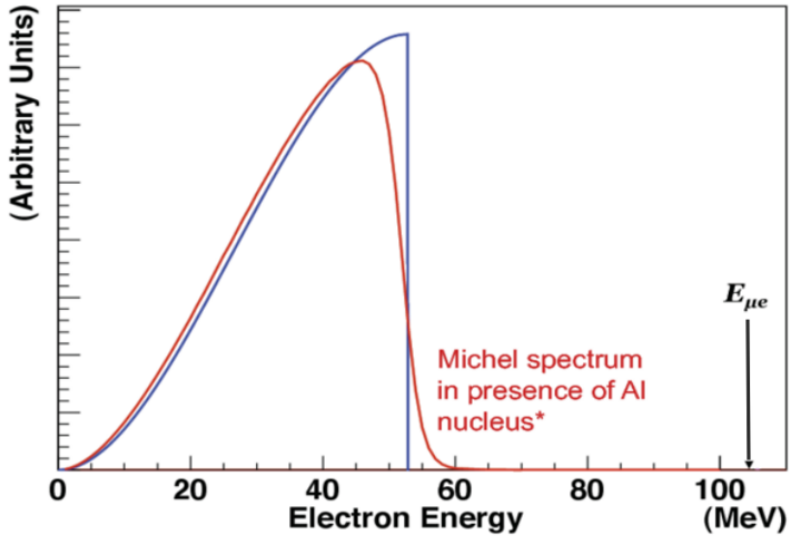
$$\frac{\mathrm{d}\Gamma_{\mathrm{free}}}{\mathrm{d}x} = \frac{G_F^2 m_\mu^5}{192\pi^3} x^2 (6 - 4x + \frac{\alpha}{\pi} f(x)) \,, \tag{2.3}$$

$$x = \frac{2E_e}{m_\mu}, 0 < x \leq 1 \,. \tag{2.4}$$

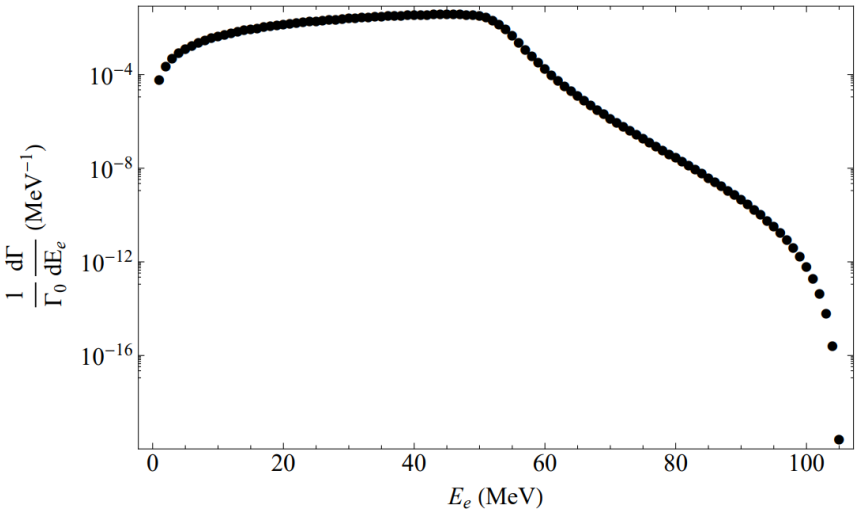
Here G_F is the Fermi constant, α is the fine-structure constant, m_μ is the muon mass, and E_e is the produced electron energy. $f(x)$ represents a complicated radiative correction term that is described in [9]. As shown by the blue line in Figure 2.2(a), the spectrum has a maximum energy at $52.8\ \mathrm{MeV}$, which is half of the muon rest energy, or the energy of the searched conversion electron.

However, when there is an Al nucleus present, the electron can exchange momentum with the nucleus, resulting in a maximum possible energy ("endpoint energy") that is approximately the same as the conversion electron (neutrino energy negligible is this case). The energy spectrum in presence of the Al nucleus is represented by the red line in Figure 2.2(a) [10, 11]. $E_{\mu e}$ in the figure is the conversion energy, which is also the endpoint of the spectrum.

Figure 2.2(b) [10] provides a closer look at the high-energy region of the spectrum. For



(a)



(b)

Figure 2.2: Michel spectrum of muon decay-in-orbit: (a) Michel Spectrum of free muon and that in presence of an Al nucleus; (b) a closer look at the Michel spectrum in log scale. The part close to the endpoint can be approximated by $(E_{\mu e} - E_e)^5$ [10].

electron energy $E_e > 85\ \mathrm{MeV}$, the leading order term scales with $(E_{\mu e} - E_e)^5$ [10, 12, 13], and the rate in the energy region very close to the endpoint is low. However, in the real experiment, due to uncertainties in reconstructions and particle energy loss when interacting with stopping target or detector materials, the monoenergetic conversion electrons would be reconstructed to have a left-skewed energy distribution with a full width at half maximum (FWHM) on the order of $1\ \mathrm{MeV}$ [6] (shown later in Figure 2.14). The high-energy electrons from muon DIO and the conversion electrons are then rendered indistinguishable, which become a source of background signals.

In order to reduce the level of backgrounds from muon DIO, the Mu2e detectors are designed to be blind to most of the low-energy electrons. A particle tracking detector with a good momentum resolution also contributes to the background reduction, which will be discussed in more details when introducing the experiment setup.

Radiative Muon Capture

Radiative muon capture (RMC) process produces an additional photon at tree level compared to the ordinary muon nuclear capture. The photon in $\mu^- N \to \gamma \nu_\mu N'^*$ can be real or virtual [6]. By interacting with matter or through pair production, it can produce electrons with energies close to the conversion electron energy, introducing backgrounds to the experiment. The emitted photon energy follows a spectrum; the maximum possible energy of the photon, the kinematic limit $k_{\max}$ is calculated by the following equation [6]:

$$k_{\max} = m_\mu c^2 - |E_{\mathrm{b}}| - E_{\mathrm{recoil}} - \Delta M\ , \tag{2.5}$$

where E_{b} is the muon binding energy on the initial nucleus, E_{recoil} is the recoil of the daughter nucleus, and ΔM is the rest energy difference between the final and the initial nuclides.

Our understanding of the RMC spectrum at high photon energy is relatively poor; the theoretical calculations need improvement and the spectrum in the interested region is poorly measured [14, 15]. Although the measurements from the TRIUMF radiative muon capture program reported that the endpoint photon energies extrapolated from data on all targets were a few MeV lower than the kinematic limits [14], which could mean RMC background events have lower energies and thus are insignificant, it is worth noting that the statistics were low in the concerned region in these measurements, and only real photons were measured. In addition, an excess in SINDRUM II positron spectrum indicated possible extra radiative muon capture contributions than the current estimation [1, 16].

Despite the large uncertainties in the RMC background, it can still be rendered negligible in the Mu2e experiment. With the selection of aluminum as the stopping target material, the

daughter nuclide of a muon-capture process of any kind is heavier than the original nuclide. The kinematic limit of the aluminum RMC energy is $101.9\ \mathrm{MeV}$ [6]. Given the planned FWHM of the conversion peak being $\sim 1\ \mathrm{MeV}$, the RMC background will be outside the signal region to search for conversion electrons. That being said, the RMC background can distort the DIO spectrum in the $80 \sim 100\ \mathrm{MeV}$ region, and bring difficulties to extrapolate the DIO spectrum to its endpoint [6, 15]. Determining the RMC background from the data will be an important task for the experiment.

2.2.2 Prompt Processes

This category of background sources can produce electrons nearly coincident in time with the incoming beam particles. Four processes will be discussed, which are radiative pion capture (RPC), pion decay-in-flight (π-DIF), muon decay-in-flight (μ-DIF), and beam electrons.

Radiative Pion Capture

The secondary muon beam contains significant amount of pions. Some pions may live long enough to reach the stopping target. These pions, when captured in the stopping target or other detector materials, can produce a high-energy photon, i.e. $\pi^- N \to \gamma N'^*$. This process is referred to as radiative pion capture (RPC), and occurs to $\sim 2\%$ of pions captured in Al [17]. Similar to RMC, the photon can convert internally to electron-positron pair, or be radiated as an on-shell photon which then undergoes pair production. The external pair-production is also dependent on the amount of materials available to induce it; either way, the electrons produced can contribute to the experiment background.

Despite the resemblance to RMC, RPC is more problematic, since the photon (and hence the produced electron) energy spectrum endpoint is no longer limited by the muon rest energy. The rest energy of a pion is $139.6\ \mathrm{MeV}$, which is much higher than the conversion energy, so there is no energy separation between the search window for conversion electrons and the energy spectrum of electrons originated from RPC photons.

The SINDRUM II experiment results were limited by the pion-induced background (along with the low intensity of its muon beam) [1]. Following a similar experiment concept, it also faced challenges from RPC background. The primary proton beam SINDRUM II used came every $19.75\ \mathrm{ns}$ for $\sim 0.3\ \mathrm{ns}$; the time interval between pulses is shorter than the pion lifetime of $26\ \mathrm{ns}$ [3], so it faced a mostly constant pion flux. SINDRUM II used a degrader to remove pions, and a veto counter in the beam to suppress RPC, which reduced the number of pions reaching their target to less than 1 out of 10^9. However, with a much more intense beam, the Mu2e experiment cannot continue to use the SINDRUM II method; the level of pion suppression these techniques provide cannot meet Mu2e's requirement.

Instead, a pulsed proton beam is used in Mu2e to solve this problem. Due to their short lifetime, virtually all pions decay or interact with materials within a short time period after the proton beam pulse. Meanwhile, the muon lifetime is more than an order of magnitude larger than the pion lifetime; by opening a live window for conversion electron search at a right time, the RPC background can be suppressed without losing much data. One caveat here, is that if there are protons coming out of the beam pulse ("out-of-time"), the resultant pions may still cause RPC background (which also apply for some other prompt background processes). As a result, the pulsed beam needs to reach a good extinction level, which means the ratio between out-of-time to in-time protons should be controlled under a certain level. More details of the pulsed beam used for Mu2e will be discussed in Section 2.3.

Pion Decay-in-Flight and Muon Decay-in-Flight

Pion decay-in-flight (π-DIF) and muon decay-in-flight (μ-DIF) are rather similar, so they are discussed together here. Free pions and muons can decay to electrons when moving from the production target to the stopping target, through the processes $\mu^- \rightarrow e^- \nu_\mu \bar{\nu}_e$ and $\pi^- \rightarrow e^- \bar{\nu}_e$. Within the center-of-mass frame of the initial particles, electrons from the former process follow an energy spectrum whose endpoint is $52.8\ \mathrm{MeV}$, and those from the latter process are monoenergetic with an energy of $\sim 70\ \mathrm{MeV}$. Since these pions and muons move at relativistic speeds, energies (and momenta) of the daughter electrons are boosted. For example, a muon with $\sim 79\ \mathrm{MeV/c}$ momentum or a pion with $\sim 70\ \mathrm{MeV/c}$ momentum can each produce an electron with $105\ \mathrm{MeV}$ [6]. As a result, such processes can create background signals for the experiment.

The usage of a pulsed proton beam and a delayed live window can mitigate π-DIF and μ-DIF backgrounds. The particles with enough momentum to boost the daughter electrons to the concerning energies move fast down the muon beamline, and are gone by the time when live search window begins [15]. In addition, the geometry of the Mu2e detectors further aid the mitigation.

Beam Electrons

There are other processes that produce electrons, both at the production target and in the muon beamline. For example, neutral pions produced at the production target can decay through $\pi^0 \rightarrow \gamma\gamma$, and subsequently the photons can pair-produce or interact with surrounding materials to produce electrons. Other beam particles can also decay or interact anywhere upstream of the muon stopping target, and produce electrons with the same energy as conversion electrons. These additional backgrounds are referred to as beam electrons.

The combination of pulsed beam and delayed live window in time again reduces the background from beam electrons. In addition, Mu2e deploys a system of solenoids that creates

a magnetic field across the muon beamline and in the region where the stopping target and the detectors are located. A charged particle moves in a helical trajectory in the magnetic field; the radius and the pitch of the helix are determined by particle's charge and momentum (more specifically, the momentum components along and perpendicular to the direction of the magnetic field). This makes it possible to install collimators in the muon beamline that suppresses the transport of high-momentum particles above $100\ \mathrm{MeV/c}$ [6]. In addition, the magnetic field in the region of the stopping target is specifically designed with a gradient, which separates trajectories of conversion electrons from those of electrons coming from upstream, provided the electrons are not scattered in the stopping target [6, 15]. The system of solenoids and the magnetic field will be discussed in more details in Section 2.3.

2.2.3 Delayed Processes from Antiprotons

When protons in the primary beam have sufficient energies, it is possible to produce antiprotons in the production target. In the antiproton production process $pp \rightarrow ppp\bar{p}$, a proton-antiproton pair is actually produced. The minimum kinetic energy of the initial beam proton can be calculated by applying 4-momentum conservation to the system. Assuming the 4 particles in the end state are all stationary in the center of mass frame, the minimal kinetic energy required for $p\bar{p}$ production is calculated to be $6m_pc^2 \approx 5.6\ \mathrm{GeV}$, where m_p is the proton rest mass. Ideally, if the beam used for the experiment has energy lower than the threshold, this background can be avoided altogether. Unfortunately, the Mu2e proton beam is beyond the antiproton production threshold and associated background needs to be considered.

The antiprotons are long-lived and massive. Different from the particles in the last category, antiprotons with momenta lower than $100\ \mathrm{MeV/c}$ move at speeds less than $0.1c$, and can take up to a few microseconds to spiral from the production target to the stopping target [6]. They have the right sign and amount of charge, as well as the right momentum to pass the aforementioned collimators between the production target and the stopping target. When they annihilate or undergo other interactions with materials, they can release significant amount of energy, and produce many secondary, even tertiary electrons. The delayed timing of such interactions is much longer than the muon lifetime, which means there is a continuous flux of antiprotons at the stopping target. The approach of using a pulsed beam along with a delayed live window cannot mitigate the backgrounds from these antiprotons.

In order to reduce the antiproton-induced backgrounds, the most effective approach is to prevent the antiprotons from reaching the region where the stopping target is located. A thin absorber is placed in the muon beamline for capturing the antiprotons. Its design is optimized, so that a balance is achieved between maximizing antiproton absorption and minimizing muon beam loss. [6, 15].

2.2.4 Cosmic-Ray-Induced Backgrounds

Cosmic-ray particles, primarily comic-ray muons, are also capable of introducing backgrounds to the experiment. Again, any process that produces, or mimics $105\ \mathrm{MeV}$ electrons can induce backgrounds. Here are a few example mechanisms [6]:

- Muon decays in the detectors or around the stopping target;
- Muon interacts with nuclei in detectors or surrounding materials;
- Muon scatters in the detectors and is misidentified as electrons;
- Muon enters muon beamline with a special initial momentum, either interacts with collimator to produce an electron, or travels down the muon beamline to be misidentified...

While in some cases the cosmic-ray muons can be correctly identified with improved reconstruction algorithms, there are other cases where the cosmic-ray-induced signals are indistinguishable from that of a conversion electron. The number of cosmic-ray-induced backgrounds scales with the length of data-taking time, and the rate is specific to the experiment setup. It was estimated from a detailed simulation that in the Mu2e experiment, without any counter-measures, conversion-like signals induced by cosmic rays occur roughly once per day, which would flood the actual conversion signals [6, 18].

The mitigation is to use a mixture of active veto and shielding in the experiment. A high-efficiency veto, as will be shown later, is capable of suppress cosmic-ray-induced background to a sufficiently low level.

2.2.5 Accidental Activities

This last background category cannot be attributed to physical interactions of a specific particle. Instead, it comes from mistakenly and accidentally reconstructing additional activities in the detectors as conversion-like signals. As previously mentioned, the muon-capture process results in nuclei in excited states, which can eject protons, neutrons, and photons. The ejected protons are highly ionizing, and can leave large signals in the detectors, adding to the probability of reconstruction error. Other coincidences, such as multiple muon decay-in-orbit that occurs within a short period of time, can also cause reconstruction errors.

In order to reduce the flux of protons from the muon-capture process that reaches the detectors, additional absorbers made of polyethylene is added to surround the stopping target in the Mu2e experiment [15]. A systematic uncertainty is evaluated as a part of background to address these accidental activities.

* * *

Table 2.1: Summary of backgrounds and mitigations of muon-to-electron conversion experiments

Background	Mechanism	Mitigation
	Category: Intrinsic	
Muon DIO	Michel spectrum distorted by nuclear recoil	Improved detector design, high momentum resolution
RMC	$\mu^- N \to \gamma^{(*)} \nu_\mu N'^*$	Aluminum stopping target
	Category: Prompt processes	
RPC	$\pi^- N \to \gamma^{(*)} N'^*$	Pulsed beam, delayed live window, good beam extinction
π-DIF, μ-DIF	$\mu^- \to e^- \nu_\mu \bar{\nu}_e$, $\pi^- \to e^- \bar{\nu}_e$	Pulsed beam, detector geometry
Beam electrons	$\pi^0 \to \gamma\gamma$, etc.	Collimator, graded magnetic field
	Category: Delayed processes	
Delayed $\bar{p}$	$\bar{p}$ annihilation, various interactions with detector materials	Absorber in muon beamline
	Category: Cosmic-ray-induced	
Cosmic-ray μ's	Cosmic-ray muon decays, interactions, misidentifications	Active veto combined with shielding
	Category: Accidental activities	
Reconstruction errors	Protons from excited nucleus after muon capture	Absorbers around stopping target

Table 2.1 summarizes the sources of backgrounds in $\mu^- N \to e^- N$ searches, along with their respective mitigations in the Mu2e experiment. The next section will show how these mitigations are integrated into the experiment setup.

2.3 Experiment Setup

Figure 2.3 provides a schematic overview of the Mu2e experiment [6, 19]. The pulsed proton beam arrives at the production target as marked by the red arrow. A solenoid system produces magnetic fields for the experiment. The Production Solenoid (PS) surrounds the production target. Moving downstream, the Transport Solenoid (TS) provides the magnetic field for the muon beamline. There are collimators and proton absorbers placed along the S-shaped TS, reducing experiment backgrounds. The stopping target is located at the beginning of the Detector Solenoid (DS). There are proton absorbers around the stopping target, which are not shown in the figure. The DS also houses the Tracker and the Calorimeter for detecting the 105 MeV conversion electrons, providing momentum and energy measurements respectively. There is also a Stopping Target Monitor (STM) downstream at end of the DS that is not shown in the figure. It is used for monitoring the stopping target, and provides an estimation for the number of muons captured on the stopping target. The Mu2e Cosmic Ray Veto (CRV) system is not shown in the figure either; it surrounds the entire DS and half of TS. This section will describe various systems of the Mu2e experiment in more details, explaining how they facilitate the background mitigation methods previously discussed.

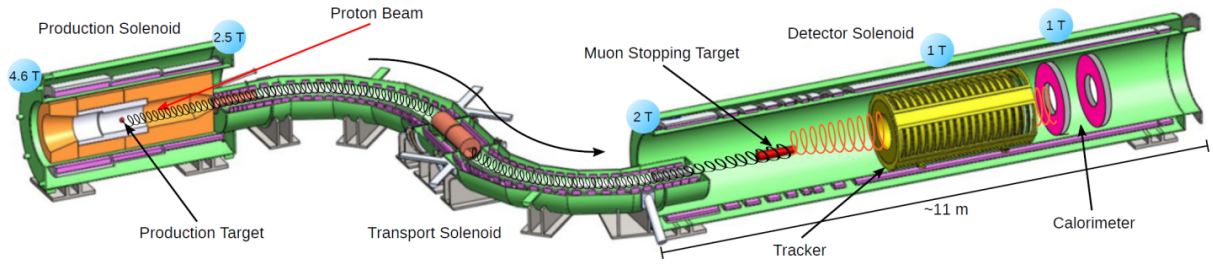


Figure 2.3: A schematic view of the Mu2e experiment [19]. Cosmic Ray Veto and Stopping Target Monitor are not shown.

2.3.1 Accelerator Systems and Pulsed Proton Beam

Pulsed Proton Beam

As discussed earlier, the Mu2e experiment uses a pulsed proton beam to reduce backgrounds from the prompt processes. This 8 GeV, 8 kW beam comes from the Fermilab Booster. In order to obtain a pulsed beam with 1695 ns between neighboring pulses, a series of beam manipulations are required [6, 20].

Figure 2.4(a) shows the Fermilab accelerator facilities involved in producing and delivering the pulsed proton beam [21]. The Fermilab Booster delivers 8 GeV protons in 20 batches over a

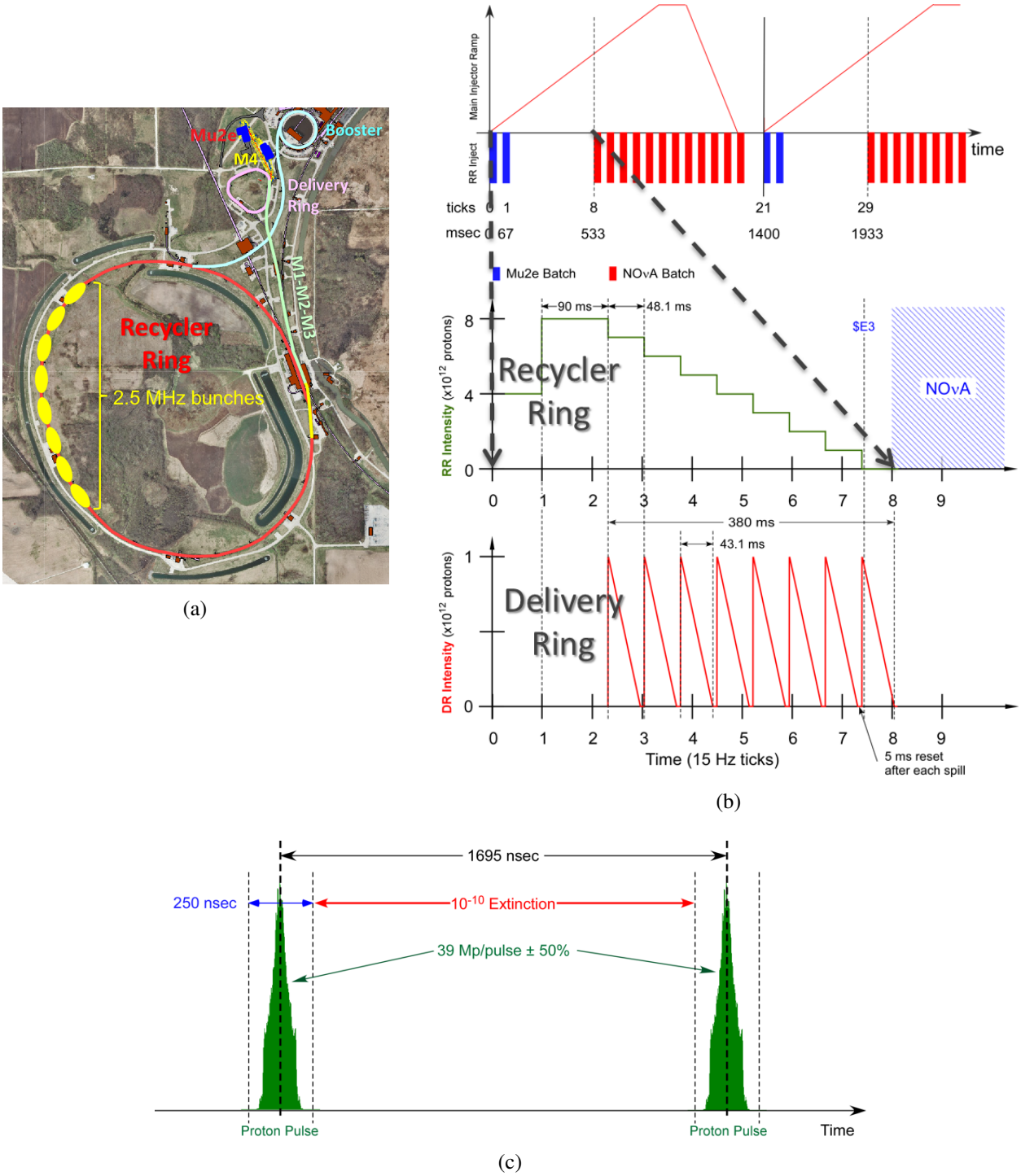


Figure 2.4: Pulsed proton beam delivery [21]: (a) Fermilab accelerator facilities involved in producing and delivering the pulsed proton beam; (b) proton beam delivery to Mu2e; (c) proton beam profile at the Mu2e Proton Target.

Main Injector cycle of 1.4 s at 15 Hz, shown in the top part of Figure 2.4(b) [21]. Henceforth, the accelerator timeline is discussed using a basic time unit of $1 \text{ tick} = 66.7$ ms.

The two Mu2e batches at tick 1 and 2 (shown as the two blue bars) are injected into the Recycler Ring, each containing 4×10^{12} protons. The protons from the two batches are then rebunched in the Recycler with a 2.5 MHz radio frequency (RF) system into 8 bunches. These bunches are extracted one by one from the Recycler and transported to the Delivery Ring every 48.1 ms (middle part of Figure 2.4(b)).

This single bunch of 1×10^{12} protons circulating the Delivery Ring then undergoes slow extraction. A resonance condition destabilizes part of the beam in a controlled manner.[2] Unstable particles stream away from the beam center, then deflected by a Electrostatic Septum (ESS). The ESS contains a thin plane of foils or wires, which separates the beam from a region with high electrostatic field used for deflection. For each turn, a small fraction of the bunch is peeled off and transported to the Mu2e experiment. The entire bunch is extracted over an interval of 43.1 ms, across ~ 25000 turns around the Delivery Ring (Figure 2.4(b) bottom).

Figure 2.4(c) shows the time structure of the beam at the Mu2e Proton Target [21]. Consecutive proton pulses are separated by 1695 ns, which is determined by the time structure of the proton beam in the Delivery Ring [20]. Each pulse lasts 250 ns, containing $3.9 \times 10^7 (\pm 50\%)$ protons. The beam pulse profile is represented in Figure 2.4(c).

The pulse separation of 1695 ns is extremely convenient for the purpose of Mu2e. Figure 2.5 shows the beam pulse together with the simulated pion flux, the muon capture rate on the stopping target, and the muon decay rate [22]. The live window for searching conversion electrons starts at 700 ns and lasts for 900 ns. This choice balances the need to minimize the pion-induced backgrounds (e.g. RPC) and that to include more muon decays. Due to the short lifetime of the pions, they are long gone by the time the live window opens. This allows a suppression of the number of pions

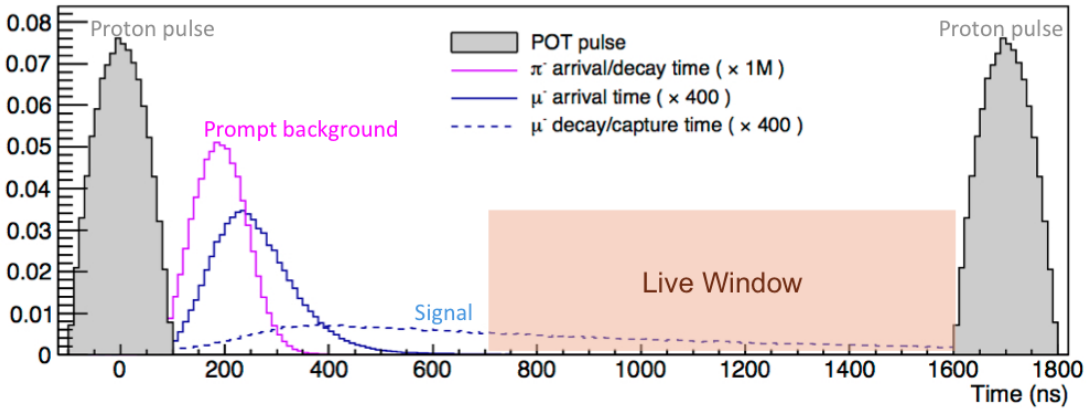


Figure 2.5: The pulsed proton beam timeline with the live search window. The simulated pion flux, the muon capture rate on the stopping target, and the muon decay rate are also shown. The live search window starts from 700 ns [22].

[2] Slow extraction is also sometimes referred to as resonant extraction.

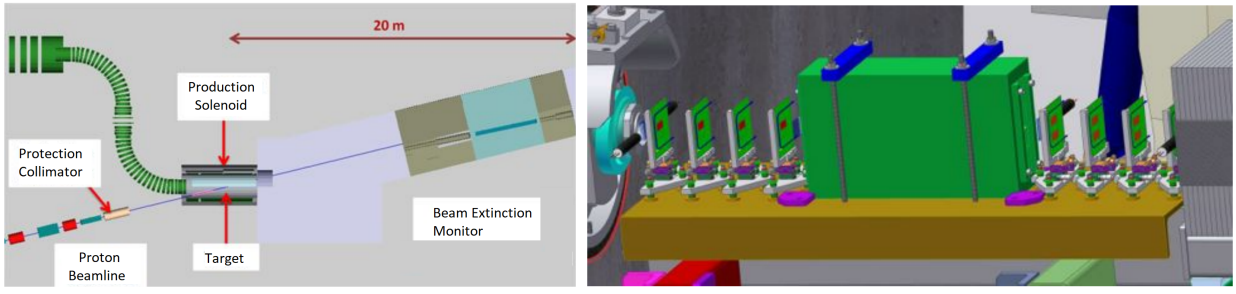
by $\mathcal{O}(10^{11})$ [15]. The pulse separation of 1695 ns is almost twice as long as the muon lifetime. It allows sufficient separation in time between the prompt backgrounds and live window, while not being too long and costing the beam intensity.

Proton Beam Extinction and Extinction Monitor

To reduce backgrounds caused by out-of-time protons (protons arriving out of the 250 ns proton pulse window), the Mu2e experiment has a requirement on the extinction level of the incoming proton beam, as mentioned in the previous section. The extinction rate, defined as the ratio between the number of out-of-time protons to that of the in-time protons, needs to be lower than 10^{-10} [6].

The RF systems used for generating the bunch structure in the Recycler Ring naturally lead to an extinction level 2.1×10^{-5}, which should be preserved in the beam transfers. To fill in the gap and to take into account the fact that some beam will leak out of the RF bucket during slow extraction, an Extinction Insert is deployed in the M4 beamline between the Delivery Ring and the Mu2e experiment. By using a set of oscillating dipoles ("AC dipoles"), the out-of-time beam particles are swept into a system of collimators. These AC dipoles are simulated to provide an additional extinction factor of 5×10^{-8}, bringing the total extinction down to 1.1×10^{-12}, leaving a margin of 10^2 [21].

An Extinction Monitor is placed further downstream of the production target along the proton beamline, as shown in Figure 2.6(a) [6, 23]. It monitors the extinction level of the incoming beam hitting the Mu2e production target, and provides a measurement at a level better than 10^{-10}. The Extinction Monitor consists of a collimator and magnetic filter system, a pixel telescope, a system of trigger scintillators, and a range-stack. The collimator and magnetic filter system transports a small fraction of the particles created at the production target to the Extinction Monitor. The pixel telescope (Figure 2.6(b) [23]) measures the trajectory of the charged particles coming



(a) (b)

Figure 2.6: The beam Extinction Monitor [23]: (a) the Extinction Monitor is located downstream of the production target; (b) the pixel telescope consists of a permanent magnet and 8 scintillators.

from the collimator. With a momentum-resolving permanent magnet separating two sets of 4 planes of scintillators, it also provides momentum measurements of the incoming particles. The range-stack sits further downstream to the pixel telescope. With layers of steel absorber plates separating layers of scintillators, it distinguishes between hadrons and muons through their penetrating capabilities.

Production Target

The Mu2e production target is another important component of the accelerator systems. It is suspended in the center of the PS bore [6]. It is made of tungsten, which provides a large enough pion production cross-section to produce the required number of stopped muons. The refractory material allows the target to be cooled radiatively without the assistance of additional cooling systems. A compact design minimizes reabsorption of pions on the production target.

2.3.2 Solenoid Systems

The Mu2e solenoid Systems consists of three magnetically coupled systems: the Production Solenoid (PS), the Transport Solenoid (TS), and the Detector Solenoid (DS). Each of the system contains multiple solenoid modules. A typical solenoid module is shown in Figure 2.7(a) [24]. The modules are made of superconducting coils wound with aluminum stabilized Nb-Ti Rutherford cables (Figure 2.7(b) [25]).

The solenoids produce a strong magnetic field of $\mathcal{O}(1\ \mathrm{T})$. The specific values at selected locations along the solenoids are marked in Figure 2.3, which are purposefully and carefully designed, so that μ^-'s are efficiently transmitted while other particles are suppressed. The magnetic



(a) (b)

Figure 2.7: Solenoid module and superconducting cable: (a) a solenoid module in the TS [24]; (b) a view of the solenoid superconducting cable cross-section [25].

field value is at its highest of 4.6 T at the upstream end of the experiment; the lowest value is 1T at the downstream end. This general trend of lowering magnetic field in the downstream direction helps to direct the muons towards the DS and the stopping target; local minima in the magnetic field are avoided to prevent trapping particles in these locations.

More specifically, in the Production Solenoid, the magnetic field has a large gradient from 4.6 T to 2.5 T at the entrance of the TS. The large gradient helps to collect the secondary pions and muons, and send them towards the DS.

The gradient across the S-shaped TS is a relatively small; the field values at the two ends only differs by 0.5 T. Nonetheless, the TS serves several purposes [15]. First, there is no line of sight across the TS, which means the photons and other neutral particles produced at the production target cannot directly enter the DS and interact with the detectors. The dimensions of the TS prevents particles with high momentum from being transmitted. These particles either spiral with too large a helical radius (large initial momenta perpendicular to the field) or cannot make the S-shaped bend (large initial momenta parallel to the magnetic field), and would hit solenoid walls or collimators. In addition, as particle drifts in the solenoid field is dependent on the particle charge as discussed in Jackson [26], positively and negatively charged muons drift in opposite vertical directions and are separated; a collimator in the central part of the TS (brown cylindrical objects in TS in Figure 2.3) blocks the unwanted μ^+'s. The TS provides sufficient length for pions to decay, which suppresses RPC backgrounds. It also keeps the detectors away from radioactively hot areas around the primary beam and the production target.

The magnetic field in the first half of the DS changes from 2 T to 1 T. In such a gently gradient field, the adiabatic invariance of the magnetic flux can be applied, and assuming a constant $p_\perp^2/B$, there is [26]:

$$v_\parallel^2 = v_0^2 - v_{\perp 0}^2 \frac{B(z)}{B_0} \,. \tag{2.6}$$

Here, $p_\perp^2$ and $v_\perp^2$ denotes the momentum and velocity perpendicular to the magnetic field B, $v_\parallel$ is the velocity parallel to the magnetic field B along the downstream $+z$ direction, and the subscript 0's in $v_{\perp 0}^2$ and B_0 indicate these are values in the initial state. From this relation, when a charged particle spirals down a decreasing magnetic field, its transverse momentum $p_\perp$ is traded for momentum along the field direction $p_\parallel$, changing the pitch of the helical trajectory. This separates beam electrons from conversion electrons by trajectory pitches, as mentioned in Section 2.2. In addition, the conversion electrons are pushed towards the detectors downstream. 105 MeV conversion electrons emitted between $90^\circ \pm 30^\circ$ with respect to the solenoid axis ($p_\perp > 90$ MeV/c) are projected forward, and pitched to have reconstructible trajectories [6].

The second half of the DS containing the Tracker and the Calorimeter has a uniform magnetic field of 1 T. This allows particle trajectories and momenta to be reliably measured.

2.3.3 *Muon Beamline*

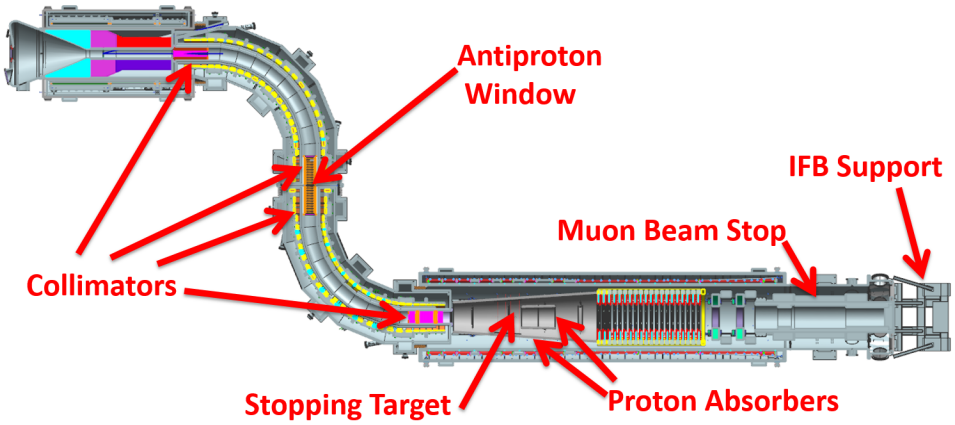


Figure 2.8: A closer look at the Mu2e muon beamline [27].

Figure 2.8 provides a closer view of the Mu2e muon beamline [27]. The Mu2e experiment has a high-intensity muon beamline. It is estimated that there are ~ 0.002 stopped negative muons per 8 GeV proton on target, and the stopped muon rate at the stopping target is 10^{10} Hz [6]. The high-intensity muon beamline is required to reach the measurement sensitivity of the experiment.

Collimators and Absorbers

A total of four collimators are placed along the muon beamline; two of them are stationary and placed at each end of the TS, while the other two are rotatable and placed in the mid-section [6]. All collimators filter the beam and select particles with desired momenta. They are designed to maximize pion transfer and suppress other particles, so certain backgrounds can be reduced as Section 2.2 discussed. Additionally, the collimator at the entrance of the TS protects the first few upstream TS coils from the high-level radiation produced at the stopping target. The two collimators at the center perform sign selection using particle drifts as previously discussed. The rotatable feature allows selection of μ^+'s instead of μ^-'s, which can be used for detector calibration. The collimator at the exit of the TS also shields the detectors from neutron backgrounds.

A thin antiproton stopping window is installed between the two rotatable collimators. It absorbs antiprotons from the production target and reduces the related background described in Section 2.2. An Inner Proton Absorber (IPA) and an Outer Proton Absorber (OPA) are placed near the muon stopping target. They are made of polyethylene. As discussed, by absorbing protons released in muon captures, they effectively reduce accidental activities seen in reconstruction. The IPA is designed to have a thin wall of 0.5 mm, in order to minimize energy losses of conversion electrons passing through.

Muon Stopping Target

The Mu2e muon stopping target consists of 37 annular aluminum foils of $> 99.99\%$ purity [15]. These foils are $100\ \mu\mathrm{m}$ thick, which ensures low energy loss when conversion electrons exit the stopping target. This optimized design helps to narrow the reconstructed conversion electron momentum distribution, and increases its separation from the momentum distribution of the DIO electrons. The annular shape reduces the level of interactions with the beam electron flash, which can give sprays of particles and cause damage to the inner portions of the Mu2e Tracker. The central hole does not influence target's ability to stop muons, as they move in helical trajectories. Muons passing through the hole of an upstream foil will stop in a downstream layer.

The selection of aluminum as the stopping target material includes several considerations. First, as discussed in Section 2.2, aluminum target has negligible RMC background, due to the massive capture daughter nuclide. Second, the muonic Al atom has a relatively long lifetime. A nucleus of a heavier element contains more protons for undergoing muon nuclear capture. It increases the ratio of capture to muon decay-in-orbit, and significantly reduces the lifetime of the corresponding muonic atom, as shown in Figure 2.9(a) [5, 28]. This is problematic, as there is more overlap in time between the prompt backgrounds and the muon decays/captures. The signal live window would either be flooded by prompt backgrounds (if it starts early) or contain very few muon decays/captures (if it waits until prompt backgrounds diminishes). In this regard, a target of light element (like Al) is preferred. The long lifetime allows separation between prompt backgrounds and a live window with a good decay rate. In addition, the muon DIO endpoint energy also depends on the type of nucleus, as shown in Figure 2.9(b) [10–13, 28, 29]. Aluminum has a relatively high endpoint energy, which means when muons are captured on other detector materials with a higher atomic number Z, their DIO electrons would have lower endpoint energies, and would not be a source of background.

Aluminum is a suitable stopping target material for muon-to-electron conversion searches. However, measurements from Al target alone cannot differentiate between possible models that give rise to the conversion. If the process is dominated by a single short-distance operator, the branching ratio (BR) of the conversion varies when stopping target material differs, due to changes in both atomic number Z and mass number A. The ratios between BRs on different nuclei reveal the type of dominating operator [30, 31]. Figure 2.9(c) [30] shows the BRs on different target materials normalized to the value in aluminum. Four operator types, scalar(S), dipole(D), vector of transition charge radius type ($V^{(\gamma)}$), and vector of effective Z-penguin type ($V^{(Z)}$), are represented by various lines in the figure. Despite the difficulty to separate prompt backgrounds from signals due to the short lifetimes of corresponding muonic atoms, materials of higher atomic number Z have better model-differentiation power. It is suggested that if Mu2e observes conversion signals, a subsequent search ("Mu2e-II") can use titanium for the stopping target. It provides some power to differentiate

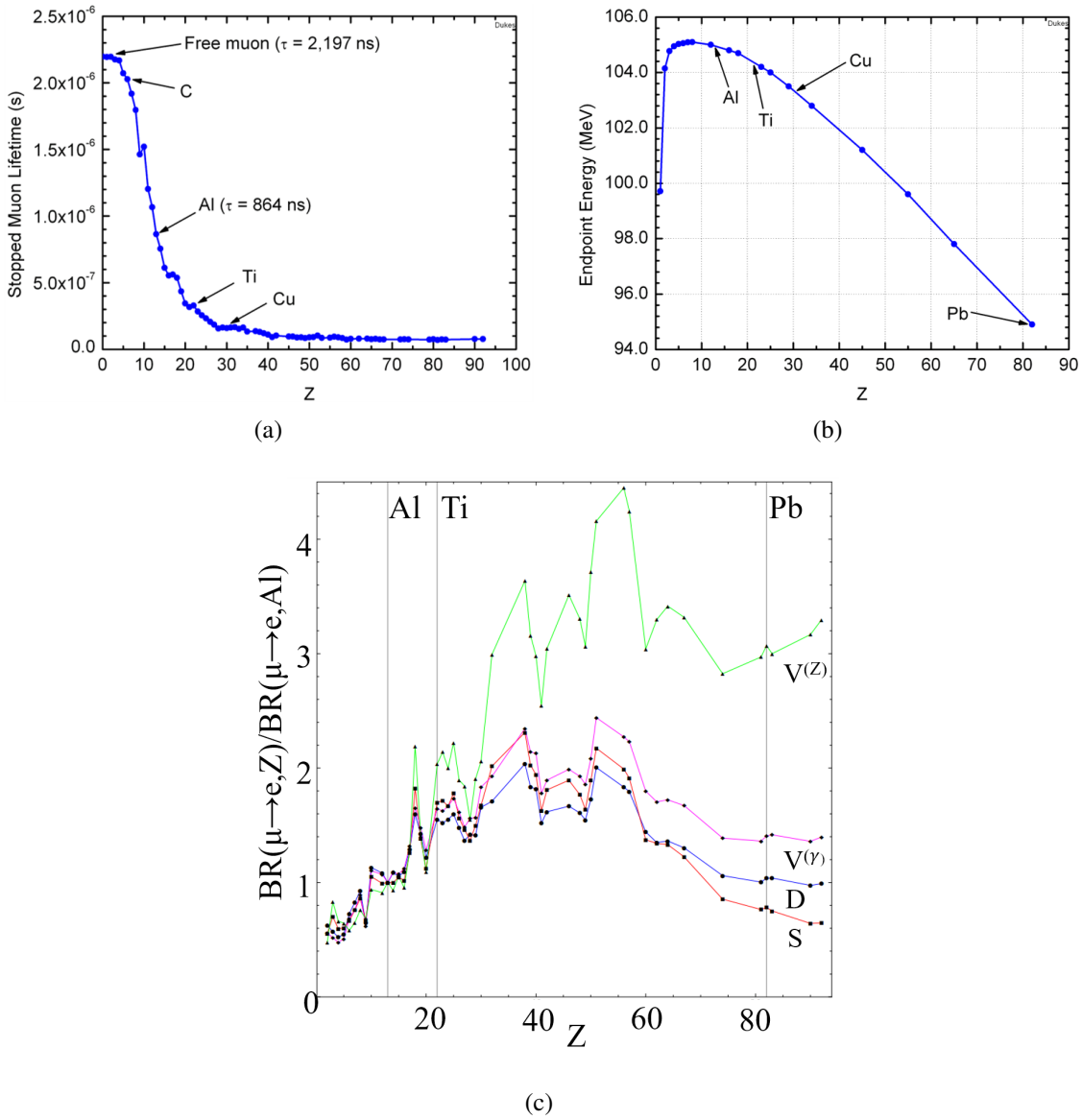


Figure 2.9: Dependence of muonic atom properties on the atomic number: the dependence of (a) stopped muon lifetime [28], (b) electron energy spectrum endpoint from muon DIO [28], and (c) relative conversion branching ratio calculated by various theories [30] on the atomic number Z are shown here.

models, while the prompt backgrounds can still be separated from the conversion signals [32].

