UNIVERSITY OF CALIFORNIA,
IRVINE

Search for $\mu^+ \to e^+\gamma$ in the MEG II Experiment's First Physics Dataset and Analysis of the MEG II Drift Chamber Data

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Physics

by

Dylan Palo

Dissertation Committee:
Professor William Molzon, Chair
Professor David Kirkby
Professor Yuri Sherman
Professor Juan Pedro Ochoa-Ricoux

2023

DEDICATION

To my family.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I really enjoyed the Ph.D. program in experimental particle physics, and even though it was often a very stressful process, I have no regrets and am thankful that I am able to continue forward in my career in academia. I learned invaluable skills and gained experience that I am very thankful for. These skills and experience would not be possible without help and time from a collection of people.

First, the Ph.D. wouldn't have been the same without my advisor, Bill Molzon. He's been an exceptional advisor, always going above and beyond what is required to help me develop in individual projects and my career in general. He is someone that I can always depend on for advice, expertise, or a new proverb. To this day, we still can and do discuss analysis, physics, or just the day-to-day on the phone for hours.

At the beginning of the program, I'd show him a draft presentation, piece of code, etc. and I'd know I'd get a slew of comments back: "too much white space on the slide","not enough in-line comments" "New York Times says to tell the reader what you're going to say, say it, and then tell them what you told them", etc. Later, I'd find myself thinking "what will Bill say about this line?". Through this process, I think that many of his philosophies on academic writing, programming, analysis, and physics in general have been instilled in me; I know they will stay with me and I'm very thankful for that.

I thank my professors at UCI for giving me a general knowledge of quantum, QFT, particle physics, and machine learning.

I thank Yusuke Uchiyama for providing me with software guidance through my Ph.D. Especially early on, taking the time to help me understand Geant, Genfit, sql, our simulation code, and the MEG software in general. I found his guidance invaluable. I also thank Terence Libeiro for helping me through understanding software installations, compilations, help with ROOT and the MEG software in general.

I thank Francesco Renga, for his help and support in the PAWG+target+physics analysis. I think he did a great job leading the PAWG, welcoming new collaborators to join, offering help and suggestions, keeping the work collaborative across the many institutions, etc.

I thank the CDCH hardware group for developing and operating the drift chamber through the 2021+2022 physics runs. This thesis would not be possible without the working detector. I also want to thank Marco Chiapinni, Marco Francesconi, Luca Galli, and Antoine Venturini for always making me feel welcome when visiting PSI.

I thank the run coordinators Malte Hildebrandt, Angela Papa, and Toshiyuki Iwamoto for always providing me with help when at the lab, especially with the operation and installation of the camera. I also thank the MEG II spokespeople, Toshinori Mori and Alessandro Baldini for my time on the MEG II experiment.

Chapter 4 is an adaptation of the material as it appears in "Precise photographic monitoring

of MEG II thin-film muon stopping target position and shape", used with permission from Elsevier. The co-authors listed in this publication are Malte Hildebrandt, Andreas Hofer, William Kyle, Deep Lad, Terence Libeiro, and William Molzon. William Molzon directed and supervised research which forms the basis for the thesis/dissertation.

The thesis contains a description of the design and construction (Chapter 3), alignment (Chapter 6), and calibration (Chapter 7) of the MEG II detectors and contains the details of reconstruction and physics analysis on MEG II datasets (Chapter 8). This is work done by the full MEG II collaboration. Much of this work is described in previous and upcoming publications: [7],[12][13],[14][15]. Details and figures have been pulled from a variety of talks and publications from the collaboration.

VITA

Dylan Palo

Education

- Jul. '17 – Present: **Physics Ph.D.**, *University of California, Irvine*, Irvine, CA. GPA: 3.84/4.00
- Aug. '14 – May '17: **Physics B.S., Mathematics B.A.**, *Miami University*, Oxford, OH. GPA: 3.73/4.00

Employment

- Jul. '17 – Present: **Ph.D. Student Researcher**, *University of California, Irvine*, Irvine, CA.

  My dissertation and research have been on the MEG II experiment under my advisor Bill Molzon. The first two years focused on developing a photographic technique used to monitor the position, orientation, and shape of the MEG II stopping target. This contained image analysis, minimization, operation of a Raspberry Pi, milling, 3D printing, Monte Carlo Geant4 simulation studies, etc. The remainder of the program has focused on software development and analysis of the MEG II drift chamber data. This consists of wire-by-wire alignment using positron tracks, neural network applications to improved track point measurements, digitized waveform analysis such as noise suppression and filtering, and track fitting optimization. This work included learning ROOT, C++, Python, MySQL,Keras, etc.

- Sep. '17 – Mar. '18: **Teaching Assistant**, *University of California, Irvine*, Irvine, CA.

  During this time I taught Physics 52A in Fall 2017 and Physics 138 in Winter 2017. Physics 52A is an optics lab course for freshman and sophomore physics students. This course was structured such that I was teaching 10 groups of two students. This required making sure each single group understood the lab work while being attentive to the class as a whole. Physics 138 is an upper-division astrophysics course; I taught its discussion section. I was responsible for creating quizzes and creating a discussion based lecture every week.

Jun. '16 Aug. '16 **Mechanical Technician**, *Fermilab National Accelerator Laboratory*, Batavia, IL.

- As a technician I was working almost exclusively with the D-0 and Muon G-2 experiments. The responsibilities as a technician ranged largely depending on the task at hand. My work at D-0 was mainly removing parts such as electronics cards that would later be reconfigured. Additional work at D-0 consisted of assisting staff scientists through jobs such as assembling GPS fibers needed to calibrate scintillator panels.The work was supplemented by the biweekly Undergraduate Lecture Series that Fermilab offers to all undergraduate workers. These lectures along with my coursework were responsible for spiking my curiosity in the field of particle physics.

Jun. '15 Aug '15 **Undergraduate Summer Scholars Researcher**, *Miami University*, Oxford, OH.

- Through the Undergraduate Summer Scholars program I was responsible for proposing a research project in the spring, then carrying the experiment out in the summer. The experiment focused on applying spectral phasor analysis to monitor the cellular metabolic responses of eukaryotes. Soft UV-excitation was used to fluoresce cellular NADH molecules found throughout the metabolic pathways. I gained experience culturing cells, managing optical arrangements, and altering code in Andor: a program responsible for recording the emission spectra. In addition, I was responsible for documenting my findings through journal-style writings.

Publications, Proceedings, and Technical Notes

- **Palo, D., Molzon, W. et al. (2023). Neural Network Applications to Improve Drift Chamber Track Position Measurements. *Plan to submit to Nuclear Instruments and Methods in Physics Research A*, under final review by collaboration**
  Hyperlink
  Here, we describe neural network applications to drift chamber data to improve the distance of closest approach measurements. The network provides a data-driven time-distance relationship estimate, conventionally estimated by simulation. The network uses more information than the conventional track position estimate by inputting the wire's digitized voltages and thus information from the ensemble of ionization clusters. The neural network application applied to the MEG II drift chamber data improves the single hit resolution and the e^+ kinematic resolutions. This technique is implemented

in the standard MEG II analysis for 2021. I was responsible for developing this technique, evaluating it's performance, and implementing it in the MEG II framework.

- **Palo, D. on Behalf of the MEG II collaboration. (2023). Status of the MEG II Experiment and Performance Results From the First Year's Data Taking. *Instruments*, Submitted.**
  Hyperlink
  Here, we describe the status of the MEG II experiment, a search for the charged lepton flavor violating decay: $\mu^+ \to e^+\gamma$. Preliminary data-driven positron and photon kinematic resolutions are shown using the 2021 dataset. Using data-driven resolutions, a preliminary sensitivity for both the 2021 dataset and the estimated full MEG II dataset are quoted.

- **Libeiro, T. et al. (2023). Novel X-ray Scanning Technique For Remote Alignment of Photo-detectors in MEG2 Calorimeter. *Nuclear Instruments and Methods in Physics Research A*, 1048, 167901**
  Hyperlink
  Here, we describe and show results of a technique to calculate the positions of photodetectors installed in the MEG II liquid xenon calorimeter. The measurement scans the front surface of the calorimeter in axial (Z) and azimuthal (ϕ) directions using a thin, collimated beam of X-rays obtained from a 57Co X-ray source to detect signals in individual photo-detectors. Position resolutions below 1 mrad in ϕ and 1 mm in Z have been achieved. I was responsible for creating the lead strips used to cross-check the calibration method; this included creating lead strip holders using a MakerBot 3D printer and designing them in autoCAD.

- **Chiappini M., et al. (2023). The Cylindrical Drift Chamber of the MEG II Experiment. *Nuclear Instruments and Methods in Physics Research A*, 1047, 167740**
  Hyperlink
  This proceedings presents a description of the MEG II cylindrical stereo drift chamber. The commissioning of the drift chamber is described along with the status of the single hit resolutions and momentum resolution estimates.

- **Palo, D., Molzon, W. (2022). Iterative Track-based CDCH Wire Alignment. *Internal MEG II Technical Note.***
  Hyperlink
  This note describes the iterative track-based wire alignment technique using Michel positron tracks. We discuss the technique's procedure, shortcomings, and remaining residuals. The final wire alignment is compared to the survey alignment by comparing the wire positions, mean residuals, and the double turn analysis. The procedure results in a wire alignment wire-by-wire of $\sim 2\mu m$ at the wire center and $\sim 15\mu m$ at the

wire edges. This is the default wire alignment in the MEG II analysis for 2021. I was responsible for developing this technique and implementing the result in the MEG II framework.

- **Palo, D., Molzon, W. (2021). CDCH Coherent Noise. *Internal MEG II Technical Note***
  Hyperlink
  During the 2020 engineering run, coherent, low frequency noise was observed in the MEG II drift chamber data. Most of the noise is coherent over 16 wires (32 channels) connected to the same pair of electronic DRS boards (upstream/downstream pair). Here, we discuss measurements of the coherent noise and techniques for subtracting it event-by-event in the data. This technique is implemented in the standard MEG II analysis for 2021. I was responsible for developing this technique and implementing it in the MEG II framework.

- **Baldini, A.M., et. al. (2021). The Search for $\mu^+ \to e^+\gamma$ with 10^{-14} Sensitivity: The Upgrade of the MEG Experiment. *Symmetry*, 13, 1591**
  Hyperlink
  The MEG II experiment is a search for the $\mu^+ \to e^+\gamma$ decay; the experiment aims to improve on the decay's sensitivity (set by its predecessor, MEG I) by an order of magnitude. Here, we present the current status of the preparation and commissioning of the MEG II detector in the 2019-2020 engineering runs. Preliminary resolutions from the detectors are presented.

- **Baldini, A.M., Berg, F., Biasotti, M. et. al. (2020). Search for lepton flavour violating muon decay mediated by a new light particle in the MEG experiment. *Eur. Phys. J. C* 80, 858.**
  Hyperlink
  Here, the first direct search for lepton flavour violating muon decay mediated by a new light particle X, $\mu^+ \to e^+X$, $X \to \gamma\gamma$ is described and presented. This search uses the full MEG experimental dataset.

- **Palo D., Hildebrandt M., Hofer A., Kyle W., Lad D., Libeiro T., Molzon W. (2019). Precise photographic monitoring of MEG II thin-film muon stopping target position and shape. *Nuclear Instruments and Methods in Physics Research A*, 944, 162511**
  Hyperlink
  Here, we present the photographic technique developed to measure continuously the position, orientation, and shape of the MEG II stopping target. The photographic technique uses a camera $\sim$1.2 m from the target inside a $\sim$1.3 T magnetic field with significant radiation to image $\sim$120 dots printed on the target. The technique calculates the six rigid body parameters that define the plane and additional deformation

parameters. I was responsible for developing this technique and implementing this technique into the MEG II software.

- **Palo, D., Molzon, W. (2019). Corner Cube AIF Contribution. *Internal MEG II Technical Note***
  Hyperlink
  Here, we discuss the analysis used to determine the optimal corner cube location on the target frame. The corner cubes are used for the optical alignment of the target position in the MEG II coordinate system. The optimization is performed to minimize the background due to annihilation in flight (AIF) in the corner cubes.

- **Maltas J., Palo D., Wong C.K., Stefan S., O'Connor J., Al Aayedi N., Gaire M., Kinn D., and Urayama P. (2018) A metabolic interpretation for the response of cellular autofluorescence to chemical perturbations assessed using spectral phasor analysis. *RSC Advances*, 8, 41526-41535**
  Hyperlink
  Here, we confirm that the spectral phasor approach (described in the below publications) can be used to infer metabolic state changes. This is confirmed by several examples. One example is shown by addition of the same metabolic inhibitor to cellular solutions with different initial metabolic states. The addition of the same metabolic inhibitor created different shifts in phase space. This suggests that the phase space shift is dependent on the initial metabolic state of the cells.

- **Palo D., Maltas J., Risal L., and Urayama P. (2017). Sensing NADH conformation using phasor analysis on fluorescence spectra. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 186, 105-111**
  Hyperlink
  The publication describes the use of spectral phasor analysis to illustrate how conformational change can be quantified using fluorescence spectra. I was responsible for quantifying the ligand-binding constant associated with MDH (malate dehydrogenase) and NADH. Similar to the methods in the publication below, the setup required sequential additions of proteins (MDH) to NADH in solution. We later fit for the binding constant by calculating an approximate fraction of bound NADH molecules at each sequential addition of MDH.

- **Maltas J., Amer L., Long Z., Palo D., Oliva A., Folz J., and Urayama P. (2015). Autofluorescence from NADH conformations associated with different pathways monitored using nanosecond-gated spectroscopy and spectral phasor analysis. *Analytical Chemistry*, 87, 5117-5124**
  Hyperlink
  Here, we apply nanosecond-gated spectral acquisition, combined with spectral phasor analysis to Saccharomyces cerevisiae in order to monitor the ensemble of protein-bound

cellular NADH (Nicotinamide adenine dinucleotide) conformations associated with long excited-state lifetime fluorescence separately from the entire fluorescence signal. I was responsible for measurements that applied the spectral phasor analysis to NADH protein binding. I created the experimental procedure, performed the experiments, and produced the analysis necessary for the manuscript.

Presentations

- **D. Palo on Behalf of the MEG II Collaboration, Status of the MEG II Experiment and Performance Results From the First Year's Data Taking, NuFact 2022: The 23rd International Workshop on Neutrinos from Accelerators, August 2022**

- **D. Palo, W. Kyle, T. Libeiro, W. Molzon, Various Titles, Global Developments of Researches in Lepton Flavor Physics With Muons, 2018-2022**

- **D. Palo, W. Kyle, T. Libeiro, W. Molzon, Various Titles, MEG II Collaboration Meetings 2018-2022**

- **S. Stefan, D. Palo, M. Gaire, N. Al Aayedi, and P. Urayama Using Real-time Quantification Of Autoflorescence Spectrum Shape To Distinguish Between Metabolic Responses Involving NADH-Utilizing Pathways, Biophysical Society 61th Annual Meeting, 2017.**

- **P. Urayama, D. Palo, S. Stefan, M. Gaire, J. O'Connor. NADH Conformation Assessed Using Spectral Phasor Analysis: Illustrating Concepts From Molecular Folding To Metabolic Monitoring, Biophysical Society 60th Annual Meeting, 2016.**

- **D. Palo, S. Stefan, P. Urayama. Bioanalytical Discrimination of Protein-bound NADH Conformations using Spectral Phasor Analysis of NADH Fluorescence, Undergraduate Research Forum, 2015**

Honors & Awards

- 2017 **University of California, Irvine**, *Chancellor's Fellowship*, Irvine, CA.

- 2017 **Miami University**, *Outstanding Undergraduate Researcher Award*, Oxford, OH.
- 2017 **Inducted at Miami University**, *Sigma Pi Sigma Member*, Oxford, OH.
- 2016 **Miami University**, *Dean's Scholar Award*, Oxford, OH.
- 2016 **Miami University**, *Miami College Of Arts and Sciences Ambassador*, Oxford, OH.
- 2015 **Miami University**, *Undergraduate Summer Scholars Recipient*, Oxford, OH.

ABSTRACT OF THE DISSERTATION

Search for $\mu^+ \to e^+\gamma$ in the MEG II Experiment's First Physics Dataset and Analysis of the MEG II Drift Chamber Data

By

Dylan Palo

Doctor of Philosophy in Physics

University of California, Irvine, 2023

Professor William Molzon, Chair

The MEG II experiment[7] is an active search for $\mu^+ \to e^+\gamma$. The experiment plans to reach a sensitivity of $6 \cdot 10^{-14}$, improving on the limit of the MEG I experiment[5], by an order of magnitude. The MEG II detector analysis was written by the MEG II collaboration; this thesis details the detector analysis focusing on portions of the analysis I contributed to including the positron analysis and target analysis. These chapters also serve as a thorough description of the analysis for current and future collaborators. The thesis ends with a description of two physics analyses performed on the MEG II experiment's first year of physics data and demonstrates the experiment's expected performance over its lifetime.

Chapter 1 is an introduction to the Standard Model of particle physics. Charged lepton flavor violation (CLFV) processes, including $\mu^+ \to e^+\gamma$ are realized in the Standard Model, but occur at an extremely small, undetectable, rate ($\mathcal{B} \sim 10^{-54}$). A variety of theoretical models beyond the Standard Model allow for CLFV (and thus $\mu^+ \to e^+\gamma$) at a potentially observable rate ($\mathcal{B} \sim 10^{-14}$). CLFV has not been observed despite extensive searches for over 50 years. It's observation would have dramatic implications on the field of particle physics. Chapter 2 discusses past and future experimental searches in the field of CLFV. Chapter 3 describes the MEG II experimental technique, including the signal/background, detectors,

electronics, trigger, etc.

The remaining chapters discuss MEG II analysis. One of the most critical points of the experiment is the background rejection, which is required to achieve the optimal experimental sensitivity. This background rejection is achieved via precise kinematic measurements of the positron and the photon at the stopping target. The precision of these kinematic measurements is common theme that will be discussed throughout this thesis. In Chapter 4, a novel technique used to monitor the MEG II stopping target using a camera is described along with its results. Chapter 5 presents the reconstruction of positrons (e^+) in the MEG II drift chamber. The relative alignment of the MEG sub-detectors and their calibrations are discussed in Chapters 6 and 7 respectively. Chapter 8 describes the details of the physics analysis of the 2021 dataset. Conclusions are described in Chapter 9.

Chapter 1

Theoretical Background

1.1 Overview

This chapter summarizes the current state of particle physics with a focus on muon physics. We start with an explanation of the "Standard Model" of particle physics. It encompasses all electroweak and strong force interactions. In this model, there are only left handed, zero mass neutrinos. We then discuss the addition of massive neutrinos to the Standard Model through the addition of right-handed neutrinos. This allows for mixing of the lepton flavor eigenstates and therefore charged lepton flavor violation (i.e. $\mu^+ \to e^+\gamma$) at an extremely small, unobservable rate ($\sim 10^{-54}$).

The chapter ends with a discuss the limitations and shortcomings of the Standard Model with massive neutrinos. These limitations strongly motivate physics beyond the Standard Model. Many additions to the Standard Model allow for charged lepton flavor violation at a dramatically higher, potentially detectable branching fraction ($\sim 10^{-14}$) and therefore motivate our search for $\mu^+ \to e^+\gamma$. Detecting this process would provide unambiguous evidence of physics beyond the Standard Model.

1.2 Standard Model

The Standard Model consists of the following gauge group:

$$G_{SM} = SU(3)_C \cdot SU(2)_L \cdot U(1)_Y \tag{1.1}$$

with there families of fermions, each with its own representation in this gauge group:

Generation	I	II	III	Representation
Lepton Doublet (l_i)	$\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L$	$\begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L$	$\begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L$	$(1,2)_{-1/2}$
Right-handed Leptons (r_i)	e_R	μ_R	τ_R	$(1,1)_{-1}$
Quark Doublet	$\begin{pmatrix} u \\ d' \end{pmatrix}_L$	$\begin{pmatrix} c \\ s' \end{pmatrix}_L$	$\begin{pmatrix} t \\ b' \end{pmatrix}_L$	$(3,2)_{1/6}$
Right-handed Quarks I	u_R	c_R	t_R	$(3,1)_{2/3}$
Right-handed Quarks II	d_R	s_R	b_R	$(3,1)_{-1/3}$
Higgs Scalar	ϕ	-	-	$(1,2)_{1/2}$

The left-handed lepton and quarks form doublets under the $SU(2)_L$ gauge. All quarks form triplets under the $SU(3)_C$ gauge. The representation subscript denotes the hypercharge, Q_Y, of the $U(1)_Y$ gauge. The hypercharge is based on the known electric charge and constraints from the other gauges. This massless neutrino Standard Model includes three gauge coupling terms, 9 fermion masses, 4 Cabibbo Kobayashi Maskawa (CKM) matrix parameters i.e. QCD mixing parameters that describe the mixing of quark mass eigenstates and mass eigenstates, three other parameters (θ_{QCD}, and two Higgs parameters).

The $SU(2)_L x U(1)_Y$ symmetry breaks at lower energies into the $U(1)_{em}$ symmetry. At low energies, the Standard Model then contains the $SU(3)_C \cdot U(1)_{em}$ symmetry.

1.2.1 Lagrangian

Using this gauge group and the mentioned representations, we write down the Lagrangian for the Standard Model. This Lagrangian describes all Standard Model interactions. Below, we write the terms relevant for the MEG II experiment, that is, excluding the quarks and $SU(3)_C$ gauge.

$$\begin{aligned}
\mathcal{L}_{SM} = & -0.5tr(W_{\mu\nu}W^{\mu\nu}) - 1/4 B_{\mu\nu}B^{\mu\nu} && U(1)_Y, SU(2)_L \text{ Gauge Terms} \\
& + \sum_{i=0}^{i<3} \overline{l_i} i\gamma^\mu D_\mu l_i + \overline{r_i} i\gamma^\mu D_\mu r_i && \text{Lepton Kinetic Terms} \\
& + \sum_{i=0}^{i<3}\sum_{i=0}^{j<3} Y^L_{i,j} \overline{l_i} \phi e_j + h.c. && \text{Yukawa Couplings to Higgs} \\
& + (D_\mu\phi)^\dagger (D^\mu\phi) - 0.5\lambda(\phi\phi^\dagger - v^2_{EW})^2 && \text{Higgs Kinetic, Mass, Quatric Terms}
\end{aligned}$$

The lepton kinetic terms in the second row contain the interactions between the leptons and the gauge $(SU(2)_L \cdot U(1)_Y)$ bosons: $A_\mu, W^{+/-}_\mu, Z_\mu$. In this equation, A_μ, and Z_μ are linear combinations of the original W^3_μ, B_μ bosons through the Weinberg angle. Similarly, W^+_μ, W^-_μ are linear combinations of the original W^1_μ, W^2_μ bosons through a $-45°$ mixing. Below, the lepton doublet kinetic term is shown in more detail (expanding the D_μ):

$$\mathcal{L}_{SM} = \sum_{i=0}^{i<3} \overline{l_i} i\gamma^\mu [id_\mu + \begin{pmatrix} Q^\nu_{em} A_\mu + Q^\nu_N Z_\mu & Q^l_C W^+_\mu \\ Q^l_C W^-_\mu & Q^e_{em,L} A_\mu + Q^e_N Z_\mu \end{pmatrix}] l_i$$

The right-handed electron kinetic term is shown here:

$$\mathcal{L}_{SM} = \sum_{i=0}^{i<3} \overline{r_i}\gamma^{\mu}[id_{\mu} - Q^{e}_{em,R}A_{\mu} - Q^{e}_{N}Z_{\mu}]r_i$$

The constants, Q, each represent different couplings between the leptons and the gauge bosons. It's important to note that these Lagrangian coupling terms for the Z boson and the left handed electron, right-handed electron, and left-handed neutrino are all different. In addition, the right-handed electron does not couple to the W bosons. Of course, this Lagrangian also indicates that the neutrinos do not interaction with the photon (A_{μ}).

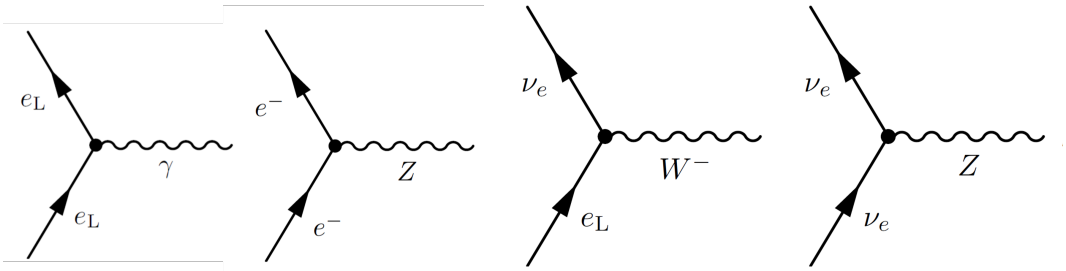


Figure 1.1: Electron, electron neutrino interaction Feynman diagrams.

The mentioned Lagrangian terms can be used to draw Feynman diagrams of lepton interactions. The interactions of electrons and electron neutrinos with the electroweak bosons are drawn in Figure 1.1. The same can be drawn for muons and taus (and their neutrinos). The first represents the coupling of electrons to the photon (A_{μ}), the second represents the coupling of electrons to the Z boson (separate coupling constant for right and left handed electrons), the third represents the coupling of left-handed electrons and left-handed neutrinos to the W boson, and the final diagram represent the coupling of left-handed neutrinos to the Z boson.

We also note that the Yukawa coupling terms in the Lagrangian give rise to the lepton masses after the spontaneous symmetry breaking of the $SU(2)_L \cdot U(1)_Y$ symmetry. The Higgs kinetic term gives rise to the W/Z boson masses.

1.2.2 Conservation of Lepton Number

Both the lepton doublets and the right-handed leptons carry an inherent additional U(1) symmetry. As an example, for the muon doublet and the right handed muon singlet:

$$\begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L \rightarrow e^{i\alpha_\mu} \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L , \quad \mu_R \rightarrow e^{i\alpha_\mu} \mu_R \tag{1.2}$$

The Lagrangian is invariant under this U(1) transformation, which implies a conserved "charge". In this case, this charge is "muon charge" or muon number, which must be conserved in all particle interactions. The muon, and its neutrino both carry muon number equal to one. Their anti-particles carry muon number equal to negative one, and all other particles carry muon number of zero. This conservation law also follows for the electron and tau generations.

1.2.3 Standard Model Muon Decays

We use the Feynman diagram rules and conservation laws described above to build interactions between the generations of leptons. As an example, we show the simplest and most common muon decay, the Michel decay ($\mu^+ \rightarrow e^+ \nu_e \bar{\nu_\mu}$) in Figure 1.2. This is the result of two Standard Model vertices and accounts for $\sim 99\%$ of all muon decays. Next, we show an example of a radiative muon decay (RMD)($\mu^+ \rightarrow e^+ \nu_e \bar{\nu_\mu} \gamma$) in Figure 1.3. This diagram simply adds an additional vertex point and thus occurs at a lower rate, suppressed by a factor of α; the branching fraction of a muon RMD decay with a photon energy of least 10 MeV is 1.4%[16][17]. These are the two most relevant processes for our experiment as nearly all positrons we observe result from one of these two muon processes. More exotic

muon decays include additional Standard Model vertices and propagators and hence have an even lower branching fraction e.g. $\mu^+ \to e^+\nu_e\bar{\nu}_\mu e^+e^-$. These are not very relevant for our experiment.

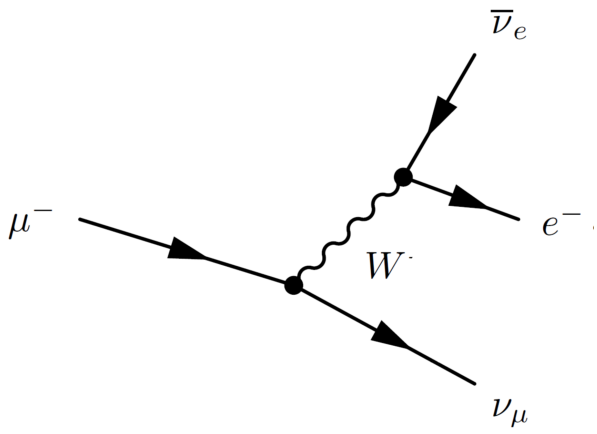


Figure 1.2: Muon Michel decay Feynman diagram, this is the most typical muon decay.

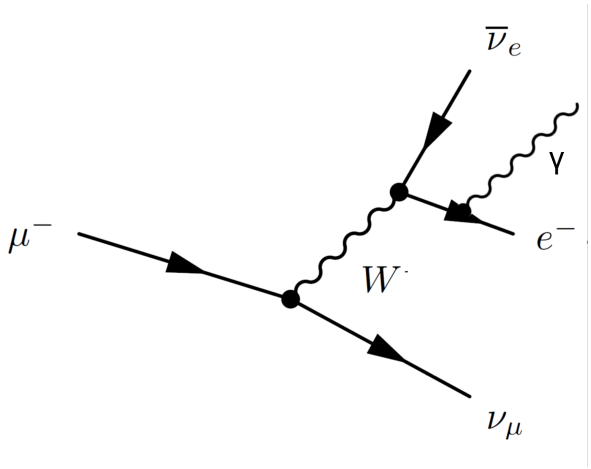


Figure 1.3: Muon radiative muon decay (RMD) Feynman diagram, the second most common muon decay.

1.3 Massive Neutrinos

The Standard Model as discussed above is an approximation assuming zero-mass neutrinos. Originally, it was thought that the neutrino was massless, however it is now known from

several experiments that neutrinos are massive (but very light)[18]. Therefore the Standard Model is clearly incomplete without some modification to include massive neutrinos.

$$\mathcal{L}_D = -Y^{\nu}_{i,j}\overline{l_i}\epsilon\phi * \nu_{R,j} + h.c. \tag{1.3}$$

Just like the Yukawa coupling term for the charged leptons, this term also gives mass to the neutrino. Again, when the Higgs ϕ is replaced by its v_{EV}, we are left with a neutrino mass :

$$\mathcal{L}_D = -\sqrt{(2)}Y^{\nu}_{i,j}v_{EV}\overline{\nu_i}\nu_{R,j} + h.c. \tag{1.4}$$

In addition, with the right-handed neutrino singlet, ν_R, we may introduce yet another Lagrangian term:

$$\mathcal{L}_M = -1/2M_{i,j}\overline{\nu_i}^C_R\nu_{R,j} + h.c. \tag{1.5}$$

Using the two right-handed neutrino singlet terms, we form two mass terms with different relative masses.

$$\mathcal{L}_{see-saw} = -\sqrt{(2)}Y^{\nu}_{i,j}v_{EV}\overline{\nu_i}\nu_{R,j} - 1/2M_{i,j}\overline{\nu_i}^C_R\nu_{R,j} + h.c. \tag{1.6}$$

This can be re-written as a 2x2 mass matrix:

$$\mathcal{L}_{see-saw} \propto \begin{bmatrix} \bar{\nu} & \bar{\nu_R}^C \end{bmatrix} \cdot \begin{bmatrix} 0 & Y^{\nu}_{i,j} v_{EV}/sqrt2 \\ Y^{\nu}_{i,j} v_{EV}/sqrt2 & -1/2 M_{i,j} \end{bmatrix} \begin{bmatrix} \nu \\ \nu_R \end{bmatrix} \tag{1.7}$$

We can calculate the mass eigenstates of this mass matrix:

$$m_{1/2} = M/2(1 \mp \sqrt{1 + \frac{4Y^{\nu}_{i,j} v^2_{EV}}{2M^2}}) \tag{1.8}$$

If M is taken to be very large, the result is very large mass right-handed neutrinos and very light left-handed neutrinos. The masses m_1, m_2 roughly represents the left and right handed neutrino respectively. This yields small mass neutrinos when M is large (10^{14} GeV to result in neutrino masses of the order 0.1 eV). This is often called the See-Saw model.

1.4 Neutrino Oscillations

By including the Yukawa coupling term with right-handed neutrinos, the lepton Lagrangian situation is now analogous to that of the Yukawa couplings in the quark sector. That is, there are two independent matrices: $Y^L_{i,j}, Y^R_{i,j}$. Previously with only $Y^L_{i,j}$, we select a basis of the flavor states such that this matrix is diagonal and therefore the flavor states are identical to the mass states. However, now we can only diagonalize one of the two matrices, i.e. in the previously defined basis, $Y^R_{i,j}$ will not be diagonal. Therefore, the interactions between the neutrinos, electrons, and the W boson will not be in the mass states, but the flavor states. Therefore, we violate the conservation laws of lepton number!

We are left with three flavor states ($\nu_l : l = e, \mu\tau$) and three mass states ($\nu_i : i = 1, 2, 3$) for the neutrinos. The mixing between the mass and flavor eigenstates is shown below:

$$U_{MNS}^{T} m_{\nu}^{Majorana} U_{MNS} = \begin{bmatrix} m_1 & 0 & 0 \\ 0 & m_2 & 0 \\ 0 & 0 & m_3 \end{bmatrix} \tag{1.9}$$

And,

$$\nu_l = \sum_i U_{MNS}^{\dagger} v_i \tag{1.10}$$

Many experiments have measured this U_{MNS} matrix, the latest values are described here[19]. Rough values are shown below:

$$U_{MNS} = \begin{bmatrix} 0.82 & 0.55 & 0.15 \\ 0.42 & 0.62 & 0.66 \\ 0.38 & 0.56 & 0.73 \end{bmatrix} \tag{1.11}$$

1.4.1 Standard Model $\mu \to e\gamma$

The mixing of flavor and mass eigenstates violates lepton flavor number and also allows for the possibility of an alternate muon decay that violates lepton number e.g. $\mu \to e\gamma$. This is shown in Figure 1.4. This includes two elements of the flavor-mass mixing U_{MNS} matrix.

However, this occurs at an extremely small rate as the branching fraction is proportional to the sum of the squared mass splittings between the neutrinos divided by the mass of the W boson, quantity squared. Here, we are summing over the three possible flavor eigenstates (clearly with identical mass neutrinos this term falls to zero). The branching fraction is shown below[20]:

$$BR(\mu \to e\gamma)_{SM} = \frac{3\alpha}{32\pi} \sum_i U^{*\mu,i}_{MNS} U^{i,e}_{MNS} (\frac{\Delta^2_{\nu,i}}{m^2_W})^2 \sim 10^{-4} \cdot (\frac{10^{-3}}{(10^{22})})^2 \sim 10^{-54} \tag{1.12}$$

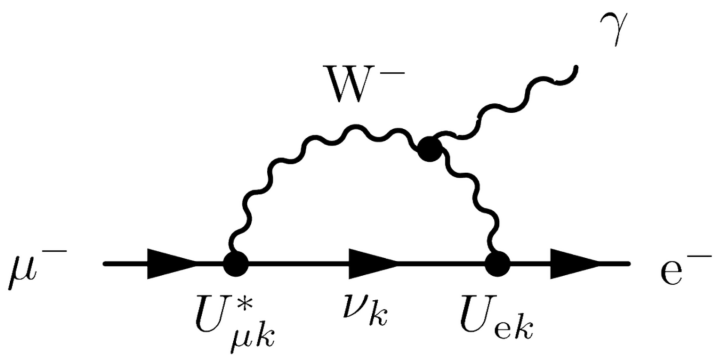


Figure 1.4: The Feynman diagram of $\mu \to e\gamma$ in the Standard Model through the mixing of neutrino flavor and mass eigenstates.

This rate is so small that it could never be detected by any conceivable experiment (nearly 40 orders of magnitude away from state of the art muon experiments).

1.5 General Physics Beyond the Standard Model

Although the Standard Model including massive neutrinos does yield a great approximate solution to a large fraction of particle physics, there are still some unsolved problems and thus it is known to be incomplete. For this reason, additional theoretical physics models beyond the Standard Model (BTSM) are highly motivated.

In the next couple sections, we discuss two BTSM theories, SUSY SeeSaw[1] and SUSY GUT[21]. In both cases, well-known issues with the Standard Model can be resolved. In these cases as well as many others, charged lepton flavor violation is expected to occur at a dramatically higher rate than that due to neutrino mixing in the Standard Model. This is a key motivation behind charged lepton flavor violating experiments like MEG II. Observing charged lepton flavor violation would be unambiguous evidence of physics BTSM.

1.6 Hierarchy Problem and Super Symmetry See Saw

The first theoretical problem is known as the hierarchy problem. In this case, we focus on the first order correction to the Higgs propagator. This is shown in Figure 1.5. The diagram contains with two fermions of momentum p and $k+p$ with Yukawa coupling y_k, and effectively mass scale Λ^2. This results in the following amplitude squared:

$$-M^2 \propto \lambda_f^2 \int d^4k \frac{\cdot p - k \cdot (k)}{[(p-k) - m_f^2][(k) - m_f^2]} \propto y_k^2 \Lambda^2 \tag{1.13}$$

This correction term quadratically diverges with the mass scale and thus results in a term significantly larger than the original propagator.

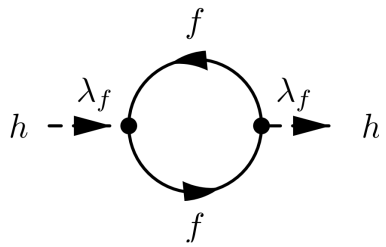


Figure 1.5: One loop correction to the Higgs propagator.

There are many ways to solve this theoretical problem. For example, this can be solved by the theory of super symmetry. In this model, we assert an equal number of bosonic and

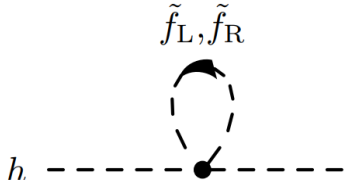


Figure 1.6: Super symmetric correction to the Higgs propagator.

fermionic degrees of freedom. That is, for each known boson, we have a "supersymmetric" fermion counterpart and for each known fermion we have a "supersymmetric" boson. This model can be tuned such that the Higgs interaction with the fermion (e.g. electron) and boson (e.g. selectron) cancel one another via relationships between their Yukawa couplings. The Higgs propagator correction due to the Higgs interaction with a super-symmetric boson is shown in Figure 1.6 and the amplitude is shown below:

$$-M^2 \propto \widetilde{\lambda}_f^2 \int d^4k \frac{1}{[(k) - m_{\tilde{f}}^2]} \propto y_k^2 \Lambda^2 \tag{1.14}$$

Combining this super symmetric theory with massive neutrinos is often called SUSY SeeSaw (SUper SYmmetry See Saw). A consequence of SUSY See Saw is charged lepton flavor violation at a dramatically higher rate. In contrast with the SM charged lepton flavor violation, the Feynman diagram loop "selectron" and "smuon" have comparable masses to the "photino" (also in the loop) and therefore charged lepton flavor violation can occur at a significantly higher rate than the SM $\mu \to e\gamma$ decay.

In short, the see-saw neutrino model above results in Majorana neutrinos (mainly left-handed) with very small masses and the Dirac neutrinos (mainly right-handed) have very large masses. This combined with super symmetry results in lepton-slepton mixing that is dependent on the mass of the Dirac neutrinos. The lepton-slepton interactions can violate lepton number (e.g. lepton $\to$ slepton $\to$ lepton') and results in a branching ratio of $\mu \to e\gamma$

given below:

$$\mathcal{B}_{\mu\to e\gamma} \propto (m_\nu^D)^4/m^8 \tag{1.15}$$

We assume a fixed photino mass of $O(0.1m)$ where m is the scale of Super Symmetry breaking. The Feynman diagram is shown in Figure 1.7.

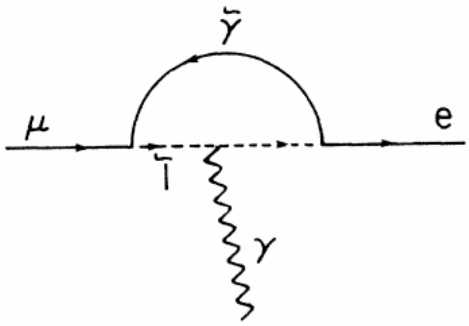


Figure 1.7: The Feynman diagram of $\mu \to e\gamma$ in the Super Symmetric See Saw model (Figure 1 ©1986 American Physics Society [1]).

In addition, we show the plot the branching ratio of $\mu \to e\gamma$ as a function of m, m_ν^D in Figure 1.8.

Since none of the super symmetric counter parts have been discovered, these would need to have masses at significantly higher energy than any of the Standard Model particles.

1.7 Super Symmetry and Grand Unification

Another issue with the the Standard Model is less quantitative. The SM contains a large number of parameters, and the representations don't really follow any underlying pattern or structure. That is, why do they quarks form a 3-plet under SU(3), the leptons form a

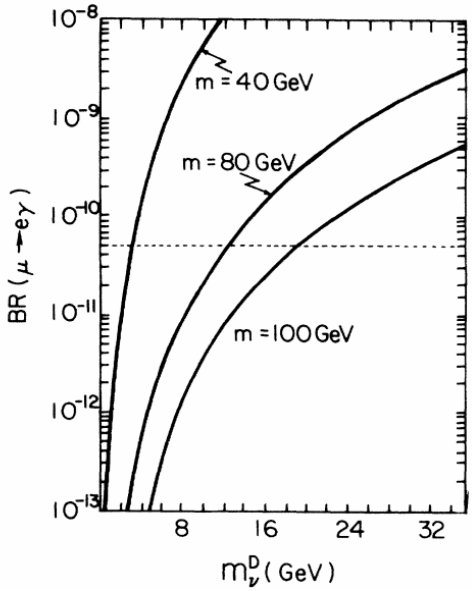


Figure 1.8: The branching ratio of $\mu \to e\gamma$ as a function of m, m_ν^D parameters (Figure 2 ©1986 American Physics Society [1]).

2-plet under SU(2)? In addition, this list of parameters in the Standard Model gets more cumbersome with the addition of massive neutrinos. One solution to this is a theorized Grand Unified model, that is all the symmetries converge into a single large symmetry at higher energies. To motivate this, first we show in Figure 1.9 the scale of the SM couplings as a function of energy.

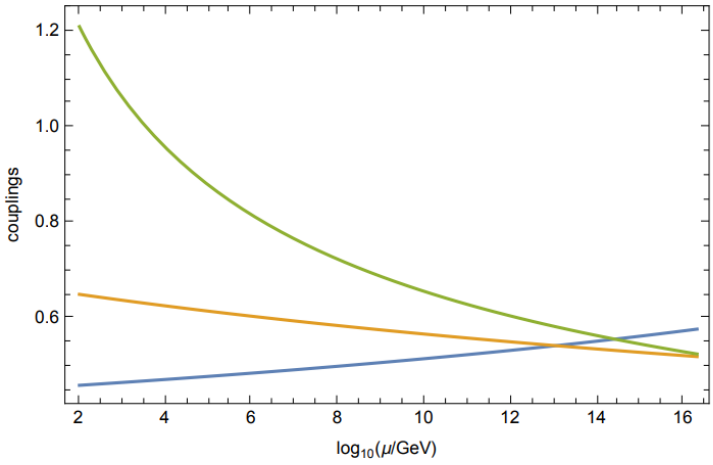


Figure 1.9: Scale of the Standard Model couplings with energy (Figure made by Michael Ratz)

At higher energies, these coupling constants are close, but not exactly identical. However,

as mentioned above, it is known that the Standard Model does not exactly describe our universe at higher energies. Therefore, it is reasonable to add new physics that may result in the couplings converging to a single coupling energy scale at high energy. In the minimal supersymmetric model (MSSM) it is possible to achieve the same scale for all three couplings at high energies! This is shown below in Figure 1.10. Here, the super symmetric partners have a common mass of the order TeV.

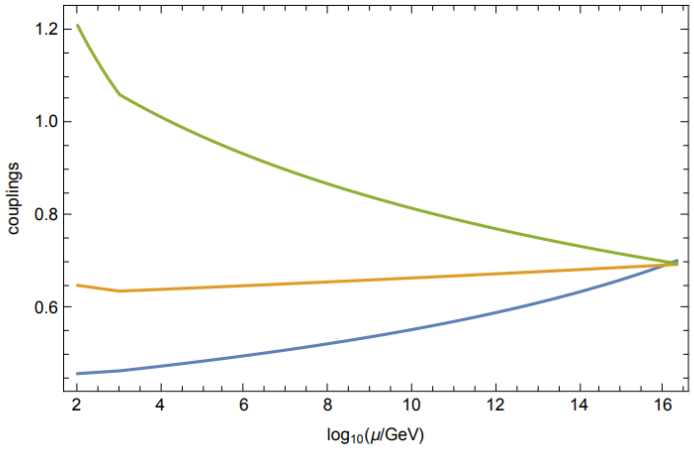


Figure 1.10: Scale of the Standard Model coupling with the addition of the Minimal supersymmetric model (MSSM) (Figure made by Michael Ratz).

This highly motivates a larger symmetry at higher energies ($\sim 10^{16} GeV$). The standard model symmetries can all be derived from a single SU(5) symmetry, which is then broken at lower energies. In SUper SYmmetry Grand Unification Theory models (SUSY GUT), there are interactions between the bosons ($X^{+,-,0}$) and the leptons/sleptons ($\tilde{\nu}$ is a neutral slepton and $L_{-/+}$ is a charged slepton). This allows for additional channels of charged lepton flavor violation. The Feynman diagram vertices are shown below in Figures 1.11 and 1.12. The $\mu^+ \rightarrow e^+\gamma$ decay is built by combining these vertices. This is expected to occur at a dramatically higher, potentially detectable rate.

The Standard Model can alternatively be composed of a SO(10) symmetry (SU(5) x U(1)), which yields the possibility for a right-handed neutrino representation; this avoids much of

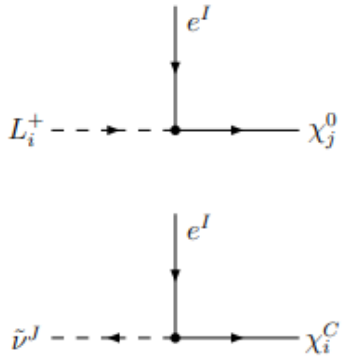


Figure 1.11: Feynman rules for electrons, neutrinos, and the MSSM particles relevant for $\mu \to e\gamma$.

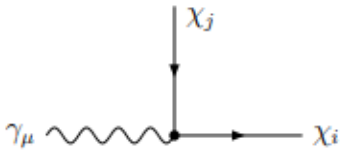


Figure 1.12: Feynman rules for photons and the MSSM particles relevant for $\mu \to e\gamma$.

the cumbersome additional treatment required to add the massive neutrinos to the Standard Model.

1.8 Closing Theoretical Points

We started this chapter with a brief summary of the Standard Model and lepton physics. We then discussed the introduction of massive neutrinos that results in charged lepton flavor violation at a extremely small undetectable rate.

We then described two issues with the Standard Model; these are examples that motivate physics BTSM. We described two SUSY examples of physics beyond the Standard Model, but there are many more. In a large fraction of these BTSM theories, charged lepton flavor is expected to occur at a significantly higher, detectable rate! This motivates searching for charged lepton flavor violation in particle physics experiments as a probe to physics BTSM.

In the next chapter we start with a theoretical comparison of different charged lepton flavor violating searches and then a provide a history of experimental searches for charged lepton flavor violation. We then discuss recent and planned charged lepton flavor violating experiments.

Chapter 2

Charged Lepton Flavor Violation Experiments

In this chapter, we discuss experimental searches for charged lepton flavor violation (CLFV). First, we motivate searching for different CLFV processes and describe the relative search power of different CLFV channels to different BTSM theories. We then discuss past searches for CLFV and upcoming searches for other muon decays or conversions e.g. $\mu^+ \to e^+e^-e^+$, $\mu^- N \to e^- N$.

2.1 Theoretical Comparison of CLFV Channels

In general, if any CLFV experiment detects a signal, it would imply that most CLFV processes occur at a significantly higher rate than the SM. Measuring only one CLFV channel would confirm physics beyond the standard model but not be able to pin-point an exact BTSM theory. Instead by measuring the branching fraction of several different CLFV modes one could start to narrow down which BTSM best reflects the experimental results.

The comparison of different BTSM theories on different CLFV channels can be done by inspecting the Feynman diagrams of different theoretical models. For example, Figure 2.1 (originally made by Bill Marciano) contains six Feynman diagrams of $\mu N \to eN$ in a variety of BTSM theories. These graphics all contain $\mu N \to eN$, but anywhere $\mu N \to eN$ occurs, $\mu \to e\gamma$ will also occur. One can tell via inspection that $\mu \to e\gamma$ will be an ideal search for SUSY as $\mu N \to eN$ requires a photon propagator. On the other hand, Leptoquark or Compositeness models are more ideal for a $\mu N \to eN$ search.

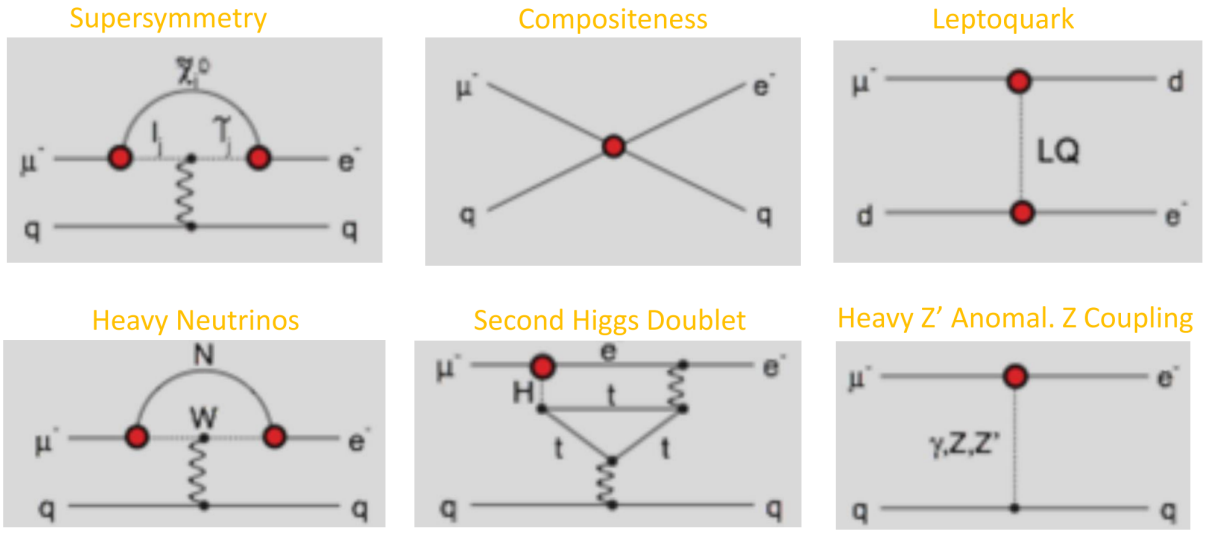


Figure 2.1: Set of six Feynman diagrams that all produce $\mu N \to eN$. All of these processes also make $\mu \to e\gamma$. (Figure originally made by Bill Marciano).

We can also arrange these classes of theoretical models into groups. This can be done through the use of a model-independent effective Lagrangian[2]. We discuss an example comparison, consider a Lagrangian with two terms: the first is a "magnetic-momentum type operator" ideal for $\mu \to e\gamma$ searches and the second is a 4-fermion Lagrangian, ideal for $\mu N \to eN$ searches. Both Lagrangian terms contain a κ parameter to dictate the relative strength of the two terms. This discussion only mentions two charged lepton flavor violating processes, but this technique could be used to compare any set of processes.

$$\mathcal{L}_{CLFV} = \frac{m_\mu}{(\kappa+1)\Delta^2}\bar{\mu_R}\sigma_{\mu\nu}e_L F^{\mu\nu} + \frac{\kappa}{(\kappa+1)\Delta^2}\bar{\mu_R}\gamma_\mu e_L(\bar{u_L}\gamma^\mu u_L + \bar{d_L}\gamma^\mu d_L) + h.c. \quad (2.1)$$

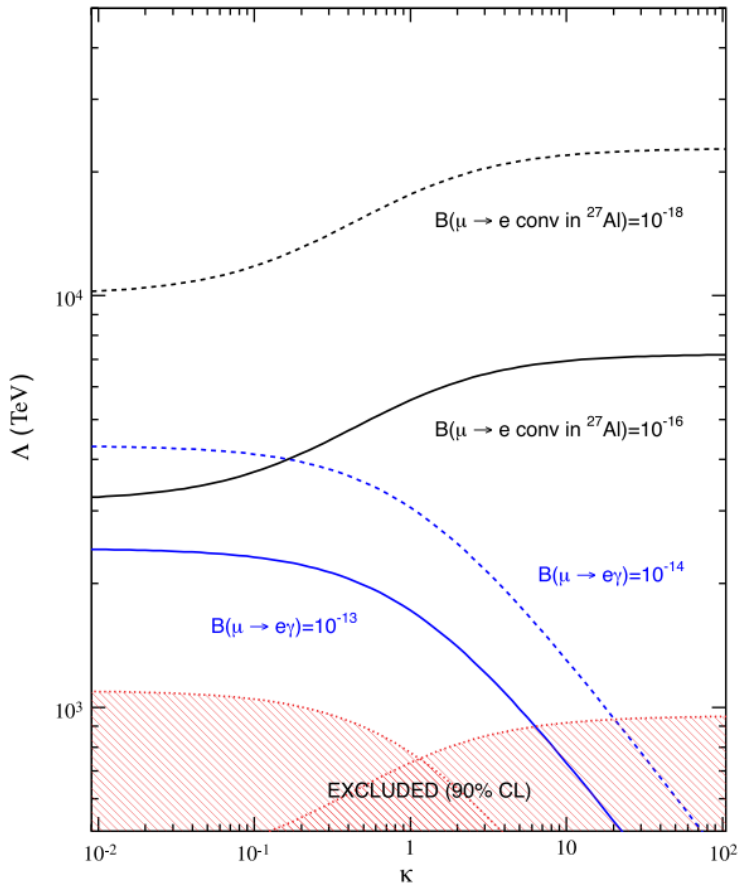


Figure 2.2: Comparison of the physics scale Δ reached by $\mu \to e\gamma$ and $\mu N \to eN$ searches at varying sensitivity as a function of κ, a term determining the relative scaling of two terms in the effective Lagrangian shown in equation 2.1. (Figure 2 ©2013 Elsevier [2])

In Figure 2.2, we show the physics scale Δ reached for $\mu \to e\gamma$ and $\mu N \to eN$ searches as a function of κ and thus the weighting of the two Lagrangian terms. The bottom of the plot represents phase space that has already been excluded by previous experiments.

Of course, the most interesting points on this plot are the end points of the x-axis. The middle κ region represents a region which coincidentally allows for two methods to achieve CLFV. The specific curves roughly represent the latest and upcoming CLFV experiments for

these muon channels. The solid/dotted blue curves represent MEG I/MEG II experiment sensitivities (MEG II expected) and the black solid/dotted curves roughly represent the expected sensitivities for Mu2e[3]/Mu2e-II[22] experiments at Fermilab.

At the small κ limit, the $\mu \to e\gamma$ branching fraction required to reach the same physics scale is lower by about two orders of magnitude than that of $\mu N \to eN$. This plot indicates that if MEG II detects the $\mu \to e\gamma$ signal, this would likely imply a signal should be observed in Mu2e! However, of course the opposite is not the case i.e. if our universe is in a large κ region, MEG II would not observe a signal, but a signal could still be detected by Mu2e. This is an explicit example of the synergy between the two CLFV searches.

In the next section we briefly discuss the history of CLFV searches.

2.2 Past Experimental Searches

The sensitivity of CLFV experiments with muons over time is shown in Figure 2.3; it has improved incrementally since the 1940s.

Although the searches have all resulted in a null signal, they still had very significant impacts on the field of particle physics. For example, in the first searches for $\mu \to e\gamma$, the null result indicated that the muon was a unique particle with respect to the electron. Today, even the absence of charged lepton flavor violation at smaller branching fractions would have a significant impact on the field since many BTSM theories include charged lepton flavor violation at a potentially detectable rate. Therefore, a null signal would significantly restrict model parameter spaces or may reject theories entirely.

Figure 2.3 includes the three most common CLFV channels: $\mu \to e\gamma$, $\mu \to eee$, and $\mu N \to eN$. They are known as the "golden" CLFV channels as they are the most sensitive searches

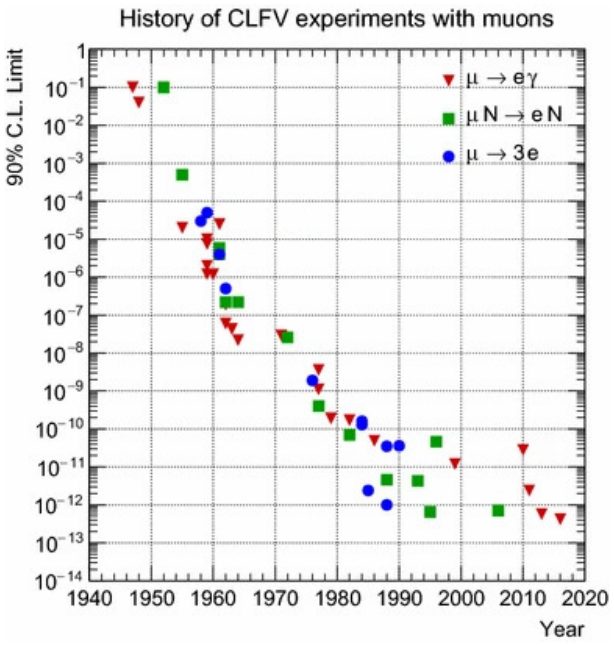


Figure 2.3: The 90% upper limit of CLFV muon experiments over time (Figure 1 ©2018 Springer).

to a broad range of BTSM theories. The experimental searches for these three processes have progressed for over 50 years. The sensitivity of the competing experiments is always improving from higher beam rates, improved technology, or design improvements.

Although these are some of the most common CLFV channels, others have also been explored. As an example, the MEG I dataset was also used to search for $\mu^+ \to e^+X; X \to \gamma\gamma$[23]. There are also ongoing discussions about using the MEG II experiment to search for a CLFV axion signal[24]. CLFV experiments have also searched for tau flavor violating processes e.g. $\tau \to \mu\gamma$, $\tau \to e\gamma$, $\tau \to \mu\mu\mu$, etc. the most recent of which were done by the Belle[25] and BaBar[26] experiments. Even more are listed here[27].

In the next sections, we discuss two examples of upcoming muon CLFV experiments, Mu3e (search for $\mu^+ \to e^+e^-e^+$) and Mu2e (search for $\mu^-N \to e^-N$). We briefly outline these experiments to give a sense of the CLFV field in the years to come. Over the full lifetime of the experiments, they plan to achieve sensitivities beyond MEG II. However as mentioned, MEG II is still competitive for BTSM theories like SUSY, which require a photon propagator

for both Mu2e and Mu3e processes.

2.3 Mu2e Overview

The Mu2e experiment[3] searches an electron resulting from a $\mu^- N \to e^- N$ conversion with a momentum of $E_e \sim m_\mu$ (105.0 MeV if the N is Al). The objective is to detect this signal while rejecting all backgrounds. The experiment is expected to start its first physics data in 2025. Over the lifetime of the experiment, Mu2e will either detect a signal or expects to set an upper limit at the 90% confidence level on the branching fraction of $\mu N \to eN$ of $6 \cdot 10^{-17}$ surpassing its predecessor (SINDRUM II[28]) by 4 orders of magnitude in sensitivity.

In Mu2e, unlike MEG II, there is no accidental background due to time coincidences between decay products. Instead, there are backgrounds from several different processes. First, there is background from the beam. This is avoided by using a pulsed beam with a width of 200 ns when the $\mu N \to eN$ conversion is expected to take on the order $\sim 200 - 2000$ ns; therefore, all of the "prompt" beam background (100-700 ns) is eliminated by using a delayed search window (800-1700 ns) at the expense of some signal efficiency. This requires very precise knowledge that the beam is completely eliminated; this is verified using a separate detector that monitors the beam 'extinction'. Second, there is background from cosmic rays that can produce an electron at the signal energy; this background is removed by the cosmic ray veto module (CRV). The CRV is an array of cosmic ray detection modules that surrounds the detector area (tracker, calorimeter); if a cosmic ray is detected, the events nearby in time are eliminated from the data sample. Finally, there are muon decays in orbit (DIO), which can result in a high momentum electron (reaching $E_e \sim 104$ MeV in the recoil tail). This is rejected by $\sigma_{E_e} \sim 100$ keV in the tracker.