Muon Beam Stop and Stopping Target Monitor

A Muon Beam Stop (MBS) is placed at the downstream end of the DS to absorb muons that are not stopped in the stopping target [6]. The MBS is designed to reduce influences of muon decays and captures inside the stop. The magnetic field gradient keeps most of the low-energy charged particles produced in the MBS from going upstream. It is constructed by high-Z materials

and polyethylene. Muons have a very short lifetime in high-Z materials as shown in Figure 2.9(a), which enables the activities to take place before the live window starting at $700\ \mathrm{ns}$. On the other hand, polyethylene absorbs protons and neutrons released by the excited nuclei produced in muon captures.

A STM is deployed further downstream of the MBS to measure the number of muons captured on the stopping target, which is then converted to the denominator in calculating the conversion-to-capture ratio $R_{\mu e}$. As mentioned, Mu2e measures the x-ray spectrum emitted when muons are stopped on the stopping target and fall to the ground state. More specifically, the experiment uses the 357-keV x-ray photons produced when muons cascade from the $2p \rightarrow 1s$ state [15], and plans to measure the number of stopped muons to an accuracy of 10% [6].

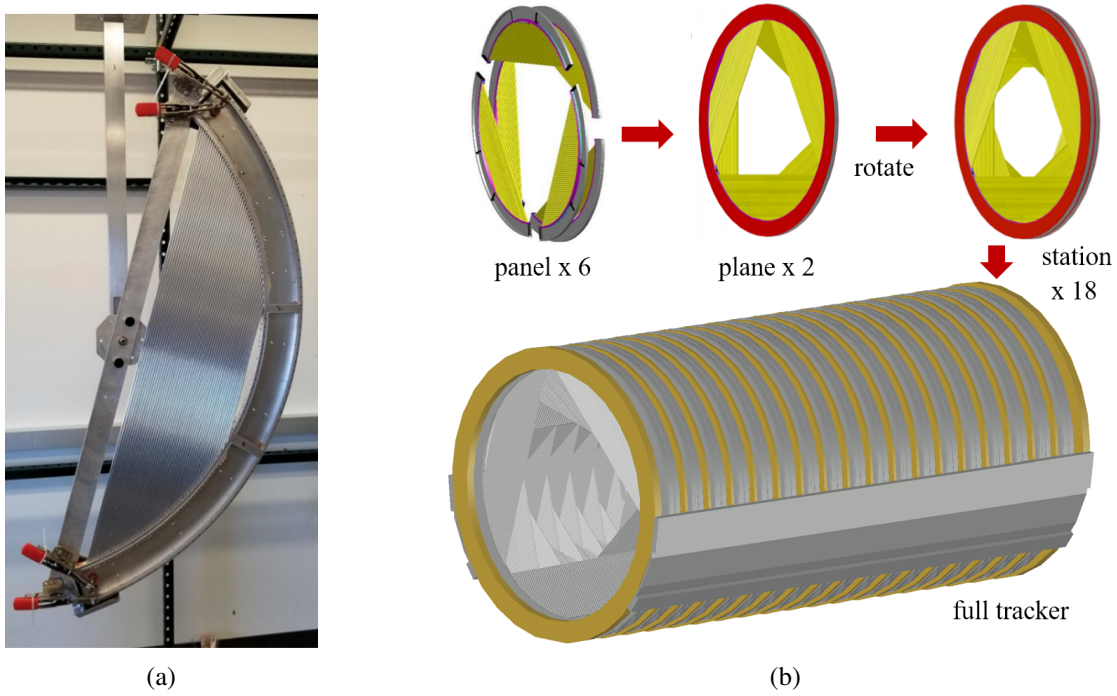
The Mu2e STM uses the standard x-ray detection device of a high purity germanium (HPGe) solid-state detector [6]. It is enclosed in a concrete housing, so the device can be shielded. It views the stopping target through a stainless steel pipe that is $10\ \mathrm{cm}$ in diameter and $3\ \mathrm{m}$ in length, and tunnels through the MBS near the DS axis. Collimators inside the pipe gives the STM a full view of the stopping target, and limits the view of other parts, such as the downstream collimator in the TS. HPGe devices are not fast enough to perform an event-by-event count. In addition, germanium lattice is susceptible to radiation damage. To make both radiation damage and rates manageable, the STM is about $35\ \mathrm{m}$ from the stopping target and is heavily shielded [15].

An alternative approach measures the delayed photons released during the beta decay of $^{27}\mathrm{Mg}$, which happens in 13% of muon captures [15]. The excited $^{27}\mathrm{Mg}$ decays to an excited state of $^{27}\mathrm{Al}$, with a half-life of $9.5\ \mathrm{min}$, which then emits an $844\ \mathrm{keV}$ photon within picoseconds, and can be detected. This method can use the macrostructure of the pulsed beam (the $1.4\ \mathrm{s}$ cycle) and detects only when proton batches are not sent to Mu2e. The rate can be manageable, and a shutter can block the pipe during beam time to avoid radiation damage to the HPGe detector. Again, this is not an event-by-event count; the process occurs at 13% of time and an average over the $9.5\ \mathrm{min}$ life-time needs to be taken.

2.3.4 Straw Tube Tracker

The experiment goes through great length to eliminates backgrounds to the $105\ \mathrm{MeV}$ conversion electrons. It is equally important to measure these conversion electrons. The Mu2e Straw Tube Tracker is located inside the DS, downstream of the stopping target. It tracks the helical trajectories of the conversion electrons in the magnetic field and provides momentum measurements.

The Mu2e Tracker uses gas drift tubes ("straw tubes") to detect particles. 96 straw tubes are arranged in two staggered layers to form a panel [6], which is the basic standalone module both mechanically and electrically, as shown in Figure 2.10(a). Each harp-shaped panel spans 120°; the straw tubes are placed like chords of the harp, and the electronics are housed in the outer



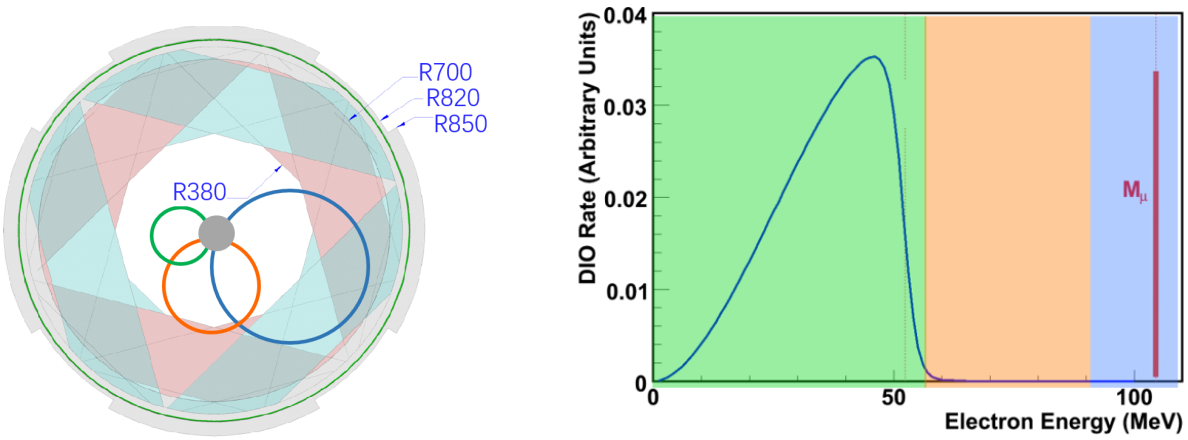
(a) (b)

Figure 2.10: The Mu2e straw tube Tracker: (a) a Tracker panel; (b) the Mu2e straw tube Tracker consists of 216 panels [6, 33].

volume. The Mu2e Tracker has very low material budget especially in the straw region, which minimizes chances of scattering the conversion electrons, and improves the momentum resolution of the device. More detailed discussions of the straw tube Tracker panel regarding its detection mechanism, mechanical design, front-end electronics, firmware and data acquisition (DAQ) system, and testing will be carried out in the following chapters.

The Tracker panels are assembled into the full Tracker following the scheme shown in Figure 2.10(b) [6, 33]. Three panels complete a full circle, and two layers of 6 panels, rotated by 30° between the layers, form a plane. Two planes, again rotated by 60°, make a station. Through the rotations, the entire annular region gets coverage by one or more panels. The whole Tracker consists of 18 identical stations. It is 3.2 m long and has a diameter of 1.7 m [6].

When viewing the Tracker along the direction of the beam axis, the Tracker straws are only in the active region with radius r in the range of $380 \text{ mm} < r < 700 \text{ mm}$, as shown in Figure 2.11(a) [33]. The central part of the Tracker is not instrumented. A similar annular design can also be found in the Mu2e Calorimeter. Such detector geometry is purposefully applied to reduce the level of backgrounds. First, the detectors are blind to the muon beam and most of the associated activities such as π-DIF, μ-DIF, beam electrons and other particle released in muon nuclear captures (a.k.a., the "beam flash"). More importantly, the annular detectors are not triggered by low-energy electrons produced in muon DIO. As shown in Figure 2.11(a), the helical trajectories



(a) (b)

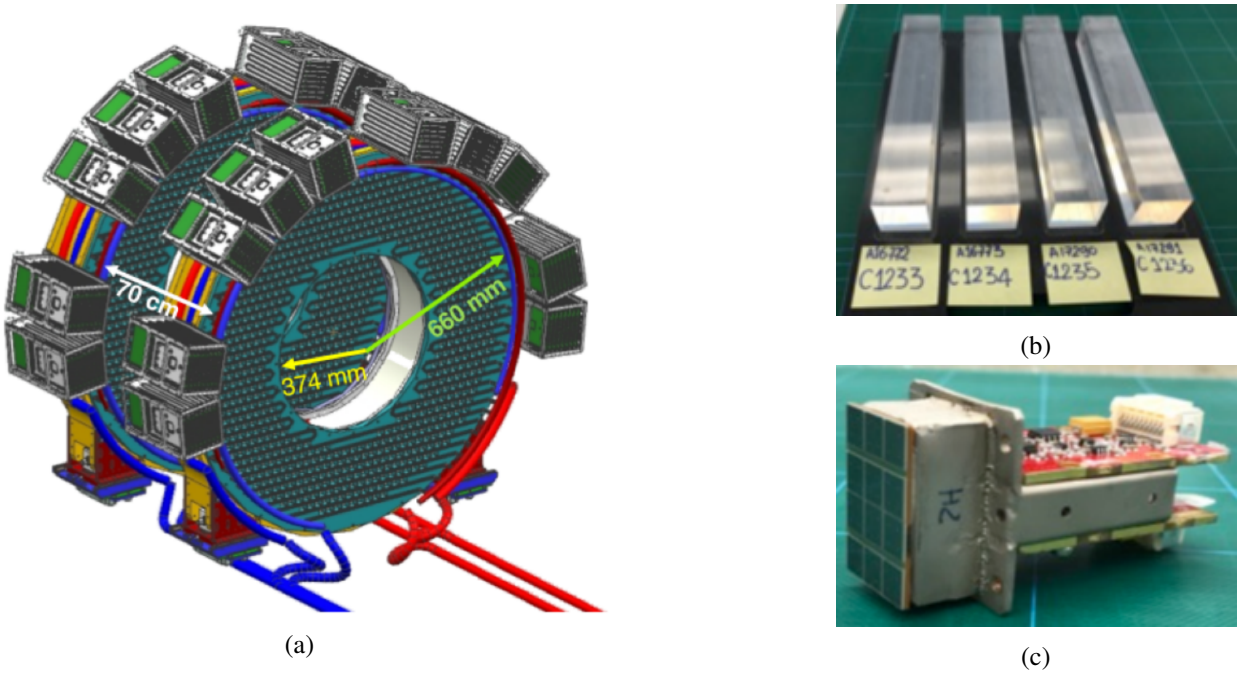
Figure 2.11: Annular design of Mu2e detectors: (a) a beam's-eye view of the Tracker [33]; (b) low-momentum ($<\sim 53$ MeV/c) electrons leave no hits on the Tracker (green); mid-momentum electrons leave some hits, but the tracks are not reconstructible (orange); only high-momentum ($>\sim 90$ MeV/c) electrons have reconstructible trajectories (blue).

of the electrons are represented by the circles of various colors passing through the stopping target in the center. For a uniform magnetic field, the radius of a helix is proportional to the transverse momentum of the electron. Thus, low momentum electrons ($<\sim 53$ MeV/c) would pass the detectors without leaving a hit. This corresponds to the green region in Figure 2.11(b), and an example trajectory is represented in Figure 2.11(a) as the green circle. Electrons with slightly higher momenta can leave some hits on the detectors. However, the few hits in the Tracker are not sufficient to reconstruct the electron trajectory (the orange region and the orange trajectory). The tracks become reconstructible only when the electron momentum gets high enough ($>\sim 90$ MeV/c), as shown in the blue region in Figure 2.11(b). It is estimated that only 100k of the muon DIO events will register in the Tracker [6]. The low rate eases requirements on the front-end electronics (FEE), and reduces reconstruction errors from accidental activities.

2.3.5 Electromagnetic Calorimeter

The Mu2e electromagnetic (EM) Calorimeter sits downstream of the Straw Tube Tracker. It consists of two annular detection disks, separated by 70 cm, as shown in Figure 2.12(a) [19]. The separation is chosen to match half of the distance between two periods in the helical trajectory of a typical 105 MeV electron. This maximizes the detection probability of conversion electrons in the Calorimeter; electrons that pass through the hole of the first disk will be detected by the second one.

Each Calorimeter disk is stacked with 674 undoped cesium iodide (CsI) crystals [19]. These crystals are $34 \times 34 \times 200$ mm^3 in dimensions, as shown in Figure 2.12(b). The scintillation light yield in the crystals is proportional to the incident particle energy. Each crystal is coupled with a



(a) (b) (c)

Figure 2.12: The Mu2e electromagnetic Calorimeter [19]: (a) the Electromagnetic (EM) Calorimeter consists of two annular rings; (b) cesium iodide (CsI) crystals used in Mu2e Calorimeter; (c) a Silicon Photomultiplier (SiPM) module used for Calorimeter readout.

readout module consisting of two ultraviolet (UV)-extended Silicon Photomultiplier (SiPM) arrays and the corresponding analog front-end electronics (FEE) boards, as illustrated in Figure 2.12(c) [19]. The electronics for digitizing signals and controlling readouts are housed in the blocks around the annulus outer perimeter, as shown in Figure 2.12(a). The Calorimeter is designed to have good energy resolution ($\sigma_E/E < 10\%$), timing resolution ($\sigma_t < 500$ ps), and position resolution (< 1 cm). It works in a vacuum of 10^{-4} Torr and a magnetic field of 1 T [15].

The Mu2e EM Calorimeter serves multiple purposes [15]. First, the energy deposition measured in the Calorimeter separates electrons and muons with the same momentum. The separation in energy-to-momentum ratio (E/p) can be used to identify particles, and suppresses part of the cosmic background caused by cosmic-ray muons. Second, the Calorimeter signals can improve the quality of track reconstruction by providing a "seed". The additional information can be used for consistency check, reducing reconstruction errors caused by accidental activities. Third, it provides triggers for the experiment independent of the Tracker.

2.3.6 Cosmic Ray Veto

As mentioned in Section 2.2, the Mu2e experiment expects backgrounds induced by cosmic rays at a rate of one per day. Although some backgrounds caused by muons with right momenta can be distinguished by the EM Calorimeter, other cases where cosmic muons decay to conversion-

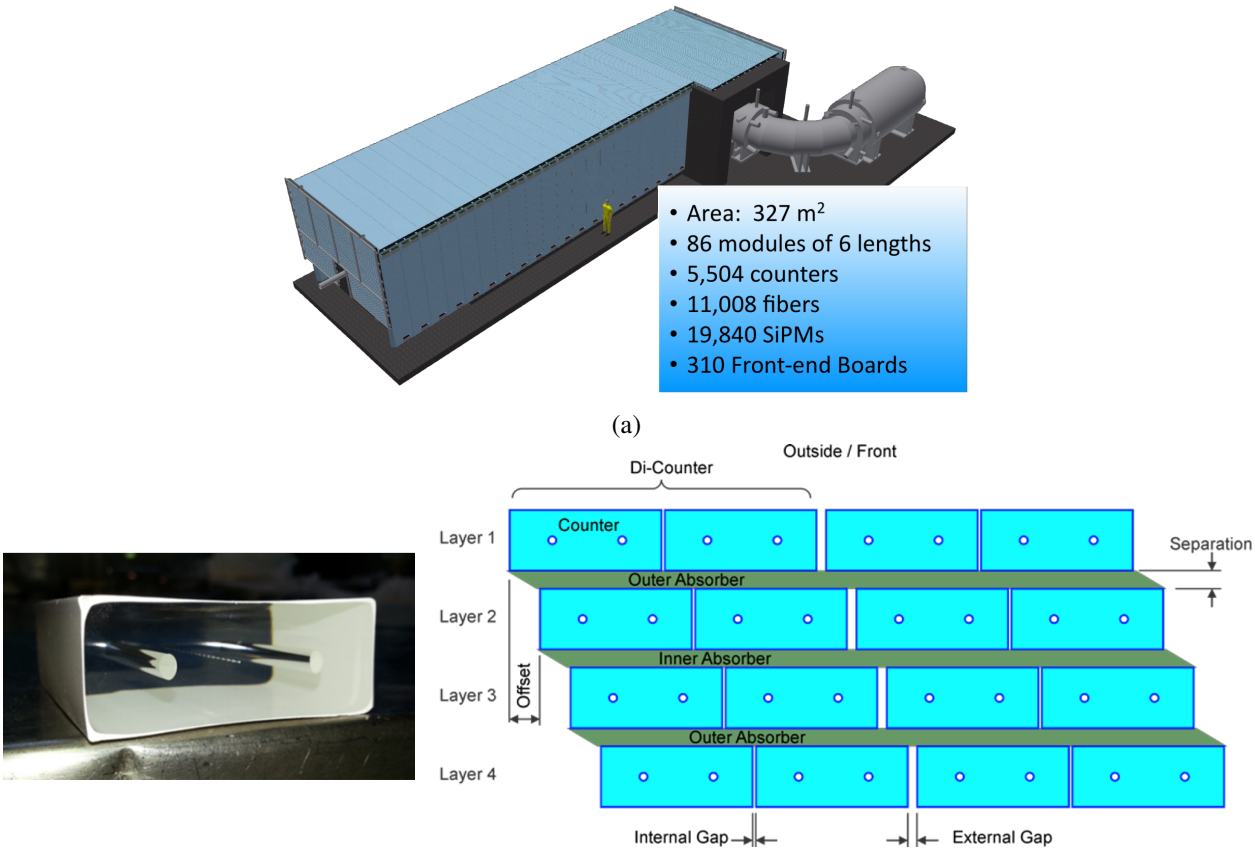


Figure 2.13: The Mu2e Cosmic Ray Veto [19]: (a) Mu2e CRV overview; (b) plastic scintillator counter used in CRV; (c) layout of a CRV module.

energy electrons remain a problem. In order to suppress these backgrounds, Mu2e adapts an active veto system combined with shielding.

Figure 2.13(a) [19] provides an overview to the Mu2e Cosmic Ray Veto (CRV) system. It covers the entire DS and half of the TS (there is no veto on the bottom), with a total area of $327\ \mathrm{m}^2$. The CRV modules are made of plastic scintillator extrusions, as shown in Figure 2.13(b). Each extrusion has a cross-sectional area of $5 \times 2\ \mathrm{cm}^2$, with various lengths. The extrusions are cladded with a titanium dioxide layer to increase internal reflections for a higher light yield. Two grooves are extruded inside the scintillator bar along the length, where $1.4\ \mathrm{mm}$ diameter wavelength shifting fibers are embedded. They transmit light to the extrusion ends where each fiber is read out by a $2 \times 2\ \mathrm{mm}^2$ SiPM on each end. Figure 2.13(c) [19] shows a cross-section of the CRV module. Each module contains four overlapping layers of plastic scintillator counters in order to minimize the effect of gaps. The layers are separated by $\sim 10\ \mathrm{mm}$ of aluminum plates as absorbers. When three out of four layers of the CRV are triggered, a veto window of $\pm 125\ \mathrm{ns}$ is issued. A conversion-like signal observed during this time window is assumed to be produced by a cosmic-ray muon [6]. It is

estimated the total dead time is $\sim 5\%$. The CRV system needs to be highly efficient. A detection efficiency of 99.99% or more is required for the success of the experiment.

2.4 Background and Sensitivity

This section outlines the process of determining the Mu2e signal region. A few terms normally used for discussing the sensitivity of the Mu2e experiment is defined, and the predictions based on the selected signal region are given. A summary of backgrounds in the signal region is also provided.

All discussions of the Mu2e experiment sensitivity are rooted in the estimation of backgrounds. Mu2e expects to collect 3.6×10^{20} protons-on-target (POT). As discussed, the live window after each proton pulse is set to start from 700 ns. Combined with detailed design geometry of the experiment, backgrounds from all sources discussed in Section 2.2 are evaluated in the high-momentum region through simulations. The readers are referred to [6] and [7] for details in background estimation. A representation of the expected backgrounds versus reconstructed track momentum is shown in Figure 2.14 [7]. Both statistical and systematic contributions to the uncertainties are shown.

The signal region of the experiment is optimized for discovery sensitivity [7]. To be precise, "discovery" means the value $R_{\mu e} = 0$ is excluded at the 5σ level, provided the level of backgrounds b present in the experiment. Consider an ensemble of pseudo-experiments with identical setup, POT, signal region, $R_{\mu e}$, and so on. In the ensemble, a certain fraction of the pseudo-experiments would report discovery. This fraction is dependent on the underlying value of $R_{\mu e}$; a higher $R_{\mu e}$ corresponds to a larger fraction. A median discovery sensitivity can be defined at the $R_{\mu e}$ value where half of the ensemble reports discovery. This is illustrated in Figure 2.15 [7]. The corresponding 1σ band is constructed using the $R_{\mu e}$ values that correspond to 1σ values in a normally-distributed discovery fraction. With the median discovery sensitivity defined, the signal window can be optimized to obtain a minimal median discovery sensitivity. The resultant value is the discovery reach of the experiment. Based on the estimated background level, Mu2e reports an optimal momentum window of $[103.85, 104.9]$ MeV/c. The corresponding median 5σ discovery sensitivity, or the discovery reach, is $R_{\mu e,5\sigma DR} = 2 \times 10^{-16}$ [7, 34]. Figure 2.14 [7] also includes the reconstructed conversion electron momentum distribution for the discovery reach.

The total background in the momentum window is 0.41 ± 0.13 events (both statistical and systematic contributions included). A detailed summary of backgrounds from various sources are shown in Table 2.2 [7]. The leading contribution is from the cosmic-ray-induced backgrounds, which corresponds to about half of all backgrounds.

Another commonly mentioned gauge of the experiment sensitivity is the single-event

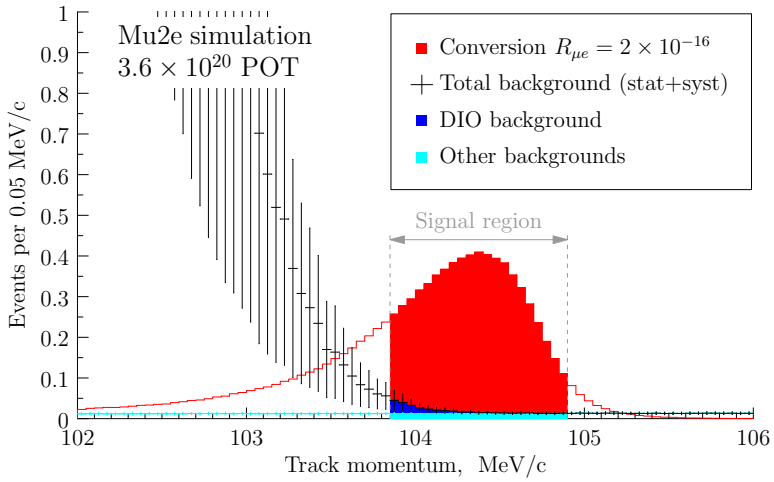


Figure 2.14: Simulated conversion electron and backgrounds [7]. There are 3.6×10^{20} protons-on-target (POT) in the simulation; signal window is optimized for discovery reach. The plotted conversion electron spectrum is based on the discovery reach of $R_{\mu e}$.

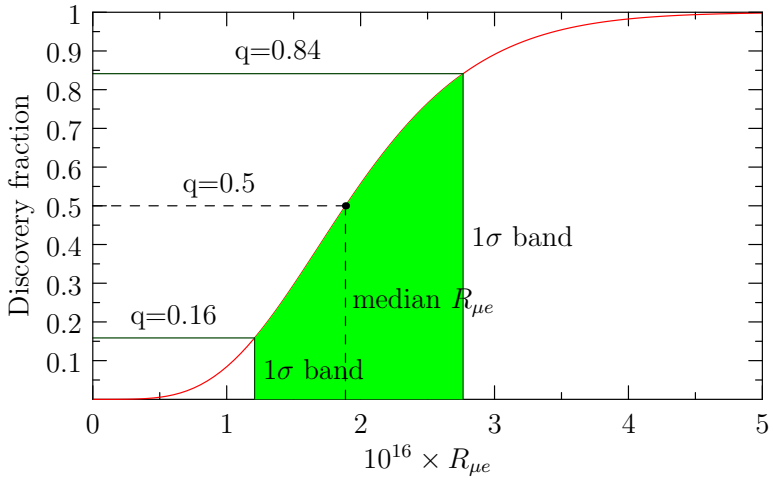


Figure 2.15: Discovery reach of the Mu2e experiment [7]. The corresponding signal region is the optimal $[103.85, 104.9]$ MeV/c.

sensitivity (SES). It is the $R_{\mu e}$ ratio when there is one signal event, which can be expressed as:

$$SES \equiv \frac{1}{N_{POT} \cdot P_{\mu\ stop} \cdot \epsilon_{CE} \cdot BR_{capture}} . \tag{2.7}$$

Here N_{POT} means the number of POT in the experiment; $P_{\mu\ stop}$ is the number of muons stopped on target per proton; ϵ_{CE} is the conversion electron acceptance, which is a product of detector efficiency (dependent on the momentum signal region) and fraction of muons interacting in the live time window; and $BR_{capture}$ is the branching ratio of muon captures, which is 60.9%. The SES value is evaluated to be 3.01×10^{-17} [7]. The ratio of $(R_{\mu e,5\sigma DR}/SES)$ gives the median of events required in the momentum signal region to make a discovery.

Table 2.2: A summary of estimated background events from various sources

Process	Expected Event Yield
Cosmic rays	0.209±0.022(stat)±0.055(syst)
Muon decay-in-orbit (DIO)	0.144±0.028(stat)±0.11(syst)
Antiprotons	0.040±0.001(stat)±0.020(syst)
Radiative pion capture (RPC)	0.021±0.001(stat)±0.002(syst)
Muon decay-in-flight (μ-DIF)	$<$0.003
Pion decay-in-flight (π-DIF)	0.001± $<$0.001
Beam electrons	$(2.1\pm1.0)\times10^{-4}$
Radiative muon capture (RMC)	$0.000^{+0.004}_{-0.000}$
Total	0.41±0.13(stat+syst)

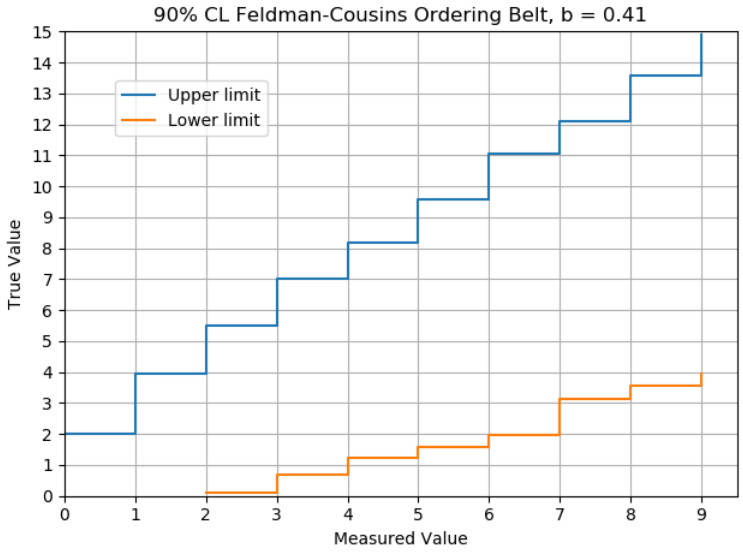


Figure 2.16: Confidence belt using Feldman-Cousins ordering at 90% CL for background $b = 0.41$. The figure is generated in accordance with [35].

Additionally, the exclusion power, or the expected upper limit at 90% CL is also referred to when describing the experiment sensitivity. The calculation of this value starts with generating the confidence belt at 90% CL using the expected background count of $b = 0.41$. Following the Feldman-Cousins ordering method [35], the Poisson-based 90% confidence belt is constructed as Figure 2.16. For a given number of measured events, the band shows a range that true number of events should be in at a 90% CL. In an ensemble of identical pseudo-experiments, if there is no conversion signals, the number of events observed should follow a Poisson distribution $Pois(b)$ of mean $b = 0.41$. Then the average upper limit of events observed at 90% CL, n_{UL}, is a weighted sum of the upper limits in the confidence band weighted by $Pois(b)$. For $b = 0.41$, calculation

gives $n_{UL} = 2.8$. Thus, the corresponding upper limit at 90% CL is given by:

$$R_{\mu e,UL} = n_{UL} \cdot SES = 8 \times 10^{-17} \, . \tag{2.8}$$

It should be pointed out that the signal region of the experiment can be optimized for upper limit as well. In this case, the signal region is selected so that the calculated upper limit is the lowest. Its deviation from the optimization for discovery is small; the resultant upper limit changes by only about 5% [7].

If the Mu2e experiment observes signals in the conversion electron region beyond the background level, it will be a clear sign of new physics. In future experiments, different stopping target materials, e.g. titanium, can be used to further understand the underlying mechanism and differentiate theories. On the other hand, if no signal is observed, the experiment can still place constraints on many BSM theories.

2.5 Lepton Number Violation Search at Mu2e

The Mu2e apparatus can also detect the lepton-number-violating process of $\mu^- N \to e^+ N'$. This process not only violates charged lepton flavor, the final state also has a different lepton number ($|\Delta L| = 2$). The conversion positrons leave helical tracks in the Tracker as well, although the chirality of these tracks are different from the electron tracks in the same magnetic field. The positron energy and momentum are comparable to the conversion electrons, so they are detectable under the same experiment running conditions. While this is an interesting side channel that Mu2e can explore, more detailed discussions of related theories, experimental history, and experiment backgrounds are included in Appendix A, so they do not interrupt the focus of the dissertation on the $\mu^- N \to e^- N'$ search.

References

[1] W. Bertl, *et al.* (SINDRUM II Collaboration), Eur. Phys. J. C . **47**, 337 (2006).

[2] T. Fazzini, *et al.*, Phys. Rev. Lett. **1**, 247 (1958).

[3] P. Zyla *et al.* (Particle Data Group), Prog. Theor. Exp. Phys. **2020**, 083C01 (2020), and 2021 update.

[4] D. F. Measday, Phys. Rept. **354**, 243 (2001).

[5] T. Suzuki, D. F. Measday, and J. P. Roalsvig, Phys. Rev. C **35**, 2212 (1987).

[6] L. Bartoszek, *et al.* (Mu2e Collaboration), *Mu2e Technical Design Report*, Tech. Rep. (Fermilab, 2015) arXiv:1501.05241 [physics.ins-det] .

[7] A. Gaponenko, "Summary of backgrounds for CD3," (Mu2e Backgrounds 2017-06-08), mu2e-doc-7464-v13.

[8] L. Michel, Proc. Phys. Soc. A **63**, 514 (1950).

[9] T. Kinoshita and A. Sirlin, Phys. Rev. **113**, 1652 (1959).

[10] A. Czarnecki, X. Garcia i Tormo, and W. J. Marciano, Phys. Rev. D **84**, 013006 (2011).

[11] A. Czarnecki, *et al.*, Phys. Rev. D **90**, 093002 (2014).

[12] R. Szafron and A. Czarnecki, Phys. Rev. D **92**, 053004 (2015).

[13] R. Szafron and A. Czarnecki, Phys. Lett. B **753**, 61 (2016).

[14] P. Bergbusch, *et al.*, Phys. Rev. C **59**, 2853 (1999).

[15] R. Bernstein, Front. Phys. **7**, 1 (2019).

[16] M. MacKenzie and P. Murat, Fermilab PUB-20-457-E 10.48550/arviv.2009.00214 (2020).

[17] J. A. Bistirlich, *et al.*, Phys. Rev. C **5**, 1867 (1972).

[18] R. Ehrlich, "Cosmic Ray Backgrounds in the Mu2e Experiment at Fermilab," (DPF 2017 Poster 2017-07-31), `https://indico.fnal.gov/event/11999/contributions/11447/attachments/7562/9672/poster_DPF2017_5.pdf`.

[19] S. Giovannella, J. Instrum. **15** (06), C06022.

[20] D. Stratakis, *et al.*, Phys. Rev. Accel. Beams **20**, 111003 (2017).

[21] S. J. Werkema, "CD-3c Review: Mu2e Accelerator Systems," (DOE CD-3c Review 2016-06-14), mu2e-doc-7530-v6.

[22] R. E. Ray, "Mu2e Project Overview for CD-3c Review," (DOE CD-3c Review 2016-06-14), mu2e-doc-7084-v23.

[23] E. Prebys, *et al.*, in *Proceedings of the 6th International Particle Accelerator Conference* (2015) [FERMILAB-CONF-167-APC].

[24] M. J. Lamm, "Mu2e Solenoid Status," (DOE IPR 2018 2018-10-02), mu2e-doc-20562-v5.

[25] M. J. Lamm, "CD-3c Review: Mu2e Solenoids," (DOE CD-3c Review 2016-06-14), mu2e-doc-7461-v9.

[26] J. D. Jackson, *Classical Electrodynamics* (Wiley, 1999) pp. 588–596.

[27] G. Ginther, "CD-3c Review: Muon Beamline Overview," (DOE CD-3c Review 2016-06-14), mu2e-doc-7531-v5.

[28] E. C. Dukes, "Mu2e-II: The Mu2e Experiment in the PIP-II Era," (NuFact 2021 2021-09-08), mu2e-doc-39318-v3.

[29] R. Abramishvili, *et al.* (COMET Collaboration), Prog. Theor. Exp. Phys **2020** (2020), 033C01.

[30] V. Cirigliano, *et al.*, Phys. Rev. D **80**, 013002 (2009).

[31] R. Kitano, M. Koike, and Y. Okada, Phys. Rev. D **66**, 096002 (2002), [Erratum: Phys.Rev.D 76, 059902 (2007)].

[32] F. Abusalma, *et al.*, "Expression of interest for evolution of the mu2e experiment," (2018), arXiv:1802.02599 [physics.ins-det; hep-ex].

[33] A. Mukherjee, "Mu2e Tracker," (DOE IPR 2017 2017-07-25), mu2e-doc-11567-v3.

[34] A. Gaponenko, "Updated Background and Sensitivity Numbers," (Mu2e Weekly 2017-07-13), mu2e-doc-12014-v1.

[35] G. J. Feldman and R. D. Cousins, Phys. Rev. D **57**, 3873 (1998).

Chapter 3 Mu2e Tracker Panels

The next part of the dissertation will be dedicated to the work performed on the Mu2e Tracker. To establish the context, this chapter will provide more details on the straw tube Tracker panel, including its detection mechanism, the mechanical design, and the FEE. Chapter 4 will discuss the Tracker data acquisition (DAQ) system and emphasize the work on the Tracker firmware. Then the Tracker ^{55}Fe source scan studies during the Vertical Slice Test (VST) will be shown in Chapter 5.

As mentioned in the last chapter, the Mu2e Tracker consists of 216 identical panels. Each panel is a standalone module both mechanically and electrically, and serves as a minimal functional unit for design, test, and operation. In each panel, there are 96 drift tubes that serve as the basic detecting elements. We will begin the discussion with the theoretic aspects of the drift tubes.

3.1 Drift Tubes

Among devices capable of measuring charged particle coordinates, a few characteristics of the drift tubes make them stand out in particular. They have good spatial resolution and high detection efficiency, while still remaining cost-efficient [1]. An overview of the detection mechanism of drift tubes is given below based on discussions in [2] and [3]; the readers are referred to the original work for more quantitative details.

Drift tubes are typical gaseous ionization detectors [2]. Figure 3.1 shows the basic configuration of a drift tube. A hollow cylindrical conducting tube is grounded and serves as the cathode.

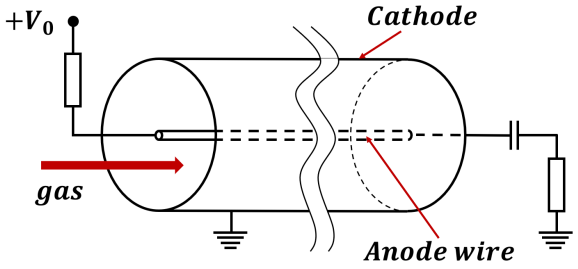


Figure 3.1: A schematic diagram of a drift tube.

A suitable gas, normally a mixture of noble gas such as argon and some quench gas, flows through the tube. Along the axis of the cylindrical cathode suspends a thin sense wire. A high voltage V_0 is applied to this wire which serves as the anode. Long thin drift tubes are also referred to as "straw tubes" due to their resemblance in shape.

Assuming a radius a of the anode wire and an inside radius b of the outer cylinder, the electric field E within the drift tube can be easily obtained using Gauss' theorem as a function of radial distance r from the axis:

$$E(r) = \frac{1}{r}\frac{\lambda}{2\pi\varepsilon} = \frac{1}{r}\frac{V_0}{\ln(b/a)} \ (a < r < b) \ . \tag{3.1}$$

Here λ is the linear charge density on the wire, and ε is the dielectric constant of the gas. The field is cylindrically symmetrical and reaches its strongest near the wire.[1] Generally speaking, thinner sense wires are preferred in drift tubes. As will be explained later, a stronger electric field near the wire leads to a higher amplification factor (gain) of the drift tube, given the same V_0 is applied. In addition, a thinner wire also leads to a better spatial resolution [4].

3.1.1 Gas Ionization

When charged particles traverse the gas volume, there are essentially two types of interactions for them to deposit energy, which are excitation and ionization. The latter is straightforward; through ionizing collisions between a charged particle p and an atom A, one or (sometimes) more electrons are ejected from the atom: $Ap \rightarrow A^+e^-p$ (or $Ap \rightarrow A^{++}e^-e^-p$, ... if more than one electron is ejected). This is referred to as the primary ionization process due to the presence of the original charged particle. On the other hand, the atoms excited by the original particle can cause ionization as well. In drift tubes, a noble gas atom often turns into a metastable state A^* after being excited through the primary interaction $Ap \rightarrow A^*p$. If the excitation energy of A^* is higher than the ionization potential of another species B in the gas (oftentimes the quencher), the Penning Effect can produce ionization through $A^*B \rightarrow AB^+e^-$. In addition, the noble gases can also form molecular ions from the metastable state through processes such as $A^*A \rightarrow A_2^* \rightarrow A_2^+e^-$. Ionization through these processes, along with that induced by secondary electrons with sufficient energy (delta electrons), is also known as secondary ionization.[2]

The mean number of electron-ion pairs created by an initial particle can be calculated as

[1] In case of small displacement of the wire from the center, due to, for example, fabrication errors or gravitational sagging of wire or straw, a perturbation to the field needs to be added for correction [3]. This can affect the gain of the drift tube as well as impact the accuracy of particle tracking. Hence, the anode wire has to be carefully aligned.

[2] It should be pointed out that neutral particles can induce ionizations as well. The case is clear for photon which can interact photoelectrically. Other neutral particles such as neutron can also ionize indirectly through secondary interactions. Nonetheless, the neutron detection efficiency in drift tube detectors is low (typically on the order of 10^{-4} [5]).

its total energy loss divided by the average amount of energy required to produce an electron-ion pair. Due to the energy lost to excitation, this average does not equal to the gas ionization potential. Measurements revealed values on the order of one electron-ion pair per $30\ \mathrm{eV}$, with some variations depending on the type of the gas and the initial particle. This value is independent of initial particle energy except for very slow particles.

Without the presence of an electric field, the electrons and the ions produced in ionizations diffuse isotropically. They lose energy through collisions and soon reach thermal equilibrium with the gas. Eventually they recombine. The maximum range an electron freed in ionization can reach is positively correlated with its initial kinetic energy. A $10\ \mathrm{keV}$ electron is stopped in roughly $1\ \mathrm{mm}$ in a gas of normal temperature and pressure (NTP). Typically, the kinetic energy of an electron obtained from ionization is lower, so the range is also smaller.

3.1.2 Drift of Ions and Electrons

In the presence of an electric field, the freed electrons and ions are accelerated towards the anode and cathode respectively, along the electric field lines. During acceleration, the charges scatter on other particles in the gas and lose energy. Their directions of motion are randomized and maximum speeds are limited. Consequently, *on average* the charges assume a constant velocity along the electric field. This is called the drift velocity of the charge. It is superimposed with the thermal motions.

Ions and electrons behave differently during drift, and their drift velocities are determined by different factors. The ion mass is much larger than that of the electron and is comparable to those of gas molecules. During collisions, a good fraction of the energy that the ions acquired during acceleration is transferred to gas molecules.[3] Consequently, under the typical electric field strength in a drift tube detector, only a small amount of electric field energy enters the energy associated with the random motions of ions, so it is not much larger than the original thermal energy before acceleration. In such circumstance, the ion drift velocity u is proportional to the reduced electric field E/N where N is the number density of the gas.[4] In drift tubes, the typical ion drift velocities are on the order of $\mathcal{O}(10^3\ \mathrm{cm/s})$ (except in the region adjacent to the anode wire, where a much stronger electric field leads to a larger drift velocity). This value is smaller compared to the typical ion thermal velocities $\sim \mathcal{O}(10^4\ \mathrm{cm/s})$ under room temperature.

[3] In gas mixtures, if the drifting ions collide with gas molecules with lower ionization potential, a charge exchange can occur, which neutralizes the original ion and creates a new ion. Through such process of charge transfer, ion species from different molecules disappear quickly, leaving only the ion species with the lowest ionization energy.

[4] In some older literatures, the "reduced electric field" sometimes refers to the ratio E/p between the electric field strength and the gas pressure, which differs by a factor of $k_B T$ (Boltzmann constant multiplied by absolute temperature) for ideal gas. The proportionality breaks when the electric field strength gets much higher compared to typical drift tube conditions. There is $u \propto \sqrt{E/N}$ at high fields.

On the contrary, electrons only transfer a small fraction of their kinetic energy during elastic collisions, so they can acquire more energy from the electric field than their initial thermal energy. Electron drift velocity is influenced by many factors. For instance, some gas molecules (e.g., H_2O, CO_2) have strong electron affinity that can combine with electrons through electron attachment and form negative ions. In some other cases, the electrons can accumulate enough energy that surpasses excitation threshold of gas molecules which leads to inelastic collisions. Different factors affect the cross-section of collision of the electrons across a wide energy spectrum, which make the electron drift velocity a complicated function of the electric field strength. Figure 3.2 shows the electron drift velocity at different electric field strengths in argon-carbon dioxide (Ar-CO_2) mixtures of different proportions [6]. Typically, electron drift velocities in drift tubes are on the order of $\mathcal{O}(10^6\ \mathrm{cm/s})$. This is much larger than the ion drift velocities, and is comparable to the electron thermal velocity under same conditions. The radial coordinate of an ionization can be calculated using the electron drift velocity and its drift time.

As mentioned, the drifting electrons and ions are scattered on gas molecules, so they also diffuse along their way.[5] Electrons diffuse much faster than ions due to their large root mean square velocity. For drift tubes measuring spatial coordinates of incoming particles, the intrinsic resolution

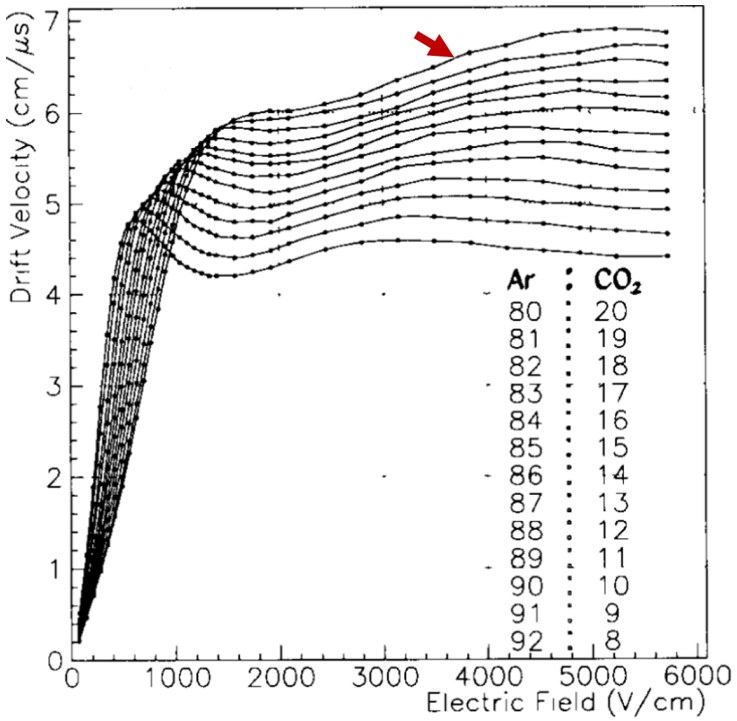


Figure 3.2: Electron drift velocity versus electric field in Ar-CO_2 mixtures of different proportions [6]. The red arrow points at the curve of 80%/20% Ar-CO_2 mixture, which is the gas used in the Mu2e Tracker [7].

[5]The presence of electric field (and magnetic field, if it exists) leads to anisotropy in both electron and ion diffusion, unlike the case when there is no electric field.

would be limited by such electron diffusion. At high electric fields, it is preferable to have low energy associated with the random motion of electrons, which helps to reduce electron random speeds, hence lowers the diffusion range. Certain gas molecules like CO_2 ("cold gas") has internal degrees of freedom accessible at low collision energies, which prevents electron energies from being much higher than thermal energy until the field strengths raise above $2\ \mathrm{kV/cm}$,[6] and helps to improve the intrinsic spatial resolution.

3.1.3 Avalanche Multiplication

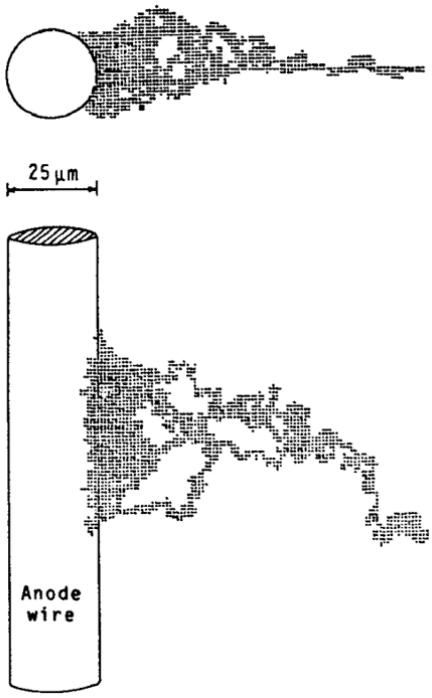
When electrons reach the region near the anode wire where the electric field is strong, they acquire sufficient energy between consecutive collisions and cause ionizations. The same goes for the freed secondary electrons, which then produce tertiary electrons and so on. The number of free electrons multiplies rapidly, forming an avalanche. For a typical drift tube, in which the electron mean free path is on the order of micrometers, an avalanche occurs when the electric field reaches $\mathcal{O}(10^4\ \mathrm{V/cm})$. According to Equation 3.1, using parameters $a \sim \mathcal{O}(10^{-3}\ \mathrm{cm})$, $b \sim \mathcal{O}(1\ \mathrm{cm})$, and a normal operating voltage V_0 of $1 \sim 2\ \mathrm{kV}$, the calculation shows that avalanches occur within $\mathcal{O}(100\ \mu\mathrm{m})$ from the anode wire. This means the electrons produced in the avalanche are collected on the anode wire within $\sim 1\ \mathrm{ns}$, leaving the positively charged ions drifting towards the cathode. As will be shown later, the drifting ions are mainly responsible for creating signals in the electrodes through induction.

Figure 3.3 shows an avalanche originated from a single electron from a Monte Carlo simulation [8]. Shading in the figure shows the avalanche electron density near the anode wire. With typical drift tube working conditions and dimensions, avalanches mainly develop on the approaching side of the initial electron. They would start to surround the wire, however, if the voltage is raised. The lateral extent of the avalanche, on the other hand, can span $\mathcal{O}(100\ \mu\mathrm{m})$, which is roughly the electron mean free path multiplied by the number of generations in the avalanche, as it is mainly due to electron diffusion and electrostatic repulsion among electrons.

3.1.3.1 Avalanche Gain

Here we derive an expression for the amplification factor of the avalanche, which is also referred to as the gain. In an avalanche, the number of electrons freed per unit path length is given by the first Townsend ionization coefficient α. It is determined by the electron cross-sections and detailed ionization processes, so is different for each gas mixture. In particular, it is positively correlated with the electric field strength E, as a larger E corresponds to a higher kinetic energy of

[6] On the other hand, an example of "hot gas" is argon, in which a field of $1\ \mathrm{V/cm}$ is sufficient to produce electron energies much higher than thermal energy.



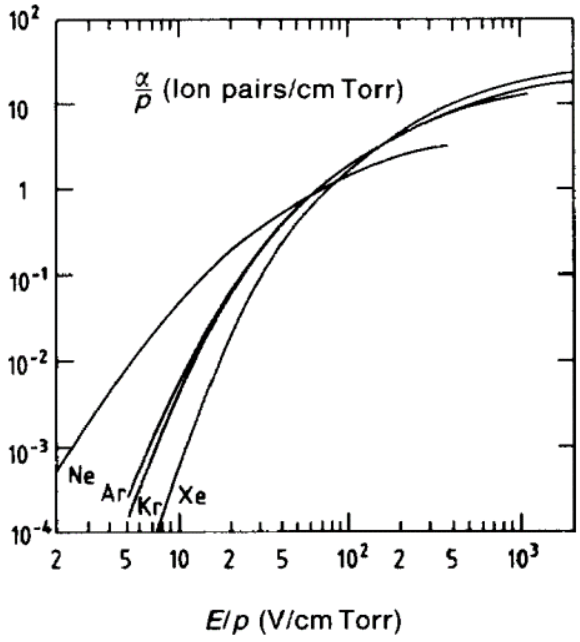


Figure 3.3: (Left) A display of electron density near an anode wire in an avalanche originated from a single electron, created by a Monte Carlo simulation [8].

Figure 3.4: (Right) First Townsend ionization coefficient of some noble gases as a function of reduced electric field [3].

the electron, which increases the ionization cross-section. Figure 3.4 [3] presents the first Townsend ionization coefficient of some noble gases as a function of reduced electric field E/p, where p is the gas pressure. Trivially, with N initial electrons, the increase number of electrons $\mathrm{d}N$ per path length $\mathrm{d}s$ is given by:

$$\mathrm{d}N = N\alpha \mathrm{d}s \; . \tag{3.2}$$

In the drift tube the field is non-uniform (Equation 3.1), so implicitly α is a function of s. By integrating Equation 3.2 between the point $S_{\min}$ where the avalanche begins and the wire radius a, the gain G can be calculated as:

$$G \equiv N/N_0 = \exp \int_{S_{\min}}^{a} \alpha(s)\mathrm{d}s = \exp \int_{E_{\min}}^{E(a)} \frac{\alpha(E)}{\mathrm{d}E/\mathrm{d}s}\mathrm{d}E \; . \tag{3.3}$$

Here $E_{\min}$ is the minimum field required for ionization to occur, and $E(a)$ is the electric field strength at the surface of the wire. As the minimum electron energy required for ionization is proportional to the product of $E_{\min}$ and the mean free path l, there is $E_{\min} \propto (1/l) \propto \rho$ (ρ is the gas density). Then for any gas density ρ, the corresponding $E_{\min}(\rho)$ can be related to a standard

$E_{\min}(\rho_0)$ of gas density ρ_0 by:

$$E_{\min}(\rho) = E_{\min}(\rho_0)\frac{\rho}{\rho_0} \tag{3.4}$$

In a drift tube filled with some noble gas, the reduced electric field E/p is typically between 10^2 and 10^3 $\mathrm{V/cm \cdot Torr}$. Observing the corresponding region in Figure 3.3, it is reasonable to approximate the first Townsend ionization coefficient using a proportional relation of $\alpha = \beta E$. Inserting this relation together with Equation 3.1 into Equation 3.3, the following relation is obtained:

$$\ln G = \frac{\beta\lambda}{2\pi\varepsilon}\ln\frac{E(a)}{E_{\min}} = \beta a E(a)\ln\frac{E(a)}{E_{\min}}\,. \tag{3.5}$$

The coefficient β can be related to ΔV, the average potential drop to multiply the gain by a factor of 2. As the voltage drop per unit path length is $\mathrm{d}V = E(s)\mathrm{d}s = (\alpha/\beta)\mathrm{d}s$, substituting this back into Equation 3.2 yields $\mathrm{d}N = N\beta\mathrm{d}V$. Integrating this over a potential that doubles the gain, one can derive that $\beta = (\ln 2)/\Delta V$. Hence, the gain of a drift tube can be expressed as

$$\ln G = \frac{aE(a)\ln 2}{\Delta V}\ln\frac{E(a)\rho_0}{E_{\min}(\rho_0)\rho}\,, E(a) = \frac{V_0}{a\ln(b/a)}\,, \tag{3.6}$$

which is known as the Diethorn's formula. The parameters $E_{\min}(\rho_0)$ and ΔV can be extracted from gain measurements with varying ρ/ρ_0, a and $E(a)$. For a drift tube of given geometry and known gas mixture, the major factors influencing its gain are gas temperature T and pressure P, as well as the operating voltage V_0. In terms of small fluctuations in pressure ΔP and temperature ΔT, the change in gain ΔG can be approximated by [9]:

$$\frac{\Delta G}{G} = \frac{V_0\ln 2}{\Delta V\ln(b/a)}\left(\frac{\Delta T}{T} - \frac{\Delta P}{P}\right)\,. \tag{3.7}$$

3.1.3.2 Quench Gas

Oftentimes, a quench gas, such as $\mathrm{CO_2}$, methane ($\mathrm{CH_4}$), and some other hydrocarbons, is added to the drift tube gas mixture to prevent additional secondary avalanches. During an avalanche, photons can be produced by processes such as deexcitation of gas molecules and electron attachment to electronegative species (forming a negative ions). Some of the photons are energetic enough to cause ionizations outside the primary avalanche region, or create free electrons through photoelectric effect at the conducting cathode surface, which then give rise to secondary avalanches. This becomes an issue as the signal produced would be unproportionately large for the amount of energy deposited by the initial particle, and it would be no longer localized to where the energy deposition occurred. Even worse, a detector breakdown (continuous discharge) may occur if there are enough energetic photons to cause a chain reaction of secondary avalanches. The addition of the

quench gas can prevent the secondary avalanches from happening, as its molecules can absorb the ionizing photons before they travel far. Under normal operating conditions, even a small amount of quench gas can largely reduce the chance of secondary avalanches and prevent breakdowns from happening.

3.1.3.3 Operation Modes of Gaseous Ionization Detectors

As demonstrated in the derivation of the gain, in drift tube detectors the number of electron-ion pairs produced in an avalanche is proportional to the initial number of electrons. This is referred to as operating in a proportional mode, which requires a suitable operating voltage so that the changes of the electric field caused by the charges produced in avalanches are negligible.

A gaseous ionization detector can operate in different modes depending on the operating voltage applied, as shown in Figure 3.5[2]. Overall, higher operating voltage corresponds to more charges collected on the electrodes. When the voltage is very low, the charges produced in

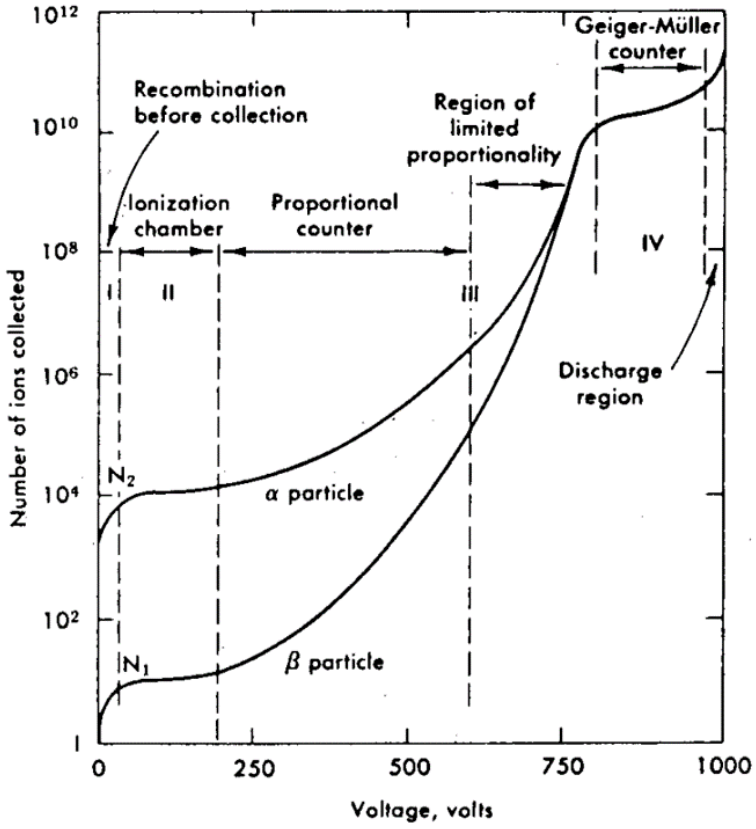


Figure 3.5: Dependence of particle gain on applied voltage in gaseous ionization detectors [2]. The values on the axes are for scale only; transitional voltages between different work modes vary with geometries and gas contents of different devices, so does the specific gain. Drift tubes work in the proportional mode which is indicated by the region above region II and below point III.

ionizations recombine before they reach any electrodes, so no signals are collected. Under slightly higher operating voltages (region II), the charges can drift to the electrodes, but the electric field is too weak for avalanches to proceed. This would be the case for ionization chambers. Further increasing the operating voltage, we reach the realm of drift tubes and other proportional counters (below point III). When the voltage gets high enough, the proportionality would be lost. Once the electrons produced in an avalanche are collected, the large number of positive ions near the anode would have sufficient space charge density to influence the electric field. For electrons in subsequent avalanches that enter the space between the positive ion cloud and the wire, they would experience a reduced electric field which leads to a smaller amplification. In the meantime, near the tail of the avalanche (i.e., in the direction of the avalanche with respect to the wire, radially further away than the ion cloud), the electric field gets stronger instead. Devices are said to work in a mode of limited proportionality in this voltage region. Finally, when the operating voltage gets very high (region IV), enough energetic photons are produced during avalanches that secondary avalanches propagate through the entire detector length regardless of the quench gas, causing a detector breakdown and saturating the output completely. In this voltage region, detectors operate in the breakdown mode, which is also referred to as the Geiger-Müller mode due to Geiger-Müller counters.

3.1.4 Signal Creation and Propagation

Contrary to intuition, the signal pulses read from drift tubes are not directly from collecting charges produced in avalanches. Were it the case, the anode wire would see a full signal within about a nanosecond, which is not true. Instead, the signal pulses are formed by charges induced on the electrodes due to the motion of the electrons and the ions. The induced charge and hence the induced current can be calculated using the Shockley-Ramo theorem [10, 11], which is based on Green's second identity and Green's reciprocation theorem [12]. A full derivation along with additional discussions can be found in [3]; additional review of the theorem can be found in [13].

The Shockley-Ramo theorem states that the charge Q_k and current I_k (pointing away from the electrode) induced on an electrode k by a point charge moving along a trajectory $\boldsymbol{x}(t)$ are given by:

$$Q_k = -q\psi_k(\boldsymbol{x}) \,, \tag{3.8}$$

$$I_k = -\frac{\mathrm{d}Q_k}{\mathrm{d}t} = q\boldsymbol{\nabla}\psi_k(\boldsymbol{x})\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = -q\boldsymbol{v}(t)\cdot\boldsymbol{E}_k(\boldsymbol{x}) \,, \tag{3.9}$$

where $\boldsymbol{v}(t) = \mathrm{d}\boldsymbol{x}/\mathrm{d}t$ is the velocity of the charge; $\psi_k(\boldsymbol{x})$ and $\boldsymbol{E}_k(\boldsymbol{x}) = -\boldsymbol{\nabla}\psi_k(\boldsymbol{x})$, respectively referred to as the weighting potential and the weighting field, are the potential and the field at the instantaneous position of $\boldsymbol{x}$, given q is removed, electrode k is set at unit potential, and all other electrodes are grounded. In other words, the dimensionless scalar field ψ_k is determined by the boundary condition of $\psi_k|_j = \delta_{jk}$ for any electrode j in the system.

The Shockley-Ramo theorem leads to several useful conclusions. First, for a charge q that moves from a position $\boldsymbol{x}_0 = \boldsymbol{x}(t_0)$ to another position $\boldsymbol{x}_1 = \boldsymbol{x}(t_1)$, the induced charge from the electrode k is simply $Q_k^{ind} = q[\psi_k(\boldsymbol{x_1}) - \psi_k(\boldsymbol{x_0})]$, which means the total induced charge is determined solely by the initial and the final position of the moving charge and is independent of the specific path. Now if a charge pair q and $-q$ is produced at some location $\boldsymbol{x}_0$ and drift to respective electrodes at $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, after they both arrive, the total induced charge on the electrode k is then the sum:

$$\begin{aligned} Q_k^{ind} &= q[\psi_k(\boldsymbol{x_1}) - \psi_k(\boldsymbol{x_0})] + (-q)[\psi_k(\boldsymbol{x_2}) - \psi_k(\boldsymbol{x_0})] \\ &= q[\psi_k(\boldsymbol{x_1}) - \psi_k(\boldsymbol{x_2})] \, . \end{aligned} \tag{3.10}$$

If k is the destination of one of the charges, say q, corresponding $\psi_k(\boldsymbol{x_1}) = 1$ and $\psi_k(\boldsymbol{x_2}) = 0$, so $Q_k^{ind} = q$. Or if none of the charges arrive at k, then $\psi_k(\boldsymbol{x_1}) = \psi_k(\boldsymbol{x_2}) = 0$ and $Q_k^{ind} = 0$. In other words, the total charge induced on an electrode by a charge pair is equal to the charge collected on it. In addition, if we treat all electrodes as a unity, all electrodes have a weighting potential of 1. If there is one electrode that encloses all the others, the consequent weighting field in the enclosed space is 0 everywhere. This means at any moment, the total induced current on all electrodes is always zero.

Now the Shockley-Ramo theorem can be applied to the drift tube. In an avalanche where N electron-ion pairs are produced, we examine the induced current signal on the anode wire (the current on the cathode simply contains an extra negative sign). The weighting potential and weighting field in the straw are functions of the radial coordinate r:

$$\psi_w(r) = \frac{\ln(b/r)}{\ln(b/a)} \, , \tag{3.11}$$

$$\boldsymbol{E}_w(r) = \frac{1}{r \ln(b/a)} \hat{\boldsymbol{r}} \, . \tag{3.12}$$

For the ions, whose drift velocity u is proportional to the electric field strength E as previously discussed, we can write $u = \mu E$.[7] Then the ion trajectory satisfies:

$$\frac{\mathrm{d}r(t)}{\mathrm{d}t} = u = \frac{\mu V_0}{r(t) \ln(b/a)} \, . \tag{3.13}$$

With the ionization occurring very close to the anode, an approximation of the initial condition is

[7] The parameter μ here is also known as the mobility of the particle.

given as $r(0) = a$. Then the solution to the above differential equation is:

$$r(t) = a\sqrt{1 + \frac{t}{t_0}},\ t_0 = \frac{a^2 \ln(b/a)}{2\mu V_0}\ . \tag{3.14}$$

Here t_0 is the characteristic time, which is typically on the order of $1\ \mathrm{ns}$. The time t is in the range $0 < t < t_{max}$ where $t_{max} = t_0[(b/a)^2 - 1]$. Then the induced current and the charge from the wire at t are given by:

$$I^{ind}_{ion}(t) = -(eN)E_w(r)\frac{\mathrm{d}r(t)}{\mathrm{d}t} = -\frac{eN}{2\ln(b/a)}\frac{1}{t + t_0}\ , \tag{3.15}$$

$$Q^{ind}_{ion}(t) = \int_0^t I^{ind}(t')\mathrm{d}t' = -\frac{eN}{2\ln(b/a)}\ln\left(1 + \frac{t}{t_0}\right)\ . \tag{3.16}$$

On the other hand, the total charge induced on the wire by the moving electrons is:

$$Q^{ind}_e = -eN[\psi_w(a) - \psi_w(r_{avg})] = -eN\frac{\ln(r_{avg}/a)}{\ln(b/a)}\ , \tag{3.17}$$

where r_{avg} represents an average position of ionizations in the avalanche. Typically, there is $r_{avg} \sim \mathcal{O}(10\ \mu\mathrm{m})$, so when compared with the total charge induced $Q^{ind}_{tot}(t_{max}) = -eN$, the movement of the electrons in the avalanche are only responsible for $1 \sim 2\%$ of it. The signal can be almost entirely attributed to the drift of positive ions from the avalanche.

From this point on, the signals can be treated as coming from nodes of ideal current sources in electrical models. A signal propagates to both ends of the drift tube from the avalanche location. For low-frequency components of the signal (for a meter-long drift tube, this translates to a frequency much less than $300\ \mathrm{MHz}$), a quasi-electrostatic approach is sufficient. The signals seen at tubes ends are influenced by impedances between and on the electrodes. On the other hand, for the high-frequency components of the signal, the tube needs to be treated as a transmission line. The propagation speed of a signal wave with frequency ω is then $c/\sqrt{\varepsilon(\omega)}$, where c is the speed of light and ε is the dielectric constant of the gas [14]. As ε is a function of ω, components of different frequencies propagate at slightly different velocities. This leads to a dispersion (widening) of the signal [12]. It brings some subtleties when using signal arrival time difference between straw tube ends to determine the longitudinal coordinate of an avalanche, which will be discussed more in Chapter 5.

3.1.5 Example Detector Responses

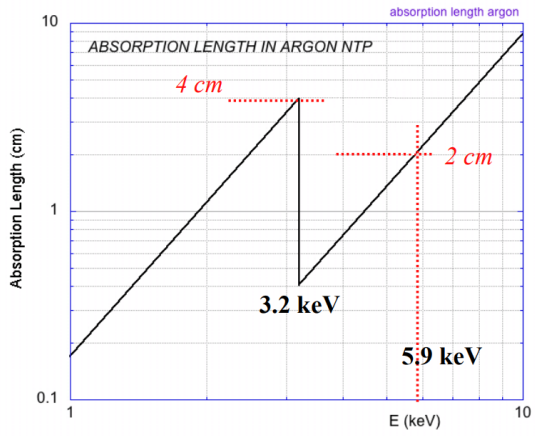
Before moving on, the respective responses of a straw tube to three types of radiation sources will be briefly outlined here to set some grounds for later discussions and to define some terms. In specific, we will make a quick overview on the detector's response to an $^{55}\mathrm{Fe}$ x-ray photon source,

the minimum ionizing particles (MIPs), and highly relativistic electrons.

3.1.5.1 X-Ray Photons from an ^{55}Fe Source

^{55}Fe is a radioactive isotope of iron. It has a half-life of about 2.7 years [15]. It decays through electron capture, and emits either an Auger electron or an x-ray photon when some outer-shell electron moves to the empty orbital in the K shell. In capsulated radioactive sources normally used in laboratories, the ~ 5 keV Auger electrons are blocked by the packaging materials, which makes ^{55}Fe effectively an x-ray source. The majority of the emitted photons are $K\alpha_1$ and $K\alpha_2$ x-rays, both with an energy close to 5.9 keV. Photons from less frequent $K\beta$ emissions are around 6.5 keV. Since the $K\beta$ decay rate is only about one seventh compared to that of $K\alpha$, and the energy resolution is limited, straw tubes normally cannot resolve the $K\beta$ emission peak from the $K\alpha$ one.

These x-ray photons from ^{55}Fe decays are not energetic enough to pair produce. When they traverse the drift tube, they can either lose energy to electrons through Compton scattering, or free bound electrons through photoelectric effect [16]. In the latter case, the excited atoms or molecules emit Auger electrons or fluorescence photons to return to the ground state. The scattered photons and fluorescence photons have various energies lower than the initial x-ray photons, which correspond to different absorption lengths. Specifically in argon under NTP, the absorption lengths of photons with different energies are shown in Figure 3.6 [16]. The drop at 3.2 keV is the K absorption edge [2, 17], which corresponds to the electron binding energy of the K shell. When the



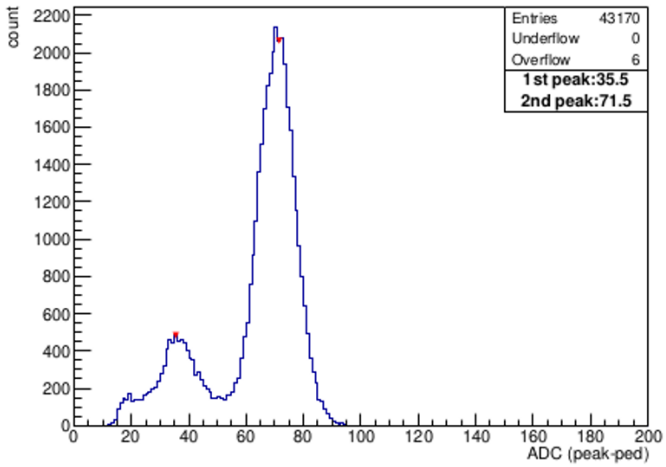


Figure 3.6: (Left) Photon absorption length versus photon energy in argon under normal temperature and pressure [16]. The drop at 3.2 keV is the K absorption edge, which gives rise to the argon escape peak in the ^{55}Fe energy spectrum.

Figure 3.7: (Right) Example ^{55}Fe energy spectrum from a straw tube measurement. The higher peak corresponds to the full energy of the 5.9 keV x-ray photon; the lower peak is the argon escape peak.

photon energy gets larger than this value, its energy becomes sufficient to free K-shell electrons photoelectrically, which increases its cross-section and leads to the dip in the absorption length. For photons with energies slightly below the K absorption edge, their absorption lengths can be relatively large compared to the straw dimensions, which means they can exit the detection volume of the straw tube and only part of the initial photon energy is deposited in the detector. This leads to the energy spectrum shown in Figure 3.7. The peak centered around 71.5 ADC counts corresponds to the photons depositing the entire $\sim 5.9\ \mathrm{keV}$ energy inside the detector. The lower peak, on the other hand, misses the energy carried away by the photons escaping the detector, which is why it is referred to as the "argon escape peak".

The energy deposition of the x-ray photons proceeds in a localized manner, where most of the ionizations are produced near the site of the Compton scattering or the photoelectric interaction, as well as the location where the Compton scattered photon or the fluorescence photon is absorbed. This differs from the behaviors of the MIPs or highly relativistic electrons, which is discussed next.

3.1.5.2 Minimum Ionizing Particles

For massive charged particles, the *mean* rate of energy loss in materials through ionization (i.e. inelastic collisions with atomic electrons) is described by the Bethe-Bloch formula [18]. It accounts for most of the energy loss of particles in materials across a wide range of incident particle energies, since elastic scatterings with the absorber nuclei generally do not transfer much energy. The formula gives the linear stopping power (unit of $\mathrm{MeV/cm}$) of materials and is applicable to relativistic particles:[8]

$$\left\langle -\frac{\mathrm{d}E}{\mathrm{d}x} \right\rangle = 4\pi N_A r_e^2 m_e c^2 \rho \frac{Z}{A} \frac{z^2}{\beta^2} \left[\frac{1}{2} \ln \frac{2m_e c^2 \beta^2 \gamma^2 W_{\max}}{I^2} - \beta^2 - \frac{\delta}{2} \right] , \tag{3.18}$$

$$W_{\max} = \frac{2m_e c^2 \beta^2 \gamma^2}{1 + 2\gamma m_e/M + (m_e/M)^2} , \tag{3.19}$$

where:

[8] In some literatures, a slightly different form of the Bethe-Bloch formula calculates the mass stopping power which is also denoted as $\langle -\mathrm{d}E/\mathrm{d}x \rangle$. In such a case, the factor of absorbing material density ρ is excluded from the formula and the unit becomes $\mathrm{MeV} \cdot \mathrm{cm}^2/\mathrm{g}$. The formula shown here includes the density effect correction term but does not include the shell correction term, which are respectively important for cases with high and low $\beta\gamma$. Additional corrections exist for specific cases including higher-order photon coupling with heavy ions, spin effects, and particle internal structure, just to name a few. [2]

N_A:	Avogadro's number;	z:	incident particle charge in units of e;
r_e:	classical electron radius;	β:	incident particle speed in units of c;
$m_e c^2$:	electron rest mass energy;	γ:	Lorentz factor $\gamma = 1/\sqrt{1-\beta^2}$;
ρ:	absorbing material density;	$W_{\max}$:	maximum energy transfer to electron in a single collision;
I:	mean excitation energy;		
Z:	absorbing material atomic number;	M:	incident particle mass;
A:	absorbing material mass number;	δ:	density-effect correction.

This form of the Bethe-Bloch formula is valid for $0.1 \lesssim \beta\gamma \lesssim 1000$, which is sometimes referred to as the "Bethe region". A particle at $\beta\gamma \sim 0.1$ has comparable velocity to that of atomic electrons so a shell correction must be included. At even lower energies, energy loss through interactions with the absorber nuclei becomes important. On the other hand, for highly relativistic particles, radiative losses (from pair production, bremsstrahlung, and photonuclear interactions) are dominant [18]. In the lower Bethe region, the stopping power is dominated by the $1/\beta^2$ factor and decreases rapidly with rising $\beta\gamma$. The stopping power reaches its minimum ("minimum ionizing") at around $\beta\gamma \sim 3$, and rises again slightly. The increase is small due to the logarithmic dependence in the equation, as well as the density-effect correction, which is related to the polarization of the medium by the charged particle [18]. Consequently, the massive charged particles all experience an energy loss similar to the minimum when they reach relativistic speeds. Thus, they are categorized together as minimum ionizing particles (MIPs). The exact value of the minimum energy loss and its location are weakly influenced by factors such as particle species (different masses lead to different $W_{\max}$) and absorbing materials (due to different excitation energies). Nonetheless, such variations are often insignificant, and a rule-of-thumb energy loss approximation of $2\ \mathrm{MeV} \cdot \mathrm{cm}^2/\mathrm{g} \cdot \rho$ can be used [18].

The Bethe-Bloch formula only provides the mean energy loss; the distribution of the energy loss, on the other hand, was calculated by Vavilov [19]. For thin absorbers or gases, the solution can be well approximated by a Landau distribution that features a long tail and a most probable value much lower than the mean. Figure 3.8 shows a spectrum of cosmic rays measured in a straw tube, where cosmic-ray muons are typical MIPs. Due to the energy resolution of the detector, this spectrum takes the shape of a Landau-Gaussian convolution (with some corrections due to the cylindrical straw shape) rather than a pure Landau distribution.

Microscopically, in drift tubes a MIP loses energy through many inelastic collisions with atomic electrons in the gas. For instance, across $1\ \mathrm{cm}$ of argon under NTP, there are on average 35 collisions [3]. The amount of energy transferred in each collision varies, which gives rise to ionizing clusters of various sizes along the MIP path. This is illustrated in Figure 3.9(a), which shows a cross-sectional view of a MIP traversing a drift tube. The ionized electrons in the clusters need to drift different path lengths before they respectively cause avalanches and arrive at the anode

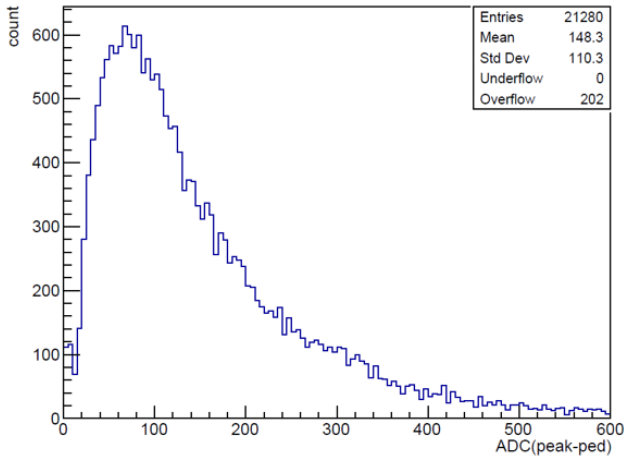
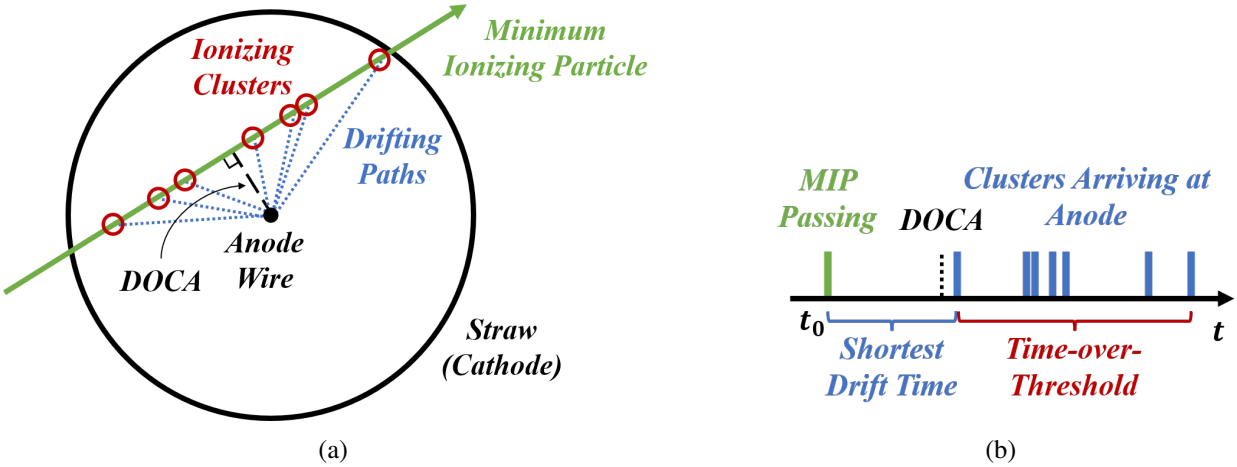


Figure 3.8: Example cosmic ray energy spectrum from a straw tube measurement. The spectrum takes the shape of a Landau-Gaussian convolution.



(a) (b)

Figure 3.9: MIP interactions in a straw tube. (a) In this cross-sectional view of the straw tube, ionizing clusters formed along the path of the MIP. The electrons from ionizations drift radially along their respective paths to the anode. (b) The processes are organized according to the time. Ionizing clusters formed along the MIP path drift along various path lengths and have different arrival times, which broaden the signals.

wire. Consequently, their respective drift time lengths also differ. On a time axis (Figure 3.9(b)), the ionizing clusters can be regarded as created at the same time t_0, since the MIP velocity is much larger than the electron drift velocities. Hence, the ionizing clusters have different arrival times. The shortest drift time is limited by the distance of closest approach (DOCA), which is the shortest distance between the MIP path and the anode wire. The span of the arrival times is determined by the drift paths difference between the furthest and closest ionizing clusters to the wire, embedding information regarding the geometry of the initial MIP path. Since the signals originated from multiple clusters are summed to produce a final signal output on the wire, this span of arrival times broadens the width of the final signal pulse, or what can be measured as the

time-over-threshold (TOT). Such timing information, along with the total charge in the signal (which is dependent on the MIP path length in the gas volume), can assist the precise reconstruction of the initial particle trajectory through the detector.

3.1.5.3 Highly Relativistic Electrons

In the Mu2e experiment, the Tracker measures $\sim 100\ \mathrm{MeV}$ electrons. The detector response to the highly relativistic electrons bears resemblances to that of the MIPs.

There are two relevant types of electron energy loss in materials in the pertinent energy range. The first one is again the ionization loss. The Bethe-Bloch formula is not applicable to calculate the electron collision stopping power, due to the difference in kinematics, as well as the fact that the collisions occur between identical particles. Instead, it should be calculated from the first moment of the Møller cross-section as [18]:

$$\left\langle -\frac{\mathrm{d}E}{\mathrm{d}x} \right\rangle = 4\pi N_A r_e^2 m_e c^2 \rho \frac{Z}{A}\frac{1}{\beta^2}\left[\frac{1}{2}\ln\frac{m_e c^2\beta^2\gamma^2\{m_e c^2(\gamma-1)/2\}}{I^2} + \frac{1-\beta^2}{2} - \frac{2\gamma-1}{2\gamma^2}\ln 2 + \frac{1}{16}\left(\frac{\gamma-1}{\gamma}\right)^2 - \frac{\delta}{2}\right]. \tag{3.20}$$

Despite the different expression, the electron energy loss due to ionization takes a similar trend to that of a MIP, as shown in Figure 3.10 [20]. The collision stopping power first decreases rapidly to a minimum and then stays mostly constant.

A second contribution comes from radiative loss through bremsstrahlung, which is the electromagnetic radiation emitted when electrons are deflected and decelerated by nuclei and electrons in the materials. This component increases almost linearly with the incident electron energy, which is also presented in Figure 3.10. At low energies, the ionization loss is significant, while the radiative loss dominates the high energy region. The point where the two contributions equal to each other is called the critical energy. For an 80%/20% Ar-$\mathrm{CO_2}$ under NTP, the critical energy is about $50\ \mathrm{MeV}$. The bremsstrahlung loss of $\sim 100\ \mathrm{MeV}$ electrons is about twice as much as the ionization loss.

In such a case, the electron energy loss distribution in thin absorbers and gases also observe the Landau distribution like the MIPs, except the corresponding mean value is the sum of the two energy loss contributions instead. Due to the ionization contribution, the electrons also leave ionizing clusters along their paths. The discussions regarding the timing of the ionizing clusters still hold true.[9]

[9]There is one small subtlety regarding the electron path. Due to its small mass, low energy electrons can be easily deflected by atomic electrons and especially nuclei, and change their directions. The highly relativistic electrons are less susceptible, as the outgoing electron is focused in the original direction of motion when translating the corresponding cross-section to the lab frame. For the $\sim 100\ \mathrm{MeV}$ electrons, most of the deflections are small.

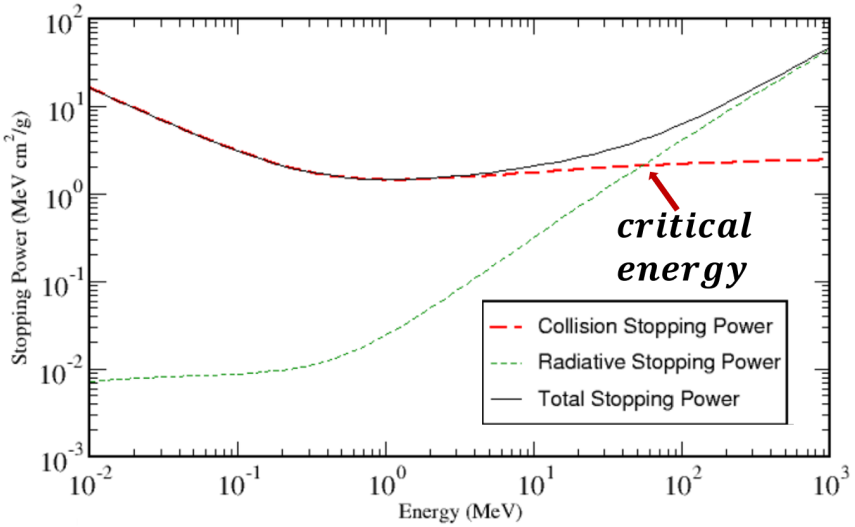


Figure 3.10: Electron stopping power in the 80%/20% Ar-CO_2 mixture. Data come from the ESTAR database of the National Institute of Standards and Technology [20]. Both contributions from collision loss (ionization) and radiative loss (bremsstrahlung) are present.

The bremsstrahlung interactions, however, do give rise to a unique feature in the detector response to the ~ 100 MeV electrons. In specific, the radiative energy loss proceeds in a discrete manner that is different from the quasi-continuous ionization loss along the path of the initial electron. Few bremsstrahlung interactions occur but each time a fraction of the initial energy is emitted through bremsstrahlung photons [21]. Most of the time a straw tube would not see any signal from bremsstrahlung interaction since the detector contains very little matter and the mean radiative energy loss is low, but there can be rare occurrences of large fluctuations in signals if bremsstrahlung interactions do occur and part of the emitted energy is captured in the detector.

The takeaway from the above comparisons between the ~ 100 MeV electrons and the MIPs is that most of the time the electrons behave very similarly to the MIPs. The more available MIPs (cosmic-ray muons, high-energy proton beams, etc.) can be used as substitutes to study detector behaviors without creating much discrepancy.

3.2 Tracker Panel Details

Despite following the same detection principles as the first gaseous ionization detectors dated back to the early 20th century [2], the straw tube arrays of the Mu2e Tracker have come a long way in design and fabrication to satisfy the precision required by the experiment.

The straw tubes used in the Mu2e Tracker (Figure 3.11 [22]) are 5 mm in diameter and between $0.33 \sim 1.17$ m in lengths [23]. The straws are wound by two layers of 6 μm-thick metallized Mylar with a 3 μm layer of adhesive in between. The 15 μm-thick straw wall is the



Figure 3.11: (Top) A picture of the Mu2e straw tube comparing with a pencil [22].

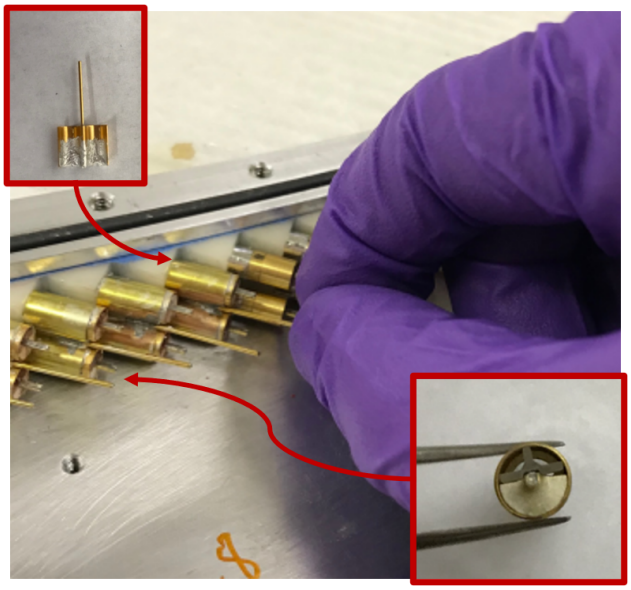


Figure 3.12: (Right) Straws fully assembled in the Tracker panel. The bottom-right pad shows the structure used to reinforce the anode pin. The small brass component ("grounding clip") in the top-left pad is glued in place using silver epoxy to provide reliable ground connections to the cathodes of a pair of neighboring straws.

thinnest among all large-scale straw tube trackers to date. This minimizes the amount of materials in the Tracker, which reduces the total amount of energy loss of the measured electrons. In addition, it reduces the probability of large deflections in the electron paths which helps with the track reconstruction. Thus, the experiment can achieve a better momentum resolution. The straw tube anodes are made of gold-plated tungsten wires that are $25\ \mu\mathrm{m}$ in diameter. Both the straws and the anode wires are tensioned, and the wires are work-hardened during manufacturing to prevent sagging [24].

The anode wires are aligned to the panels with a precision better than $25\ \mu\mathrm{m}$ in the radial direction and about $50\ \mu\mathrm{m}$ in the perpendicular direction, in order to achieve an ideal electric field and to ensure good spatial resolutions. In addition, all panels are x-ray scanned after manufacture so that the exact positions of the wires of all straw tube channels are measured and documented for later reconstruction use. Figure 3.11 also shows the termination structure that hold the wire in place. On each end of the straw, a brass tube is attached by silver epoxied to the straw to reinforce the mechanical strength [23]. A Kapton sleeve is placed inside the brass tube for electrical isolation, and an injection-molded plastic insert (the white plastic part in the picture) is attached inside the sleeve. The insert contains a semicylindrical duct that allows gas to flow in and out of the straw tube. There is a groove along the axis of the insert, and a U-shaped brass anode pin is attached to the end. The anode wire is epoxied to the groove and soldered in the anode pin to prevent slippage.