Figure 2.4 shows an overview of the experimental design. The electrons (muon decay prod-

uct) are measured in a straw tube tracker. The straw tube tracker consists of $\sim 20k$ straw tubes in an array of panels each with 96 straw tubes that are oriented perpendicular to the beam axis. The straw tubes are 5 mm in diameter, filled with a 80/20 Ar/CO2 mixture, and are expected to achieve a $\sim 100\mu$m resolution. The electrons eventually intersect the calorimeter which is used for the trigger and T_0 measurements for the electron tracks. The calorimeter consists of an array of scintillators and SiPMs to detect the scintillation light.

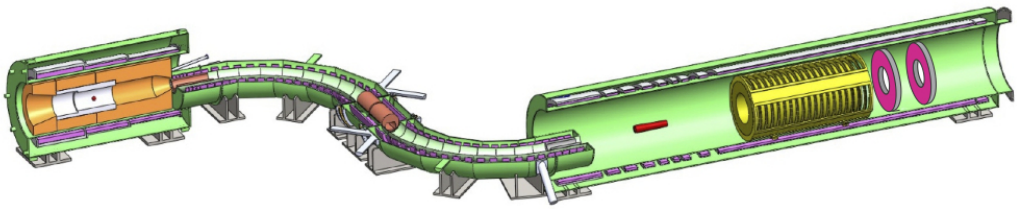


Figure 2.4: The experimental design of the Mu2e experiment[3].

Note that there is an additional experiment searching for $\mu N \to eN$ conversion, COMET[29], which plans for data taking on a similar time schedule with a different experimental design. In addition, Mu2e upgrades (Mu2e-II)[22] are planned in the future to achieve further sensitivity improvements of another order of magnitude.

2.4 Mu3e Overview

The Mu3e experiment[4] is a search for the decay of a muon to three electrons ($\mu^+ \to e^+e^-e^+$). The experiment aims to detect a signal or set an upper-limit on the branching fraction of $\mu^+ \to e^+e^-e^+$ of 10^{-16} at the 90% confidence level.

Similar to MEG II, the experiment expects background due to accidental time coincidences

and $\mu^+ \to e^+e^-e^+\nu_e\bar{\nu_\mu}$ decays (analogous to the RMD decay for MEG II : $\mu^+ \to e^+\nu_e\bar{\nu_\mu}\gamma$). The signal is discriminated from the background using precise measurements of the electrons/positrons at the stopping target (position, time) and using the momentum of all three electron/positrons (must sum up to m_μ and be consisted with a muon decay at rest).

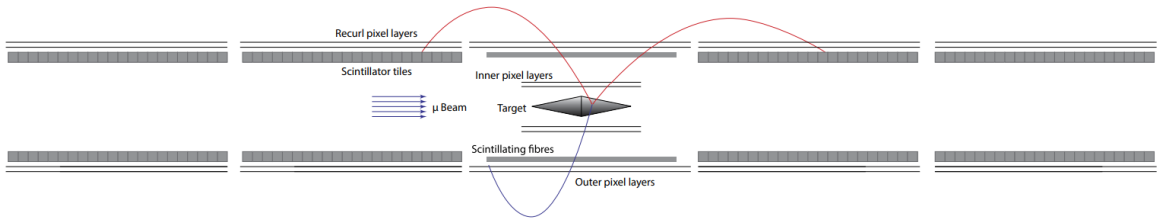


Figure 2.5: The experimental design of the Mu3e experiment[4].

Figure 2.5 depicts the experimental design. The experiment proceeds with a muon stopping beam up to $\sim 2 \cdot 10^9$ Hz. A majority of the muons are stopped and decay in a stopping target. The positrons and electrons which are the result of muon decays are then detected by an array of scintillating fibers/tiles for timing, and an array of active pixel sensors to measure the particle's position at the sensor and therefore the momentum.

Chapter 3

MEG II Experiment

In this chapter, we outline the MEG II experiment and its search for $\mu^+ \to e^+\gamma$. We start with the signal signature, the physics background, and how we distinguish between the two. We then discuss the MEG II predecessor, MEG I, which currently has the most stringent limit on $\mu^+ \to e^+\gamma$. Finally, we discuss the design of the MEG II experiment including the beam, target, detectors, electronics and the trigger. The experiment aims to reach a sensitivity of $6 \cdot 10^{-14}$ at the 90% confidence level, improving upon the upper-limit of MEG I by an order of magnitude (upper-limit of $4.2 \cdot 10^{-13}$ at the 90% confidence level).

The experiment is based at the Paul Scherrer Institut (PSI) in Villigen Switzerland. The experiment is run by an international collaboration with about 50 collaborators from Italy, Japan, Russia, Switzerland, and the US.

3.1 Signal vs. Background

The signal consists of a e^+/γ pair that is time-coincident at the stopping target back-to-back each with an energy equal to half the muon rest mass. Due to the high decay rate and the

kinematics, the primary background is due to a high-energy photon from a radiative muon decay (RMD, $\mu \to e\gamma\nu\bar{\nu}$) or annihilation in flight (AIF, $e^+e^- \to \gamma\gamma$) that is time-coincident with a high-energy Michel e^+ ($\mu \to e\nu_\mu\overline{\nu_e}$) from two separate decays. The other source of background is the result of true RMD decays in which the two neutrinos carry off a small amount of energy.

As the signal and background result in the same particles (positron and photon), the experiment requires precise kinematic measurements to distinguish between the signal and background. The upgraded design with respect to MEG I is motivated in part by this precision requirement.

The positron kinematics are measured in a drift chamber. The positron position, momentum, and direction $(y_e, z_e, p_e, \theta_e, \phi_e)$ at the target are estimated by propagating the positron trajectory from the chamber to the target using a Kalman Deterministic Annealing Filter (DAF) method[30][31]. The time of the positron (t_e) is measured using a set of pixelated timing counters (SPX); the time at the target is measured by propagating the track from the SPX to the target. A liquid xenon detector (LXe) measures the photon kinematics at the calorimeter $(\vec{X_\gamma}, E_\gamma, t_\gamma)$. We then assume that the photon originates at the positron target vertex (y_e, z_e). The analysis uses $t_{e\gamma}, \phi_{e\gamma}, \theta_{e\gamma}, E_\gamma, p_e$ to discriminate signal from background.

In general, the analysis is designed to maximize the number of signal events while minimizing the number of backgrounds. The number of signal events is proportional to the following where R_μ is the beam rate, T is the experiment lifetime, Ω is the geometric acceptance, $\epsilon_\gamma, \epsilon_e$ are the photon, positron efficiencies, and ϵ_{select} is the final event selection criteria.

$$N_{SIG} \propto R_\mu \cdot T \cdot \Omega \cdot \epsilon_\gamma \cdot \epsilon_e \cdot \epsilon_{select} \tag{3.1}$$

Whereas the number of accidental events is proportional to the following where σ_{E_γ},σ_{p_e},$\sigma_{t_{e\gamma}}$,$\sigma_{\phi_{e\gamma}}$, $\sigma_{\theta_{e\gamma}}$ are the resolutions of the photon energy, the positron momentum, the relative positron photon timing, and the relative positron photon angles respectively.

$$N_{ACC} \propto R_\mu^2 \cdot T \cdot \sigma_{E_\gamma}^2 \cdot \sigma_{p_e} \cdot \sigma_{t_{e\gamma}} \cdot \sigma_{\phi_{e\gamma}} \cdot \sigma_{\theta_{e\gamma}} \tag{3.2}$$

Note, from the equations above it's clear that increasing the beam rate increases the number of accidentals at a higher rate than the increase in expected signal. Therefore, a careful balance between the number of expected signal and accidentals is required. Roughly, an experiment with a 10% acceptance and 50% efficiency is required to run at $5 \cdot 10^7 \mu/s$ at 50 weeks of physics data-taking to reach a single event sensitivity (SES) of $1 \cdot 10^{-14}$. The SES is defined as the signal branching fraction such that we expect to observe one signal event in the full data set.

Therefore, the MEG I and MEG II experimental goal is to design an experiment and analysis procedure to optimize the signal efficiency and kinematic resolutions while handling the beam rate required to reach the goal sensitivity. The signal efficiency and kinematic requirements will be repeatedly discussed through the thesis as they are critical to achieving our goal.

3.2 MEG I Overview

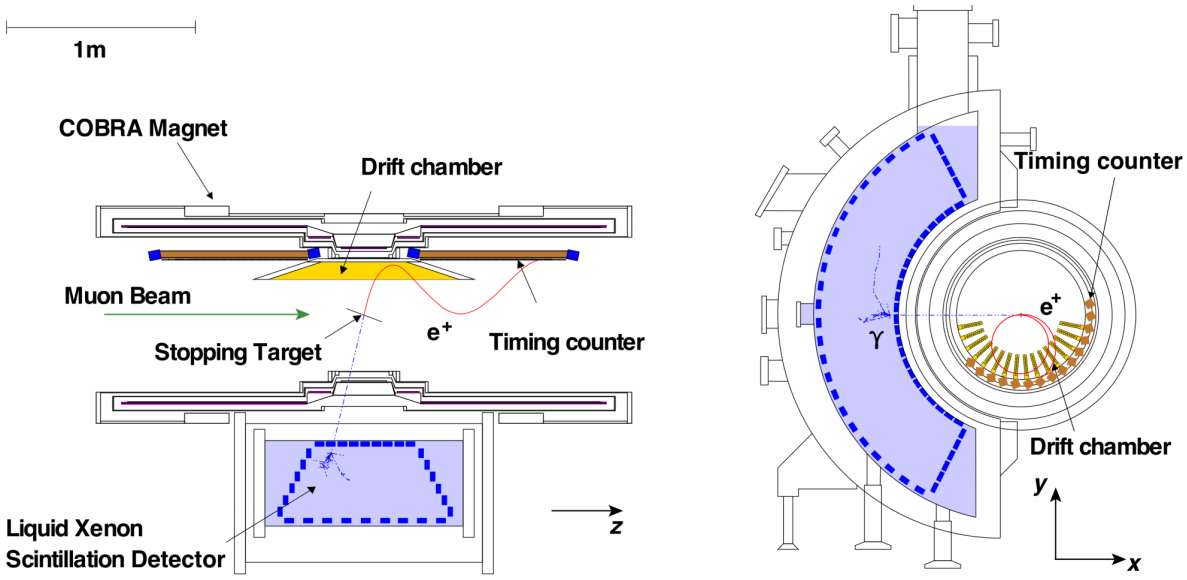


Figure 3.1: The experimental design of the MEG I experiment (Figure 1 ©2016 Springer [5]).

The MEG I experiment was a search for $\mu^+ \to e^+\gamma$ that collected data from 2009-2013 and published final results in 2016[5]. The experiment achieved an upper-limit of $4.2 \cdot 10^{-13}$ at

the 90% confidence level, which is the most stringent limit on the decay.

The MEG I experimental design is shown in Figure 3.1. The experiment proceeds with a muon beam that results in muon decays in a stopping target. The muon decay products are then measured in a series of detectors.

The positrons are measured in a magnetic spectrometer. The magnetic field is generated by the COBRA magnet[8]. The COBRA magnetic field is grated with a maximum field of ~ 1.3 T near the target and drops off along the beam axis. This magnet design has the key advantage of moving positrons away from the target in the direction of the beam axis and thus out of the detector. This was based on one of the key shortcomings of the previous search for $\mu^+ \to e^+\gamma$, the MEGA experiment[32], in which positrons emitted nearly perpendicular to the beam axis would stay in the area taking many turns through the detector. In MEG I, the positrons were measured in a drift chamber with 16 modules, each with two layers of drift cells. The cells consisted of two cathode foils separated by ~ 7 mm with a layer of sense and potential wires (the sense/potential wires in the two layers had an alternating geometry) and was filled with 50:50 mixture of $He : C_2H_6$. The positron timing was measured by an array of scintillating bars connected to PMTs. The positron reconstruction faced significant degradation in the tracking efficiency as the positrons would often intersect material (drift chamber electronic, wires, etc.) prior to intersecting the timing counter. This is discussed later as it in part motivated the design of the MEG II experiment.

The photons are measured in a single volume liquid xenon calorimeter. The photon shower is fully contained inside the liquid xenon, the shower is detected by an array of PMTs surrounding the liquid xenon. One of the shortcomings of the LXe detector was the inner-face; there were instances when the shower would initiate very close to the inner face and a large fraction of the light would escape without hitting the active area of the PMTs; this is addressed in the design of the MEG II experiment as well.

The MEG I experiment improved upon the previous most stringent upper-limit on $\mu \to e\gamma$, set by MEGA, by a factor of ~ 30. However, as will be discussed in the next section, the experience and issues observed in MEG I led to the design of an upgraded detector system, MEG II.

3.3 MEG II Overview

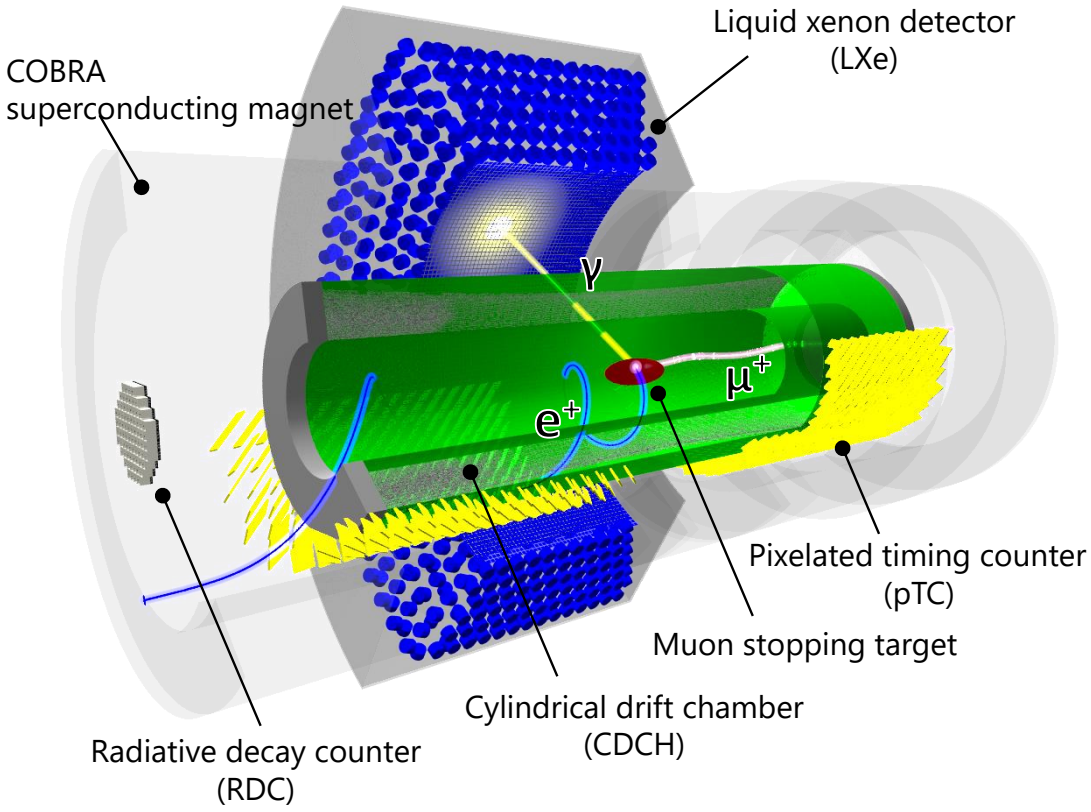


Figure 3.2: MEG II experimental design.

The MEG II experiment is an ongoing search for $\mu^+ \to e^+\gamma$. The experiment maintains the general procedure used by MEG I, but includes both hardware upgrades in all previous sub-detectors and additional sub-detectors to overcome limitations faced in MEG I.

The experiment proceeds by stopping a muon beam (μ^+) into a thin film target; a majority of the muons are stopped in the target and decay at rest. The decay products (i.e. e^+, γ) are measured in a series of detectors. Figure 3.2 shows the MEG II experimental layout. The positrons are measured in a magnetic spectrometer. The core of the magnetic spectrometer is the Cylindrical Drift Chamber (CDCH). The CDCH is a completely redesigned cylindrical ultra-light weight stereo wire drift chamber with full ϕ coverage. The CDCH is centered about the stopping target. The positron momentum is measured using the Constant Bending Radius (COBRA) graded magnet. The magnetic field configuration has a maximum field of $\sim$ 1.3 T, centered at the target resulting in approximately equal positron maximum radius independent of emission angle. In addition, the field configuration moves positrons with small axial momenta away from the target and out of the detector region, ideal for avoiding backgrounds. The COBRA magnet is that used in MEG I. The positrons are either emitted upstream or downstream of the target depending on the positron kinematics. The positron time is measured in a new detector: the pixelated timing counter (SPX). The SPX consists of two semi-cylindrical modules (upstream and downstream) each containing 256 timing counter tiles. The SPX is located at larger radius than the CDCH outer-radius. The photons are measured in the same liquid xenon calorimeter (LXe) used in the MEG I experiment, but the detector has been upgraded. The LXe is a fully absorbing liquid xenon calorimeter (LXe). The LXe inner face is now covered by 4096 Multi Pixel Photon Counters (MPPCs) and the other sides remain covered by Photo Multiplier Tubes (PMTs). The MEG II experiment contains a new radiative decay counter (RDC) downstream of the target used to eliminate events due to a time-coincidence between a Michel positron and an RMD photon from the dataset. Finally, the experiment operates two cameras to photographically image the stopping target continuously to monitor all target motion. The following sections contain details on all MEG II systems.

3.4 Accelerator and Beam line

The MEG II muon beam is generated from PSI's proton accelerator. The core of the accelerator is the proton ring cyclotron shown in Figure 3.3. The cyclotron accelerates protons to 590 MeV; with each pass, the protons are accelerated using electric fields and reach a larger radius due to the larger momentum in the constant magnetic field.

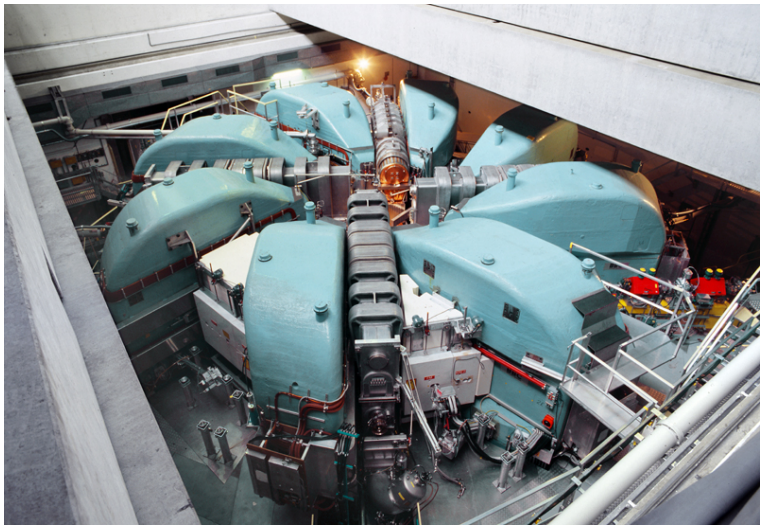


Figure 3.3: PSI's proton ring cyclotron.

Proton collisions in a fixed carbon target result in π^+ that enter PSI's $\pi E5$ line (Figure 3.4). Some π^+ quickly decay at the surface of the target, thus creating μ^+. This produces a "surface" μ^+ beam with a maximum rate nearing 10^8 Hz. The surface muon beam is ideal as it results in a narrow momentum distribution ($FWHM \sim 7\%$) of low momentum muons (28 MeV/c). The low momentum is ideal for achieving a high stop rate in a thin target (ideal for suppressing background rates). Note, this "surface" muon beam is only possible for μ^+ production as π^- will interact with target nuclei, therefore μ^- production relies on pions that decay in flight.

The μ^+ beam is brought into the MEG II experimental hall and directed at the MEG II stopping target. At the MEG II stopping target, the beam is approximately gaussian in both directions normal to the beam axis (X_{MEG}, Y_{MEG}) where $\sigma_{X_{MEG},Y_{MEG}} \sim 1$ cm. The

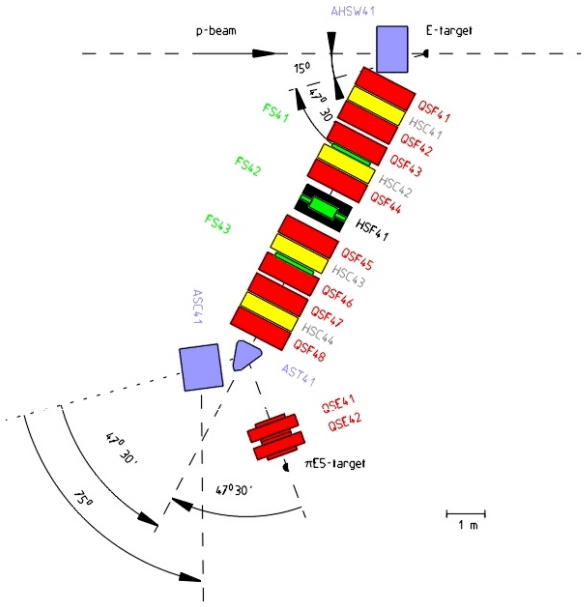


Figure 3.4: $\pi E5$ beam line.

beam spot was measured using an auxiliary detector mounted at the target position during commissioning. The profile is shown in Figure 3.5. This narrow beam width is ideal as it allows the use of a smaller target and thus less material, less background (e.g. AIF).

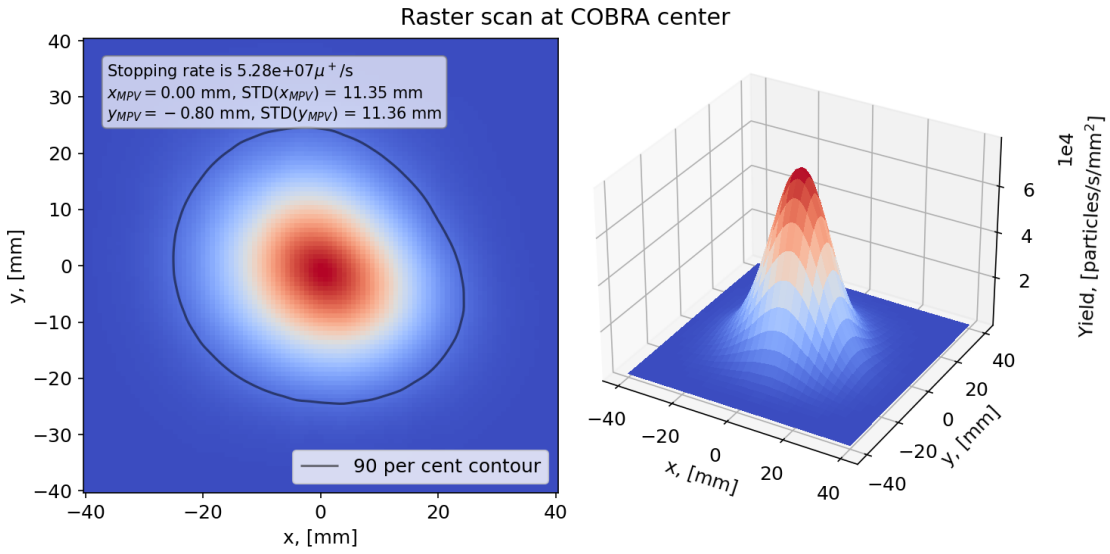


Figure 3.5: Beam spot in 2021 with beam intensity of $5 \cdot 10^7$ Hz.

3.5 Stopping Target

The MEG II stopping target is made of a thin scintillating plastic film with average thickness of 174 μm ranging from 155-194 μm and elliptical in shape, with width 270 mm and height 66 mm (shown in Figure 3.6). It is supported between two hollow carbon-fiber box frames and allowed to float to avoid stress on the target due to dimensional changes in the frames or foil. Ideally, this avoids deformations in the foil shape.[33]

The target is rotated $\sim 85^{\circ}$ with respect to the beam axis (muons along the beam axis intersect $\sim 700\mu m$ material) in order to maximize the stopping rate while minimizing the amount of material in the area.

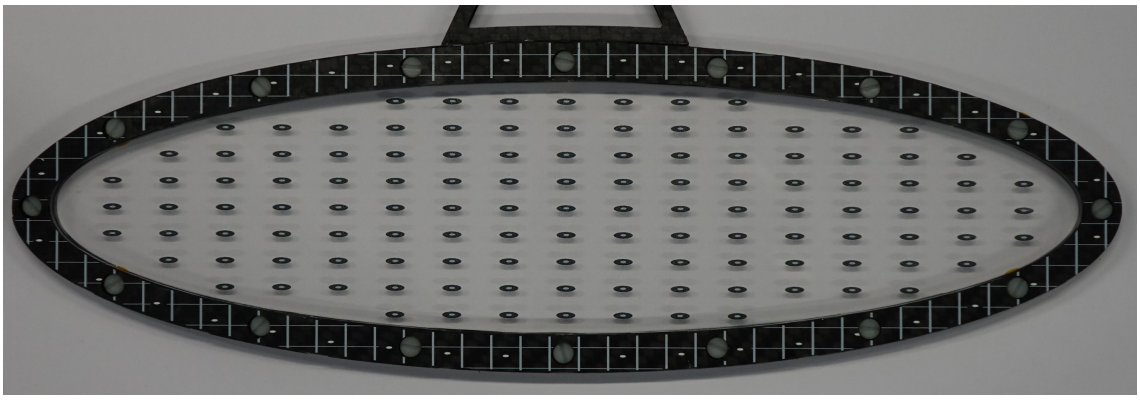


Figure 3.6: Head-on image of the target.

Precise knowledge of the position, orientation, and shape of the stopping target with respect to the magnetic spectrometer is required in order to have a high quality kinematic measurement of the positron tracks at the target (in particular the ϕ_e measurement, described below). Experience with a similar target in the MEG I experiment[5] showed that the target shape changed over a period of ~ 1 year of operation, developing a bowing with maximum deviation from the plane of approximately 1 mm. The time dependence of the bowing was not well monitored. Possible time dependent target motion might also result from the periodic pneumatically actuated extraction and insertion of both the MEG I and MEG II targets; this is done for the purpose of acquiring special data used to calibrate the detectors described

in Chapter 6. These two time-dependent effects motivate the requirement to monitor the position, shape and orientation of the target continuously[33].

The e^+ direction is determined by projecting the helical trajectory measured in the spectrometer to the target plane. An error in the position of the target in the direction normal to the target plane would result in an error in the e^+ direction due to the incorrect path length to the target and hence an incorrect amount of curvature. The precision with which the target position must be measured is set by the requirement that the impact of any error on the e^+ direction be less than that of other contributions to the error in the relative $e^+\gamma$ angle. For MEG II, this angle is measured with a precision of ~ 6 mrad and the goal is for the uncertainty in the $e^+\gamma$ angle due to uncertainty in the target plane position to be < 0.6 mrad. This corresponds to an error in the target position or shape of 85 μm normal to the target surface, giving a path length error of 120 μm for a 53 MeV/c e^+ incident at 45° with respect to the film's surface in a 1.3 T magnetic field. The consequence of a target position error on the positron track's ϕ_e kinematic measurement is shown in Figure 3.7.

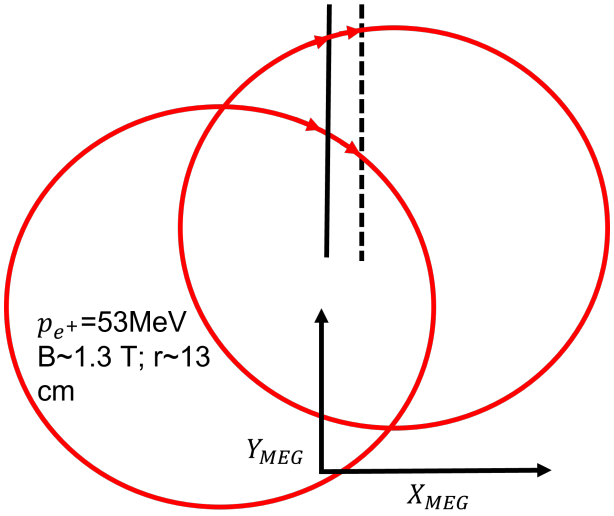


Figure 3.7: Two positron tracks intersecting the target at different ϕ_e values. The solid/dotted lines represent the true/nominal survey target position respectively, clearly this results in a systematic ϕ_e error dependent on the positron ϕ_e.

3.5.1 Analysis Description

In MEG II, the stopping target position, rotation, and shape is measured continuously by a pair of photographic cameras. The details of the camera analysis is described in Chapter 4 and here[33]. The other camera analysis is described here[34].

3.6 Drift Chamber: CDCH

3.6.1 Design Motivation

MEG II uses a new lightweight cylindrical stereo wire drift chamber (CDCH) with full ϕ coverage. The design of the MEG II drift chamber overcomes the efficiency and resolution shortcomings of MEG I. In MEG I there was a significant number of positron tracks that would intersect electronics and other material prior to intersecting the timing counter; this resulted in a significant degradation of tracking efficiency. Therefore a key component of the MEG II design was to reduce the material to improve both the tracking efficiency and the momentum resolution. Further, the MEG II drift cells use an "open cell" geometry i.e. the cell's electric field is generated using cathode wires. This further suppresses the material and thus the scattering. Similar to MEG I, the CDCH starts at large radius ($\sim$ 19 cm) to suppress the material and avoid low momentum positron interactions; the low momentum positrons are swept out of the area along the beam axis by the COBRA magnet.

In addition, the MEG I drift chamber consisted of 16 modules each with two cell layers; the MEG II drift chamber design contains $\sim$ 1000 readout drift cells and thus there are significantly more track space points per track and thus an improved resolution.

3.6.2 Basic Principal

A drift chamber detects charged particles using the fact that they can ionize gas atoms they intersect. Drift chambers are typically filled with a noble gas (no valence electrons) and thus have a high ionization energy and a long radiation length (e.g. Helium $X_0 = 5300$ m), this helps avoid multiple scattering contributions to the momentum resolution. The ionized electrons are accelerated towards a sense wire using an electric field incorporated in the cell. The time of the ionization electrons arriving at the sense wire is recorded using electronics connected to both wire ends. The time of the ionization electrons is converted into a 3D position using the position and orientation of the wire and drift velocity of the electrons at all points in the cell. By measuring the track position in many cells inside a magnetic field, the track's radius of curvature and thus its momentum is estimated.

3.6.3 CDCH Geometry

The CDCH consists of 9 radial layers of drift cells, each with an alternating stereo angle of $\sim \pm 7°$. Each layer has a total of 192 drift cells where only 128 were readout in 2021 (a majority of the other cells are not intersected by $\mu \to e\gamma$ events due to the acceptance of the LXe and the SPX). The CDCH is shown in Figure 3.8.

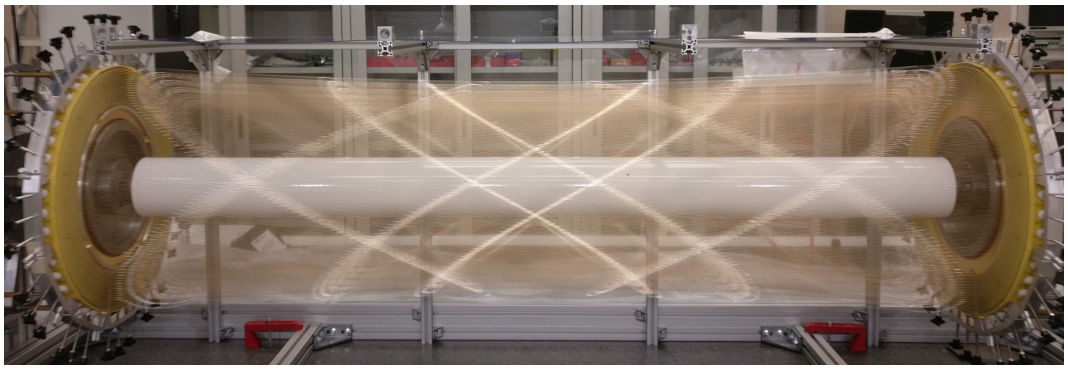


Figure 3.8: Picture of the MEG II cylindrical stereo drift chamber.

The electric field inside the drift cells is created by a combination of high voltage cathode

and guard wires surrounding the sense wires. The high voltage wires are typically operated at $\sim 17V$. A slice of the CDCH wires at the center of the chamber is shown in Figure 3.9; the cell diameter ranges from $\sim 6-9$ mm. The sense/field wires are 20 μm diameter gold plated W/40 μm diameter silver plated Al respectively. Note that O(100) field wires broke during the chamber construction due to chemical corrosion from humidity exposure[35]. Based on Monte Carlo studies, this is expected to have a negligible effect on the experiment, however as a precautionary it also prompted the construction of a new chamber using new wire material (CDCH II). The plan is to potentially use the CDCH II in 2024, but this has yet to be decided.

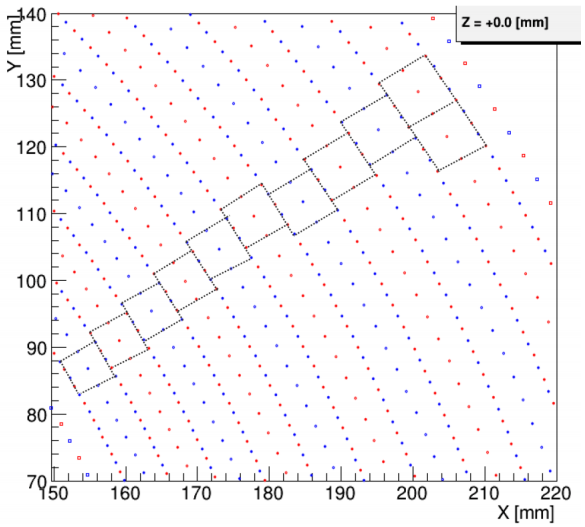


Figure 3.9: A CDCH slice. The drawn boxes represent drift cells created by the high voltage wires surrounding the sense wires.

As mentioned above, in MEG II the total material the positrons intersect is highly suppressed; this improves the positron efficiency and resolution. In MEG II, the material the positron tracks intersect consists of the target region (filled with helium), a thin inner Al Mylar foil ($r \sim 17$ cm), the CDCH gas mixture (described below), potentially the sense/field/guard wires, and an outer shell ($r \sim 28.53$ cm). The outer shell consists of a 100 μm Aluminum foil and then a ~ 1.2 mm outer carbon fiber shell. The construction procedure is described in detail here[7].

3.6.4 Gas Mixture

The chamber is nominally filled with a 90/10 helium isobutane mixture; the helium is selected for a small Coulomb scattering contribution to the positron momentum resolution. The isobutane is added in order to quench self-sustained discharge. Typically a positron track ionizes ~ 15 Helium atoms. This ionization results in ionization "clusters" of electrons (1 or sometimes 2 electrons are ionized) in a cell. These electrons maintain a relatively constant drift velocity until they get close to the sense wire. The avalanche process near the sense wire has a typical gas gain of $\sim 1-5\cdot 10^5$. The gas gain follows a Polya theta function (Figure 3.10) where n and $\bar{n}$ are the number electrons and the average number of electrons respectively. Typically drift chambers with a $O(10^5)$ gas gain have a $\theta \sim 0.4$[6], however the exact value is difficult to measure. As will be discussed in Chapter 5.2, this non-zero probability of very low gas gain results in difficulty detecting isolated ionization sites (particularly near the wire).

$$P(n) = \frac{1}{\bar{n}} \frac{(\theta+1)^{\theta+1}}{\Gamma(\theta+1)} (\frac{n}{\bar{n}})^{\theta} e^{-(\theta+1)n/\bar{n}} \tag{3.3}$$

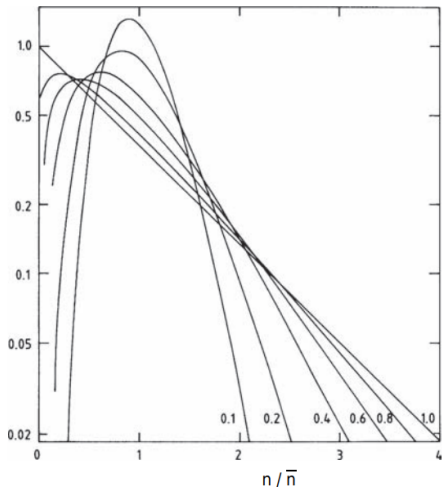


Figure 3.10: Poly theta function as a function of $f = (\theta + 1)/1$ (Figure from 2008 Springer [6]).

In the 2020 engineering run, the CDCH experiences current spikes, an example is shown

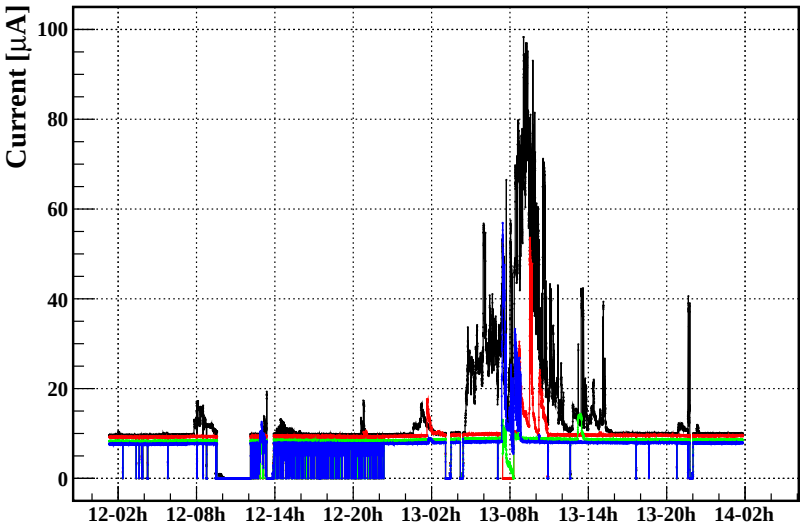


Figure 3.11: A spike in the current on the sense wires. Each curve represents the current on a set of sequential 16 sense wires.

in Figure 3.11. These currents were mitigated by including a small amount of isopropanol alcohol and oxygen. However, oxygen was found to lower the number of drift electrons that arrived on the sense wires, which could be a consequence of both lowering the gas gain and capturing single drift electrons prior to the avalanche. Therefore, an optimization was done to lower the oxygen while maintaining enough oxygen to operate without high currents. Figure 3.12 shows an example of current in a sector of 16 sense wires moving from $2\% \rightarrow 1\%, 1\% \rightarrow 0.5\%, 0.5\% \rightarrow 0.1\%$ oxygen[12]. Lowering or removing the isopropanol again resulted in current spikes as well. In the 2021-2023 physics runs, the CDCH was operated in stable conditions with $He : C_4H_{10} : C_3H_8O : O_2$ (88.2:9.8:1.5:0.5) with minimal current spikes. This gas mixture results in less signal on the wires (with respect to no additives), thus suppressing the signal/noise ratio on the wires. This makes it difficult to detect low gain ionization clusters; this is discussed in detail in the CDCH analysis sections.

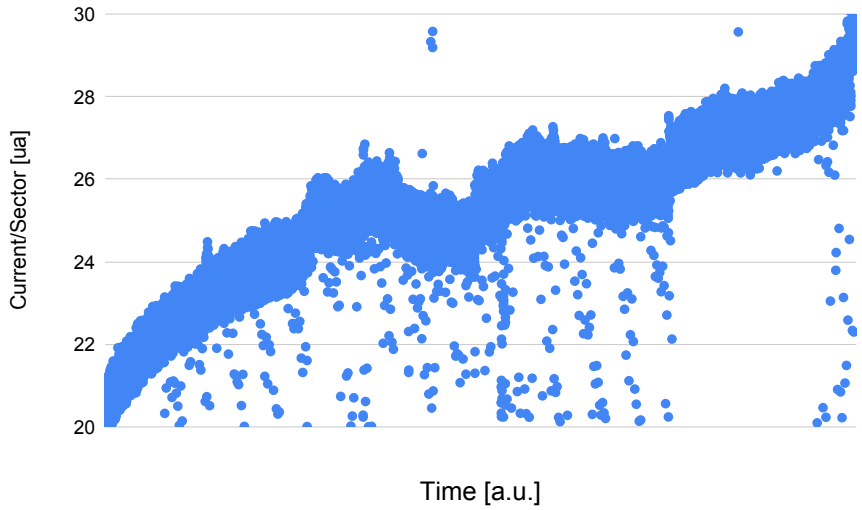


Figure 3.12: The current moving from $2\% \rightarrow 1\%, 1\% \rightarrow 0.5\%, 0.5\% \rightarrow 0.1\%$ oxygen. The solid lines represent a separation in time between the three data samples.

3.6.5 Analysis Description

The drift chamber analysis is described in detail in Chapter 5.

3.7 Timing Counter: SPX

3.7.1 Design Motivation

The core of the detector upgrade is to have a significantly higher hit multiplicity to achieve an improved positron timing resolution. This is a completely redesigned detector with respect to its MEG I counterpart.

3.7.2 Basic Principal

The active detector component of the SPX is scintillator; the positrons intersect the scintillator and interact with electrons in the scintillator atoms, exciting them into higher states. The electrons drop back to lower energy states, releasing photons.

These photons are then captured by Silicon Photo-Multipliers (SiPMs). This detector design is used to achieve high timing precision. One draw back is that the positrons lose significant energy when intersecting the scintillating material, thus making tracking post-intersection difficult. In our case, the particle tracking is done by the drift chamber before the scintillator is reached.

3.7.3 Detector Design

The SPX consists of two semi-cylindrical modules placed upstream and downstream of the stopping target. The detector is shown in Figure 3.13[36]. The modules are installed at the outer radius of the CDCH centered each starting at $|z| \sim 23$ cm and extending to $|z| \sim 117$ cm. The modules have nearly π coverage in ϕ from $165.8° < \phi < 5.2°$ (centered below the CDCH).

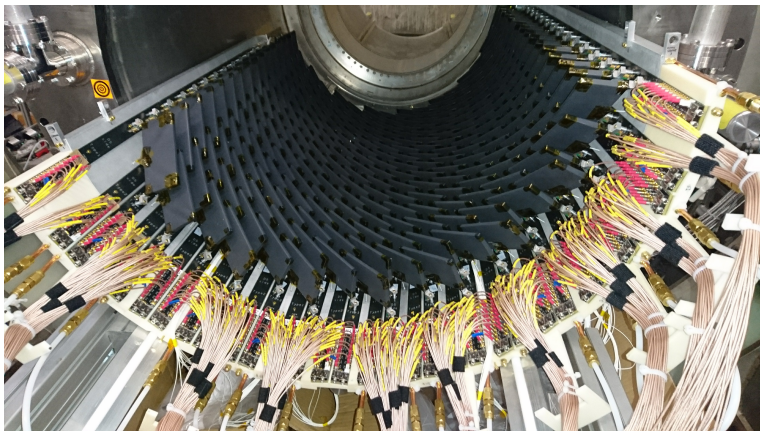


Figure 3.13: The downstream SPX module.

Each module consists of 256 timing counter tiles, each with a hit timing resolution of ~ 90 ps. The timing counter tiles consist of a piece of plastic scintillator (L x W x T $= 120 \cdot (40$ or $50) \cdot 5mm^3$) with 6 SiPMs on each of the sides of the scintillator. An example pixelated timing counter is shown in Figure 3.14. The tiles are tilted to be typically interested by a signal positron perpendicular to its surface.

Figure 3.14: A single pixelated timing counter tile.

In contrast with MEG I, on average, the signal positron tracks intersect 9.3 timing counter tiles. Therefore achieving a timing resolution of $\sim 90/\sqrt{9} = 30$ ps.

3.7.4 Analysis Description

The analysis is described in detail in Miki Nishimura's thesis[37]. Relevant kinematic resolutions for the 2021 physics analysis are described in Chapter 8.

3.8 Liquid Xenon Calorimeter: LXe

3.8.1 Design Motivation

The MEG I liquid xenon detector is one of the world's largest liquid xenon detectors. The choice of liquid xenon was made due to the advantages of a short radiation length and large density including a high photon yield and a short response time.

The MEG I LXe performance was degraded due to non-uniform covering of the liquid xenon on the inner face (incoming photon direction). The PMTs surrounding the liquid xenon detector had gaps in the active area of the order $\sim$ 15 mm due to the PMT geometry. Therefore, there were instances when the photons, particularly on the inner face, missed the active PMT area. This in turn degraded the energy and position resolution. The objective of the MEG II upgrade was to remove this non-uniformity by replacing the PMTs with SiPMs on the inner face.

The design aims to significantly improve the energy resolution for all classes of events ($\sim$ $2\% \rightarrow 1\%$) and improve the position resolution near the inner-face ($\sim$ 5 mm $\rightarrow$ 2 mm).

In addition, this upgraded inner face can be used to aid in pileup discrimination.

3.8.2 Basic Principal

The core of the detector is the scintillating property of liquid xenon. Photons enter the single volume of liquid xenon and interact with electrons in the xenon atoms. This produces a shower of particles as the photon moves through the xenon.

Some photons excite electrons in Xenon atoms from lower orbitals to higher orbitals (leaving a "gap"). This gap is eventually filled resulting in the emission of a photon. In addition,

electrons ionize and interact with other xenon atoms, which can then excite other electrons, etc. The signal photon eventually loses all its energy through interactions with the liquid xenon.

The liquid xenon is surrounded by an array of Silicon Photo-Multipliers (SiPMs) and photo multiplier tubes (PMTs), which collect the photons. By collecting all the energy collected on the SiPMs and PMTs, the energy of the signal photon is estimated. By measuring the average position and time of the detected photons, the position and time of the signal photon at the calorimeter are also estimated.

3.8.3 Detector Design

The liquid xenon detector is a homogeneous single volume vat containing $\sim$ 800 L of liquid xenon. The detector is shown in Figure 3.15. The LXe is 'C' shaped with an inner radius starts at the outer radius of the COBRA magnet $\sim$ 60 cm extending radially outward another 38.5 cm. The detector inner face extends vertically $\sim \pm 68$ cm and $\sim \pm 24$ cm along the beam axis.

As mentioned above, the outer faces (5) are covered by a total of 668 PMTs (Hamamatsu R9869). The PMTs are VUV-sensitive with a quantum efficiency of $\sim$ 16%. These are more or less consistent with the MEG I design with an upgraded support structure.

The inner-face is now covered by 4096 SiPMs called Multi-Pixel Photon Counters (MPPCs). The MPPCs have an active area of 12x12 mm^2 with a total area of 15x15 mm^2. This allows for the inner face to have significantly more uniform coverage with gaps of the order $\sim$ 6mm between active regions. These MPPCs have no protection coating, allowing for VUV detection (light expected from the LXe). Figure 3.16 depicts an MPPC and its active region.

Beyond simply improved spatial resolution, the inner-face upgrade results in improved pileup

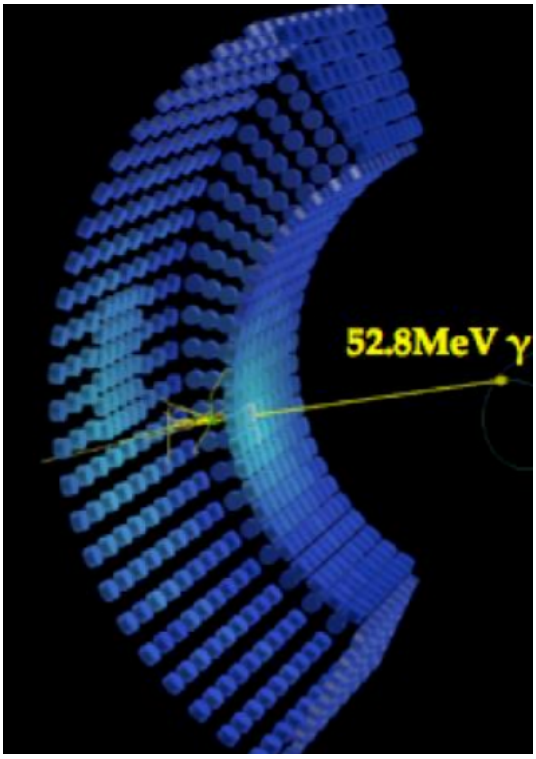


Figure 3.15: The liquid xenon detector.

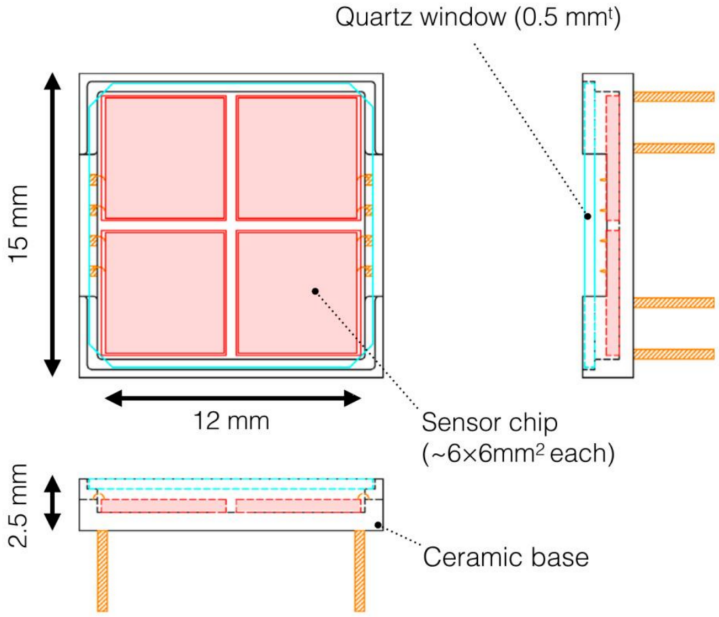


Figure 3.16: A schematic of an MPPC.

discrimination and thus an improvement in all kinematics. An example Monte Carlo pileup event simulated in the MEG I and MEG II LXe detectors is shown in Figure 3.17. It's clear

that in MEG I, it would be very difficult to determine that this is a pileup event, but in MEG II the energy clearly comes from more than one photon.

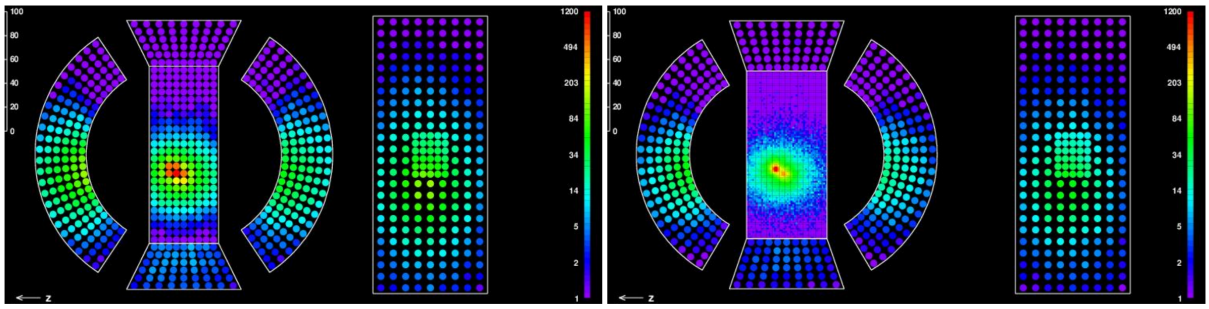


Figure 3.17: The liquid xenon detector in MEG I (left) and MEG II (right). This is a MC simulation of a two-photon event. The two photons are clearly distinguished only in MEG II.

3.8.4 Annealing

During the 2021 physics run, it was identified that the LXe MPPCs were losing significant quantum efficiency as a function of beam exposure. This is shown in Figure 3.18[38]. It is expected to be a consequence of removing the protecting coating on the MPPCs. This loss of gain was a $\sim 10\%$ effect at the end of the 2021 run, but the 2022 run was significantly longer. After the 2021 run, annealing options were investigated. The best annealing procedure was running the MPPCs at high current to heat up the MPPCs to recover the quantum efficiency (i.e. the Joule Method). The annealing resulted in comparable quantum efficiency before and after the run. This was then repeated at the end of 2022.

3.8.5 Analysis Description

The analysis is described in detail in Satoru Kobayashi's thesis[39] and Shinji Ogawa's thesis[40]. Relevant kinematic resolutions for the 2021 physics analysis are described in Chapter 8.

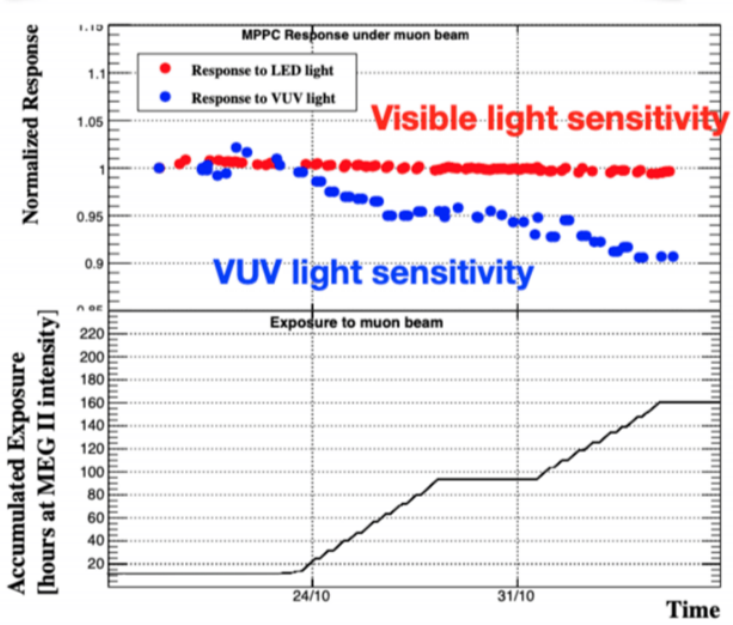


Figure 3.18: The quantum efficiency of the MPPCs as a function of time (and thus beam exposure) during the 2021 physics run.

3.9 Radiative Decay Counter: RDC

3.9.1 Design Motivation

Due to the significant reduction in the CDCH material, the number of accidental time coincidences between a high energy Michel positron and a high energy photon produced from AIF was significantly suppressed with respect to MEG I ($\sim 1/2 \rightarrow 1/3$ of all accidental time coincidences). However, the other dominant source of accidental background, an accidental time coincidence between a high energy Michel positron in the CDCH and a high energy photon from an RMD muon decay, was not lowered (except for a narrower signal region due to the resolution improvement).

In these accidental RMD events, there is a low momentum positron that doesn't enter the CDCH, but exits upstream/downstream via the magnetic field. By adding an additional detector downstream of the target, a significant fraction of these RMD accidentals (those

where the RMD positron travels downstream) can be eliminated from the dataset using the relative time of the photon in the calorimeter and the positron in this detector. This detector is known as the Radiative Decay Counter (RDC).

3.9.2 Basic Principal

The detector uses fast scintillating counters with SiPMs on the two ends (similar to the pixelated timing counter used in the SPX detector) to estimate the time of the low energy positrons. The positron energy is then absorbed in an array of LYSO crystals. The positrons create scintillation light, which is then absorbed by SiPMs attached to the LYSO crystals.

3.9.3 Detector Design

The detector is situated $\sim$ 1 m downstream of the target along the beam axis. The detector is roughly $\sim$ 22x22 mm^2. As mentioned above, the detector consists of a first layer of scintillating tiles to estimate the positron timing and a second layer of LYSO crystals to estimate the positron momentum. Figures 3.19 and 3.20 show the scintillating tile layer and the LYSO crystal layer respectively.

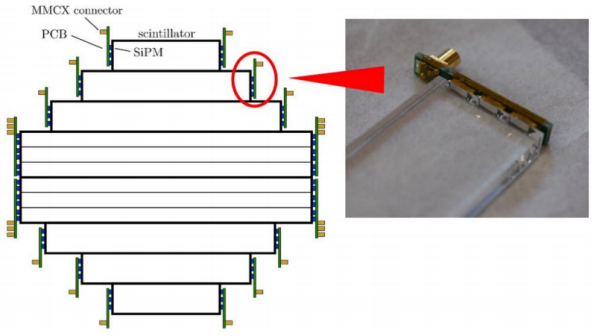


Figure 3.19: The RDC scintillating tile array.

Clearly, the RDC counter will also be crossed by Michel positrons that are accidentally intersected (a random positron). This is a "background" in the RDC counter detection

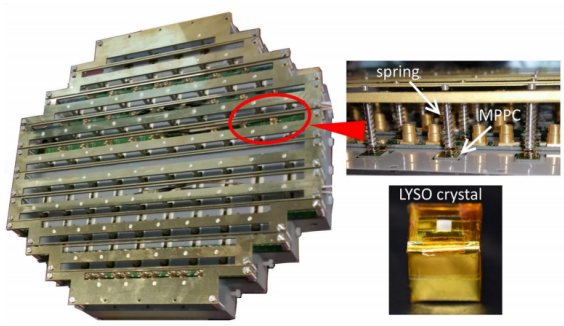


Figure 3.20: The RDC LYSO crystal array.

(not to be confused with the physics background). These will have no preferential time coincidence with the LXe and the RDC and they will be dominated by higher momentum positrons as RMD events result in a significantly lower momentum positron on average.

Eliminating events based on this accidental time-coincidence between the LXe and the RDC (non-true RMD events) would eliminate potential signal events and thus degrade the sensitivity of the experiment. Therefore, we additionally use the RDC positron momentum to optimally determine if the RDC intersection is a true RMD event. In Figure 3.21, we show the expected Monte Carlo momentum for true RMD positrons and accidental positrons at the RDC counter.

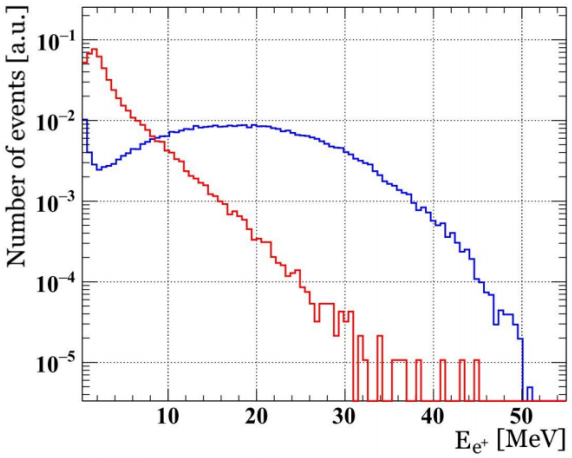


Figure 3.21: The energy expected from Michel and RMD positrons in the RDC detector.

The RDC was tested in a pilot run using a different photon detector to mimic the timing of

the LXe (BGO, described in Chapter 7). In this run, the relative time of the photon in the BGO and the positron in the RDC is measured. In Figure 3.22, we show the relative time distribution between the positron and the photon with and without a cut on the positron momentum in the RDC (< 5 MeV). It's clear that after the positron momentum cut, a vast majority of the accidental time-coincident RDC positrons are eliminated resulting in almost exclusively true RMD events. Eliminating events based on this relative time coincidence and the estimated energy, we expect to eliminate a significant fraction of RMD accidentals at the expense of a small amount of signal efficiency. This is expected to improve the sensitivity by $\sim 15\%$[7].

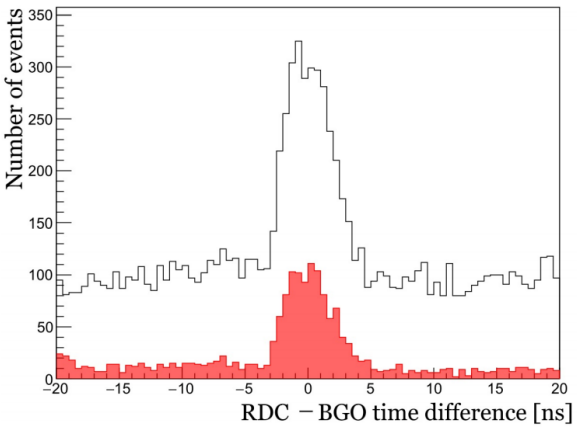


Figure 3.22: The relative time of a positron in the RDC and the time of a photon in the BGO. The top histogram contains no cut on the positron momentum and the bottom requires the positron momentum to be less than 5 MeV.

3.9.4 Analysis Description

The analysis is described in detail in Rina Onda's thesis[41]. Relevant probability density functions required for the 2021 physics analysis are described in Chapter 8.