Additional reinforcements to the straw terminations are assembled after the straws are attached to the panels and the anode wires are installed in place. On each end of the straw, a T-shaped pin protector (bottom-right of Figure 3.12) covers the groove in the plastic insert and

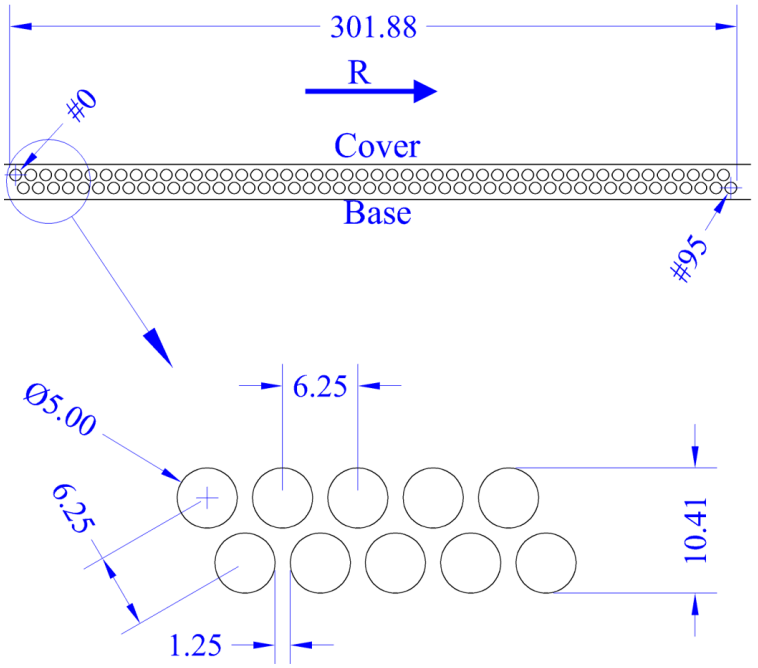


Figure 3.13: Straw arrangement in a tracker panel [25].

prevent anode pin from snapping. The pin protector is epoxied to the insert and an additional brass ring bonds them together. A ground clip (top-left of Figure 3.12) is silver-epoxied to a pair of neighboring straws on the brass tubes and the brass rings, and provide a shared ground connection for the straws. The anode pin pair and the grounding clip (through the pin attached to the clip) feed signals to a shared pre-amplifier (preamp) board, which will be introduced later with the front-end electronics (FEE). The fully assembled straws inside a Tracker panel is shown in Figure 3.12. I was involved in standardizing and optimizing the procedures to install the reinforcement structures, so they can be assembled properly in a timely fashion during mass production.

The 96 straw tubes on each panel are arranged in two staggered layers for better tracking.[10] The detailed spatial arrangement of the straws are presented in Figure 3.13 [25]. The channels are numbered from 0 (the longest) to 95 (the shortest) starting from the radially innermost straw. Neighboring straws are separated by $1.25\ \mathrm{mm}$ for manufacturing tolerance, which also allows straw expansion under gas pressure [23].

The Mu2e Tracker panels uses a gas flow of 80%/20% Ar-$\mathrm{CO_2}$ mixture. The pressure inside the gas volume is $1\ \mathrm{atm}$. As the Tracker reside in a vacuum environment inside the Detector Solenoid (which again minimizes the influence of electron scatterings on the trajectories), the Tracker panels, especially the straws, need to endure the pressure difference. An average leak rate limit of $0.014\ \mathrm{cm^3/min}$ under standard temperature and pressure per panel is required [24].

[10] In a single straw, it cannot be determined which side of the wire a particle traverses. Two layers of straws reduce the degree of freedom and resolve the ambiguity when the hits across the full Tracker are fit by a Kalman filter fit [23].

The nominal operating voltage of the straw tube channels is 1450 V. From an earlier study on a prototype of the Tracker, the straw tube gain was reported to be 1.25×10^4 at 1250 V, and 7×10^4 at 1425 V [26]. At 1450 V, the gas gain is approximated to be 1×10^5 from the Diethorn's formula (Equation 3.6).[11]

3.3 Tracker Front-End Electronics

The raw analog signals from the Tracker straw tube channels need to be amplified, digitized, and packaged before they can be accessed by the Mu2e data acquisition (DAQ) system. The Tracker front-end electronicss (FEEs) are specifically designed to accomplish those goals. The Tracker FEE consists of multiple printed circuit boards (PCBs) as shown in Figure 3.14 [29]. Signals from the straw tube channels are first read out from both ends by the pre-amplifiers (preamps). The figure

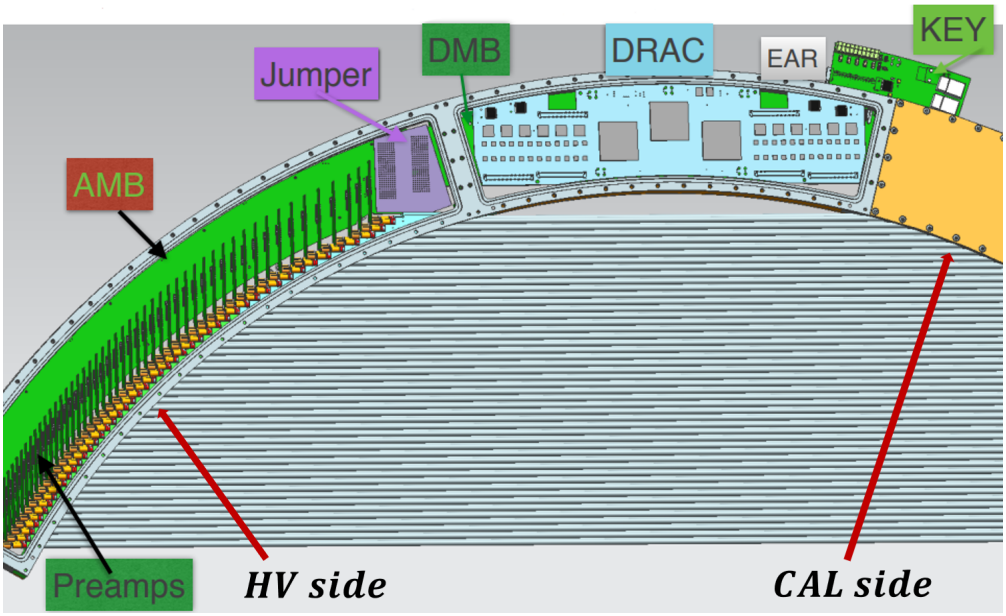


Figure 3.14: An overview of the Tracker FEE [29]. The preamps and the Analog Motherboard (AMB) on the other side of the straws are not shown in the figure.

[11]There is a pair of Diethorn's parameters available for the 80%/20% Ar-CO_2 mixture in [27], which gives $E_{\min}(1\text{ bar},\ 295\text{ K}) = 22\text{kV/cm}$ and $\Delta V = 49\text{V}$ from an earlier ATLAS study. However, the values were mentioned to contain about 20% systematic errors along with statistical errors, and a direct insertion of the values to Equation 3.6 yields gains that are about an order of magnitude smaller than the gains measured in the Mu2e Tracker prototype, so the calculated value from this piece of literature is not presented here. Instead, from the Mu2e prototype gains, a pair of Diethorn's parameters can be estimated that agrees both with the data and with the literature values (up to errors), which is $E_{\min}(1\text{ bar},\ 295\text{ K}) \simeq 18\text{kV/cm}$ and $\Delta V \simeq 41\text{V}$. The estimated gain for the 1450 V operating voltage is calculated using these values. Diethorn's formula can provide good estimations of gain changes when gas conditions fluctuate, but due to the linear approximation made in its derivation, extra caution should be taken when adapting the Diethorn's parameters obtained from one detector to others with different geometries (as the discrepancies rooted in the linear approximation are "absorbed" in the parameters). An alternative nonanalytic approach to estimate the gain of a detector geometry utilizes simulation packages such us Garfield++ from CERN [28].

here only shows the preamps on the HV side of the panel. On each side, there is an AMB and a Jumper board that route the signals from the preamps to the Digital Motherboard (DMB) in the center, then to the Digitizer Readout and Assembler Controller (DRAC) (mounted on top of the DMB) to be processed and temporarily stored. In addition, the AMBs and the DMB handle low voltage distribution; the AMB on the HV side is responsible for distributing the high voltage to the straw anode wires (hence the name "HV side"). Several sensors are placed at various location on the AMBs and the DMB to monitor environment variables including temperature, pressure, and humidity. All the PCBs mentioned so far reside in the outer part of the panel. The EAR, on the other hand, is a feedthrough that connects the DMB inside the panel volume and the KEY outside. The low voltage power supply is fed into the panel through the KEY. Besides, the KEY contains an optical fiber link and a JTAG interface for communication.

The FEE components were selected to sustain high levels of radiation and tested extensively with about 10 times of the maximum dosage expected throughout the experiment lifetime [30]. Almost all active components in the FEE are concentrated on the preamps and the DRAC, which will be introduced in more details next. An explanation of the signal flow in the Tracker FEE will also be provided.

3.3.1 Pre-Amplifiers

As mentioned, the straw tube channels are read out from both ends, and two neighboring channels are connected to the same preamp (Figure 3.15 [29]). So each Tracker panel contains 48 preamps on the HV side and another 48 on the other side (which is referred to as the CAL side). The preamps are installed vertically on the AMBs. All preamps are parylene-coated to prevent high voltage discharges between neighboring ones that are compactly installed.

Current signals from the straw tubes are turned into voltage signals by the preamps. The signals are amplified and shaped. The voltage gain for each side of a channel (which is different from the gas gain in the avalanches!) can be individually set through a control signal from the DRAC. Moreover, a bias voltage is applied to the signals, whose value is also adjustable for each side of a channel. As the signal triggers are later determined when the biased signals pass through zero-crossing comparators on the DRAC, the bias voltages effectively set the triggering thresholds for the straw channels.

The preamps on the two ends of the straws use the same circuitry for signal processing, but they are not exact mirror image copies to each other as additional functions are performed. The preamps on the HV side distribute the high voltage supply to individual straw tubes. It contains a simple custom-built high voltage disconnection mechanism consisting of a spring wire and a heating circuit.[12] If a straw tube develops some high voltage problem over time during data collection, it

[12]The experiment chose not to use a commercially available fuse mainly out of the concern of halogen contamination,

would be implausible to dismantle the entire detector merely to fix a single channel. In case of a straw channel being deemed unrecoverable, a signal will be issued electronically to activate the heating circuit. The spring wires are installed in a tensioned manner with the ends closer to the high voltage inputs bonded by low-temperature solder. Once melted, the spring wire will disconnect the malfunctioning channel (and its neighboring channel on the same preamp) from the high voltage. The preamps on the CAL side, on the other hand, contain a circuitry that can inject calibration pulses to the channel (hence "CAL side"). This feature allows the channel readout electronics to be conveniently tested without a high voltage supply.

3.3.2 The Digitizer Readout and Assembler Controller

The Digitizer Readout and Assembler Controller (DRAC) is the brain of a Tracker panel. All straw readouts are channeled here for digitizing, packaging, and temporary storage. It also controls panel operations. Figure 3.15 [29] shows a picture of the DRAC board. The arrays of analog-to-digital converters (ADCs) and comparators, as well as the DDR3 memories, are marked in the picture. The three large chips in the center are Microsemi PolarFire Field-Programmable Gate

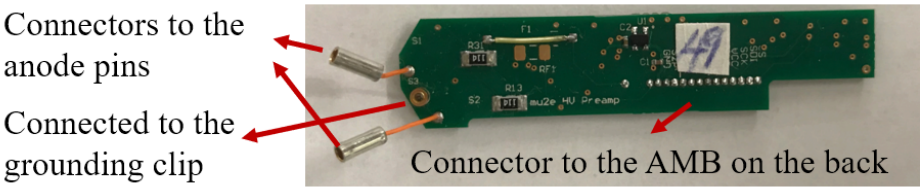


Figure 3.15: A picture of the Mu2e Tracker preamp [29]. Each preamp reads out two channels on one end. The connectors to the straw terminations are marked.

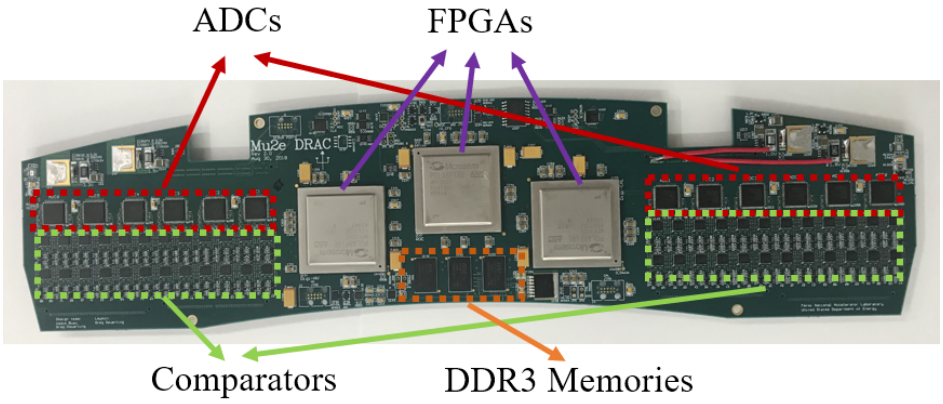


Figure 3.16: A picture of the DRAC board [29].

which can lead to problems of faster detector aging. The size of the fuses also pose an additional problem as the preamps are installed so compactly that there is not enough clearance.

Arrays (FPGAs). The two outer ones, which are referred to as the DIGI FPGAs, are responsible for controlling data output, buffering data, and assembling data packages. Each of them takes care of 48 straw channels. The center FPGA is the Readout Controller (ROC), which manages communications, monitors slow control variables, and controls panel operations.

Figure 3.17 [23] shows the signal flow in the Tracker FEE. The signals from both ends of a straw are processed by the preamps and then routed to the DRAC through the PCBs. In the DRAC, the two biased signals are individually fed to zero-crossing comparators, which yield square pulses when the signals surpass their respective triggering thresholds. The square pulses are directed to

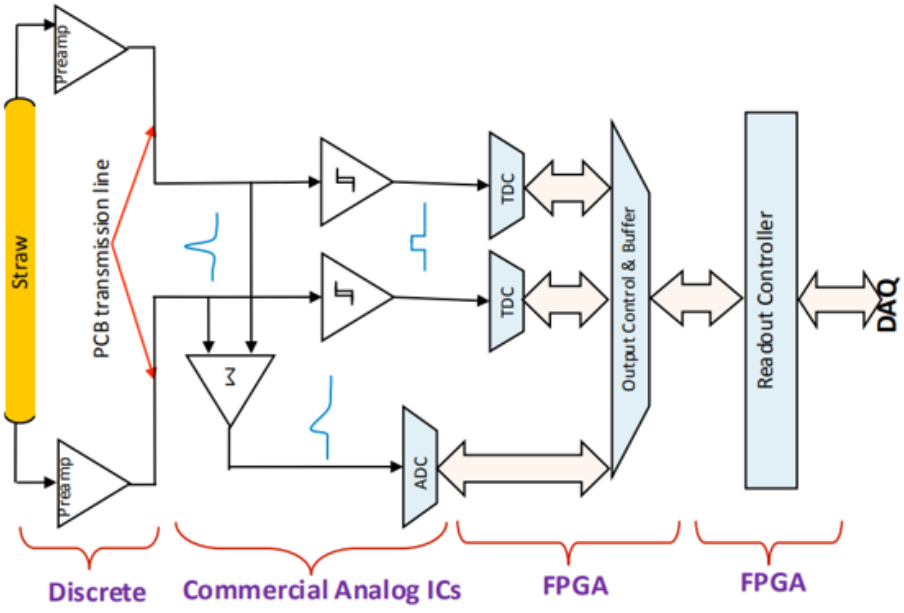


Figure 3.17: A chart of the signal flow in the Tracker FEE [23].

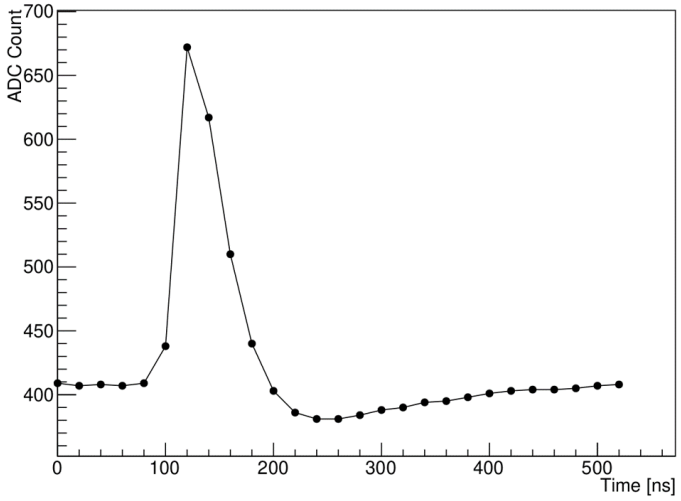


Figure 3.18: Example digitized signal from a Tracker straw channel. The signal is the sum of outputs from both ends of a straw and is digitized at 50 MHz.

firmware-based time-to-digital converters (TDCs) that reside in the DIGI FPGAs, which digitize the trigger timings, including the arrival times of the triggers and the TOTs, at a rate of $20\ \mathrm{ps}$ per count; the timing resolution is about $70\ \mathrm{ps}$ [31]. On the other hand, the signals from the two straw ends are added by an integrator. The voltage signal sum is then digitized by a 10-bit ADC at a rate of $50\ \mathrm{MHz}$ and the data are streamed to the corresponding DIGI FPGA. In data taking, a hit requires both ends of a straw channel to be coincidentally triggered. Only then the DIGI FPGA assembles a data packet for the hit using the TDC and the ADC information. This suppresses false triggers by random electronic noises. Figure 3.18 shows an example hit from a straw channel in a cosmic dataset. Then the data packets are transferred to the ROC, and temporarily stored in the DDR3 memories for later access by the DAQ system. There is a total of $8\ \mathrm{Gb}$ of memory space available on each DRAC. More details of the DAQ system will be discussed together with the firmware in the next chapter.

3.4 Requirements on Tracker Performance

There are requirements that the Tracker must satisfy to ensure the success of the experiment [32]. Some of the operational requirements have already been discussed, including the abilities to operate in an ambient vacuum ($< 10^{-4}\ \mathrm{Torr}$) with a low leak rate, to sustain high levels of radiation, to hold high voltages without breakdowns, and to function in a magnetic field of $1\ \mathrm{T}$. More importantly, it must meet the following two tracking requirements.

First, the Tracker must have an acceptance of approximately 20% for conversion electrons. Simulations, as well as tests on the earlier prototype, have demonstrated a nearly full acceptance [22, 25].

In order for the experiment to reach its designed sensitivity, the most crucial requirement of the Tracker is to achieve a momentum resolution of better than $180\ \mathrm{keV/c}$ for a $105\ \mathrm{MeV/c}$ electron [32]. From simulations, this is translated to spatial resolution requirements on a single straw that the longitudinal resolution along the straw length is $\sim 4\ \mathrm{cm}$ and the transverse resolution of drift path lengths is $\sim 200\ \mu\mathrm{m}$ [22]. The resolution studies performed during the Vertical Slice Test will be discussed in detail in Chapter 5.

References

[1] P. Gianotti, *et al.*, J. Instrum. **8** (8), T08001.

[2] W. R. Leo, *Techniques for Nuclear and Particle Physics Experiments* (Springer-Verlag, 1994).

[3] W. Blum, W. Riegler, and L. Rolandi, *Particle Detection with Drift Chambers* (Springer-Verlag, 2008).

[4] M. Deile, *et al.*, Nucl. Instrum. Methods Phys. Res. A: Accel. Spectrom. Detect. Assoc. Equip. **449**, 528 (2000).

[5] T. Muller, *et al.*, in *2008 IEEE Nuclear Science Symposium Conference Record* (2008) pp. 2284–2287, arXiv:0908.1562 [ins-det] .

[6] T. Zhao, *et al.*, Nucl. Instrum. Methods Phys. Res. A: Accel. Spectrom. Detect. Assoc. Equip. **340**, 485 (1994).

[7] R. Bernstein, Front. Phys. **7**, 1 (2019).

[8] M. Matoba *et al.*, IEEE Trans. Nucl. Sci. **NS-32**, 541 (1985).

[9] D. Pinci and A. Sarti, *Study of the MWPC gas gain behaviour as a function of the gas pressure and temperature*, Tech. Rep. (CERN LHC, 2005) [CERN-LHCb-2005-079].

[10] W. Shockley, J. Appl. Phys **9**, 635 (1938).

[11] S. Ramo, Proc. IRE **27**, 584 (1939).

[12] J. D. Jackson, *Classical Electrodynamics* (Wiley, 1999) pp. 35–37;40–43;52;295–351.

[13] Z. He, Nucl. Instrum. Methods Phys. Res. A: Accel. Spectrom. Detect. Assoc. Equip. **463**, 250 (2001).

[14] P. Horowitz and W. Hill, *The Art of Electronics* (Cambridge University Press, 2015) p. 1116.

[15] M.-M. Bé *et al.*, *Table of radionuclides. Vol. 1 A=3 to 244* (BIPM, 2006) pp. 5–9.

[16] F. Sauli, "Gaseous Detectors Fundaments," (Excellence in Detectors and Instrumentation Technologies Workeshop, CERN, Geneva, Switzerland, 2011-2-1), `https://indico.cern.ch/event/124394/`.

[17] J. H. Scofield, *Theoretical Photoionization Cross Sections from 1 to 1500 keV*, Tech. Rep. (Lawrence Livermore Laboratory, 1973) [UCRL-51326].

[18] P. Zyla *et al.* (Particle Data Group), Prog. Theor. Exp. Phys. **2020**, 083C01 (2020), and 2021 update.

[19] P. V. Vavilov, Sov. phys. JETP **5**, 749 (1957).

[20] M. J. Berger *et al.*, "Estar, pstar, and astar: Computer programs for calculating stopping-power and range tables for electrons, protons, and helium ions," (2017, version 2.0.1), [Online] Available: `http://physics.nist.gov/Star` [2022-6-2]. National Institute of Standards and Technology, Gaithersburg, MD.

[21] S. M. Seltzer and M. J. Berger, At. Data Nucl. Data Tables **35**, 345 (1986).

[22] R. Bonventre, "Searching for muon to electron conversion: The mu2e experiment at fermilab," (Lake Louise Winter Institute 2020-2-14), `https://indico.cern.ch/event/846070/`.

[23] L. Bartoszek, *et al.* (Mu2e Collaboration), *Mu2e Technical Design Report*, Tech. Rep. (Fermilab, 2015) arXiv:1501.05241 [physics.ins-det] .

[24] K. Heller, "Mu2e Tracker Panel Assembly," (DOE IPR 2019 2019-12-11), mu2e-doc-30538-v6.

[25] A. Mukherjee, "Mu2e Tracker," (DOE IPR 2017 2017-07-25), mu2e-doc-11567-v3.

[26] R. Bonventre *et al.*, "Gain Variations in the 8-Straw Tracker Prototype," (2016-07-02), mu2e-doc-7702-v1.

[27] M. Aleksa and W. Riegler, *Non-Linear MDT Drift Gases like Ar/CO_2*, Tech. Rep. (CERN LHC, 1998) [ATL-MUON-98-268].

[28] H. Schindler, "Garfield++," (2022, version 2022.2), [Online] Available: `https://garfieldpp.web.cern.ch/garfieldpp/` [2022-6-10]. CERN, Geneva, Switzerland.

[29] V. L. Rusu, "Mu2e Tracker Front End Electronics," (DOE IPR 2019 2019-12-10), mu2e-doc-30505-v3.

[30] M. Kargiantoulakis, "Radiation Studies for the Mu2e Tracker FEE," (Tracker Preamp Mini-CRR 2021-05-17), mu2e-doc-11381-v2.

[31] R. Bonventre, "Tracker Firmware Status," (Tracker Meeting 2016-11-21), mu2e-doc-8336-v1.

[32] R. Bernstein and A. Mukherjee, "Tracker Requirements Document," (2016-04-07), mu2e-doc-732-v22.

Chapter 4 Firmware-Based Data Acquisition and Tracker Operation

This chapter will first introduce the Mu2e data acquisition (DAQ) system and its interactions with the Tracker DAQ. The implementation of the firmware-based Tracker control and operation will then be reviewed. The highlight will be on the work carried out on an alternate control and operation system specifically designed for testing.

4.1 Experiment and Tracker Data Acquisition

The Mu2e DAQ system is based on a "streaming" readout approach, which means *all* detector data from the experiment are digitized and zero-suppressed in respective FEE, and then transmitted off the detectors for processing and storage [1]. Such approach leads to a high data rate in the DAQ system, but provides more flexibility in data selection and analysis. Besides, the required DAQ architecture can be relatively simple [1].

Figure 4.1 [2] shows the overall architecture of the Mu2e DAQ system. The left blocks are the Readout Controllers (ROCs) in various detectors. The middle block contains the online part of the DAQ system, which mainly consists of the Run Control Host, 40 DAQ servers, the Detector Control System (DCS), and the Event Building Switch. To handle the high data rates, this part of the DAQ system relies heavily on firmware operations. The right block includes the offline components of the DAQ system for data storage and offline processing.

The experiment DAQ processes are coordinated by the Run Control Host operating some experiment-defined Run Plan. Throughout an active spill (the first ~ 43 ms in the 48 ms bunch extraction cycle discussed in Chapter 2), the experiment receives RF Zero-Crossing Markers from the Accelerator that are synchronous to the 1695 ns proton pulse cycles (a.k.a., the event windows[1]) [3]. Based on these markers, The Command Fan-Out (CFO) module inside the Run Control Host fabricates a 40 MHz system clock and encodes Event Window Markers (EWMs) in the system clock to signify the beginnings of the event windows. The CFO then distributes the encoded system clock, along with run control packets, to Data Transfer Controllers (DTCs) in the DAQ servers. The DTCs further distribute the encoded clock to the ROCs in the detectors, where

[1]Event windows are also assigned during the time periods between spills and when the experiment receives no beam batches (i.e., "off-spill"), except that the event window length is much longer due to the lower data rate. The exact length has not been decided, but there was a suggestion to use a 14 μs off-spill event window.

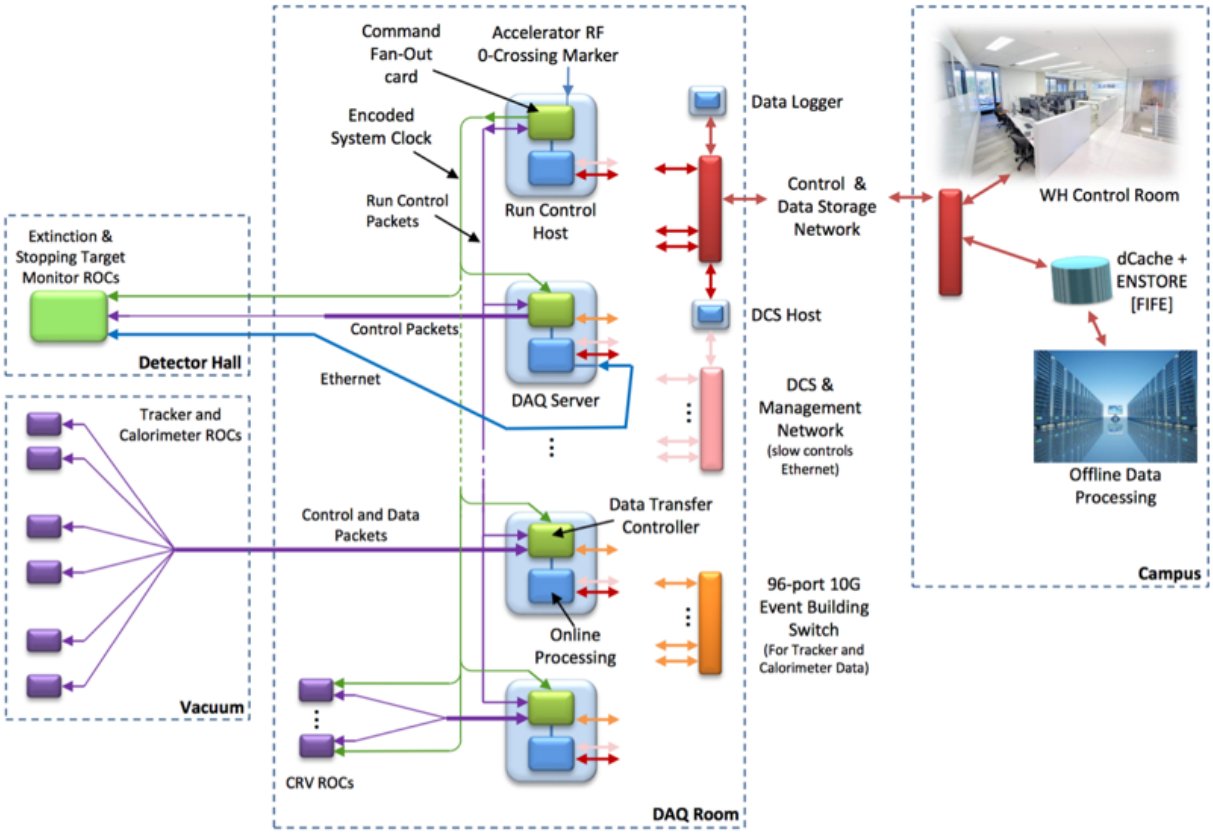


Figure 4.1: Architecture of the Mu2e DAQ system [2].

the EWMs are recovered with fixed latency relative to the initial RF Zero-Crossing Markers, and are utilized in the local ROCs to differentiate data collected during contiguous event windows. Specifically in the case of the Tracker, a DDR3 memory address is generated at the beginning of each event window. The corresponding memory space is assigned to store the Tracker hits received during that event window [4].

Data readouts from the ROCs to the DAQ system are initiated by Data Requests. The Data Requests are configurable through the CFO as defined by the Run Plan, but tentatively they are issued to the Tracker and Calorimeter ROCs after each event window through the DTCs [2]. Meanwhile, the CRV Data Requests are issued by software as will be explained in a moment. The ROCs then respond to the Data Requests and send correspondent data to the DTCs. Data from multiple DTC sources are forwarded through the Event Building Switch to a certain DTC in a round-robin manner, then pre-processed and passed to the Online Processing module in the same DAQ server. If a trigger decision is made, the corresponding CRV data would also be read out from the ROCs through a software-issued Data Request. Event data from all detectors are aggregated at the Data Logger, and transferred to long-term data storage where they can be accessed for further offline studies. Calculations showed an expected data rate of 35 GBps (with 30% overhead) from

the ROCs, and about 7 PB per year at the storage [2].[2]

Apart from detector data, the DTCs are also responsible for acquiring slow control variables from the detectors. In the case of the Tracker, these include panel temperatures, pressures, humidities, voltages and currents at selected PCB nodes, high voltages, and channel threshold settings from all panels. The data are managed by the DCS Host and logged in long-term storage as well. Additionally, they are available to the Control Room for real-time monitoring. Additional detector controls and operations are handled through DCS commands via DTCs. Again in the context of the Tracker, these include panel configuration, calibration using preamp-generated pulses, and channel high voltage disconnection when needed.

4.2 Alternate Panel Control and Operation System

The formal Mu2e DAQ system is still under active development. In the meantime, needs of Tracker panel FEE testing and firmware development urge for an alternate panel control and operation system. Such a system is also useful for quality checking the mass production panels, making diagnostics for panel issues, and assisting with the detector commissioning. The system needs to be easy to access, meaning it can be set up without the DAQ servers; yet it should provide efficient and robust panel control functions. It also should be easily migratable to the DTC with minimal changes when the formal DAQ system is online.

The alternate system is minimalistic compared to the formal one, as shown in Figure 4.2. It shares the same firmware-based components in the DIGI and ROC FPGAs as the formal DAQ. However, instead of connecting to the DTC via the optical fiber link on the KEY (Section 3.3), the ROC is connected to a Raspberry Pi device through a serial link based on the Universal Asynchronous Receiver-Transmitter (UART) protocol via JTAG connectors.[3] Here, the Raspberry

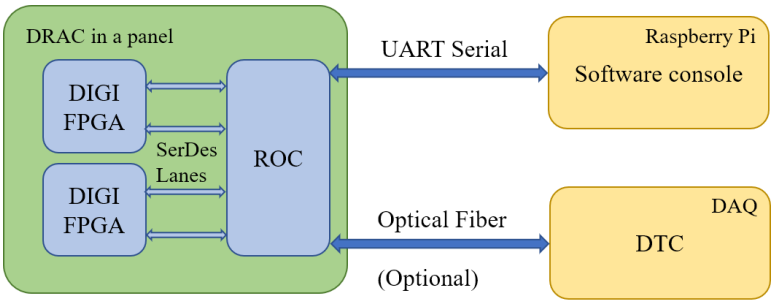


Figure 4.2: A block diagram of the alternate panel control and operation system.

[2]The calculation assumed a 80% experiment up-time. A total rejection ratio of $\sim 125:1$ is required during pre-processing, event building, and filtering [2].

[3]Apart from the JTAG connector on the KEY, there is another one reserved on the DRAC board so the system can run solely on the DMB-DRAC assembly without the rest of the panel.

Pi operates a software control console. It issues control commands to the ROC and reads out its responses. Only one panel (ROC) is connected here but the system allows more at the same time.

In the alternate system, the optical fiber connection to the DAQ server is optional. If connected, the ROC can still use the system clock and the EWMs from the DTC, and read out to the DTC from the DDR3 memories as in the formal system. This is the configuration used during the Vertical Slice Test (VST) which will be discussed in the next chapter. On the other hand, the ROC can be configured to generate a system clock and emulate EWMs of its own without the optical fiber link. In this case, straw hit data can be read out through the serial link. This configuration is more convenient for the mentioned testing and quality check purposes.

Next we will take a closer look at various parts of the alternate panel control and operation system. The firmware designs of DIGI and ROC FPGAs will first be reviewed, then the micro-controller command set driving the ROC operations will be presented. The details of the software console will follow.

4.2.1 Tracker Firmware

As mentioned, the Tracker firmware is the same for the formal DAQ and the alternate one. The two DIGI FPGAs operate on the same firmware design with only differences in chip pin-out (and hence the routing), which handles data buffering and packaging. The ROC FPGA takes care of communications and realize the control functions. Microsemi provides a design software named Libero, where the firmware is built up with custom modules written in Very High-Speed Integrated Circuit Hardware Description Language (VHDL) and Verilog, along with provided FPGA Intellectual Property (IP) Cores.

4.2.1.1 DIGI Firmware

Figure 4.3 presents an overview of the DIGI firmware. The system clock relayed from the ROC is used; an IP Core fabricates local clocks with various frequencies and phases that are locked on to the system clock, and distributes them to other modules. In addition, The ROC supplies the values for an array of control registers. These registers configure behaviors of other modules. For instance, there are registers that trigger reset processes of various modules, while some others mask active channels for readout, or change the length of output data packets. The ROC is also capable of reading back the register values, so some control registers capture outputs of other modules and are used to monitor their status.

Each DIGI FPGA is responsible for processing data from 48 straws, so it receives data from 48 ADC channels and 96 comparators. The ADC data are transferred from the ADC chips through a Serializer/Deserializer (SerDes) interface. A Bit Alignment Module monitors and adjusts the capture

phases of the incoming signals using an IP Core that monitors the signal eye diagrams, ensuring the quality of the SerDes data capture. In addition, it is responsible for controlling bit slippage, which adjusts the starting bit of a word in the serialized data stream (illustrated in Figure 4.4). This step is crucial to guarantee the ADC data are properly captured in the DAQ system.

On the other hand, the comparator signal pairs are fed to firmware-based TDC modules.

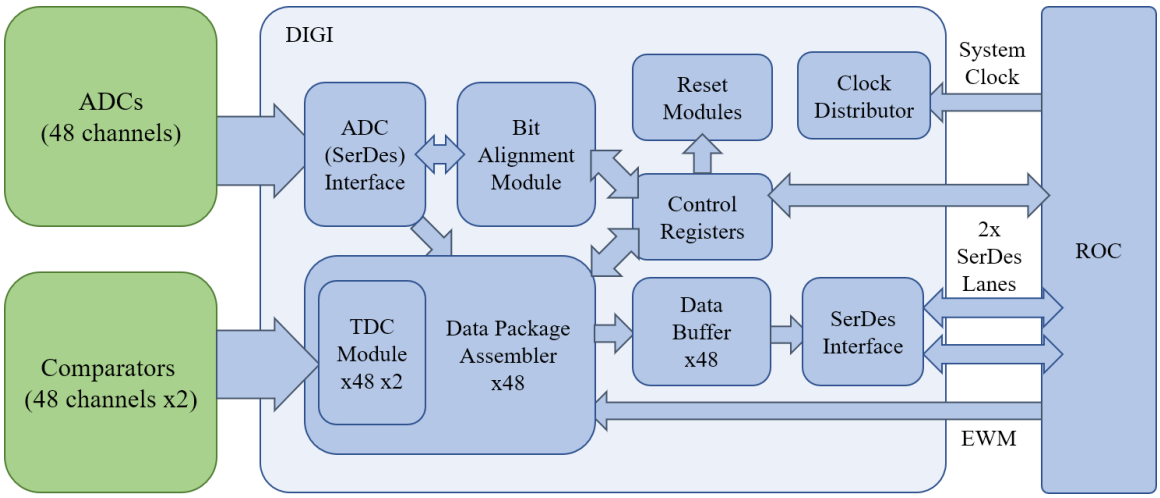
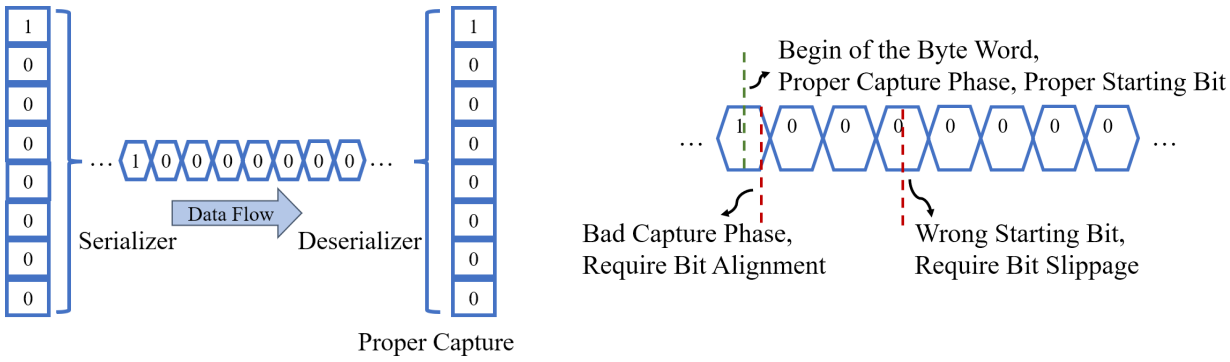


Figure 4.3: A block diagram of the firmware on the DIGI FPGAs.



(a) (b)

Figure 4.4: Bit alignment and bit slippage in the SerDes interface. (a) In a SerDes interface, some parallel data from the first chip are serialized before transmission. The data are transmitted to a second chip via a single differential pair. The second chip deserializes the transmitted data and obtains the original parallel word if the transmission is properly restored. (b) A proper data capture on the Deserializer side requires a reasonable capture phase and a correct starting bit. The eye diagram monitor deals with the former (bit alignment) and a custom written module adjusts the latter (bit slippage). Failure to achieve either results in meaningless data.

These modules average multiple clock delay chains to achieve sub-100 ps time resolutions. A coincidence between the triggers from both ends of a straw initiates the data packet assembly process. The timing information and the ADC data are packaged together with a header including additional information (straw channel number, EWM number supplied by the ROC, etc.) of the straw hit by the Data Package Assembler of the corresponding channel. The data packets are temporarily stored in the data buffers and are streamed to the ROC through one of the two SerDes lanes between the FPGAs.

4.2.1.2 Readout Controller Firmware

A block diagram of the ROC firmware is shown in Figure 4.5. It packs more communication modules and contains a micro-controller to coordinate various operations.

The micro-controller of RISC-V architecture is embedded in the firmware design using an IP Core supplied by Microsemi. It reacts to the commands from the control console (or DTC in the future) and executes a series of actions accordingly, which include making simple calculations, reading or writing some registers in the DIGI or ROC FPGAs, configuring and controlling various parts such as Digital-to-Analog Converters (DACs) and calibration pulse generators in the preamps, and moving data among various blocks and interfaces. The schemes of operations are defined by a command set stored in a secure Non-Volatile Memory (sNVM), which retains its contents after

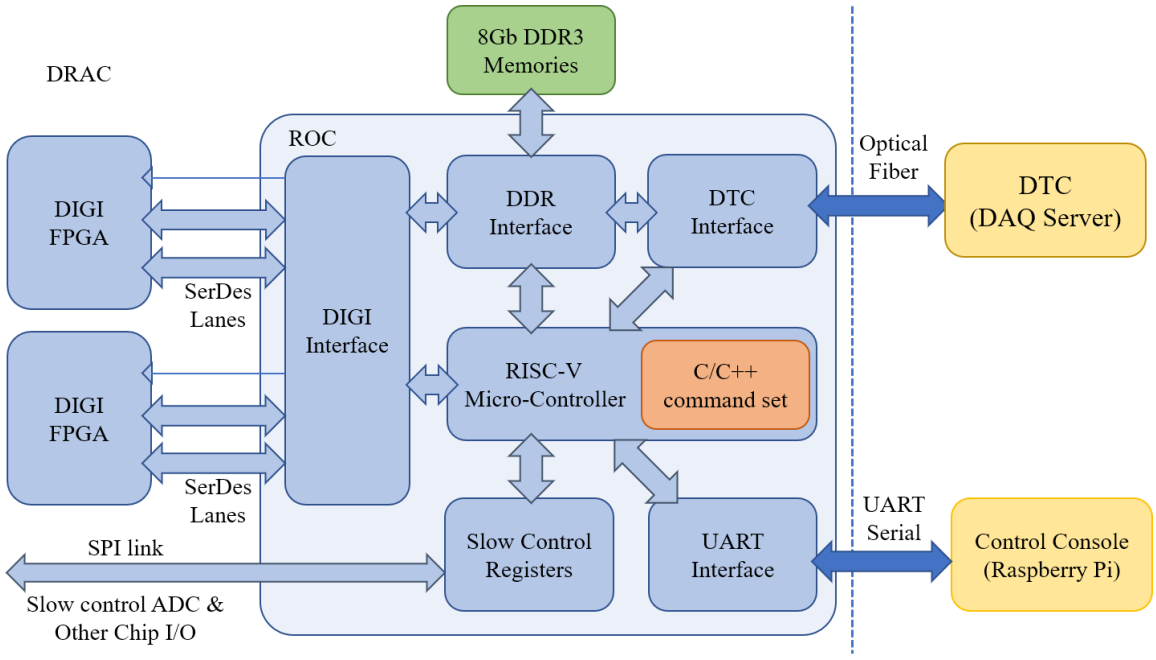


Figure 4.5: A block diagram of the firmware on the ROC.

power disconnections. The command set is coded in C/C++. The amount of work it took deserves a separate section of its own; it will be discussed in the following section.

When the DTC is configured as the destination, all data from the DIGI FPGAs are moved to the DDR3 memories for storage until the memories are filled to a set level. Like in the formal system, the DTC then issues Data Requests to acquire the memory contents. The data are transferred from the memories to the DAQ server and are written out as binary data files. On the other hand, if the DTC is disconnected, the data from the DIGI FPGAs are deposited in a small First-in, First-Out (FIFO) memory in the firmware. A register drains the FIFO 32 bits at a time and is read out by the control console through the UART link. This approach is slow compared to the optical fiber link and can only handle a small amount of data, but it is sufficient for test use. It does not involve the DDR3 memories in the process.

The slow control variables are mostly managed through the slow control registers. When prompted by the control console, the micro-controller requires data from sensors and ADC chips (which measures, for example, voltages across diode-based temperature sensors, those between certain PCB nodes, and those across known resistors for current measurements). The returned values are stored in respective registers and are read out through the UART interface. Meanwhile, some registers are utilized in the same manner as the control registers in the DIGI, which keep track of operation status of various modules.

From the architecture of ROC firmware, most of the modules will remain unchanged when the DAQ is moved to the formal system. The focus of the migration work will be rerouting the micro-controller inputs from and outputs to the UART interface towards the DTC interface, once a proper handshake protocol for slow control data is established.

4.2.2 Micro-Controller Command Set

As mentioned in the previous section, the C/C++ command set in the ROC firmware drives the operations of the micro-controller, and instructs its responses to the commands from the control console (or DTC in the future). Microsemi provides a software named SoftConsole that compiles the C/C++ codes into a hexadecimal source file that can be interpreted by the micro-controller and is stored in the sNVM.

The sNVM space on the FPGA used for the ROC is extremely limited. There is only a total of 56 kB available. Consequently, the codes have to be carefully structured and extensively optimized in order for the compiled product to fit in the given space. A modularized design principle was followed throughout the code development.

The command set mainly consists of three portions. The first part is comprised of utility functions that address communications and signal relays. Drivers for various interfaces and chips, along with various reading and writing protocols, are included in this category. For the UART

interface, Microsemi provides a stock driver allowing data reception and transmission. As for the standard chip-to-chip interfaces such as Serial Peripheral Interface (SPI) and Inter-Integrated Circuit (I^2C), there are drivers available along with the corresponding IP Cores or from open-source code repositories. Due to the complexity of its design, the DRAC also includes a few instances of SPIs daisy-chained to the output pins of some Input/Output (I/O) expanders. They are supported by a custom written driver that bit-bangs the pin outputs of the I/O expanders and mimics the behaviors of a regular SPI port. The chip drivers are then constructed upon these chip-to-chip communication interfaces. As chip operations are encoded in command sequences defined by the manufacturers, the chip drivers can be boiled down to writing command sequences and reading back returned values. In addition, there are functions reading and writing the (slow) control registers in the ROC and the DIGI firmware, which manage different modules and relay their operation status, as well as the slow control readouts.