3.10 Electronics and Trigger

All detectors modules are connected to custom Waveform Domino REAdout Module (WaveDREAM) electronics boards[42]. This includes all O(10k) channels on the CDCH wires on both ends, LXe MPPCs and PMTs, SPX SiPMs, and RDC SiPMs. Each channel contains 1024 'sample-and-hold' cells that sample and temporarily hold signals from a particular detector module. The frequency of sampling is $\sim$ 1.2 GHz, thus resulting in a total recorded time of $\sim$ 850 ns. The signal (charge) in each of the 1024 cells is digitized after receiving a trigger. This results in a voltage and a time associated with each cell (bin). The circuit diagram is shown in Figure 3.23 and a WaveDREAM board is shown in Figure 3.24. Each WaveDREAM board contains two Domino Ring Sampler (DRS4) chips; each DRS4 chip contains 8 channels.

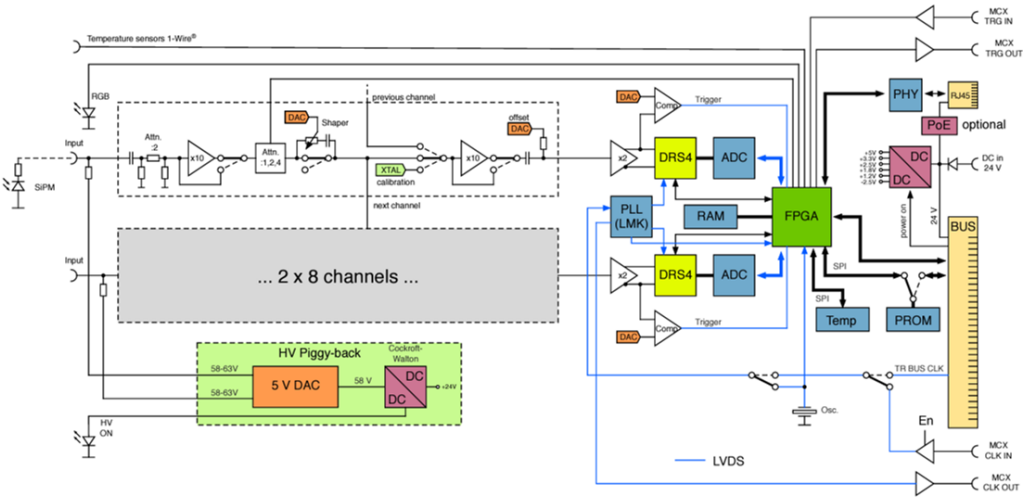


Figure 3.23: Circuit diagram of the MEG II WaveDREAM boards.

The MEG physics trigger[43] consists of three main components in order to suppress the number of background triggers, they are listed below.

- Energy threshold in the LXe ($E_{Threshold} < E_{\gamma}$), where $E_{Threshold} \sim 45$ MeV.
- Time match between the intersected pixelated timing counter and the LXe detector

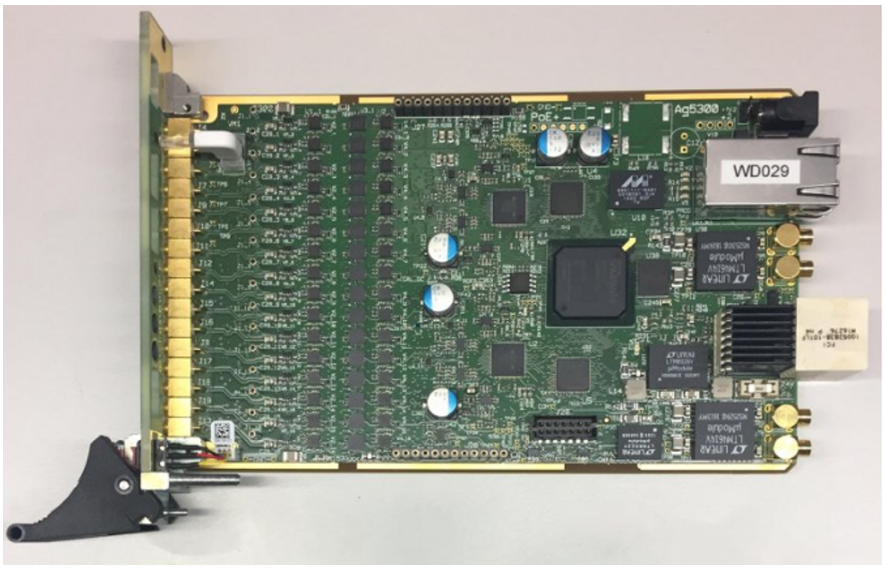


Figure 3.24: Graphic of the MEG II WaveDREAM boards.

($|T_{e\gamma}| < T_{Threshold}$), where $T_{Threshold} \sim 12$ ns

- Spatial match between the intersected pixelated timing counter and the LXe detector region.

The energy threshold suppresses the number of low energy photon triggers that would otherwise overwhelm the trigger rate. The energy threshold requires an online energy reconstruction estimate, which also requires time-dependent energy calibrations. The online energy resolution is $\sim$ 2.4%. It is not possible to have a very precise $t_{e\gamma}$ measurement as there is no online positron track reconstruction, only a hit in the pixelated timing counter. As mentioned, the positron track can intersect the drift chamber 3,5,or 7 times before hitting the timing counter, where each turn (2 chamber intersections) typically takes $\sim$ 2 ns; the number of turns is not known in the online analysis and thus forces a rather wide criteria on the relative timing of the photon at the LXe and the positron at the timing counter. The wide $t_{e\gamma}$ region is also used in the physics analysis to study accidental events away from the signal. Finally, the spatial match between the positron and the photon position is calculated from a Monte Carlo simulation by simulating $\mu^+ \to e^+\gamma$ events. At minimum,

one can see that a roughly a factor of two can be achieved through a spatial match as a true signal positron intersecting the upstream timing counter would require a signal photon in the downstream region to conserve momentum. One caveat is that this is not the case if the positron is close to $\theta_e \sim 0$, where the positron can turn around due to the magnetic field.

Combining the three trigger elements results in a trigger rate of ~ 12 Hz at standard beam intensity. All three trigger elements require online reconstruction that requires CPU time, this must be complete before the sample and hold cells containing the signal we would like to trigger on are replaced by new signals (this signal is constantly being replaced until the online reconstruction is done and the trigger is sent).

3.10.1 Performance Description

The performance of the trigger is described in detail here[44] and here[13].

Chapter 4

MEG II Stopping Target Analysis

4.1 Introduction

In this chapter, we describe a novel photographic technique to increase the precision of the e^+ kinematic measurements at stopping target by precisely measuring the position, orientation, and shape of the stopping target with respect to the magnetic spectrometer. Much of this information is described in the paper[33]; some details are only described in this thesis.

The technique described monitors continuously the *change* in the target's shape and in its position and orientation with respect to the camera. It does this by imaging approximately 120 dots printed on the stopping target using a camera located about 1.2 m from the target. This technique does not measure the absolute position of the target. To determine the absolute position and orientation of the target with respect to the magnetic spectrometer, MEG II uses two independent techniques described later.

The first technique is an optical survey of the target with respect to the magnetic spectrometer, done only at the beginning of every run. The correlation with the photographic

results is done by analyzing a sequence of very precise photographic measurements taken simultaneous with the optical survey. The second technique (described in Appendix A.1) was first used in the MEG I experiment[5]; it uses momentum analyzed positrons recorded during data-taking to image, in 3 dimensions, small holes in the target by detecting a deficit of positrons originating from the position of the holes. The MEG I experiment achieved a precision in the measurement of the target position normal to its surface of $0.3-0.5$ mm. The uncertainty was primarily due to lack of statistics available to measure the time dependence of the position. By correcting for time dependence in the target geometry using the photographic technique, the full data-set of MEG II will be used for measuring relative alignment of the spectrometer and the target. This second technique has an additional advantage that it is independent of the optical survey.

4.2 Methods

4.2.1 Camera Installation and Operation

The implementation of the photographic alignment system is complicated by several factors. Access is limited since the stopping target is at the center of a tracking detector ~ 2 m long in a solenoid with nominal field of 1.3 T at the target location. No simple optical path from the target to a position outside the solenoid exists. The camera system cannot be closer than ~ 1.2 m from the target, with the camera axis nearly along the magnet axis (Figure 4.1 and 4.2), and it must be located at the incoming muon beam end of the spectrometer. The magnetic field at this location is ~ 0.8 T. Further, there is significant positron flux at the camera location from muon decay in the target. This presents the possibility of radiation damage to the sensor and camera electronics. Finally, the available space for the camera and lens is limited.

Figure 4.1: Drawings of the assembly including including a) the camera used in this technique, b) the stopping target, c) the pneumatically actuated target support, and d) the LED lights. The camera is positioned at the vertical (MEG II y-axis) center of the target, and is rotated to center the target in the image plane. Left: Beam-axis view of assembly. Right: Top view of the assembly, with some support pieces, a second camera not used in this analysis, and cabling omitted for clarity.

The camera is mounted approximately 1.2 m from the target, at the same height as the target, and offset horizontally from the target by approximately 10 cm (Figure 4.1 and 4.2). The camera axis is at an angle of 5.60° with respect to the MEG II beam axis and 69.40° with respect to the vector normal to the target plane. It is mounted to a support structure that is rigidly attached to a spool piece attached to the cryostat of the COBRA magnet[8]. The structure also supports two LED lamps for illuminating the target, another camera not used for the position monitoring described here, and a pneumatically controlled target support that moves the target between the inserted position when it is being used and an extracted position during certain calibration data taking. We verified that there was no motion of the target with respect to the spectrometer by imaging and monitoring the position of a flange on the spectrometer structure.

In 2021, the alignment system was implemented with an industrial camera[45] with a 1/2.3-inch CMOS sensor with 3856×2764 pixels, each 1.67 μm square, and a 50 mm lens. The interface to an acquisition computer is by USB3. Note, Ethernet interfaces failed in the 0.8 T magnetic field. We use an active USB extender cable to allow the acquisition computer to be at a sufficiently small magnetic field. The manufacturer provides a graphical user interface and provisions for operating the camera with scripts that can be written in C++ or Python. The software provides the capability to set the frame rate, exposure, gain, and

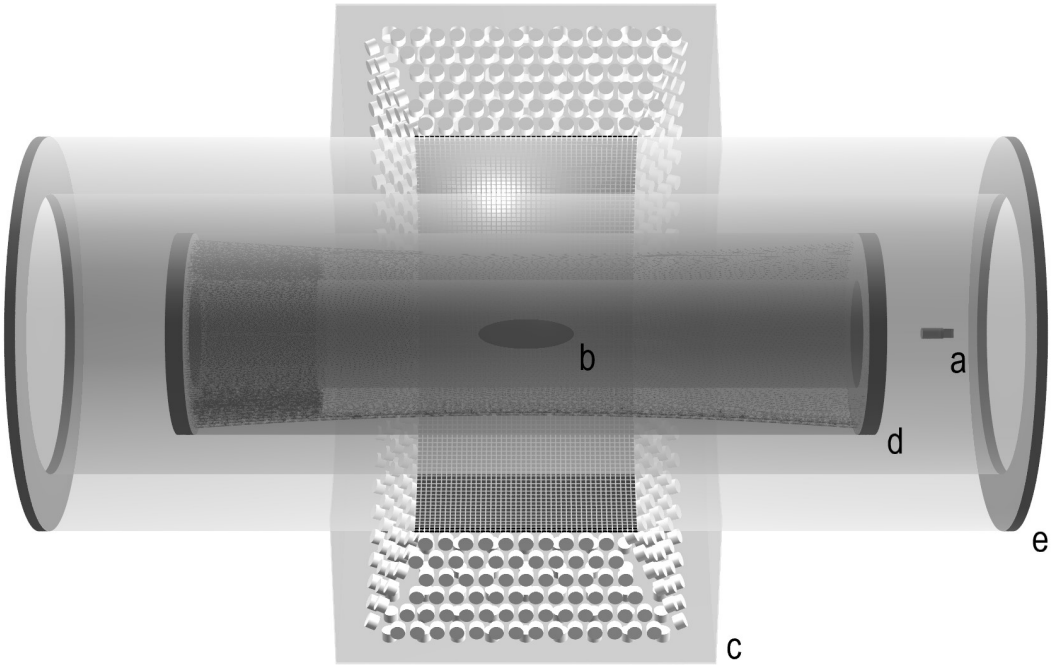


Figure 4.2: A schematic perspective view[7] of relevant parts of the MEG II experiment including a) the camera used in this technique, b) the stopping target, c) the liquid xenon photon detector, d) the drift chamber, e) the COBRA magnet. For clarity, the arrays of scintillation counters at each end, the camera support structure, and the spool piece are not shown. The COBRA magnet is depicted as a single hollow cylindrical shell for clarity; the inner bore is in fact stepped with a smaller diameter in the center[8].

fraction of the image plane to be read out. The focus and aperture are set manually.

The camera control and image acquisition are implemented using Python scripts combined with scripts to control the LED lights and sequencing of image acquisition. All the acquisition and control software is integrated with the MEG II data acquisition and control system. During operation, we acquire a dark field image and a set of three normal images every 15 minutes. The dark-field image is used to subtract the background intensity in 'hot'pixels. The hot pixels remain in fixed position and comprise 0.05% of all pixels. The hot pixels appeared after a couple weeks of operation with the muon beam on; we assume they are due to radiation damage from muon decay positrons.

New cameras were installed prior to both the 2021 and 2022 runs to eliminate any previous damage due to beam exposure. Unfortunately, both cameras failed after 3-5 weeks of beam

exposure, losing connection between the camera and any computer to which it was connected. The operational period accounts for a significant fraction of the short 2021 physics data but only a small fraction of the 2022 data. The apparent failure due to radiation damage from the beam motivated a search for radiation hard cameras for future runs. In the 2023 run, we instrumented a radiation tolerant camera from Spectral Instruments (RVT-100 1700S[46]) equipped with 4MP, a pixel size of 5.5x5.5 μm, and a 50 mm lens. This camera connects directly to a server communicating via fiber optic and removes much of the signal processing to outside the area of high radiation. The camera has now operated successfully during the 2023 physics run for $\sim$ 16 weeks without any issue. Note that this camera contains less pixels and has a wider field of view than that of the original camera and therefore we expect a slight reduction in the resolution, but still beyond that needed by the technique.

4.3 Coordinate Systems

First, we describe the relevant coordinate systems to the camera analysis. The coordinate systems are shown in Figure 4.3. The target frame has X_{TGT}/Y_{TGT} parallel to the target surface where the target frame is nominally rotated by 15 degrees with respect to the MEG frame about the Y axis. The CT scan of the target (described later) contains a coordinate system extremely well aligned with the target frame, the only difference is that the CT scan data is offset by 4.295 mm in Z_{CT}. In all practice, we only use the target frame, simply offsetting the CT scan data by this factor. In addition, the camera analysis relies on the camera frame, which is centered about the camera's effective lens. The Z_{CAM} axis is aligned with the camera direction and the X_{CAM}/Y_{CAM} axes are aligned with the camera's CCD. In the MEG coordinate system, the camera is located nominally at negative X (8 cm), and negative Z (120 cm), and the camera is rotated by 5 degrees about the Y axis with respect to the MEG frame. Therefore, we expect a nominal rotation of (90-20) degrees about the Y

axis between the target and the camera frame. Finally, the CCD frame is a 2D frame on the camera's image plane aligned with the camera's CCD.

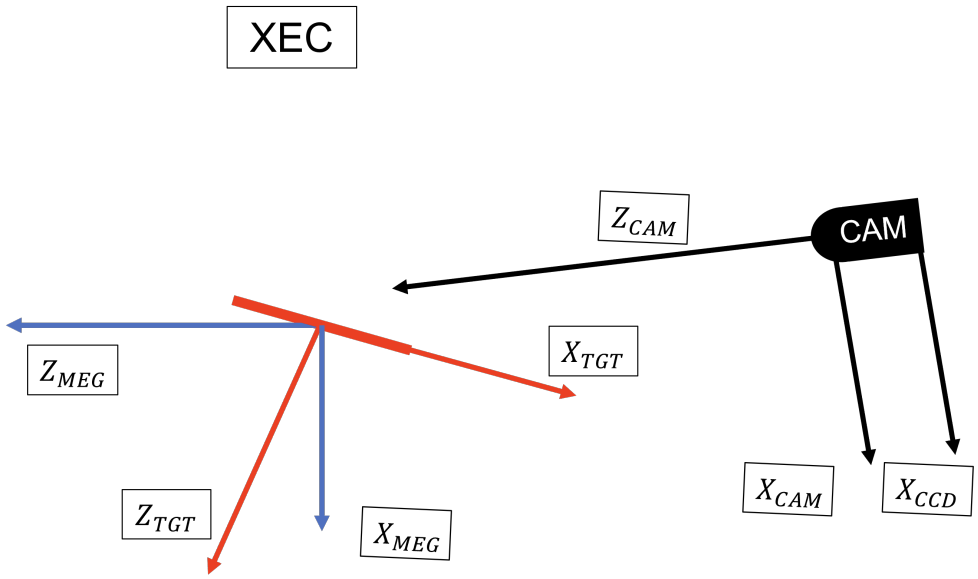


Figure 4.3: Graphic of the relevant coordinate systems.

4.4 Camera Analysis

The general objective of the camera analysis is to minimize a χ^2 function which compares the measured dot coordinates on the 2D image plane (X_{CCD}) to the estimated 3D dot coordinates in the camera frame $\vec{X}_{CAM}$. The 3D coordinates in $\vec{X}_{CAM}$ are projected onto the image plane using optical projection equations so they can be compared with the measured coordinates in X_{CCD}.

4.4.1 Image Dot Finding

The analysis is initiated by searching for the dot coordinates on the image (X_{CCD}). This is done using OpenCV[9] contour analysis. The analysis uses a binary threshold (e.g. 50/255)

to find all contours at that selected threshold on the image. As we are searching for dots on the target foil, we reject contours based on the contour area and aspect ratio. To avoid the cluster's centroid depending on details of pixels association near the cluster's edge, each dot's centroid is calculated using an intensity weighted mean pixel position. This method determines the centroid of each dot with a dispersion of $\sigma = 0.2$ μm at the image plane (approximately 4.8 μm at the target) determined from a series of sequential images taken close in time. We then assign a 2D index to map each selected contour to a unique printed dot based on its 2D position on the image plane. This results in an array of dots with a 2D position on X_{CCD} and an X/Y index (4xN array). If there are two contours within a single dot region (e.g. the dot is broken into two contours due to lighting issues), both contours are removed from the analysis. An example of the contours at an optimal threshold is shown in Figure 4.5. We only use the smaller contours surrounding the inner white dots.

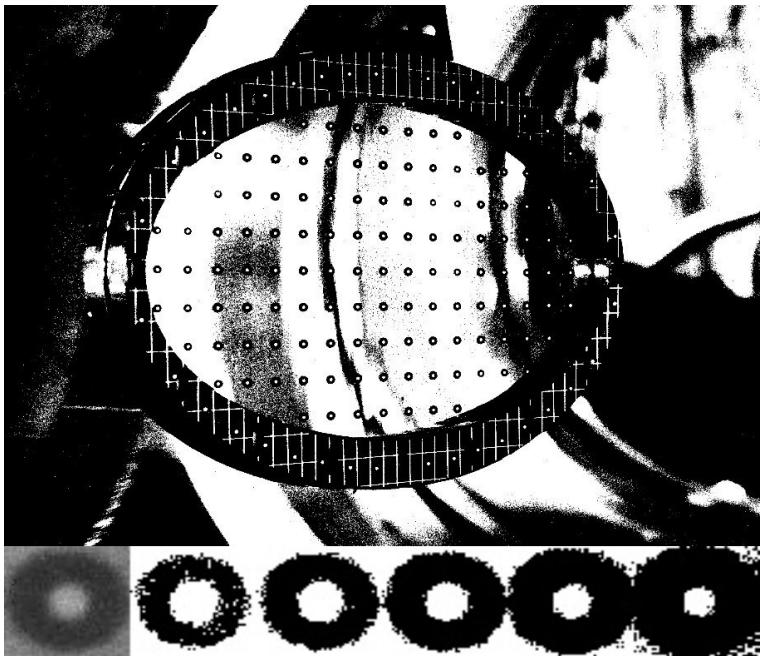


Figure 4.4: Top: A typical binary image of the target[9]. Bottom: From left to right we show an original (non-binary) dot followed by binary versions with increasing threshold parameter values.

We have verified that the lighting intensity and the threshold parameter do not affect the measured positions in a systematic way, although they do affect the cluster size (Figure

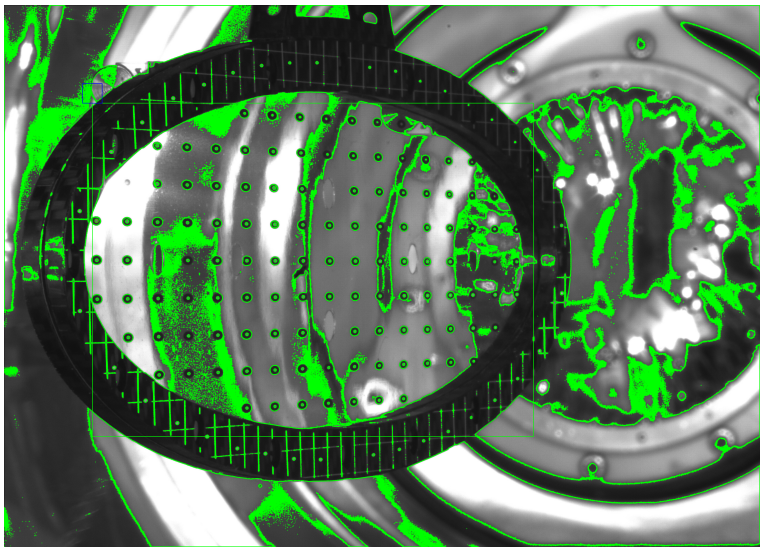


Figure 4.5: The measured contours at an optimal threshold drawn in green.

4.4). The dispersion in the measured dot positions does marginally increase with decreased lighting. Additionally, we analyzed approximately 20 images while varying the threshold parameter; this did not have a systematic effect on the resulting fit for the target shape, position, and orientation (see Appendix A.2).

4.4.2 Dot Coordinates in Target Frame

To estimate 3D dot coordinates in $\vec{X}_{CAM}$, we rotate and translate the 3D dot coordinates in the target coordinate system ($\vec{X}_{TGT}$) into the $\vec{X}_{CAM}$ frame. To measure the 3D dot position in the target frame $\vec{X}_{TGT}$, we use an independent 3D CT scan. This scan was done in the months prior to the 2021 dataset. The CT scan contains several million data points on the target frame and foil. Data points estimate the 3D position on both sides of the target foil. To estimate the dot coordinates in $\vec{X}_{TGT}$, we first remove all CT data points associated with the other side of the film. We then zoomed in on each dot. An example is shown in Figure 4.6. Here, Z_{TGT} or the coordinate normal to the target surface is represented by color. We estimate each dot's 3D coordinate using a python GUI that gives the 3D coordinate at the position of the cursor. We use this technique as it was unclear how to optimize a standard

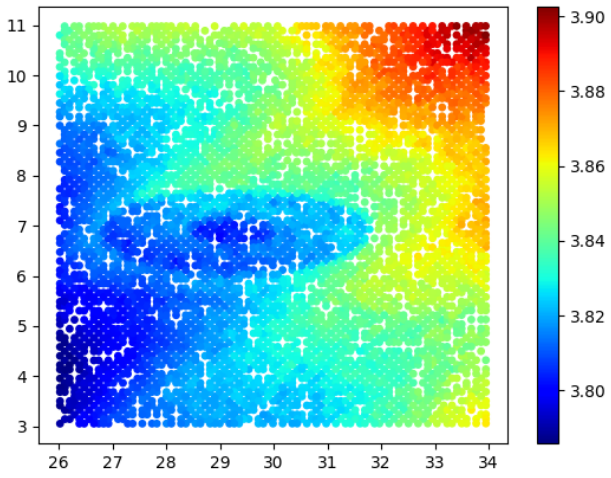


Figure 4.6: An example dot in the CT scan (mm). The color represents depth i.e. deformation.

fitting procedure that works for every dot, whereas by eye the position normal to the surface is well estimated. We estimate that for most dots, our precision normal to the surface is ~ 10 μm. We then average each column's X_{CT} coordinate and each row's Y_{CT} coordinate. We use these CT coordinates as a starting grid in $\vec{X}_{TGT}$. These dot coordinates are indexed, resulting in a 5x117 array of dot coordinates (X/Y index + 3D coordinate); this defines the dots in the target frame. The CT scan dot coordinates are shown in Figure 4.7

4.4.3 Rotation and Translation

We then rotate and translate the dot coordinates from $\vec{X}_{TGT}$ into $\vec{X}_{CAM}$. We use the MEG standard ZYZ rotation convention. We apply the rotation to the object (dots), not the axes so the ordering of ψ, ϕ is flipped with respect to other MEG conventions.

That is:

$$\vec{X}_{CAM} = Z(Y(Z(\vec{X}_{TGT}, \phi), \theta), \psi) + d\vec{X}_{CAM}$$

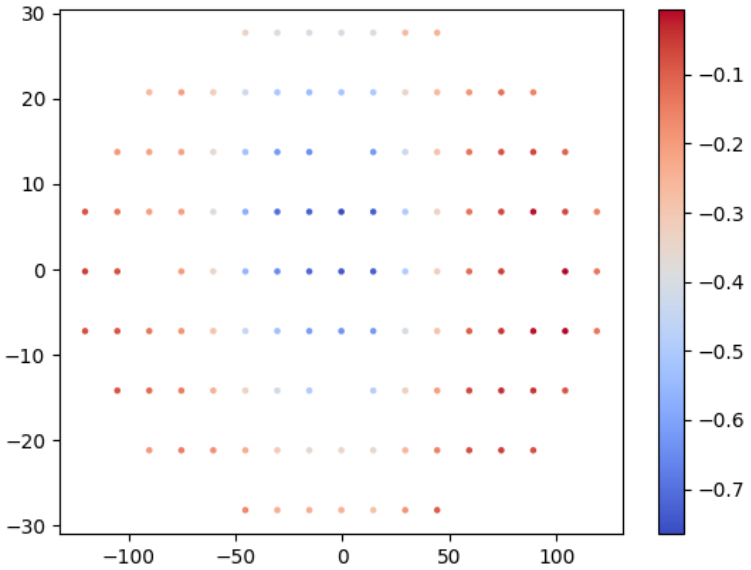


Figure 4.7: The target deformation at the dot coordinates in the CT scan (mm). The full range of the deformation is $\sim 800\mu m$.

Where:

$$Z(\vec{X}, a) = \begin{bmatrix} cos(a) & -sin(a) & 0 \\ sin(a) & cos(a) & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

$$Y(\vec{X}, a) = \begin{bmatrix} cos(a) & 0 & sin(a) \\ 0 & 1 & 0 \\ -sin(a) & 0 & cos(a) \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

4.4.4 Optical Projection

To compare the estimated 3D coordinates in $\vec{X}_{CAM}$ with the measured dots on the camera's CCD, we rely on optical equations to project the 3D coordinates onto the camera's CCD. We use the following optical equations, where f is the camera's focal length and d_I is the distance from the lens to the image plane.

$$\frac{1}{f} = \frac{1}{d_I} + \frac{1}{Z_{CAM}}$$
$$\frac{d_I}{Z_{CAM}} = \frac{X_{CCD}}{X_{CAM}}$$
$$\frac{d_I}{Z_{CAM}} = \frac{Y_{CCD}}{X_{CAM}}$$

Converting that into projection equations and thus eliminating the distance to the image plane, we are left with the following projection equations. Note that although the camera is a complex lens, we are using a simple lens equation. This likely creates some systematic effects. In addition, this projection requires a focal length estimate. We found that the residuals are minimized by using a focal length of 51.2 mm with a nominal focal length from the manufacturer of 50 mm.

$$X_{CCD} = \frac{X_{CAM} \cdot f}{Z_{CAM} - f}$$
$$Y_{CCD} = \frac{Y_{CAM} \cdot f}{Z_{CAM} - f}$$

4.4.5 Minimization

The χ^2 is listed below where the X'_{CCD}/Y'_{CCD} are the measured coordinates on the image plane and X_{CCD}/Y_{CCD} are the projected coordinates from the above equations:

$$\chi^2 = \sum_i^N (X_{i,CCD} - X'_{i,CCD})^2 + (Y_{i,CCD} - Y'_{i,CCD})^2 \tag{4.1}$$

We minimize the χ^2 in the python Scipy[47] package using the Nelder-Mead minimization routine. Below, we show the 2D residuals after the minimization. The residuals are scaled

using the projection equation. By scaling these residuals to the object, we are assuming that they represent an error in X_{CAM}/Y_{CAM}.

In general, we observe minimal residuals in Y_{CCD}/Y_{CAM}. This is expected as there is not a physical explanation to these residuals. In contrast, we observe systematic residuals in X_{CCD}/X_{CAM}. The X_{CAM} coordinate is nominally aligned with Z_{TGT}, so a systematic X_{CAM} residual can correspond to a deformation of the target with respect to the deformation at the time of the CT scan (Z_{TGT}). We see a $\sim$ 50 μm deformation around the target holes centered at index (2, 4).

Originally, we observed some systematic errors in the residuals column-by-column or row-by-row i.e. each dot in a single row/column had a mean residual of 30 μm (at the object) for all images taken near the survey. This indicates that our estimate of the dot's coordinate parallel to the target surface contained a systematic error (CT scan estimate). Since it is unlikely to get a residual exclusively for every dot in a single column/row (and no other nearby dots), we shift the CT scan measurement's accordingly.

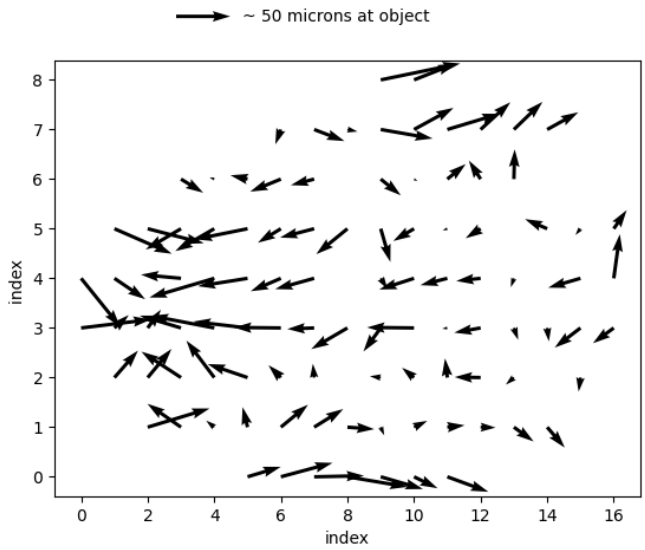


Figure 4.8: Example 2D residuals for an image taken at the time of the survey. The residuals are scaled using the projection equation to show the error size at the object.

The minimization fit results in the 6-parameter rigid body rotation and translation to move

the dot coordinates from the target coordinate system into the camera coordinate system. At this time we do not fit for the deformation, but this is easily incorporated by including additional drum head mode degrees of freedom into the minimization scheme. This was tested by starting with a nominally flat target (setting $Z_{TGT} = 0$ for all dots) and indeed we are able to recover a comparable deformation to that observed in the CT scan. However, at least in 2021 we do not observe large residuals and so we have left the fit without any deformation parameters.

4.5 Transformation Into the MEG Coordinate System

In order to determine the target foil's position in the MEG coordinate system from the optical survey, we require the relative position of the target foil and the corner cubes. We (Yusuke Uchiyama) used the CT scan to extract the position of the 'corner' in each corner cubes in the CT scan (i.e. $\vec{X}_{TGT}$). The optical survey of the corner cubes on the target results in the corner cube positions in $\vec{X_{MEG}}$. We then fit for the $\vec{X}_{TGT} \rightarrow \vec{X}_{MEG}$ transformation (6-parameter rotation/translation). We can then use the same transformation to move the target dots from $\vec{X}_{TGT}$ to $\vec{X}_{MEG}$.

We use images taken on the same day of the survey to estimate target position at the time of the survey in $\vec{X}_{CAM}$. We can then fit for the 6-parameter transformation from $\vec{X}_{CAM}$ to $\vec{X}_{MEG}$. As we have not fit for any additional deformation of the target, this is an exact transformation. This calculates the unique transformation from the camera coordinate system to the MEG coordinate system for all future images in 2021. We calculate the rotation to be (-0.01861, -0.07746, 0.018103) radians and the translation to be (-10.930,-6.497, -1231.13) mm. The two points of interest are the Z translation, which we expect to be approximately 1.23 m from the survey, and the Y axis rotation (second angle) which is expected.

For any future image, we transform the dot coordinates from $\vec{X}_{CAM}$ into $\vec{X}_{MEG}$ using this transformation. That is, camera analysis on future images results in an updated target position in $\vec{X}_{CAM}$, we then apply this same transformation to move the updated target position into $\vec{X}_{MEG}$.

In addition, using the just calculated $\vec{X}_{MEG}$ coordinates of the target, we calculate the transformation: $\vec{X}_{TGT} \rightarrow \vec{X}_{MEG}$. This yields the 6-parameter transformation from the nominal target coordinate system (CT scan) into the MEG coordinate system at any given image. This is used in the MEG analysis code as described later.

4.6 Survey Check

In 2021, the camera and the target were both surveyed. However, the relative position of the camera corner cube and the camera's CCD or the camera's effective lens position have not been estimated. Regardless, we can use the survey as a sanity check of the dot coordinates in $\vec{X}_{CAM}$. In the camera frame using the camera analysis, we observe that the center dot is located at (-11.03, -6.49, -1231.30) or 1231 mm away from the effective lens. The survey suggests that the camera corner cube is $\sim$ 1228 mm away from the target foil center. Since we don't know the distance from the corner cube to the camera's effective lens to this precision, we conclude that the relative position of the camera and the target is consistent in the two techniques. To be clear, this is not critical for the camera analysis as we do not use the camera to estimate the global position of the target, but only to identify the relative movement of the target with respect to a reference point.

4.7 Results

The 6-parameter transformation from the target coordinate system into the MEG coordinate system as a function of time is shown in Figure 4.9.

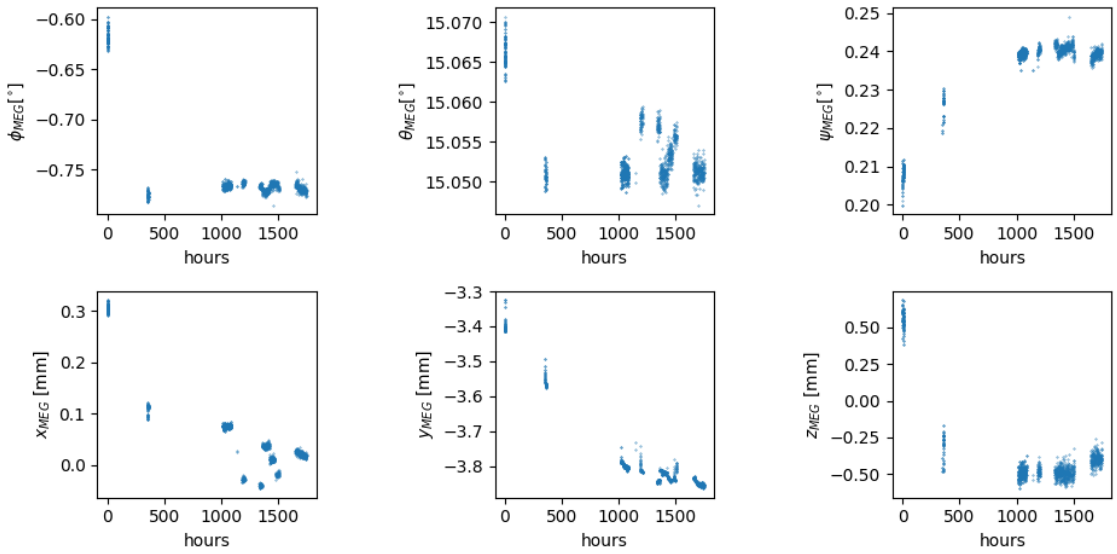


Figure 4.9: The 6-parameter transformation from X_{TGT} into X_{MEG} as a function of time in the 2021 dataset.

There are several interesting features to discuss. First is the scale of the translations. We are mainly interested in the translations normal to the target surface. We observe total shifts of ~ 300 and ~ 1000 μm in X_{MEG} and Z_{MEG} respectively. Both of these move the target in positive Z_{TGT} (normal to the target surface). Approximating the slant angle as 15°, this corresponds to a total movement of $\sim 600\mu$m in Z_{TGT}. To convert this into an error in ϕ for a 53 MeV positron, roughly an error of 140 μm normal to the target corresponds to an error of 1 mrad. Therefore, if there are no corrections with respect to the survey position, we would have a systematic error of 4.3 mrad in ϕ.

Next, we found that there is some systematic motion of the target after an insertion/extraction. We give 3 examples of extraction/insertions in Figure 4.10 where the parameters are normalized such that the parameters are they are equal to zero before the insertion. It's clear that

in all cases, there is a relaxation of the target Y position. This relaxation seems to occur on the time scale of 1-3 hours.

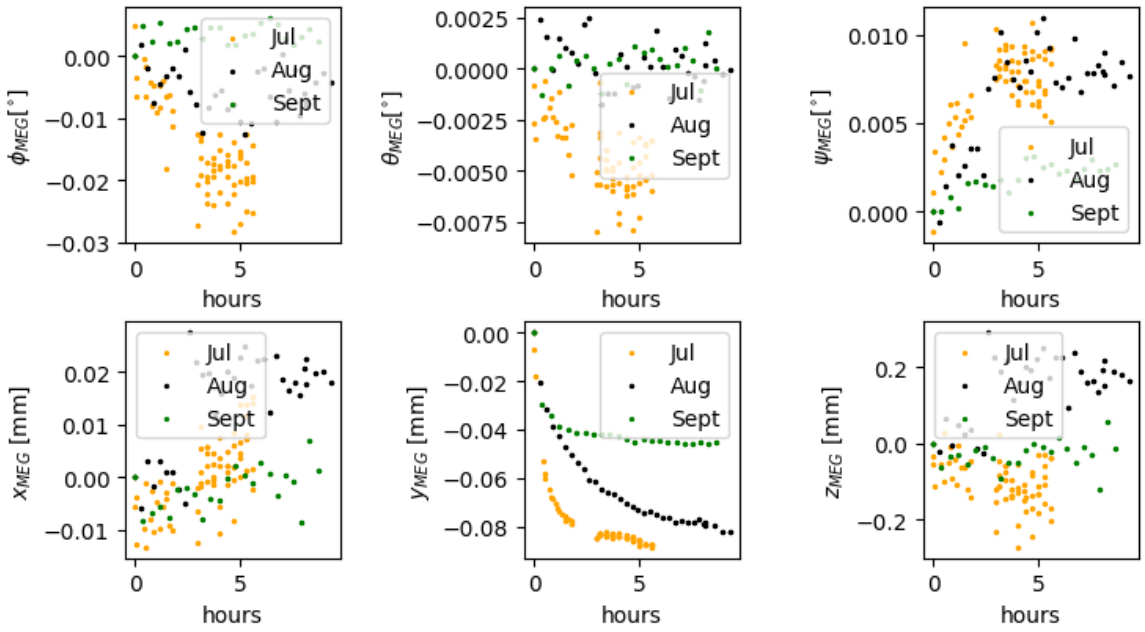


Figure 4.10: The 6-parameter transformation from X_{TGT} into X_{MEG} as a function of time in the 2021 dataset for three separate periods immediately after a target re-insertion.

Finally, we show an example of the data with several distinct regions in Figure 4.11. Using the dispersion in each region, we observe a precision in the angles of $\sim 0.006, 0.003, 0.002$ degrees for ϕ, θ, ψ respectively. In the translations, we observe a precision of $10, 3, 80\mu$m respectively, these are very approximate numbers.

4.8 Deformed Target Implementation

Implementing the target foil deformation into the positron analysis is required for optimal kinematic measurements at the target. The positron kinematic estimates at the target foil are estimated by propagating the positron track using the Genfit package[48]. Genfit uses a Kalman[30]/DAF[31] filter to fit for the positron state vector (and covariance matrix) in the drift chamber and then propagate the state vector through the magnetic field and geometries.

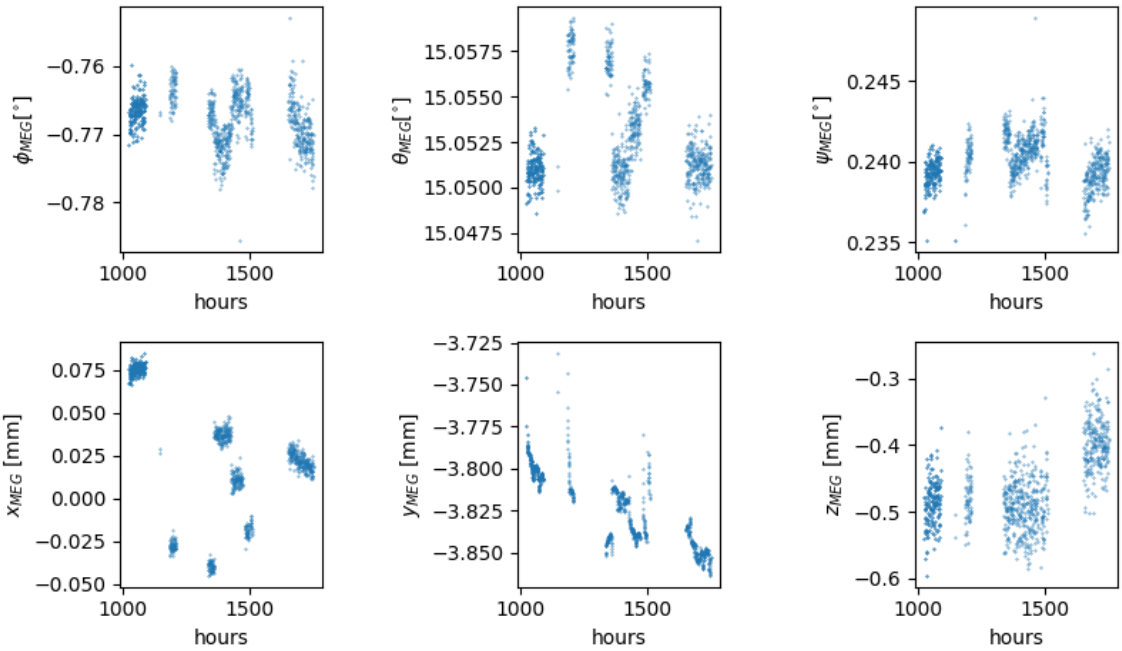


Figure 4.11: The 6-parameter transformation from X_{TGT} into X_{MEG} as a function of time in the 2021 dataset. This is zoomed in to highly the motion between insertion/extractions.

This section describes the technical details of the implementation.

First, we must account for the target material at the correct position. This is to correctly account for the energy loss and adjust the covariance matrix elements according to the material the positron has intersected. Second, we must propagate the positron tracks to the correct vertex position on the target surface; as mentioned deformations create a ϕ_e bias if left uncorrected. In this section we describe the details of both of these procedures. These details are outlined here mainly for any future user of the deformed target implementation.

4.8.1 Tessellation

To include the deformed target in the positron analysis, we define the target as a tessellated grid of triangles. Triangles are ideal as each set of three vertices defines a plane. The tessellation is described in this subsection.

First, we spline[47] the 3D dot coordinates over the foil's surface so we can choose our tessellation grid spacing. The spline is shown in Figure 4.12; the non-smooth edges are due to the lack of dots near the edge of the foil (and inside the frame). A smaller grid spacing results in more computation time to select/propagate to the correct tessellated triangle, whereas too large of a spacing results in errors in the deformation.

To discuss the scale of changes in the deformation along the target surface, we point again to the example graphic of a dot in the CT scan (Figure 4.6). We observe a gradient of roughly 100μm from the bottom-left corner to the top-right corner. Non-linear deformations are observed at a much lower level $\sim 10 - 30\mu m$. At the current moment, we use the spline to create a 60x25 array of points in X_{TGT}/Y_{TGT}, calculating the Z_{TGT} coordinate using the spline. Errors due to the granularity of the tessellation are negligible.

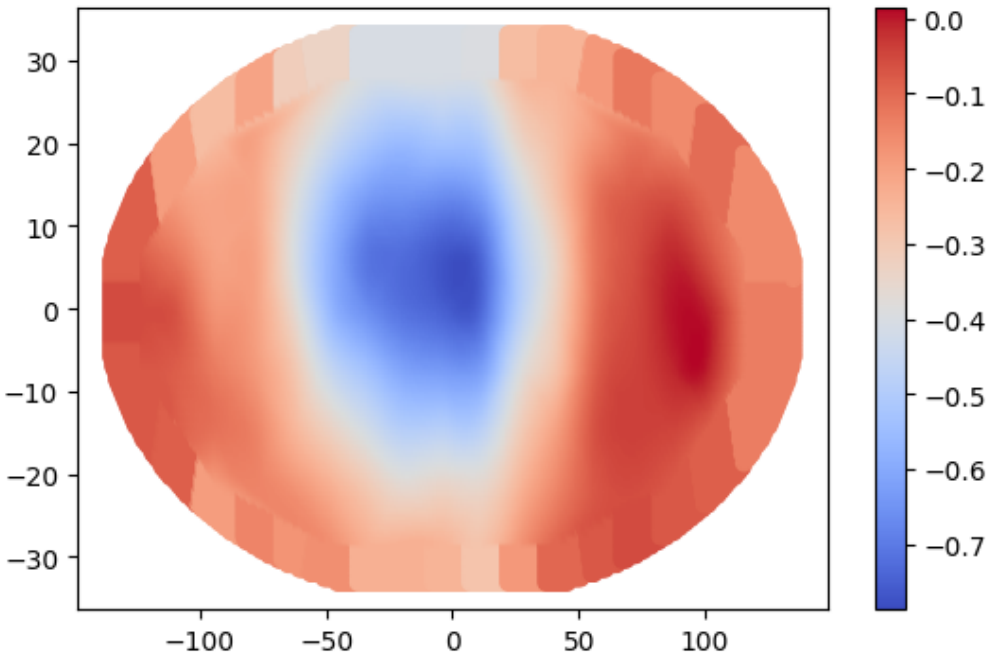


Figure 4.12: Spline function of the deformation of the target surface.

We need to convert this grid of points into an array of triangles (each of which defines a plane). We choose to cut the grid into two types of right triangles. The first type connects

points (i,j),(i+1,j-1),(i,j-1), the second connects points (i,j),(i+1,j),(i+1,j-1). Both sets of points are connected clockwise. This 'fills' the grid with an array of triangles without gaps. For each triangle, we save the three-vector for each vertex (9 total values). As a reminder, this defines the target foil coordinates on the face directed at the camera. This face points in negative X_{MEG}/negative Z_{MEG}. We require a target surface in the middle of the foil (for the propagation) and we need the location of both foil faces for the target geometry implementation.

The CT scan measures the foil thickness to be $\sim 250\mu m$. This is significantly thicker than independent measurements of the foil thickness (Malte Hildebrandt), which resulted in an average thickness of $179\mu m$ ($\sigma \sim 9\mu m$). The thickness was measured at 15 points on the foil. The manufacturer of the CT scan concluded this is likely a systematic in the CT scan. Currently, the implementation assumes a thickness of 180 μm for all points on the surface. The dot coordinates define the negative Z_{TGT} face, we then define the positive Z_{TGT} face at $\vec{X_{TGT}}$+(0,0,180) μm in the target frame.

The positron tracks are propagated to a triangular face (plane) which is in the middle of the two faces (averaging the three vertices individually over the two faces). Finally, to form the triangular prism geometry, we concatenate the additional three vertices on the positive Z_{TGT} face with those on the negative Z_{TGT} face to form the single triangular prism in the target coordinate system.

4.8.2 Target Geometry TGeo Definition

There is an additional difficulty associated with implementing the target geometry into the ROOT[49]/Genfit[48] software. The target geometry must by defined in the TGeometry ROOT framework. The only class that fits the requirements for a triangular prism is the TGeoArb8 class. To create a TGeoArb8 object, one must define 3-4 vertices on a XY plane

(clockwise) at -dZ and then 3-4 additional vertices on an XY plane (clockwise) at +dZ. The geometry combines the vertices on the two planes such that there are no 'twists'. By defining 3 vertices on each plane, a triangular prism is created.

We call this reference plane the 'bench-top' reference frame (X_{BT}). The 'bench-top' reference frame is **not** equal to the target frame unless the target is exactly flat (no deformation over the whole surface), which is not the case. Therefore, we must first fit for the rotation/translation from the 'bench-top' reference frame into the target frame.

To make the bench-top frame coordinates, we first assume that the (i,j) vertex is centered at $X_{TGT}, Y_{TGT}, -dZ$, then we assert that the second vertex, (i+1,j), is at $X_{TGT}+L_{1,2}, Y_{TGT}, -dZ$ where $L_{1,2}$ is the distance between vertices (i,j),(i+1,j). We then calculate the coordinates X_{BT}, Y_{BT} for the vertex (i+1,j-1) using the restrictions set by $L_{1,3}$ and $L_{2,3}$. A similar procedure is followed for the triangles oriented in the other direction (i,j),(i+1,j-1),(i,j-1).

We do the same for the other foil face, it is effectively the same except Z_{BT} and therefore the dZ offset sign is reversed. Note, since the two triangular faces are identical except displaced in Z_{TGT}, in the bench-top frame the two triangle faces must be slightly offset in X_{BT}/Y_{BT} as Z_{BT} is not parallel to Z_{TGT}. This offset is the thickness of the foil times the dot product $Z_{BT} \cdot Z_{TGT}$. For the same reason, the distance dZ is not exactly the distance in Z_{TGT} between the two surfaces, but the dot product $Z_{BT} \cdot Z_{TGT}$ (nearly 180μm).

We verify that the distance between all 6 vertices on the triangular prism in both the target frame and the bench-top frame are equal. We can now fit for the rotation/translation from the bench-top frame into the target frame. This is done using a python script. We verify that the residuals in the fit are drastically less than 1 μm ($\sim 1nm$), which is expected from a perfect rigid body rotation/translation.

We fill the 6 vertices in the bench-top frame and the rotation/translation from the bench-top frame to the target frame in a csv that is to be read by ROOT (18+6=24 entries/row). The

tessellated geometries are shown in Figure 4.13 and again in Figure 4.14.

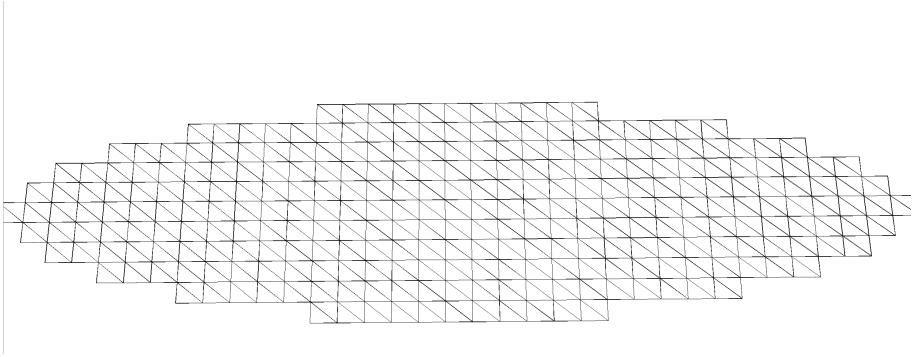


Figure 4.13: The full tessellated geometry is shown.

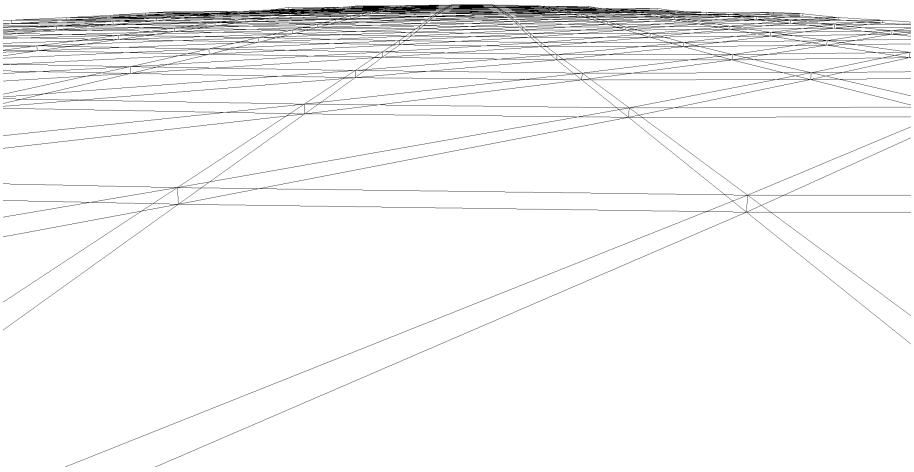


Figure 4.14: A zoomed in view of the tessellated geometry.

In the meganalyzer code, a ROOT macro (deformedTarget.c) is called by the MEGMain task. The task imports parameters from the sql database including the target position, rotation, and the deformed target csv filename. First, the macro makes a box object large enough to contain the full target. Second, all the TGeoArb8 objects are created (in the bench-top frame) and then rotated/translated into the box object (now in the target frame). Finally, this box object is rotated/translated from the target frame into the MEG frame using the database parameters. This system conveniently allows us to only adjust the database parameters to rotation/translate the target throughout the run.

This box (containing the tessellated target) is added to a larger geometry including all of

the materials in the experiment. That is, the timing counter, drift chamber, etc. The target geometry included with all other geometries is shown in Figure 4.15.

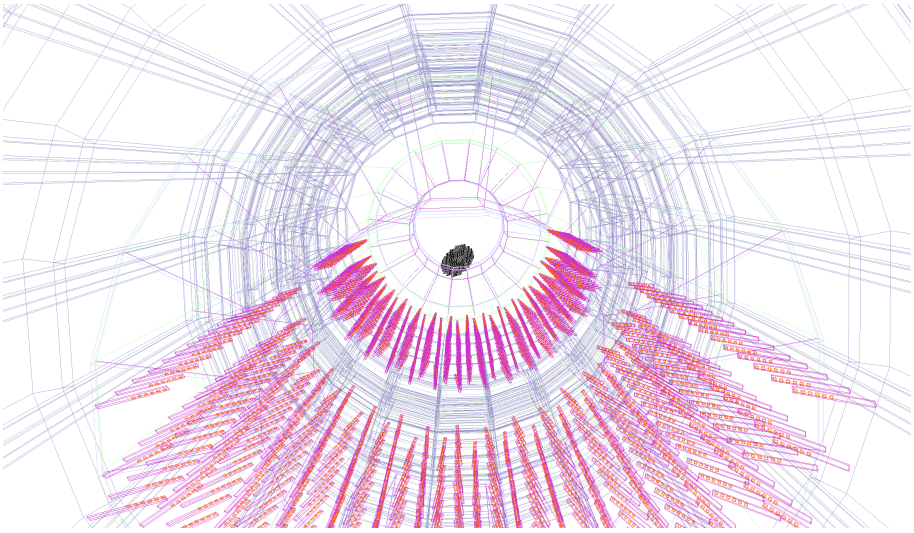


Figure 4.15: The full TGeometry of the MEG II experiment is shown. The LXe/RDC is excluded as it is not relevant for the positrons.

Note that both inserting the TGeoArb8 objects into the 'target box' and inserting the 'target box' into the world (ensemble of MEG geometries) requires use of the TGeoRotation class. This does not have Euler angle ZYZ implementation, but instead rotations about fixed axes. We found that if we instead flip the ordering of the angles i.e. (3-2-1 instead of 1-2-3) we get an identical rotation matrix to the euler angles matrix in the standard Rotation3D class.

The target scintillator material is included in the TGeoArb8 object creation in order to have the correct average energy loss of positrons as they move through the target. Currently, the carbon fiber frame and target support (positive Y_{MEG}) are not implemented in the Kalman. In order to account for the larger loss of energy and the increased energy uncertainty from intersecting the carbon fiber frame, we scale the triangular TGeoArb8 scintillator density. We estimate the carbon fiber to be 4 sheets of 200 μm with a density of $1.4 g/cm^3$ compared to the standard 180 μm with density of $1.03 g/cm^3$. This results in a relative density scaling of 7.2.

To summarize, the tessellated geometry contains the correct thickness and position of the target foil, including its deformations. The carbon fiber was approximated by scaling the material in the carbon fiber frame region. This implementation results in an optimal estimate of the energy loss in the target and the optimal covariance matrix estimation.

4.8.3 Genfit Foil Definition

The Genfit MEG code is initiated by pulling the same target database parameters used for the target geometry (previous subsection). After using the transformations described in the previous section, we have 6 MEG coordinate vertices defining each triangular prism (TGeoArb8). We need to define a plane in the middle of the triangular prism using these coordinates. This is defined using a DetPlane in Genfit. This requires a TVector3 normal to the plane's surface, an origin point, and a size (-L,L,-W,W). To calculate the normal vector we create two TVector3 objects defined as the vectors between the three vertices i.e. $V_1 \vec{-} V_2$, $V_1 \vec{-} V_3$. We then calculate the cross product of these two vectors and normalize it: this is the vector normal to the surface. For the origin, we use the mean of vertices. For the length, we choose an arbitrarily large plane (the rational for this will be described later). This creates an arbitrarily large rectangular plane centered about the mean of the triangular prism and parallel to the two triangular faces.

4.8.4 Propagation in Genfit

In the last section, we defined the target geometry in Genfit using the DetPlane class. In this subsection, we describe how we propagate to the target plane. All of the propagation to the target occurs in the DCHKalmanFilterGEN task. This function is initiated by filling the track initial state, hit information, etc. The tracks are then fit. Once a track is successfully fit, it is propagated to the target surface: this is the code that has been modified.

With a flat target, the positrons are simply propagated to a single DetPlane defined by the vectors normal to the target surface. Instead, we must choose what triangular plane to propagate to. Note, the scale of the deformation gradient is $\sim 10^{-2}$ i.e. moving 1 mm on the target surface typically contains $\sim 10\mu m$ deviations in the normal coordinate. As the triangular array is defaulted to $\sim 1.5mm$ spacing, propagating to the wrong triangle (off by 1 triangle) creates an extremely small and negligible error in ϕ_e. This motivates the use of a very large plane as propagating to the wrong triangle will not create a large error in ϕ_e.

All MEG signal events have the photon emitted towards the LXe (negative X_{MEG}) and the positron emitted in positive X_{MEG}. We start with an initial propagation of the positron tracks from the drift chamber to a flat target surface +1 mm displaced in X_{MEG} from the origin. The current 2021 deformations are all smaller than 1 mm and so this should get the state vector very close to the target surface. This displacement can be altered for specific time periods if necessary in the future.

Note, we found that even a target surface displaced by 1 mm does not always result in a vertex position close to the target surface. To avoid this issue, in instances when the $|Y_{MEG}|$ vertex coordinate is larger than the target's minor ellipse axis, we continue propagating the positron iteratively until this criteria is satisfied. In addition, we require that the positron track must be within 2 mm of the target surface. This iterative procedure helps a small fraction of the events, but is required to achieve optimal efficiency.

The above procedure results in a positron vertex position very close to the deformed target, we then propagate the track from this vertex to the triangular face with the closest center (mean of the three vertices). This is usually a propagation of ~ 5 ps (slightly larger than 1 mm). This is the value that is stored as the final target position. To check this procedure, we transform the target vertex from the MEG frame into the target frame by first subtracting off the translation and then applying the inverse rotation. Figure 4.16 shows the mean Z_{TGT} coordinate of positrons as a function of the X_{TGT} and Y_{TGT} coordinates. When comparing

the vertex coordinates to the CT scan dot coordinates, the vertex is slightly more positive in Z_{TGT}. This is the result of propagating the target to a surface in the middle of the foil whereas the CT scan dot coordinates are at the negative Z_{TGT} foil face.

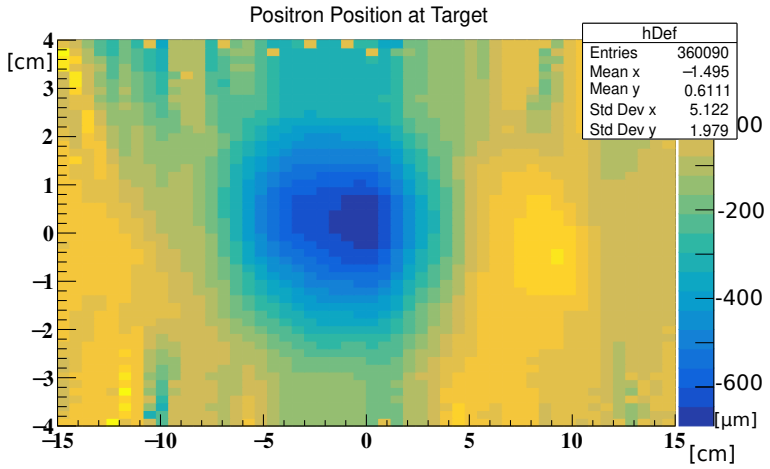


Figure 4.16: The positron position on the deformed target. The mean coordinate in Z_{TGT} is represented in color here as a function of the other two target coordinates [cm] (deformation has a maximum of $-700\mu m$).