The second part of the command set provides a framework that continuously monitors instructions from the UART interface, interprets the instructions, and react accordingly. In the codes, this is realized through an infinite loop in the main body of the program. Every microsecond, the micro-controller initiates a read from the UART interface and appends the returned values to a temporary buffer that serves as a command queue. As each of the issued console commands always consists of a communication verifier,[4] a uniquely assigned identification number, a command length count and optionally some arguments specifying parameters used in executing the operation, the program can determine whether a received command is complete. When it is, according to the provided command identification number, the program invokes a scheme describing the step-by-step operations in response to the command (which will be described as the third portion of the command set). When the operations are completed, the command buffer is updated to remove the executed command and the next one (if it is already received) is moved up the queue.

The rest of the command set comprises schemes of step-by-step routines. Included are instructions to initialize various interfaces and chips, as well as the operation schemes that fulfill the requests of the console commands. The individual steps in the schemes are mostly calling reading or writing modules of various objects described in the first part of the command set. If a group of actions is shared among multiple schemes, it is always packaged into a module for code optimization.

The essence of the command set is hereby summarized in Pseudo Code Snippet 4.1.

[4]The communication verifier is a pair of alternating-patterned bytes of 0xAA. When the UART communication malfunctions and loses alignment, the pattern is very likely to change, so it can be used to verify the serial communications.

Pseudo Code Snippet 4.1 Main Structure of RISC-V Micro-Controller Command Set

```
1: Define interface drivers
2: Define chip drivers
3: Define reading and writing modules for registers, etc.                    ▷ Portion 1
4:
5: Package repeated routines as modules                          ▷ For code optimization
6:
7: function MAIN( )
8:     Execute initialization schemes                                        ▷ Portion 3
9:     while True do
10:        Wait 1 μs                                                         ▷ Portion 2
11:        Read from UART interface, append to command buffer
12:        if Command complete then
13:            Extract verifier, command ID, command length, optional parameters,
14:        else
15:            Continue
16:        end if
17:        if verifier ≠ 0xAAAA then
18:            Report communication error
19:        end if
20:        if command ID = scheme #1 ID then
21:            Execute scheme #1 using optional parameters                   ▷ Portion 3
22:        else if command ID = scheme #2 ID then
23:            Execute scheme #2 using optional parameters
24:        else if ... then
25:            ...
26:        end if
27:        Remove executed command, move the next up the queue
28:    end while
29: end function
```

4.2.3 Software Control Console

The last remaining component of the alternate control and operation system is the software console, which is responsible for issuing the control commands and interpreting the returned responses from the panels. The software console contains a package of Python scripts running on the Raspberry Pi device.

The script package also follows a modularized design which gives it a lot of robustness. The most fundamental modules include communication utilities to send and receive data packets through the serial port, a command assembler, and a data interpreter. Following the same set of command identification number assignments, the command assembler turns command inputs to byte arrays that the micro-controllers can understand, while the data interpreter unpacks the returned

byte arrays according to commonly-agreed data organizing schemes assigned to the commands. Conversions and calculations are made if necessary (for example, voltages are converted from ADC readings, and temperatures are calculated from voltages across temperature diodes). In the case when straw hit data are returned, the data interpreter can write out the data to files.

A general-purpose ROC control script wraps the above fundamental modules and provides a human interface. Inputs to the script use combinations of command keys and parameter flags, which are processed by the command assembler and are sent out in a queue. Each command input expects a set of printouts parsed from some return package by the data interpreter. Input parameters are always repeated in the returned data as proofs of successful communications and correct command interpretations on the micro-controller's end.

Apart from the basic console, the package also contains automation scripts for certain tasks. For instance, there is a script developed for electronics diagnostics. It contains checks for data paths in the DRAC (through bit alignment and straw channel data readout tests), tests for preamp settings and operations (through finding appropriate thresholds settings, which involves toggling bias voltages and gains of the channels), noise level tests (for electronics noise or high voltage micro-sparks), and anode wire continuity tests (using the calibration pulses). Automatic error detections in the script quickly singles out channels with issues, and the test printouts provide information to diagnose the problems. A single execution of the program takes about fifteen minutes, as compared to hours if the same tasks are performed manually. This automation script helped to reduce the preparation time of a single panel from days to hours, which will be useful for conducting tests and diagnostics when commissioning mass-production panels.

4.3 Realization of Panel Operations

The panel control system performs a series of functions, including but not limited to what are listed in Table 4.1. This wide range of functions are realized by one or more levels of the alternate Tracker panel operation system, which can be in the firmware, in the micro-controller command set, or in the software console. Before answering the question of at which level of the system a function should be realized, a comparison among the three is constructive.

Each of the three levels in the system has its own strengths and faces its own challenges. The firmware operates with synchronous parallel processes and gives the fastest responses. However, it is constrained by the FPGA chip resources and has to meet various timing requirements to ensure performance. It is less robust compared to the other two due to its longer development time. The command set, on the other hand, can be developed fast compared to the firmware and is sufficiently robust. Also, it runs reasonably fast. Its limitation comes with the size requirement to be fitted into the 56 kB sNVM. Despite the efforts of code modularization and optimization (and compiler

Table 4.1: Functions realized in the alternate panel control and operation system

Categories	Functions
Interface utilities	Read DDR3 status and perform operations (reading, writing, etc.) Interact with DCS
Preamp configurations	Set preamp gains Find appropriate channel threshold settings (bias voltages) Set and read channel thresholds Read preamp settings
Slow control monitoring	Read temperature diodes, humidity and pressure sensors Read PCB currents and voltages
Data taking and transmission	Perform bit alignment Read straw hits to DDR3 memories Read straw hits to console Read from multiple panels simultaneously Plot straw hit signals
Diagnostics	Configure and toggle calibration pulses Measure channel rates (noise rates or hit rates) Check anode wire continuity Activate high voltage disconnection mechanism
Other utilities	Read ROC serial number Reset interfaces, TDCs, chips, FPGAs, etc. Read and write DIGI and ROC registers

optimization), the driver modules are irreducible to start with, and the ROC needs to control all the chips. The software console has no concern with resource limitations, and the Python platform along with the modularized design gives it great robustness. However, it has to deal with the bottleneck of the UART serial communication. Unlike the other two levels where most processes happen inside the FPGA chips, complex procedures controlled by the software console require multiple transmissions through the serial link, which not only are limited by the serial data rate, but also introduce extra communication delays, making the processes peculiarly slow.

Hence, the question of "what goes where" requires careful considerations for a better-optimized system. The choices made during the development balanced the considerations of efficiency and resource usage. The solutions combined the use of firmware and software, and prioritized functions with great importance and frequent use.

Here we will take the functions managing the preamp thresholds as an example. The thresholds are rather important in the sense that they control what goes into the DAQ. A straw hit is triggered by both ends of a straw simultaneously surpassing their respective thresholds. The "streaming" approach of the DAQ system means the thresholds determine the channel hit rates,

influencing detection efficiencies of the straws. Moreover, as will be demonstrated in Chapter 5, the threshold levels also affect the longitudinal resolution of a straw, which in turn influences the detector's momentum resolution. The thresholds will be regularly checked during the experiment data taking.

As mentioned, thresholds are controlled by the bias voltages individually applied to the ends of the straws. These bias voltages are supplied by DAC chips. Hence, if one knows what digital value to use, setting the threshold to one end of a straw is rather straightforward: one only needs to pass the value to the corresponding channel of the correct DAC. This is clearly most suitable to be handled on the micro-controller level.

Reading back the threshold voltages of straw ends receives high priority due to the checking needs. Due to the relation between the thresholds and the bias voltages, the threshold voltages can be measured as the differences between zero voltage and the signal pedestals, or simply the absolute values of the output pedestals. An additional subtlety is that the ADC readouts are sums of both straw ends, so the gain of the opposite end needs to be turned off before measuring the threshold on one end (where the gain control is accomplished on the micro-controller level). The actual measurement involves a dedicated DIGI firmware module, which taps into the data streams of all channels from the ADC interface. The module takes running averages of the pedestal readings and returns the values as the measured thresholds. The fast parallel processing of multiple channels in the firmware allows the procedure to be squeezed in the off-spill time of the proton beam if necessary.

Now the only loose end is to decide what digital values to use to set up the thresholds. Previous simulations suggested a set of balanced thresholds of $12\ \mathrm{mV}$ that optimizes the Tracker momentum resolution. Due to small discrepancies (e.g., component tolerances) in preamp electronics, the DAC digital values corresponding to the $12\ \mathrm{mV}$ offset are not all the same; the values are normally distributed with a standard deviation of $15 \sim 20\ \mathrm{DAC}$ units. As each DAC unit increase corresponds to roughly $1\ \mathrm{mV}$ decrease in the threshold voltage, the DAC digital value on every straw end needs to be found individually. The algorithm to find the DAC setting value corresponding to a target threshold voltage on one end of a straw channel is outlined in Pseudo Code Snippet 4.2. As the threshold settings only need to be found once under the appropriate running conditions,[5] taking its complexity and other competing priorities into consideration, the scheme to find threshold settings is executed in the software control console. For cross-checking purposes, the threshold voltage measurements in this scheme were conducted slightly differently by reading out traces of pedestal data samples to the console and taking sample averages.

In the same spirit as demonstrated, all panel functions are realized in appropriate levels of

[5]Temperature changes of the electronic components can shift the bias voltages by a few millivolts, so ideally the procedure should be conducted after the electronic components reach thermal equilibrium.

Pseudo Code Snippet 4.2 Algorithm to Find Threshold Setting Value on Single End of a Straw

Turn off gain on the opposite side of the straw
Set DAC value to a reasonable value x
while $step < 100$ **do** ▷ Prevent infinite loops. Errors reported if the maximum step is reached.
    Measure threshold voltage V_m
    if $V_m < V_t$ (target voltage) **and** V_p (measured voltage in the previous step) $< V_t$ **then**
        $x \leftarrow x - 1$
        Set DAC value to the new x
    else if $V_m > V_t$ **and** $V_p > V_t$ **then**
        $x \leftarrow x + 1$
        Set DAC value to the new x
    else ▷ The previous and the current measurements sandwiches the target value.
        Set the x corresponding to the highest threshold voltage below the target value
        Log the DAC value in a file of settings
        break
    end if
    $step \leftarrow step + 1$
end while
Turn on gain on the opposite side of the straw

the system. Frequently used ones (threshold voltage readings, bit alignment, etc.) are handled by dedicated firmware modules. The ones involving simple register and interface reading or writing are normally the micro-controller's responsibilities. And the complex ones with lower priorities (diagnostics scripts) are tended by the control console through the UART link. Treatments using different levels are combined when deemed beneficial.

A few extra examples of the function implementations are presented in Appendix B if the readers are interested.

* * *

In summary, the firmware-based operation and data acquisition provide the Tracker with high data throughput capabilities and robust control options. In the absence of the formal experiment DAQ, the alternate control console through a serial link provides powerful functions and great conveniences for detector development, testing, diagnostics, and beyond.

References

[1] L. Bartoszek, *et al.* (Mu2e Collaboration), *Mu2e Technical Design Report*, Tech. Rep. (Fermilab, 2015) arXiv:1501.05241 [physics.ins-det] .

[2] B. Echenard and R. Rivera, "Trigger and DAQ Working Group Report," (Collaboration Meeting 2018-03-02), mu2e-doc-16229-v1.

[3] R. Kwarciany and R. Rivera, "TDAQ Timing Distribution Specification," (2018-09-26), mu2e-doc-7048-v11.

[4] M. Tecchio, "DRAC DDR3 Readout," (2019-06-23), mu2e-doc-26902-v1.

Chapter 5 Tracker Vertical Slice Test

This chapter will discuss the Tracker VST efforts and present the results from panel ^{55}Fe source scan studies. In the studies, the straw longitudinal resolutions were verified to satisfy the experiment requirement, and effective signal propagation speeds across a panel were mapped. The influence of channel thresholds on the longitudinal resolution is also presented in this chapter.

5.1 Vertical Slice Test and Setup

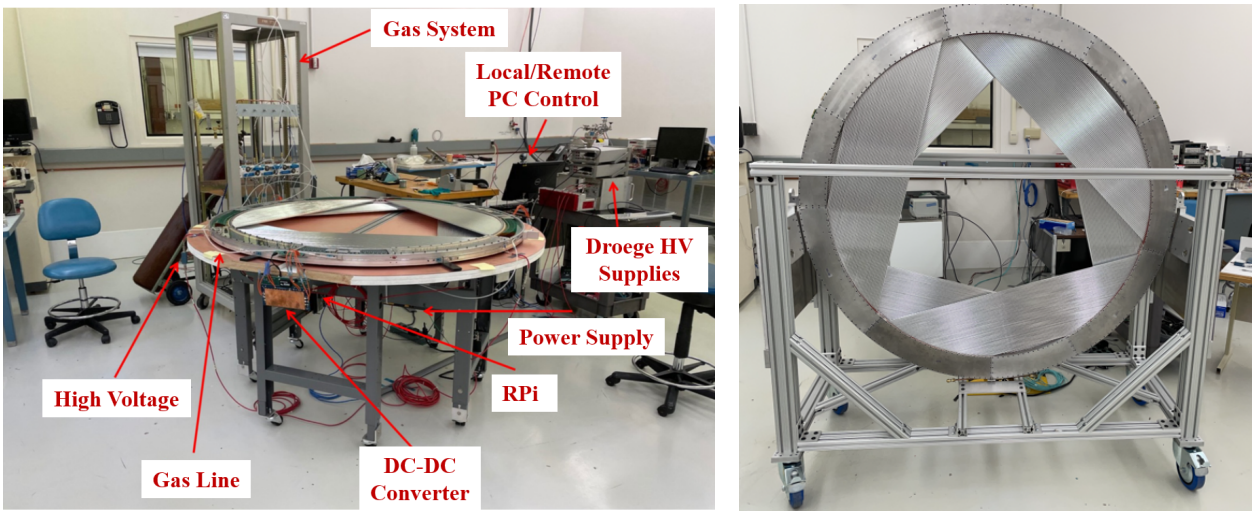
As discussed in Chapter 3, the Mu2e Tracker panels need to satisfy a series of performance requirements for successful operations and the designed resolution. While various individual component and single panel tests in the past confirmed their respective effectiveness, it was of great interest to extend the tests to a system of multiple panels over a longer period of time under a more realistic setup (i.e., similar to that of the actual experiment setup), and to get better quantitative understandings of the detector performances. Hence, a VST was conducted.

The overall goal of the VST was to "exercise the full tracker operation and readout chain, from the amplified wire signal to its storage in digitized format" [1]. The test used a whole plane of six panels from panel pre-production,[1] which was 1/36 of the full Tracker.

The VST contained two phases. In the first phase, the plane was laid horizontally on a test bench for diagnostics and horizontal position runs, as shown in Figure 5.1(a). In the figure the supporting systems, including gas lines, DC-to-DC converters for low voltage supplies, high voltage supplies, and the Raspberry Pi (marked as RPi) for panel controls, are all labeled.[2] In this phase, the plane operations and performances were checked. The plane data output to the DAQ server through the optical fiber links (see Chapter 4 and Appendix B) were finalized and tested. Small cosmic-ray datasets were taken to verify the ability to synchronize data taking among the panels. And ^{55}Fe source scan studies were performed for calibrations. The second phase of the

[1] These pre-production panels were the earliest to make while the manufacturing processes were still maturing. Their performances were not expected to be among the best of the panels due to later manufacturing improvements based on experience and better understandings.

[2] A Faraday shielding cover was placed over the plane, which shielded the detector from the EM noises in the environment. The long straws and wires in the panels are great antennas that are highly susceptible to such noises. It will not be a concern in the real experiment since the Transport Solenoid solenoids and the walls of the vacuum cavity around the detectors will shield such ambient EM noises. The shielding is not included in Figure 5.1(b) either to give a better view of the setup.



(a) (b)

Figure 5.1: Pictures of VST setup. (a) Setup of the VST plane in the horizontal position. The arrows point to various supporting systems. (b) The tested plane installed on the mechanical support for the vertical position runs.

VST, on the other hand, collected large cosmic-ray datasets with the plane in a vertical position, as shown in Figure 5.1(b). This was the same plane orientation as in the full detector; cosmic-ray track reconstruction techniques were exercised so as to prepare for detector commissioning. Throughout the year-long VST campaign, the operations of the full plane and data taking through the optical fibers were extensively exercised. The long-term reliabilities of the mechanical and electronic components were confirmed.

Now we will focus on the ^{55}Fe source scan studies. Results obtained from the cosmic-ray datasets will be referenced for necessary comparisons.

5.2 Source Scan Studies with Iron-55

In the experiment, longitudinal coordinates of incident particles on a straw are reconstructed from time differences between signal arrival times at the straw ends (a.k.a., "time divisions"). In the source scan the correspondence between the coordinates and the time differences were studied. In addition, the straw longitudinal resolutions at various panel locations were mapped to investigate how they vary with straw lengths and locations along the straws.

As illustrated in Figure 5.2, when a particle creates a signal in a straw, it propagates towards both ends. The leading edge is determined by some high-frequency component that has a large propagation speed v_{prop}. The components with lower frequencies have slower propagation speeds, and lag behind, which lead to a dispersion of the signal.[3] When the signal reaches an end of the

[3]Here only the dispersion behaviors in the straws are discussed, but those during signal propagation from the

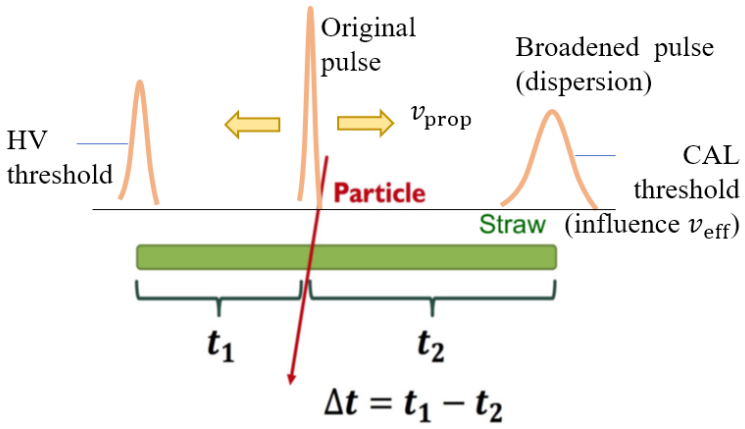


Figure 5.2: Signal dispersion and effective speed: signal dispersion while propagating along a straw, together with the triggering thresholds, affects the effective speed of a signal pulse and subsequently the longitudinal resolution.

straw, the time of the rising edge surpassing the threshold level on that side is taken as the signal arrival time. By this moment the fast components of the pulse have already arrived at the straw end, so it would appear as if the pulse propagates slower than v_{prop} at some effective propagation speed v_{eff}. This effective propagation speed is affected by the level of signal dispersion, the threshold level, as well as the initial pulse size. Consequently, the signal time division $\Delta t = t_1 - t_2$ (where t_1 and t_2 are the propagation time to the two ends) deviates from a simple linear relation to the longitudinal coordinate of the incident particle along the straw. And when the longitudinal coordinate is reconstructed from the time division, the associated uncertainty, or the longitudinal resolution, is not uniform across the entire panel. Without a clear analytic form to describe the higher order terms in the relation between the longitudinal coordinate and the time division, or that and the longitudinal resolution, actual panel measurements are necessary to understand the behaviors, to improve the reconstruction accuracy, and to assist tuning of more realistic Monte Carlo simulations of the Tracker and the experiment.

Conducting the measurements with an ^{55}Fe source was a decision based on practicality. Following the discussions in Section 3.1.5, MIPs are the best substitutes for highly relativistic electrons in studying straw tube responses. The most accessible MIPs, the cosmic-ray muons, were also used in the VST. However, the cosmic-ray tests are abided by certain limitations. The hit rates on the straws are relatively low ($\sim \mathcal{O}(1\text{ Hz})$), and the exact coordinates of the incoming cosmic-ray muons are not known without additional detecting elements.[4] On the contrary, radioactive sources

preamps to the DRACs are ignored. This comes with a good reason, as the preamps shape the narrow pulses from the straws to much wider ones. As discussed in Jackson, the broad wave packets distort relatively little, while the narrow ones broadens rapidly.

[4]Due to the way how the Tracker DAQ is set up, it would require some major changes in the firmware in order to add additional detecting elements with spatial resolution capabilities and synchronize their signal readouts with those of the straws, so it was preferred to avoid including such additional elements altogether. Another work-around is through

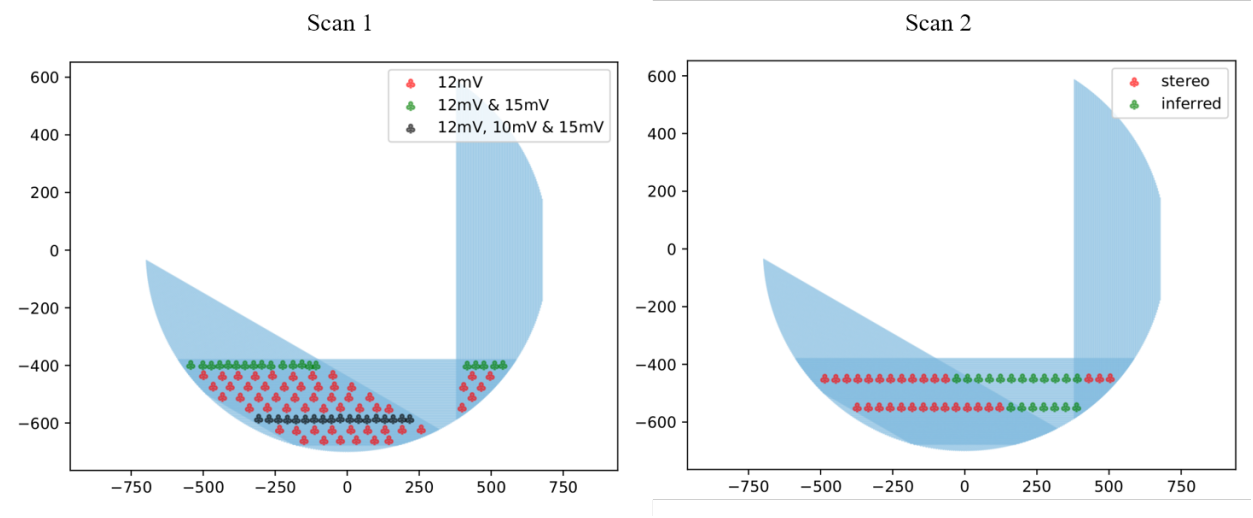


Figure 5.3: Source positions in the two collected scan datasets. The positions are indicated by the markers.

can provide a high flux of radiation at a relatively well-defined location. Especially in the case of ^{55}Fe, both the x-ray photon energy spectrum and the interactions in the straw tubes are well known. In addition, the previous prototype measurements demonstrated that the longitudinal resolution values measured from ^{55}Fe sources were similar to those measured through cosmic rays [2, 3]. X-ray photons from ^{55}Fe sources do interact differently and leave no ionization tracks; yet for the purpose of longitudinal resolution scans, along with some other tests not specifically investigating the ionization tracks and their charge drifting behaviors, the ^{55}Fe sources are arguably more suitable for the tasks. The source used for the scans was cased in a small (roughly 1 cm in diameter) metal cylinder with a thin window on top, and was considered azimuthally symmetric.

Because the x-ray photons from ^{55}Fe typically deposit more energy in the straws than the MIPs (or the ~ 100 MeV electrons), in the source scans, the high voltage applied to the straws was lowered to 1250 V. This led to a smaller gas gain, and the amounts of charges collected on the wires were comparable to those of the MIPs.

There were two scan datasets collected from two different VST panels. The source was placed about 1 cm away from the closest layer of straws at various locations across the panels, as mapped in Figure 5.3. In the figure, each marker represents a source location. Due to the spatial spread of the x-ray photons from the source, at each source location roughly 20 straw channels per panel received sufficient numbers of hits for analysis. In Scan 1, all source locations were in the regions with overlapping panels. Such "stereo" locations allowed reconstruction of the source positions through the channel occupancy plots from the panels, which will be shown in a moment.

using cosmic-ray muons traversing multiple straws. Through fitting, the trajectories of the original particles can be reconstructed. This was in fact the method used to study the spatial resolutions from the cosmic datasets. Yet there is no control of the trajectories of the incoming cosmic-ray muons, and it is time-consuming to accumulate sufficient amount of data to map straw longitudinal resolutions at various locations of the panel, as what was done in the source scan.

All functional straws in the tested panel were covered in this scan. Datasets with balanced $12\ \mathrm{mV}$ thresholds on both ends were collected at all source locations. In addition, at selected locations as marked in the figure, datasets with balanced $10\ \mathrm{mV}$ and $15\ \mathrm{mV}$ thresholds were also collected for comparison. A proposal to add mechanisms to constrain the source locations to a straight line was raised, so that the locations in the non-stereo region of the panel can be precisely reconstructed for scans. It was not approved in Scan 1 out of concerns of damaging the straws due to limited space in the setup, but was adopted in Scan 2 after specifying the potential hazards and standardizing the operational steps. A metal bar was fixed to the bench, and by sliding the source along the bar by a fixed amount of distance each time, the additional source locations marked in the figure were surveyed and their locations known to better than $1\ \mathrm{mm}$ in each direction. In Scan 2, at each source position, a dataset with unbalanced thresholds ($15\ \mathrm{mV}$ on the HV side and $12\ \mathrm{mV}$ on the CAL side) was taken in addition to the standard dataset of balanced $12\ \mathrm{mV}$ thresholds to assess the influence on longitudinal resolutions. If not otherwise stated, the analyses and results presented in the following section are from datasets taken with balanced $12\ \mathrm{mV}$ thresholds.

5.3 Scan Analyses and Results

5.3.1 Reconstruction of Source Locations

In order to know the longitudinal coordinates of the incident x-ray photons, the source locations need to be precisely reconstructed. In both scans, the source locations were primarily calculated using the overlapping panels. For each stereo location, the coordinates along the radial directions of the panels were separately extracted from channel occupancy plots of each panel, and the intersection was used to pin the location.

An example channel occupancy plot is shown in Figure 5.4(a). It shows the number of hits in each straw channel on a panel. As illustrated, the straws directly above the source were closer to the source and corresponded to larger solid angles, so they received more hits than the channels further away. A cut based on signal amplitude was applied to exclude hits triggered by high-frequency electromagnetic noises picked up from the environment, which in this case only cut out a very small fraction of hits.

This occupancy plot exhibits a clear even-odd phenomenon as expected. As the straws in the panels are arranged in two layers, one is closer to the source and receives more hits. Consequently, the channels with odd straw numbers were treated separately from the even ones. In each case, the occupancy plot after separation was fitted with a Gaussian function, which happened to describe the shape of the occupancy plot relatively well, to extract the fitting mean and the standard deviation. The former corresponded to the source location, or in other words the mean coordinate of the emitted photons, in the unit of straw separation distances along the radial direction perpendicular

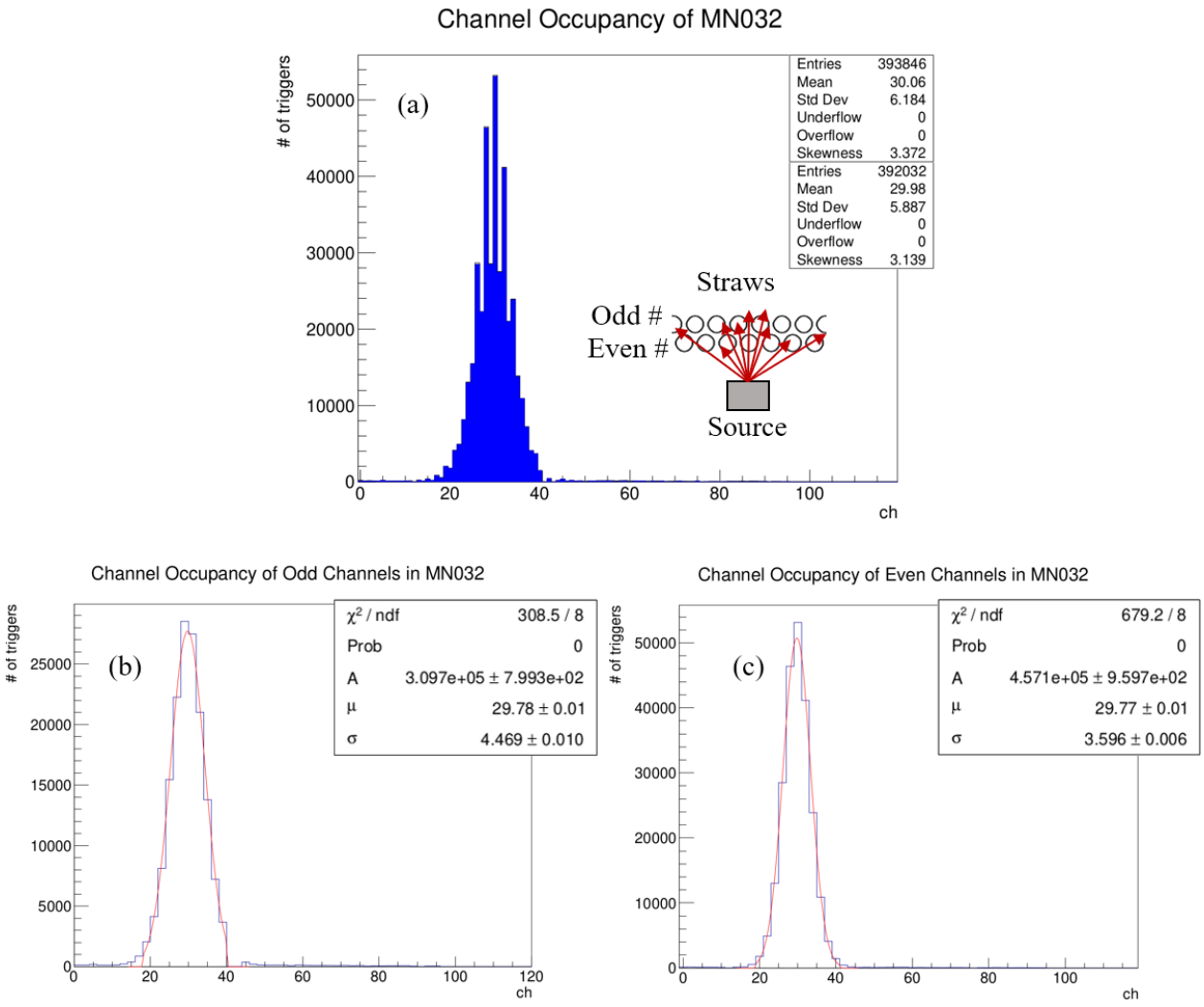


Figure 5.4: Example channel occupancy plot and fittings. (a) An example channel occupancy plot of a panel corresponding to some specific source location. A simple cleaning cut based on signal amplitude was applied to exclude empty hits triggered by high-frequency noises. The two statistics box correspond to the data before and after the cut. (b)(c) Odd and even channels were separated. Each occupancy plot was fitted by a Gaussian function to extract the mean position and the spread.

to the straws. The standard deviation, on the other hand, described how wide the emitted photons spread. Due to the azimuthal symmetry, the same width was expected in the direction of the straw lengths, which was later used in evaluating the straw longitudinal resolutions. The mean values from the fitting results always had good agreements between even and odd channels, so averages were used for later calculations. The widths of the photon spread increased as the straw layer got further away from the source. For instance, in the given example, the odd layer had slightly larger standard deviation than the even layer. A high level of agreement was observed among the standard deviations in the same layer of straws, regardless of the source locations.

Similar occupancy plots were obtained and fitted for the overlapping panel. In the example,

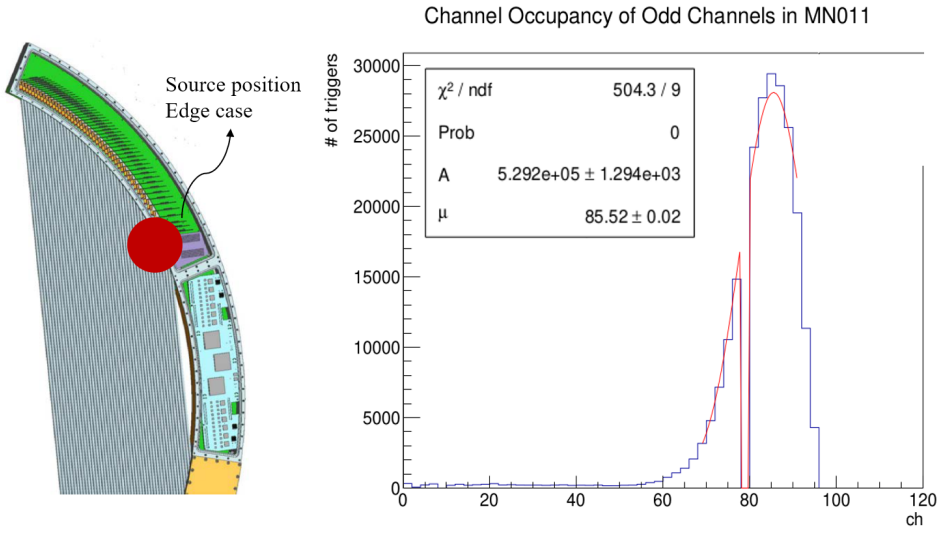


Figure 5.5: Channel occupancy plot of a source position close to panel edge. It is asymmetric due to uncaptured photons out of the straw region. Special treatment is needed in fitting.

as the second panel was on the opposite side of the plane from the source, the fitted spreads were even larger than those in Figure 5.4(b) and (c). While the standard deviations in the bottom panel were on average $11.2\ \mathrm{mm}$ and $14.0\ \mathrm{mm}$ after converting the straw separation distances to millimeters, the overlapping panel saw spreads of $23.5\ \mathrm{mm}$ and $24.6\ \mathrm{mm}$ among the photons.

The larger spreads required special treatments for source locations close to the panel edges. As shown in Figure 5.5, the red circle represents the area with incident photons. As some of the photons were incident out of the straw region, the higher-numbered channels received fewer number of hits than usual, and the channel occupancy plot became asymmetric as a result. Using the same fitting scheme would lead to a large error for such locations. The solution was to fit the occupancy plot only on the left half of the peak, and to fix the standard deviation using the value obtained from normal locations. The resultant fit as shown in the figure still captured the source location relatively well. Another detail worth mentioning in this figure is the empty bin at Channel 79. In the panels, there were some inactive channels due to various defects, including high voltage malfunctions, manufacturing mishaps, damages, and so on. A total defect rate below 5% of all straw channels is allowed while still maintaining an efficient Tracker. When such channels appeared in the occupancy plots, the corresponding bins were excluded from the fits as shown here to avoid fitting errors.

The radial coordinates perpendicular to the straws on the panels confined the source locations to lines parallel to the straws, as shown in Figure 5.6. Since the two overlapping panels were rotated by either 30° or 90°, the intersections provided the source locations. Their coordinates in the Cartesian plane (where the center of the VST plane was set to be the origin and the straws of the scanned panel were defined to be along the x-axis, on the lower half of the Cartesian plane) were calculated as already shown in Figure 5.3.

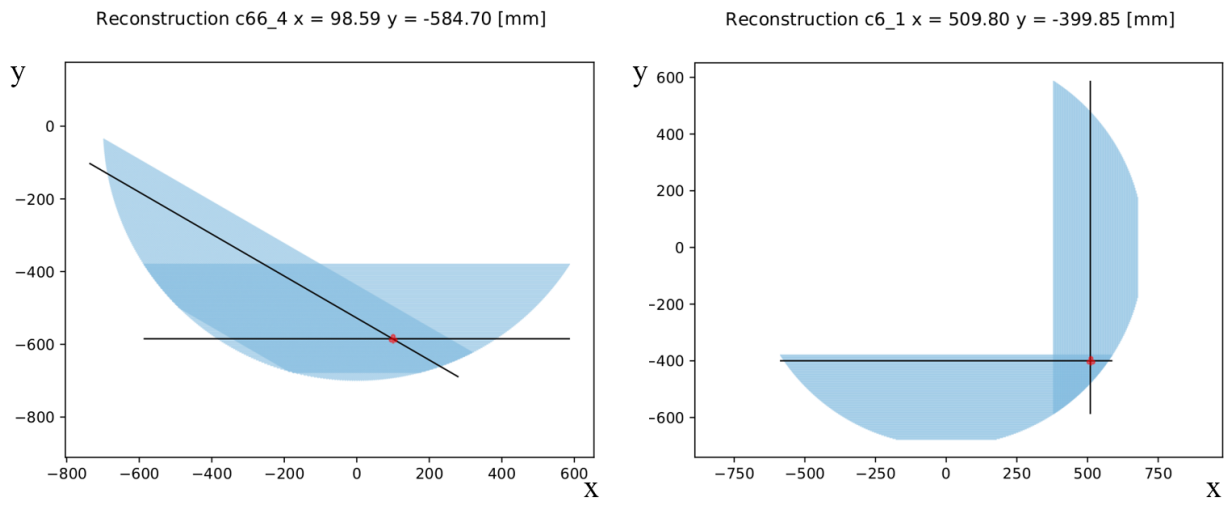


Figure 5.6: Two example plots of source location reconstruction.

For the locations in the non-stereo region in Scan 2, their coordinates were then calculated from the other locations on the same straight line as they were equidistant between neighboring locations. The reconstruction errors of the coordinates were estimated from the occupancy plot fitting uncertainties, as well as the differences between fitted means from even and odd channels. All locations were reconstructed to precisions better than $1\ \mathrm{mm}$, which were sufficient for the following analyses.

5.3.2 Time Division and Effective Propagation Speed

Now we can take a look at the hits collected at each source location. To minimize the influence of the photon spread widths on later resolution evaluations, the analyzed panel is on the same side of the plane as the source.

Figure 5.7 shows the hits collected at one of the source locations. Figure 5.7(a) is the spectrum of signal pulse amplitudes. Rigorously speaking, it is the total area under the pulse that reflects the total charges collected during a hit. As the pulses are almost congruent to one another (up to very small fluctuations) after being normalized by the peak heights, the amplitudes alone can be conveniently used to represent the signal sizes. The features of spectrum were previously explained in Chapter 3. Figure 5.7(b) shows the distribution of hit time divisions. Here Δt was defined as the TDC time on the HV side subtracting that on the CAL side. So, the HV side of the straw corresponds to more negative time division values, and the CAL side has more positive ones. It should be noted that the center of the straw does not necessarily correspond to $\Delta t = 0$. As the PCB trace lengths from the preamps to the DRACs differ slightly for the two ends of a straw, the time divisions include some offsets specific for each channel. The offset values can be

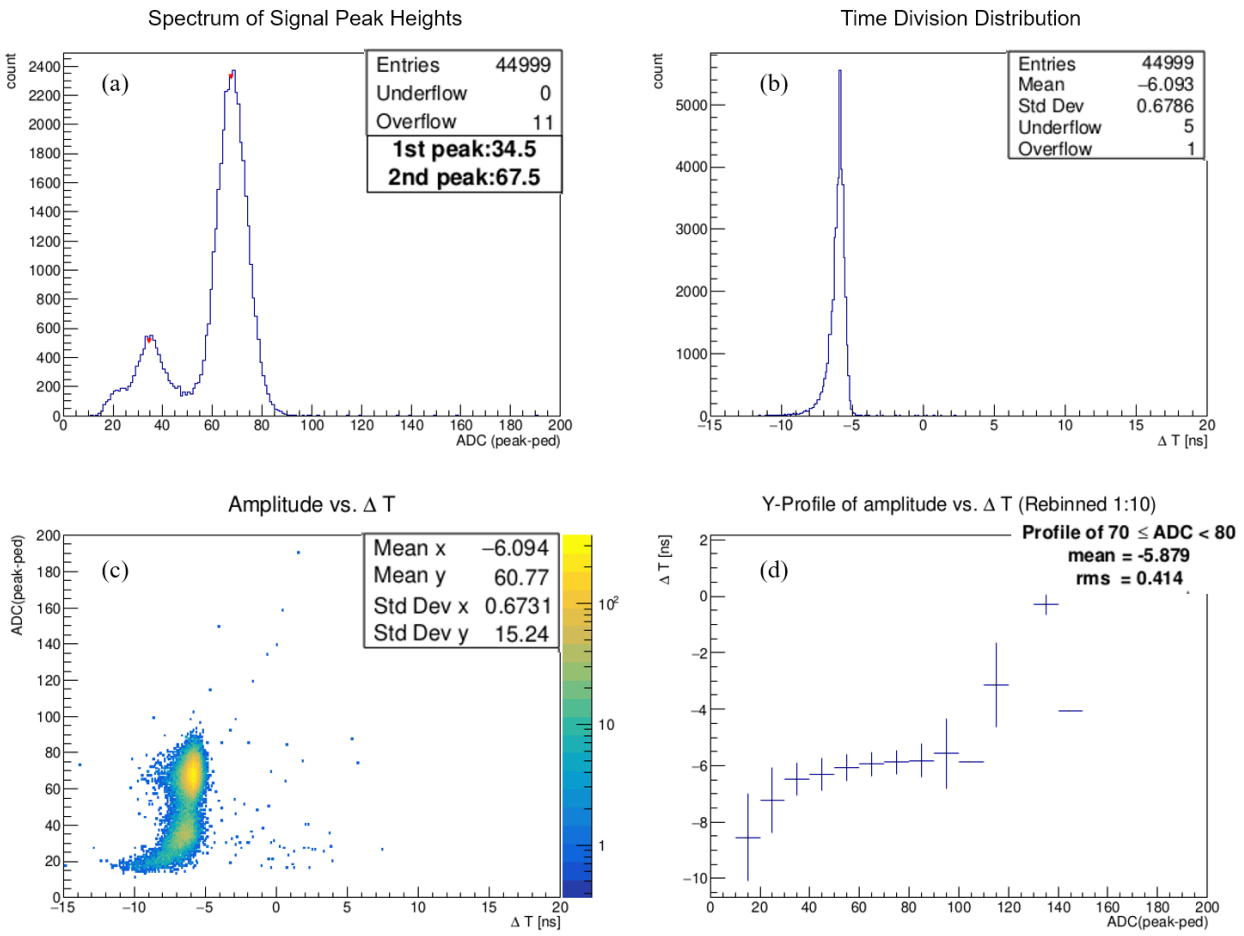


Figure 5.7: Spectrum and time distribution of hits at a fixed source location. (a) The spectrum of signal pulse amplitudes. The 5.9 keV photon peak and the argon escape peak were explained in Chapter 3. (b) The distribution of time division. The skewness is determined by the source location along the direction of the straw length. (c) A 2-D histogram binned by time division and signal amplitude. Notice the center of the time division moves with an increasing signal amplitude. (d) The y-profile of the 2-D histogram emphasizing the mean time divisions and the time division widths for different signal sizes. The histogram was rebinned for statistics. Each bin corresponds to 10 ADC counts.

extracted from fittings that will be shown in a moment. It is worth noticing that the time division distribution is skewed. When viewing the distribution in a 2-D histogram binned by time division and signal peak height as shown in Figure 5.7(c), one notices that the skewness mainly comes from the low-amplitude hits. This can be understood relating back to the explanation of Figure 5.2. For signals whose amplitudes are not much larger than the threshold level, there is a larger time difference between the fast component arrival time and the trigger time, and the effective signal propagation speeds are slower compared to larger signals. The time division value hence shifts along with the signal trigger times. This is more clearly manifested in the y-profile of the 2-D histogram in Figure 5.7(d), which presents the mean time divisions and the time division widths

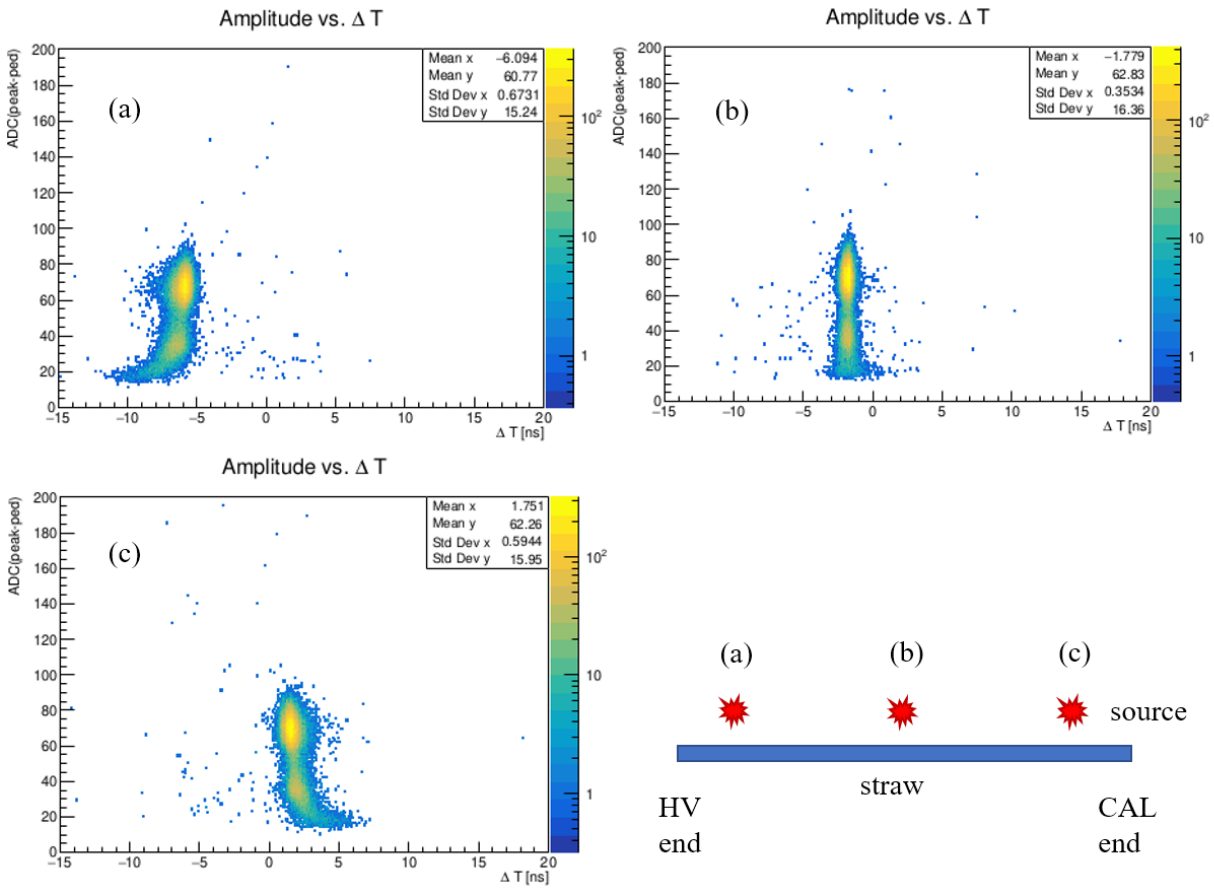


Figure 5.8: 2-D histograms of hit time division and amplitude for various source locations along a straw.

for different signal sizes. The low-amplitude hits have most widely spread time division values furthest deviated from the rest of the population. The time divisions become relatively stable when the signal amplitudes reach 3 or 4 times the threshold level (roughly $40 \sim 50$ ADC units). Due to this phenomenon, the straw longitudinal resolution is a signal-size-specific value; hits with different amplitudes correspond to slightly different longitudinal resolutions. To capture the characteristics of the hits coming from the highly relativistic electrons (whose spectrum is similar to the MIP spectrum shown in Figure 3.8), the following analyses used the mean time divisions and time division widths of signals with amplitudes between 70 and 80 ADC units. This range corresponds to the most likely hit sizes in the MIP spectrum, and hits with even higher amplitudes expect similar longitudinal resolutions.

The discussed behaviors at low hit amplitudes were observed across the entire straw length. Figure 5.8(a) to (c) show 2-D histograms similar to the one in Figure 5.7(c), each corresponding to a different source location along the straw as illustrated. In general, the samples closer to the straw ends had larger time division widths, and larger skews in mean time divisions for the low-amplitude hits. For multiple source locations along a straw, the respective core time division

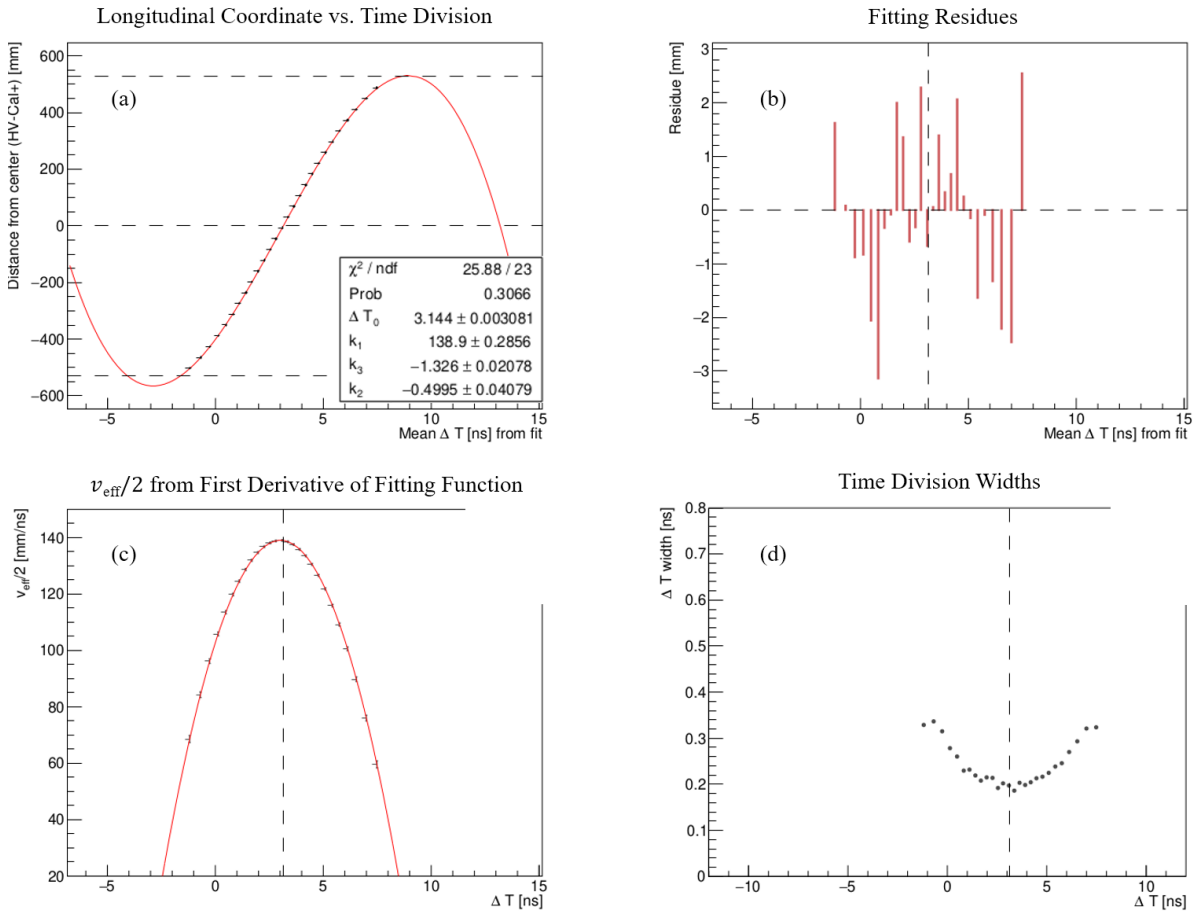


Figure 5.9: Source coordinates, time divisions, and local effective speeds. (a) Source coordinates along a straw were plotted against the mean time divisions (of hits between 70 and 80 ADC units) and fitted using an expression of Taylor series expansion. Dashed lines indicate the ends of the straw. (b) The fitting residuals were typically on the order of $\mathcal{O}(1\ \mathrm{mm})$. The vertical dashed line corresponds to the time division offset from the fit. (c) Local (half) effective propagation speeds were extracted as the first derivatives in the previous fit. (d) Core time division widths of hits between 70 and 80 ADC units at various source locations.

widths of hits between 70 and 80 ADC units were plotted against the mean time divisions, as shown in Figure 5.9(d).

For multiple source locations along a straw, the distances between the source and the center of the straw were calculated from the reconstructed source coordinates, and plotted against the mean time divisions, as shown in Figure 5.9(a). Here the HV end of the straw was assigned a negative distance. From the graph, one can clearly observe its deviation from a straight line. To describe the relation between the source coordinates along the straw and the values of mean time divisions, a Taylor series expansion (up to third order) was fitted to the graph. The exact form is given as the following:

$$y = k_3(x - \Delta T_0)^3 + k_2(x - \Delta T_0)^2 + k_1(x - \Delta T_0)\ . \tag{5.1}$$

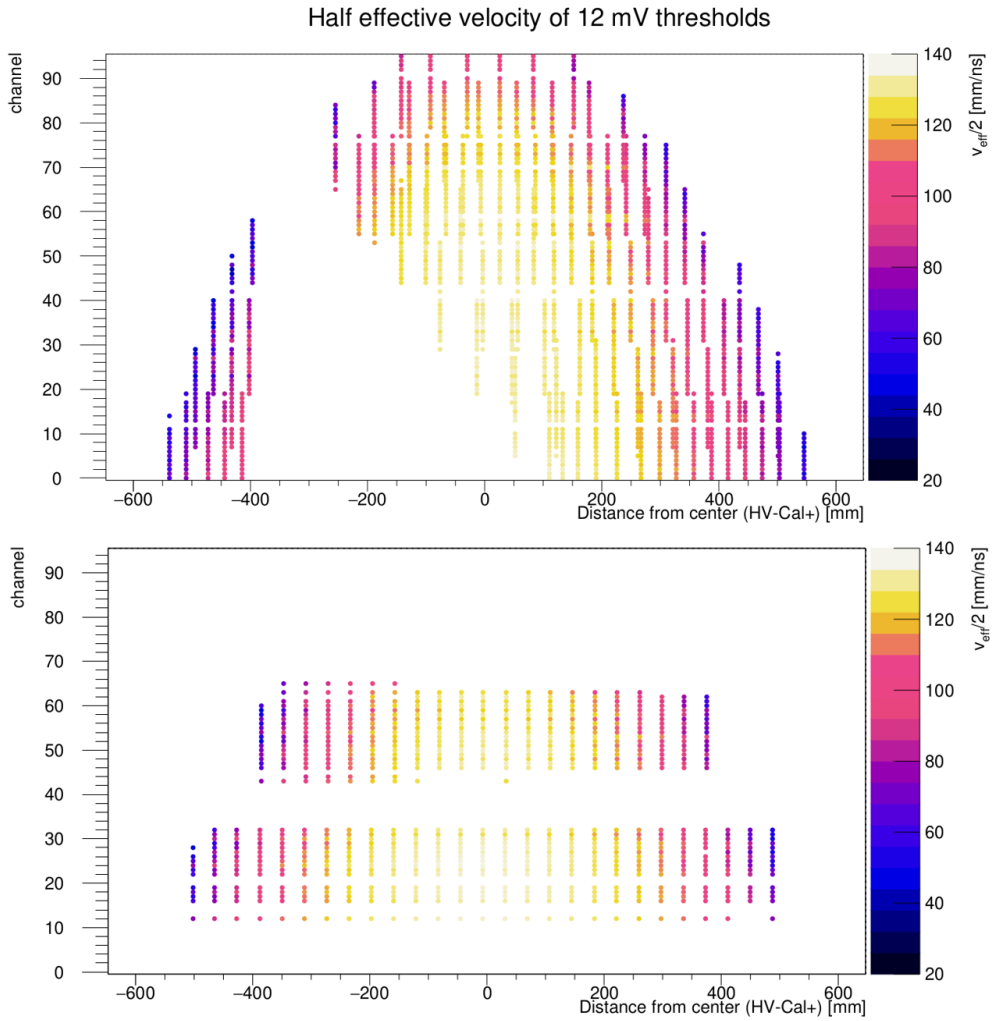


Figure 5.10: Maps of local half effective propagation speeds across the entire panel. One was obtained from Scan 1 (top), and the other Scan 2 (bottom).

Here k_1, k_2, and k_3 are the fitted Taylor expansion coefficients, and ΔT_0 accounts for the time division offsets caused by PCB trace length differences as mentioned. The expansion has no explicit constant term due to its construction; if a signal reaches both ends of a straw simultaneously, it must originate from the center of the straw. Ideally if two ends of the straw are totally symmetric, the Taylor expansion should only contain odd-power terms. The second-order term was only added to account for asymmetries in the straw, such as small threshold differences between the ends, which will be discussed in later sections.

The fits in general had good qualities; the fitting residues, as shown in Figure 5.9(b), were typically on the order of $\mathcal{O}(1\ \mathrm{mm})$, which were comparable to the coordinate reconstruction errors. Taking the first derivative of the fitted function yielded a quadratic expression that corresponded to the local half effective propagation speeds (the "half" part came from the fact that the x-axis here was the time differences between the two ends). This is shown in Figure 5.9(c). In general, the

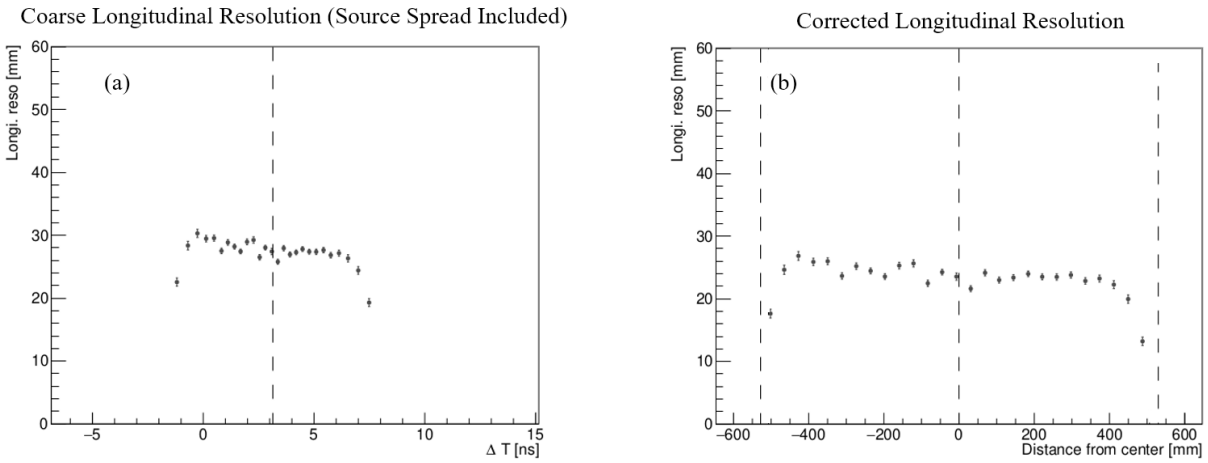


Figure 5.11: Longitudinal resolutions at various straw locations. (a) Coarse longitudinal resolutions along a straw. The values are direct products of local effective propagation speeds and time division widths. The source photon spread is included. (b) Longitudinal resolutions after correcting for the source spread. The values are plotted against the straw coordinates.

effective speeds are larger near the central regions of the straws. On the leading order, the values are symmetric with respect to the straw midpoints.

When combining the results obtained from multiple channels and plotting the values against the straw coordinates, a map of local (half) effective propagation speeds across the entire panel was obtained. Figure 5.10 shows the maps from both scans. From the maps, the longer straws have larger maximum effective speeds. The results from the two panels were consistent with each other, and Scan 2 confirmed there was no unexpected behavior in the non-stereo region of the panel. These maps can be utilized to improve reconstruction algorithms that give incident particle coordinates in the real experiment, and provide better reconstruction accuracy.

5.3.3 Straw Longitudinal Resolution

Using the local effective speeds and the time division widths (Figure 5.9(d)), a coarse longitudinal resolution was calculated by taking the product of the two at each source location, which is presented in Figure 5.11(a). It should be noticed that these values include the source spread widths. As no correlation was expected between the longitudinal resolution and the source spread, the values were corrected to obtain the longitudinal resolutions at various source locations. Plotting them against the corresponding straw coordinates gave Figure 5.11(b). The longitudinal resolution values are similar for the central part of the straw and improve slightly in the last 100 mm approaching the straw ends.

Figure 5.12(a) shows the longitudinal resolution values of a few arbitrarily selected straw channels. Also, similar to the effect propagation speeds, the longitudinal resolutions across the

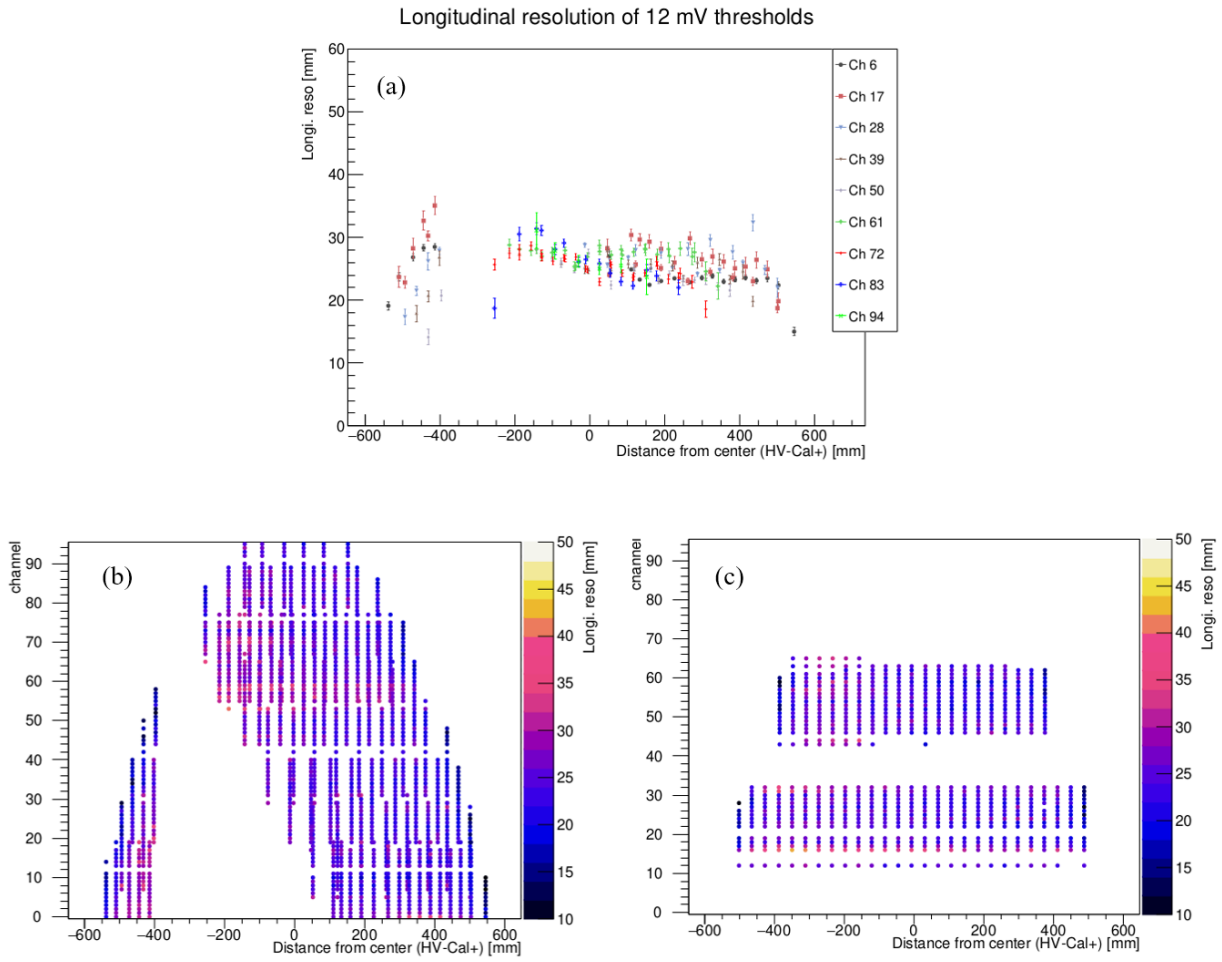


Figure 5.12: Longitudinal resolutions across panels: (a) longitudinal resolutions of some straw channels. (b)(c) the maps of longitudinal resolution across the entire panel obtained from Scan 1 and 2.

panels were mapped as in Figure 5.12(b) and (c). From the figures, when the thresholds on the straws were set to $12\ \mathrm{mV}$ on both ends, the bulk of the panels demonstrated similar longitudinal resolutions slightly below $30\ \mathrm{mm}$ (with some fluctuations $\sim 5\ \mathrm{mm}$), which satisfied the Tracker resolution requirement.

The results were compared to those from the cosmic-ray datasets. As shown in Figure 5.13(a), in the dataset the straight cosmic-ray muon tracks traversing multiple straws in the plane were fitted. Using a fixed particle incident time and fitting for ionizing cluster drift time in the straws, both longitudinal and transverse coordinates of the incoming particles were reconstructed [4]. The fitting results were improved using two iterations. The fitting residues are shown in Figure 5.13(b) and (c); in each figure the red histogram represents the outcome of the first iteration and the blue line shows the improved results from the second one. After the second iteration, the fitting residues for drift time were reported to have a core width of $2.4\ \mathrm{ns}$, which was translated to a transverse resolution of $150\ \mu\mathrm{m}$ [4]. The residues for longitudinal positions, on the other hand, had a core resolution of

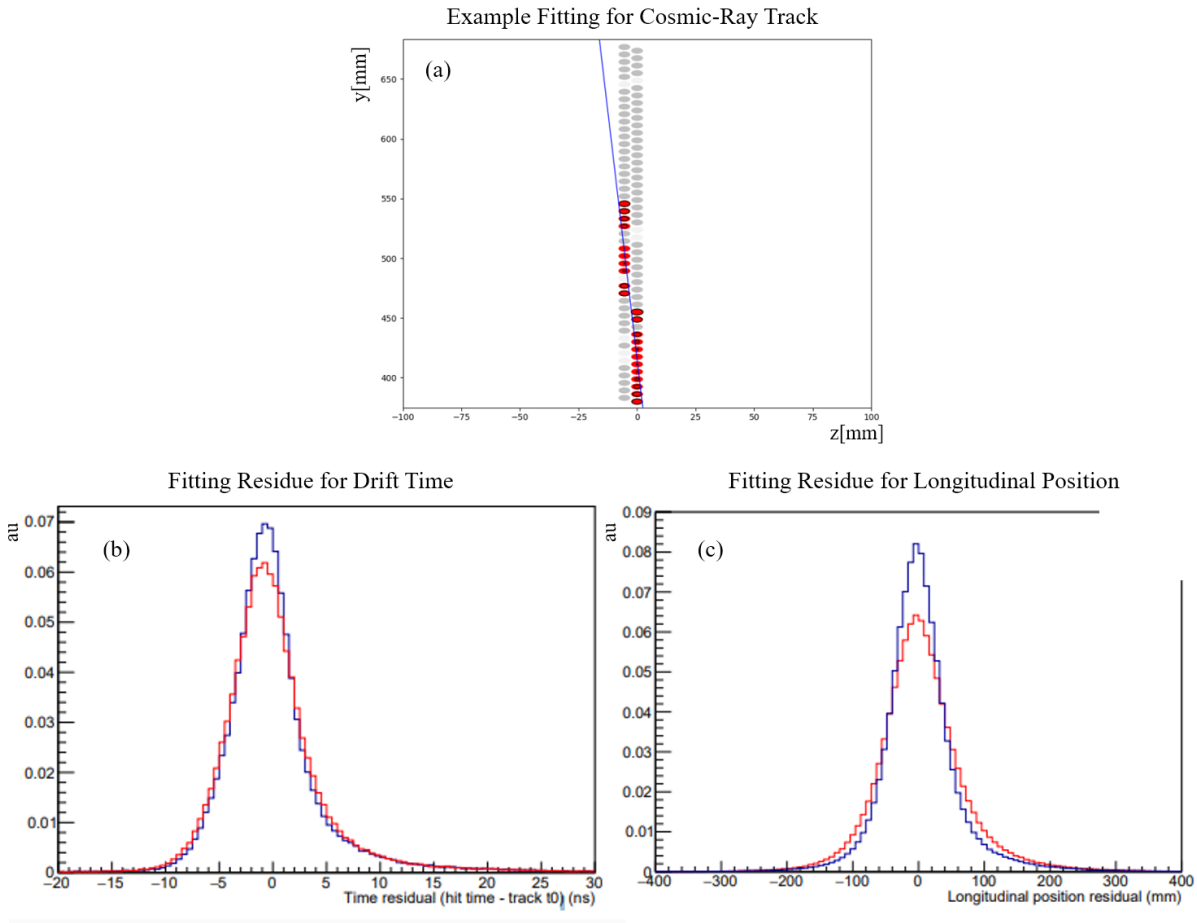


Figure 5.13: Reported cosmic-ray track fitting and resolutions [4]. (a) Example fitting for a straight cosmic-ray track causing hits in multiple straws in a cross-sectional view of the straws. (b) A histogram of fitting residues for drift time. Red line represents the results from the first iteration and the blue line comes from the second one. (c) Fitting residues for the longitudinal positions along the straws.

~ 30 mm at straw centers, and 37 mm across the whole straw length for longer straws [4]. The reported longitudinal resolution from the cosmic-ray datasets agreed with the results from the source scans. The transverse (drift) resolution also satisfied the experiment requirement.

Apart from the nominal balanced 12 mV thresholds, in Scan 1, datasets with balanced 10 mV and 15 mV thresholds were also collected. The goal was to make comparisons between different thresholds and to examine how strict the constraints on them should be. Using the same analysis method, the straw longitudinal resolutions under different thresholds were obtained. As an example, Figure 5.14(a) shows a comparison of longitudinal resolutions along a straw among different threshold settings. In the figure, the resolutions along the straw consistently deteriorate when the threshold settings change from 15 mV to 12 mV then to 10 mV. The longitudinal resolution differences between 15 mV and 12 mV thresholds at all source locations in all straws were tallied and shown as the histogram in Figure 5.14(b). As the mean of the histogram shows, on

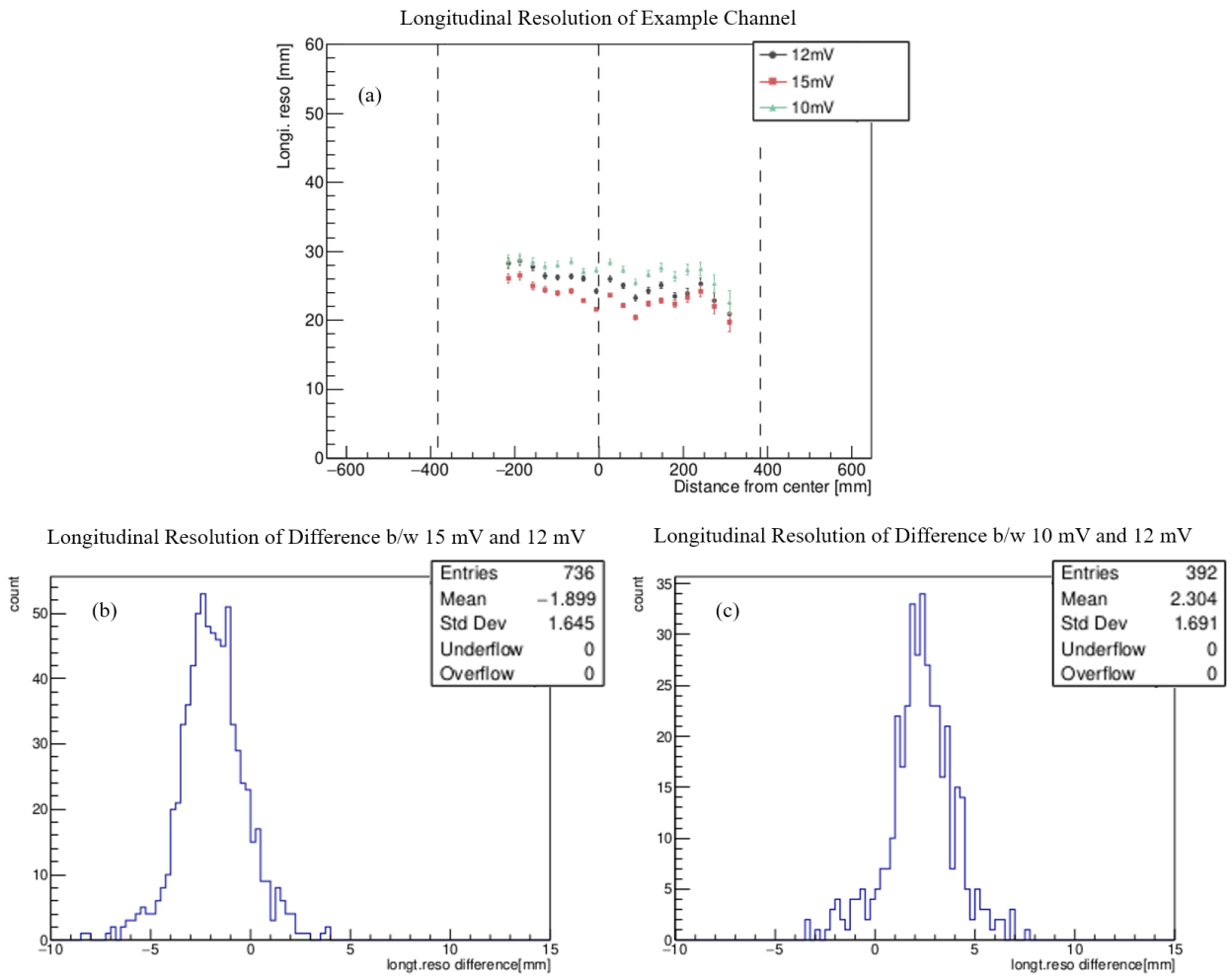


Figure 5.14: Longitudinal resolutions with different threshold settings. (a) Longitudinal resolution comparison under different threshold settings in an example channel. (b) Histogram of differences between the resolutions under $15\ \mathrm{mV}$ and $12\ \mathrm{mV}$ thresholds at each source location in all measured straws. (c) Histogram of longitudinal resolution differences between $10\ \mathrm{mV}$ and $12\ \mathrm{mV}$ thresholds.

average the straw longitudinal resolution of $15\ \mathrm{mV}$ thresholds is about $1.9\ \mathrm{mm}$ finer than that of $12\ \mathrm{mV}$ thresholds. A similar comparison between $10\ \mathrm{mV}$ and $12\ \mathrm{mV}$ thresholds in Figure 5.14(c) showed the longitudinal resolution of the former was more coarse by $2.3\ \mathrm{mm}$ on average. In all cases, the resultant overall longitudinal resolutions satisfied the experiment requirement.

5.4 Longitudinal Resolution Asymmetry between Straw Ends

In examining the straw longitudinal resolution results, it was noticed that in many channels the resolution values near the HV ends were worse than those near the CAL end, despite the threshold settings were supposed to be balanced. Figure 5.15(a) provides one example where the longitudinal resolution difference between straw ends is close to $10\ \mathrm{mm}$.

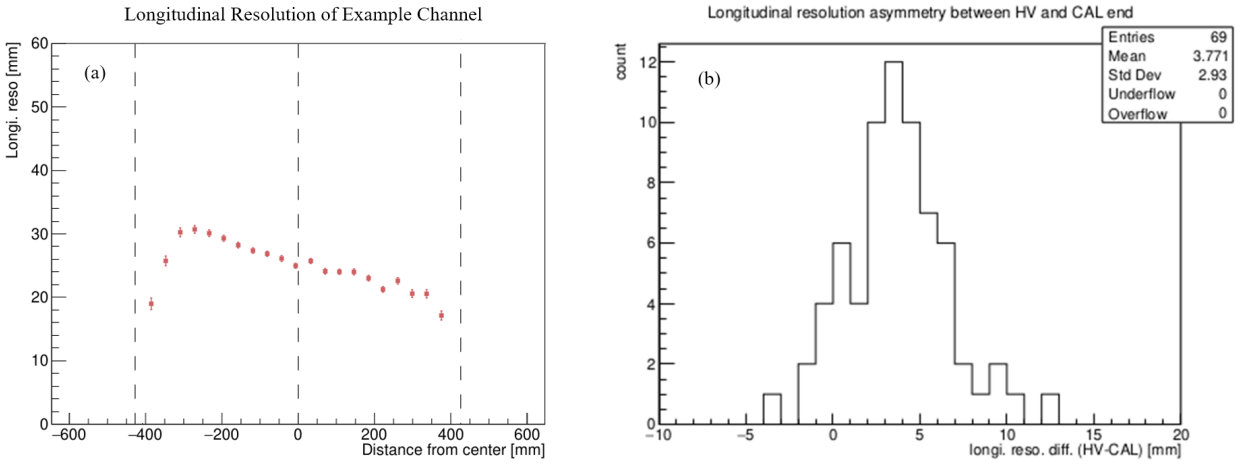


Figure 5.15: Longitudinal resolution asymmetry between straw ends. (a) An example of a straw channel with large longitudinal resolution differences between the ends. The longitudinal resolution near HV end is much larger than that near the CAL end. (b) Histogram of longitudinal resolution asymmetry (HV−CAL) levels in Scan 1. On average the difference between the longitudinal resolutions of the two ends was $3.8\ \mathrm{mm}$.

To quantitatively discuss the encountered issue, the notions of longitudinal resolution at the CAL end and that at the HV end needed to be defined. Due to the behavior near straw ends, the average longitudinal resolution between $100\ \mathrm{mm}$ and $200\ \mathrm{mm}$ from an end was chosen to represent the resolution on that end. Following this definition, the resolution differences in straws measured in Scan 1 were summarized in Figure 5.15(b). On average there was a longitudinal resolution asymmetry of $3.8\ \mathrm{mm}$ and the HV ends tended to be worse. This appeared alarming at the time. The most plausible cause was that the zero-crossing comparators were referencing a zero voltage not exactly the same as that of the ADCs, which made the actual thresholds applied different from the readings. A naive comparison to the results from various balanced threshold settings would suggest a $5\ \mathrm{mV}$ or larger difference between the ends in the actual thresholds. An offset this large, *if real*, was very unlikely to be caused by non-ideal circuits. Two questions needed answers, one being how channels with unbalanced thresholds would behave, and the other being what caused the seen differences. Both questions were answered through more measurements.

5.4.1 Unbalanced Channel Thresholds and Longitudinal Resolution

The first clues came from the datasets in Scan 2 with unbalanced thresholds. In the measurements the HV side thresholds were set to be $15\ \mathrm{mV}$ while the CAL side remained $12\ \mathrm{mV}$. Then the results were compared to the previous ones to study the influence from a single-ended change in threshold.

The analysis steps were the same as those of the datasets with balanced thresholds. For each straw channel, the longitudinal resolutions were evaluated. One example channel is presented

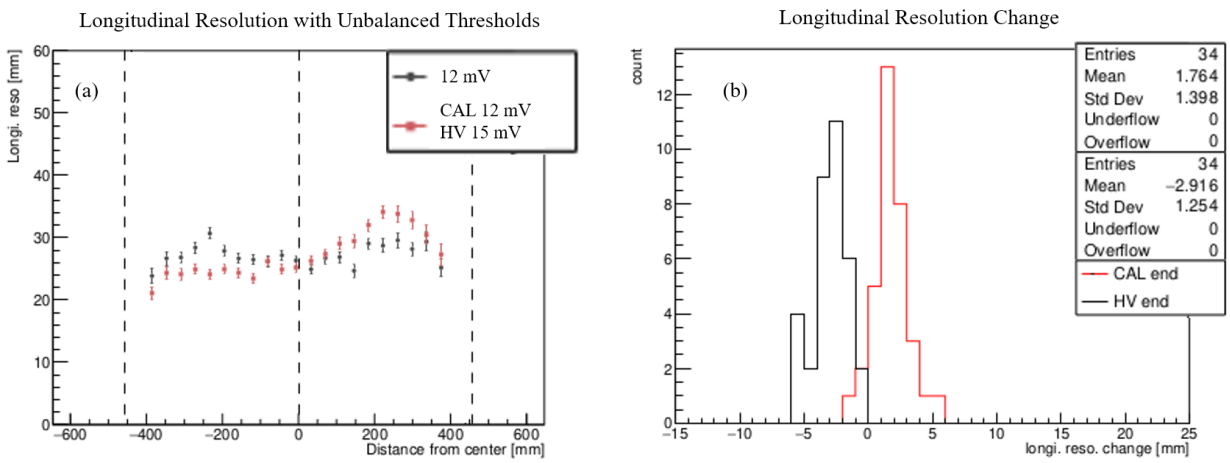


Figure 5.16: Unbalanced threshold settings and longitudinal resolution asymmetry. (a) Longitudinal resolutions under unbalanced threshold settings in an example straw. (b) Changes in longitudinal resolutions at each end after a single-ended change in threshold (HV end from 12 mV to 15 mV).

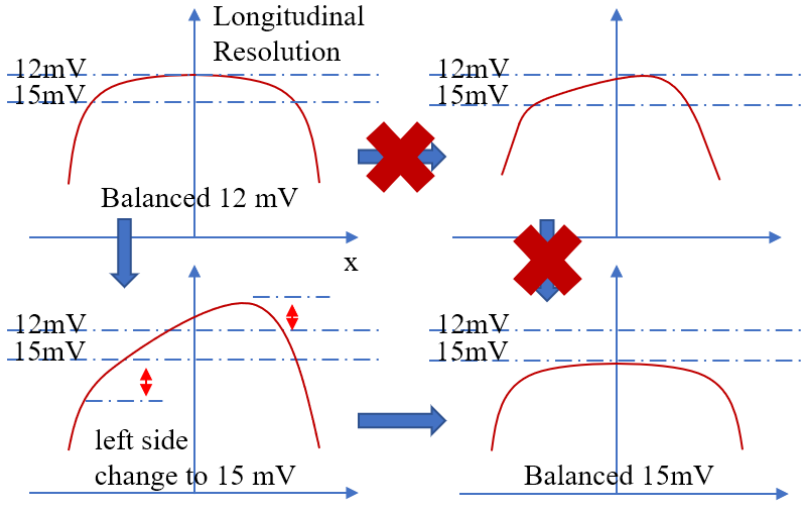


Figure 5.17: An illustration demonstrating the longitudinal resolution changes after a single-ended threshold change. Resolution changes are induced at both ends of the straw, and the resultant longitudinal resolution asymmetry between the two ends exceeds the resolution difference between the two balanced thresholds.

here in Figure 5.16(a) (red markers). The resolutions obtained from balanced 12 mV thresholds are also present in the figure (black markers) for comparison. In this channel, after a single-ended HV threshold change from 12 mV to 15 mV, the longitudinal resolution near the HV end did improve as expected. In the meantime, the resolution near the CAL end worsened. Such behavior was common in all channels, as shown in Figure 5.16(b). With the single-ended change, on average the CAL end resolution worsened by 1.8 mm, while the HV side had an improvement of 2.9 mm. On the HV side with the increased threshold, the signals originated near the CAL end appeared to trigger in a less consistent part on the rising edges of the dispersed pulses. The increased time division widths

close to the CAL side drove the change in the longitudinal resolution.

The illustration in Figure 5.17 provides a better demonstration of what happens after a single-ended threshold change. The two comparisons are the longitudinal resolutions along the straw under balanced $12\ \mathrm{mV}$ and $15\ \mathrm{mV}$ thresholds. The dashed lines show the corresponding typical resolution levels. When the left end threshold changes from $12\ \mathrm{mV}$ to $15\ \mathrm{mV}$, contrary to the previous impression, the resolution values on the close end are decreased below the level of balanced $15\ \mathrm{mV}$ thresholds, and those on the far end are increased beyond the level of $12\ \mathrm{mV}$. Switching from this configuration to that of balanced $12\ \mathrm{mV}$ thresholds involves self-consistent changes following the same guidelines. The longitudinal resolution asymmetry introduced between the two ends after the single-ended threshold change can exceed the resolution difference between those of balanced $12\ \mathrm{mV}$ and $15\ \mathrm{mV}$ thresholds. Using statistics from Figure 5.16(b), a total difference of $4.7\ \mathrm{mm}$ was introduced between the longitudinal resolutions near the two ends, and this value exceeds the level of resolution asymmetry observed in the data.

From the above analyses, it was made clear that $2 \sim 3\ \mathrm{mV}$ discrepancies between the threshold readings and actual threshold levels on the correspondent ends of the straws were sufficient to cause the $3.8\ \mathrm{mm}$ resolution asymmetries we saw. A discrepancy at this level is entirely possible as a result of parasitic impedance variances among different regions on a DRAC board influenced by temperature or other factors.

5.4.2 Examination of Channel Thresholds

While the above section explained the behaviors of channels with unbalanced thresholds, and clarified that small threshold discrepancies *can* cause the resolution asymmetry in the data, it did not confirm there *were* discrepancies between thresholds readings and actual threshold levels.

To verify whether the actual thresholds deviated from the measured values, a set of dedicated measurements were made. The setup is shown in Figure 5.18. In these measurements the high voltage was tuned down to $1090\ \mathrm{V}$ in order to get more hits with sizes comparable to the thresholds. When the thresholds change, the hit rates can change significantly. First, a baseline measurement was made. The $^{55}\mathrm{Fe}$ source was placed about $20\ \mathrm{cm}$ from the panel edge closer to the CAL end. The CAL thresholds were set to nominal $12\ \mathrm{mV}$[5] while the HV ends had thresholds of $\sim 6\ \mathrm{mV}$. With such low thresholds on the HV side, hits were triggered depending primarily on the CAL side thresholds. The hits were then collected for about 5 minutes. Then the setup was changed to the test configuration, where the source was placed the same distance away from the HV end. In these runs, the CAL side thresholds were set at low values of $\sim 6\ \mathrm{mV}$. The tested HV side thresholds were set to $12\ \mathrm{mV}$, as well as the discrete DAC digital values in the range of $[n-5, n+2]$, where n is the

[5]The values from threshold readings at this moment, as mentioned in Chapter 4 while explaining the algorithm to find the threshold settings, would be slightly lower than $12\ \mathrm{mV}$.

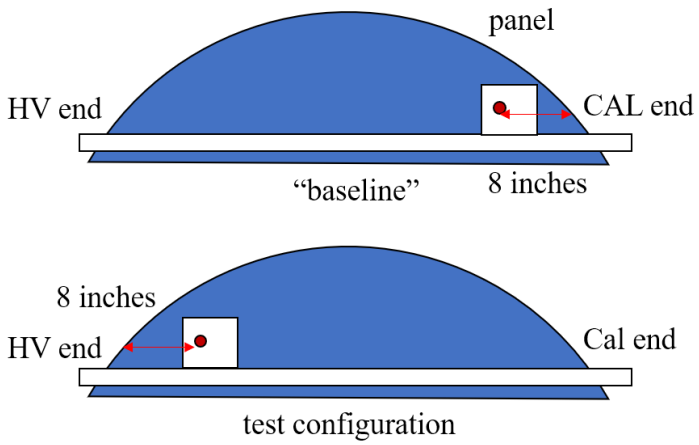


Figure 5.18: Setup for measurements to compare actual thresholds and threshold readings.

setting for $12\ \mathrm{mV}$. As mentioned in Chapter 4, each unit in the digital value corresponds to roughly $-1\ \mathrm{mV}$, so the tests included thresholds from roughly $10\ \mathrm{mV}$ to $17\ \mathrm{mV}$. And in these runs, they determined the triggering decisions. For each configuration, a 5-minute run was recorded.

The data processing approach was based on the following understandings. If the preamp gains on the two straw ends are identical, the underlying spectra for all runs should be the same after normalization (Figure 5.19(a)). The symmetry in the source locations ensures that any potential influences from signal attenuation are cancelled out since the ADCs take signal sums of the two ends. Now depending on the threshold level, only hits with sufficiently large amplitudes can cause a trigger. This means the lower ends of the spectra expect different levels of deficits in the number of triggers depending on the threshold levels, while the higher ends remain the same. As waveforms are reshaped in the electronics, the cutoffs corresponding to the thresholds are not clean cuts; the boundaries are slightly smeared out. If we compare a spectrum with a higher threshold against one with a lower threshold, at the lower end the former would see none of the triggers the latter sees. It sees a fraction of the triggers of the latter around its threshold, and the same numbers at even larger amplitudes. Taking a bin-wise division between the two yields an activation curve as illustrated in Figure 5.19(b). The transition position (i.e., the threshold) can be extracted through fitting a Sigmoid function to the activation curve. Then comparing the transition positions gives a comparison of thresholds. Even in the case when there is a preamp gain difference between the two ends and at higher ends the spectra no longer match, the process to find the threshold can still be applied after a scaling of signal amplitudes, which equivalently stretches the spectra along the x-axes.

For data processing, all datasets needed to be cleaned and normalized first. Apart from the usual cut on hit amplitude (amplitude larger than 5 ADC units) that rejects high-frequency noises, additional cuts on time division were applied to each channel to deal with cosmic-ray hits, as shown

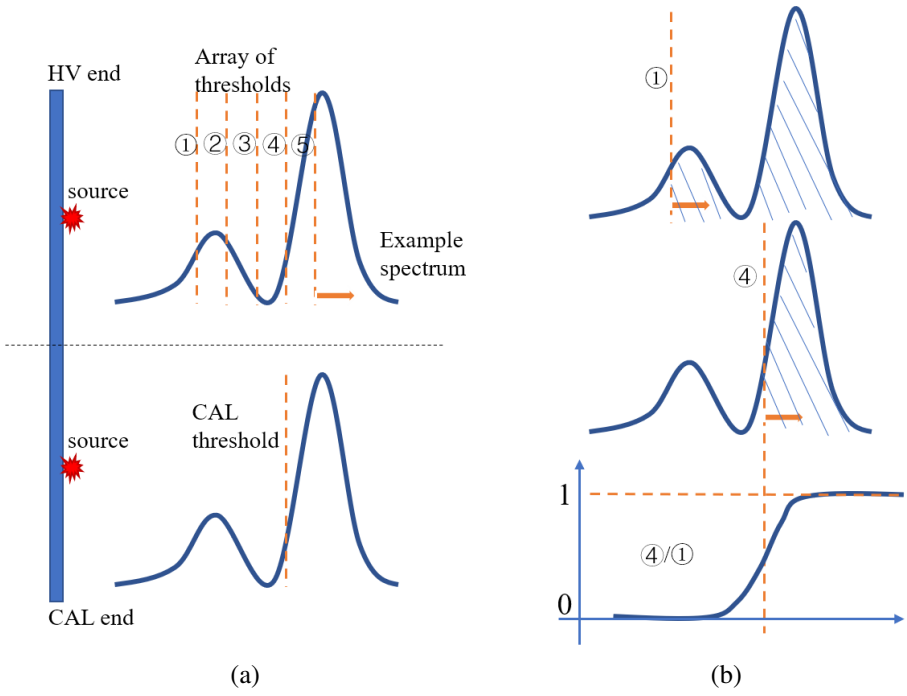


Figure 5.19: Threshold and activation curve. (a) Assuming the preamp gains on the two ends are identical, the symmetry in the data-taking approach ensures the underlying spectra are the same for the runs taken with different source locations. For different thresholds, the lower ends of the resultant spectra expect different levels of deficits in the number of triggers. (b) A bin-wise division between the spectrum with a higher threshold and that with a lower one yields an activation curve, which provides information on the threshold.