The energy loss in the target material (foil or the frame) is shown in Figure 4.17

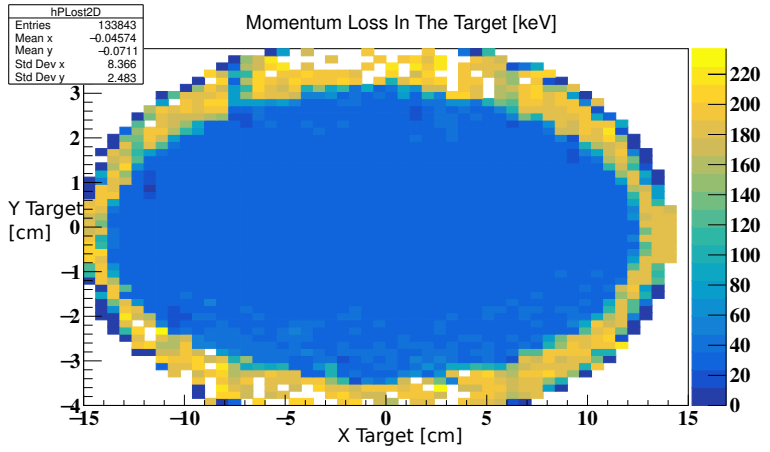


Figure 4.17: Momentum loss in the target [keV] as a function of position [cm]. The large energy loss at large radius is due to the carbon fiber frame implemented by scaling the triangular tessellated array density.

4.9 Conclusion

The camera analysis results in a precision normal to the target surface of $\sim 10\mu m$, far better than that the goal of $\sim 100\mu m$. Errors in the time-dependent target motion should have a negligible effect on the positron ϕ_e resolution. The analysis observed large shifts of the target position of the order $\sim 500\mu m$ from the time of the optical survey to the time of the MEG data-taking; these were accounted for in the final analysis.

In 2023, a new radiation-hard camera was implemented. It has not experienced any issues in the first ~ 3 months of the run. The camera has taken an image every 10 minutes, a total of roughly 13k images.

The deformation of the target was estimated using a CT scan. This was then checked using the camera analysis by looking at the residuals with respect to the CT scan deformation. The residuals with respect to the CT scan have a maximum size of $\sim 50\mu m$, still contributing a negligible systematic error to ϕ_e (< 0.5 mrad). We observed minimal change in the deformation over the 2021 run. The deformation of the target was modeled using a tessellated grid of triangles. The grid was used to model the 3D target geometry and propagate the positrons to the correct target vertex position. The geometry was implemented in the ROOT TGeometry class and the propagation was implemented in Genfit.

Chapter 5

MEG II Drift Chamber Analysis

5.1 Overview

In this chapter we describe the MEG II drift chamber analysis[14]. The general overview is depicted in Figure 5.1. The final objective of the positron analysis is to estimate the kinematic variables $y_e, z_e, \phi_e, \theta_e, p_e, t_e$ at the stopping target so that the kinematics can be compared to that of the photon measured in the LXe detector. To reiterate, the resolution of the positron and photon kinematic measurements is critical to achieving the optimal sensitivity of the experiment. Throughout the next several subsections, we describe in detail the procedures in that are used to optimise resolutions.

The drift chamber analysis is initiated by searching for drift cells intersected by potential particle trajectories (tracks). The digitized, time dependent signal (waveform) of such a particle passage is referred to as a hit. For each hit, we estimate the arrival time of each hit's first ionization cluster in the waveform. Using the relative time of ionization clusters and the relative amplitude of the signals on the two wires ends, we also estimate the position of the hit along the wire axis. This procedure is conventionally called the waveform analysis

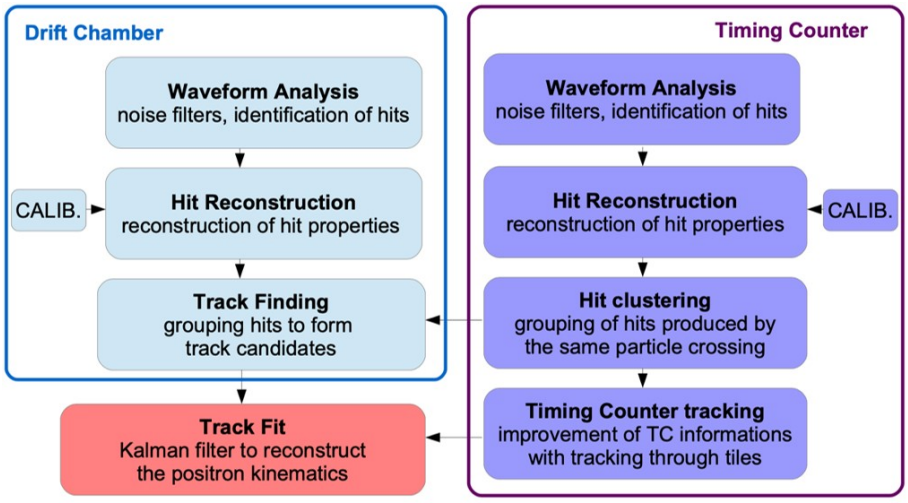


Figure 5.1: Overview of the positron analysis (Graphic made by Francesco Renga).

and is discussed in detail in Section 5.2; this involves noise suppression, signal detection, hit arrival time estimation, and the estimation of the position along the wire axis.

The hit arrival time is equal to the particle time at the cell (referred to as the T_0) plus the drift time of the ionization clusters. At this stage the T_0 for the individual hit is not yet known. The drift time of the first (primary) drift electron is converted into the track's distance of closest approach (DOCA) to the cell's sense wire using the position dependent drift velocity that is a calculated or measured value, which depends on the local electric and magnetic fields in the cell and on the gas mixture used. This is known as a time-distance relationship (TXY). The drift velocity is separately measured or calculated using a code such as Garfield++[50]. In Section 5.3 a novel technique to estimate the TXY using a neural network trained on data is presented; this results in improved resolution at the hit and kinematic resolution level. The DOCA and the hit's position along the wire axis (z) are both inherently measurements in the wire's local coordinate system, to convert them into the global MEG II coordinate system, we require precise knowledge of the wire position, direction, and sagitta. These wire positions, directions, and sagittas are measured using Michel positron tracks; the alignment is described in Section 5.5.

The hit information is input to a track finding algorithm used to assign hits into a particle track. Finally, a Kalman Deterministic Annealing Filter (DAF) method[30][31] is used to fit the tracks and provide a kinematic measurement at the target. The tracking procedure is detailed in Section 5.4. This includes a refitting algorithm that uses a track found in the first attempt at fitting to search for additional hits that should be on the track, but were not included by the track finder. The track is then projected to the target to provide a 3D decay position. Details on the propagation to the target surface including the deformation were discussed in Section 4.8. The tracking relies on a magnetic field map and the relative alignment of the drift chamber and the magnetic field map. In addition the final estimates of the relative kinematics of the positron and the photon rely on a precise estimate of the relative alignment of the target, the drift chamber, and the LXe detector. These alignments are discussed in Chapter 6.

In the positron dataset, there is a subset of positron tracks that contain very low resolution; these contribute very little to the physics analysis and can contaminate the signal with unwanted additional accidental time-coincident events. Section 5.6 outlines a machine learning approach trained on data to optimize the positron track selection algorithm. The algorithm minimizes the number of low quality tracks in the dataset while maximizing the overall tracking efficiency.

Finally, the physics analysis requires determining the probability of being signal or background (probability density functions) given a set of kinematic variables. The probability density functions are built using data driven kinematic resolution estimates at the target and the correlations between those kinematic variables, this is detailed in Section 5.7.

5.2 Waveform Analysis

5.2.1 Overview

This section describes the digital waveform analysis of the CDCH, the first step in the CDCH analysis. It consists of noise suppression, detection of ionization clusters (signals), estimation of the arrival time of the primary (first) ionization cluster, and estimation of the signal position along the wire axis.

The CDCH waveforms contain a broad frequency range noise that significantly contaminates the signals from ionization clusters. A majority of the low frequency noise is coherent over a pair of WaveDREAM boards. The first reconstruction step is removing this noise while preserving the signal amplitude and shap and is describe in Subsection 5.2.2. The high frequency filter is a discrete Fourier transform (DFT) high frequency cutoff filter, this is is briefly described in Subsection 5.2.3. This general procedure is not optimal for the channels with known issues; treatment for these cases are discussed in Subsection 5.2.8.

After filtering the waveforms, we start the signal detection procedure. The algorithm starts at the first waveform bin and searches for the first instance of a signal. This is done on both ends independently. The objective is to detect the maximum number of signals due to ionization clusters while minimizing the number of signals due to high amplitude noise; the algorithm is described in Subsection 5.2.4.

Once a signal on a single end is detected, we estimate the signal's leading edge time. This is the time associated with the rise in voltage due to the first ionization cluster to reach the wire. The algorithm is described in Subsection 5.2.5. In addition, we estimate the signal's coordinate along the wire axis using the difference in the signal arrival time and the difference in the collected charge on the two ends. This process is described in Section 5.2.6.

The signal is considered complete (no later ionization clusters associated with that signal) if a fixed number of sequential bins after the signal are below a voltage threshold. After the signal detection procedure is complete, the analysis results in an array of signals on both wire ends. The analysis proceeds to combine these signals into "hits". For example, one expects to find the same signal on the two ends. In addition, there may be a sufficiently large time delay between two ionization clusters and thus the signal detection procedure results in two "signals" on a single end that should be combined into a single hit. However, the analysis also must avoid combining signals from two separate tracks or combining a signal from noise with a signal from a set of ionization clusters. This procedure is described in Subsection 5.2.7.

After combining the signals into a final list of hits, the waveforms on the two ends are added to make a final hit arrival time estimate with the maximum signal/noise ratio. The sum waveform is built by displacing one end in time based on the fitted time difference on the two ends for this individual hit. This time difference is approximately $\pm$0-7 ns.

The hit arrival time, time difference and charge division on the two ends are passed to the hit reconstruction task. This task converts the hit arrival time, time difference, and charge division into a final hit time and hit z measurement using estimated wire-by-wire and end-to-end calibrations (described in Chapter 7). Finally, the hit time is converted into a hit distance of closest approach (DOCA) using a time-distance relationship, this is described in Section 5.3. The hit DOCA and z estimates are passed to the track finder to initiate the positron track reconstruction.

5.2.2 Coherent Noise Calculation

The objective of the coherent noise suppression is to suppress the noise that is coherent over the CDCH electronics. A majority of the noise is coherent over an upstream/downstream

pair of WaveDREAM boards or 4 DRS chips (4x8 channels). This coherent low frequency noise varies event-by-event and board-by-board. As an example, we show the average voltage bin-by-bin for the four chips in an upstream/downstream pair of boards in Figure 5.2; this is data without beam (pedestal data), where wires 424-431 have a higher front end gain than wires 416-423 by a factor of 2. The wires with different gain are shown to illustrate that the noise is clearly amplified by the front end preamp. It's clear that the low frequency noise on the four chips has the same phase. The upstream/downstream pair of boards has no connection other than the physical wires. This implies that the noise is generated on the sense wire or it is generated in one board and coherently travels across the sense wires.

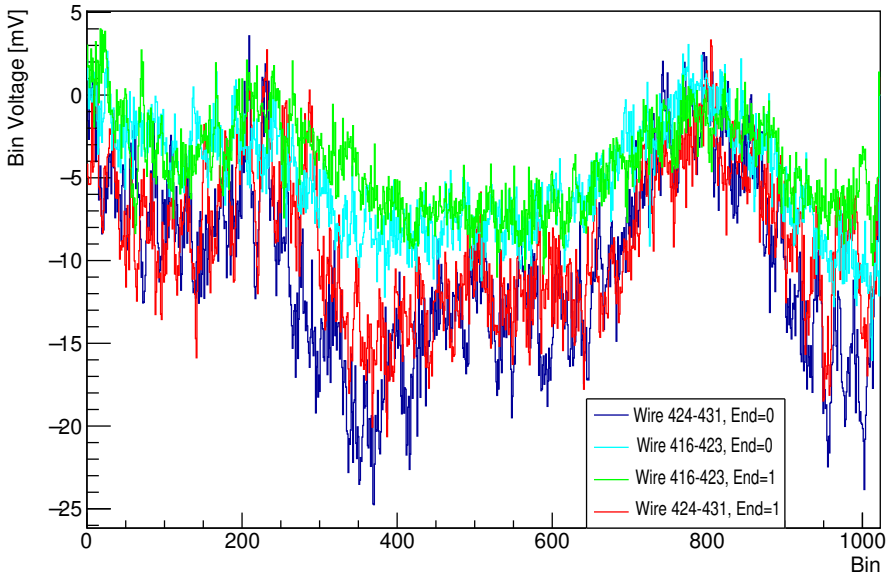


Figure 5.2: An example average voltage on a pair of upstream/downstream DRS boards. Wires 424-431 have an amplified electronic gain on the pre amplifier.

Note that attempts to suppress the noise was investigated by running with different front-end gains. As expected, the signal amplitude scaled with front-end gain. The RMS of the noise was measured using pedestal runs without the particle beam; the noise scaled at slightly less than the electronic gain. Thus the highest front-end gain ($\times 4$) had the best signal to noise ratio. The frequency spectra of the noise at different front end gains ($\times 1$,$\times 2$,$\times 4$) was measured using a discrete Fourier transform on data without any beam (pedestal), this is

shown in Figure 5.3.

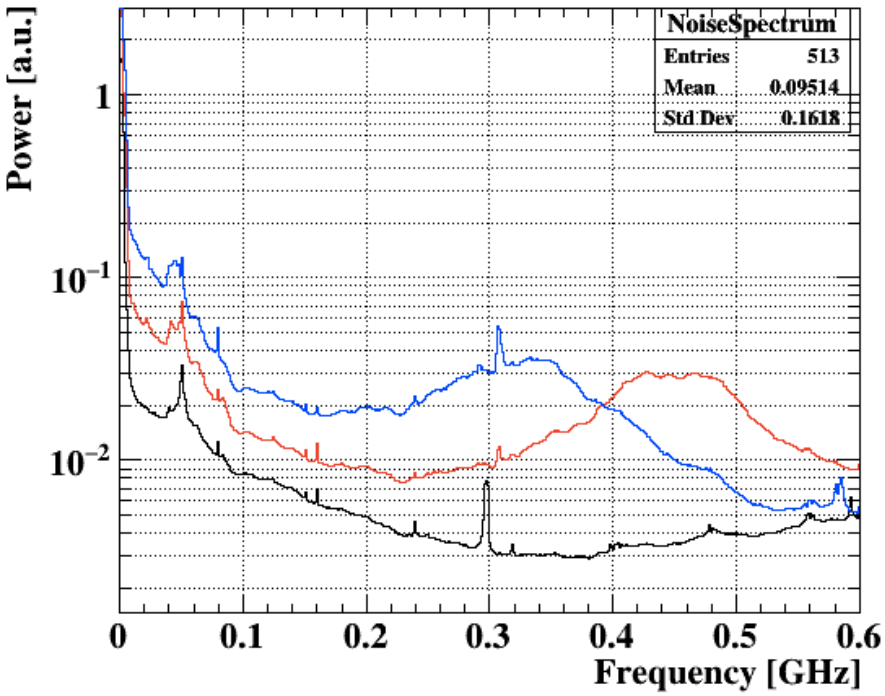


Figure 5.3: Power spectra of pedestal data with varying front-end gain. The gains $\times 1, \times 2, \times 4$ are shown in black, red, and blue respectively.

In the pedestal data, calculating and thus eliminating the low frequency noise can be done by simply subtracting the chip's average voltage bin-by-bin. However, clearly in physics data the calculation is complicated by the ionization cluster signals. This requires an algorithm that calculates the coherent noise while excluding signals. The variable names are given in parentheses to aid collaborators in understanding the code.

This procedure is initiated by first calculating a moving average baseline wire-by-wire. The moving average baseline is the average voltage over a wide number of bins (MovingAverageBaselineNBins), summing over the two ends wire-by-wire. Voltages above a threshold (VoltageThreshold) [what is the purpose of giving these names] are excluded from the moving average baseline calculation.

Next, we perform a rough "hit [potential confusion with other use of " search on the non-filtered waveform. We consider there to be a hit if a binned voltage minus the moving

average baseline at the bin is above a fixed threshold (VoltageThreshold redundant). Again, the binned voltage is the sum on the two ends. If a hit is detected, we exclude the nearby bins (± ExcludeNBins) from the coherent noise calculation on both ends.

Finally, the coherent noise is calculated bin-by-bin for each DRS chip, excluding the bins with a detected hit. To clarify, if a hit is found in bin A on wire i, bin A on wire j on the same DRS chip is not excluded from the calculation. If every channel on the DRS chip has a hit in a particular bin, the coherent noise is taken as the average coherent noise on the chip over a wide range of bins (AverageNBins). This coherent noise is then subtracted bin-by-bin for each DRS chip.

As an example, we show a set of waveforms on the same DRS chip in Figure 5.4. A majority of the low frequency noise has been eliminated on all channels.

To quantify the amount of the noise that is coherent, we show the average power spectra in pedestal data over all DRS chips with and without the coherent noise subtraction in Figure 5.5. In addition, we show the average power spectra in pedestal data for individual DRS chips with and without the coherent noise subtraction in Figure 5.6. We observe that each individual pair of DRS chips has a unique noise spectra. In all cases, a large fraction of the noise is removed by the coherent noise subtraction.

In addition, we histogram the binned voltages in Michel data with and without the coherent noise subtraction in Figure 5.7; the positive tail is due to signal.

The variables (MovingAverageBaselineNBins,VoltageThreshold, ExcludeNBins, AverageNBins) have all been optimized based on the 2021 physics data by maximizing the number of hits on high quality fitted tracks (this is a combination of maximizing the tracking efficiency and the number of hits per track). The analysis is not very sensitive to MovingAverageBaselineNBins; it should remain a large value to account for the very low frequency noise (wavelength is greater than the length of waveform). The analysis is sensitive to the Volt-

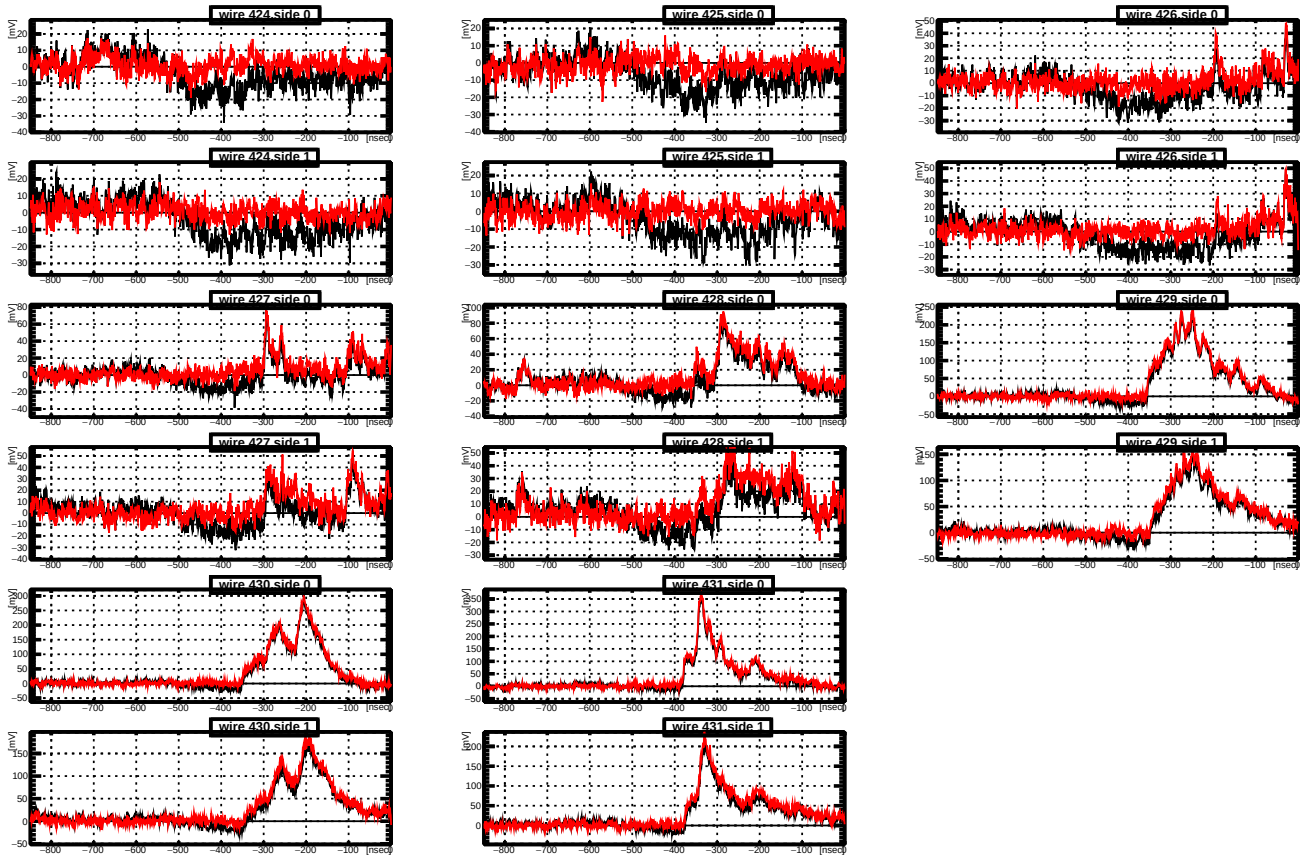


Figure 5.4: Michel data waveforms with the coherent noise subtraction filter applied (no filter) shown in red (black).

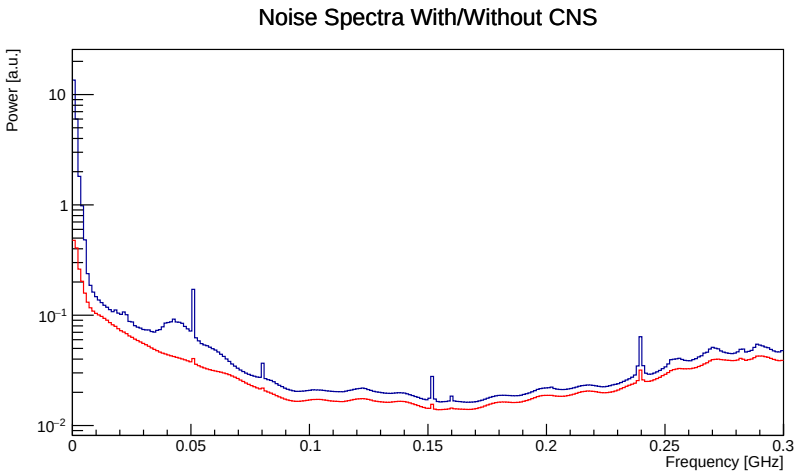


Figure 5.5: A fast Fourier transform of the waveforms with and without the coherent noise suppression.

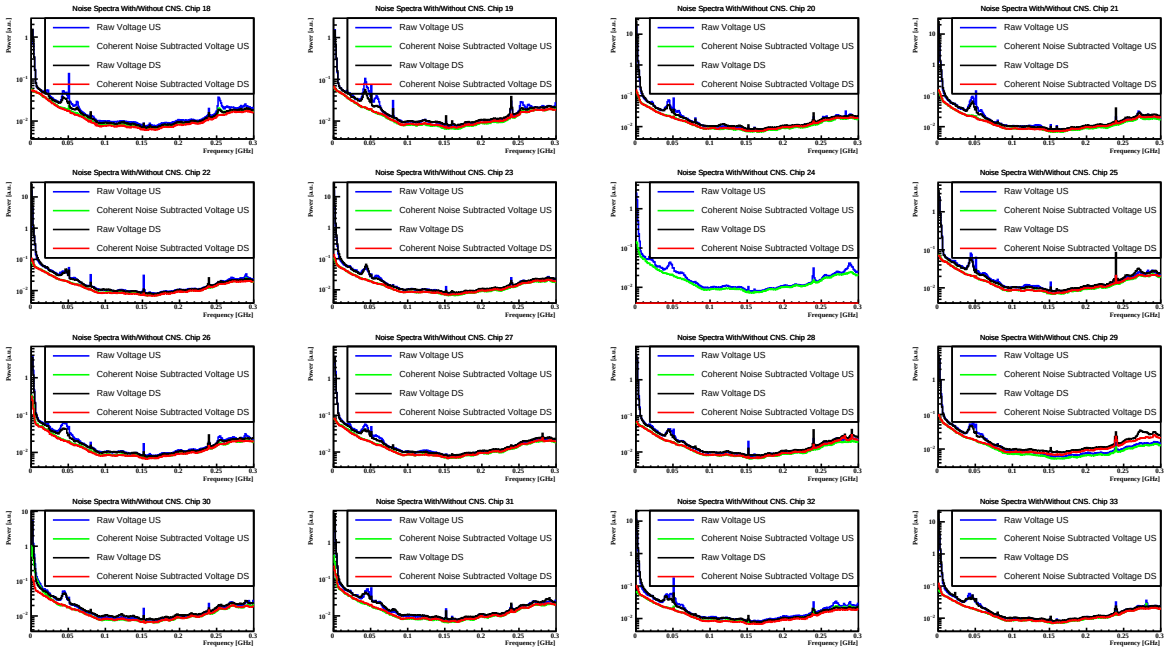


Figure 5.6: A fast Fourier transform of the waveforms with and without the coherent noise suppression for individual DRS chips.

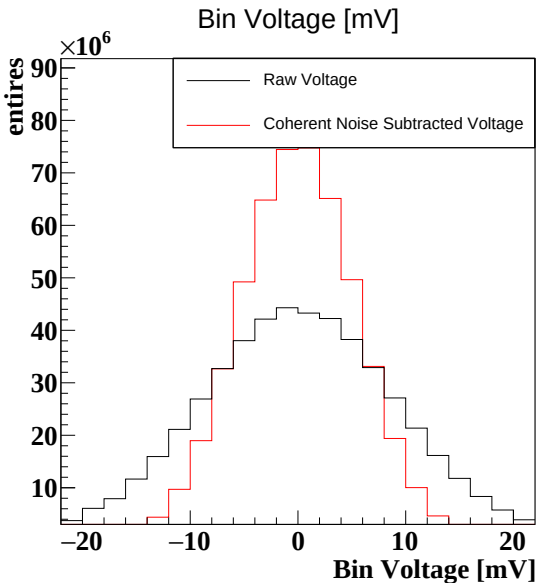


Figure 5.7: Voltage bin-by-bin with and without the coherent noise subtraction.

ageThreshold parameter. If this is too small, high amplitude noise is mistaken as a signal and thus the noise is excluded from the coherent noise calculation and thus maintained on the wire. If this is too large, signals are mistaken as noise; the signal is thus added into the coherent noise calculation resulting in an incorrect average positive voltage, which when

subtracted channel-by-channel, creates a negative baseline. Both instances negatively affect the signal detection efficiency on the channel and may result in picking up more noise hits on channels with no signals. ExcludeNBins is set to 50 bins, i.e. if a bin is above threshold, the surrounding 50 bins are all excluded from the coherent noise calculation. If this is too small, there are instances when a hit is detected, but not all voltages from the signal are excluded, thus some low amplitude signals will contaminate the coherent noise subtraction. If this region is too large, an excess of noise bins on that wire are removed from the coherent noise calculation. Finally, the analysis is not very sensitive to the AverageNBins parameter, which aids in instances when all channels have signals at the same time; this rarely happens.

5.2.3 High Frequency Noise Suppression

The standard waveform analysis applies a discrete Fourier transform to eliminate the high frequency components of the waveform. The optimal frequency cutoff removes the noise while maintaining the signal shape and amplitude. To aid in selecting the optimal high frequency cutoff, we show the spectra of pure noise in pedestal data (Figure 5.5) and the spectra of digitized ionization clusters with zero noise based on the MEG II Geant4 Garfield++ simulation (Figure 5.8). The pedestal noise spectra is shown with and without the coherent noise subtraction. The noise spectra has several narrow peaks (e.g. 155, 160, and 240 MHz) and several wider peaks (e.g. 255, 270, and 290 MHz). In addition, the power gradually increases with frequency from 200-300 MHz. The signal loses power rapidly until it plateaus at $\sim$ 250 MHz. Ideally, we would eliminate all noise peaks, however cutting at $\sim$ 155 MHz would remove significant signal power. This suggests an optimal cutoff frequency of $\sim 175 - 225$ MHz to maintain the maximal signal power while eliminating the noise peak at 240 MHz and the slow rise in power starting at $\sim$ 200 MHz. We additionally tuned the cutoff frequency by maximizing the number of hits on fitted tracks; this achieved a similar optimal frequency cutoff of $\sim$ 225 MHz.

Future analysis should be tuned on this cutoff, especially if it is known that the noise conditions have changed.

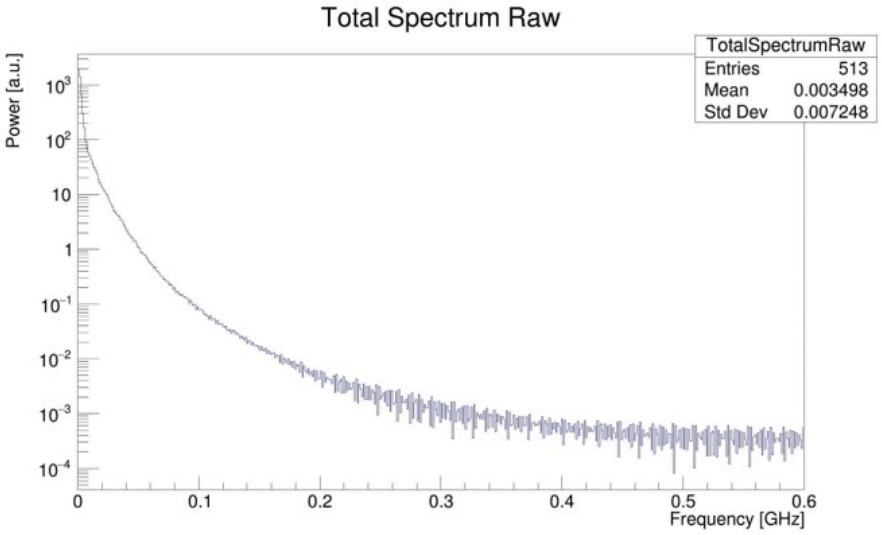


Figure 5.8: Fast Fourier transform of the digitized signals in the Monte Carlo simulation without any noise.

5.2.4 Discriminator

The discriminator code detects signals on the waveform. The algorithm has been tuned to maximize the number of true detected signals while minimizing the number of fake signals due to noise. This function is applied after the noise suppression. The current discriminator shape is roughly based on the signal shape estimated by fitting signals presumed to be single drift electrons. An example of a fitted signal is shown in Figure 5.9. The average over many fitted signals results is shown in Figure 5.10.

There is still some uncertainty, but it's clear the signal shape consists of a sharp peak over $\sim$ 2 ns from the 10%-90% level; this corresponds to $\sim$ 2 bins (time/bin =1.67 ns). The pulse shape has a wide tail dropping to the 20% level after $\sim$ 25 ns. In most occasions, the waveform is not due to a single drift electron, but a sequence of ionization clusters that create a signal from overlapping pulses.

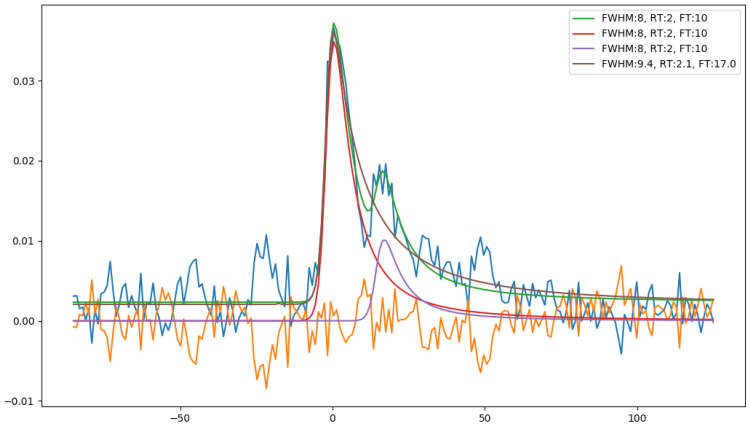


Figure 5.9: An example hit with the two-pulse fit (green), the two individual pulses in the two-pulse fit (red, purple), the one-pulse fit (brown), and the residuals to the two-pulse fit (orange). RT is the 20%-80% rise time and FT is the 80%-20% fall time; both are listed in nanoseconds.

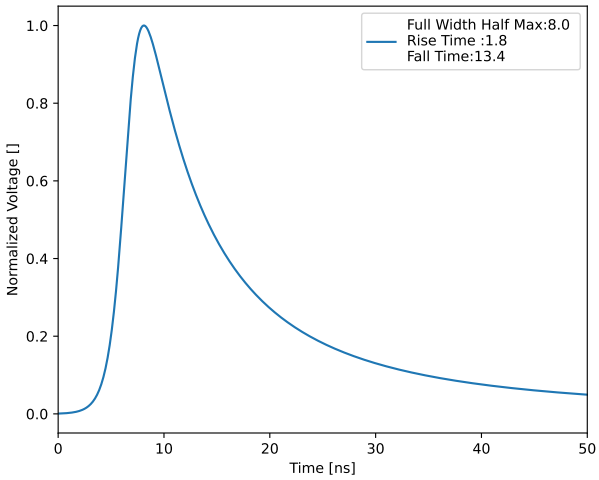


Figure 5.10: The pulse shape estimated by averaging over fitted signals in the 2020 Michel data. RT is the 20%-80% rise time and FT is the 80%-20% fall time; both are listed in nanoseconds.

The standard discriminator requires voltage at bins i and i+1 above a voltage threshold (BinVoltageThreshold) plus the sum voltage over sequential N bins starting at bin i to be over an integrated voltage threshold (IntegratedVoltageThreshold over $\sim$ 25 ns). Both thresholds are a fixed threshold minus the estimated end-by-end, event-by-event baseline, which should be highly suppressed by the noise suppression.

A signal is considered complete by using the same criteria to start a hit (two bins below a threshold and an integrated voltage over many bins below another threshold). If this criteria is met over a fixed number of sequential bins (UnderThresholdNBins), the signal is ended. However, if this criteria is met again, we assume the signal has later ionization clusters and thus resets the search for the end of the signal.

We make a few general points to close this section. First, the two thresholds (BinVoltageThreshold and IntegratedVoltageThreshold) are critical to the tracking efficiency. If the thresholds are too low, there will be an excess of noise hits in the track finder which degrades the tracking efficiency and results in a very long track finder routine (CPU time). Signal hits are missed if the thresholds are too high, resulting in a degraded tracking efficiency as well. If UnderThresholdNBins is too small, later ionization clusters will be mistaken as a new hit, however this can still be resolved by an algorithm described later that combines "hits" on a single end. However, if this parameter is too large, multiple hits (e.g. one from a pileup track and another from the triggered on track) will be combined into a single hit. Therefore, it is best for this parameter to be small (default is 40 ns).

5.2.5 Hit Arrival Time

This algorithm estimates the first ionization cluster arrival time. Frequently, the discriminator does not find the first ionization cluster, but a later higher amplitude signal. More accurately, the discriminator detects a series of ionization clusters creating voltage on top of one another thus creating a larger voltage. This is particularly true in instances when the track nearly intersects the wire; in these cases there is often a lone first ionization cluster (small voltage amplitude) that can be missed by the discriminator. Similarly, there is a large time-distance gradient near the edges of the cell, therefore if a track only intersects the corner of a drift cell, we are more likely to get an isolated first ionization cluster. The algorithm

proceeds by searching backward in time from the bin triggered by the discriminator. This task aims to find the earliest, smallest signal due to an ionization cluster without mistaking noise as a fake signal. Furthermore, the function must avoid moving too far back that it detects an earlier hit from a different positron track.

Similar to the discriminator, the standard leading edge time finder requires 2 bins above a voltage threshold plus an integrated voltage threshold. The voltage thresholds are fixed thresholds with respect to the calculated baseline. These thresholds are slightly lower than that of the discriminator; this is ideal as unlike the discriminator, having too low of a threshold does not result in an excess of hits, but too early of a hit arrival time (i.e. too small a DOCA). A lower threshold should be used as we have more information; the discriminator was passed and thus are confident we have a signal so a small amplitude peak is more likely to be a lone drift electron that just a noise spike.

The leading edge time finder only searches back to the end of the previous hit (to avoid picking up late signals from an earlier hit) and only searches back a fixed amount of time; the smaller of these two is selected. Default maximum look back time is 20 ns.

The track distance of closest approach to the wire resolution (primary track position measurement) is sensitive to the two leading edge time thresholds.

5.2.6 Hit Z: Cross Fit

The objective of this function is to estimate the signal's coordinate along the wire axis. This is done by "fitting" one end of the waveform to the other using three free parameters: relative scaling, time difference, and a baseline difference. Due to the wire resistivity, the charge on the two wire ends is different (more charge if the signal is closer to the electronics). Due to the signal propagation speed along the wire, ionization cluster arrival time at the electronics

on the two ends has a time difference of ±0-6 ns with a signal propagation speed of c.

As there can be other signals on the waveform that could contaminate this measurement, the cross fit uses only a section of the waveform. In theory, we should use the full hit duration, from the start to the stop of the hit. However, this doesn't result in a large improvement with respect to the default, which is a fixed 30 ns, 20 ns before the start time and 10 ns after.

The main advantage of this technique is that it uses information from all ionization clusters whereas the conventional approach only uses the first bin. The cross fit results in comparable z resolution from charge division with respect to the conventional approach (relative integrated charge in the two hits); this is as expected as the conventional approach to z from charge division already uses information from all bins.

The cross fit uses limits in the scaling/ time difference which are slightly larger than the actual limitations of the detector (e.g. a time difference limit corresponding to a larger z than the actual length of the wire). The analysis is not particularly sensitive to the cross fit parameters and the analysis doesn't need to be tuned on them. The current resolution in z for charge division and time difference is shown in Figure 5.11. Note, the z resolution has nearly zero effect on the positron kinematic resolution, but it is critical for reducing the combinatorics in the track finder.

5.2.7 Combine Clusters

The objective of the combine clusters algorithm is to merge the signals found on the two wire ends. Ideally, the algorithm combines all signals associated with the same hit, but keeps signals from two individual hits separate. Additionally, this algorithm aims to not merge a noise signal with a signal from an ionization cluster.

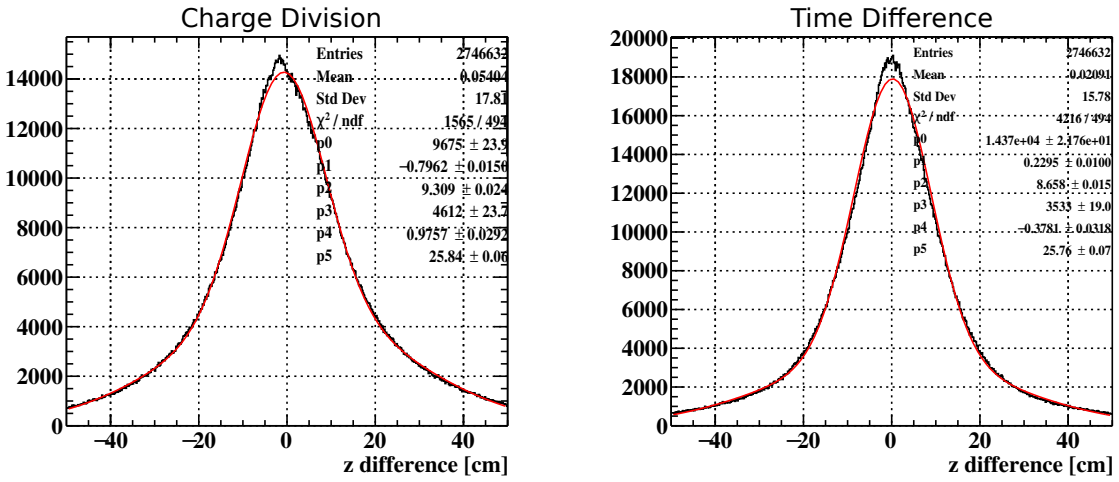


Figure 5.11: The z resolution for the two measurements: time difference and charge division.

The algorithm starts by looping through all possible signal pairs. The pair is merged if the difference between their leading edge times is below a threshold (TimeThreshold, default of 300 ns) and the difference between their cross fit end-to-end time difference is below a threshold (TimeDifferenceThreshold, default of 2 ns).

The hit leading edge time is taken as the earliest signal leading edge time. The stop time is replaced by the latest stop time. The relative time difference and relative scale (z measurements) are replaced by a weighted average of the two signals being combined. The weighting is the uncertainty from the cross fit routine (Minuit).

Once two signals are combined, they are eliminated from the signal list and replaced by the hit (combination of the two). Thus, the resulting hit can still be combined with additional signals.

The analysis is sensitive to the TimeDifferenceThreshold parameter. If this parameter is too small, the analysis doesn't combine enough clusters, leaving a large number of hits in the track finder (increased CPU time). If this parameter is too large, the analysis is more likely to combine the signals from two separate positron tracks or the signal from an ionization

cluster and a noise hit. In general, the smaller this value is, the higher the CPU time and the higher the tracking efficiency; the tracking efficiency eventually reaches a plateau.

5.2.8 Wire Noise and Bad End Treatment

In this subsection we list the categories of known wire issues. The wire lists are contained in a mysql database that needs to be tuned every year, or in instances when a time period is known to have increased noise.

5.2.8.1 Single Channel Readout

Some wires only have one end read out. As the bulk of the analysis is done on single ends, this is a minor issue. When creating the sum waveform, we instead sum two copies of the read out end. This is to have the peak amplitudes at the same level as the standard wires. In addition, we assert the z of the hit to be 0. We implement a larger σ_z in the track finder/fitter for these hits.

5.2.8.2 Non-Coherent Noise Channels

If a wire has non-coherent noise on either end, the wire is excluded from the coherent noise calculation. This wire would otherwise contaminate the coherent noise calculation. In order to suppress the low frequency noise on these wires, we instead implement a high pass moving average filter of 50 ns. This has been shown to lower the peak voltage and the signal/noise (this is why we don't use this on all waveforms). We then proceed with the standard analysis. These have a degraded track position resolution due to the worse signal/noise ratio. Note, we also tried a discrete Fourier low frequency cut off, this gave a degraded tracking efficiency.

5.2.8.3 High Noise Channels

An additional set of channels have non-coherent noise on a single end that is consistently high amplitude (> 50 mV). For this reason, it is very difficult to distinguish the signal from noise. We remove these channels from the analysis and treat the wire as if we only have a single end.

5.2.9 Analysis Status

The analysis for the 2021 physics data has been tuned over a few iterations optimizing the number of hits on tracks, the tracking efficiency, and the tracking resolution. There are still some methods that can be improved upon, these are listed below.

- The optimal treatment for waveforms with known pileup has not been fully optimized. e.g. if we see a signal before the T0 (e.g. after tracking), it's known that we have an early pileup hit on the waveform. The optimal method for detecting the leading edge time method in these situations has not been investigated thoroughly.

- The cross fit uses a fixed window of -20 ns before the hit time and 10 ns into the hit. This is to avoid pileup issues. However, we can use the total hit length (-20 ns before the hit arrival time to the hit stop time) to increase statistics. On average, hits are drastically longer than $\sim$ 20 ns. By adding in more bins, we should improve the measurement in z by sqrt(NBins). This should be explored.

- In wires with non-coherent noise we use a moving average high pass filter. This has not been optimized thoroughly. In addition, these wires use the same discriminator even though it is well known that the signal shape has been distorted.

- The calculation of the baseline after the coherent noise calculation is applied has not

been investigated in detail. In addition, using a fixed threshold compared to a threshold with respect to the RMS (and the RMS calculation) has not been investigated. Currently we are using a fixed threshold with respect to the baseline.

- One technique that could be implemented in order to improve the detection of low amplitude hits would be to use a sum waveform for hit detection. e.g. at z=0 the hit peak amplitude improves by a factor of 2 and the noise (if non-coherent) increases by a factor of sqrt(2). For this reason this should allow us to detect hits with smaller amplitude. These will be mainly hits near the edge of the cell with a small number of clusters. To be clear, this does not affect the leading edge as we use the sum waveform for already detected hits.

In addition, we briefly note here that there is an alternative hit detection algorithm, developed by Yusuke Uchiyama, which uses a neural network for hit detection. The algorithm is trained using noise from pedestal data and signal shapes from the pulse fitting described previously. This algorithm results in a very large amount of hits with respect to the conventional waveform analysis. Adding these hits into the hits found by the conventional analysis results in very high CPU time in the tracking algorithms, higher tracking efficiency, and degraded kinematic resolutions. The precise treatment of mixing the hits and the tracks from the two hit detection algorithms is describes in the track selection chapter.

5.3 Neural Network Applications to Improve Drift Chamber Track Position Measurements

5.3.1 Overview

This chapter describes applications of two neural networks to improve drift chamber position measurements. One network calculates a data-driven estimate of the drift cell time-to-distance relationship that is conventionally estimated by a numerical calculation based on the anode and cathode wire geometry, wire potentials, and gas properties. The second network additionally uses the full digital waveform of the signal in the drift chamber, hence accessing information on the full ensemble of ionization clusters. This network uses more information than the conventional position estimate that relies exclusively on the arrival time of the first drift electron. In principle, this technique improves resolution even when multiple ionization clusters cannot be separated, in contrast with a cluster-counting technique. The performance of both networks when applied to MEG II drift chamber data is reported and compared to that of a conventional approach.

5.3.2 Introduction

The primary position measurement of an ionizing particle traversing a drift chamber is the distance of closest approach (DOCA) of the particle to the cell's sense wire. Optimizing the accuracy and precision of the measurement is critical to a variety of modern day particle physics experiments ranging from high energy lepton collider experiments to rare decay searches at low momentum. In many applications, the positron measurement resolution in a magnetic spectrometer has significant contributions from both the position measurement and scattering in the drift chamber material. The position resolution has contributions from

ionization statistics, longitudinal diffusion, electronic noise, and effects that contribute to tracking errors (e.g. signals from an out-of-time ionizing particle). This paper focuses on improving the position measurement by applying neural network techniques. In Section 5.3.2.1, we discuss limitations to the conventional DOCA estimation and how using a neural network might improve it. In Section 5.3.3.1, we give a brief overview of the MEG II experiment and an explanation of how the neural network application could improve its DOCA estimation. We then present the neural network methods in Section 5.3.3.2. In Section 5.3.4, we discuss the improvements in kinematic measurements resulting from applying the neural network approach to MEG II drift chamber analysis. Finally, in Section 5.3.5, we discuss potential improvements to the technique.

5.3.2.1 Improving the Conventional DOCA Estimate

Here, we list the issues with the conventional DOCA estimate that may contribute to systematic errors in the most probable value and dispersion of the DOCA estimate, and how using a neural network avoids or minimizes these effects.

First, the conventional approach systematically overestimates the DOCA in any drift chamber with a low cluster density. This is illustrated in Figure 5.12, which shows the mean DOCA bias (i.e. DOCA of the 1st ionization cluster - the track DOCA) as a function of the track DOCA in the MEG II Geant4-based simulation[51]. At small DOCA (¡2 mm), the first ionization site is farther from the wire than the track DOCA by 50-200 μm. By training on data, the neural network approach can learn and thus remove this bias as a function of drift time, track angle, cell size, etc.

Next, in many analyses, including that of MEG II, the DOCA estimate relies on Garfield++[50] to accurately estimate the ionization pattern/statistics and the drift velocity in the gas mixture in use. This comes with uncertainty that is eliminated by a data-driven TXY function.

Garfield++ is inherently two dimensional; if the cell size or relative anode cathode geometry is changing along the wire axis, the algorithm necessitates an interpolation of the TXY function from calculations at fixed positions along the wire axis. Additional uncertainty in the TXY function arises from assembly variations in wire positions and the effects of bowing of wires from electrostatic and gravitational effects. These variations in the TXY function can in principle be corrected in the machine learning analysis given a sufficiently large training sample.

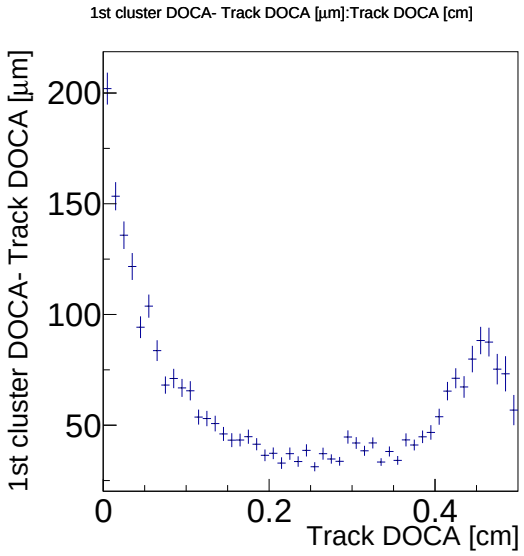


Figure 5.12: The mean value of the difference between the first ionization cluster position and the distance of closest approach (DOCA) of the track to the sense wire (DOCA bias) as a function of the track DOCA, from a Geant4 based simulation of the MEG II CDCH. It is summed over all wires to show the general trend of the DOCA bias. This bias shows the ionization statistics effect in the gas mixture.

Finally, the conventional DOCA estimate uses only the drift time of the first (primary) drift electron. An improved DOCA resolution can be achieved by identifying ionization clusters and their times (i.e. cluster counting). Cluster counting algorithms map the cluster time distribution to an improved final DOCA estimate relying on a correct determination of the number and pattern of clusters in a hit. In drift chambers with a low signal/noise ratio, it is difficult to distinguish overlapping clusters and differentiate noise from low amplitude

clusters. Since all available ionization cluster information is stored in the digital waveform, a neural network may use the available information without explicitly counting clusters to determine the best DOCA estimate.[52][53]

5.3.3 Methods

5.3.3.1 MEG II Experiment Limitations

Here, we list limitations of the conventional DOCA estimation specific to the MEG II experiment.

First, in the CDCH, the location of the cathode wires surrounding a given anode wire varies with position along the wire due to the stereo geometry in the drift chamber. In addition, the cell size is changing along the wire axis and layer-by-layer. Since Garfield++ is an inherently 2D program, the analysis relies on slices along the wire axis; the magnetic field and therefore the TXY function is then approximated along the wire. This explicit binning is not required by the neural network.

Simulating any gas mixture comes with uncertainties; this is especially true when using a non-standard gas mixture like the 4-gas mixture used in MEG II. The standard MEG II Garfield++ simulation of the TXY function does not include oxygen or isopropanol; including them did not yield any clear improvement in the DOCA estimate. Training a machine learning algorithm on data removes reliance on the simulation. Further, the isopropanol and oxygen lower the gas gain and therefore the signal/noise ratio in the waveform, worsening the precision with which the first ionization cluster time is measured and making estimating the number and time distribution of clusters (cluster counting) more difficult.

5.3.3.2 Neural Networks

In this section, we describe the two neural network approaches to DOCA estimation. First, a dense neural network (DNN) is trained to create a data-driven TXY function. That is, converting the waveform analysis hit arrival time into a DOCA estimate depending on other variables like cell size and track angle. Here, there is no explicit dependence on these other variables, but the network can implicitly learn how the DOCA estimation depends on these variables. Second, a convolutional neural network (CNN) also uses the full waveform to access information from all ionization clusters. The convolutional network layers are used to incorporate the temporal correlations between the waveform bins.

5.3.3.3 Training

The networks are trained using MEG II data taken at low beam rate. We use the fitted track DOCA (unbiased by the hit itself) as an estimator of the true DOCA for the training. This technique results in a data-driven DOCA estimator. Both networks were trained in Keras[54]. The input variables (X_i) have been normalized using their mean (μ_i) and standard deviation (σ_i): $X_i \to \frac{X_i - \mu_i}{\sigma_i}$. The track DOCA (Y) estimate has the same normalization: $Y \to \frac{Y-\mu}{\sigma}$. This normalization is applied for optimal training (i.e. find the optimal network weights and thus the minimal loss function). The network was trained with a mean absolute error loss function ($\sum |Y - \hat{Y}|$) to avoid a large dependence on outliers (Y represents the fitted track DOCA and $\hat{Y}$ represents the network's DOCA estimate).

The use of low beam rate data ($1 \cdot 10^6$ Hz, ~ 40 times lower than standard beam intensity) effectively removes all out-of-time or "pileup" tracks that otherwise contaminate the signals. We found that using low intensity beam rate data improves the DOCA resolution even at standard beam intensity. This implies that the network has difficulty learning the TXY function at the same time as distinguishing between in and out of time signals (i.e. pileup

discrimination). We come back to this point in Section 5.3.5.

Below, we list the input variables with some justification. All variables are included in both networks. Some of these variables are specific to the MEG II application. For example, if the cell size and relative anode cathode geometry was independent of the position along the wire axis, it would not be necessary to include this as an input variable.

5.3.3.3.1 Input Variables

- Hit arrival time - This reconstructed hit time is the result of conventional waveform analysis and is the primary input to the TXY function.
- Layer number - The cell size decreases with layer number. There could be unique properties to specific layers (e.g. the first/final layers) so it's also included as a hot encoded variable (binary length 9 array).
- Wire number in the layer - The average track angle varies with the wire's global ϕ. There may be trends in the anode cathode geometry as a function of wire number in a layer.
- T_0 at the wire - This is required for any DOCA estimator. It is the SPX-based estimate and has an uncertainty, σ_{T_0}, of 35-90 ps with a full T_0 range of $\pm$10 ns.
- Track angle at the wire - Isochrones are not circular, thus the angle is required for the optimal estimate.
- Kalman filter stereo reconstructed longitudinal wire coordinate - Cell size and thus the TXY function vary along the wire.
- Waveform channel time and gain calibrations - These are useful for the convolutional network to normalize the wire-by-wire and channel-by-channel differences due to gain differences, cable lengths, etc.

- Binary flags to represent if a given wire end has been disconnected or is noisy - These ends are excluded from the analysis.

- First stored waveform time bin - Each bin contains a stored voltage and time; this binned time is included to avoid the possibility of mapping the binned voltages to an incorrect drift time and thus an incorrect DOCA.

5.3.3.4 Dense Neural Network

The dense neural network calculates a DOCA estimate given a hit arrival time; this is the TXY function. This network is shown in Figure 5.13. Since it is trained on data, it can learn the true TXY function as a function of the input variables. In addition, this network can learn the bias due to ionization statistics.

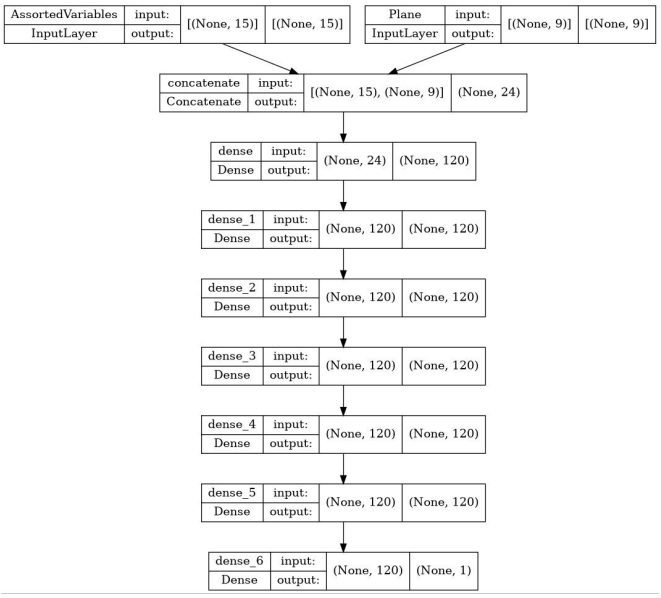


Figure 5.13: The dense neural network architecture is shown. The neural network estimates the time-distance relationship using the hit arrival time and tracking information.

5.3.3.5 Convolutional Neural Network

All available information about the time distribution of the clusters in a hit is stored in the digital waveform. Here, the neural network is trained to "learn" how to extract the useful information from the waveform in order to optimize the DOCA estimate. The convolutional layers take advantage of the temporal relationship between waveform bins. We use 300 bins of data per waveform (300/0.6 GHz $\sim$ 500 ns). The starting bin is 60 ns before the estimated track time (T_0) at the hit. The model is initiated by reshaping the binned waveform data into a 300x2 array such that, for each of the 300 time bins, there are two waveform values (upstream/downstream wire ends). The convolutional neural network is shown in Figure 5.15. To show how the waveform itself can be used to estimate the time-distance relationship, in Figure 5.14, we show the average voltage bin-by-bin in the waveform for several slices of track DOCA. The waveforms are alignment such that the waveform bin time equal to the track T_0 is set to zero for every hit. Each rise of the voltage roughly represents the drift time associated with the smallest track DOCA in each track DOCA window. In addition to just the drift time, the structure of the voltage in the waveform can be used to differentiate between different values of DOCA, for example the smallest track DOCA measurements on average contain more voltage $\sim$ 200 ns after the track T_0 than those with a track DOCA of $\sim$ 2 mm. We note that at least in MEG II, this is clearly a small effect and thus there will be difficulty in using this information to improve at the $\sim 100\mu$m level.

5.3.4 Results

Here, we compare results for the conventional DOCA estimator, the DNN, and the CNN. All results here use independent test data to avoid any training sample dependence. For all three DOCA estimators (NNs and the conventional approach), the hit DOCA estimates are re-evaluated after the first pass of tracking and then the tracks are refit. The first pass

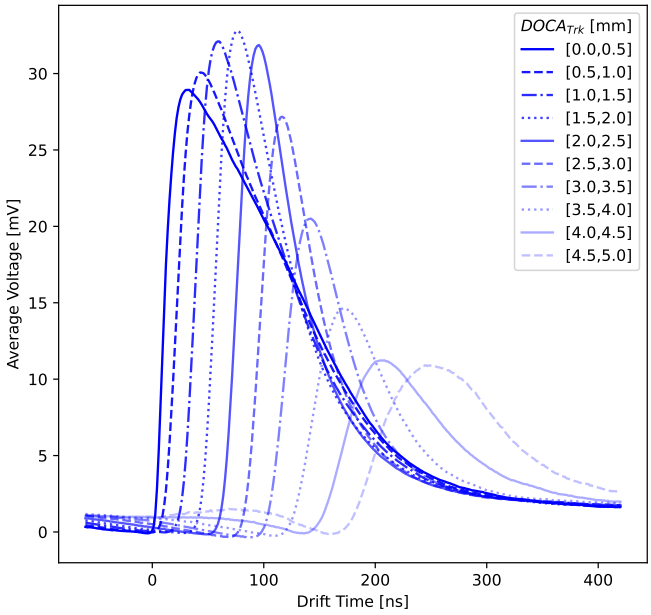


Figure 5.14: The average voltage bin-by-bin in the waveform for several slices of track DOCA. The waveforms are alignment such that the waveform bin time equal to the track T_0 is set to zero for every hit.

analysis uses the conventional analysis. We refit the tracks in all three cases for consistency.

The Kalman filter requires both a DOCA and a σ_{DOCA} estimate to fit the track. The σ_{DOCA} is used by the Kalman filter for both the DOCA uncertainty and for removing outliers. For all time-distance approaches, we use a 4th order polynomial to fit the squared DOCA residuals as a function of track DOCA to estimate σ_{DOCA}. We then scale the σ_{DOCA} constant parameter in the fit to achieve a comparable number of hits per track and a comparable χ^2/DOF (Figure 5.16). We use the same σ_{DOCA} for both neural networks. This yields nearly the same tracking efficiency with a fixed track selection; facilitating the comparison among the different approaches. The neural network σ_{DOCA} is $\sim 9\%$ smaller than that of the conventional approach. Alternatively, using the same σ_{DOCA} in the three cases results in a lower tracking efficiency, less hits per track and a larger χ^2 in the conventional approach. The results using the same σ_{DOCA} are compared in Appendix B.1.