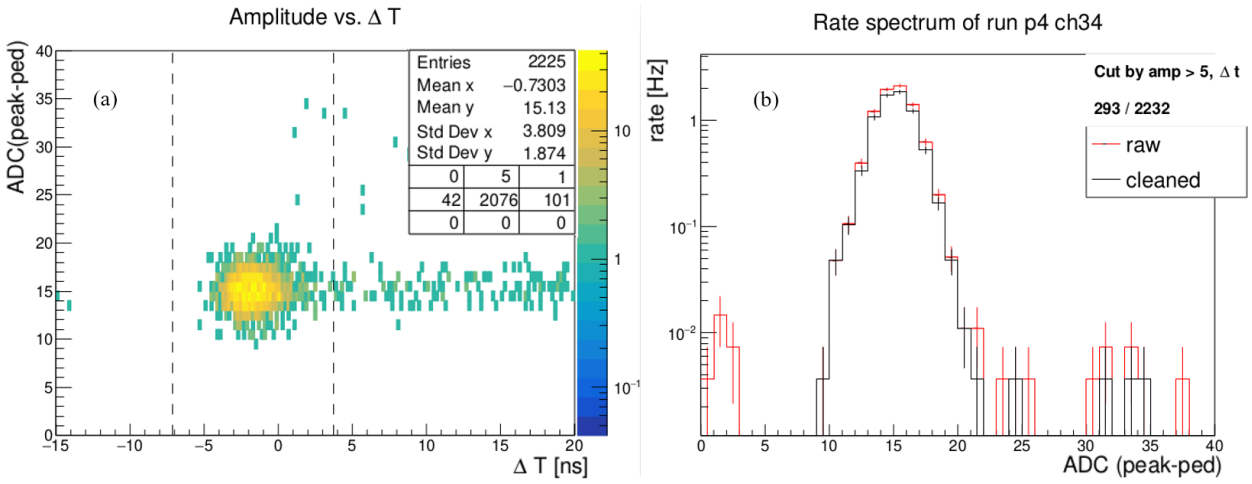


Figure 5.20: Threshold examination data cleaning and normalization. (a) Cosmic-ray hits were cleaned using cuts on time division. In this 2-D histogram of time division and amplitude, the nonlocalized hits away from the cluster corresponded to cosmic-ray hits. The cuts are indicated by the two vertical dashed lines. (b) An example of a rate spectrum normalized by time.

in Figure 5.20(a). The spectra were then normalized by time to obtain what are referred to as the rate spectra. Instead of the number of hits on the y-axes, the rate spectra use the hit rate. This

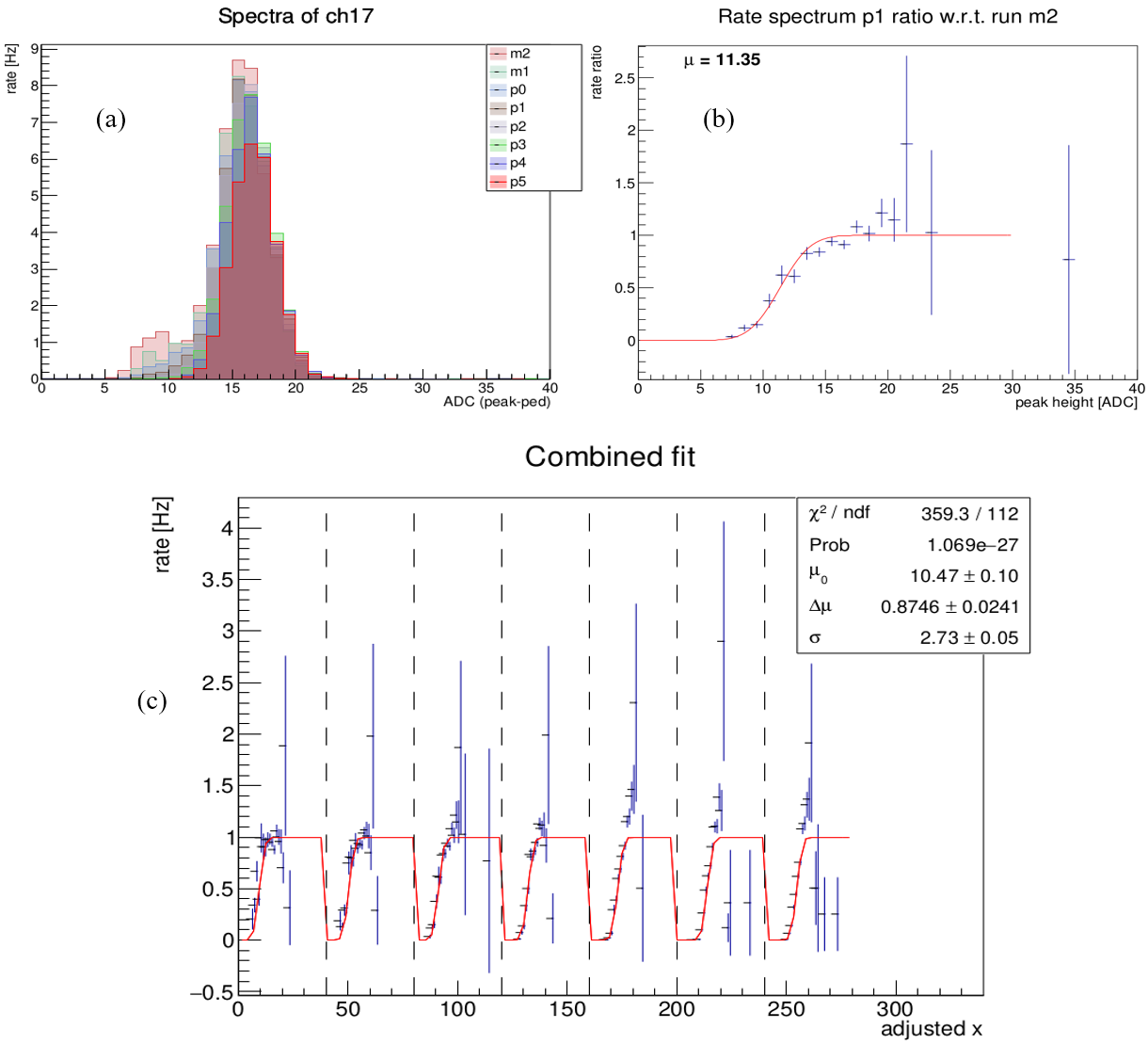


Figure 5.21: Extraction of thresholds from activation curves. (a) Example rate spectra of a channel with different thresholds are stacked together. The deficits on the low ends corresponding to the thresholds are clearly shown. The marked "m2", "m1", ..., "p5" correspond to the threshold digital settings from $n+2$ to $n-5$, where n is the setting for 12 mV. (b) An example of an activation curve. All the spectra were compared to the run "m2" ($\sim$ 10 mV HV threshold). (c) To limit the degree of freedom of the fitting function, the activation curves were fitting together using 3 physical parameters.

normalization was based on the fact that the source emission rate is approximately constant over a short period of time. Figure 5.20(b) shows an example of a rate spectrum.

With the obtained rate spectra, those of the same channel under different HV thresholds were collected for comparison (Figure 5.21(a)). In each channel, the spectra were compared to the one with $\sim$ 10 mV threshold on the HV side. The activation curves were calculated through bin-wise division. One example is shown in Figure 5.21(b). To extract the thresholds, the activation curves needed fits with some Sigmoid function. In this case a Gauss error function was used. In the

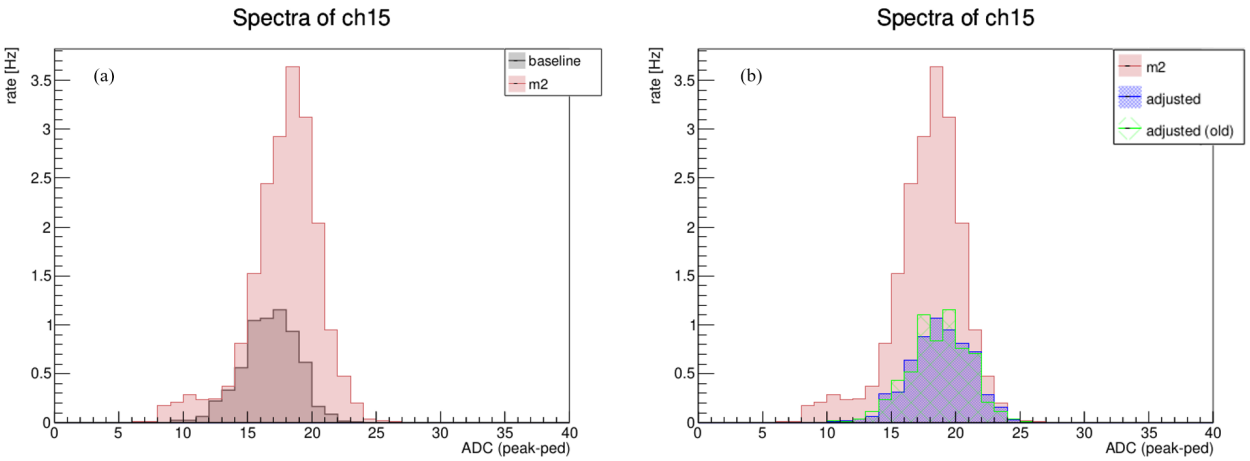


Figure 5.22: An example of a channel with relatively large gain correction. The rate spectra both before ("baseline" in (a)) and after the scaling ("adjusted" in (b)) are shown. The scaling (along x-axis) makes the higher end of the corrected spectrum match better with the others. The "adjusted (old)" tag in (b) represents another corrected spectrum using an earlier dataset before the scan studies. Similar results were obtained.

trials, it was noticed that fitting each activation curve individually yielded relatively poor results due to large uncertainties in some bins. Considering the thresholds were supposed to be evenly spaced, instead the curves were fit together. Only three parameters with actual physical meanings were used. The parameter μ_0 represents the corresponding ADC value of the $\sim 12\ \mathrm{mV}$ HV threshold, $\Delta\mu$ is the shift in ADC units with each equal increment in threshold settings, and σ is the width of all individual error functions. By assembling the activation curves side by side as shown in Figure 5.21(c), the combined plots were fitted using a piecewise function constructed by the three parameters mentioned above. The fitting results were then used to calculate the transition positions in each activation curve. For instance, the run in Figure 5.21(b) with an HV threshold of about $13\ \mathrm{mV}$ had the transition at $\mu = \mu_0 + \Delta\mu = 11.35$ ADC units.

On the other hand, the spectra in the baseline runs are first scaled to account for possible gain differences between the HV and CAL preamps. In this case the scan data were used; in the spectra of corresponding channels and appropriate source locations, the positions of the $5.9\ \mathrm{keV}$ peaks were extracted using local Gaussian fits, and a gain scaling factor was obtained as the ratio between the peak ADC positions. While most of the channels had scaling factors very close to 1, few channels needed the correction to make appropriate comparisons. Figure 5.22 shows one of them. As the figure shows, the higher end of the rate spectrum of this channel aligns poorly with the others before the correction, and significantly improved after the scaling was applied.[6]

After scaling, the activation curves of the baseline spectra corresponding to a (nominal) $12\ \mathrm{mV}$ CAL threshold were obtained following the same process as previously discussed. Fits using

[6]While the individual hit amplitudes were scaled with the factor, the overall spectrum shape changes slightly due to different binning.

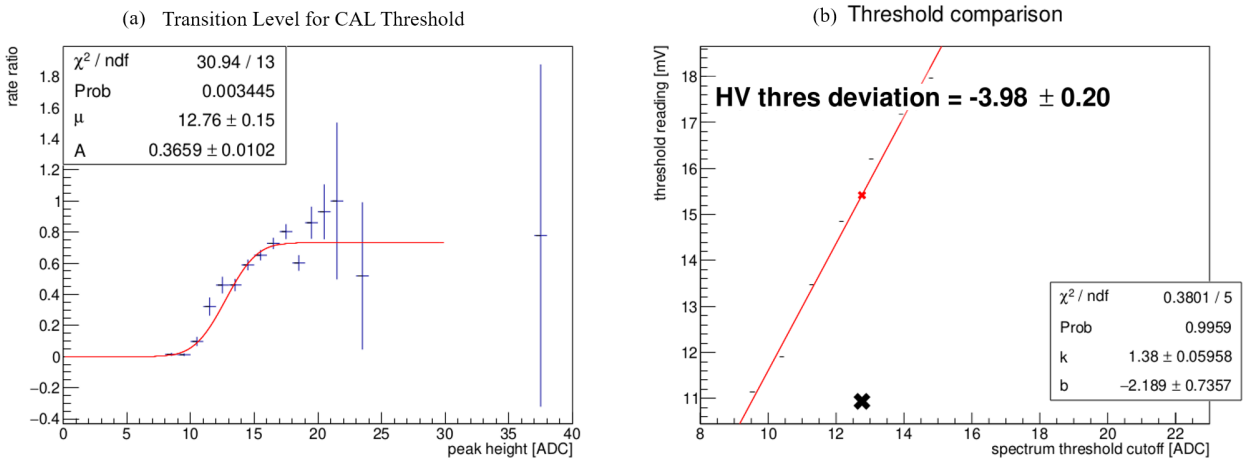
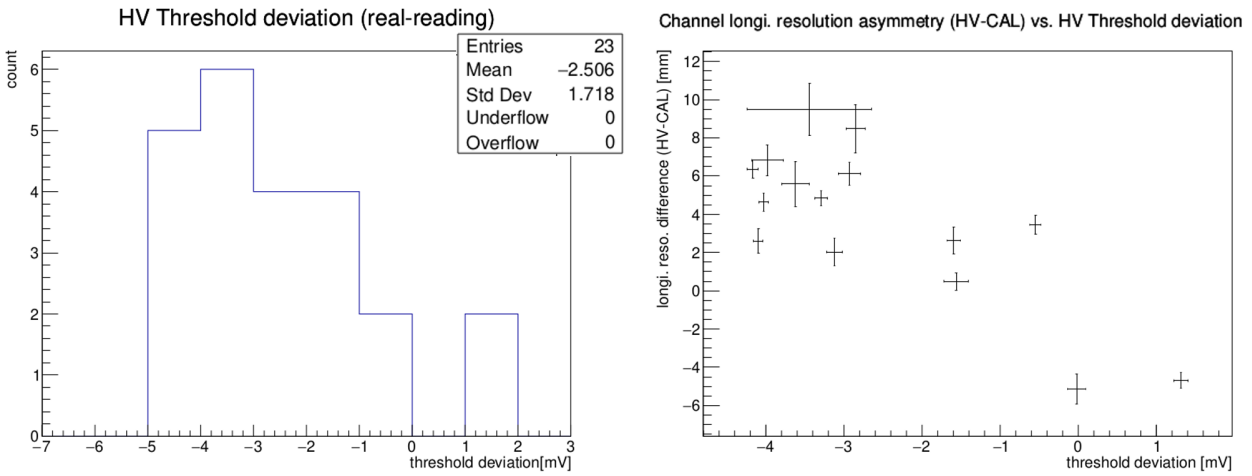


Figure 5.23: Threshold comparison using transition positions in activation curves. (a) The CAL thresholds were fitted using the same error function widths from the HV side. (b) HV thresholds were compared to the CAL threshold used as the baseline.

error functions with fixed σ values inherited from the HV threshold fitting results provided the CAL thresholds in the ADC scales. One example is provided in Figure 5.23(a). Then comparisons can be made. An example is shown in Figure 5.23(b) for one straw channel. The x-axis here represents the transition positions from the fits and the y-axis shows the threshold readings. The data points corresponding to the HV thresholds are in a straight line as expected; the slope of the straight line tells us that for every ADC-unit shift in the fitted transition position, the threshold reading changes by $1.38\ \mathrm{mV}$. Meanwhile, the black cross marker represents the CAL threshold. It was fitted to be at 12.76 ADC unit, and the threshold reading from the slow control was about $11\ \mathrm{mV}$. Since the gain have already been corrected, the same transition position in activation curve fits should correspond to the same actual threshold level (up to a gain scaling factor that is almost 1). If we trust the CAL threshold readings (or just to choose it as a reference), the actual threshold on the HV side corresponding to 12.76 ADC unit should be about $11\ \mathrm{mV}$ as well. However, the readings on the HV side was roughly $15\ \mathrm{mV}$. The detailed calculation factoring in the gain scaling factor showed that the actual HV threshold in the channel was $3.98\ \mathrm{mV}$ lower than the reading.

Among the straws measured in this test, on average the actual thresholds on the HV side were $2.5\ \mathrm{mV}$ lower than the readings, assuming the CAL side threshold readings were correct. From a different perspective, the threshold readings on the HV side would be systematically higher than those on the CAL side by $2.5\ \mathrm{mV}$ when they were actually set to the same. This was shown in the histogram in Figure 5.24(a). Recalling the results from the previous section, such threshold deviation was the right amount to cause the level of longitudinal resolution asymmetry saw in the scans. In addition, when plotting the channel longitudinal resolution differences between the two ends against the threshold deviations from the readings as shown in Figure 5.24(a), there was a



(a) (b)

Figure 5.24: Threshold deviations from the readings. (a) Histogram of the HV threshold deviation as differences between those from activation curve fitting and those from the readings. (b) When the longitudinal threshold differences between the HV and the CAL side were plotted against the threshold deviations from the readings, the plot showed a good level of correlation between the two.

good level of correlation between the two.

In conclusion, the threshold examination revealed the existence of a $2.5\ \mathrm{mV}$ deviation (on average) in the HV threshold readings. It accounted for the most, if not all of the $3.8\ \mathrm{mm}$ straw longitudinal resolution asymmetry observed during the scans. Although its influence on panel longitudinal resolution was not large enough to drive it above the upper limit required to reach the designed Tracker momentum resolution, one may wish to avoid the resultant regions with worse resolution values. Mitigation options may include some type of calibration during detector commissioning.

* * *

Despite the small deviation in HV threshold readings and the resultant longitudinal resolution asymmetry, in the $^{55}\mathrm{Fe}$ source scan studies, it was verified that the straw longitudinal resolutions ($\sim 30\ \mathrm{mm}$) satisfied the $40\ \mathrm{mm}$ experiment requirement, and a map of effective signal propagation speeds on a panel was generated to improve reconstruction accuracy. The obtained data, along with those from the cosmic-ray studies, will be used to assist future simulation tuning.

From the VST, valuable lessons on Tracker behaviors and detector operations were learnt. Experiences with the Tracker have been accumulated in preparation for the upcoming detector and experiment commissioning. All the efforts are contributing to a successful Mu2e experiment.

References

[1] R. Bonventre and V. O'Dell, "Mu2e Tracker Vertical Slice Test Plan," (2020-04-15), mu2e-doc-32725-v18.

[2] R. Bonventre, "Searching for muon to electron conversion: The mu2e experiment at fermilab," (Lake Louise Winter Institute 2020-2-14), `https://indico.cern.ch/event/846070/`.

[3] R. Bonventre *et al.*, "Longitudinal dependence of Fe-55 signals with the 8-straw tracker prototype," (2016-07-11), mu2e-doc-7722-v2.

[4] R. Bonventre, "Tracker Vertical Slice Test Status," (2022-06-01), mu2e-doc-42370-v1.

Chapter 6 Research and Design: Beam Extinction Monitor

Before conclusion, this second final chapter will be spent to discuss a non-Mu2e but still related Research and Design project. The goal was to design a Beam Extinction Monitor for Proton Improvement Plan-II (PIP-II) which would provide the primary beam for the contemplated Mu2e-II experiment. The approach to the beam extinction measurement will be introduced, and then the focus will be on the investigations into a suitable electron detector used in the system. The conducted proof-of-principle measurement will also be discussed, along with proposed further testing plans.

6.1 Proton Improvement Plan-II and Mu2e-II

The Proton Improvement Plan-II (PIP-II) is an enhancement project of Fermilab's existing accelerator complex including a series of upgrades and improvements currently under active construction. While PIP-II also involves modifications to the Fermilab Booster, Recycler and Main Injector, its central piece is a new superconducting Linac accelerating negative hydrogen ions (H^-) [1]. The Linac is designed to accelerate up to $2\ \mathrm{mA}$ of beam current to $800\ \mathrm{MeV}$; the peak currents can be up to $10\ \mathrm{mA}$ for less than a few microseconds.

In the starting stages, the Linac contains a radio frequency quadrupole (RFQ) that accelerates $30\ \mathrm{keV}$ H^- ions from an ion source to $2.1\ \mathrm{MeV}$, which operates at $162.5\ \mathrm{MHz}$ [1]. The $162.5\ \mathrm{MHz}$ bunches coming out of the RFQ then pass through a bunch-by-bunch chopper consisting of two kickers, which is capable of removing any unwanted bunches to an absorber while causing minimum distortions to the ones left in the Linac [1]. The chopper allows the bunch sequence to be set by requirements coming from the experiments.

The PIP-II Linac will operate in a quasi-continuous wave (CW) mode [1]. In a $20\ \mathrm{Hz}$ cycle (which is determined by the Booster cycle), the first $\sim 3\ \mathrm{ms}$ will be reserved for pulsed Booster injection. The Booster and other accelerator facilities further accelerate the beam for experiments requiring higher beam energies (e.g., the Long Baseline Neutrino Facility/Deep Underground Neutrino Experiment (LBNF/DUNE) program). For the rest of the $20\ \mathrm{Hz}$ cycle, the Linac operates in a full CW mode. With each user acquiring different bunches in the sequence separated by a three-way radio frequency (RF) split, the PIP-II Linac can supply beam for multiple experiments

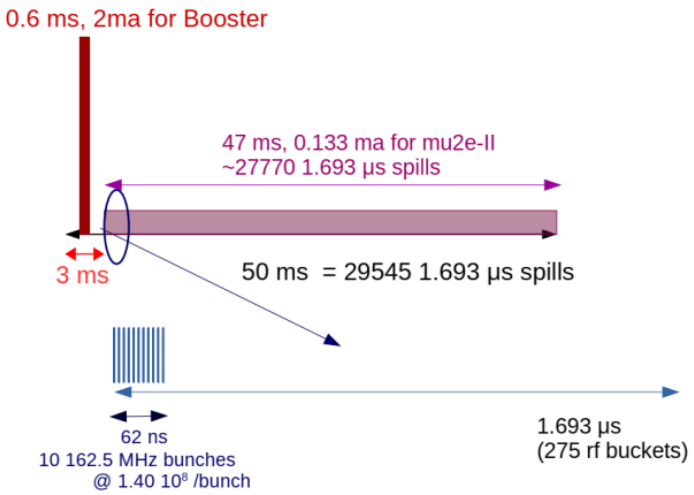


Figure 6.1: A schematic illustration of a possible bunch sequence for Mu2e-II. The beam structure can be configured at will when the Linac operates in the CW mode.

simultaneously [1, 2].

Among the potential users is Mu2e-II, a possible upgrade of the Mu2e experiment. With multiple detector upgrades being discussed and a further order of magnitude improvement in sensitivity in mind, it is a natural continuation of the Mu2e experiment [2]. As mentioned in Chapter 2, if Mu2e (or COMET) finds a CLFV signal, further experiment using different stopping target materials will give better understandings of the underlying physical mechanism; otherwise a search with an extra order of magnitude sensitivity can provide stronger theory exclusion power [2].

Using the PIP-II Linac beam for the Mu2e-II experiment will bring several advantages. The PIP-II beam has better quality then the current proton beam [1]. The $800\ \mathrm{MeV}$ energy level is well below the antiproton production threshold so backgrounds related to the delayed antiprotons will not be an issue anymore. Without the batch structure in the old proton beam, the experiment no longer needs to sit in idle when the beam is delivered to other experiments. In addition, PIP-II can deliver more protons (after stripping the electrons of H^- ions) to the experiment in each pulse. In a tentative plan, Mu2e-II will use $100\ \mathrm{kW}$ out of the $1.6\ \mathrm{MW}$ beam power of the PIP-II Linac. Within the $47\ \mathrm{ms}$ Linac CW operation, the experiment will use ten $162.5\ \mathrm{MHz}$ bunches every $1.693\ \mu\mathrm{s}$, each bunch containing 1.4×10^8 particles, as illustrated in Figure 6.1 [2]. In each $1.693\ \mu\mathrm{s}$ pulse, Mu2e-II receives roughly 40 times more protons than the Mu2e experiment. And even if the experiment decides to use a different stopping target material that corresponds to some different muonic atom lifetime, which requires modification to the intervals between pulses, the PIP-II Linac in the CW mode can easily provide the desired beam structure through configuring the bunch sequence.

The desired upgrade in the experiment sensitivity requires an order of magnitude improvement in the beam extinction to reach $< 10^{-11}$ [2]. Similar to Mu2e, the extinction of Mu2e-II will

also be achieved in two stages. The first comes from the PIP-II Linac itself in the chopping efficiency of the bunch-by-bunch chopper. It is set to be 10^{-4} as required by the Linac specifications [1], but a better level is expected in the actual performance [2]. The second stage will use an upgraded version of the Mu2e extinction system (as discussed in Chapter 2), which needs to have an extinction level of 10^{-9}, leaving a two-orders-of-magnitude safety margin. The goal of our project was to investigate for a Beam Extinction Monitor that monitors the extinction level in the Linac. The monitor was tentatively planned to be inserted into a $25\ \mathrm{MeV}$ section of the beamline.

6.2 Beam Extinction Measurement of Negative Hydrogen Ion Beam

Typically, beam monitoring for proton beams are achieved through wire targets lowered into the beamlines. An example is the Bunch Shape Monitor (BSM); one of them was used in the 400 MeV section of the Fermilab Linear Accelerator [3, 4]. As shown in Figure 6.2, when a proton bunch passes through the wire target, low energy secondary electronics are emitted. The secondary electron flux is proportional to the proton flux in the bunch. Typically, the wire is connected to a negative high voltage, so the electrons are accelerated. Some of them pass through the input collimator. A deflector with a periodic AC sweeping voltage (phase-locked to the beam period) deflects the electrons and translates the flux difference in time to that in spatial coordinates. Some type of electron collector behind an output collimator capable of moving along the direction of signal separation measures the electron flux at various locations, from which the initial proton bunch shape can be reconstructed. While the device was specifically designed for bunch shape measurements, some modifications (e.g. changes on sweeping voltage frequency, output collimator and electron detector area) can be applied to allow the electron detector to measure the total rates during and out of a bunch, which provides an extinction measurement for the proton beam.

The above approach, however, cannot work for a H^- beam due to its special characteristics. In a H^- ion, the first ionization potential is $0.756\ \mathrm{eV}$ while the other electron has a binding energy of $13.6\ \mathrm{eV}$ [5]. If a wire target is in the H^- beamline, the loosely bound electron can be easily stripped during interactions, causing a beam loss. The Beam Extinction Monitor (BEM) system for H^- should be based on different physical principles. Following the approach in [5, 6], a BEM utilizing H^- electron photoejection was envisioned. A pulsed laser is used to strip the loosely bound electrons in H^- ions when needed, and then the electrons are detected for the extinction measurement.

A schematic drawing of the BEM system is shown in Figure 6.3. The $25\ \mathrm{MeV}\ \mathrm{H}^-$ ions ($\beta = 0.226$) first pass through a magnetic dipole that deflects and separates the ones that have already lost the second electrons. As the electron binding energy is so low, the slightly-bound electron can be stripped by black-body radiation, magnetic field, and residue gas [7]. This first

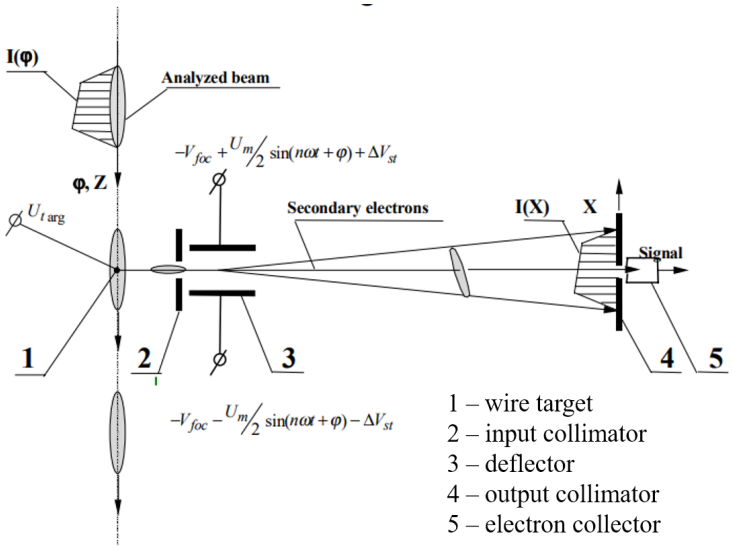


Figure 6.2: A schematic drawing of a Bunch Shape Monitor [3].

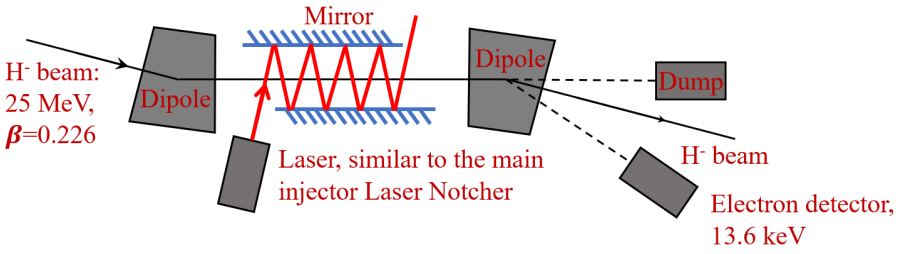


Figure 6.3: A schematic drawing of the envisioned Beam Extinction Monitor.

dipole disposes the free electrons that possibly exist from these stripping processes.

Then the beam passes through a laser system, which is used to detach the second electrons for later measurements. Unlike the wire target, such a laser system has no additional mechanical structure in the beamline; the occurrence of the electron detachments can be controlled using a pulsed laser, which is only activated when there is no bunch in the beamline, avoiding disturbing the normal particle bunches. Similar to the reported system in [5, 6], a Q-switched Nd:YAG laser can be used. It can produce photons with a wavelength of $1064\ \mathrm{nm}$ in $100\ \mathrm{mJ}$, $10\ \mathrm{ns}$ pulses. Inspired by the laser notcher design used in the Fermilab Main Injector [8], a mirror system is installed to maximize the electron photoejection rate in the system. As the parallel mirrors reflect the laser beam, it crosses the H^- beamline multiple times and multiplies the effective intensity of the laser, allowing a reduction in the beam pulse energy. Another factor influencing the interaction rate is the photoejection cross-section. As shown in Figure 6.4 [6], the interaction cross-section is a function of photon wavelength. A peak is reached for photons of wavelength $\lambda \simeq 820\ \mathrm{nm}$. While the laser wavelength is unchangeable, in the center of mass frame of the ions, the appeared photon

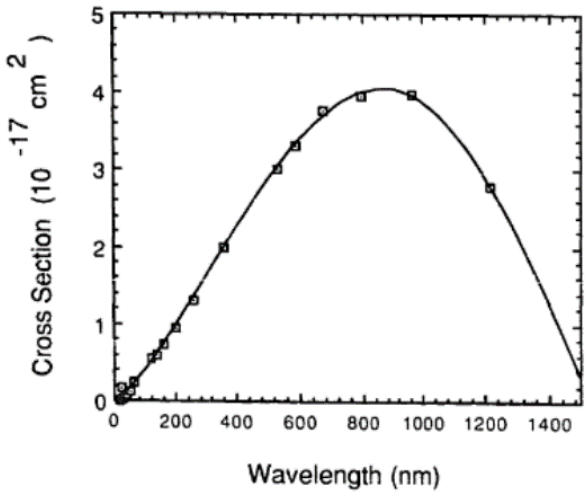


Figure 6.4: (Left) One-electron photoejection cross-section of H^- ions. The curve peaks at a photon wavelength of about 820 nm [6].

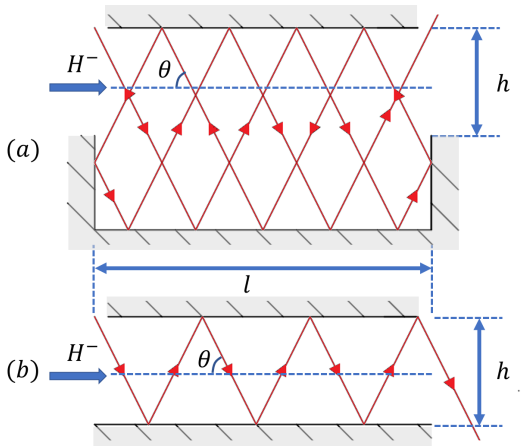


Figure 6.5: (Right) An example of Beam Extinction Monitor mirror system optimization. More exotic mirror designs can potentially further improve the efficiency of the laser system given a fixed beam power. In the illustration (a) and (b) have the same length l, the same clearance between mirrors d and the same laser incident angle θ, but the implementation of the corner reflectors increases the number of laser beam passings.

wavelength can vary due to the relativistic Doppler effect. At $\beta = 0.226$, it can be calculated that the smallest possible Doppler-shifted wavelength from a source of $\lambda = 1064$ nm is 845.4 nm, when the laser points along the same direction as the ions. While this is unrealistic, a smaller angle between the laser and the H^- beam is preferred from this perspective. However, with a limited space to install the mirror system, in a parallel mirror design with some fixed length, a smaller incident angle sacrifices the number of beam passings. Some optimizations can be made considering the beam speed, the mirror reflectivity, and the available installation space. And if more exotic mirror systems are considered as shown in Figure 6.5, further optimizations are possible. A combination between a more complicated mirror system and a less powerful laser can be a more cost-efficient option while remaining the same level of electron photoejection rate.

After the H^- ions pass through the laser system, some fraction of the ions are neutralized to hydrogen atoms and free electrons, all moving at almost the same speed as the H^- ions. With the assistance of a second magnetic dipole, these species can be separated. The hydrogen atoms are directed to a dump, the remaining ions continue their path down the accelerator, and the electrons with 13.6 keV kinetic energies are channeled to an electron detector for detection. From the number of electrons detected during the time intervals when there is supposed to be no beam in the beamline, the extinction rate can be calculated accordingly. A high-performance electron detector is desired.

6.3 Selection of Electron Detector

The electron detector used in the BEM needs to have good performance characteristics when detecting $\sim 10\ \mathrm{keV}$ electrons. A good candidate should have a high detection efficiency and a low background rate, and have the capability to detect a single electron. It should be able to function in a vacuum, near a magnetic field, and sustain high levels of radiation. In addition, a small-sized device with a simple system is preferable due to space limits.

There are two major categories of often-used low-energy electron detectors, which are secondary electron multipliers and solid state detectors. They will be separately discussed below. A decision was made though, to use a vacuum-ultraviolet (VUV)-Multi-Pixel Photon Counter (MPPC) for the electron detection. This is a type of solid state detector that appear to best satisfy the requirements.

6.3.1 Secondary Electron Multipliers

As the name suggests, the secondary electron multipliers are based on the detection principle of secondary electron multiplication. Examples in this category include electron multipliers and microchannel plates (MCPs).

As shown in Figure 6.6 [9], the electron multipliers contain a series of secondary emission electrodes, a.k.a. dynodes [9]. When an electron impacts a dynode, its energy is transferred to the electrons in the dynode material, and some secondary electrons escape the dynode surface. The dynodes are connected to incrementing high voltages, which form electric fields between successive dynodes, directing and accelerating the escaped electrons from one stage to the next. The voltage differences between successive dynode stages, typically around a few hundred volts, are optimized for a high secondary electron emission coefficient (stage gain) in the dynodes. The total number of electrons is multiplied in each dynode and grows exponentially over multiple stages. The electron multipliers usually have high gains; typically one with 10 to 14 stages can reach a total gain of 10^7 [9]. The noise levels in electron multipliers are normally low. They have good single-electron detection efficiency at low energies ($\sim 1\ \mathrm{keV}$, depending on the dynode material), but due to reasons that will be discussed in a moment, do not maintain the same efficiency level when primary electron energy increases above the optimal range. In addition, they are susceptible to influences from magnetic fields.

An MCP is essentially an array of tiny electron multipliers [10]. As shown in Figure 6.7, it consists of a glass plate with microscopic channels oriented parallel to each other [9, 10]. A typical MCP contains 10^4 to 10^7 microchannels each with a diameter of $10\ \mu\mathrm{m}$ to $100\ \mu\mathrm{m}$ [9]. The two surfaces of the MCP are conductive; a high voltage is applied between the two sides to provide electric field for the microchannels. The inside walls of the microchannels are treated so that they

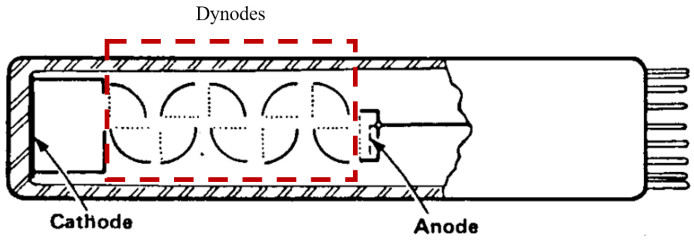


Figure 6.6: An example electron multiplier in a photomultiplier tube (PMT) [9]. The electron multipliers are common components in PMTs. The only difference is that in a PMT, there is an additional photocathode converting primary photons to electrons. Here the dynodes arranged in a box and grid configuration are highlighted.

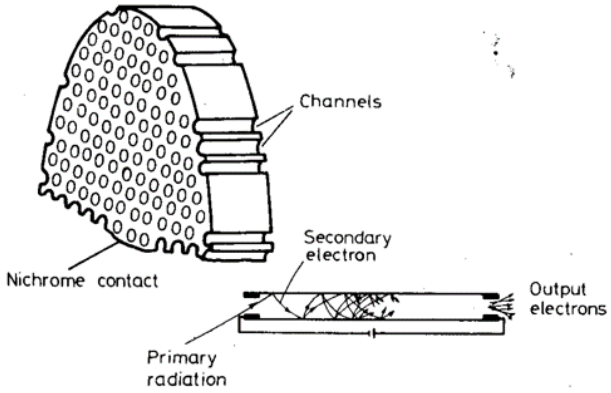


Figure 6.7: A schematic illustration of a microchannel plate [9]. A cross-sectional view is also provided for a single microchannel to show its detection mechanism.

can emit secondary electrons. Each microchannel acts like an electron multiplier with a continuous dynode; incident electrons are multiplied while bouncing between walls and emitting secondary electrons. A typical gain of the MCP is in the range of 10^3 to 10^4; overlaying two or three MCPs in what is referred to as a Chevron configuration can increase the gain significantly [9]. The MCPs are also capable of single-electron detection. Due to their compact sizes, they are more robust with magnetic fields. However, the MCPs also lose efficiency with an increasing incident electron energy.

The reasons to the decreasing detection efficiency in the secondary electron multipliers originate from the electron stopping power in materials. As previously shown in Figure 3.10 as a typical example, in the low-energy range the collision stopping power decreases with the increasing electron energy. This means in the first dynode stage (or in the first dynode collision in the case of an MCP), electrons with higher energies in this range travel further in the material. For primary electrons with higher energies, they tend to produce secondary electrons deeper inside the dynodes. As secondary electron ranges do not increase as much as the primary electrons, the probability of secondary electrons escaping the dynode decreases. When there is no escaped secondary electron in the first stage, there will be no detection for the primary electron, leading towards a lower detection

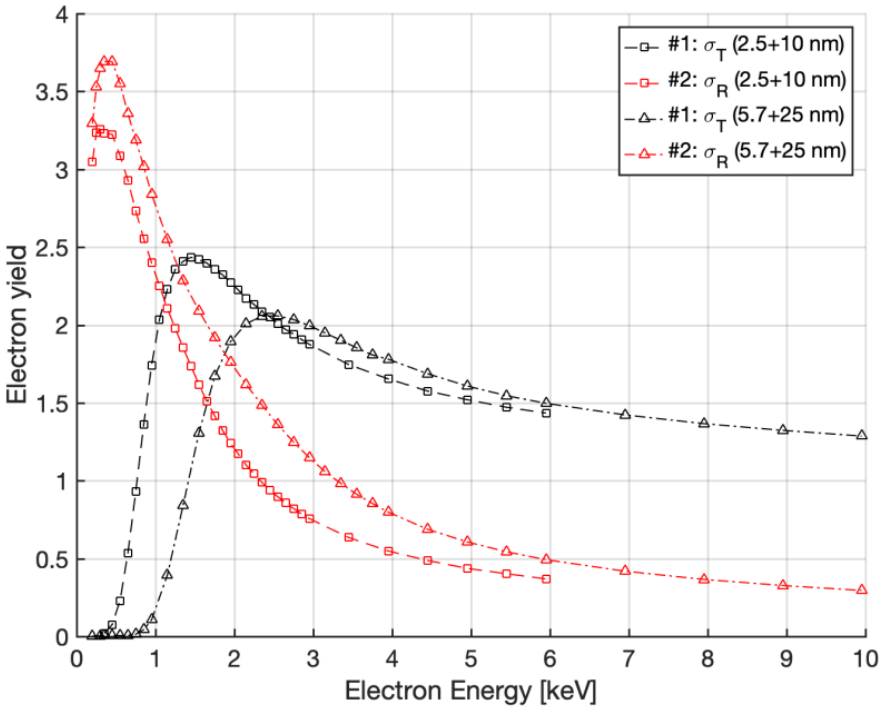


Figure 6.8: An example of secondary electron emission coefficients for different primary electron energies [11]. The test multilayered Al_2O_3/TiN membranes were developed for transmission dynodes in novel vacuum electron multipliers. σ_T and σ_R respectively represent transmission yield and reflection yield, while triangle and square differentiate different membrane thicknesses. Although dynodes with different materials and geometries have different secondary electron emission coefficient curves, these are representative enough to show the trend.

efficiency. This can be confirmed in the trend of the secondary electron emission coefficient. An example is shown in Figure 6.8 [11]. Secondary electron emission coefficients of some novel transmission dynodes are plotted for different primary electron energies. The reflection yields are optimized below an incident electron energy of 1 keV, while the transmission yields are optimized around 2 keV. With an increase in the primary electron energy above those optimal values, the electron yield decreases. At 10 keV, the highest transmission yield is roughly 1.3. Using Poisson statistics, there is an intrinsic deficiency of 0.27 from the detection mechanism alone. For the higher electron energy we are looking at, the deficiency will be higher; and commercial devices are likely to be worse. If an MCP or some other secondary electron multiplier is used for the BEM, there will also be little robustness in terms of using the device in other sections of the beam. In most other beam sections the electron energies obtained would be only higher, meaning even lower intrinsic detector efficiencies. This gives poor compatibility with other installation options. In addition, the high voltage supply required to operate the device will bring additional constraints in clearance and so on; extra efforts will be required to channel the high voltage supplies into the beam vacuum.

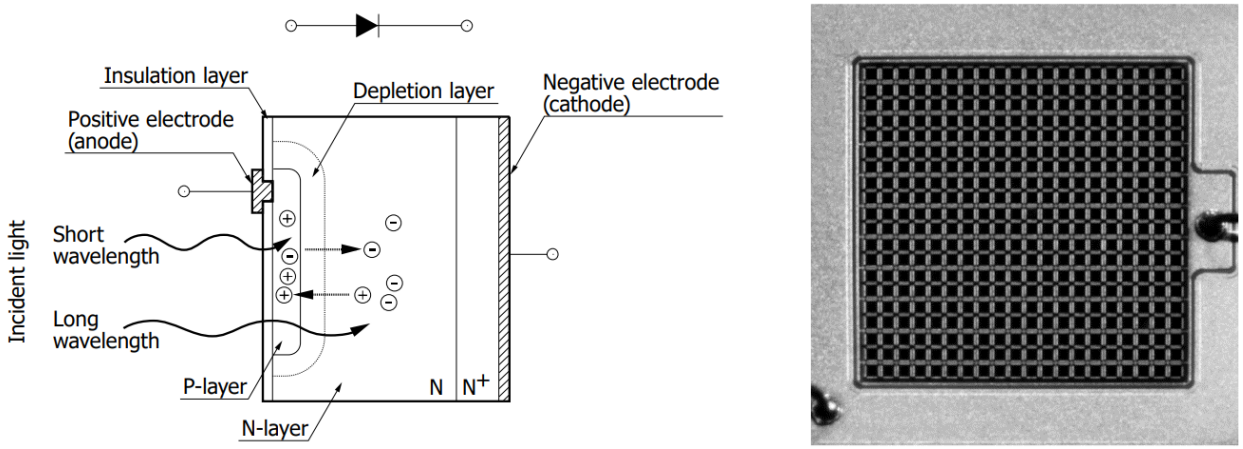
6.3.2 Solid State Devices

Solid state devices provide an alternative to the discussed secondary electron multipliers.