The description of results of the neural network analysis applied to MEG II data are organized

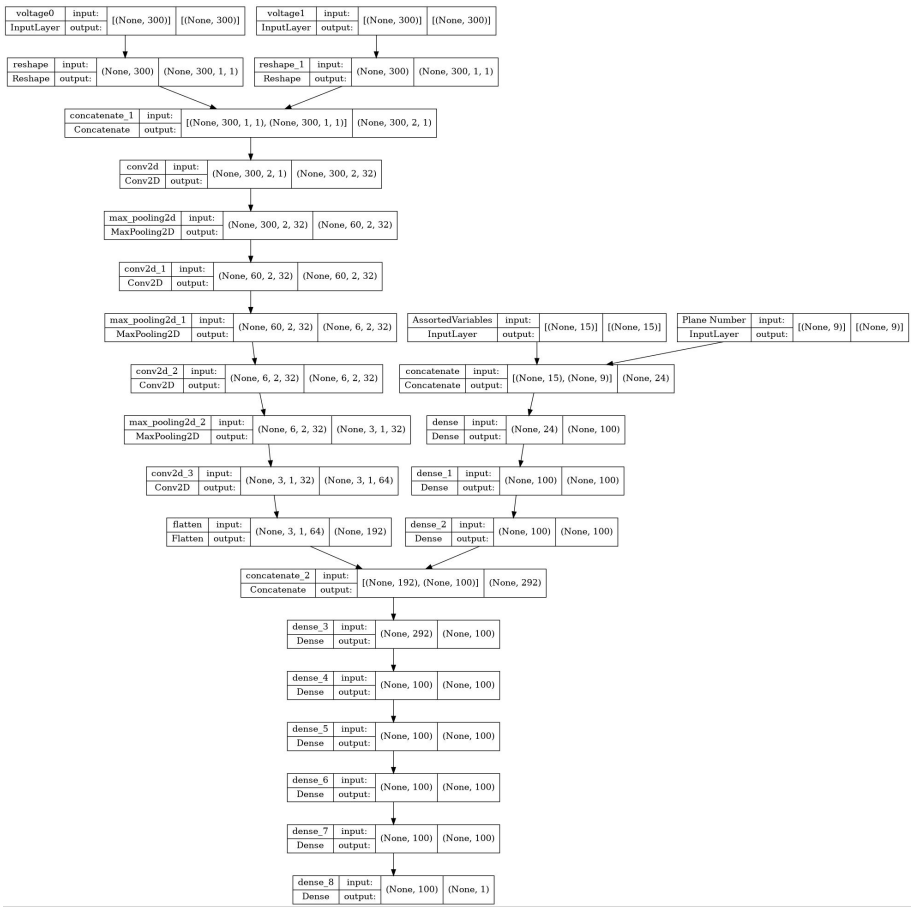


Figure 5.15: The convolutional neural network architecture. It estimates the time-distance relationship using the hit time and the CDCH digitized voltages.

as follows. Section 5.3.4.1 compares the hit residuals for the three techniques. Section 5.3.4.2 compares the Garfield++ TXY function to that of the DNN. In Section 5.3.4.3, a data-driven positron measurement resolution technique is used to compare the DOCA estimator's effect at the kinematic level. The full MEG II analysis is ongoing and improvements in hit resolutions and kinematic measurements presented here may change as improvements to the MEG II analysis (e.g. tracking, wire alignment, magnetic field mapping, etc.) are incorporated. The final MEG II detector performance will be given in [13].

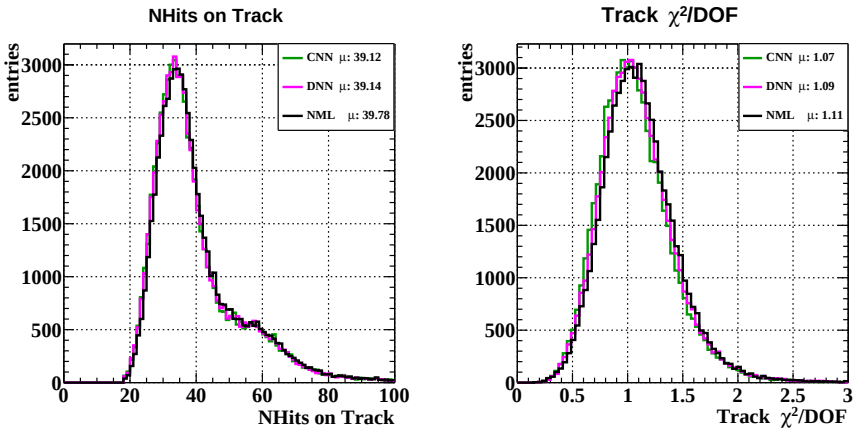


Figure 5.16: The number of hits per track and the χ^2/DOF are compared in the three techniques. The conventional DOCA estimate uses a larger σ_{DOCA} in the track fitting in order to get a comparable χ^2/DOF and number of hits. NML represents the standard or conventional DOCA estimation, DNN represents the dense neural network, and CNN represents the convolutional neural network that uses the full waveform information .

5.3.4.1 Hit-Level Results

Here we present and discuss the results at the hit residual level, shown in (Figure 5.17). The DOCA residual distributions for all approaches are shown top-left. As expected, all distributions have a positive tail due to the asymmetry created by the ionization statistics (track is closer to the wire than the first ionization site) and instances when the first cluster is not detected in the waveform analysis. Both neural network techniques result in less bias, with a slightly negative mode. The CNN has a slight reduction in the positive tail and a slight increase in the number of entries in the core compared to those for the DNN. We infer that the CNN uses the full waveform information to further suppress the ionization statistics bias, whereas the DNN is simply removing the average bias. In training, the CNN resulted in a $\sim 3\%$ smaller loss with respect to the DNN ($\sum_i |Y_i - \hat{Y}_i|$). In the top-right, we show the DOCA residual distribution near the wire (track DOCA $< 800\mu m$). Here there is the largest ionization statistics bias and also the largest improvement seen by the neural network approaches.

We show the root mean squared residuals as a function of DOCA in the bottom-left. To

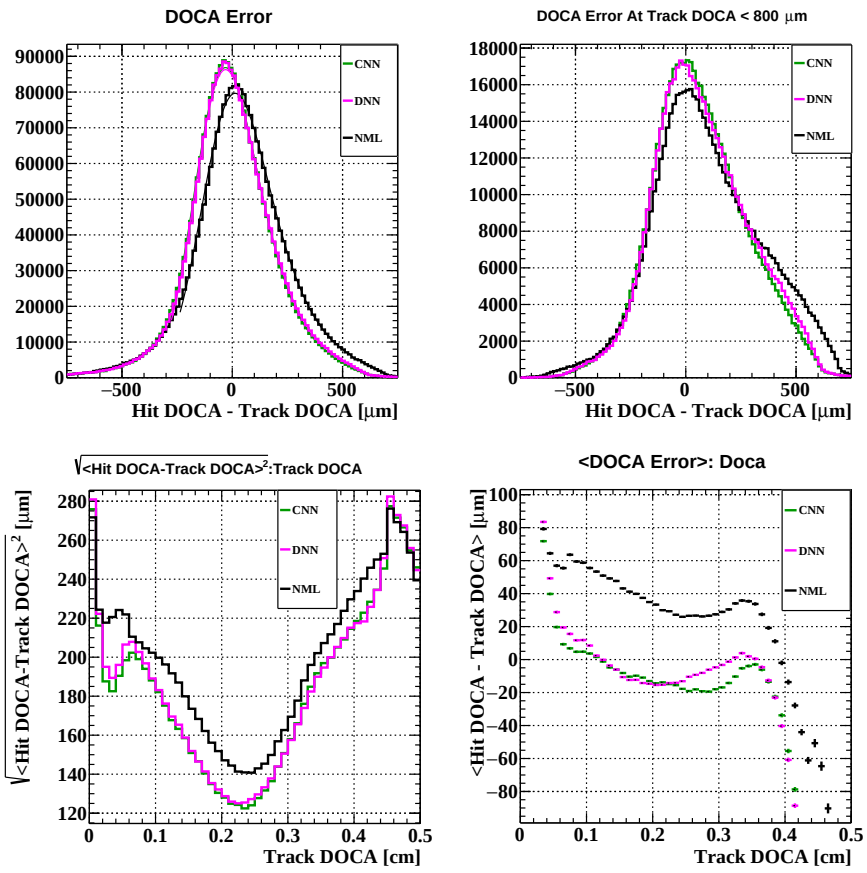


Figure 5.17: The hit level results for the three DOCA estimators. NML represents the standard or conventional DOCA estimation, DNN represents the dense neural network, and CNN represents the convolutional neural network that includes the full waveform information.

avoid the effect of outliers, hits with a residual larger than $500\mu m$ were excluded from the calculation. The improvement due to the two networks is largest($\sim 18\%$) close to the wire due to the larger ionization statistics bias there. The CNN does better than the DNN mostly close to the wire (track DOCA < 1 mm).

We show the ¡hit DOCA - track DOCA¿ (DOCA bias) as a function of track DOCA for the three techniques in the bottom-right. The non-zero mean DOCA bias in the conventional case is comparable to the DOCA bias expected from the MC from ionization statistics (Figure 5.12); this is a good indication that at least on average, we are correctly estimating the TXY function. In the CNN and DNN cases the ¡DOCA bias¿ has been highly suppressed in the middle of the cell, with a bias between $\sim [-20, 20]\mu m$ for all hits with track DOCA between

$\sim [0.06, 0.38] cm$.

Since both the hit and track DOCA are positive definite, when the track DOCA is close to zero, the only possibility is that the residual (hit DOCA - track DOCA) is positive. Hence, we expect a large bias simply due to this effect. The opposite is true at the edge of the cell. The neural network DOCA estimates should always return a value inside the cell to minimize the loss function and therefore we expect a negative ¡DOCA bias¿.

In Appendix B.2, we discuss whether a simpler technique can suppress the DOCA bias in the conventional TXY function approach by using the transformation: $t_{drift} \rightarrow t_{drift} - 2ns$. On average, this corrects the bias and gives a $< DOCAbias >$ similar to that of the the dense neural network. However, there are still fewer hits on tracks at small DOCA and a positive tail in the hit DOCA - track DOCA distribution with respect to the NN approaches.

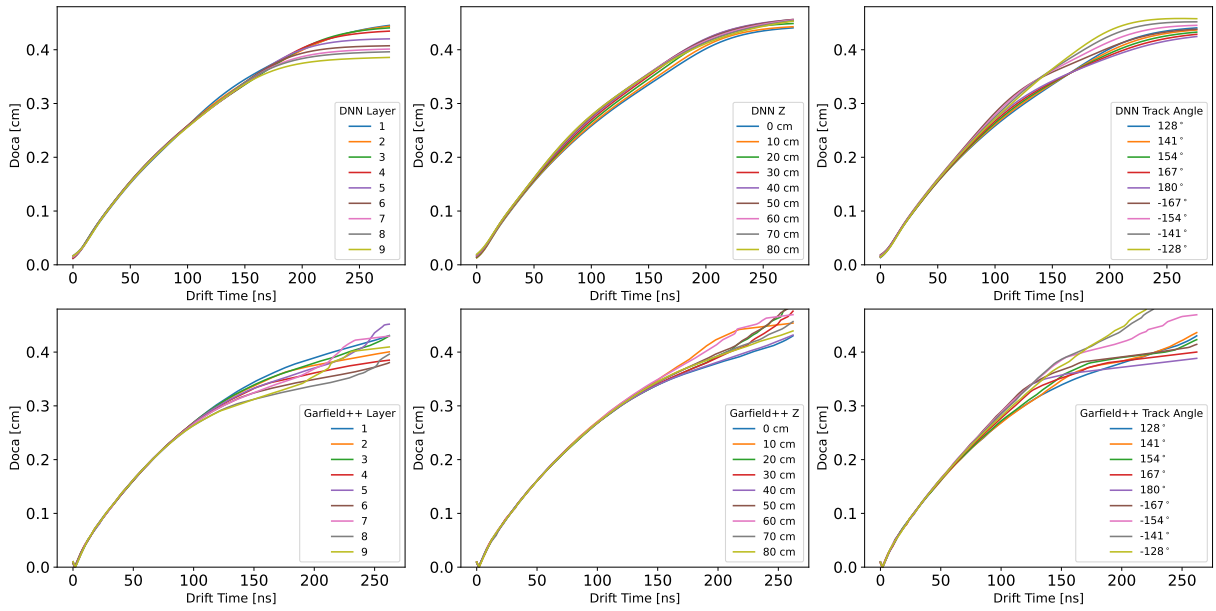


Figure 5.18: The time-distance relationship as a function of layer (column 1), position along the wire axis, (column 2), and track angle (column 3) are shown for the dense neural network (top row, DNN) and the conventional Garfield++ TXY function (bottom row, No ML).

5.3.4.2 Time-Distance Comparison

Here we compare the TXY function from the conventional Garfield++ calculation to that of the dense neural network as a function of the layer number, the position along the wire axis (z), and the track angle. This comparison is shown in Figure 5.18. Since the convolutional network requires digital waveforms to make a DOCA estimate, it's difficult to compare the TXY function in the same way. As a reminder, both layer number and position along the wire axis affect the cell size.

The Garfield++ TXY function results in non-smooth functions near the edges of the cell. This may be an effect of the positions of the cathode wires. The TXY function dependence on the CDCH layer is suppressed in the DNN case.

Comparing the DNN case to that of Garfield++, the hit's position along the wire has a larger effect on the TXY function at drift times around $\sim$ 100 ns, but a smaller effect at the edges of the cell. This implies that the dense neural network "learned" a different TXY function dependence on cell size than that simulated by Garfield++.

Finally, we observe that the TXY function dependence on track angle is significantly different at large drift times ($>$ 150 ns). This is likely the result of smoothing between the anode cathode geometry over many cells. An interesting point is that both the Garfield++ TXY and the DNN TXY show a significant difference between positive and negative track angles once reaching a drift time of $\sim$ 120 ns.

We conclude that the observed improvement in the root mean squared residuals and a suppressed DOCA bias at all track DOCA (Figure 5.17) indicates that the dense network produces the more accurate TXY function.

5.3.4.3 Evaluating the Effect on Positron Kinematic Measurements

Here we verify that the improved DOCA resolution improves the resulting e^+ measurements.

We use the following procedure to estimate the *relative* e^+ track measurement quality using the three DOCA estimators. In the MEG II experiment, $\sim 15\%$ of e^+ tracks pass through the drift chamber 5 times (9x5 layers intersected). An example double turn event is shown in Figure 5.57. For these tracks, we independently fit and measure the first turn (2x9 layers intersected) and the second turn (3x9 layers intersected). We then use a Kalman filter to propagate both state vectors (one forward and one backward) to the extended target plane between the two turns. Improved agreement between the kinematics implies improved resolution. This double turn analysis method was originally developed and implemented in the MEG I experiment[5].

These distributions are not the MEG II signal positron resolutions, but the resolution of the first and second turns of Michel positron tracks added in quadrature. The first turn only interests the chamber two times and thus has a degraded resolution with respect to the standard Michel track (three intersections). We fit each histogram to the convolution of two double Gaussians. We fix the two double Gaussians to be identical. Extracting out the eventual MEG II signal resolutions requires corrections from the Monte Carlo simulation. Nonetheless, this fit gives resolution estimates that can be used to compare the DOCA estimators.

All positron resolutions have improved when incorporating either neural network DOCA estimator. Here, the relative amplitude of the core and the tail of the two Gaussians are fixed for all DOCA estimators. When comparing the conventional approach to the convolutional neural network, the CNN core resolutions improve by 4.0%, 11.7%, 7.5%, 5.8%, 0.6% for y_e, z_e, ϕ_e, θ_e, p_e respectively. The uncertainties in the core σ are 0.0073 mm, 0.019 mm, 0.066 mrad, 0.076 mrad, and 1.16 keV for y_e, z_e, ϕ_e, θ_e, p_e respectively, well below the differences

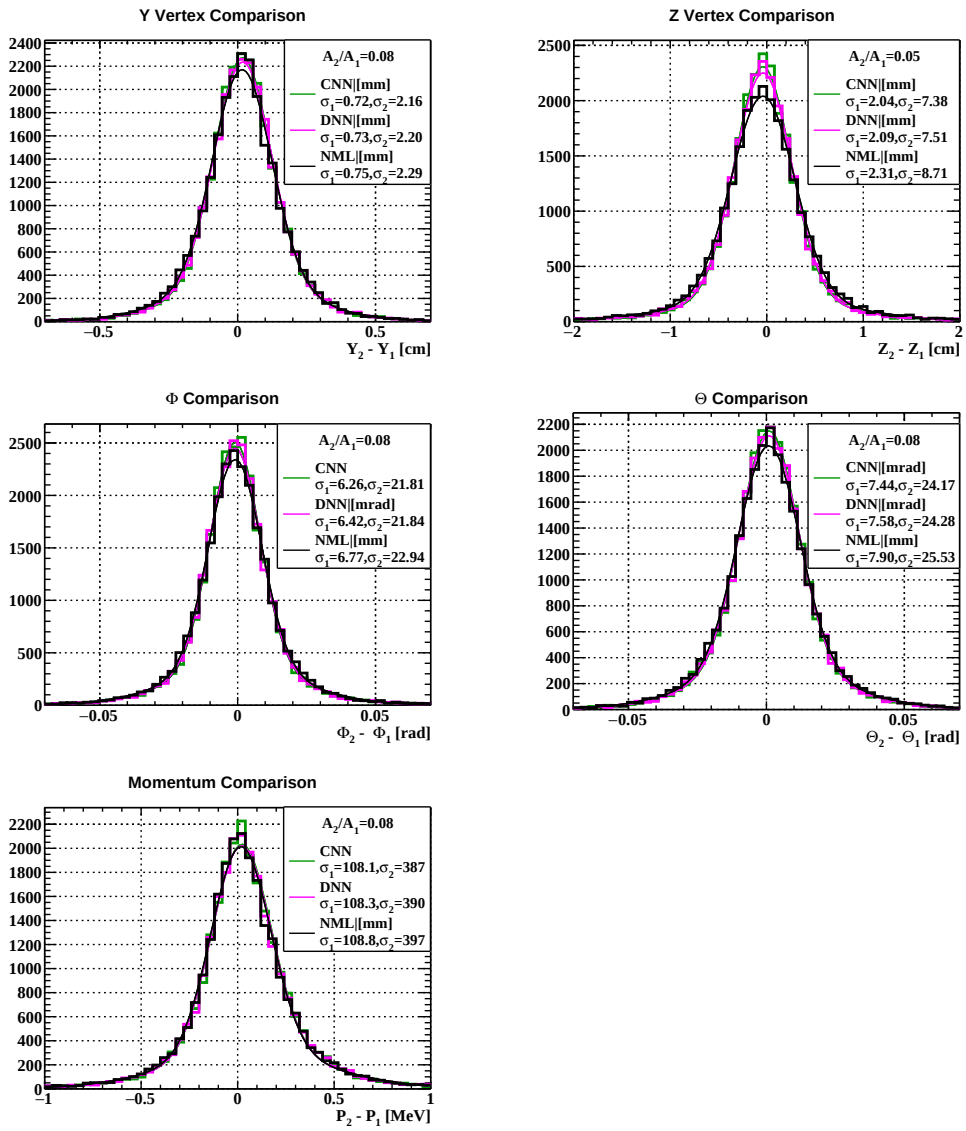


Figure 5.19: Here, we show the kinematic comparison at a common plane between two independently fit and measured turns of a single two-turn track. NML represents the standard or conventional DOCA estimation, DNN represents the dense neural network, and CNN represents the convolutional neural network that inputs waveform voltages. These distributions are preliminary and do not represent the signal positron resolution, but the resolution of the first and second turns of Michel positron tracks added in quadrature.

between the network resolutions and the conventional resolutions. In MC studies, improved DOCA resolutions had the largest fractional improvement on the z_e kinematic resolution; this supports our results, which also show the largest fractional improvement in z_e. The tails are also wider in all kinematic distributions in the conventional approach. In Appendix

B.1, we note that if the same σ_{DOCA} is used in all cases, the double turn analysis results in less entries in the peak and degraded resolutions with respect to the neural networks. We also mention that the double turn resolutions improve when using the CNN instead of the DNN. This implies that the network's use of the waveform results in an improvement at the kinematic level, though this is a minor effect.

5.3.5 Discussion

In this subsection, we discuss potential improvements to the technique.

5.3.5.1 Potential Improvements

We discuss potential improvements to the networks and their use in particle tracking software, some particular to the MEG II and some more general. In the MEG II application, the algorithm has access to the unique layer number through a hot encoded array (Figure 5.15). Alternatively, the neural network could input a hot encoded array containing the unique wire number. This would require significantly more training data (factor of ~ 100) in order to have the same precision i.e. creating the TXY function for each wire as a function of drift time, position along the wire, track angle, etc. This would allow the network to learn wire-specific subtleties such as the relative anode cathode geometry or signal/noise.

In some applications, a hybrid application using both the conventional and the neural network TXY functions might improve performance. This was tested for the MEG II case by including the Garfield++ based DOCA estimate as an additional neural network input variable. This did not yield significant performance improvement, but might lead to improvements for other experiments.

In the MEG II application we found that the convolutional neural network could achieve a

DOCA estimate without using the waveform analysis-driven hit arrival time but performance was worse than even that of the conventional approach. Achieving optimal resolution without the waveform analysis might be possible, but developing such an analysis is beyond the scope of this paper. It may require a multi-step neural network. For example, a first step that finds the hit arrival time and then a second step of converting that into a DOCA estimate.

Another possible approach to improve performance is to use a hybrid of conventional cluster counting (e.g. finding peaks in the waveform) with machine learning. This could involve adding a list of the waveform peak times (possibly not resolving all clusters) as an additional input to the neural network. Just as the waveform analysis hit time is required to have the optimal neural network, maybe the found peak times could improve the network further.

Finally, we return to the point that our training was optimized by using low intensity beam rate data. The improvement using low intensity data implies that the neural networks are not able to learn how to distinguish between the contributions from in-time and out of time (pileup) tracks to the waveform in the training process. One possible method to improve the performance is to implement a conventional pileup detection algorithm that produces a binary flag indicating evidence of more than one hit on the waveform. By developing separate machine learning algorithms for the two cases, the network might improve performance for the signal DOCA in the pileup case.

5.3.6 Conclusions

In this chapter, we describe the application of neural network models to improve the distance of closest approach measurement in drift chambers. The application results in a data-driven time-distance relationship, unlike conventional approaches. The data-driven approach accounts for all detector-specific properties (e.g. longitudinal diffusion, gas properties, ionization statistics, electronic noise, etc.). In an example application to the MEG II drift chamber

data, the networks improve the track position measurements with respect to those of the conventional approach. The optimal configuration is a convolutional neural network that uses the signal waveforms. This gives the network access to information from all ionization clusters, which is not conventionally used in DOCA estimation.

5.4 Tracking

5.4.1 Overview

In this section we describe the positron tracking algorithm. It entails grouping CDCH hits into tracks, fitting the tracks, and propagating them to the SPX and target. Note, here the primary objective is to maximize the number of hits on tracks, optimize the tracking resolution (DOCA), and maximize the tracking efficiency.

In the MEG II drift chamber, positron tracks intersect the drift chamber three, five, or seven times (1.5,2.5,3.5 turns respectively) before leaving the chamber's outer carbon fiber shell and intersecting the timing counter. The track finding task is to identify the CDCH drift cells intersected by the particle track on a single turn or half turn through the CDCH. The set of CDCH hit in the turn are matched to an SPX cluster, yielding a T_0 at each hit by correcting for the propagation time from the SPX to the CDCH hit. The TXY tables (and the machine learning algorithm described previously) are then used calculate 3D CDCH hit position estimates. The hits are fit to yield a preliminary estimate of the track state vector in the CDCH (i.e. the momentum and track position at each hit).

The track fitting stage uses a DAF filter which is an application of a Kalman Filter, to fit the CDCH hit positions and produce the best value of the positron state vector along the track. This DAF filter aids in solving the hit-by-hit left-right ambiguity discussed later. The

track fitting involves multiple updates to identify the maximum number of hits on the track, eliminate outliers that might be the result of noise hits or mismeasured hits, and derive the optimal position of the track at each hit. The final fitted track is eventually propagated to the target to estimate the kinematics $y_e, z_e, \phi_e, \theta_e, p_e, t_e$ and the Kalman covariance matrix at the target.

5.4.2 Positron Track Trajectory

First, we briefly discuss the typical trajectory of positron tracks.

A majority of positron tracks (signal momentum) do not reach the carbon fiber outer shell on the first "turn" through the CDCH due to the higher magnetic field (maximum near the target, dropping off in $|Z_{MEG}|$). On the second turn, the track is at larger $|Z_{MEG}|$ and thus reaches a larger radius. The typical positron track leaves the target region, reaches maximum radius, returns to small radius, then exits the chamber and intersects the timing counter (a total of ~ 50 drift cells). About 5% intersect the carbon fiber shell on the first turn; these tracks are difficult to reconstruct with precision and are mostly rejected.

Tracks close to $\theta_e = 0$ take more turns prior to intersecting the SPX. About 14% of tracks make 2.5 turns and $< 1\%$ of the tracks make 3.5 turns. In Figure 5.20, we show the $z_{e,TGT}$: $\theta_{e,TGT}$ distribution for reconstructed 1.5,2.5,3.5 turn tracks (black, purple, red respectively). At large $|z_{e,TGT}|$ multiple turns are achieved when the positron reverses its axial momentum (i.e. starts near the downstream end of the target going upstream, reverses direction in the graded field and reaches the downstream SPX).

The prominent band of events with negative slope in the distribution corresponds to physical 2.5 turn tracks missing the first turn; this was an effect observed in MEG I and can be used to estimate the number of double turn tracks with a missing first turn (this is an inefficiency).

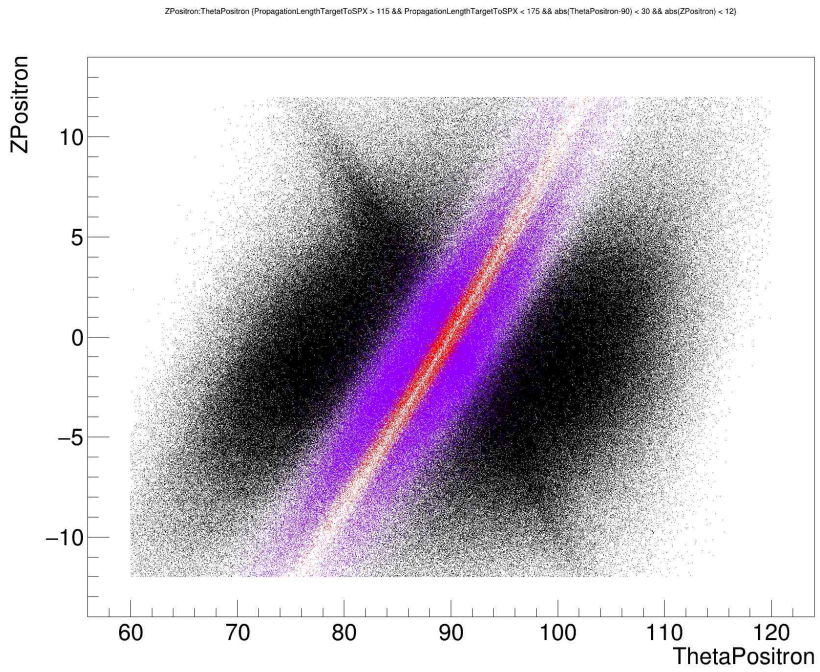


Figure 5.20: The $z_{e,TGT}$ [cm]: $\theta_{e,TGT}$ [°] distribution for 1.5,2.5,3.5 turn tracks (black, purple, red respectively). The black line with a negative slope was found in MEG I to represent 2.5 turn tracks missing the first turn.

5.4.3 Track Finding

5.4.3.1 Initialization

The track finder is initiated by importing the set of CDCH hits and the SPX clusters (set of SPX tiles intersected by a single positron). CDCH information consists of the wire number, arrival time of the first ionization cluster (hit time) on the wire and the hit position estimate along the wire axis (z). The hit time requires a wire-by-wire time calibration described in Chapter 7. The hit arrival time is converted into a DOCA estimate at this stage using the conventional TXY assuming a T_0 based on the mean trigger time. Clearly, at this stage, the DOCA resolution will be very degraded due to the large errors in the T_0.

The z estimate (z_{local}) uses a combination of two estimates from the time difference in the arrival of ionization clusters on the two wire ends and the difference in charge on the two

wire ends. The z estimates require calibrating the two ends of the wire, described in Chapter 7.

At this stage, the DOCA can be converted into a x_{local}, y_{local} position estimate by asserting the tracking angle to be the average track angle of a positron track in the cell. The position, angle and sagitta of each wire is required to transform the local hit position into the global MEG coordinate system; the alignment procedure is explained in Section 5.5.

The track finder also imports the SPX cluster time and the position of the first intersected SPX tile (ϕ_{SPX}, z_{SPX}). The track finder additionally imports the magnetic field map, but not the geometry file including the detector and target materials. This is only used to achieve a rough momentum estimate that will be replaced by the track fitting task.

5.4.3.2 Track Finding Algorithm

This process is describe in more detail here[14]. After importing the CDCH hit and SPX cluster information, the track finder first groups them into turns. Due to the magnetic field, low momentum positron tracks ($\sim<$ 45 MeV) do not reach the outer layers and so it is intersected by less positron tracks. This makes the outer layers ideal for initializing the tracks. A cluster of CDCH hits in the outer layers is used to "seed" a positron track "candidate".

This seeding is a difficult combinatorics problem as we still have a very degraded DOCA resolution and the z resolution is $\sim<$ 10 cm. The z resolution of the hits is extremely useful for suppressing the combinatorics (distinguishing between two tracks with similar hit arrival times).

We replace the nominal DOCA estimate with a reasonable approximation of the DOCA to help find an eventual track. This requires a T_0 estimate better than the nominal trigger time,

which could be displaced by as much as 10 ns from the T_0 at the cell. At a drift velocity of 20μm/ns, a 10 ns T_0 error would result in a 200 μm DOCA error, larger than the resolution.

The current track finding algorithm does not contain a fit for the T_0 using the CDCH hits. Instead, the finder relies on the SPX cluster time. For each SPX cluster, we try setting the T_0 at the CDCH hits to $T_{0,SPX} - F_i$, where F is a set of fixed offset corresponding to a typical offset between the SPX cluster time and the CDCH turn (F=[0.6, 4.5, 8.5] ns). The best offset yields the lowest errors between the track and hit DOCA estimates. Using the correct "best" offset, this simple T_0 approximation should results in maximum errors of the order $\pm 2ns \sim 70\mu m$. This is sufficient for the track finder stage and will be replaced by a more precise measurement in the fitting stage.

The track searches for hits nearby the cluster using a "road width", that is, using a rough estimation of the track kinematics at the seed, the algorithm extrapolates the seed towards lower radius and searches for hits along the trajectory within the road width consistent with the track seed. If the hit has a DOCA and a z consistent with the seed, the hit is added to the track candidate. Note, the track kinematic resolution is improving as this procedure is continued.

The end result is a track candidate that contains a list of all hits on the turn and an estimate of the track state vector $(y_e, z_e, \phi_e, \theta_e, p_e)$ at the first and final hit on the turn. Note, a large fraction of the turns will be the final half turn (moving from r=0 until hitting the SPX at maximum radius). These half turns have significantly degraded resolution and will be discussed later.

5.4.3.3 Status

The track finder was written before the CDCH was ready for data taking (written before 2019). The signal and noise situation has changed dramatically since that time (expected

that the CDCH signal/noise ratio would be significantly higher). However, many of the parameters had not been tuned. In addition, through trial and error, the tracking efficiency has improved significantly by removing hits with a hit arrival time before the nominally trigger T_0 (i.e. known pileup hits). The tracking efficiency also significantly improved by including the ML hit finding algorithm mentioned in Section 5.2 that resulted in significantly more hits/event. Both changes to the hit distribution resulted in track finder efficiency improvements of the order $\sim 10 - 20\%$, neither of these were expected and it is unclear how much higher of a track finding efficiency could be achieved. In addition, the track finder CPU time is a very large fraction of the overall MEG II CPU time ($> 50\%$), which is not ideal for reprocessing the data. For these reasons, one of the main focuses on the CDCH analysis in future years will be on modifying the track finder and developing alternate track finders that may yield significant efficiency improvements.

5.4.4 Track Fitting

In this subsection, we describe the track fitting procedure. As mentioned, the procedure uses a DAF to fit the positron tracks. The DAF is used to solve the CDCH hit left-right ambiguity problem, that is, given a hit DOCA and a track angle, there are two unique solutions to the hit x_{local}, y_{local} coordinates. In general, this increases in difficulty as a hit is closer to the wire (two solutions are very similar).

The task is initiated by fitting single and half turns output from the track finder. The task then merges the turns into a full positron track. At each stage of track fitting, the track angle at the intersected drift cell is known more precisely and therefore the DOCA can be improved further by reevaluating the DOCA estimate. Similarly, the propagation length from an intersected drift cell to the SPX and thus the T_0 at each drift cell is known more precisely. We additionally update the DOCA accordingly and refit the track.

At the end of the track fitting, we propagate the positron track from the CDCH to the target and to the SPX. This yields the final kinematics at the stopping target $(y_e, z_e, \phi_e, \theta_e, p_e, t_e)$. Note, the target is thin film subject to deformations O(1 mm); the deformed target has been implemented as described in Chapter 4.

After the track has been fit, we search for hits nearby by the positron state vector, but are not on the track. If these hits are found, they are added to the track and the track is refit.

5.4.5 Updating DOCA using T_0 and Track Angle

After fitting the turns, the tracks are propagated to the first intersected tile in the SPX cluster used to estimate the T_0 (SPX seed). If the track successfully propagates to this tile, the SPX cluster and the track are considered a match if the distance between the position of the track on the tile and the SPX hit position estimate is less than 10 cm and the track propagates within 3 cm of the physical timing counter tile. This is a rather loose selection criteria. However, this requires propagating through the drift chamber outer carbon fiber shell where the tracks can scatter, adding uncertainty to the trajectory. In addition, some tracks without the final half turn can be propagating $\sim 80cm$ to this SPX match.

If this CDCH-SPX matching is a success, we set the T_0 at each CDCH hit to the time of the SPX cluster minus the propagation time from the intersected wire to the SPX. This is successful for $\sim 99\%$ of positron tracks that pass our standard selection criteria.

We then recalculate the hit DOCA with this improved T_0. At this stage, the T_0 is known extremely precisely ($< 100ps$) and therefore completely eliminates its contribution to the DOCA resolution. Each time we re-calculate the drift time as the arrival time of the first ionization cluster - hit T_0 and convert the drift time into a DOCA estimate using the time-distance relationship. We additionally input the updated track angle at the drift cell into

the DOCA estimator to improve the resolution further.

Note that some hits contain a negative drift time (arrival time is before the T_0). On a hit-by-hit basis, the cause is unclear. The main causes are mistaking noise on the waveform as an early cluster or picking up the signal from an earlier cluster due to a pileup track. Currently, we simply assert that the drift time is equal to zero. This is not fully optimized.

This update is repeated before and after the track turn merging procedure that is described in the next section.

5.4.6 Positron Turn Merging

For the $\sim$ 85% of tracks than contain 1.5 turns, the first full turn and the final half turn need to be merged to achieve the optimal precision. Additional turns are merged for longer length tracks.

To decide which turns should be merged, the turns are propagated to the closest point to the beam axis (r=0) in both directions. Then, in a double for loop over all turns, the kinematics of the two turns at the beam axis are compared loosely. The comparison requirement is very loose (e.g. $|p_2 - p_1| < 8$MeV and $|t_2 - t_1 < 3$ ns, $|z_2 - z_1| < 12$ cm). The requirements mainly exist to avoid attempting to merge poorly measured tracks or attempting to merge pileup tracks.

If the comparison succeeds, then one of the tracks is propagated slightly further to the plane perpendicular to the other's state vector (at $r = 0$). The kinematics are compared using a χ^2 incorporating the covariance matrix elements of both turn state vectors and the kinematic comparison. If the χ^2 is below a threshold ($\chi^2/DOF < 150$) the merging is attempted. Regardless if the merging was a success or a failure, the two original non-merged tracks and the final merged track are stored in the track folder. Therefore even if the merging fails,

there is still a full first turn and other methods to recover the final half turn.

Unfortunately, this technique often fails to successfully merge the final half turn, this is likely caused by the difficulty in estimating the momentum of a half turn track. The track only propagates ~ 7 cm or 60° and can have as little as 10 hits.

5.4.7 Positron Refit Algorithm

After the turn merging procedure, we still observe a large fraction of tracks that are missing the final half turn. This is a combination of difficulty from the track finder finding the half track (less hits in the "seed" region) and the difficulty of merging low quality half turns. This loss of the final half turn and hits in general motivates a search to find additional hits that are associated with the positron track. This algorithm is described in this subsection.

The algorithm propagates the positron track from the first CDCH hit to the intersected SPX cluster searching for intersected drift cells that should be on the track, but are missing. This falls into two main categories: hits that were not added by the track finder (e.g. due to a bad z measurement) or hits on additional turns (full or half turns) not properly merged.

5.4.7.1 Track Propagation to Search for Intersected Wires

The algorithm starts by propagating the track between the first CDCH hit and the SPX cluster in time steps (3 mm or 10 ps); this distance is appropriate as the mean cell diameter is roughly 8 mm. The only hits missed by the algorithm are those with large DOCA where the track only intersects the edge of the cell.

To have the best track estimate at each time step, we determine the closest hit (in T_0 space) to the given time step. We propagate from the state vector at the closest hit to that time

step. If the propagation length is less than 5 cm, we assume we are searching for a hits on a new turn. If we are searching for hits on an existing turn, our track position estimate is significantly more accurate and thus we apply tighter criteria for determining if the hit should be added to the track (discussed in detail later).

At each time step, we determine the closest drift chamber layer. At each time step, we search for a wire in the layer we such that the distance between the track position and the wire position at this Z_{MEG} coordinate is smaller than ~ 6 mm. If this is satisfied, we flag this as a hit candidate.

If the hit is considered on the track, we propagate further to the wire POCA (point of closest approach). All the details of the propagation e.g. the 3D position at the POCA, T_0 at the POCA, track angle, propagation length from the hit to the wire, etc. are all saved.

5.4.7.2 Matching Hits in Preexisting Hit Folder

For all hit candidates, we determine if there is an existing hit in the CDCH hit folder (found and saved by the waveform analysis) with a comparable position to the hit candidate.

To determine if there is a match, we compare the DOCA and z of the track (hit candidate) and hit. We use the hit candidate T_0, the hit arrival time, track angle, and the conventional TXY table to convert the hit time into a DOCA estimate. If both the DOCA and z conditions are met, we add the hit onto the track. If there are multiple hits on the same wire that pass these thresholds, we add the hit with the smallest difference in DOCA.

5.4.7.3 Refit

After adding all hits, we refit the track. We assign the initial value of the state vector at each hit to be the point of closest approach with the calculated momentum vector (previous

section), and a dummy covariance matrix. We sort all hits on the track by the T_0 at the hit.

Similar to the merging algorithm, both the "old" and the "refit" track are saved. Here is another instances of populating the track folder with multiple reconstructed tracks associated with the same physical track (ghost tracks). Of course in general, the one with the best number of hits, χ^2/DOF, etc. is the "best" version. Selecting the best ghost track to go into the final physics analysis is described in Section 5.6.

5.4.7.4 Sensitive Parameters

The analysis is most sensitive to the DOCA thresholds for adding hits onto existing turns/ new turns. These have been tuned using data using the following procedure.

First, we set the thresholds to be larger than expected. We separate the hits added to the track into two categories: hits successfully added to the fit and those rejected by the fit. For these two categories, we then histogram: (hit DOCA - hit candidate DOCA) in Figure 5.21. The hit candidate measurement is effectively the track DOCA (but prior to the track fit incorporating the new hit). We additionally show the histogram of (hit DOCA - track DOCA) for hits added to the fit after the fit is complete.

In general, we'd like to maximize the number of hits successfully added to the fit and minimize the number of hits rejected by the fit. Note, if too many hits are rejected, the Kalman occasionally fail completely, and thus the refit fails. On the other hand, forcing tighter criteria results in rejecting some hits that should be on the track.

The result suggests that for hits on new turns even an error of ~ 1 mm prior to the refit has a significant probability of being added to the track. Whereas if the hit is on an existing turn, it should only be added if the error is less than $\sim 500\mu m$.

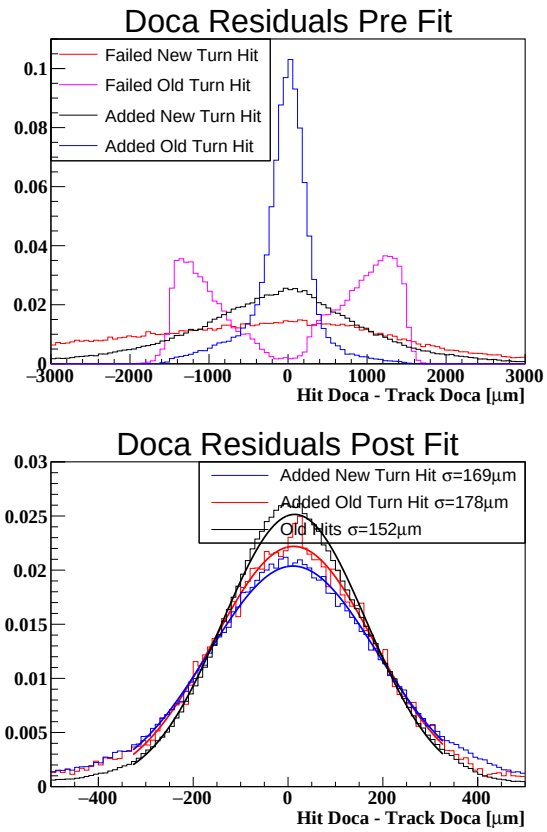


Figure 5.21: The DOCA residuals before and after the fit routine. The data is sliced based on two binary categories hits successfully/unsuccessfully added to the fit and whether the hit is on a new turn or an existing turn.

5.4.7.5 Results

In this subsection, we compare the measurables for tracks with and without the refit algorithm applied. In instances when the refit fails, we are showing the non-refit version of the track.

In Figure 5.22, there is a category of tracks where the refit adds a very small amount of hits. This is due to the track already having the full 1.5 turns or instances when the refit effectively fails. The other category is tracks where ~ 10 hits are added. These are instances where a large fraction of the final half turn is recovered.

This is verified by the change in the propagation distance between the last hit in the CDCH

and the SPX hit. After the refit, a very large fraction of the tracks have some hits on the final half turn. Note, in the final analysis of the 2021 data, there are still some instances with a propagation length to the SPX greater than 1 ns due to tightened fitting criteria. In particular, we tightened the σ_{DOCA} for all hits due to improved resolution observed in the double turn analysis (Section 5.7).

We then show the Kalman estimated momentum resolution. The momentum resolution is two peaked without the refit algorithm; this is the difference in momentum resolution for single turn tracks and tracks with one and a half turns. With the refit, the degraded resolution peak is eliminated, but some tracks still have a core resolution greater than ~ 100 keV.

Finally, in Figure 5.23 we show the Michel edge fit using the "best" track without the refit task and the "best" track with the refit task for each event in the final 2021 dataset. The ranking of tracks is based on the ML track ranking described in Section 5.6. We use the standard selection criteria except that we loosened the requirement for the distance between the final hit in the CDCH and the SPX cluster to be less than 150 cm. This avoids a difference in efficiency (many tracks without the refit have a long propagation to the SPX as they are missing the last half turn). The main difference between the two Michel edge fits is the sigma1 parameter or the core uncertainty ($101 \rightarrow 87$ keV) or an improvement of 14%. We also observe sigma2 is lower by $\sim 10\%$. Roughly, looking at the covariance matrix plot in Figure 5.22, we see a $\sim 15\%$ improvement, which is consistent with the Michel edge improvement.

5.4.8 Tracking Efficiency

The tracking efficiency is difficult to measure in data as there are triggers based an SPX hit that is not actually associated with a high quality positron track. Without any selection

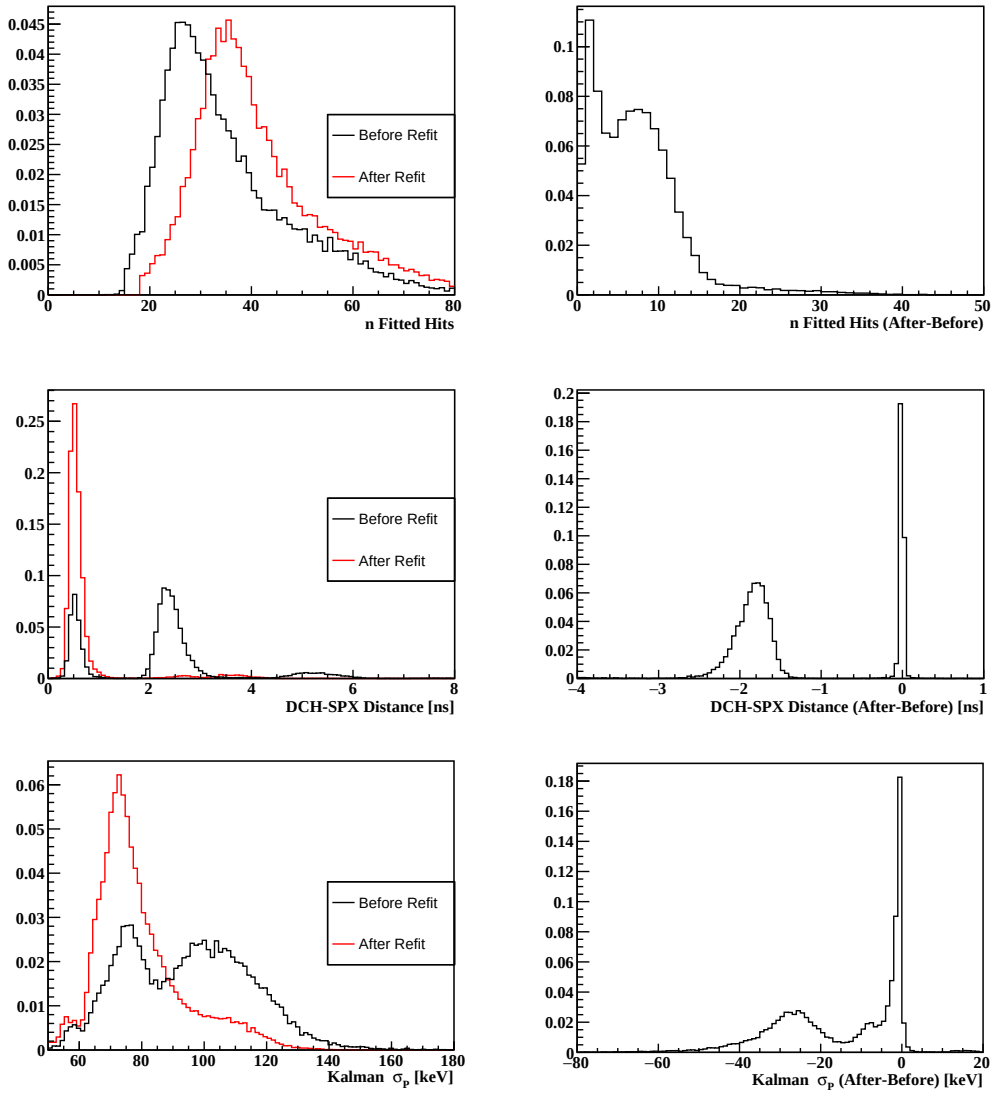


Figure 5.22: Comparison of the track kinematics of all tracks with and without the refit algorithm. The left-hand side shows the results with and without the algorithm and the right-hand side shows the relative improvement. The top row is the number of hits on the fitted tracks, the middle row is the propagation length from the CDCH to the SPX detector, and the final row is the Kalman estimated resolution.

criteria, the mode of the number of SPX hits in a cluster is 1 for a physics trigger whereas a high momentum positron should have ~ 9 on average.

Note that the normalization of the MEG II dataset and thus the experimental sensitivity does not require a precise measurement of the positron tracking efficiency, but only the

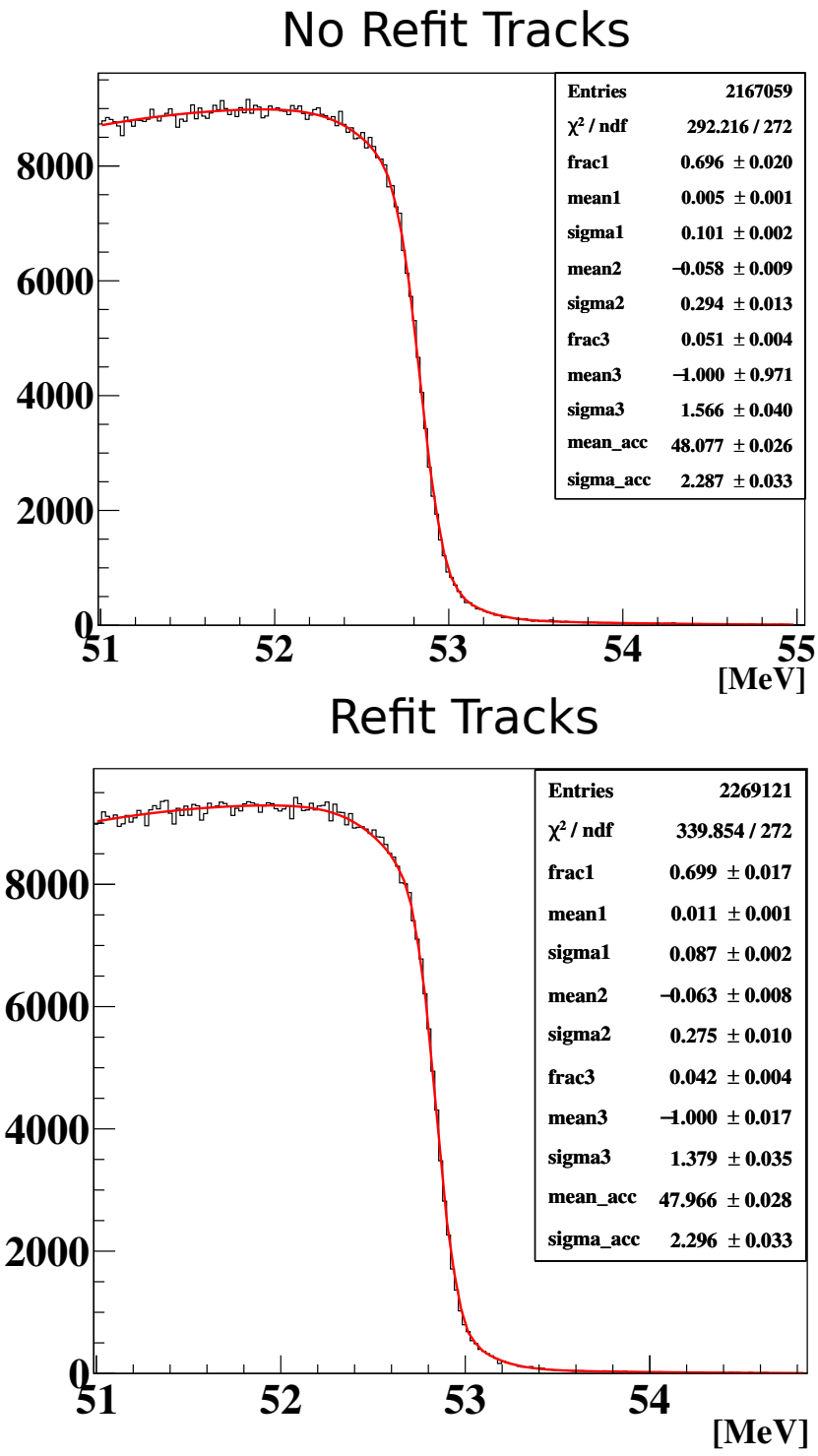


Figure 5.23: The Michel edge fit used to estimate the momentum resolution for tracks that successfully refit compared to those without any refit.

relative efficiency of 52.8 MeV positrons and positrons greater than 50 MeV (this is a minor correction). The normalization is described in Section 8.10.

Although not needed for the normalization, of course it is critical to optimize. During the optimization of the waveform analysis, track fitting, TXY, etc. we calculate a "tracking trigger efficiency" where the numerator is the number of events with at least 1 high quality positron track in the signal region and the denominator is the number of physics triggers. By maximizing this tracking trigger efficiency, we can optimize our analysis without knowing the absolute efficiency.

We can also rely on the Monte Carlo to estimate the absolute tracking efficiency for signal positrons. Here, this efficiency is such that the positron is in the "signal region" i.e. back-to-back with the calorimeter and the positron has a momentum of 52.83 MeV. This numerator only requires that the positron is inside the final analysis region described in Section 8; these are very wide cuts. This results in a positron efficiency of 52.4%, however this does not include the ML hit finding algorithm that resulted in a $\sim 28\%$ efficiency improvement (this algorithm was not tested on the Monte Carlo). This suggests a signal efficiency of $\sim 67\%$.

In addition, the efficiency has been measured using an independent trigger, triggering only on a single hit in the pixelated timing counter. Here, we measure the total number of positrons reconstructed in these triggers with a momentum greater than 50 MeV. We then divide this by the total number of expected positrons greater than 50 MeV (this includes several corrections including the trigger efficiency, the branching fraction for Michel positrons with a momentum greater than 50 MeV, the beam rate, etc.). We then make a minor correction due to the fact that signal positrons have a higher efficiency than Michel positrons with an energy greater than 50 MeV (~ 1.09). This results in an efficiency of $\sim 70\%$ that is dependent on the beam rate. This shows good agreement with our MC result.

5.5 Wire Alignment

5.5.1 Overview

This chapter describes the iterative track-based alignment technique using the 2021 data. The alignment corresponds to CYLDCHConfid=41 (MEG database entry). We discuss the procedure, shortcomings, the resulting residuals, and the required translations, rotations and sagitta with respect to the survey position. The iterative-track based wire alignment is compared to the survey alignment by comparing the mean residuals and the double turn analysis.

5.5.2 Motivation

The CDCH global position was measured in both 2021,2022,2023 using an optical survey. Additionally, the wire-by-wire alignment of the chamber with respect to the CDCH global position was calculated from a variety of measurements spanning back to the construction of the chamber.

After incorporating all survey information into the wire-by-wire alignment, systematic errors in the hit residuals wire-by-wire were still observed of the order $\sim 100\mu m$. The wire-by-wire residuals are shown in Figures 5.24 and 5.25 in the X and Y global coordinate systems respectively. These systematic errors degrade the overall CDCH resolution (in addition to others e.g. TXY tables, magnetic field alignment, etc.) or create systematic biases in the positron kinematics $(p_e, \theta_e, \phi_e, y_e, z_e)$. This motivates improving upon the survey alignment. We discuss the use of Michel positron tracks to align the drift chamber wires.

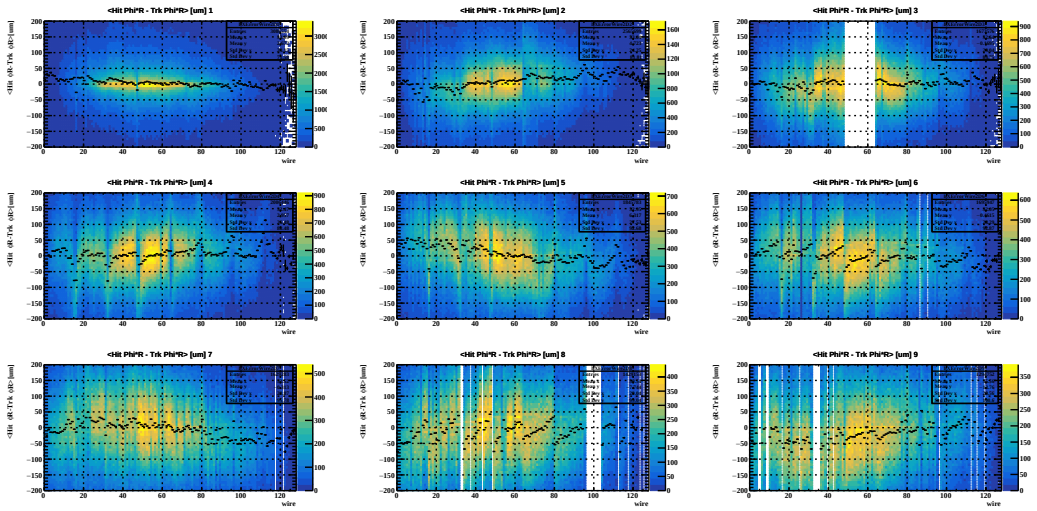


Figure 5.24: (Hit - Track) X global errors per wire using the survey alignment.

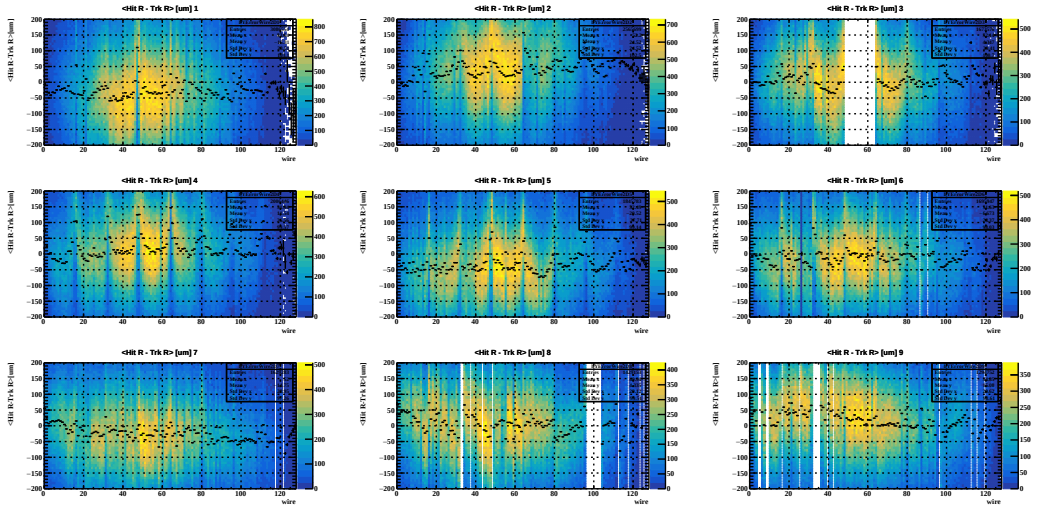


Figure 5.25: (Hit - Track) X global errors per wire using the survey alignment.

5.5.3 Track-Based Alignment

In this section, we describe the procedure used to improve the alignment of the drift chamber. To be clear, the track-based alignment is only used to improve the *relative* wire-by-wire alignment.

First, Figure 5.27 illustrates how tracks can be used to align the drift chamber. The solid and dotted X represent the true and misaligned wire position respectively. The solid and dotted

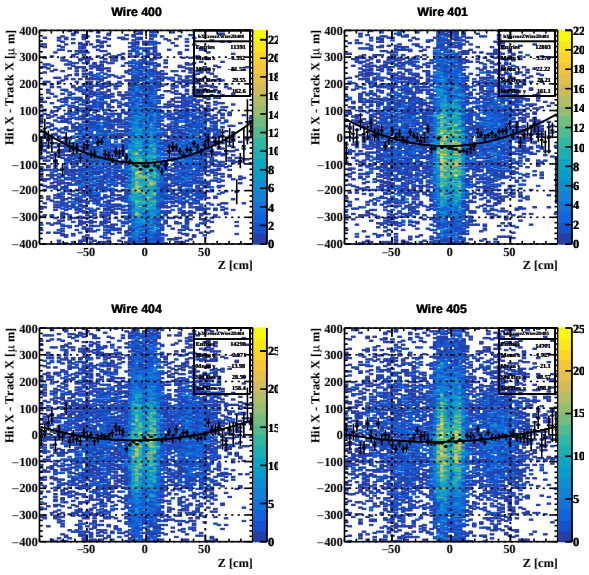


Figure 5.26: Sample of the X global residuals as a function of the position along the wire axis per wire using the survey alignment. Here, wires 400-415 are shown (L2S1).

circle represents the hit distance of closest approach isochrone (approximated as a circle). The misalignment creates a systematic residual in the direction of the misalignment for all track angles except that parallel with the misalignment direction (case D). By iteratively adjusting the wire coordinates based on the mean residual, we iteratively improve the mean residuals and thus the wire alignment.