These devices are based on PN junctions, which are formed between P-doped and N-doped semiconductors [12]. A charge carrier concentration gradient forms across a PN junction. As a result, electrons in N-doped layer (or holes in the P-doped layer) flow towards the other side of the PN junction to recombine. The immobile ionized donors remain in their original places and form a depletion layer. As a result, both sides of the PN junction are charged, which give rise to an electric field between the P and the N layer. The field eventually balances with the charge carrier concentration gradient so charges stop flowing, bringing the system into an equilibrium. However, when an external source (for instance, thermal agitation, photoelectric effect, and electrons form the exterior) creates a surplus of charge in the device, the equilibrium is disturbed. A current then flows through the device. If an additional reverse bias voltage is applied across the device, the electric field in the depletion layer is intensified. The charge carriers will obtain larger kinetic energy during their drift. With a sufficiently large reverse bias voltage, collisions between charge carriers and lattice atoms in the material will cause additional ionizations. An avalanche multiplication of charges will follow, which gives rise to readout signals through charge induction. From the above descriptions, all solid state devices in principle can perform electron detection.

There are profound resemblances between the detection mechanism in the solid state devices and that in gaseous ionization detectors as discussed in Chapter 3. Similarly, the level of carrier multiplication in the solid state devices is influenced by the reverse bias voltage. At close-to-zero reverse bias voltages, the charge carriers do not have enough energy to cause additional ionizations. PN diodes fall in this category, parallel to ionization chambers among the gaseous ionization detectors. Avalanche Photodiodes (APDs) work with higher reverse bias voltages below the breakdown voltage, which allow proportional multiplications of the charge carriers. This corresponds to proportional counters. And at bias voltages above the breakdown voltage, devices such as Geiger-mode Avalanche Photodiodes (GAPDs) (Figure 6.9(a) [12]), as the name suggests, function in the Geiger mode. The output signal sizes are saturated and independent of the incoming particle energy. In particular, when connecting a matrix of GAPDs (pixels) to a common reverse bias voltage and a common signal output in parallel, one obtains a Silicon Photomultiplier (SiPM), a.k.a., MPPC, as shown in Figure 6.9(b) [12]. From the detection principles, the signal size from an MPPC is solely determined by the number of pixels activated [12].

Although solid state devices in principle have electron detection powers, when it comes to detecting ~ 10 keV electrons in practice, some of them cannot perform the task. Oftentimes on the surface of a device (Figure 6.9(b)), a dead layer of silicone resin or other materials are applied and serve as a form of mechanical protection and quenching resistor [12]. For a quick estimation, at 13 keV, total electron stopping power in silicon can be found from [13] to be



(a) (b)

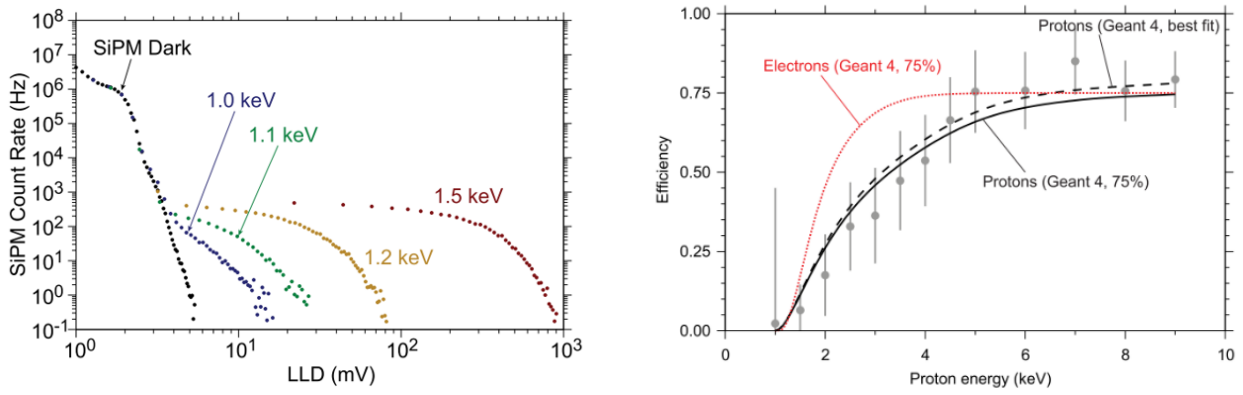
Figure 6.9: Geiger-mode Avalanche Photodiode and Multi-Pixel Photon Counter. (a) A cross-sectional view of a GAPD device [12]. (b) MPPCs contain a matrix of GAPDs [12].

$S(13\ \mathrm{keV}) = 13.92\ \mathrm{MeV}\cdot\mathrm{cm}^2/\mathrm{g}$. Using a continuous slowing down approximation (CSDA) approach as described in [9], the mean range of the particle can be approximated as:

$$\Delta x = \int_0^E \frac{1}{S(E')}\mathrm{d}E' \,. \tag{6.1}$$

Since for $E' < 13\,\mathrm{keV}$, $S(E') > S(E)$, there is $\Delta x < 13\,\mathrm{keV}/13.92\,\mathrm{MeV}\cdot\mathrm{cm}^2/\mathrm{g} \sim 0.001\,\mathrm{g/cm}^2$. As electrons generally undergoes multiple scattering, the range from the CSDA approach is always over-estimated. Approximating silicone density as $1\ \mathrm{g/cm}^3$, the $13\ \mathrm{keV}$ electrons cannot penetrate $0.01\ \mathrm{mm}$ in silicone. On the other hand, typical silicone resin thickness on, for example, regular MPPC is $\sim 0.5\ \mathrm{mm}$. The electrons will be stopped before they even enter the active volume of the device.

A candidate was suggested in [14] to use a VUV-MPPC. These are devices developed for liquid argon and liquid xenon detectors [15]. The peak emission wavelengths in these experiments are respectively $128\ \mathrm{nm}$ and $175\ \mathrm{nm}$, which are absorbed in regular MPPCs in the protective layer. In the VUV-MPPCs, the protective layers are removed, leaving an unsealed detector window. This largely improves the detection efficiencies for VUV photons, and makes low-energy electron detection possible. The studies in [14] reported detections for $\sim 1\ \mathrm{keV}$ electrons. In the quoted Figure 6.10(a), the x-axis represents the triggering threshold, and the count rates with $\sim 1\ \mathrm{keV}$ electrons under various triggering thresholds are compared to that of the device dark current, which is the triggering of pixels in the device by thermally agitated electrons. Clear rate excess was observed in the figure, showing triggers caused by the electrons. In addition, simulation results presented in the work suggested high electron detection efficiencies for $\sim 10\ \mathrm{keV}$ and above



(a) (b)

Figure 6.10: Low energy electron detection in VUV-MPPC from literature. (a) The studies in [14] demonstrated ~ 1 keV electron detection capabilities in an early generation VUV-MPPC device. (b) Simulations in [14] suggested high detection efficiency for electrons of ~ 10 keV and above; the limiting factor here is the geometrical fill factor of the device.

electrons (Figure 6.10(b)).

For our application, the multi-pixel feature will assist in noise suppressing. The solid state devices are susceptible to dark noises, which are caused by thermally agitated electrons initiating an avalanche in a pixel. While lowering the device temperature can lower the dark noise rates, for single-electron detection purposes, they become a problem if there is no way to differentiate the signals from the dark noises. This would particularly be a problem for single-pixel solid state detectors. In MPPCs, the signal amplitudes (i.e., the number of triggering pixels) appear to provide some differentiation power between the two. When electron avalanches occur in pixels, photons in the visible range can be produced. If traveling to neighboring pixels, such photons can cause extra triggers in them, which are referred to as crosstalks [12]. In cases of thermally agitated electrons (dark noises) or visible light photons, the primary particle energies are typically on an eV scale. The resultant number of photons produced in avalanches are few and the crosstalk is controlled at a low level. Consequently, dark noise pulses are often of low amplitudes corresponding to few pixel hits. On the other hand, the ~ 10 keV electrons deposit more energies in the cells. More secondary photons appear to be produced and more pixels are triggered. (In this case, high-energy secondary electrons can travel to neighboring pixels and cause triggers as well.) This was reflected in the quoted Figure 6.10(a); the majority of the signals of 1.5 keV electrons have amplitudes much larger than the dark noise pulses. Using this difference in pulse amplitude, it is believed that the signals of single electrons can be separated from the noises. In addition, since these VUV-MPPCs were designed for cryogenic temperatures, additional cooling (e.g., through a thermoelectric cooling device) can be safely applied to reduce dark noise rates for further background suppression.

The VUV-MPPCs have more desirable features. They have high gains of $\mathcal{O}(10^6)$ and can

function in magnetic fields and radioactive environments. They are small in size, and do not need high voltages to operate.

6.4 Proof-of-Principle Measurement and Discussions

The VUV-MPPCs were chosen to be used for electron detection in the BEM design. A Hamamatsu 4th generation VUV4-MPPC (S13370-3050CN [15]) was obtained to carry out a proof-of-principle measurement. The goal was to confirm the electron detection capability of the device.

The test setup was placed inside a dark box to avoid influences from photons. A $^{90}\mathrm{Sr}$ source was used as the beta source of the measurement. It emits electrons at $546\ \mathrm{keV}$ [16], which is not the most ideal, but is sufficient to show the device is capable of electron detection. The source was placed about $5\ \mathrm{mm}$ from the MPPC surface to maximize the hit rate.

A test board was made to bias the VUV-MPPC and amplify the output signals. As shown in Figure 6.11, the inverting amplifier design uses a single operational amplifier (op-amp) for a gain factor of $(-)91$. The output was then fed to a 12-bit $500\ \mathrm{MHz}$ ADC and data processing module that was designed for the KOTO experiment at J-PARC [17]. The module contained an Altera FPGA and a firmware program was written to tally the numbers of hits over an array of thresholds. Data were also taken without the $^{90}\mathrm{Sr}$ source to make a comparison with dark noises.

The results are presented in Figure 6.12. Here the red dots indicate data with the presence of the $^{90}\mathrm{Sr}$ source, while the blue ones show the dark noise rates at each threshold level. When compared to the dark noise rates, the rates with the source contained an excess, especially in the high-amplitude bins where the dark noise rates were supposed to be very low. This was a clear indication that it is possible to do single-electron detection with a VUV-MPPC device.

The plot showed similarities to the quoted Figure 6.10(a). However, the amplitudes of the electron signals do not seem to be as high as those observed in Figure 6.10(a). The probable cause is the new optical trench feature in the VUV-MPPCs. For photon measurements, a low level of crosstalk is desired. After the second generation VUV2, the trench structures were added around the pixels to absorb most of the photons [18], which in our case also reduced the single-electron signal size. The studies in [14] were stated to use the older VUV2 so a better result was obtained. Future investigations may wish to obtain a VUV-MPPC of the older VUV2 generation for comparison.

Also possible to cause a decrease in signal size is the combination of the MPPC structure and the detected electron energy. There are two alternative structures to build a MPPC pixel, one being the P-on-N type as showed in Figure 6.9(a). This type has higher gain and photosensitivity for photons with shorter wavelengths, because the shallower absorption depth of such photons leave electrons to traverse further in the silicon column before reaching the cathode at the back of the

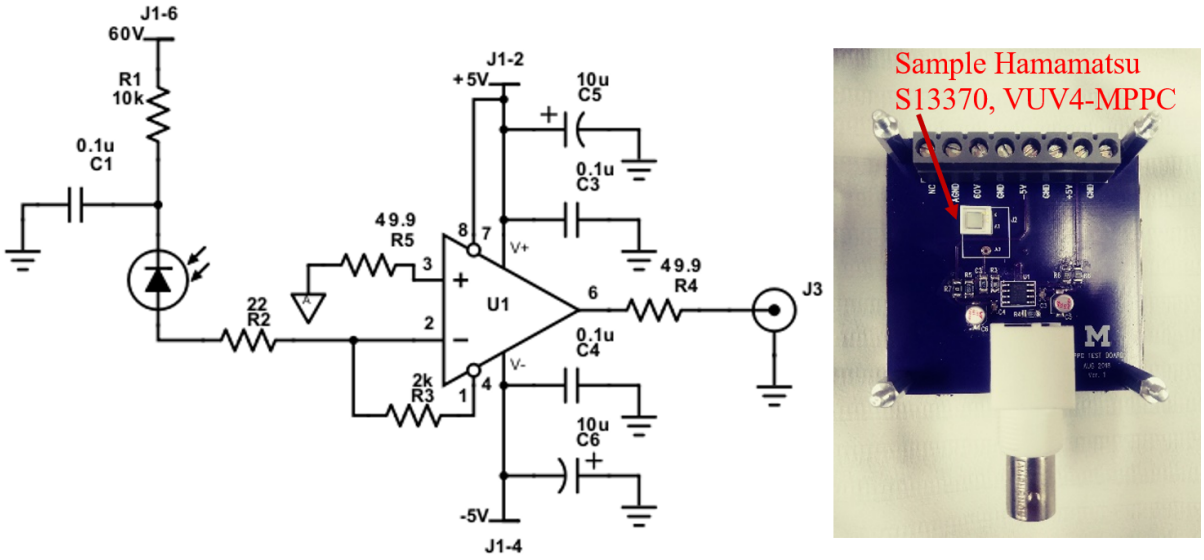


Figure 6.11: The test circuit that biases the VUV-MPPC and amplifies the output. It was based on an inverting amplifier design using an op-amp. The gain was $-2000/22 \simeq -91$. A picture of the board is also shown.

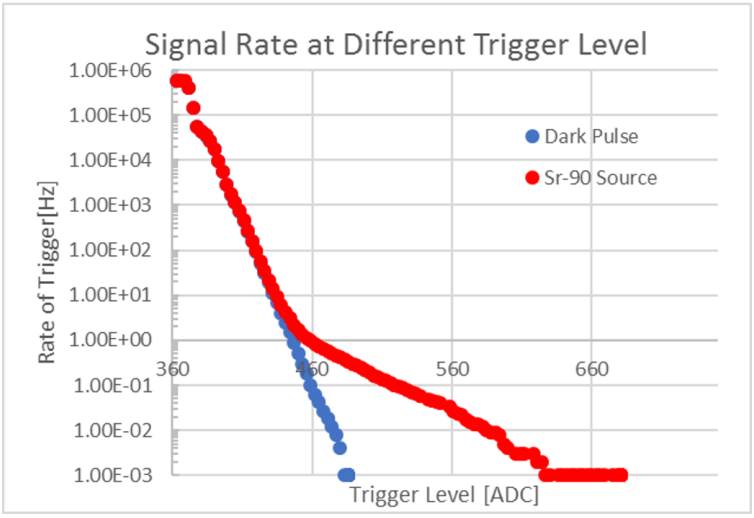


Figure 6.12: VUV-MPPC responses to electrons. Trigger rates above an array of thresholds were measured. Excess in the hit rates when comparing to the dark noise data indicated electron detection in the device.

device, producing more ionizations along the way. VUV-MPPCs use this structure to optimize for ultraviolet photon detection. The other is the N-on-P type which has a reversed orientation of the PN junction, which is optimized for long-wavelength photons. In this structure, electrons are collected at the surface. When it comes to electron detection, higher-energy electrons penetrate deeper into the material and ionize more at larger depth. Between the two structures, the P-on-N type used in VUV-MPPCs is thus not optimal for the higher energy electrons. It is hard to evaluate which factor contributed more to the phenomenon seen. Future measurements with electron energy

of both $\sim 1\ \mathrm{keV}$ and $\sim 10\ \mathrm{keV}$ should supply more information on this issue.

6.5 Development Plans

While the above measurement demonstrated qualitatively the single-electron detection capability of the VUV-MPPCs, there are questions remaining. In addition to the amplitude discussion in the last section, it still stays unknown what level of detection efficiency the devices have for $\sim 10\ \mathrm{keV}$ electrons. Subsequent plans were made for further advancements of the electron detector test and beyond.

The focus of the next-stage development of the electron detector will be on making VUV-MPPC detection efficiency measurements for electrons ideally between $1\ \mathrm{keV}$ and $20\ \mathrm{keV}$. Investigations have identified a method to obtain the electrons of desired energies, through refurbishing a BSM from the Fermilab Accelerator Department. In the schematic shown in Figure 6.2, the wire target can be biased to a desired negative high voltage while running a small current through it. This turns it into a thermionic electron source, and the negative high voltage can accelerate the emitted electrons to the appropriate energy. A beam of electrons can be obtained after passing through the front collimator, and the deflectors can be used to focus and deflect the produced beam by applying DC voltages. A straight line between the thermionic source and the detector should be strictly avoided to reduce measurement backgrounds from the photons, and to prevent damage. Measured rates should be compared to a baseline measurement with the same current flowing through wire but not turning on the high voltage to cancel out the remaining photon backgrounds. The tested VUV-MPPC will replace the electron collector and take similar runs as in the proof-of-principle test. A PCB board has been designed and manufactured to mount the VUV-MPPC in place. The circuit has been redesigned for better signal quality and amplification. More details on this will be discussed in Appendix C. The same method of triggering and counting through a firmware as used before can make the trigger rate measurements. The VUV-MPPC efficiencies can be measured through relative measurements. With a fixed current through the thermionic source, the produced electron flux should be constant. At low electron energies, the electron rate can be measured by some type of secondary electron multipliers with known efficiency as a calibration, then the detection efficiencies of the VUV-MPPC can be calculated accordingly.

Following the efficiency measurement results, the understandings of the phenomenon described in the last section hopeful can be verified. Some optimizations and follow-up tests can be carried out accordingly using VUV-MPPCs of different designs. With better understandings, it remains an option to customize a device combining all desired specifications. Further investigations and designs on the laser system and the magnetic system are also needed before bringing the BEM alive.

References

[1] M. Ball *et al.* (PIP-II Collaboration), *The PIP-II Conceptual Design Report*, Tech. Rep. (Fermilab, 2017) [FERMILAB-DESIGN-2017-01;FERMILAB-TM-2649-AD-APC].

[2] K. Byrum *et al.* (Mu2e-II), in *2022 Snowmass Summer Study* (2022) arXiv:2203.07569 [hep-ex] .

[3] A. Feschenko, in *PACS2001. Proceedings of the 2001 Particle Accelerator Conference (Cat. No.01CH37268)* (2001) pp. 517–521 vol.1.

[4] J. Hurd, *et al.*, in *Proceedings of International Conference on Particle Accelerators* (1993) pp. 2426 – 2428 vol.3.

[5] R. Connolly, *et al.*, J. Instrum. **7** (02), P02001.

[6] R. Connolly, *et al.*, Nucl. Instrum. Methods Phys. Res. A: Accel. Spectrom. Detect. Assoc. Equip. **312**, 415 (1992).

[7] J.-P. Carneiro, "H^- stripping equations and application to the high intensity neutrino source," (2007), [BEAMS-DOC-2740], `https://beamdocs.fnal.gov/AD/DocDB/0027/002740/001/fnal_2740.pdf`.

[8] D. E. Johnson, *et al.*, "The linac laser notcher for the fermilab booster," (2017), arXiv1704.04270[physics.acc-ph], [FERMILAB-CONF-16-388-AD].

[9] W. R. Leo, *Techniques for Nuclear and Particle Physics Experiments* (Springer-Verlag, 1994).

[10] T. Gys, Nucl. Instrum. Methods Phys. Res. A: Accel. Spectrom. Detect. Assoc. Equip. **787**, 254 (2015), in New Developments in Photodetection NDIP14.

[11] H. Chan, *et al.*, J. Instrum. **16** (07), P07024.

[12] A. Ghassemi *et al.*, *MPPC*, Tech. Rep. (Hamamatsu, 2021).

[13] M. J. Berger *et al.*, "Estar, pstar, and astar: Computer programs for calculating stopping-power and range tables for electrons, protons, and helium ions," (2017, version 2.0.1), [Online] Available: `http://physics.nist.gov/Star` [2022-6-2]. National Institute of Standards and Technology, Gaithersburg, MD.

[14] K. Ogasawara, *et al.*, IEEE Trans. Nucl. Sci. **64**, 2733 (2017).

[15] Hamamatsu, "Vuv-mppc 4th generation (vuv4)," (2017), datasheet, https://hamamatsu.su/files/uploads/pdf/3_mppc/s13370_vuv4-mppc_b_(1).pdf.

[16] M.-M. Bé *et al.*, *Table of Radionuclides. Vol. 3 A=3 to 244* (BIPM, 2006) pp. 43–46.

[17] M. Bogdan, J.-F. Genat, and Y. Wah, in *2010 17th IEEE-NPSS Real Time Conference* (2010) pp. 1–2.

[18] S. Ogawa, *Liquid xenon detector with highly granular scintillation readout to search for $\mu^+ \to e^+\gamma$ with sensitivity of 5×10^{-14} in MEG II experiment*, Ph.D. thesis, The University of Tokyo (2020).

Chapter 7 Conclusion

Further advancement of the Standard Model (SM) is not a question of *if*, but a question of *how*. Through measuring the CLFV neutrino-less muon-to-electron conversion in the presence of aluminum nuclei ($\mu^- N \rightarrow e^- N$), the Mu2e experiment hopes to use this clean rare muon decay channel to test the underlying symmetry of the charged lepton number conservation laws and probe parameter spaces for a wide range of BSM theories.

The experiment searches for the single monoenergetic $105\ \mathrm{MeV}$ electron signature from the conversion using high-performance detectors. With an electron momentum resolution of better than $180\ \mathrm{keV/c}$, the total measurement background will be suppressed to a sub-event level, allowing the experiment to reach a single-event-sensitivity of 3×10^{-17}, and to set an upper limit for the muon conversion-to-capture ratio of $R_{\mu e} < 8 \times 10^{-17}$ at 90% confidence level. This is a 4-orders-of-magnitude improvement to the current limit.

The Mu2e straw tube Tracker performance is crucial for reaching the desired momentum resolution. Using the arrays of straw tubes to measure incident electron coordinates, the particle momenta are reconstructed from their tracks inside the known magnetic field. Accurate reconstruction requires high spatial resolutions in the electron coordinates, which are $\sim 40\ \mathrm{mm}$ in the longitudinal direction and $\sim 200\ \mu\mathrm{m}$ in the transverse direction. The Tracker mechanical and electronic designs all revolves around this goal.

The Mu2e Tracker uses firmware-based systems for DAQ and operation to handle the "streaming" DAQ approach. Efficient firmware and a light-weighted alternative control system were developed. They have proven their effectiveness over the many tests, and are expected to expedite the quality checking and commissioning process for large panel volumes.

A VST has been conducted using a full Tracker plane to verify its performances and exercise the readout and the operations. In source scan studies using a radioactive $^{55}\mathrm{Fe}$ source during the VST, a map of effective signal propagation speeds across the panel was produced to assist future particle reconstructions and simulations. The straw longitudinal resolutions across different locations on a panel were also measured, giving a better view of the variation in the longitudinal resolutions. The overall resolution measured was around $30\ \mathrm{mm}$, concurring with the results from the cosmic-ray data studies, and confirming the spatial resolutions satisfy the experiment requirements. Studies also revealed the cause of the minor longitudinal resolution asymmetries observed in the scans, and

decided it would not be an issue with mitigation suggested.

The Mu2e experiment will soon move towards commissioning and operation phases. The first run is proposed to happen between 2025 and 2026, reaching a 10^3 improvement to the current limit. The full dataset is expected by the end of the decade.

As a natural continuation to Mu2e, the contemplated Mu2e-II experiment will have an extra order of magnitude increase in the sensitivity. It will be used to investigate the nature of CLFV were signals observed in Mu2e, or further exclude theory parameter spaces otherwise. It will use the Fermilab PIP-II Linac, requiring a beam extinction measurement. In our Research and Design project, a detection concept was established to use a laser system for H^- ion beam extinction measurements, and a candidate for the electron detector was selected to use a VUV-MPPC. Proof-of-principle measurements have established its capability to perform single-electron detections. Future measurements were planned to specify the detection efficiencies.

The Mu2e experiment and other Charged Lepton Flavor Violation searches alike will provide us with more information to advance the Standard Model and our knowledge of the universe to the most fundamental level. The party is *not* over yet.

Appendix A Lepton Number Violation Search at Mu2e

Apart from the primary goal of searching for Charged Lepton Flavor Violation (CLFV) through neutrinoless muon-to-electron conversion, the Mu2e apparatus is also capable of detecting the Lepton Number Violation (LNV) process of

$$\mu^- + A(Z, N) \to e^+ + A(Z-2, N+2), \tag{A.1}$$

which challenges the (total) lepton number conservation.

In the SM, conserved individual lepton family numbers L_e, L_μ, and L_τ directly lead to the conservation of the (total) lepton number $L = L_e + L_\mu + L_\tau$. The conservation of the lepton number corresponds to another accidental symmetry of the SM Lagrangian. However, when non-renormalizable operators are added in extensions of the SM, the conservation can be violated. In general, effective operators of dimension five or higher add new degrees of freedom and lead to LNV [1].

In the specific case of LNV with $|\Delta L| = 2$, such operators can come from, for example, neutrino Majorana mass terms [2] or R-parity-violating Supersymmetry models [3]. Additional model-unspecific discussions using effective operator approaches can be found in [1, 4, 5]. These theories allow LNV processes such as neutrinoless double beta ($0\nu\beta\beta$) decay, $\mu^- \to e^+$ conversion, and $K^+ \to \pi^- l^+ l'^+$ decays (here $l^+ l'^+ = e^+e^+$, μ^+e^+, or $\mu^+\mu^+$) [6]. While the $0\nu\beta\beta$ decay search remains the most powerful probe for LNV due to its high sensitivity [7], the $\mu^- \to e^+$ conversion search provides an opportunity to simultaneously study the Majorana phases (LNV) and the lepton flavor structure (CLFV). Besides, $\mu^- \to e^+$ can be searched concurrently with $\mu^- \to e^-$, as long as the detectors are configured to measure both electrons and positrons. This is the case in the Mu2e experiment. Positrons, if produced in the muon stopping target, also travel in helical trajectory within the DS magnetic field. They can be measured in the Mu2e Tracker and Calorimeter, and the chirality of the helical tracks will separate them from the electrons.

A list of historical $\mu^- N \to e^+ N'$ search results can be found in [8]. The current best limit was obtained by the SINDRUM II experiment. In titanium, the conversion-to-capture ratio

$R_{\mu^- e^+}(\mathrm{Ti})$ was measured to be (at 90% CL) [9]:

$$R_{\mu^- e^+}(\mathrm{Ti}) = \frac{\Gamma[\mu^- + \mathrm{Ti} \to e^+ + \mathrm{Ca}]}{\Gamma[\mu^- + \mathrm{Ti} \to \nu_\mu + \mathrm{Sc}]} < \begin{cases} 1.7 \times 10^{-12}\ (\mathrm{GS}) \\ 3.6 \times 10^{-11}\ (\mathrm{GDR}) \end{cases} . \tag{A.2}$$

The values are different depending on final nuclear states. GS indicate the final calcium nucleus is in the ground state, and GDR stands for an excited giant dipole resonance state with both a mean energy and a width of $20\ \mathrm{MeV}$ [9].

The signal in the $\mu^- \to e^+$ search is a single conversion positron, whose energy can be calculated as [8]:

$$E_{e^+} = m_\mu c^2 - (\mathrm{B.E.})_{1s} - E_{\mathrm{recoil}} - \Delta_{Z-2}\ . \tag{A.3}$$

Comparing with Equation 2.1, an additional Δ_{Z-2} term represents the change in nuclear binding energy due to different nuclear states. Due to the giant dipole resonance final state, the positron is not monoenergetic; the downward shift and the width of the positron energy distribution match the mean energy and the width of the giant dipole resonance respectively. Consequently, both radiative muon capture (RMC) and radiative pion capture (RPC) backgrounds are significant in the conversion positron search. The photons from these processes can produce positrons with energies inside the signal region through pair production. Figure A.1 [9] shows the simulated positron momentum spectra overlapping with the intrinsic RMC backgrounds in the SINDRUM II experiment. With the aluminum stopping target, the Mu2e experiment expects similar challenges to

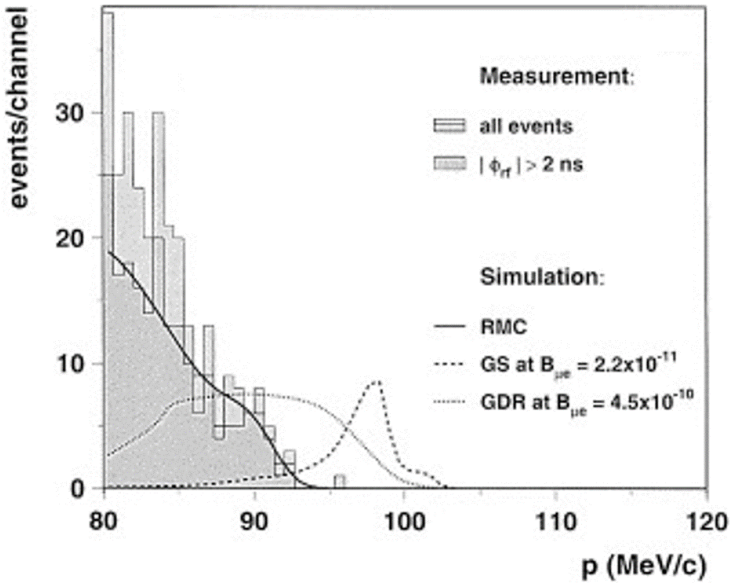


Figure A.1: Simulated conversion positron signals and RMC backgrounds in SINDRUM II. The figure is adapted from Fig. 2 in [9]. Positron momentum spectra of the two final states (at the quoted branching ratios) are overlaid with a simulated RMC spectrum.

identify possible signals from the backgrounds. Nonetheless, an improvement to the current upper limit is anticipated. Further improvement in the sensitivity by a few orders of magnitude can be achieved if a stopping target material specifically chosen for $\mu^- \to e^+$ search is used [10].

References

[1] A. de Gouvêa, *et al.*, Phys. Rev. D **100**, 075033 (2019).

[2] P. W. Angel, N. L. Rodd, and R. R. Volkas, Phys. Rev. D **87**, 073007 (2013).

[3] H. Haber and G. Kane, Phys. Rep. **117**, 75 (1985).

[4] T. Geib, A. Merle, and K. Zuber, Phys. Lett. B **764**, 157 (2017).

[5] J. M. Berryman, *et al.*, Phys. Rev. D **95**, 115010 (2017).

[6] L. S. Littenberg and R. Shrock, Phys. Lett. B **491**, 285 (2000).

[7] W. Rodejohann, Int. J. Mod. Phys E **20**, 1833 (2011).

[8] R. Bernstein and P. S. Cooper, Phys. Rep. **532**, 27 (2013).

[9] J. Kaulard, *et al.* (SINDRUM II Collaboration), Phys. Lett. B **422**, 334 (1998).

[10] B. Yeo, *et al.*, Phys. Rev. D **96**, 075027 (2017).

Appendix B Functions of Alternate Tracker Control and Operation System

The alternate Tracker control and operation system discussed in Chapter 4 performs a wide range of functions. While many of them are realized by the micro-controller command set by writing to or reading from registers, interfaces, or chips, there are some others requiring more sophisticated treatments involving dedicated firmware modules or control console scripts. Apart from the scheme used to read channel threshold voltages and to find appropriate threshold settings as discussed in Chapter 4, here are a few more selected examples.

B.1 Finding Bit Alignment

As Section 4.2.1.1 already touched upon, the bit alignment and the bit slippage ensures the ADC data are properly captured in the firmware. The main processes occur in the dedicated firmware modules. Pseudo Code Snippet B.1 describes the entire flow of processes. Appropriate ADC patterns are used in different stages for better results. The level at which a step is handled is marked in the comment (**F** = Firmware, **M** = Micro-Controller, and **C** = Console).

B.2 Taking Data from Multiple Panels through Serial Links

The procedure to take data from a single Tracker panel was briefly discussed in Section 4.2.1.2. The data are temporarily stored in a FIFO buffer and are read out by a register draining the buffer contents in order. In more advanced tests where multiple panels are read out simultaneously, the basic console cannot command and handle data from multiple ports at the same time, so a special script was developed.

The script utilizes the multi-threading feature of Python to handle the simultaneity. As the temporary buffers in the ROCs have very limited sizes (only enough for 500 hits), the script needs to take hundreds or thousands of runs to get a reasonable dataset. A configuration file is used to supply run parameters for the panels. For each panel ID, the channel masks, the sample length per hit, the latency between the beginning of a trace to the signal trigger time, the number of hits per run, together with other information, are stored in the file. It is used by the main script to configure panels for runs. At the beginning of each run, the script also logs slow control variables of each panel to gather information on their long-term performances. Data are read out using the same

Pseudo Code Snippet B.1 Procedure to Find Channel Bit Alignment and Bit Slippage

```
// Stage 1: Bit alignment                                   ▷ M, unless otherwise noted
Configure the ADC chip for checkerboard pattern 0x2AA (10 bits)
Configure eye monitors and initiate firmware module
Perform bit alignment                                       ▷ F
Return firmware module status

// Stage 2: Bit slippage                                    ▷ F, unless otherwise noted
Configure the ADC chip for one-bit-high pattern 0x200 (10 bits)   ▷ M
for 0 ≤ slippage < 10 do
    Set slippage to the ADC interface
    Compare output to the preset pattern
    if output agrees with preset pattern then
        Store slippage in reserved register
    end if
end for
Set slippage from reserved register
Return slippage numbers and errors                          ▷ M

// Stage 3: Verification                                    ▷ M, unless otherwise noted
Configure the ADC chip for mixed-frequency pattern 0x263 (10 bits)
Compare output to the preset pattern                        ▷ F
Return errors
```

Pseudo Code Snippet B.2 Procedure to Simultaneously Read from Multiple Panels through UART

```
Load run parameters from configuration file
for N runs do          ▷ Typically hundreds or thosands of runs for a reasonable dataset
    Initiate multi-threading, one for each panel
    in thread
        Create output file
        Read settings, thresholds, slow control variables   ▷ M
        Log the returned values in the output file
    end thread
    Initiate multi-threading, one for each panel   ▷ Restarted for data taking simultaneity
    in thread
        Configure panels using run parameters               ▷ M
        Read data through serial link                       ▷ M
        Store data in the output file
    end thread
end for
```

method as in the single panel, and each run of a panel is stored in a separate data file. This entire flow is summarized in Pseudo Code Snippet B.2. The processes are executed on the console level

unless otherwise noted.

This method is not very efficient, as there are long periods of dead time between runs for the data to be transmitted through the serial links, which is a very slow process. It is soon replaced by the next method where data are read out to the DAQ servers through optical fibers, while the panels are still controlled through the serial ports. Nevertheless, many of its features were inherited by the new data taking scheme.

B.3 Taking Data from Multiple Panels through DAQ Server

This method allows data taking with multiple panels through the DAQ server. Data transfer through the optical fiber is much faster than through the serial link, which makes taking large datasets ($\sim$ 1 million hits per panel) much more convenient. In this method, the ROC system clock and the EWMs are supplied by the DAQ server, instead of being emulated in the ROC. Since the Data Packet Assemblers only activate when there are EWMs present, multiple panels can be synchronized for data taking by toggling the common EWMs sent towards them from the DAQ server. In the current iteration of the firmware, the panel controls are still managed by the console, so it needs to be coordinated with the DAQ server. The script for data taking through the UART link is recycled to manage panel configurations and log slow control data. When the EWMs are turned off, it is issued for one single data run with a very short timeout. Due to the lack of EWMs, none of the panels would return any hits, and all time out quickly. But the panel configurations are written into the respective registers, and the slow control data are taken. Then turning on the EWMs to the panels kicks off the data taking, and all data in this case are stored in the DDR3 memories. Pseudo Code Snippet B.3 describes the procedure. The letter **D**'s in the comments stand for steps operated by the DTC in the DAQ machine.

Pseudo Code Snippet B.3 Procedure to Simultaneously Read from Multiple Panels through Optical Fiber

1: **Turn off** EWMs ▷ **D**
2: **Load** run parameters, **read** settings and slow control variables, **configure** panels
3: ▷ **C**, **M**, recycling the previous script
4: **Configure** DDR3 memories (specify amount of data to take) ▷ **D**, **F**
5: **Turn on** EWMs ▷ **D**
6: **Store** data to DDR3 memories ▷ **F**
7: **Read** data through optical fiber link, **store** data in output files ▷ **D**

B.4 Measuring Channel Hit Rates

Hit rate measurements are frequently used for diagnostics. The implementation uses counters included in the Data Package Assemblers in the DIGI firmware, which tally the numbers of hits on either side of the straw, as well as the coincident hits. A register relays the control signal that zeros and starts the counters. The same signal also stops the counters depending on the status of the state machines. On the other hand, the elapsed time is read from a global time count kept in the firmware. Multiple samples can be taken if average hit rates are of interest. To avoid floating point divisions, all hit counts and the elapsed time are passed on by the micro-controller to the control console for final rate calculations. As a surprising fact, there is not a single variable of floating point or double data type anywhere in the micro-controller command set. Consequently, the related C/C++ libraries are not invoked during compilation, which largely reduces the size of the compiled hexadecimal source file. Indeed the code optimization was this extreme in order to squeeze all necessary functions into the limited sNVM space. The steps for measuring the straw hit rates are again shown in Pseudo Code Snippet B.4.

Pseudo Code Snippet B.4 Procedure to Measure Channel Hit Rates

1: **for** N samples **do**
2: **Send** control signal and **read** start time ▷ **M**
3: **Zero** and **start** hit counters in Data Package Assemblers ▷ **F**
4: **Wait** some time specified by command parameter
5: **Send** control signal and **read** end time ▷ **M**
6: **Stop** hit counters ▷ **F**
7: **Read** counters Δn, **calculate** elapsed time Δt ▷ **M**
8: n_{total}(total hit counts) $\leftarrow n_{total} + \Delta n$, t_{total}(total elapsed time) $\leftarrow t_{total} + \Delta t$
9: ▷ **M**, counts are for each channel, single sided and coincident
10: **end for**
11: **Return** n_{total}, t_{total} ▷ **M**
12: **Calculate** single sided and coincident hit rates for each channel ▷ **C**

B.5 Checking Channel Noises and Electrical Continuity

The last example to mention here is the diagnostics script that utilizes the channel hit rate measurements. It checks for the channel noises. Additionally, it also checks each channel for the electrical continuity from the preamp on the CAL side, to the anode wire, and all the way to the preamp on the HV side. The channel noise checks are very straightforward. Without the presence of any additional radioactive sources and with the high voltage turned off (so the straws do not get cosmic-ray muon hits), the remaining hit rates represent the rates of electronic noises. If the high voltage is connected, as the expected cosmic-ray rates on single straws are low (< 10 Hz), channels with much higher hit rates may suffer from high voltage issues like micro-sparks. When the calibration pulses are turned on, which can be sent at some arbitrarily given rate, the hit rate increases from the electronic noise rates should match the rate of the calibration pulses for both single-ended rates as well as the coincident rate. Much lower coincident and HV side rates, while the CAL side rate agrees with the expectation, may indicate poor electrical connection between the preamps and the anode pins, which is an installation issue rather commonly seen. Another almost-never-seen cause to this issue is a broken or missing anode wire. With the help of such diagnostics, the related problems can be quickly identified and mitigated if possible. Channels with unfixable problems are documented for records. The diagnostic routine is outlined in Pseudo Code Snippet B.5; by default steps are executed by the software control console.

Pseudo Code Snippet B.5 Procedure to Check Channel Noises and Electrical Continuity

1: **Measure** channel hit rates ▷ Without radioactive sources or high voltage
2: **Set** calibration pulse amplitude and frequency ▷ **M**
3: **Turn on** calibration pulse ▷ **M**
4: **Measure** channel hit rates
5: **Examine** hit rates and **compare** hit rate increases with calibration pulses
6: **Report** channels with issues for fix

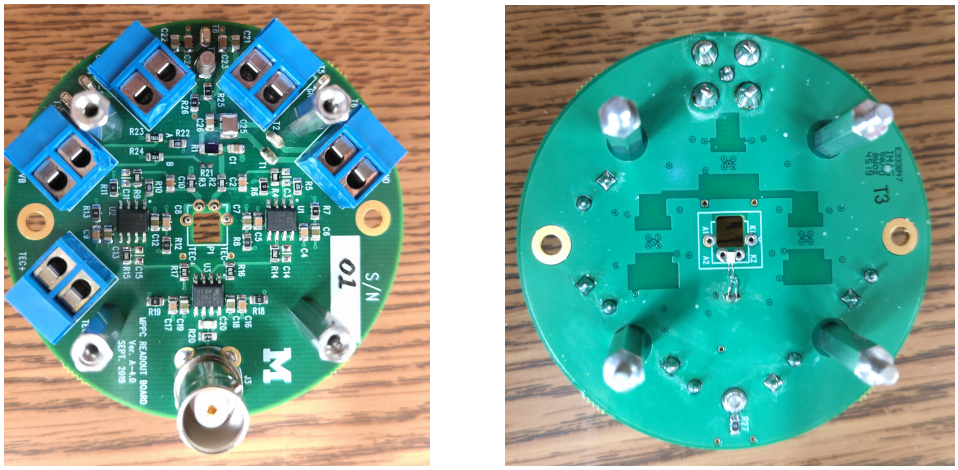
Appendix C VUV-MPPC Measurement PCB Design

To perform the electron detection efficiency measurements described in Chapter 6, a spare Bunch Shape Monitor (BSM) will be reconfigured to an electron beam. A 6-layer PCB board (Figure C.1) was designed to hold the tested VUV-MPPC in place. It will provide the reverse bias voltage and amplify the output signals. The board included some features for better efficiency measurements.

The MPPC biasing and signal amplification circuit uses a differential transimpedance amplifier design as shown in Figure C.1 [1]. This design suppresses pick-up noises from the environment. The gain is larger than the one used in the previous test; two stages of 3 op-amps provide a total gain of 915. This hopefully can help resolve single photoelectron peaks in the readout.

The PCB geometry is restricted by the cavity size of the BSM. The VUV-MPPC will be installed on the back side over the hole in the center of the board. The electron beam will come from this direction. As one may have noticed, there are no other electronics components on this side of the board, which prevents influences on the components and the signals from the beam. The design went even further by including an additional ground layer as the top layer of the back side (apart from a few locations below op-amp traces to lower parasitic capacitance as required by the op-amp datasheet), which prevents incident electrons from accumulating. This made the front side of the board rather crowded and added constraints during the routing process.

To test the idea of thermoelectric cooling of the VUV-MPPC in vacuum, which is supposed to reduce the dark noise rate of the device, the board reserved a space for Peltier cooler installation. The tested Peltier cooler fits into the central hole of the board and cools the VUV-MPPC from the back side. A cooper tape or a copper braid will be needed to connect the hot side of the Peltier cooler to some surrounding metal part as a heat reservoir, as there will be no convection cooling of the device in the vacuum inside the BSM. This will keep the hot side at relatively low temperatures and improve the cooling efficiency. A thermal diode on the board can sense the temperature on the back side of the VUV-MPPC, which is read out using a built-in Wheatstone bridge.



(a) (b)

Figure C.1: PCB for VUV-MPPC efficiency measurements. Both the front (a) and the back (b) side of the PCB are shown.

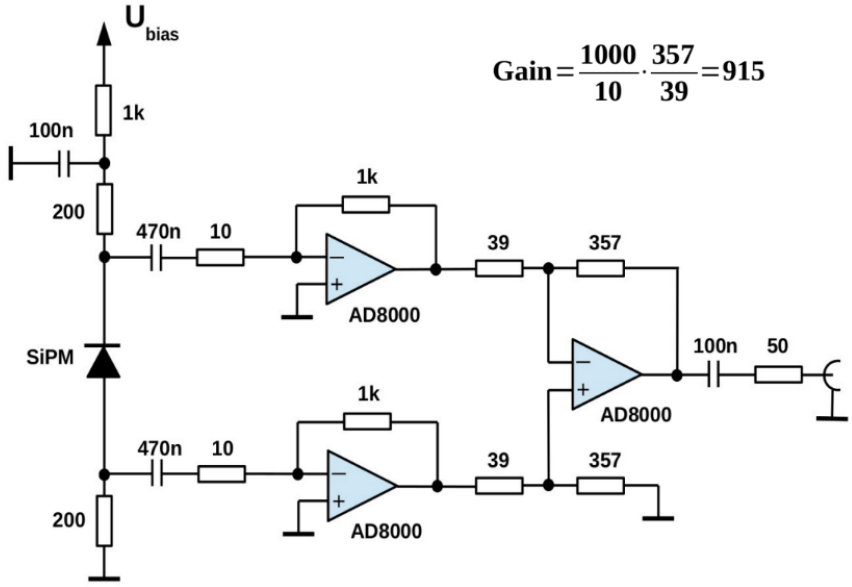


Figure C.2: A diagram of the upgraded MPPC biasing and signal amplification circuit design [1].

References

[1] S. Gundacker, "Practical guide to using sipms," (2015), online presentation.