The general approach is to fit the mean residual per wire as a function of the position along the wire axis (z). An example of the mean residual is shown in Figure 5.26. The residual plots in X,Y, and Z global are fit using the equation below.

$$p_0 + p_1 \cdot z + p_2 \cdot [(z/\frac{L}{2})^2 - 1] \tag{5.1}$$

The p_0, p_1, and p_2 correspond to a translation, rotation, and a wire sagitta respectively. This is the sagitta equation used in the MEG II software. The sagitta equation conveniently sets

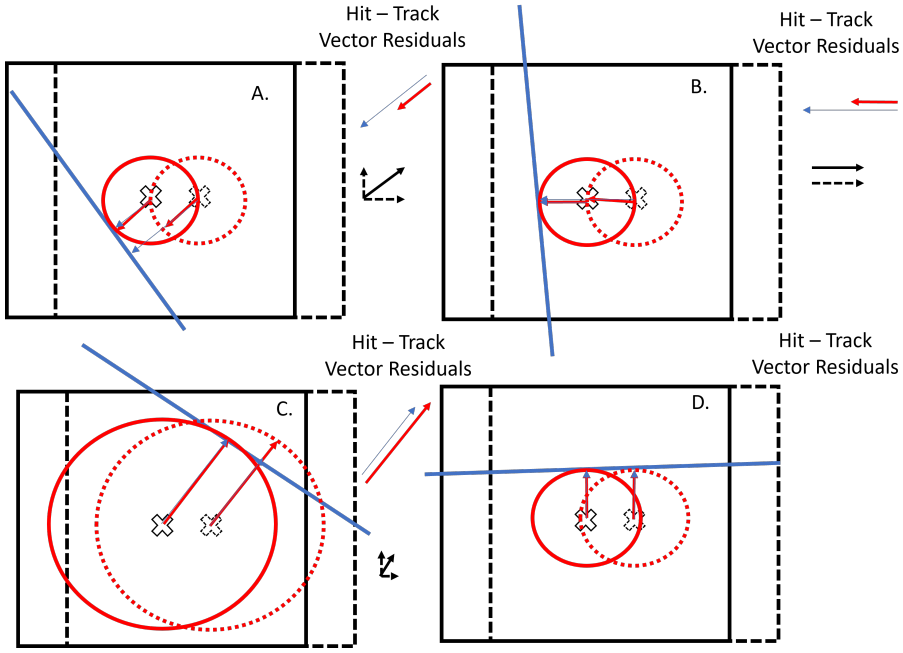


Figure 5.27: Four example tracks with the same misalignment. It's clear that in all cases except when the track is parallel to the misalignment, the misalignment creates a mean residual in the direction of the misalignment. This mean residual is used to align the track.

the sagitta to p_2 at z=0 and 0 at the wire edges ($\frac{L}{2}$). The sagitta is calculated in the X and Y global coordinate system and converted into the local coordinate system; this has a small and negligible error by not taking into account the global Z error due to the sagitta.

The sagitta can be due to electrostatics and gravity. In Figure 5.26, some wires have a sagitta as large as $\sim 100\mu$m.

In the MEG II software, the angles θ_3 and ϕ_3 are stored in the database to define the wire direction. The direction is obtained by the following transformation.

$$\begin{bmatrix} Dir_x \\ Dir_y \\ Dirz \end{bmatrix} = \begin{bmatrix} 0 & 0 & S(\theta_3)C(\phi_3) \\ 0 & 0 & S(\theta_3)S(\phi_3) \\ 0 & 0 & C(\theta_3) \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{5.2}$$

5.5.4 Limitations

As mentioned in the previous section, the method relies on tracks that are *not* parallel with the wire misalignment in order to align the wire. As shown in Figure 5.27, tracks parallel to the misalignment cannot be used to align the wire.

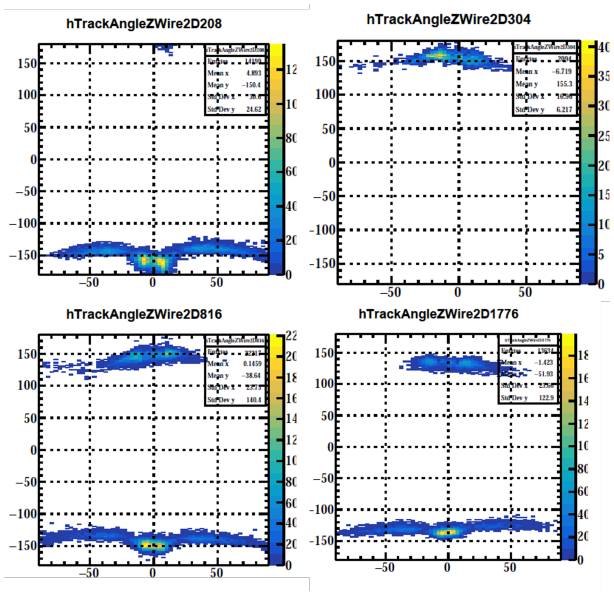


Figure 5.28: Examples of the track angle as a function of wire number and the wire axis (z).

Ideally, we'd use tracks incoming at all angles. However, we are using MEG triggers which often have only a limited range of track angles in a given cell. Some example incident track angles wire-by-wire are shown in Figure 5.28. There is clearly a large variation in the track angles as a function of wire number. This is best explained alongside Figure 5.30. The MEG trigger requires the positron to be nominally back-to-back with a γ intersecting the LXe detector. Therefore, in small sector numbers (positive X or the right side of the graphic), the track must be on the first half turn or the third half turn. e.g. wire 208 (top-left in Figure 5.28) are incident at $\sim -170°$ and z=0 and $\sim -140°$ and large —z—. Even this 30° difference is enough to align the wire in all directions. In the central sectors (e.g. wire 816, bottom-left in Figure 5.30), tracks intersect the cell on all three half turns. This results in three distinct track angles: $-150°, -130°, 150°$. This is the optimal case as $-130°$ and 150°

track angles are separated by 80° and thus almost perpendicular to one another. Whereas at large sector numbers (e.g. wire 304 - top-right in Figure 5.28), tracks only intersect on the second half turn and therefore at only one track angle. This is the least optimal case and we have minimal sensitivity to the radial alignment.

We also note that at large layer number (small radius, e.g. wire 1776) there is the largest separation between the track angles of the first and second half turn ($-130°$ and $130°$). As these are almost perpendicular to one another ($100°$ difference), these wires are easily aligned in all directions. This is more difficult in smaller layer numbers.

To summarize, our alignment is most limited in large sector numbers where tracks only intersect the cells on the second half turn. Further, in general it's easier to align wires in all directions at large wire number / small radius due to the larger difference in the track angle between the first and second half turns.

We note that this discussion is for the standard 1.5 turn tracks in the MEG trigger data. Using out-of-time half turn tracks would allow us to gain sensitivity in regions with limited track angles, but with a degraded resolution.

Further, we note that for the same reason we are not sensitive to these alignment errors, the alignment errors shouldn't actually affect the track quality. To give a simple example, in a simple drift chamber where all tracks through cell A are perfectly vertical, if the cell is misaligned vertically, it has zero impact on the track quality. However, it is possible that the combination of misalignment over several wires could potentially affect the alignment. This still needs to be investigated.

Additionally, we note that this alignment requires a mixing of tracks with varying intersected wire distributions in order to properly align the chamber. To give a simple example, in a hypothetical drift chamber intersected by three distinct categories of tracks, if all of the three categories share no common wires, the wire-by-wire alignment is not a global alignment, but

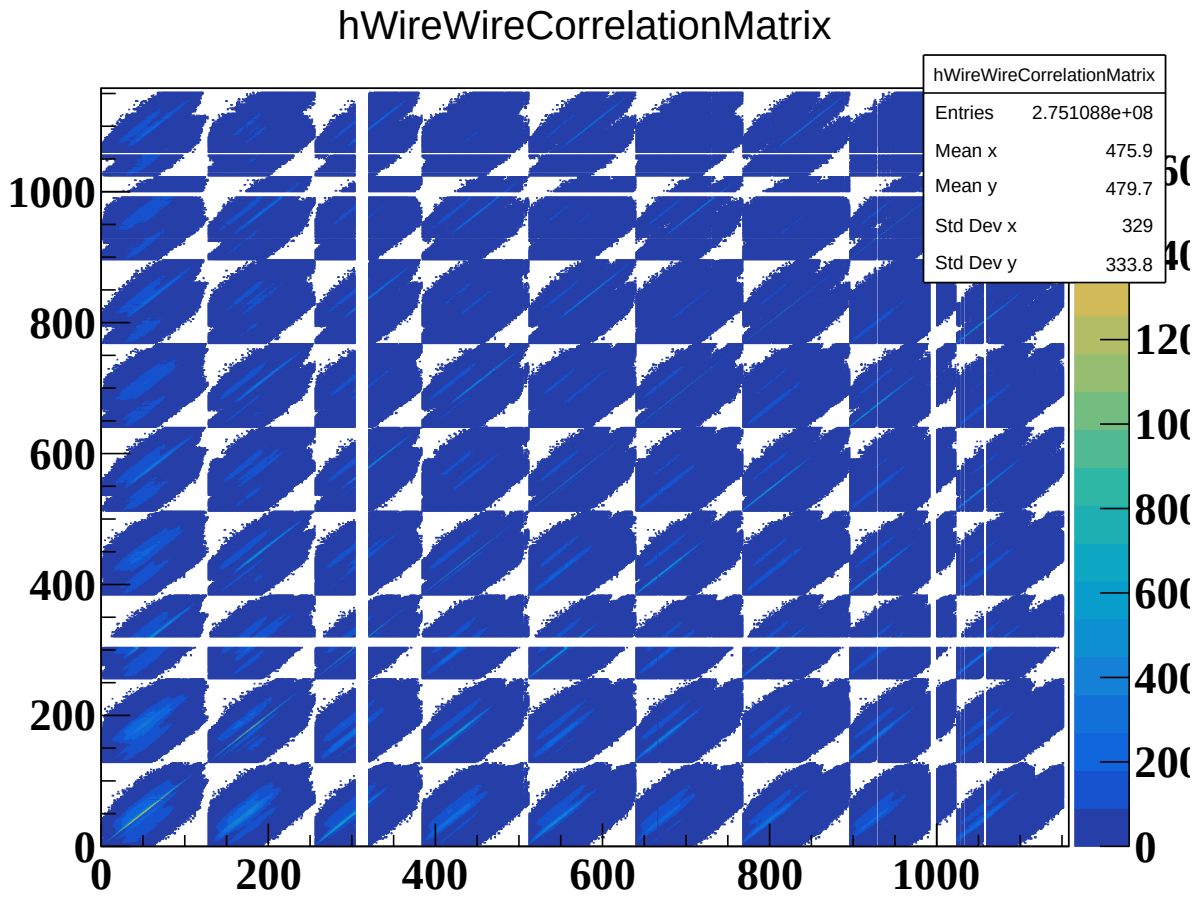


Figure 5.29: The two-dimensional histogram containing the number of tracks intersecting wire y given a track intersects wire x.

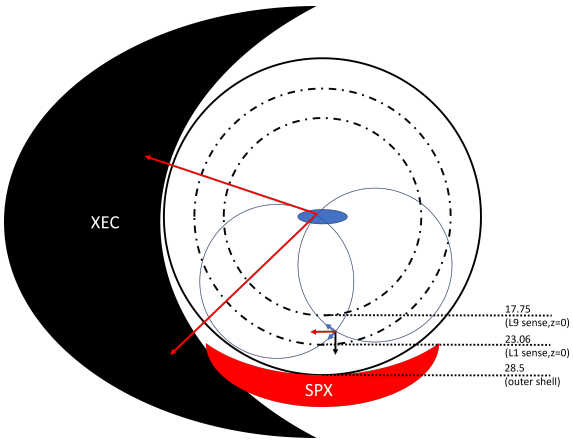


Figure 5.30: A slice of the MEG II experiment with two sample positron tracks. It's clear these are the possible angles to intersect this particular cell in the central region of the acceptance.

an alignment of three separate sections of the drift chamber. Thus we would still require the relative alignment of the three sections. Similarly, if all tracks intersect wires at $z < -30cm$ and $30cm < z$, we still require the information on how the two ends of the chamber are connected (upstream,downstream).

As for our chamber, we are lucky that the mode of the hit distribution is around z=0 and thus the z=0 region is connected to the upstream and downstream regions with upstream and downstream tracks. The upstream and downstream regions are thus connected via higher order mixing of upstream/downstream tracks with the common intersection point at z=0.

However, for the individual wires or sectors, the situation isn't as clear. As mentioned above, we benefit from the fact that the central sectors connect the first and final sectors via the tracks that intersect the central sectors on the first, second, and third half turns. Figure 5.29 shows the number of hits on wire A given the same track intersects wire B. For a more condensed plot we don't use the true wire number, but $i_{wire} - [192 * i_{layer} - 128 * (i_{layer} - 1)]$ in order to show exclusively readout wires. It's clear that the largest region with no mixing is the first and final sectors; this has some variation on the layer number, but isn't very significant. This is expected and implies that the alignment here requires the second order of mixing by tracks intersecting the first and central sectors with tracks intersecting the central and final sectors.

It's interesting to note that there are clear regions with high coupling. e.g. region (500, 700) has three high coupling sub-regions. These are likely the most common sector/layer intersections for a 1.5 turn track through that chamber region.

5.5.5 Results

We used approximately 500 runs each with 2000 events. With a trigger tracking efficiency (not the true efficiency) of $\sim 43\%$ and ~ 40 hits/tracks, the analysis contains ~ 17 M hits. The final mean residuals as a function of wire number in the X and Y global coordinate systems is shown in Figures 5.31 and 5.32 respectively. The residuals are highly suppressed with respect to the survey alignment. This contains 14 iterations of track-based alignment. To be clear, the edges of each layer have a very small amount of hits (< 2000) and thus have been excluded from the alignment procedure.

In addition, Figure 5.33 displays the same example wires shown above in 5.26. It's clear that when compared to the survey, these errors have been highly suppressed not only in the average error, but the average error along the wire.

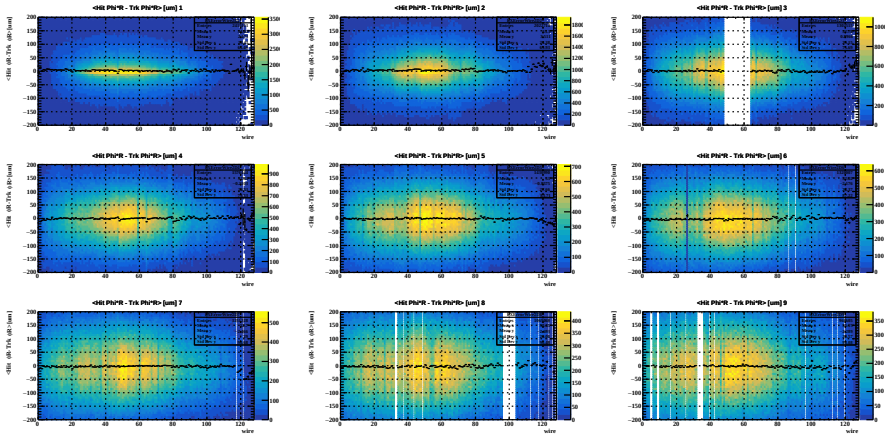


Figure 5.31: X global residuals as a function of wire number after the track-based iterative alignment.

Next, Figures 5.34 and 5.35 show the fitted intercept (translation of the wire or p_0) and the sagitta (p_2) wire-by-wire after all track-based alignment iterations for both the X and Y coordinates. For both X and Y, the two follow a linear correlation with a slope of 1. This is expected, as this implies that the center of the wire is well aligned and centered at zero error, but there are still some fluctuations in the sagitta term, which then forces a

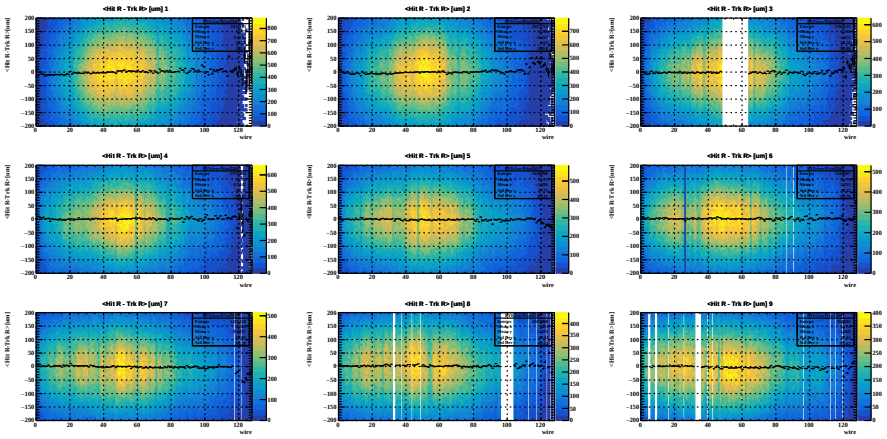


Figure 5.32: Y global residuals as a function of wire number after the track-based iterative alignment.

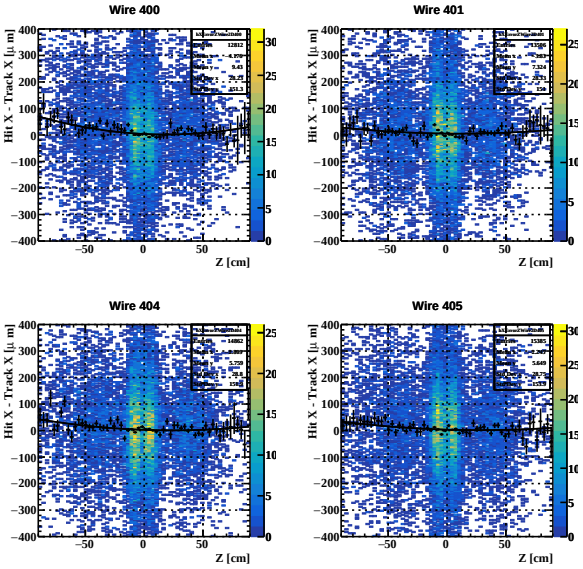


Figure 5.33: Sample residuals wire-by-wire after the track-based alignment.

fluctuation in the translation term. The *remaining* sagitta has a magnitude of $\sigma \sim 13\mu$m. Wires with large remaining sagitta are usually associated with wires with a small amount of hits; even with 17M hits in the sample, many wires have only a few thousand hits and thus have a large uncertainty in the sagitta Of course, this can be suppressed with higher statistics. Exactly 1000 wires pass our number of hits criteria (2000 hits/wire). Roughly 54 wires contain zero hits due to electronics issues, leaving $9 \cdot 128 - 50 = 1100$ wires, or roughly 100 readout wires that are excluded from the analysis. A majority of the excluded wires are

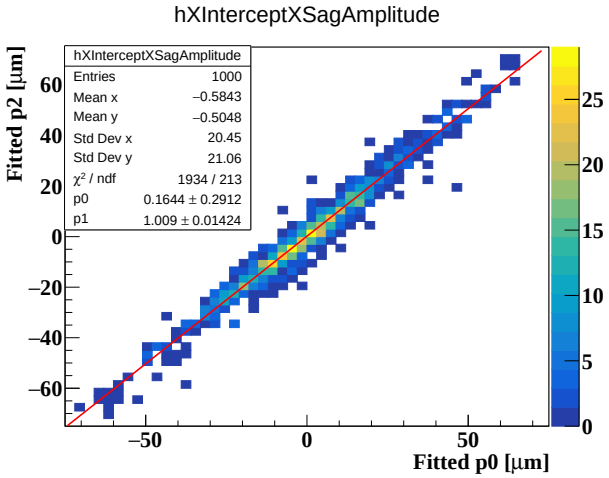


Figure 5.34: The fitted intercept (translation of the wire or p_0 and the sagitta (p_2) wire-by-wire after all track-based alignment iterations for the X global coordinate.

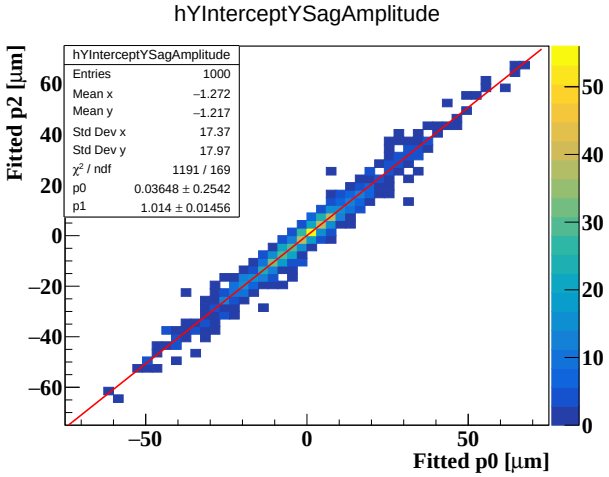


Figure 5.35: The fitted intercept (translation of the wire or p_0 and the sagitta (p_2) wire-by-wire after all track-based alignment iterations for the y global coordinate.

in the final half sector of each layer (9x8=72); these wires have an extremely low hit rate. The other excluded wires are in the first or final sector.

5.5.6 Discussion

Here we describe and quantify the changes in the wire position, rotation, and sagitta as well as the remaining alignment errors.

Figure 5.36 shows the final translations, rotations, and sagitta of all wires with respect to the survey. It's clear that the translations are centered at zero by coincidence. This is slightly expected as any time a wire is misaligned, this will create a large residual on that wire and then a smaller residual on nearby wires in the opposite direction (to account for the large residual in the fit). Whereas with the ϕ_3 angle, we observe a mean rotation of 1.1 mrad. It would be interesting to make an identical alignment with and without this mean rotation and see how it compares. Additionally, we note that the θ_3 requires a minimal rotation ($\sigma \sim$ 0.1 mrad).

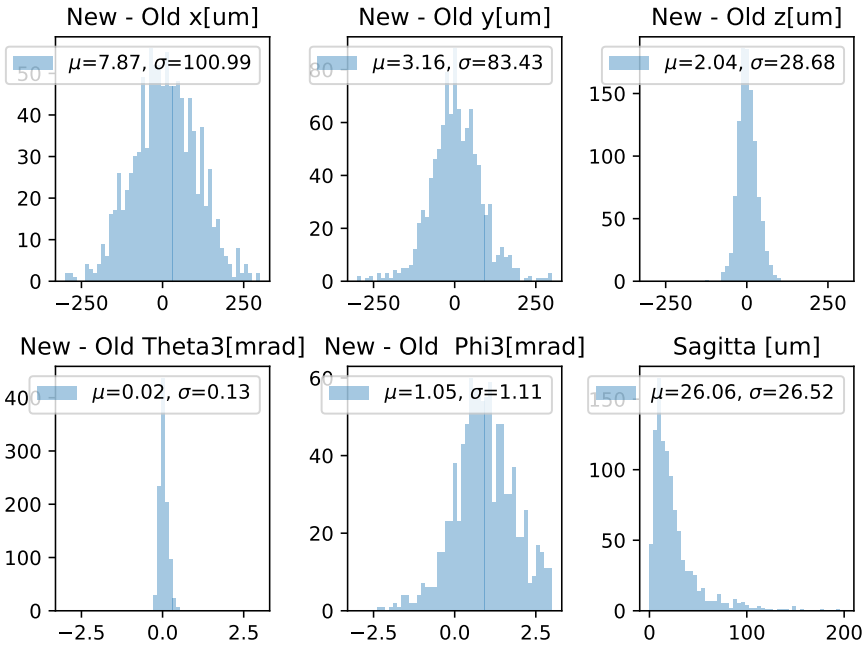


Figure 5.36: The final translations, rotations, and sagitta in a set of histograms.

Further, Figure 5.37 shows the change in position at z=0 cm, z=100 cm, and the difference between the change in position at z=100 cm and z=0 cm (this extracts the change in position due to the change in wire rotation). This confirms that this alignment doesn't shift the mean wire position in X or Y global or the mean radius of the wire at the extrema.

We also include the translations and rotations as a function of wire number in Figure 5.38. It's clear there is some structure layer-by-layer as well as sector-by-sector. The explanation

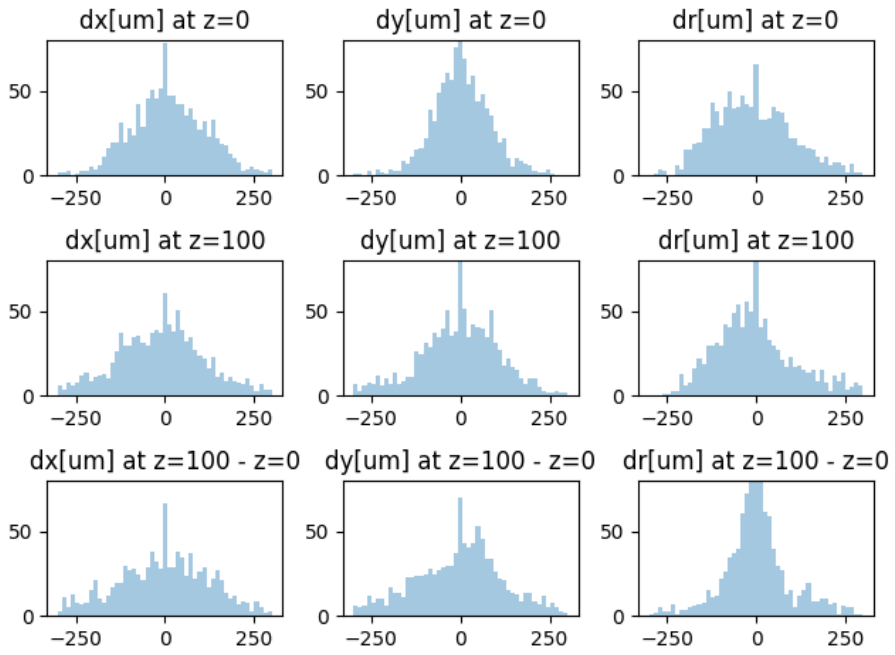


Figure 5.37: The final change in wire position in X, Y, and R. These are shown at z=0, z=100 cm, and the difference between z=0 and z=100 cm.

behind the discrepancy between this alignment and the original survey alignment is not well understood.

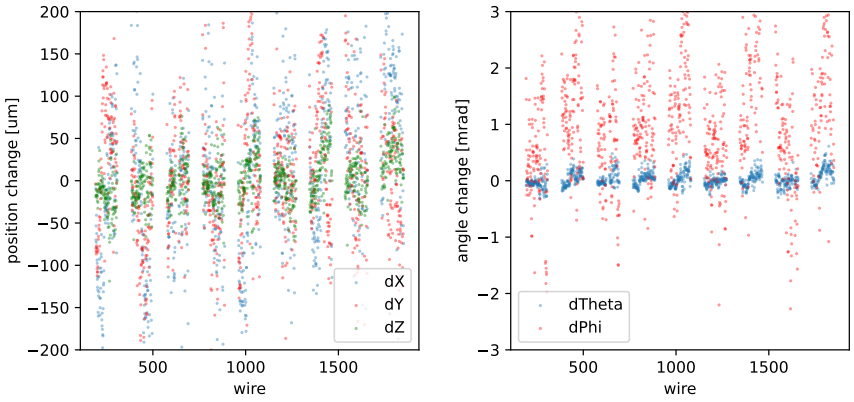


Figure 5.38: The final translations, rotations, as a function of wire number.

Next, the orientation of the sagitta and the sagitta amplitude as a function of wire number is shown in Figure 5.39 for all wires with a sagitta larger than 30 μm to avoid plotting the angle of a negligible sagitta. The most interesting feature is that the sagitta has a very polarizing angle depending on the layer number, this has not been explained.

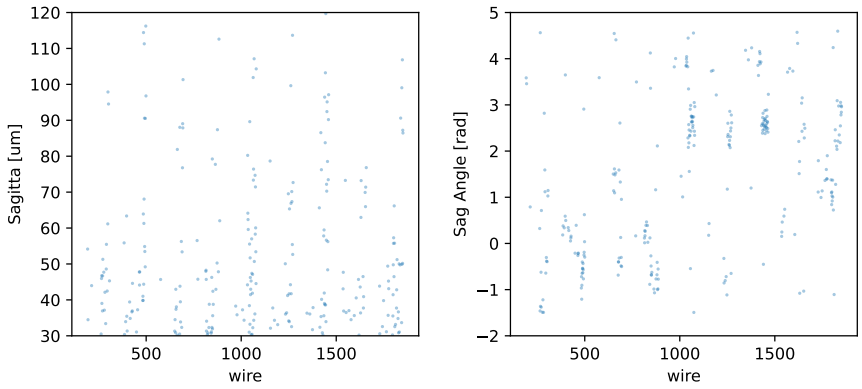


Figure 5.39: The final sagitta and the angle of the sagitta.

After the final alignment, we again fit for the required translation, rotation, and sagitta to quantify the remaining error in the wire position wire-by-wire. The fit parameters wire-by-wire are all shown in Figure 5.40. To be clear, to get the value of the error in the intercept, we take the value of $p_0 - p_2$ as this is the fitted value of the average error at z=0.

At z=0, the error in the intercept is 2 μm at the 1 σ level. At the 1 σ level, the remaining errors due to remaining sagitta are 0, 6, 12 μm at z=0, z=50, and z=100 cm respectively in both X and Y global. Similarly, at the 1 σ level wire-by-wire due to remaining angular misalignment results in errors of 0, 5, 10 μm at z=0, z=50, and z=100 cm respectively.

5.5.7 Double Turn Analysis

In order to evaluate how the change in the wire alignment affects the resolutions, we use the double turn analysis. This is a data-driven approach to estimate the positron kinematic resolution $(p_e, \theta_e, \phi_e, y_e, z_e)$. The technique uses positron tracks that intersect the chamber on two "turns". An example of a two turn track is shown in Figure 5.57. The first turn (2 chamber intersections) and the final turn (3 chamber intersections) are independently measured and fit. The two positron turns are then propagated to a common layer between the turns that is parallel with the target surface. Comparing the kinematics at this common

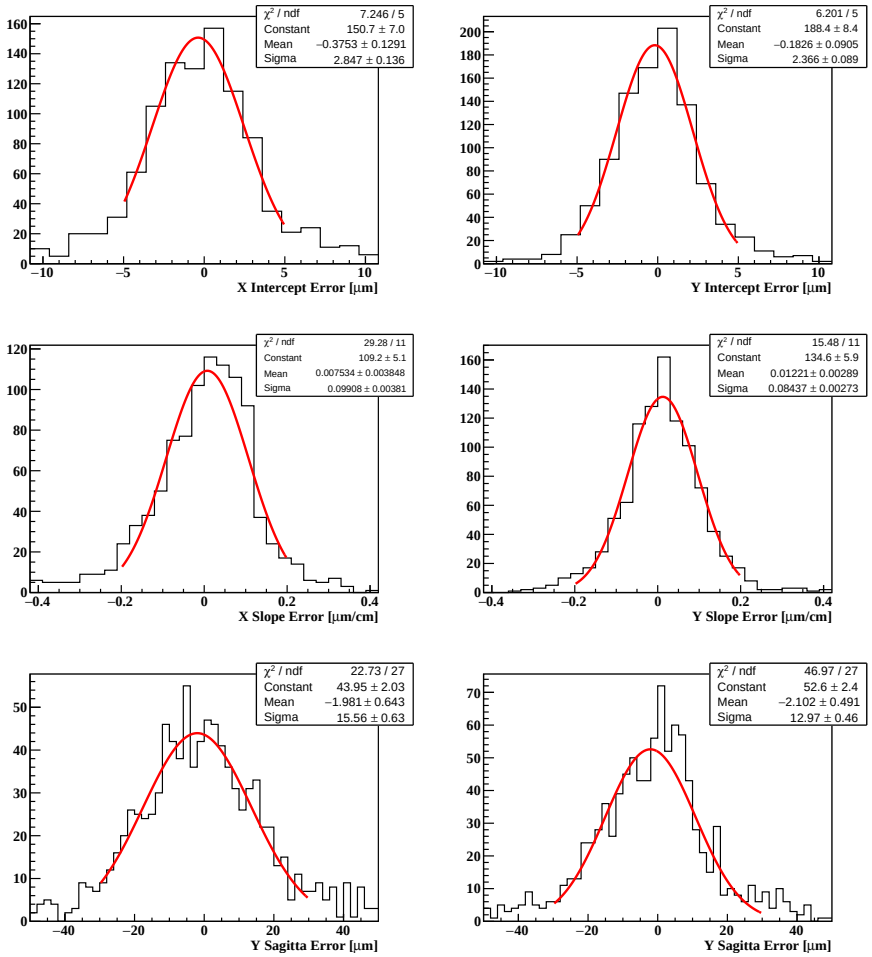


Figure 5.40: The resulting fit parameters wire-by-wire after all track-based alignment iterations.

layer yields a resolution estimate of the positron tracks: the better the comparison, the better the resolution.

Figure 5.41 shows the resulting double turn analysis for the survey alignment, the fifth iteration of the track-based iterative alignment, and the 12th iteration of track-based iterative alignment (CYLDCHConf.id=40).

It's clear that the ϕ_e, y_e, z_e mean errors are higher suppressed, whereas the θ_e comparison gains a 0.5 mrad offset. The offset in the momentum is suppressed by 10 keV. We also observe a dramatic improvement in the overall resolutions.

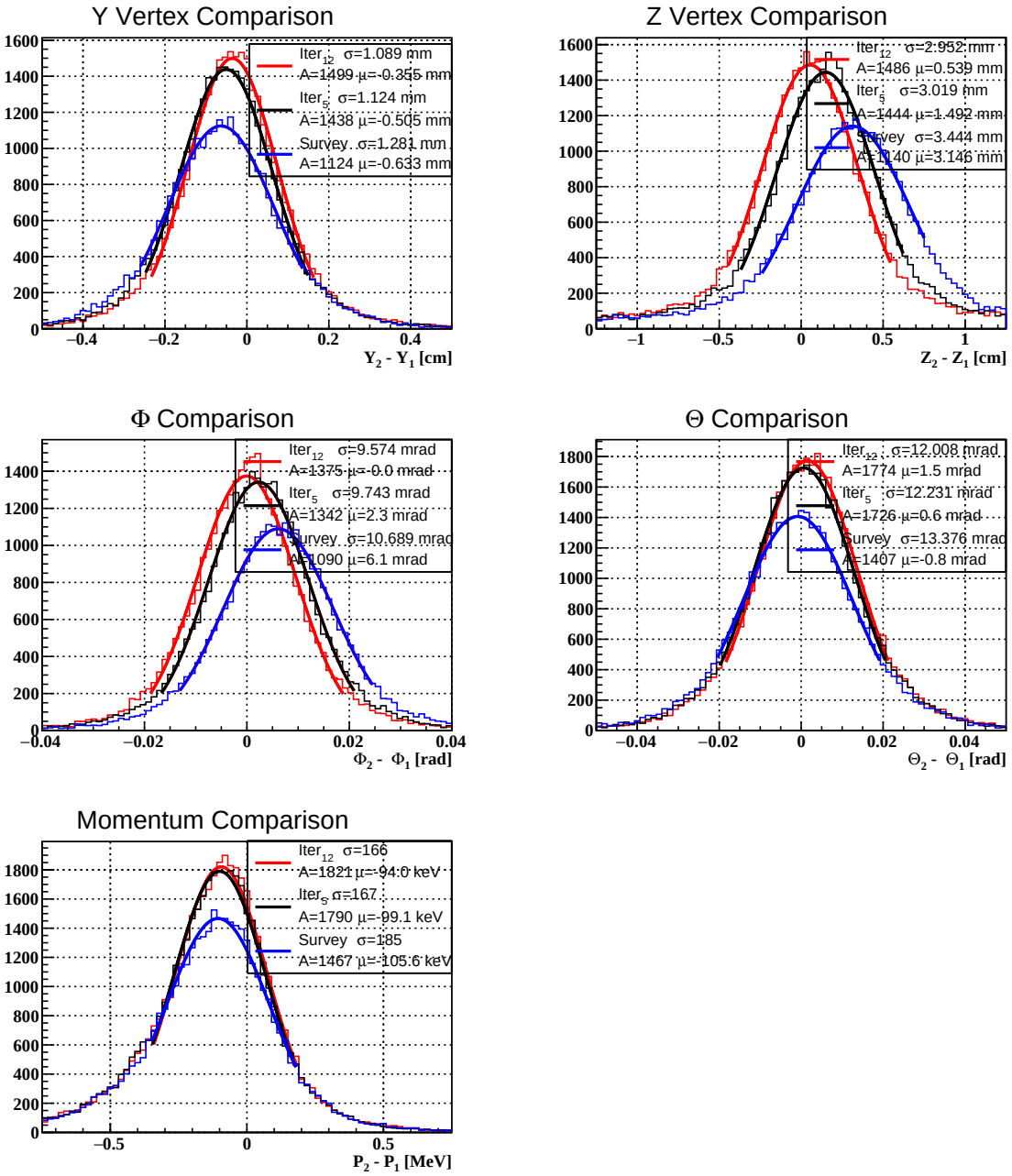


Figure 5.41: The resulting double turn analysis for the survey alignment, the fifth iteration of the track-based iterative alignment, and the 12th iteration of track-based iterative alignment (CYLDCHConf.id=40).

These resolutions and biases can be further improved by improving other alignments and calibrations. For example, Figure 5.42 contains the double turn analysis using CYLDCH-Conf.id=41 (final alignment) with the default magnetic field and an alternative magnetic

field optimized by suppressing the biases observed in the Michel edge fitting and the double turn analysis (kCalculated, scale=0.999, Y Shift = 1 mm). All kinematic comparisons have a suppressed bias when using the optimized magnetic field. Note that the momentum double turn bias still persists ($\sim$ 50 keV).

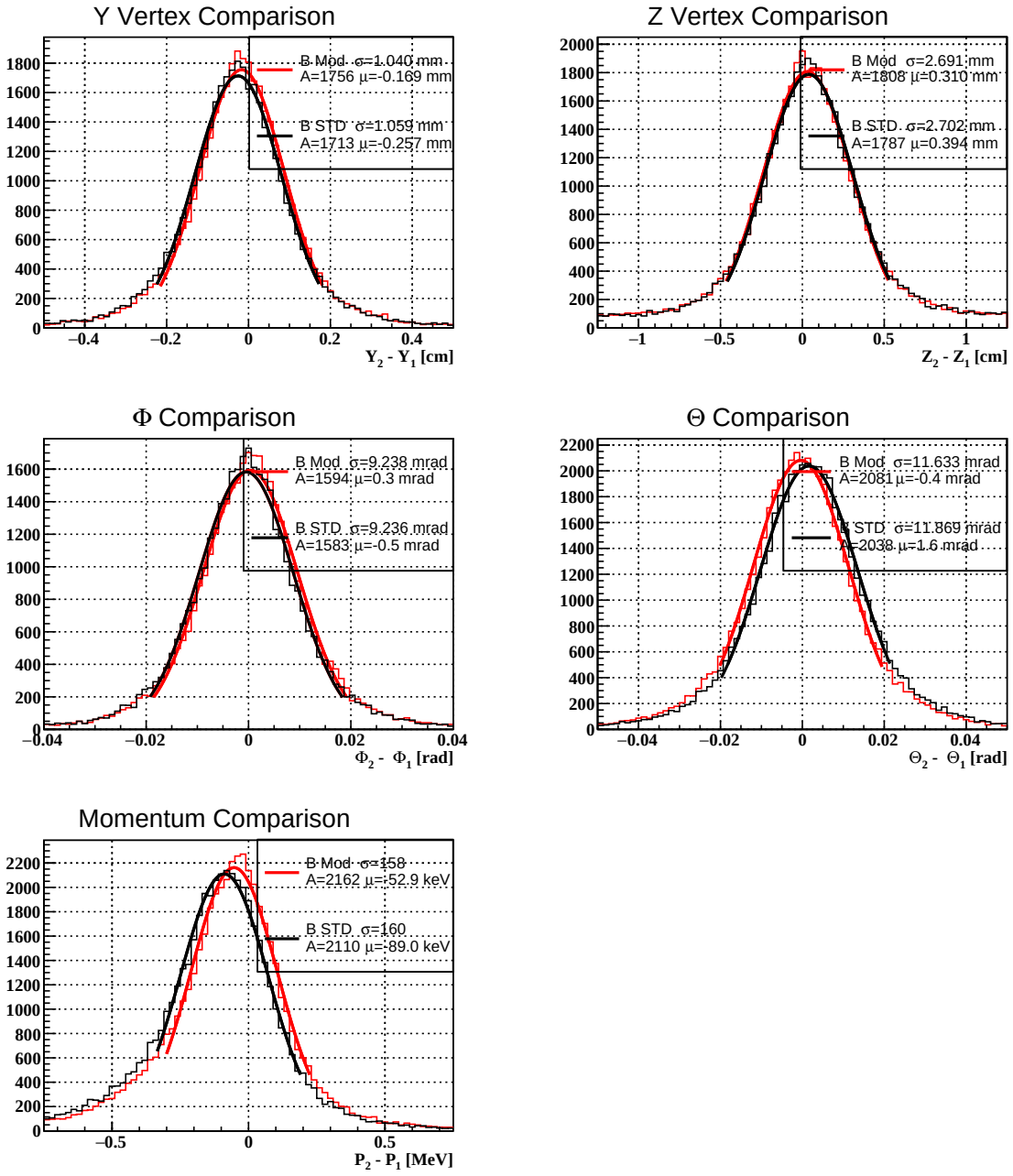


Figure 5.42: The resulting double turn analysis for the final wire alignment (CYLDCH-Conf.id=41) with the default magnetic field and a modified magnetic field to suppress the biases.

5.5.8 Possible Improvements

This analysis could be improved at the edges of the acceptance by including more hits into the analysis. Further, to improve on the alignment in regions with a limited range of track angles, we can try adding out-of-time half turn tracks into the alignment procedure. This allows us to have a larger range of track angles and thus align the wires in all directions. Finally, we mention that the wire alignment does not attempt to align the wires globally, but only the relative wire-by-wire alignment of the chamber.

5.5.9 Conclusion

This note presents a track-based iterative approach using MEG trigger data. The alignment includes wire translations, rotations, and a wire sagitta. We estimate that the remaining wire-alignment errors are 2, 5, 15 μm at z=0, z=50, and z=100 cm respectively. Results and improvements are shown at the double turn analysis and at the wire-by-wire level. We conclude by presenting some methods that could be used to improve the alignment further.

5.6 Track Selection

5.6.1 Introduction

A subset of positron tracks contain low quality reconstruction and therefore have wide kinematic resolutions (e.g. $\sim$ 1 MeV tails where the typical core resolution is $\sim$ 0.09 MeV). For signal, the tracks are often far from the signal energy and therefore do not significantly contribute to the signal efficiency. Instead, lower momentum accidental positrons can be mismeasured to high momentum resulting in increased background and thus a degradation of

the signal/background. This chapter discusses the usage of a neural network trained directly on data to remove these low quality tracks while still maintaining high signal efficiency.

In theory, the low quality tracks could be correctly handled by the physics analysis (maximum likelihood analysis) and thus should be maintained. However, for these tracks it is difficult to have a quality kinematic resolution estimate event-by-event as the wide resolutions are not observed in the covariance matrix. As an example, we show the mean MC pull (σ of the $(E_{Rec} - E_{MC})/\sigma_{KalmanE}$) as a function of different measurables (Atsushi Oya 10/6/22). In all cases when the track is low quality (low hits, high χ^2/DOF, or high diagonal covariance matrix elements, etc.), the mean pull increases. Therefore, to optimally define the resolution event-by-event, additional corrections as a function of a multi-dimensional measurable space would need to be taken into account. This comes with uncertainty and difficulty. If instead, we simply define the pull as the pull integrated over all measurables, we are underestimating the resolution for low quality events and overestimating the resolution for high quality events. This in turn degrades our likelihood analysis.

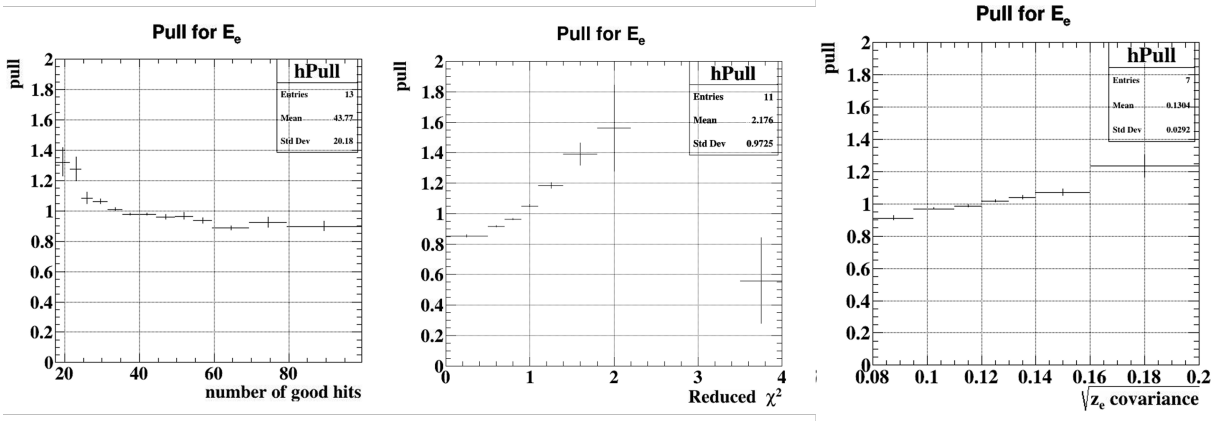


Figure 5.43: The MC momentum pull $[(E_{Rec} - E_{MC})/\sigma_{KalmanE}]$, as a function of different kinematic variables (Figure made by Atsushi Oya).

To summarize, the track selection is motivated as the events do not significantly contribute to the signal efficiency, can result in additional mismeasured accidental background events in the signal region, and it is difficult to estimate their true resolutions. Further, optimized track selection significantly simplifies the analysis and there is significantly less reliance on the

handling of tails in the final physics analysis. However, we are also attempting to maximize the number of tracks in the data sample to maximize the overall sensitivity. Therefore, our summed objective is to remove tracks with wide tails where the true resolution does not agree well with the pull, while still maintaining a high efficiency.

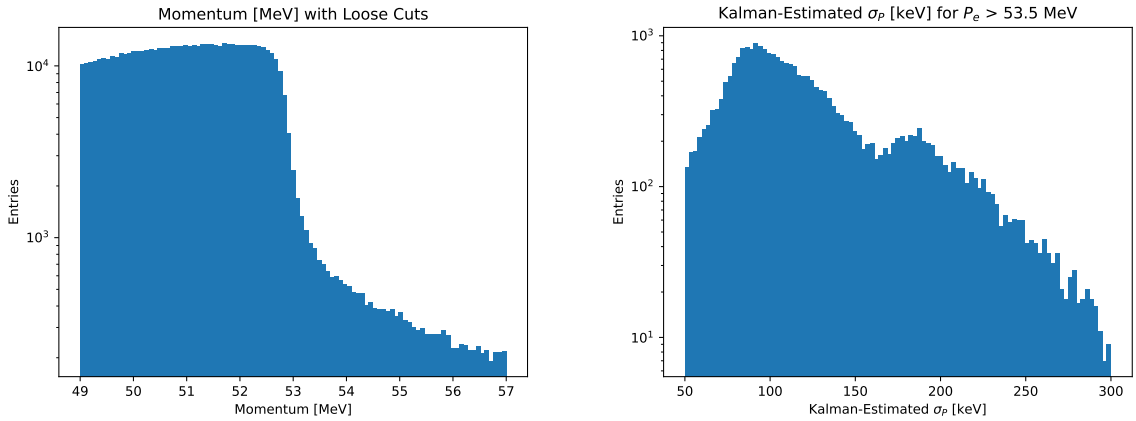
Note, the tracking procedure results in the possibility of two reconstructed tracks corresponding to the same physical positron track i.e. ghost tracks. The final physics analysis will only use one reconstructed positron track per event; the neural network track selection is additionally used to rank these ghost tracks.

5.6.2 Conventional Approach

Here we discuss a conventional approach to track selection. Selection criteria can be made by looking at measurable distributions in the Monte Carlo (MC) and removing events accordingly. For example, we can remove tracks with less than 30 hits, $\chi^2 > 1.7$, etc, such that the pull distributions aligns have a width of ~ 1. We could additionally search for measurable values that typically have very wide kinematic tails and cut accordingly.

However, this doesn't take into account correlations between the measurables. We could instead make selection based on the multi-dimensional histogram between relevant measurables (e.g. a linear selection criteria), but it is difficult to do this in a multi-dimensional space. This technique has another shortcoming in that it relies on the MC; we've found discrepancies between the data and the MC in the waveform noise level, hit resolution, kinematic resolution, etc.

Alternatively, similar selection could be based on the double turn pull distributions in data, e.g. the pull of the energy distribution ($\frac{P_2 - P_1}{\sqrt{sigma_{p,1}^2 + sigma_{p,2}^2}}$) as a function of the number of hits on the 1st turn, 2nd turn, average χ^2/DOF, etc. However, this would identify selection



(a) (b)

Figure 5.44: (a) The momentum distribution at [MeV] integrated over all beam rates. (b) The square root of the Kalman covariance diagonal element [keV] event-by-event integrated over all beam rates.

criteria for a quality double turn comparison, not selection for a standard 1.5 turn tracks so it's unclear what systematics to expect.

5.6.3 Machine Learning Data-driven Approach

5.6.3.1 Description

Here we describe a data driven machine learning approach to track selection. Positrons from muon decay can not have true energy larger than 52.83 MeV. Positron momentum larger than 53.5 MeV (for example), have been mismeasured by at least 700 keV. In Figure 5.44a, the positron momentum distribution of events that satisfy relatively loose track quality selection criteria is shown on a log scale; many events with momentum errors greater than 700 keV remain.

The data-driven momentum resolution estimates from the Michel edge fit and the double turn distributions have a core resolution of $\sigma_{p_e} \sim 90$ keV with a significant tail component extending to > 1 MeV. Events with momentum larger than 53.5 MeV are mismeasured by

at least 7 σ_{core}. In Figure 5.44b, we show the square root of the Kalman diagonal element representing the momentum resolution for tracks with momentum greater than at least 53.5 MeV. The mode of $\sigma_{KalmanE}$ for these events is $\sim$ 90 keV. Such badly measured events also exist in the acceptable momentum region. The goal of the machine learning selection procedure is to devise a method to identify these events with large momentum errors. Since the Kalman covariance diagonal does not identify these events, an alternative method is needed.

The dataset is split into two categories, "good"/"bad" with momentum <53.5 and >53.5 MeV, respectively, to measure track characteristics for mostly well measured tracks ($<$ 53.5 MeV) and those known to have large errors ($>$ 53.5 MeV). The "bad" sample all have at least a 700 keV error. On the other hand, a "good" track includes tracks with the full range of quality, that is quality tracks and a small fraction of the "bad" quality tracks.

The objective is to use the measurables to learn how to distinguish between these low quality tracks and the standard quality tracks. The machine learning approach trains directly on data to learn the correlations between the measurables and the probability of being a "bad" track. The end result is a network that inputs all measurables and returns a 0-1 float defined as the probability of being a "bad" track (P_{bad}). We then remove tracks with P_{bad} greater than some threshold, chosen to remove bad events at high probability with small probability of removing well measured events.

The current training set uses a combination of tracks found both with and without the ML hit finding algorithm. This set contains about 1M positrons that pass the standard loose selection criteria to select events to be used in the final analysis. We train using all input variables mentioned in Subsection 5.6.3.2 except those that are known to or could bias the momentum distribution. The obvious example of bias is that the ML algorithm could not use the momentum itself, as the network would not be learning to distinguish between high and low quality, but instead high and low momentum. This is discussed later in more detail.

5.6.3.2 Measurables

Here we describe the track observables that could be used in any track selection approach and describe whether or not each potentially biases the momentum distribution. We explicitly state if a measurable is not used.

- **Number of hits on the fitted track** - More measurements indicates an improved kinematic resolution given the same track path length from the target-SPX. This variable may create a small bias in the momentum distribution due to the fact that low momentum positron tracks do not reach the outer layers and therefore will have less hits.

- **Hit efficiency and plane efficiency** - As above, given the same χ^2 and path length, we expect high hit efficiency to have high quality. We describe how these efficiencies are determined in the refit task (described in a previous chapter provide the section).

- **Number of hits on the first and last half turn** - A low number of hits on either half turn indicates a degraded kinematic resolution and hence the quality of projection to the target or SPX. We found that including these variables in the neural network biases the momentum distribution and hence is not used.

- χ^2/DOF - A large χ^2/DOF will have worse resolution.

- **Covariance matrix elements** - These are explicit indications of the Kalman fitter kinematic resolution.

- **Propagation length from the target to the CDCH** - A longer path length to the target results in worse vertex position resolution.

- **Propagation length from the CDCH to the SPX** - A large propagation length to the SPX indicates a degraded extrapolation and thus a degraded timing resolution.

This shouldn't affect the other kinematic variables significantly (negligible effect on the CDCH T_0 and thus the DOCA resolution).

- **Propagation length from the target to the SPX** - Combining this with the number of hits is comparable to a rough hit efficiency. This value biases the momentum distribution due to the large correlation between energy and overall path length. This should not be used, this is described in later sections.

- **Kalman-estimated energy loss** - In general, the more energy the track loses, the more uncertainty there is in the track kinematics due both to scattering and to dispersion in the energy loss since the fitter assumes median energy loss as the particle traverses matter. The energy loss is due to the positron intersecting the target between the first and second turn, intersecting the inner mylar shell and the outer carbon-fiber shell, and to loss in the CDCH gas. This summed energy loss is split in the next three measurables.

- **Kalman-estimated energy loss from the target to the drift chamber** - Losing energy in this region would significantly degrade the track resolution at the target.

- **Kalman-estimated energy loss from the end of the first turn in the DCH to the start of the second turn** - Large energy loss between the turns could result in kinematic differences between the first and second turns. This could indicate hitting the target between turns.

- **Kalman-estimated energy loss from the start of the DCH to the end of the DCH** - An overall higher energy loss would indicate a lower resolution.

- **Positron target position, angle, energy** - Selecting positron tracks based on the kinematic variables is more likely to introduce bias than other variables as it may bias the final positron track kinematic distribution. Therefore, the technique described in this note does not use the position angle, vertex, or energy.

- **Match distance and χ^2 between the DCH track and the SPX hit measurement** - If a positron track completely misses a physical SPX tile it is suspected that the track is low quality. Similarly, if the distance between the track projection at the SPX and the SPX hit position itself is large, one of the measurements has a large error.

- **Boolean ML Hit Finding** - This is described in section 5.6.5. The 2021 dataset uses a combination of tracks from two reprocessings (with and without ML hit finding). The machine learning algorithm may find different optimization for these two classes of events.

- **Number of Physical Turns** - The positron tracks typically have 1.5 turns in the CDCH; $\sim 15\%$ of positron tracks emitted at small θ_e make an additional turn and an even smaller fraction make 3.5 turns. These positrons will obviously have more hits, different covariance matrix elements, etc. We've included this as a hot-encoded array e.g. (0, 1, 0) for a double turn track.

Figure 5.45 contains histograms of the measurables for "good" and "bad" tracks. This demonstrates that some variables have significantly better discrimination power than others. Figure 5.46a shows a 2D histogram of the hit efficiency vs. the χ^2/DOF, as expected, "bad" tracks are more often associated with a high χ^2 and a smaller hit efficiency. In Figure 5.46b, we show the 2D histogram of the χ^2/DOF and the Kalman σ_p for "good and "bad" tracks; on average "bad" tracks have a larger covariance diagonal. However, as shown above, there are also many instances when the covariance diagonal is ~ 100 keV.

5.6.3.3 Machine Learning Optimization

Here we list several optimizations that were performed.

- **Dataset weighting towards "good" tracks** - The training sample is heavily weighted

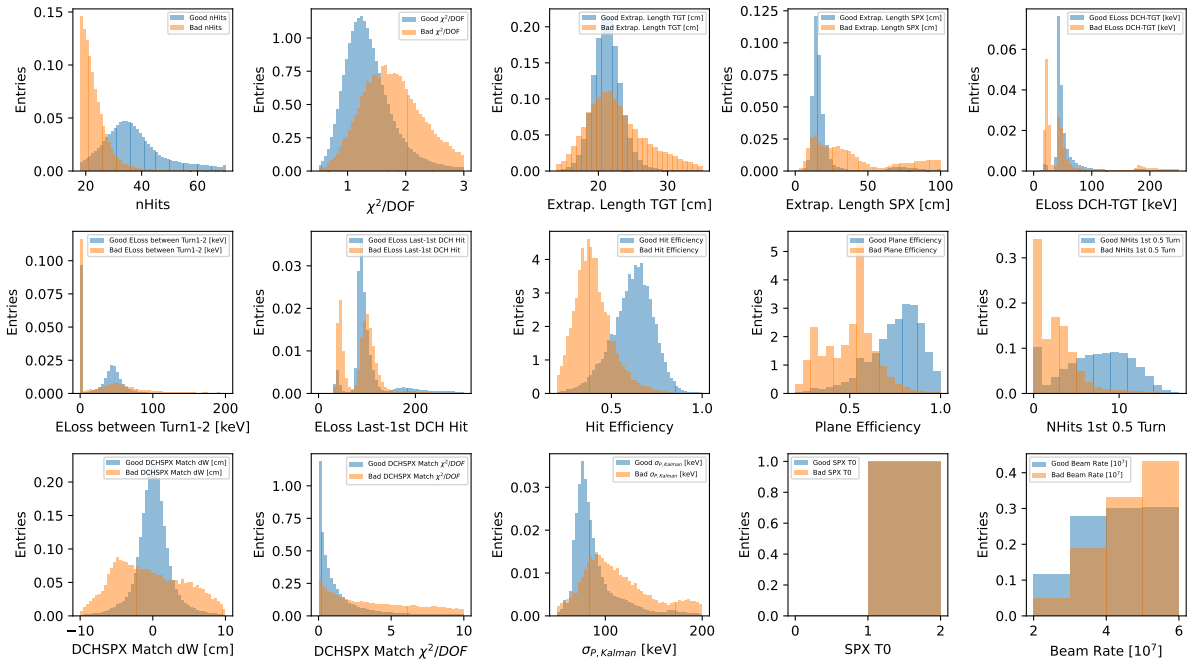
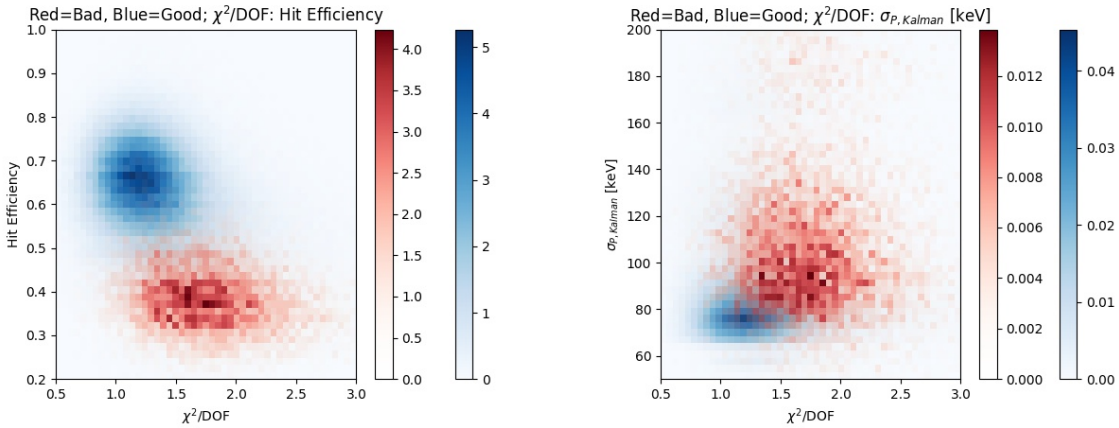


Figure 5.45: Histograms of all variables used to classify the "good" and "bad" tracks.



(a) (b)

Figure 5.46: (a) The pair scatter plot matrix of the number of hits and the χ^2/DOF."good" and "bad" tracks. (b) The 2D histogram of the Kalman σ_P and the χ^2/DOF for "good and "bad" tracks.

with "good" tracks. This results in a machine-learning algorithm that avoids labeling a track as "bad" as it is too unlikely. To remove this feature, we eliminate a large fraction of "good" tracks (75%). These are later used for validation studies. One could also adjust the loss function, but we are only limited in the statistics of the "bad"

tracks, not the "good" tracks, so this data removal shouldn't degrade our overall track selection power.

- **Optimizing the neural network structure** - The default network is shown in Figure 5.47. It is a simple series of dense layers. We found that increasing the number of nodes per layer beyond the number of input variables improved the prediction accuracy. Currently, the number of nodes per layer is set to six times the number of inputs.
- **Optimizing the neural network loss function** - We found the best loss function to be the binary crossentropy loss function (compared with a mean absolute error or a categorical crossentropy).
- **Comparison with other ML algorithms** - We also attempted the following classification algorithms in sklearn: GradientBoostingClassifier, RandomForestClassifier, SupportVectorMachine (SVM), but the dense neural network out-performed them in the validation classification. Some exploration of model parameters was done, but it was not exhaustive.
- **Avoiding Bias** - We found that an increased number of parameters (number of nodes) or increased number of training epochs resulted in a biased momentum distribution. We avoided biases by checking the momentum distribution of the prediction. We also found that momentum distribution can be biased by adding specific input variables, this is described in Section 5.6.4.1.

5.6.4 Results

5.6.4.1 Momentum Bias Check

Before discussing the quality of the selection, we discuss a verification that we are not biasing the momentum distribution with this track selection approach. That is, we want to avoid

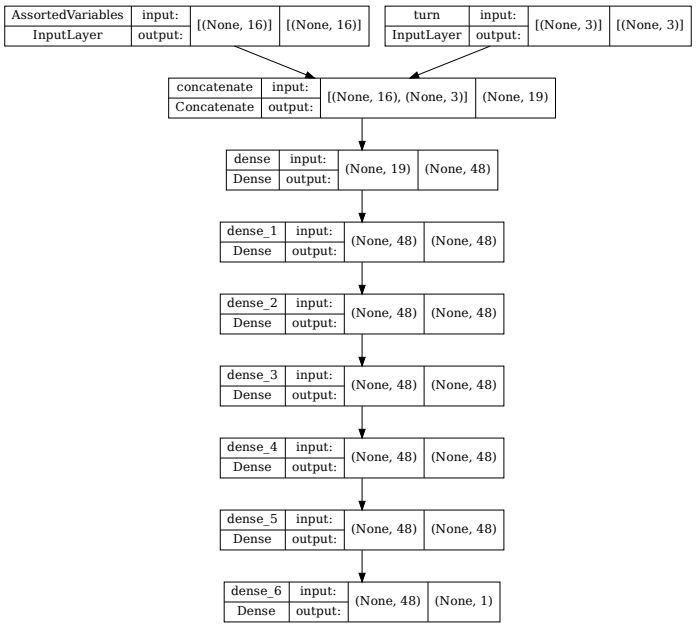
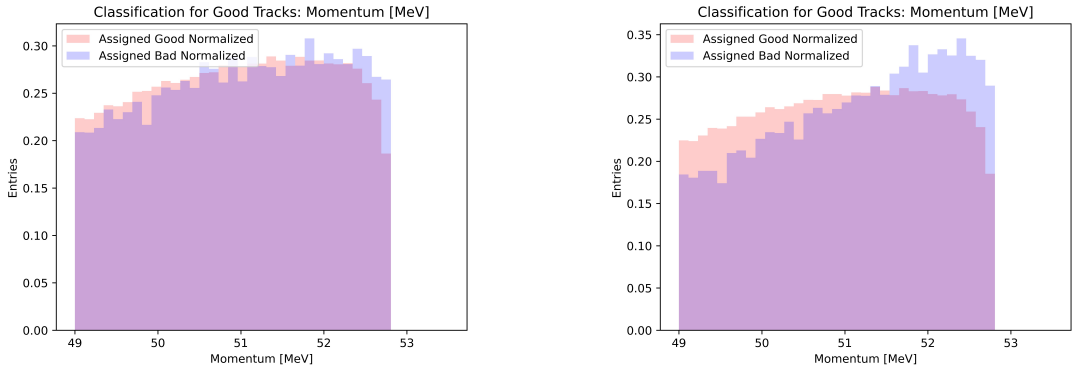


Figure 5.47: The neural network used for training/prediction.

removing a disproportionate amount of tracks at high momentum. This would be the result of a correlation between a measurable and momentum. This is a *potential* negative side effect of using the Michel edge to discriminate between high and low quality tracks. This bias would indicate that we are not selecting tracks based on quality alone, but a combination of quality and probability of high momentum.

In Figure 5.48a, we plot the momentum distribution for "good" and "bad" tracks with the standard set of input variables in test data. We select a threshold such that we remove 6% of the "good" tracks. Both histograms are normalized to the number of entries. The two histograms follow the same shape as a function of momentum, indicating that the momentum distribution is not biased.

In Figure 5.48b, we add the following variables: maximum radius on the first turn, the number of hits on the last half turn, and the total propagation length from the target to the SPX. It's clear that these variables bias the momentum. We added these variables one at a



(a) (b)

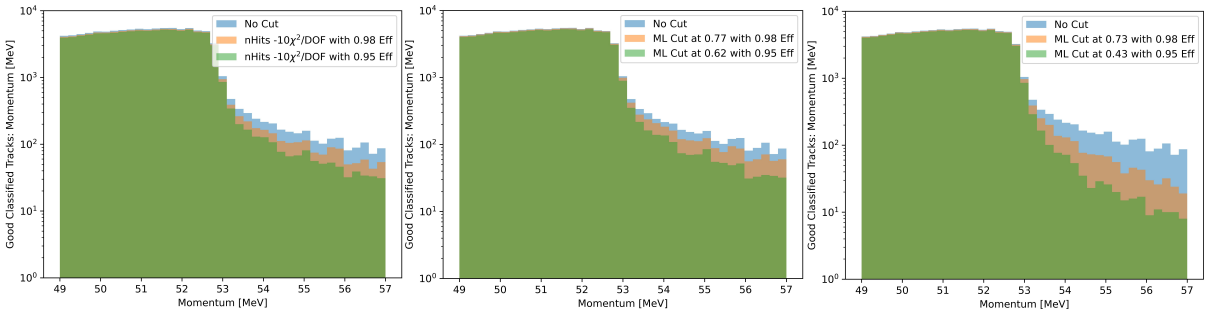
Figure 5.48: The momentum distribution for "good" and "bad" tracks. Both histograms are normalized to the number of entries. (a) The histograms follow the same shape as a function of momentum indicating that the momentum distribution is not biased. (b) In this subplot we add the maximum radius on the first turn, the number of hits on the last half turn, and the total propagation length from the target to the SPX into the training. Clearly the distribution gets biased.

time to verify that they each bias the momentum distribution. These inputs are not used in this technique. As noted above, we also observed this bias when the data was over-trained (too many epochs or parameters).

5.6.4.2 Momentum Spectrum With Varying Selection

Here we show a few example momentum spectra with different selection. First, in Figure 5.49a, we show the momentum distribution using a conventional cut: ngoodhits - $10\chi^2/DOF$. This was previously optimized using the Monte Carlo; of course, a more sophisticated conventional cut could be made, but was not investigated. In Figure 5.49b, a neural network was trained using only the number of hits on the fitted track and the χ^2/DOF. In Figure 5.49c, the neural network uses all variables (that don't bias the momentum). In each figure, we plot the momentum spectra with no cut, removing the "worst" 3% and the "worst" 6% of the data. That is, we select a threshold X that Y% of the data is removed. As expected, using a neural network that only uses ngoodhits and χ^2/DOF has similar selection power

to a conventional cut on the combination of the two input variables. The improvement from the neural network procedure comes from using a combination of all input variables.



(a) (b) (c)

Figure 5.49: The histograms show the momentum spectra for three different track selections for the two neural networks. The momentum spectra is shown with no cut, removing the "worst" 2% and the "worst" 5% of the "good" sample. That is, we select a threshold X that removes only Y% of the data. (a) A conventional cut (nHits - $10 \cdot \chi^2/DOF$). (b) A neural network only using the number of hits on the fitted track and the χ^2/DOF. (c) A neural network using all variables that do not bias the momentum.

5.6.4.3 Kinematic Resolution Comparison

Here we discuss the relative kinematic resolutions of positron tracks with and without the ML track selection. The current default threshold eliminates tracks with $P_{bad} > 0.1$, this results in an efficiency loss of 7%.

First, in Figure 5.50, we show the results of fitting the Michel edge distribution with and without the ML track selection.

As a reminder, the main objective is to not improve the core resolution, but eliminate the long tails. The Michel edge is fit to the known Michel momentum distribution, an acceptance function, and a resolution function (sum of three gaussians). The core gaussian width is effectively the same in the two cases, but σ_2 narrows from 443 $\rightarrow$ 336 MeV, the fraction in the widest gaussian, frac3, drops from 0.084 $\rightarrow$ 0.033 and the width of the widest Gaussian drops from 2.509 $\rightarrow$ 1.723 MeV. All values indicate a high suppression of the tails.

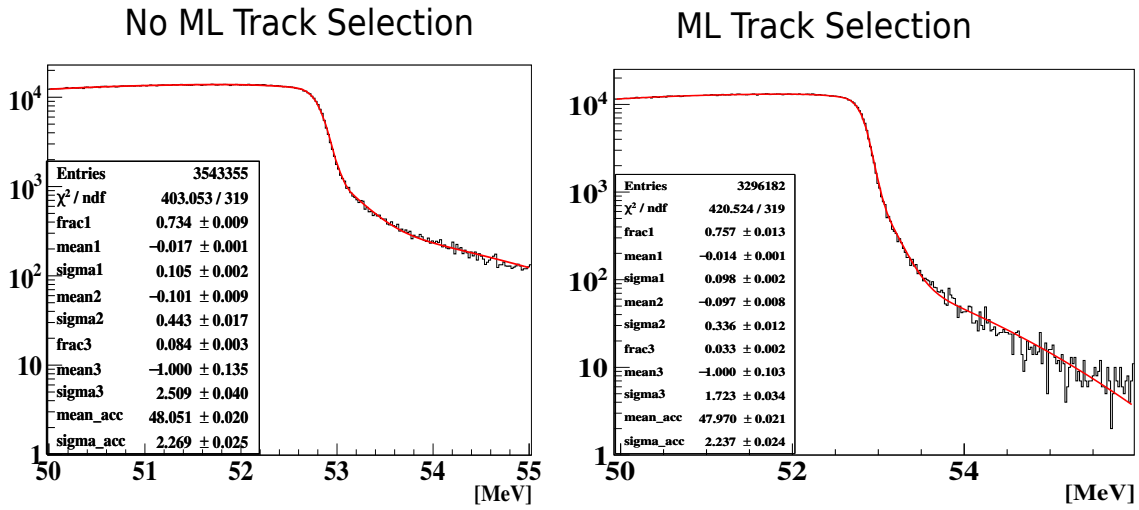


Figure 5.50: Fit of the Michel edge in the momentum distribution used to estimate the momentum resolution with and without the ML track selection applied. The distribution is shown on a log scale.

We can also use the double turn analysis to determine if the ML selection removes the wide tails in the kinematic resolutions. However we found that the ML track selection did not have a significant impact on the double turn resolutions. This is due to the fact that there are a very small number of double turn tracks with momentum greater than 53.5 MeV even without the ML track selection. Only $\sim$ 0.4% of the full dataset is a double turn track that had a momentum outside the allowed region (after already removing 75% of the good tracks). There are several steps that are required to have a successful double turn track that are not required for a single turn: successfully fit both turns, successfully propagate them to a common plane, pass merging criteria, and successfully fit the two turns together. None of these steps are required for the single turn tracks and therefore we do not observe the extremely long tails in the double turn analysis that are seen in the true MC resolutions or the Michel edge (dominated by single turn tracks). For this reason, the double turn analysis underestimates the wide resolution tails of single turn tracks. However, after getting rid of these tails using the ML track selection, the tail (and the core) of the kinematic resolutions are well represented by the double turn analysis. Note, this was one of the objectives of the ML track selection technique.

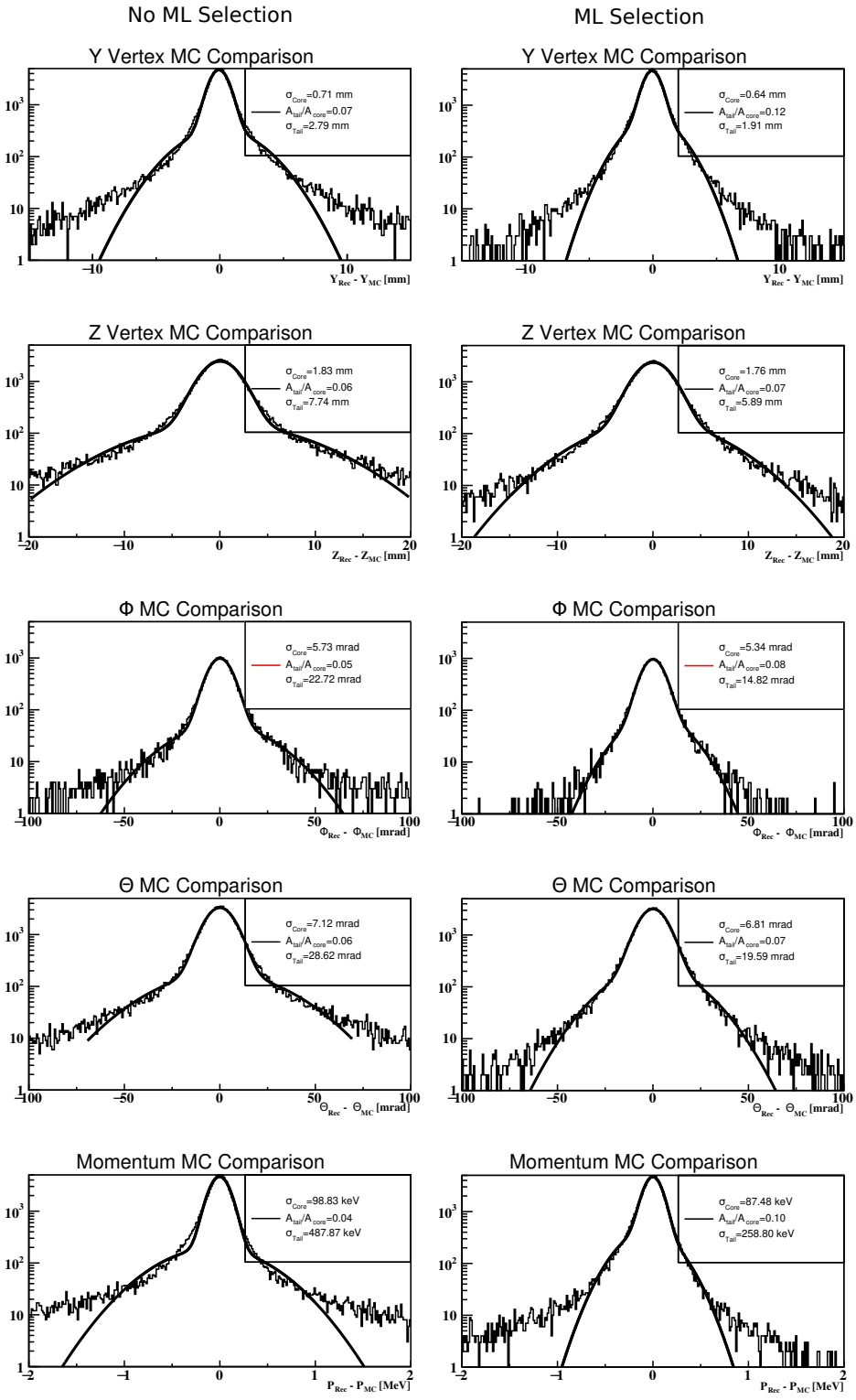


Figure 5.51: The Monte Carlo resolutions with and without the ML track selection. The distribution is shown on a log scale.

Below in Figure 5.51, we show the true MC resolutions for all kinematics with and without the ML track selection applied for signal positrons at $5 \cdot 10^7$ Hz. The identical neural network (threshold, weights, etc.) was applied to the Monte Carlo simulation. There is a large improvement in the tail resolutions. In the MC, the standard threshold results in a 7.5% reduction in events. This reduction results in a 2.4% reduction in the number of events that pass all kinematic cuts at the 3σ level (e.g. $|p_{Rec} - p_{MC}| < 3 \cdot 90keV$) and a 19.1% reduction in the number of events that fail at least one kinematic cut at the 3σ level.

To summarize, even though the neural network was trained and therefore optimized for data, in the Monte Carlo, we see a reduction in the tail resolutions by $\sim 25 - 45\%$ and a reduction in the number of events that fail kinematic cuts by 19.1%. This comes at the expense of 2.4% of signal efficiency (3σ region). In addition, we note that the ML track selection has an additional advantage that it yields a similar tail resolution in the Monte Carlo (true resolution) to the double turn analysis; without this selection the double turn analysis underestimates the tail resolutions. This comparison with the double turn analysis is shown in the next section.

5.6.4.4 Variable Importance

Here we estimate the importance of all input variables. One technique used to measure each variable's importance is to start with a simple neural network only using the number of hits on the fitted track, then adding other variables into the training sequentially. In each neural network, we measure the accuracy of validation data.

To illustrate the accuracy, we use an accuracy curve, that is, for a given accuracy in "good" tracks, what is the accuracy for "bad" track prediction i.e. given a total relative tracking efficiency (with respect to the standard cuts) of 96%, the selection removes $\sim$ 70% of the "bad" tracks that would otherwise be in the final dataset (passed the standard cuts). To

reiterate, this selection removes "bad" tracks over the entire momentum space. It is clear from the curves that the first few input values have a large amount of power, but even the last few variables continue to improve the neural network. This is another check that adding in additional variables improves our track selection.

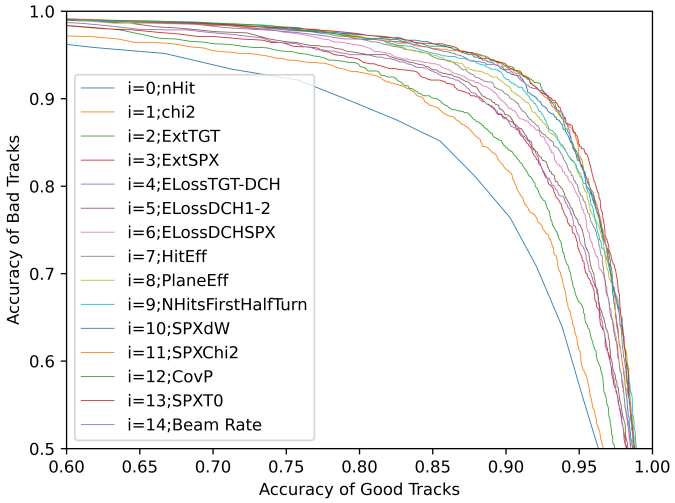


Figure 5.52: The accuracy of "good" tracks vs. the accuracy of "bad" tracks. We additionally plot this as a function of the number of variables in the neural network.

One criticism of machine learning when compared to a conventional approach is that it's often treated as a "black box", not knowing how the variables are used in the prediction. We use validation data to probe how the variables are used in the neural network.

In Figure 5.53, we histogram the average prediction value for each input variable shown in Figure 5.45. We integrate over all other variables. We expect that a track with measurable values closely associated with a "bad" track will have a high probability of being a "bad" track (P_{bad}). Note, the neural network probability output has a full range of 0-1. In addition, Figures 5.54a/5.54b use the same 2D variable pairs as in Figures 5.46a and 5.46b, but now are filled with the average prediction for a pair of variables. The histograms are as expected; given a low hit efficiency or a bad match between the positron track and the SPX hit, $< P_{bad} >$ is higher. The higher the peak value of $< P_{bad} >$, the more "power" each input

variable has. It's clear that the hit efficiency being less than 50% is one of the most powerful indicators of track quality; this is not surprising.

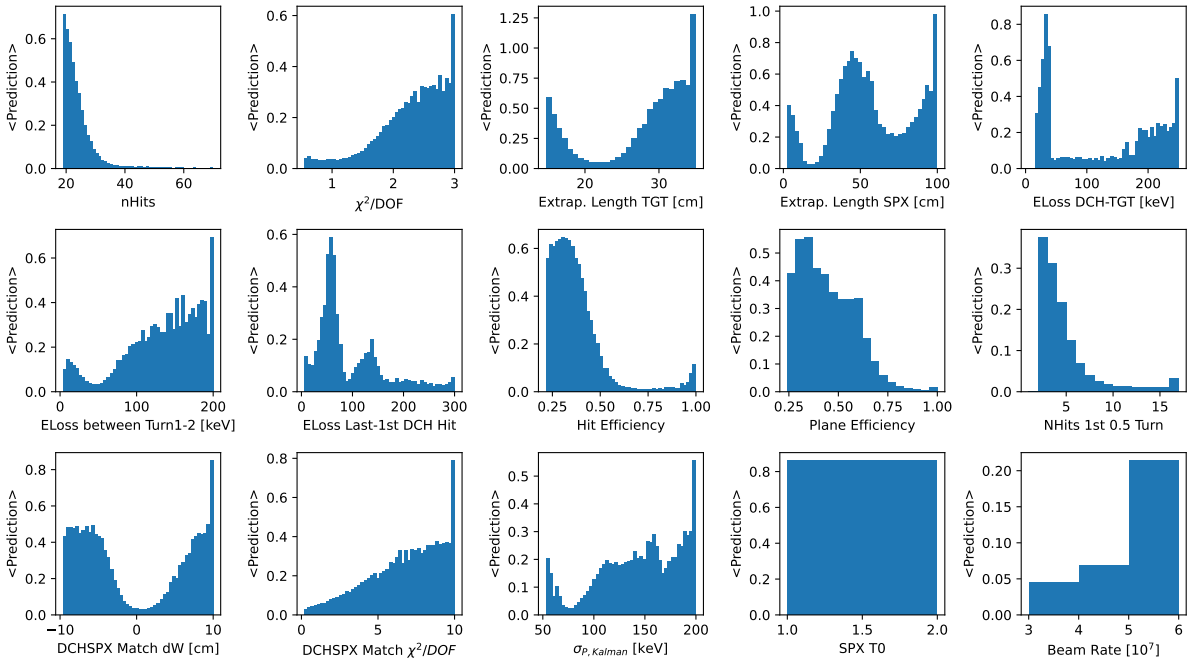


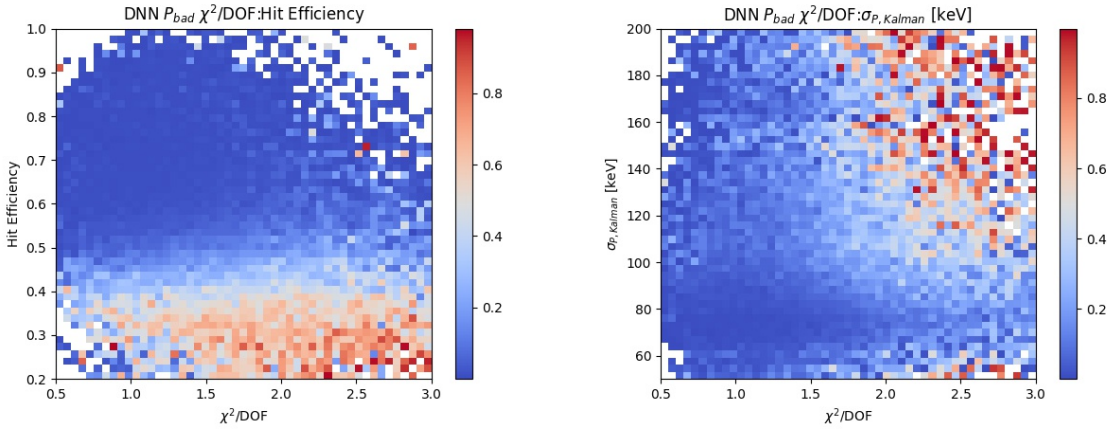
Figure 5.53: The average prediction as a function of all variables used to classify the "good" and "bad" tracks.

5.6.5 Final Track Selection and Ranking in the 2021 Dataset

Here we discuss the details of the implementation of track selection and track ranking in the final 2021 data set.

In the final 2021 analysis, the data was reprocessed twice: with and without the ML hit finding analysis (Yusuke Uchiyama). The ML hit finding yielded an improved tracking efficiency ($\sim$ 25%), but at a degraded resolution ($\sim$ 15% in all kinematics). Therefore, we decided to combine the datasets preferentially selecting "ghosts" (reconstructed tracks associated with the same physical track) from the reprocessing without ML hit finding.

Since tracks with and without ML hit finding have slightly different distributions (number of



(a) (b)

Figure 5.54: The two plots show the average prediction for two pairs of meaurables (a) The mean prediction is shown as a function of the number of hits and the χ^2/DOF.(b) The mean prediction is shown as a function of Kalman σ_P and the χ^2/DOF.

hits, χ^2, etc.), a Boolean was added to the ML track selection indicating whether or not the track was from the ML hit finding reprocessing. As mentioned, the likelihood analysis only uses one reconstructed positron track for each event. In the analysis, there can be instances of several physical positron tracks in the region of interest. Further, there are frequently multiple versions of the same physical positron track (i.e. ghost tracks). Therefore we require a procedure to select a single reconstructed track.

In this subsection, we discuss the track selection criteria, the ghost ranking and then the selection of a single positron track.

5.6.5.1 Final Track Selection

First, the standard MEG II analysis applies a set of selection criteria to all reconstructed tracks. This includes selection on all the variables mentioned above (number of hits on the fitted track, χ^2/DOF, diagonal elements of the covariance matrix, propagation length to the target, etc.) and the ML selection threshold. These thresholds are listed below. The ML selection threshold is currently set by removing a large fraction of the high momentum tails

without removing too many tracks: $\sim 7\%$ efficiency loss at the default threshold of 0.1.

5.6.5.1.1 Track Criteria

- Successful fit, propagation to the target
- Distance between the CDCH track at the 1st SPX hit and the estimated SPX hit position < 10 cm
- Time between the CDCH track at the 1st SPX hit and the estimated SPX hit time < 15 ns
- CDCH track must propagate within 3 cm of the physical SPX tile
- Fiducial target elliptical limit $(A_{In.Frame} - 2 \cdot \sigma_{z_e}\ , B_{In.Frame} - 2 \cdot \sigma_{y_e})$
- Number of hits on the fitted tracks $>= 18$
- Number of hits on the fitted track on the first half turn $>= 5$
- $chi^2/DOF < 13/3$ - number of hits on fitted track/60
- $\sigma_{p_e} < 0.3$ MeV
- $\sigma_{\phi_e} < 50$ mrad
- $\sigma_{\theta_e} < 12$ mrad
- $\sigma_{y_e} < .5$ cm
- $\sigma_{z_e} < .5$ cm
- Extrapolation length from the CDCH to the target < 45 cm
- Extrapolation length from the CDCH to the SPX < 80 cm
- T_0 from successful propagation to SPX cluster
- ML Prediction < 0.1

The standard selection criteria are loose in order to maximize the positron tracking efficiency. The track selection is then cleaned up by the ML track selection described previously. Note, even after the ML track selection, there is some $\sim 20 - 30\%$ (depending on the beam rate) of tracks have a propagation length to the SPX greater than 50 cm, which implies that both algorithms designed to add these hit have failed. This will be explored in the future.

5.6.5.2 Ghost Ranking

We only use the "best" track in the final analysis, this relies on the same ML track selection. The tracks are ranked based on the ML track selection's prediction probability of being a "good" track ($1 - P_{bad}$). Determining which reconstructed tracks correspond to the same physical track is based on the matched SPX cluster. It's highly unlikely that two physical tracks will hit the same SPX cluster of SPX tiles at the same time. This has an error rate significantly less than 1%.

However, there are two additional points. First, as mentioned above, the ML hit finding tracks have degraded kinematic resolution (double turn analysis). We found that without an explicit force, there are instances when the ML track selection chooses a ML hit finding track (e.g. with more hits), but it actually has a degraded resolution (using the double turn analysis or the Michel edge fit). We therefore force the track ranking to preferentially select tracks without ML hit finding over any track with ML hit finding. The ML track predictions are between 0-1 (1 being a bad track); we add 1 to the ML hit finding tracks. Second, some physics double turn tracks (2.5 turns) are missing the first full turn. This results in a target vertex time displaced by roughly 2 ns (and an incorrect position and angle at the target). The hit efficiency, plane efficiency, and propagation length to the target all appear consistent with a 1.5 turn track and therefore the track appears to be of good quality (from the point of view of the selection criteria). There are instances with the following ghosts on a single physical 2.5 turn track: ML hit finding results in a full 2.5 turn track and the no ML hit finding reprocessing only results in a 1.5 turn track (matched to the same SPX). In this case, regardless of the ML hit finding flag, we select the 2.5 turn track. The 2.5 turn track is the only track with a reasonable time, vertex, etc. at the target, even if it is lower quality. We add 2 to the ranking per turn.

The final track ranking is below where MLRank represents the ML ranking (0-1, 0 being

the best track), the MLHit is a Boolean representing if the track is from the ML hit finding reprocessing, and NTurns represents the number of turns on the fitted track.

$$\text{Rank} = \text{-MLRank - MLHit} + 2 \cdot \text{NTurns} \tag{5.3}$$

5.6.5.3 Physical Track Selection

At this stage, each physical track only contains one reconstructed track. We then apply the following kinematic cuts using both the positron and photon kinematics:

5.6.5.3.1 Kinematic Cuts

- $|t_{e,TGT} - t_{\gamma,TGT}| < 3ns$
- $50 < p_e < 56MeV$
- $48 < E_\gamma < 58MeV$
- $-0.988 < cos(\theta_{e,\gamma}) < -1.01$
- $|\theta_{e,\gamma}| < 0.1$ rad
- $|\phi_{e,\gamma}| < 0.1$ rad

An extremely small fraction of events have two physical positron tracks that pass the above selection criteria. At the current moment, the track is chosen which has the smallest relative angle between the positron and the photon.

5.6.6 Conclusions

In this section, we presented a neural network application to track selection and track ranking. In the MEG II dataset, positrons with a momentum $p_{e^+} > \frac{m_{\mu^+}}{2} + X \cdot \sigma_p$ are mismeasured by at least $X \cdot \sigma_p$. The neural network learns the differences in the measurable distributions

between "good" tracks ($p_{e^+} < \frac{m_{\mu^+}}{2} + X \cdot \sigma_p$) and "bad" tracks ($p_{e^+} > \frac{m_{\mu^+}}{2} + X \cdot \sigma_p$). The neural network creates a 1-dimensional probability of being a "bad" track (0-1) using a combination of the many measurable input variables (16). By training directly on data, the selection has no reliance on the Monte Carlo simulation unlike other approaches. We show that a ML track selection threshold that removes 5-10% of the data nearly completely eliminates the high energy tail seen in the Michel momentum distribution; we also show that the selection did not bias the momentum distribution. In addition, when applying the neural network to the Monte Carlo, we observe a large suppression of tracks with low kinematic resolution in all kinematic variables, not only the momentum (i.e. tails). This is applied in the standard 2021 data.

5.7 Kinematic Resolutions and Correlations

5.7.1 Introduction

The physics analysis relies on kinematic resolution estimates to calculate the probability of an event being signal or background. In this chapter, we estimate the kinematic resolution, width of the pull distributions (described later), the correlations between the kinematic variables and the correlations between the pull distributions in the 2021 data set. We present this study at $3 \cdot 10^7$ Hz beam rate, but it was done for all beam rates (here). The resolutions are estimates using the double turn analysis technique on both the data and Michel MC samples.

In the MEG II detector, approximately 15% of all positron tracks intersect the drift chamber 5 times before leaving the CDCH outer shell (2.5 turn track). For these events, we independently fit and measure the two turns on these tracks (1 turn and 1.5 turns respectively). By propagating both turns to a common plane parallel to the target between the two turns,

we compare the kinematics and thus estimate the kinematic resolution for standard tracks. That is, histogram $(X_2 - X_1)$ for kinematics $p_e, z_e, y_e, \phi_e, \theta_e$. Note, the final momentum resolution estimate uses the Michel edge fit, this was described in the previous chapter.

The physics analysis takes advantage of measured correlations between the kinematic variables to properly assess the likelihood any given positron is signal or background. Some of these correlations can be observed directly in the data using the double turn analysis. Other correlations require the MC and thus have additional uncertainty in the measurement.

In the 2021 data, all MC distributions are well-described by a double gaussian. To properly estimate the resolution of a single turn in the double turn analysis, we fit the double turn distributions to the convolution of two equal double Gaussians. This will be described in more detail below. This likelihood analysis requires the resolutions for signal positrons, and the correlations between the kinematic variables, both are described here.

The final physics analysis takes into account the Kalman filter event-by-event covariance matrix. That is, using the hit uncertainties, position of the hits, track residuals, propagation length to the target, and the number of hits on the track, the Kalman filter estimates the kinematic resolutions at the target $(p_e, z_e, y_e, \phi_e, \theta_e)$. The covariance matrix is a good estimator of the resolution, but needs corrections resulting from the covariance matrix elements on average being too large or too small. These corrections are estimated by measuring the pull distributions in both the double turn distributions $((X_2 - X_1)/\sigma_{X,1,2}$ where $\sigma_{X,1,2}$ is the square root of the relevant Kalman filter covariance matrix diagonal element of the first and second turns added in quadrature) and the MC $((X_{Rec} - X_{MC})/\sigma_{X,Rec})$ where $\sigma_{X,Rec}$ is the square root of the reconstructed track's relevant diagonal element of the covariance matrix). In both cases, the pull distribution is expected to have a zero mean with a width (Gaussian σ) of 1. However, the widths are often in the range of $0.9 - 1.3$ and therefore we must correct the covariance matrix elements accordingly. In addition, the pulls are often not described by a single Gaussian; in our case, the pulls are well described by a sum of

two Gaussians. The widths of the pull distributions are measured here. This analysis also requires the correlations between the pull distributions, not the kinematic variables. A separate set of correlation parameters (describing the same physical correlations) for the pulls is provided here.

5.7.1.1 Kinematic Distributions and Track Selection

For all data, we apply the standard track selection using the MEGPhysicsSelection class; this includes the ML selection described in the previous chapter. We do not include the cuts on the relative positron photon kinematics. In the case of "ghost" tracks (described in the previous chapter) we only use the "best" track. The "best" track is based on track selection and ranking described in the previous section.

Figure 5.55 presents the reconstructed positron kinematic distributions for Monte Carlo signal and Michel positrons, and for 2021 MEG trigger data. The data has a different beam center, significant entries in the carbon fiber frame of the target, and a significantly different ϕ_e distribution. The angular difference is due to the fact that we do not simulate events based on the MEG trigger, but only restrict the angular distributions of the positrons.

5.7.2 Kinematic Resolutions

In this subsection we estimate the kinematic resolutions using both the double turn analysis and the Monte Carlo comparisons. The goal of the section is to estimate the resolution in the data for signal.

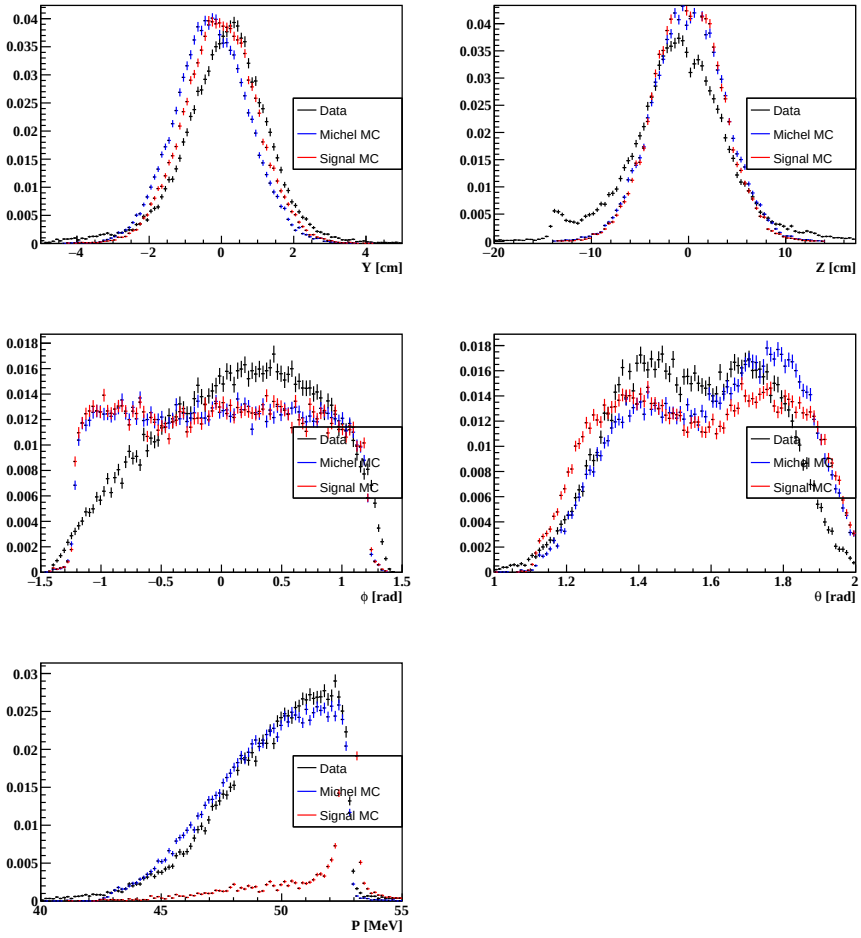


Figure 5.55: The kinematic distributions for data and MC.

5.7.2.1 MC Resolutions

First, we present the kinematic resolutions in the Monte Carlo. To avoid pileup tracks entering the histograms, we also include a simple cut that requires the reconstructed time of the positron at the target to agree with the MC within 2 ns. Of course a more sophisticated technique could be used to remove accidentals, but the current technique should not bias the distributions.

The distributions are shown below for signal positrons in Figure 5.56 at $3 \cdot 10^7$ Hz; the distributions are fit to double gaussians. Note that we found that the histograms are well fit

to a when $A_2/A_1 = 0.1$ (fit $\chi^2/DOF < 2$) in nearly all cases. For simplicity this has been fixed for all histograms except p_e, θ_e. In the case of θ_e, p_e, we restrict $A_2/A_1 = 0.15$. In these cases, allowing a variable A_2/A_1 did not significantly improve the fit's χ^2/DOF.

The tails eventually diverge from the fit on the order of < 10 entries/bin with a maximum of ~ 5000 entries/bin. The most significant divergence from the fit occurs when the reconstructed momentum is too small. This is likely due to the fact that some positrons experience bremsstrahlung. The positrons have a small probability of losing a large amount of energy in a material whereas Kalman only removes the typical amount of energy from different physics processes.

Note, in the case of the ϕ_e resolution, we present the resolution for positrons where $|\phi_e| < 0.2$ rad. We also include the correlation between the errors in ϕ_e, p_e however this is a very minor effect. In addition, the y_e resolution contains the correlation between the errors in y_e, p_e. Finally, we note that these wide tails can be further suppressed by including a more aggressive cut using the ML track selection, but this results in a lower efficiency.

All of the resolutions can be found in the attached Google Sheets. As an example, we tabulate the signal and Michel resolution at $3 \cdot 10^7$ Hz below in Table 5.1.

Table 5.1: We list the resolutions as determined by the MC comparison. We fit the distributions to a double gaussian where the tail has an amplitude fixed to 0.1 the amplitude of the core (except θ_e where $A_2/A_1 = 0.15$).

MC Double Gaussian Components $X_{Rec} - X_{MC}$					
Data	Y [mm]	Z [mm]	Φ [mrad]	Θ [mrad]	P [keV]
Signal Core	0.64	1.65	5.18	6.47	84.4
Michel Core	0.75	1.79	5.39	6.76	87.09
Signal Tail	1.95	4.65	13.10	16.33	235.6
Michel Tail	2.20	4.79	14.33	17.42	240.8

In the MC signal sample, the vertex resolution is effectively at the design of the experiment, the momentum is significantly improved with respect to the design, and the angular reso-

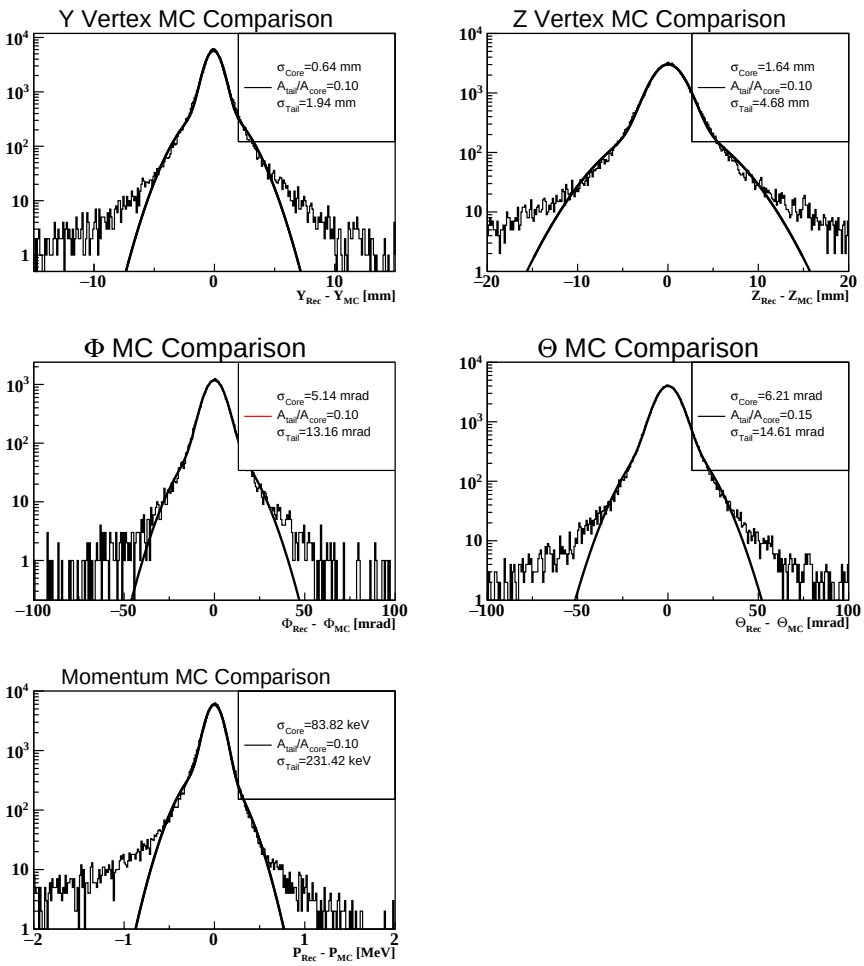


Figure 5.56: The reconstructed kinematics are compared to the true MC kinematics. The distributions are fit to a double gaussian in a wide region. Distributions are on a log scale.

lutions are degraded with respect to the design. Even the angular resolutions are now not far off ($\sigma_{\theta_e}, \sigma_{\phi_e} = 6.5, 5.2$ mrad compared with 5.3,3.7 mrad respectively). We note that the ϕ_e resolution in the design paper is likely also taking into account the correlation between errors in ϕ_e, z_e; this is a large effect. Taking into account this correlation, ϕ_e resolution ~ 3.66 mrad (achieving the design resolution).

5.7.2.2 Double Turn Resolutions

Here we discuss the results of the double turn analysis. First, Figure 5.57 shows an example of a 2021 data example of a double turn event.

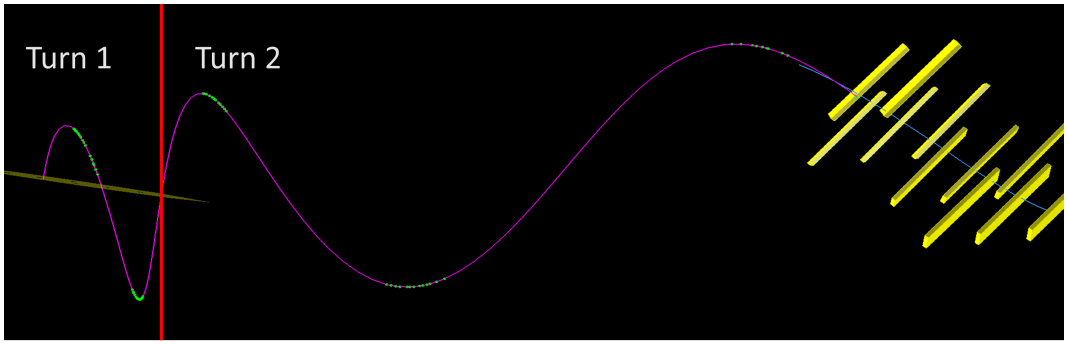


Figure 5.57: An example of a double turn positron track. The green dots show intersected wires with signal in the drift chamber; the yellow tiles show the pixelated timing counter tiles with signal.

In addition to the standard selection, we also require both turns to have at least 18 hits on the fitted track. This hit requirement is the same as that in the standard track selection. The double turn comparison is done after all tracking algorithms have been applied (turn merging, updating of T_0, refit for additional hits, machine learning doca estimate, etc.).

The 1st turn (2 intersections) and the final 1.5 turns (3 intersections) have different resolutions. This is particularly true for the momentum; without the final half turn, the momentum is degraded significantly. As mentioned above, the kinematic difference between the first and second turns contains the resolution of first and second turns added in quadrature. We fit the distribution to a convolution of two double gaussians; we assume the two double gaussians are identical for the first and second turns. This is of course, an approximation, but ends up yielding a quality resolution estimate for both the tail and the core of the kinematic resolutions. The distributions are shown below for the 2021 data 5.58.

We tabulate the double turn resolutions σ for 2021 data and Michel MC events both at $3 \cdot 10^7$ Hz in Table 5.2. We also tabulate the MC corrections to convert this double turn resolution

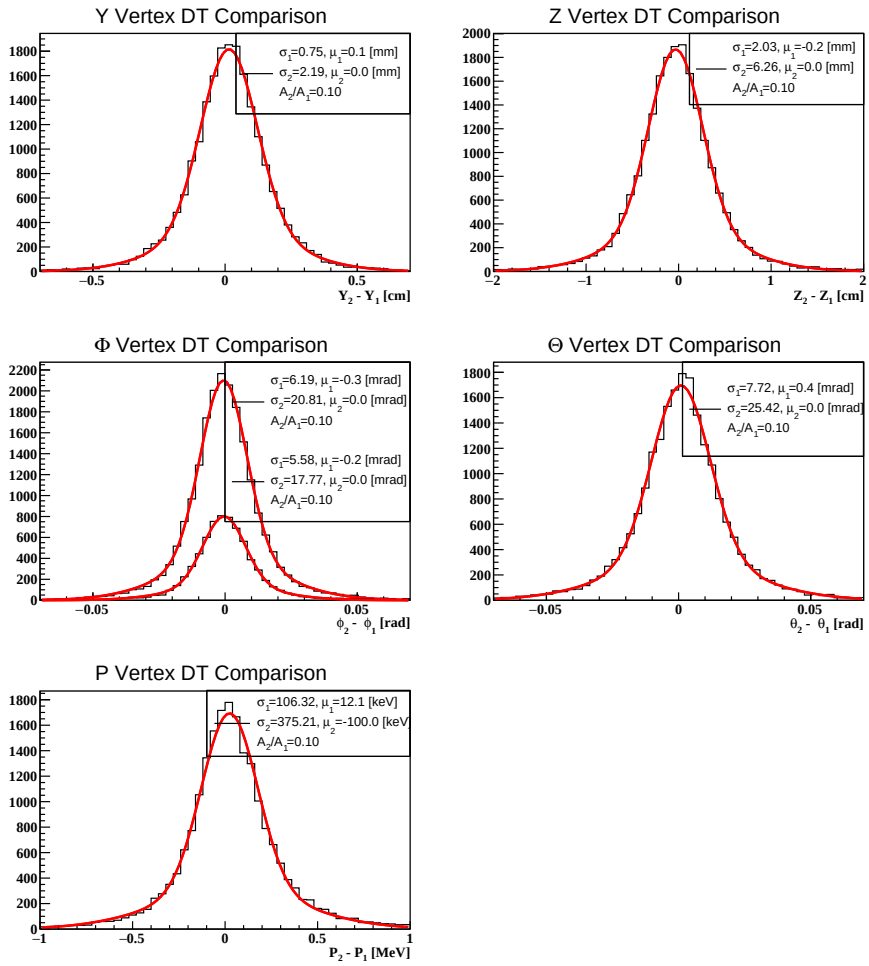


Figure 5.58: The double turn resolutions for data at $3 \cdot 10^7$ Hz. The technique compares the kinematics of the 1st and 2nd turns at a common plane parallel with the target. The smaller of the two ϕ_e comparisons only contains comparisons where the first turn's $|\phi_e| < 0.2$ rad.

into an estimate of the signal positron resolution in data.

Ideally, there is good agreement between the true MC resolutions and those estimated by the double turn resolutions for Michel MC data. The calculated ratio of the two resolutions (true resolution and that estimated by the double turn analysis) in Michel MC data are between 0.93-1.12 for the core and 0.84-1.15 for the tail components i.e. corrections are of the order 10%. There are two main effects that influence these corrections. The first is that the double turn resolution assumes the two turns have equal resolution, therefore, we expect the double turn resolution to be worse (particularly for the momentum). The table indicates

Table 5.2: The resolutions as determined by the double turn analysis in both data and Michel MC data. We list the relevant ratios of kinematic comparisons in order to extract a data estimate of the signal resolution. The $*$ represents the ϕ_e resolution including the correlation between ϕ_e, z_e.

Double Turn Gaussian Components $X_2 - X_1$					
Data	Y [mm]	Z [mm]	Φ [mrad]	Θ [mrad]	P [keV]
2021 Data Core	0.75	2.03	5.58	7.72	106.32
2021 Data Core !MLHit	0.70	1.88	5.24	7.02	101.25
Michel MC Core	0.67	1.71	4.85	6.85	93.6
2021 Data Tail	2.19	6.26	17.77	25.42	375.21
2021 Data Tail !MLHit	2.09	5.95	17.58	24.18	359.85
Michel MC Tail	1.92	5.24	15.77	19.60	288.12
Ratios to Estimate Signal Resolution in Data					
Data	Y	Z	Φ	Θ	P
MC Signal/MC Michel Core	0.85	0.88	0.96	0.96	0.97
MC Michel/DT Michel Core	1.12	1.05	1.08	0.99	0.93
MC Signal/MC Michel Tail	0.89	0.93	0.92	0.94	0.98
MC Michel/DT Michel Tail	1.15	0.91	0.91	0.89	0.84
2021 Data-driven Single Gaussian Signal Resolution					
Data	Y [mm]	Z [mm]	Φ [mrad]	Θ [mrad]	P [keV]
Data Signal Core	0.72	1.86	5.76/4.07*	7.29	95.87
Data !MLHit Signal Core	0.67	1.73	5.41/3.81*	6.63	91.30
Design Signal Core	0.7	1.6	3.7	5.3	130
Data Signal Tail	2.22	5.30	14.83	21.18	306.76
Data !MLHit Signal Tail	2.12	5.04	14.67	20.15	294.20

that the true MC core resolution is $\sim 7\%$ narrower than that predicted by the double turn ($\sim 16\%$ narrower than the tail resolution). The second effect is that the double turn tracks have already had to pass intrinsic analysis "cuts" to output a double turn track: the track fitter must successfully fit the two full turns, the two turns must successfully propagate to a common plane between the tracks, satisfy a loose merging criteria, and the two turns must

successfully merge. Therefore we'd expect the double turn resolutions to not contain the long tails; this ends up not being a dominant effect as in general the double turn tails are actually wider than that of the MC.

Finally, we note that the resolutions in data are $\sim 10\%$ degraded with respect to the Michel MC. This difference is the result of many differences between the data and the MC. This includes the signal and noise on the waveforms, magnetic field errors, wire alignment errors, etc. Further, a critical point is that the analysis on the Monte Carlo excludes a few analysis techniques that were developed based on the data. In particular, the ML hit finding technique is not included; excluding this software (only using double turn events without ML hit finding) results in narrower double turn distributions in data by $\sim 7\%$. This is also shown in Table 5.2. We observe a high level of consistency between the data without ML hit finding and the Michel double turn resolutions.

5.7.3 Pull Distributions

The final maximum likelihood analysis relies on the covariance matrix from the Kalman filter. The likelihood analysis includes a variable resolution function based on the number of fitted hits on the track, the χ^2/DOF, etc. This yields a large improvement to the analysis sensitivity. However, we must verify that the resolution estimated by the covariance matrix match the true resolution. The general technique to modify the Kalman estimated resolution is to calculate the pull distributions. In the MC, the pull for a kinematic variable X is defined as:$(X_{Rec} - X_{MC})/\sigma_{X,Rec}$ where $\sigma_{X,Rec}$ is the relevant diagonal element of the covariance matrix. In the double turn analysis, the pull is defined as: $(X_2 - X_1)/\sigma_{X,1,2}$ where $\sigma_{X1,2}$ is the Kalman filter covariance matrix diagonal element of the first and second turns added in quadrature). Below we calculate the width of the pull distributions in the MC for Michel and signal positrons. We then calculate the width of the pull distributions in Michel MC

and the 2021 data using the double turn analysis. We then use the combined information to estimate the pull distribution of signal positrons in the data.

5.7.3.1 MC Pull Distributions

Here we present the kinematic pull distributions in the Monte Carlo. All selection criteria are the same as described previously. The pull distributions are shown for signal positrons at $3 \cdot 10^7$ Hz beam rate in Figure 5.59. The resolutions are tabulated for MC Michel and signal positrons in Table 5.3.

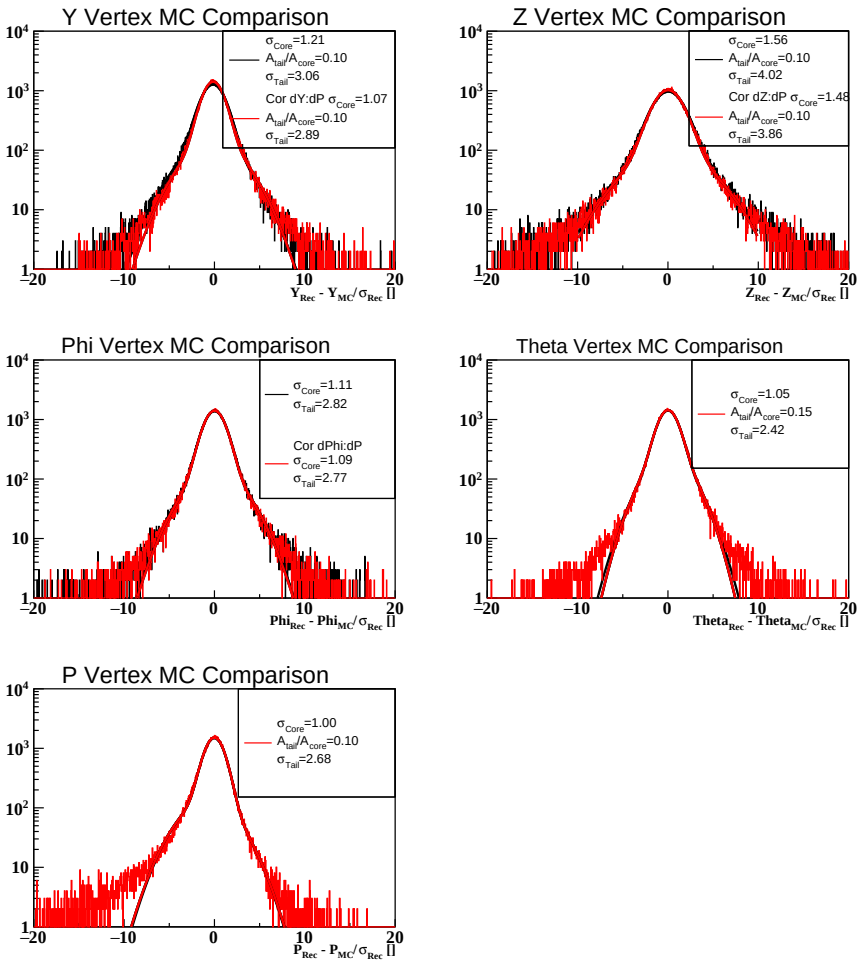


Figure 5.59: The pull distributions for signal MC positrons. The distributions are fit to a double gaussian in a wide region. Distributions are on a log scale.

Table 5.3: The pull distribution widths as determined by the MC comparison. We fit the distribution to a double gaussian where the tail has an amplitude fixed to 0.1 the amplitude of the core (except θ_e where $A_2/A_1 = 0.15$).

MC Pull Double Gaussian Components $X_{Rec} - X_{MC}$					
Data	Y []	Z []	Φ []	Θ []	P []
Signal Core	1.07	1.56	1.11	1.05	1.00
Michel Core	1.06	1.61	1.13	1.05	1.00
Signal Tail	2.89	4.65	2.82	2.42	2.68
Michel Tail	2.84	4.02	2.87	2.46	2.51

The distributions are fit to a double gaussian using the same restriction that $A_2/A_1 = 0.1/0.15$. Unlike the phi_e resolution, the ϕ_e pull uses all ϕ_e values as the covariance matrix should be handling the difference in resolution at different ϕ_e. In addition, the y_e pull contains the correlation between $y_e : p_e$. The most significant divergence from the fit occurs when the reconstructed momentum is too small; this was described in the previous section. In general the pull widths are close to 1, except for z_e, which is around 1.6. It's unclear why the core of the z_e pull is significantly wider than the others.

5.7.3.2 Double Turn Pulls

We use the same selection stated above when calculating the double turn pulls. The Kalman should already include information indicating that the resolution of the first turn (two chamber intersections) is degraded with respect to the second turn (three chamber intersections). Therefore, the double turn pull distributions should account for the resolution of both turns, and so we fit the histograms to a sum of two gaussians. The double turn pull distributions for the 2021 data is shown in Figure 5.60.

We tabulate the double turn pull widths for 2021 data and Michel MC events both at $3 \cdot 10^7$ Hz in Table 5.4. We also include corrections from the MC comparison needed to convert this into a pull width for signal positrons.

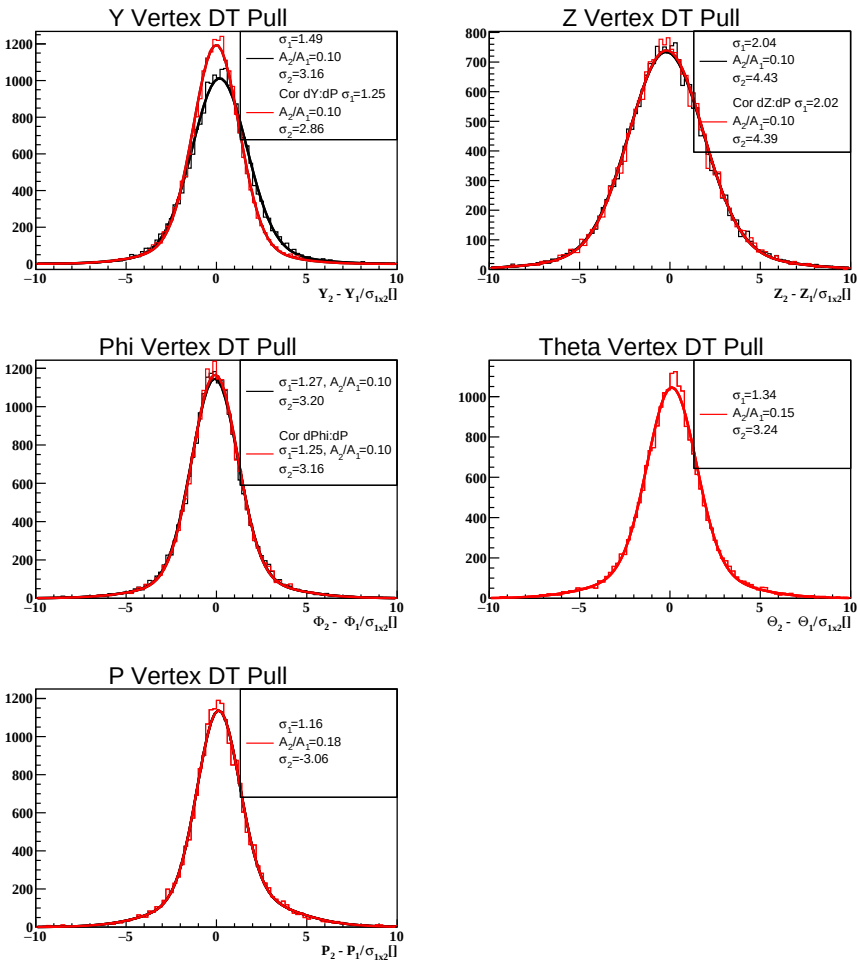


Figure 5.60: The double turn pull distributions for data at $3{\cdot}10^7$ Hz. The technique compares the kinematics of the 1st and 2nd turns at a common plane parallel with the target divided by the Kalman estimated resolution of the first and second turns added in quadrature.

The closer the width is to 1, the closer the covariance matrix is to correctly estimating the resolution. For the most part, the covariance matrix underestimates the resolutions by $\sim$ 20% in the core. Further, the covariance has no tail component, which we observe is typically $\sim$ 3x the core. Note that again, all core σ for the z_e pull is between 1.5-2. This could be a consequence of the σ_{DOCA}, which is also a single Gaussian and does not take into consideration the tails of the DOCA resolution. The typical value of σ_{DOCA} is about $110\mu m$, but our DOCA resolutions contain a significant tail of $300\mu m$. We have observed in the MC that the DOCA resolution has the largest fractional effect on the z_e resolution.

Table 5.4: The results of fitting the pull distributions in the double turn analysis in both data and Michel MC data to a sum of two gaussians. We list the relevant ratios of kinematic comparisons in order to extract a data estimate of the signal pull widths.

Double Turn Gaussian Components $(X_2 - X_1)/\sigma_{X1,2}$					
Data	Y []	Z []	Φ []	Θ []	P []
2021 Data Core	1.25	2.04	1.27	1.34	1.17
Michel MC Core	1.13	1.62	1.19	1.19	1.17
2021 Data Tail	2.86	4.43	3.20	3.24	3.06
Michel MC Tail	2.50	3.82	2.65	2.57	2.97
Ratios					
Data	Y	Z	Φ	Θ	P
MC Signal/MC Michel Core	1.01	0.97	0.98	1.00	1.00
MC Michel/DT Michel Core	0.94	0.99	0.95	0.88	0.85
MC Signal/MC Michel Tail	1.02	1.03	0.98	0.98	1.07
MC Michel/DT Michel Tail	1.14	1.02	1.08	0.96	0.85
2021 Data-driven Double Gaussian Pulls					
Data	Y []	Z []	Φ []	Θ []	P []
2021 Data Signal Core Pull	1.18	1.96	1.21	1.18	1.00
2021 Data Signal Tail Pull	3.18	4.78	3.42	3.07	2.66

Next, we note that the differences between the pull width in the Michel MC truth comparison and the Michel MC double turn comparison yield a very comparable result in all cases; the ratio of the cores is between 0.85-0.99 in all kinematics i.e. the double turn technique provides a slightly wider core. We also see a small between the signal and Michel MC positrons, especially in the core of the distributions. This implies that the covariance is correctly identifying the improved resolution for signal positrons. Finally, we note that the data double turn pulls are about $\sim 10\%$ wider than those of Michel positrons in the MC. This is consistent with the fact that the data resolutions are also $\sim 10\%$ wider than that of the Michel MC in the double turn analysis. This implies that the covariance is comparable in the data and the MC.

5.7.4 Correlations

In this subsection, we estimate the positron correlations observed in MEG II. All correlations were observed in MEG I. An exhaustive search for correlations between all pairs of kinematic variables as a function of positron angles (θ_e, ϕ_e) has not been done. The correlations with a $*$ symbol indicate that the correlation is observed in the double turn analysis. In MEG I, correlations 5.4, 5.5,5.6, 5.7 are explained in detail in TN70[55]. Correlations 5.9 and 5.8 were found later and are only described briefly in TN84[56] for MEG I. These correlations are listed below:

$$*\delta\phi = [p_0^\phi + p_1^\phi \cdot tan(\phi)] \cdot \delta E \tag{5.4}$$

$$*\delta Y = p_0^Y + p_1^Y \cdot \delta E \tag{5.5}$$

$$*\delta Z = [p_2^Z + p_3^Z \cdot cot(\theta)] \cdot \delta E \tag{5.6}$$

$$\delta Z = p_0^Z + p_1^Z \cdot \delta\theta \tag{5.7}$$

$$*\delta\phi = [p_5^\phi + p_6^\phi \cdot \phi + p_7^\phi \cdot \phi^2] \cdot \delta Z \tag{5.8}$$

$$\delta\phi = [p_2^\phi + p_3^\phi \cdot \phi + p_4^\phi \cdot \phi^2] \cdot \delta\theta \tag{5.9}$$

Several kinematic variables are correlated with the positron momentum (z_e, y_e, ϕ_e). If the positron momentum is mismeasured to be larger than the true momentum, the radius of curvature will be larger; this on average results in a systematic change in the positron vertex and emission angle at the stopping target. For example, in a simple case where a positron has a radius of curvature of ~ 25 cm, a fractional error of 100 keV/52.8 MeV ~ 0.0019 and thus changing the radius of curvature by $25 \cdot 0.0019 = 500\mu m$; this mainly affects the y_e coordinate and the angle in ϕ_e.

We are searching for the $\mu \to e\gamma$ signal with a known momentum of 52.83 MeV. Therefore, in the signal positron probability density function, the positron z_e, y_e, ϕ_e kinematics will be shifted based on the difference between the measured momentum and the signal momentum. All of these correlations are observable in the Monte Carlo and the double turn analysis.

Next, correlation 5.7 between the positron's vertex z_e and θ_e at the target is another geometrical consequence: given the positron is measured with an error in θ_e, one should expect a linear relationship between the error in θ_e and z_e. This is only visible in the Monte Carlo, not the double turn technique as explained here[55].

Finally there are the 5.9 and 5.8 correlations. Both correlations are dependent on the positron ϕ_e at the target. The fact that there are two correlations is almost exclusively due to the previous mentioned correlation between z_e, θ_e. That is, simply taking into account the correlation between z_e, ϕ_e makes the correlation between θ_e, ϕ_e negligible. Only the z_e, ϕ_e correlation is visible in the double turn technique therefore, we choose to first account for the correlation between z_e, ϕ_e and then inspect the θ_e, ϕ_e correlation. We will show later that the remaining θ_e, ϕ_e is negligible. The correlation can be thought of as producing a similar effect as an error in the target position. That is, an error in z_e nominally shifts the target's normal coordinate by $sin(15^\circ) * dz_e$; this shifts the ϕ_e coordinate. This effect is accentuated at large $|\phi_e|$. Roughly $dz_e = 2mm; dx_{TGT} \sim 0.5mm$, thus changing the ϕ_e by ~ 5 mrad.

5.7.4.1 Correlations between Kinematic Variables

First, we start by showing the correlations between the kinematic variables in the data (beam rate $3 \cdot 10^7$ Hz) in Figure 5.61. It's clear that the 5.4, 5.5,5.6, 5.8 correlations are all visible in the data double turn analysis. We additionally show the correlation between the positron ϕ_e and the squared error in ϕ_e (i.e. $(\phi_{e,2} - \phi_{e,1})^2$) as this is also incorporated into the average PDF likelihood analysis.

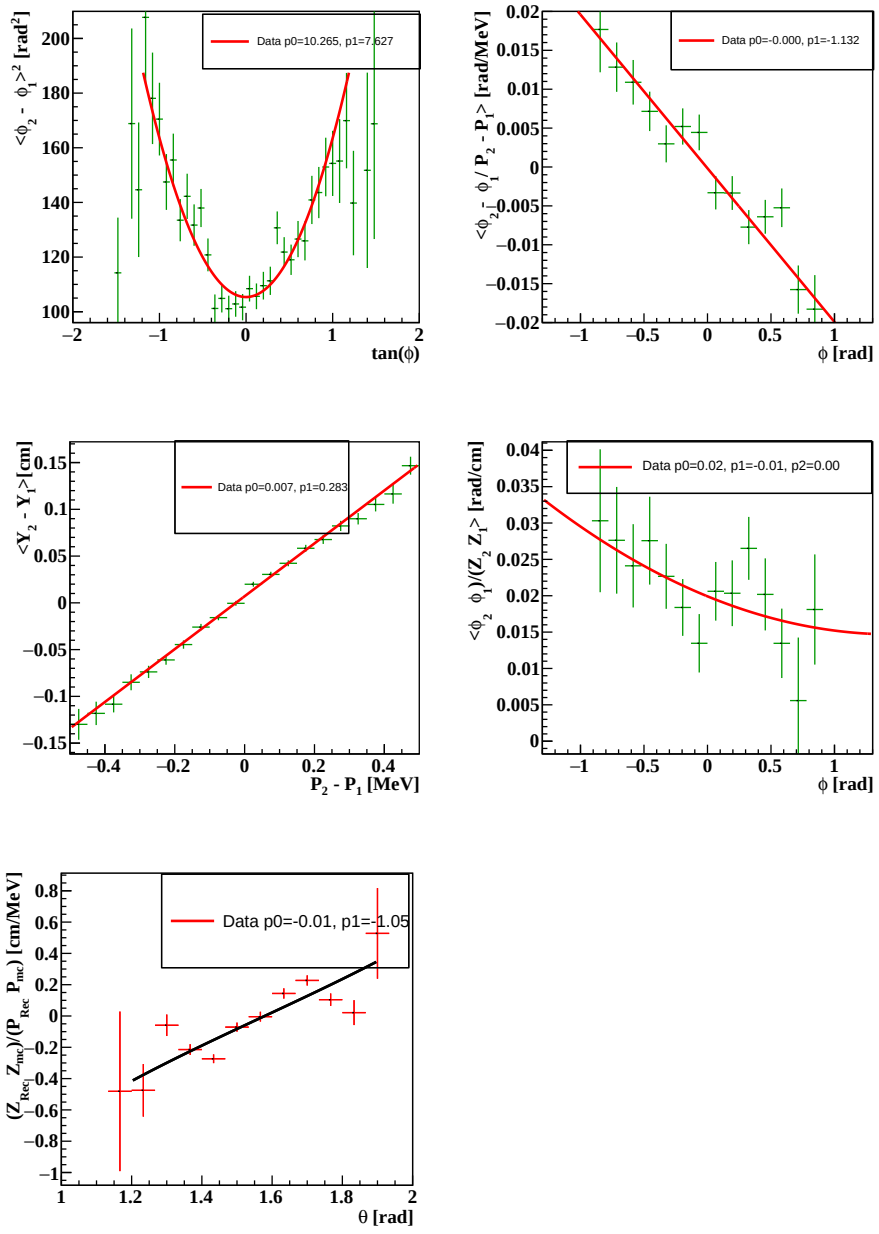


Figure 5.61: The correlations in data using the double turn analysis. This does not include all possible correlations as some are not accessible by this technique.

Next, Figure 5.62 shows the correlations in the MC comparison for signal positrons at $3 \cdot 10^7$ Hz.

To be clear, we calculated the correlations between the kinematic variables for the data using the double turn technique, the Michel MC using the double turn technique, and in the MC

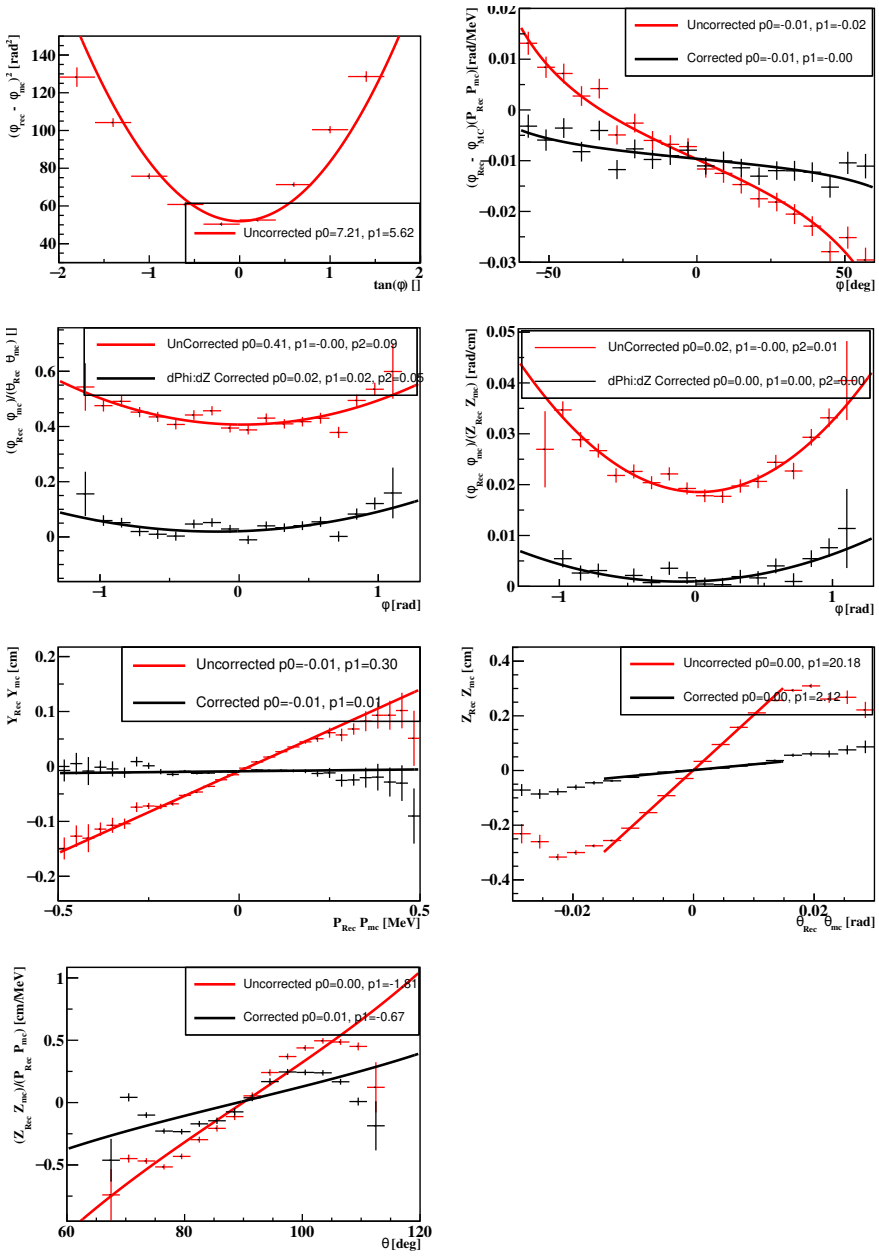


Figure 5.62: The correlations between the kinematic variables by comparing the reconstructed kinematics to the true MC kinematics for signal positrons.

using signal and Michel MC data (4 total data types). We then estimated these correlations for all beam rates. Therefore, we made 16 versions of the correlations.

5.7.4.2 Comparison of Datasets and Techniques

Here we describe the details of the agreement in the correlations between all datasets. Frequently, we calculate the typical value in data, then apply a small correlation based on the MC. In other cases when the data disagrees significantly with the MC, we rely on the MC.

For correlation 5.4, using all 16 versions, we calculate $\sigma_{p_1^\phi}/\mu_{p_1^\phi}$ or the fractional fluctuations in this parameter to be 9%. We also calculate: $\frac{p^\phi_{1,MCMichel}}{p^\phi_{1,DTMichel}} \cdot \frac{p^\phi_{1,MCSignal}}{p^\phi_{1,MCMichel}} = 0.946$. This factor is used to convert the data double turn correlation parameter into one for signal. We calculate the data $p_1^\phi = -0.0143$ rad/MeV. We rely on the average MC signal value for $p_0^\phi = -0.0096$ rad/MeV as it is significantly different in the signal MC than other data sets.

Next, we discuss correlation 5.5. We calculate $\sigma_{p_1^Y}/\mu_{p_1^Y}$ or the fractional fluctuations in this parameter to be 6.30%. We calculate: $\frac{p^Y_{1,MCMichel}}{p^Y_{1,DTMichel}} \cdot \frac{p^Y_{1,MCSignal}}{p^Y_{1,MCMichel}} = 1.010$ to convert the data double turn correlation parameter into one for signal. We calculate the data $p_1^Y = 0.314$ cm/MeV. In all techniques we find p_0^Y to be consistent with zero.

For correlation 5.8, the most relevant parameter is p_5^ϕ. This is effectively the correlation between $d\phi_e, dz_e$ at $\phi = 0$. We calculate $\sigma_{p_5^\phi}/\mu_{p_5^\phi} = 9.7\%$ and $\frac{p^\phi_{5,MCMichel}}{p^\phi_{5,DTMichel}} \cdot \frac{p^\phi_{5,MCSignal}}{p^\phi_{5,MCMichel}} = 0.88$. We calculate $p_5^\phi = 0.0176$ rad/cm using this correction factor. On the other hand, the linear and quadratic terms (p_6^ϕ, p_7^ϕ) do not have the same structure in the double turn technique and the MC comparison; in particular in the Monte Carlo, the $d\phi_e, dz_e$ dependence on ϕ has a significant quadratic term and a negligible linear term. We set $p_6^\phi = 0$ and take p_7^ϕ from the signal MC ($p_7^\phi = 0.01313$ 1/(cm*rad)).

Next, the $\phi_e : \theta_e$ correlation is effectively eliminated once the $\phi_e : z_e$ correlation is accounted for with a correlation $d\phi_e/d\theta_e \sim 0.02$ at $\phi_e = 0$ or $d\theta_e \sim 14 mrad$ (2 σ) creates a 0.3 mrad systematic. This has a very small effect on the eventual ϕ_e resolution ($\sim 90\%$ of events has a correction smaller than 0.4 mrad). In addition, this correlation is not visible in the double

turn analysis and therefore has a larger uncertainty.

Correlation 5.6 was excluded in MEG I physics analysis as it was a small correction. In MEG II including this correlation improved the z_e resolution of $\sim 5-10\%$. The slope of the correlation's dependence on θ_e is different in the Monte Carlo comparison and the double turn comparison by ~ 2; for this reason, we use the signal MC to measure $p_3^Z = -1.907$ cm/MeV. The curve does not fit perfectly to a cot function and therefore a higher order polynomial fit could be used in the future. All values of p_2^Z are consistent with zero.

Finally, we discuss correlation 5.7. As mentioned above, this is only observable in the MC. Therefore, we simply take the average values calculated in the MC for signal ($p_1^Z = 20.29$). The intercept (p_0^Z) is consistent with zero.

The qualitative and quantitative results discussed are summarized in Table 5.5.

Table 5.5: The correlation parameter values. The DT row represents if the parameter is visible in the double turn analysis. The zero row is Boolean depending on if the parameter is zero consistent. The origin row states how the value was measured. DTC abbreviates the double turn analysis data value corrected using the MC. MCS abbreviates the signal MC value.

Correlation Parameters														
Param	p_0^ϕ	p_1^ϕ	p_2^ϕ	p_3^ϕ	p_4^ϕ	p_5^ϕ	p_6^ϕ	p_7^ϕ	p_0^Y	p_1^Y	p_0^Z	p_1^Z	p_2^Z	p_3^Z
Units	$\frac{rad}{MeV}$	$\frac{rad}{MeV}$		$\frac{}{rad}$	$\frac{}{rad^2}$	$\frac{rad}{cm}$	$\frac{}{cm}$	$\frac{}{cm*rad}$	cm	$\frac{cm}{MeV}$	cm	$\frac{cm}{rad}$	$\frac{cm}{MeV}$	$\frac{cm}{MeV}$
DT	N	Y	N	N	N	Y	Y	Y	Y	Y	N	N	Y	Y
Zero	N	N	Y	Y	Y	N	Y	N	Y	N	Y	N	Y	N
Origin	MCS	DTC	-	-	-	DTC	-	MCS	-	DTC	-	MCS	-	MCS
Final Value	$-9.6\cdot 10^{-3}$	$-14.3\cdot 10^{-3}$	-	-	-	$17.6\cdot 10^{-3}$	-	$13.1\cdot 10^{-3}$	-	0.314	-	20.29	-	-1.91

5.7.4.3 Correlations Between Pull Distributions

In this subsection we revisit the same six correlations described in the previous section, but measure the correlations between the kinematic pulls. These correlations are needed for the event by event physics analysis.

Note that correlation 5.4 and the Kalman σ_ϕ are both functions of ϕ. Therefore we fit correlation to an alternate function shown in Equation 5.10. Here, the s represents the pull. This effectively assumes that the ϕ resolution has a $tan(\phi)^2$ dependence on ϕ and a resolution of $\sigma^2_{\phi,0}$ at $\phi = 0$. In TN84 $p^\phi_0 = c_\phi$ and $p^\phi_1 = k_\phi$. The c_ϕ is the correlation at $\phi = 0$, which is non-zero especially in the case of signal positrons. The k_ϕ represents the ϕ dependence of the ϕ resolution. We simply fit for all three parameters: $p^\phi_0, p^\phi_1, \sigma^2_{\phi,0}$.

$$\frac{s\phi}{sE} = \frac{p^\phi_0 + p^\phi_1 \cdot tan(\phi)}{\sqrt{\sigma^2_{\phi,0} + (p^\phi_1 tan(\phi))^2}} \tag{5.10}$$

For the remaining correlations: 5.6,5.7,5.8,5.9, we maintain the same fit, but replacing the kinematic errors with the pulls.

Same as before, correlation 5.7 and 5.9 are not visible in the double turn distributions. Figure 5.63 shows the correlations visible to the double turn analysis in the 2021 data at $3 \cdot 10^7$ Hz. Note, correlation 5.4 is well fit to the equation in 5.10.

The correlations between the pulls for Monte Carlo signal at $3 \cdot 10^7$ Hz are shown in Figure 5.64.

5.7.4.4 Comparison of Datasets and Techniques

Here we compare the correlations between the pull distributions from the double turn analysis and the Monte Carlo comparison.

First, we discuss correlation 5.10. Here, $\sigma_{\phi,0}$ is effectively the same in all techniques with a mean of 6.46 and a fractional RMS of 4.6%. We extract p^ϕ_0 using MC signal positrons; the mean is -1.056. When calculating p^ϕ_1, an outlier with a significantly larger correlation is

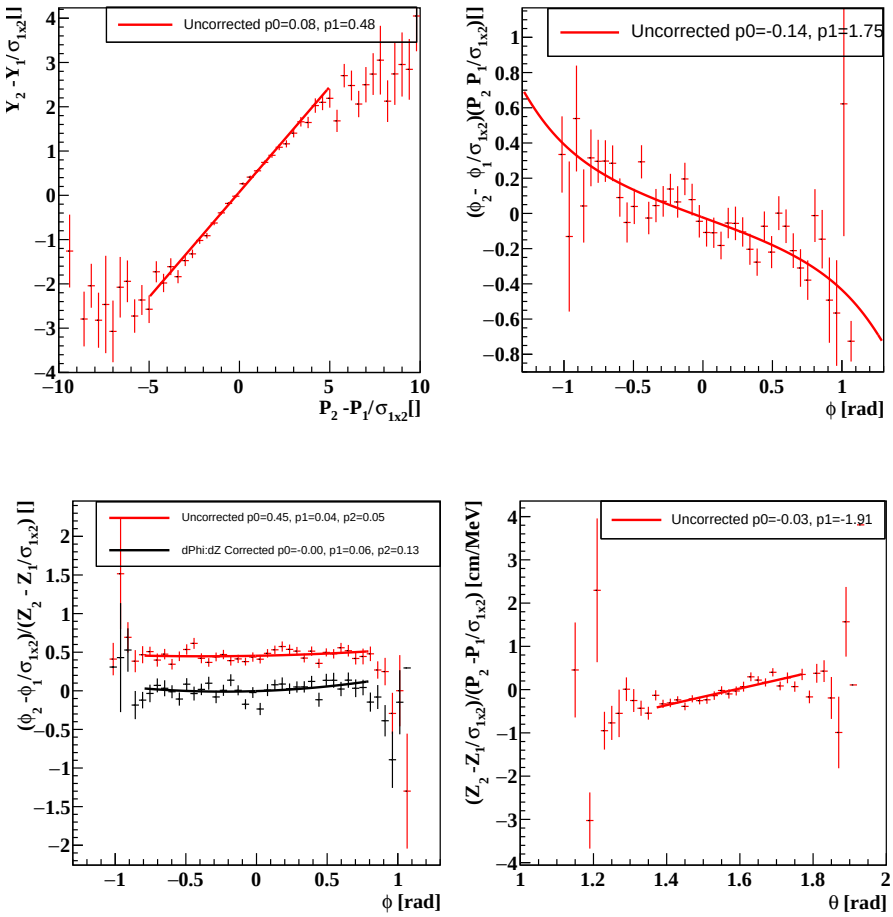


Figure 5.63: The correlations between the kinematic pulls in the 2021 data using the double turn technique. The denominator contains an uncertainty contribution from both turns.

observed only in the double turn analysis on Michel MC data at $5{\cdot}10^7$ Hz. All other values (15 total) agree within 16% of one another (full RMS). We calculate $\frac{p^{\phi}_{1,MCMichel}}{p^{\phi}_{1,DTMichel}} \cdot \frac{p^{\phi}_{1,MCSignal}}{p^{\phi}_{1,MCMichel}} \sim 0.75$. We then correction factor to the data estimate: $p^{\phi}_1 = 1.46$.

The correlation between the p_e and z_e pulls (correlation 5.6) has a θ_e dependence that is smaller in the double turn than that observe in the Monte Carlo. Therefore, we rely on the Monte Carlo signal value of -2.56.

The measured correlation between the pulls of ϕ_e, z_e (correlation 5.8) has a fractional RMS of 5.19% between all 16 measurements. We calculate the correction factor: $\frac{p^{\phi}_{5,MCMichel}}{p^{\phi}_{5,DTMichel}}$ ·

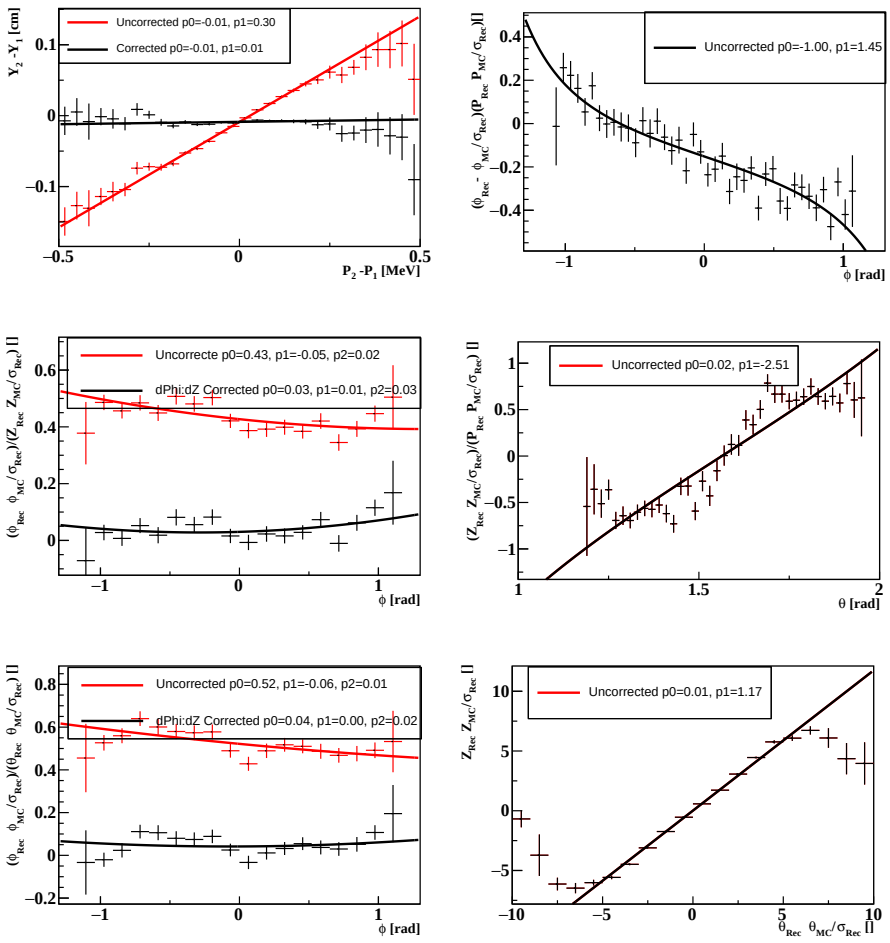


Figure 5.64: The correlations between the kinematic pulls in the Monte Carlo for signal at $3 \cdot 10^7$ Hz. The denominator is the Kalman uncertainty in the reconstructed kinematic.

$\frac{p^{\phi}_{5,MCSignal}}{p^{\phi}_{5,MCMichel}} \sim 0.926$. We apply this correction to the average data value resulting in $p^{\phi}_5 = 0.421$. We also see a nontrivial linear dependence of the correlation on ϕ in all MC samples that is not observed in the double turn analysis. Therefore, we set $p^{\phi}_6 = -0.0578$, the average of all signal MC comparisons. The quadratic dependence on the correlation is negligible.

The correlation between ϕ_e, θ_e or correlation 5.9 is almost completely eliminated after taking into account correlation 5.8. We think it is best to just remove this correlation in 2021.

The correlation between θ_e, z_e is only observable in the MC. Using MC signal positrons, $p^{Z}_1 = 1.188$.

Finally, we tabulate the results in Table 5.6.

Table 5.6: The correlation parameter values between the pulls. The DT row represents if the parameter is visible in the double turn analysis. The zero row is Boolean depending on if the parameter is zero consistent. The origin row states how the value was measured. DTC abbreviates the double turn analysis data value corrected using the MC. MCS abbreviates the signal MC value.

Correlation Parameters															
Param	$\sigma_{\phi,0}$	p_0^ϕ	p_1^ϕ	p_2^ϕ	p_3^ϕ	p_4^ϕ	p_5^ϕ	p_6^ϕ	p_7^ϕ	p_0^Y	p_1^Y	p_0^Z	p_1^Z	p_2^Z	p_3^Z
Units										cm	$\frac{cm}{MeV}$				
DT	Y	N	Y	N	N	N	Y	Y	Y	Y	Y	N	N	Y	Y
Zero	N	N	N	Y	Y	Y	N	N	Y	Y	N	Y	N	Y	N
Origin	DTC	MCS	DTC	-	-	-	DTC	MCS	-	-	DTC	-	MCS	-	MCS
Final Value	6.46	-1.056	1.46	-	-	-	0.421	-0.0578	-	-	0.314	-	1.188	-	-2.56

5.7.5 Conclusions

In this chapter, we discussed the MEG II CDCH resolutions and the widths of the pull distributions in data using the double turn analysis and the Monte Carlo. We also estimate the correlations between the kinematic variables and the pulls using the double turn analysis and the Monte Carlo.

First, we discuss the signal positron resolutions with respect to the design of the experiment. Removing the ML hit finding tracks results in a kinematic resolution improvement of roughly $\sim 5-10\%$ as seen in Table 5.2. This results in a signal positron core resolution of 0.67 mm, 1.73 mm, 5.40 mrad, 6.7 mrad, and 95.1 keV for $y_e, z_e, \phi_e, \theta_e, p_e$. The vertex position and the momentum are already comparable with the resolutions proposed in the design of the experiment. The ϕ_e resolution quoted in the design of the experiment has actually not been achieved even in Monte Carlo simulations since shortly after the design paper (Francesco Renga Sept. 2018 $\sigma_\phi \sim 5.5$ mrad over all ϕ). Personally, it seems that the 3.7 mrad goal seems difficult to achieve unless this resolution estimate takes into account the correlation between $\phi_e : z_e$. The θ resolution is still large with respect to that of the design, but

hopefully could be improved via improved pileup discrimination, improved noise reduction on the waveforms, or improved track fitting algorithms.

Further, we note that the ML hit finding algorithm is still a new algorithm. It's possible that this could result in comparable resolutions to that of the tracks without this algorithm. In future analyses we likely want to assign a different resolution/ pull value for the two classes of tracks.

We'd also like to comment on the agreement between the data and the MC Michel tracks. Comparing the resolutions without any ML hit finding, the resolutions are very comparable with the data having a wider resolution in $z_e, \phi_e \theta_e, p_e$ by roughly $\sim 8 - 10\%$ and nearly the same resolution in y_e. It's possible that there are still differences between the signal and noise on the waveforms, etc. between the data and the MC. The ML DOCA estimator is also not applied to the MC data (yielding improvements in the kinematics of the order $5 - 13\%$). But it is clear that further improvements in the wire alignment, TXY, magnetic field map update etc. should not yield a very dramatic improvement in the resolutions.

Next, we discuss the resolution as a function of beam rate. This information is only available in the attached Google Sheets document. Comparing the data at $2{\cdot}10^7$ Hz to that of $5{\cdot}10^7$ Hz results in a degradation of the resolutions by roughly $\sim 10\%$. This suggests that suppression of pileup effects at higher beam rate could potentially result in improvements of this order.

Finally, we discuss the beam rate dependence of the correlations. We observe a slight beam rate dependence in p_1^Z however the $(min - max)/max$ of all values has a difference of $\sim 3\%$. Since this value was taken from the MC, it must be assigned an uncertainty significantly larger than the observed differences so it seems acceptable to use the mean value for all conditions. This is also true for p_3^Z, which has a $(min - max)/max$ of $\sim 10\%$, again this is also taken from the signal MC so it seems acceptable to use one value for all beam rates.

Chapter 6

MEG II Alignment

In this chapter, we give an overview of the internal detector alignment for the LXe, SPX, CDCH detectors and the relative alignments of the mentioned detectors, the target and the magnetic field map. We link several documents (some only available to MEG II collaborators) that describe the alignments in more detail.

6.1 Introduction

The alignment of the MEG detectors is critical to achieve the desired kinematic resolutions of the muon decay products and thus our sensitivity to $\mu \to e\gamma$. First, the detectors must be aligned internally i.e. LXe MPPC-to-MPPC, CDCH wire-to-wire, and SPX tile-to-tile. The importance varies detector-by-detector, but in general an internal misalignment degrades the kinematic resolution and can create biases in the kinematic measurements.

In addition, the detectors must be aligned with respect to one another. In general, the relative detector alignment shifts the relative kinematics of the positron with respect to the photon at the target (mainly $\theta_{e\gamma}, \phi_{e\gamma}$). Table 6.1 tabulates the detector alignments.

It contains the technique used to measure the alignment, the desired resolution (σ_D), and the measured resolution (σ_M). We estimate the desired resolutions in the later sections by calculating how misalignments affect the positron or photon kinematic measurements. In some cases, for example the relative wire-by-wire alignment in the drift chamber, this is not as straightforward and so we make an estimate based on double turn analysis results. The achieved resolutions have been pulled from a variety of talks.

σ_M= Measured σ, σ_D= Required σ	COBRA Cryostat	Magnetic Field	CDCH	LXe	SPX	Target
COBRA Cryostat	Optical Survey $\sigma_D < 1mm$ $\sigma_M < 1mm$	COBRA design coil positions	Survey	Survey	Survey	Survey
Magnetic Field	$\sigma_D = 250\mu m$ $\sigma_M \sim 1mm$	Field Map $\sigma_D \sim 2 \cdot 10^{-4}$ $\sigma_M \sim 1 \cdot 10^{-3}$	Michel Edge Double Turn			
CDCH	-	Ad-Hoc Corrections $\sigma_M \sim 250\mu m$	Construction, Iterative (Sect 5.5) $\sigma_D/\sigma_M \sim 10\mu m$	Cosmics in LXe+CDCH	e^+ Match at SPX	Hole Analysis
LXe	-		dZ_{COSMIC} $\sim 1mm$ $\sigma_M \sim 1mm$ $\sigma_D \sim 500\mu m$	FARO + X-ray $\sigma_D < 1mm$ $\sigma_M \sim 100\mu m$		
SPX	-		$\sigma_D \sim 1mm$ $\sigma_M < 1mm$		FARO + Laser $\sigma_D \sim 1mm$ $\sigma_M \sim 400\mu m$	
Target	-		($\perp$) $\sigma_D \sim 100\mu m$ $\sigma_M \sim 200\mu m$			Camera Analysis ($\perp$) $\sigma_D < 100\mu m$ $\sigma_M \sim 10\mu m$

6.2 MEG Coordinate System

All measurements in the MEG coordinate system refer to the alignment of the detectors with respect to a standardized set of fiducial markers defining the MEG coordinate system. For example, some of these markers are on the COBRA exterior and support structure, positions on the floor, walls, etc. This defines the "local area coordinate system". All detectors, the COBRA cryostat and the target were measured by an optical survey in this coordinate system.

The coordinate system is right handed with the z-axis (Z_{MEG}) nominally aligned with the beam axis. At the time of each survey, the local coordinate system is estimated using a laser tracker (i.e. the position and orientation of the laser tracker in the area coordinate system). The LXe detector is located at negative X_{MEG}, and thus signal positrons are emitted from the target at positive X_{MEG}. The Y_{MEG} coordinate is aligned against gravity.

Since there are many fiducial markers in the area, some are used to achieve the coordinate system and some are used later to validate the found coordinate system (i.e. the coordinate system can be repeated with a different set of fiducial markers as a check). This validation results in a measurement precision of $< 1mm$; this unfortunately results in ~ 1 mrad uncertainty in the mean value of $\theta_{e\gamma}, \phi_{e\gamma}$; this will be discussed later.

6.3 Target

6.3.1 Motivation

A target misalignment normal to the target surface results in an incorrect path length to the target and thus a ϕ_e error. Roughly a 90 μm misalignment normal to the target surface results in a 1 mrad ϕ_e error. Therefore, a resolution significantly less than a mm is required. This is illustrated in Figure 3.7 and was described in detail in Chapter 4.

6.3.2 Description

This alignment is determined by a variety of techniques. More detail can be found in Chapter 4. Briefly, an optical survey and a CT scan of the optical corner cubes yields the position of the target in the MEG coordinate system. The time-dependence of the target position is taken out via the target camera analysis. The time-dependence is expected due to the

insertion/extraction of the target for weekly (or biweekly) LXe calibrations.

6.4 CDCH Internal Alignment

6.4.1 Motivation

Misalignment of the relative wire positions, orientations, etc. results in systematics in the track residuals wire-by-wire. This degrades the overall positron kinematic resolution. In addition, we observed that wire misalignments of the order $\sim 100\mu m$ resulted in biases in the double turn analysis (e.g. $< \phi_{Turn2} - \phi_{Turn1} >! = 0$). For this reason, wire alignment to a precision of $< 10\mu m$ is desirable to suppress the misalignment's effect on the kinematic resolutions.

6.4.2 Description

The relative wire-by-wire alignment of the drift chamber was done by Macro Chiappini using a variety of measurements dating back to the construction of the drift chamber. The details are attached in these slides: Chiappini Alignment Slides. First, the relative position of the wire ends to the CDCH end plate spokes (16 total) was calculated. This was done using a precise camera at the time of the chamber's construction. An image is shown in Figure 6.1.

The wire coordinates in this local spoke-based frame are transformed into the global CDCH frame in each sector using the nominal values of end plate mechanical features and measurements of the spoke edges. This is done for both the upstream and downstream side independently. This effectively results in the wire coordinates at both end plates.

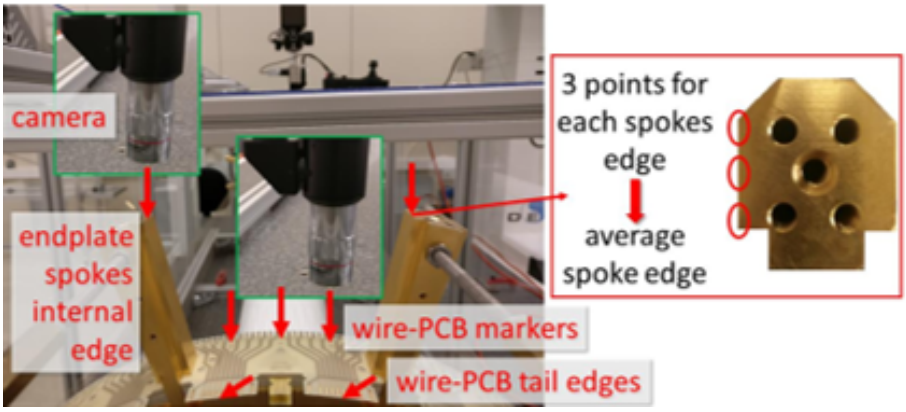


Figure 6.1: The relative spoke-wire alignment (Graphic made by Macro Chiappini).

Finally, the end plate is measured in the MEG coordinate system. This is shown in Figure 6.2. The measurement is done using a CMM touching probe at many points on the end plate. This measurement assumes the end plate is a flat plane; residuals of the order $\sim 100\mu m$ were observed. This results in the wire position/direction in the MEG coordinate system.

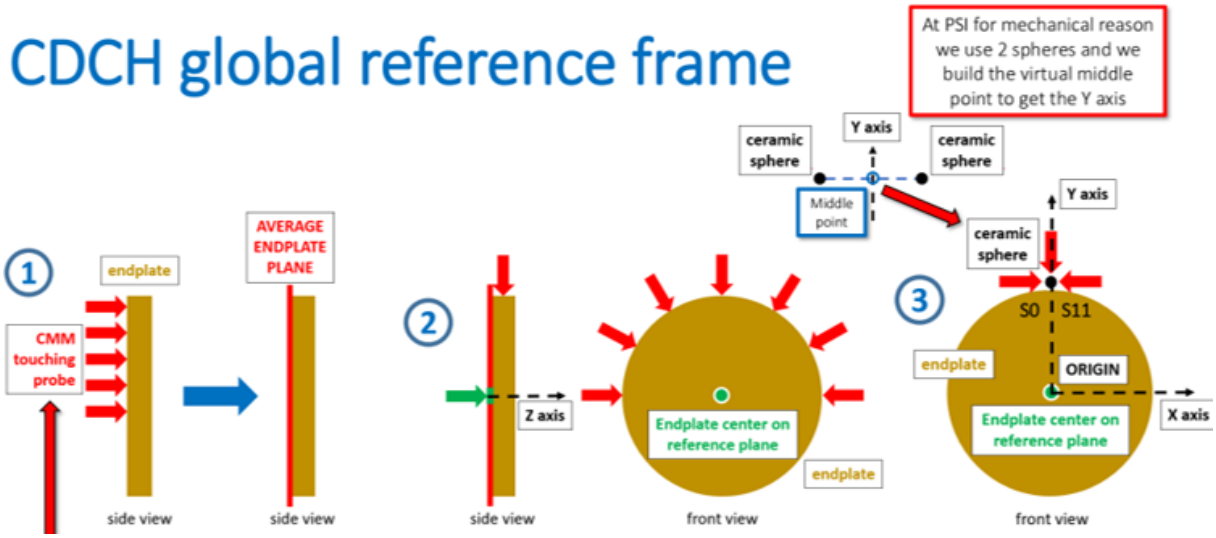


Figure 6.2: Here, we show a graphic illustrating how the end plate position is measured (Graphic made by Macro Chiappini).

When using this alignment, fitted positron tracks contained mean residuals in the coordinates normal to the wire axis that varied wire-by-wire. This implies the wire alignment had errors of the order $\sim 100\mu m$. These wire positions were corrected by translating, rotating, and applying a wire sagitta iteratively based on the observed residual. This is described in

Section 5.5. This resulted in improved track residuals and improved positron kinematic resolutions. This effectively eliminated biases observed in the double turn analysis.

6.5 SPX Internal Alignment

6.5.1 Motivation

Misalignment of the relative tile-by-tile alignment results in degraded positron timing resolution. The relevant position measurement is that normal to the tile surface i.e. perpendicular to the positron path.

The same level of precision required by the CDCH wire alignment is not needed for this alignment as we are only interested in the timing. For an e^+ traveling at c, a misalignment of 1 mm results in a timing error of 3 ps. The resolution per tile is $\sim$ 100 ps, therefore a 1 mm misalignment has a negligible contribution to the timing error. However, the goal precision was set to $\sim 1mm$ in order to have no contribution to the relative positron photon timing.

6.5.2 Description

A FARO technique was applied to estimate the relative position and orientation of the SPX tiles, this is described here: SPX FARO Measurement. An example graphic is shown in Figure 6.3.

A laser tracker is then used to measure the position of reference markers on the timing counter geometry inside the MEG coordinate system and at the time of the FARO measurement (a local coordinate system). This yields the position of the SPX tiles in the MEG coordinate

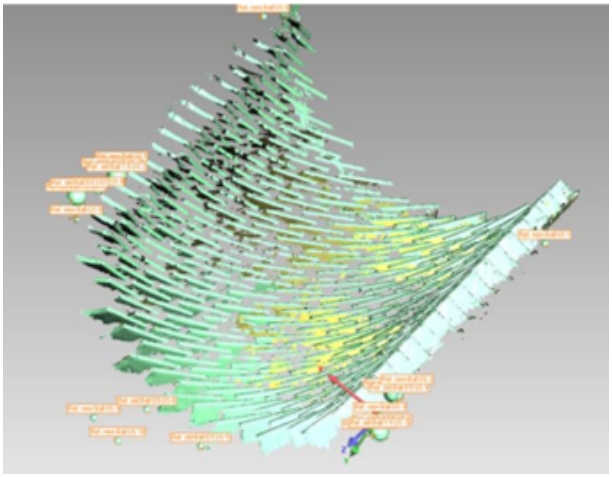


Figure 6.3: FARO SPX measurement (Graphic made by Yusuke Uchiyama).

system. Additional information is available here: SPX Alignment.

6.6 LXe Internal Alignment

6.6.1 Motivation

The position of the MPPCs on the inner-face must be known precisely to have a precise measurement of the photon intersection point (or the initial shower position) on the detector's inner-face in the global coordinate system. This photon position measurement is eventually propagated back to the positron vertex position at the target to compare the relative angles (and time) of the positron and the photon. The goal was set to measure the relative MPPC alignment to a precision of $\sim 100\mu$m to have a negligible effect on the relative angles of positron photon pairs. The photon energy and timing resolution degradation will be model-dependent. e.g. a simple charge integration over all LXe MPPC/PMTs will not affect the energy resolution.

6.6.2 Description

The LXe MPPCs were measured using two independent measurements. The first is a FARO measurement of the relative MPPC alignment when the detector is at room temperature, thus the positions may be different than that during MEG physics data collection. The technique did not scan all MPPCs as it was not possible to access them. The second technique used a collimated X-ray source to estimate the MPPC positions. The X-ray interacts in the calorimeter, producing a 1 mm^3 interaction measured by the MPPCs. The position and orientation of the X-ray source in the MEG coordinate system is known precisely using an optical survey. Using the relative position and orientation of the X-ray source and thus the intersection point on the LXe inner face, the MPPC positions in the MEG coordinate system are estimated. It is described in detail in this paper[10]. The X-ray source and the alignment setup is shown in Figure 6.4. It is also described briefly in Shinji Ogawa's thesis page 53[40].

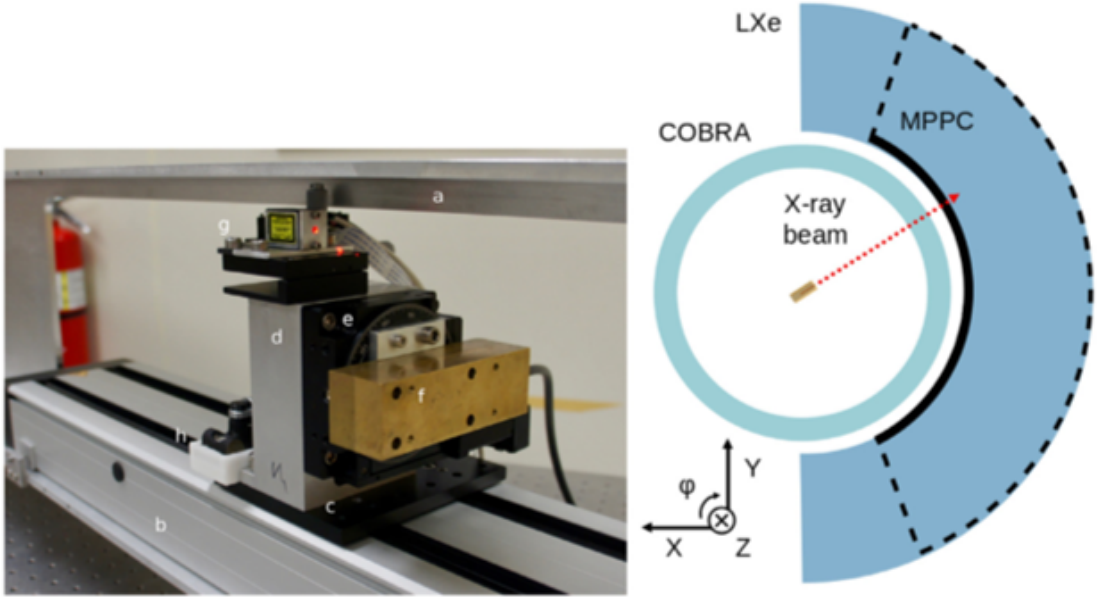


Figure 6.4: An image of the X-ray source and a graphic of the experimental setup (Figure 3 ©2023 Elsevier [10]).

6.7 Magnetic Field Measurements and CDCH Magnetic Field Map Alignment

6.7.1 Motivation

The magnetic field measurement is critical for measuring the positron tracks. This includes three vector $\vec{B}$ at all points inside the magnetic spectrometer. In general, a precision of $< 1/2000$ is desirable as this corresponds to a fractional resolution of ~ 50 MeV$/2000 = 25$ keV or adding $\sim< 4\%$ to the resolution when added in quadrature with a resolution of ~ 90 keV, even better would be ideal.

6.7.2 Discussion of the Magnetic Field Maps

There are several magnetic field maps available at the time of the 2021 physics data. We outline them here.

First, there is a calculated magnetic field based on the expected position and orientation of the magnet coils with respect to the exterior of the COBRA cryostat. The main difficulty of using this measurement is that it relies on a precise knowledge of both the CDCH and the COBRA magnet coils in the MEG coordinate system (more precisely, their relative position and orientation). This is difficult as the CDCH position in the MEG coordinate system is only quoted at $\sim 500\mu$m. In addition, the magnetic coils may shift when in use with respect to the relative position of the coils and the COBRA exterior estimated at the time of construction. Furthermore, of course, this field could include systematic errors with respect to the true magnetic field.

There are also three measured magnetic field maps. The first was taken in 2008, another

in 2014, and a third in 2018. Below in Figure 6.5, we show the discrepancy between the 2008 magnetic field map and the calculated field. There are systematics of the order $\sim 10^{-3}$. During the 2008 and 2014 magnetic field measurements, there was suspicions that the rail used to move the position of the Hall probes along the magnetic spectrometer may have bowed, thus resulting in systematic effects. In addition, it was found that the 2008 field map did not satisfy Maxwell's equations. This was handled by only using the B_z from the 2008 magnetic field measurement and relying on a fit to Maxwell's equations to yield the other two components. This fit to Maxwell's equation could be tried with the 2014 map, but was never investigated. The Hall probes used in the 2018 measurement were never successfully calibrated and thus the map was never used.

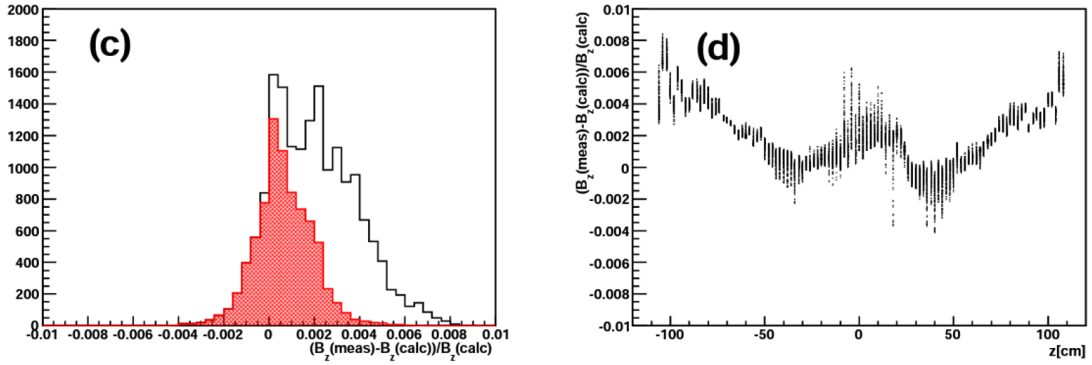


Figure 6.5: The discrepancy between the calculated magnetic field from the magnet coil positions and the measured magnetic field (Figure made by the MEG I collaboration).

Regardless of the map, we use the double turn analysis and the fit to the Michel edge were both used to determine the "best" magnetic field map. That is, in the double turn analysis, we optimize the double turn resolutions, the double turn biases (e.g. minimizing $| < \phi_2 - \phi_1 > |$), and the double turn biases as a function of different parameters (e.g. the two turn match point). In the Michel edge fit, we optimize the momentum resolution (in particular the core resolution), and the mean position of the Michel edge as a function of the positron emission angles (ϕ_e, θ_e). Ideally the mean position of the Michel edge is centered at 52.83 MeV and is flat as a function of emission angles.

In addition, these two techniques can be used to find the relative position and orientation of the CDCH and the COBRA magnetic field map and the optimal scale of the magnetic field map. This is a multidimensional optimization that requires the optimal magnetic field map, the optimal relative position and orientation of the magnetic field map and the CDCH, the optimal scale of the map, and a previous optimized wire-by-wire alignment.

When using the calculated field map without shifts or scaling, all of the double turn resolutions except the momentum ($P_2 - P_1 \sim -80$ keV) were near zero bias. In addition, it was quickly found that a shift in the relative position of the magnetic field map and the CDCH in Y_{MEG} resulted in a shift of $P_2 - P_1$. Thus a shift of 1 mm was applied in Y_{MEG}. This nearly centered all double turn histograms. This is shown in Figure 5.42.We then applied a scaling of 0.999 to force the Michel edge to be centered correctly at 52.83 MeV (1/1000 scaling).

The shift of 1 mm was ad hoc, but considering the optical survey of the CDCH precision ($\sim<$ 1 mm) and the uncertainty in the relative position of the magnetic coils with respect to the COBRA cyrostat, this seems warranted. This magnetic field measurement was later optimized further by Atsushi Oya by flattening the Michel edge further as a function of the positron emission angle. The optimal shift of the calculated field was found to be $(100, 700, 300)\mu m$ with an optimal scale of 0.9991. Note that the 2014 magnetic field map without any use of Maxwell's equations resulted in a comparable Michel edge momentum resolution, but larger biases in the double turn distributions. Applying Maxwell's equations to the 2014 field map in the future could lead to further improvements. The relative orientation of the calculated field with respect to the CDCH was not explored in much detail.

In conclusion, in the 2021 physics analysis the optimal magnetic field map was that calculated using the expected position of the COBRA coils with respect to the CDCH, with a shift of $(100, 700, 300)\mu m$ and a scale factor of 0.9991.

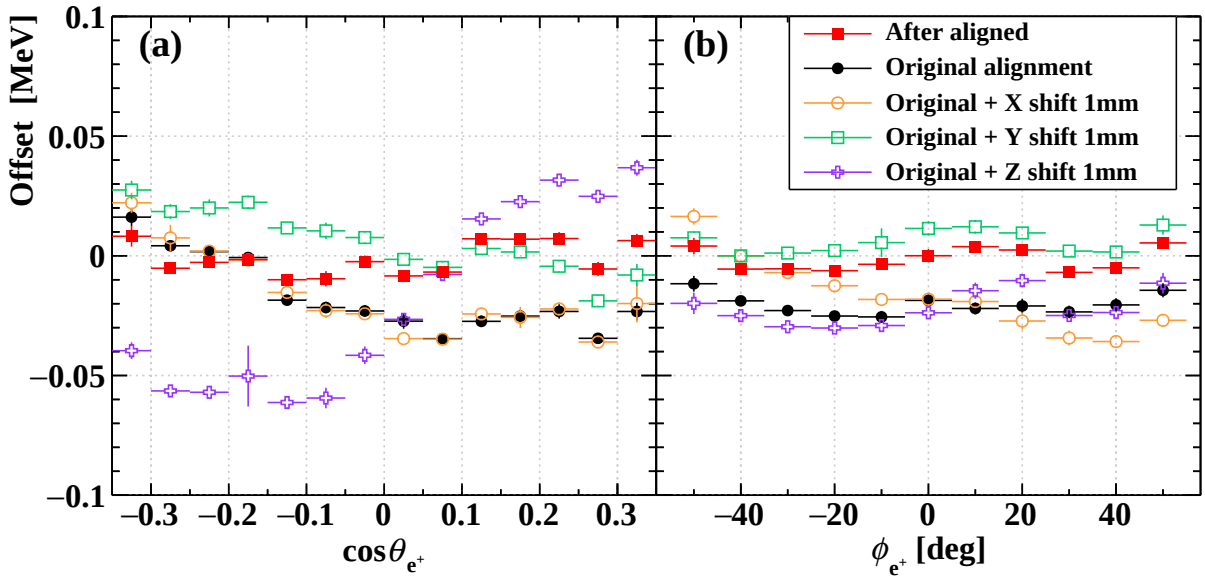


Figure 6.6: The location of the Michel edge as a function of the emission angles at the stopping target for varying shifted magnetic fields.

To investigate how errors in the magnetic field would effect the resolutions, we ran the MC with a magnetic field map (2008 map with Maxwell's equations) and then ran the analysis with varying shifts, scales, and different magnetic field maps. We observed that in general the $\phi_e, \theta_e, z_e, y_e$ resolutions were not significantly effected ($\sim< 5\%$), but the momentum resolution could be significantly degraded up to $\sim 25\%$. Therefore, with a perfect magnetic field map we could expect to improve the momentum resolution (rough guess of a few or 10%), but should not expect much improvement in the other kinematic resolutions.

6.8 CDCH SPX Alignment

6.8.1 Motivation

This alignment is required to optimize the path length of the positrons measured in the drift chamber to the timing counter. This requires the relative position of the drift chamber, the target, and the timing counter. A relative translation of the two detectors would result in

an incorrect path length. This is true for all positron tracks (except higher order effects e.g. dependence on track angle). At first order, the relative position of the CDCH and the SPX (tiles on one end) creates an offset in $t_{e\gamma}$; this is completely removed by using the $t_{e\gamma}$ RMD peak (separately for each SPX side of 256 tiles). e.g. the upstream SPX is offset with respect to its nominal position by 5 mm in Z_{MEG}; this would result in a movement of the RMD peak by 15 ps, which can be corrected for. Further, this extremely large offset in distance would produce a negligible effect on the timing even if uncorrected.

6.8.2 Description

This is most easily determined via fitted positron tracks. The positron track position at the SPX tile is estimated using the CDCH track information. For each SPX cluster, independent of the CDCH, there exists a "fitted" SPX hit position (using all the hits in a cluster) at each SPX tile in the cluster. For a given positron track at the first intersected SPX tile, we have a SPX-based hit position, and a CDCH-based hit position (propagating the track to the tile) that can be compared. However, one point is that this this error tile-by-tile can be corrected by two unique translations. One parallel to the tile surface and one perpendicular to the tile surface. In the current state of the analysis, these contributed a negligible effect to the timing resolution and thus a choice was made arbitrarily.

6.9 CDCH Target Alignment

6.9.1 Description

The target position relative to the CDCH is monitored using a technique known as the 'hole analysis'. This analysis images holes cut in the target surface by using the positron

vertex distribution at the target; the physical holes create a deficit in positron target vertex distribution. This checks the relative alignment of the CDCH and the target from the optical survey and the camera analysis. The critical point is to measure the position normal to the target surface. This is estimated using the correlation between ϕ_e at the target and the apparent y_e hole position on the target. This measurement is described in[11] and was performed by Atsushi Oya.

6.10 CDCH LXe Alignment

As mentioned above, the CDCH is aligned with respect to the target via the hole analysis. Second, the CDCH is aligned with respect to the SPX by comparing the track position at the SPX with the SPX hit positions. Therefore, the positron tracks measured in the CDCH can be used to align the CDCH-SPX-target components of the experiment. Therefore, the CDCH-LXe relative alignment completes the relative CDCH-SPX-LXe-target alignment.

The primary measurement of the relative alignment uses the internal alignments mentioned above, which yield the position of the CDCH wires/ LXe MPPCs in the MEG coordinate system. However, some of the relative CDCH-LXe alignment can be checked using cosmic rays.

6.10.1 Motivation

It is critical to know the relative position of the CDCH to the LXe in order to have precise measurements of the relative positron and photon angles. We motivate the precision requirements below.

6.10.2 X_{MEG} Mis-Alignment

First, we investigate the effect of a global misalignment of the LXe in the X_{MEG} direction. Here, we know the location of the signal $t_{e\gamma}$ from the RMD $t_{e\gamma}$ peak, therefore a uniform shift in $t_{e\gamma}$ caused by a misalignment of CDCH-LXe in X_{MEG} is irrelevant as it can be calibrated away.

In addition, a dX_{MEG} error produces a negligible position dependent effect on $\phi_\gamma/\theta_\gamma$ for a similar reason. e.g. 1 $Y_{MEG} = 500mm$,$dX_{MEG} = 1mm$, $Z_{MEG} = 0mm$, e.g. 2 $Y_{MEG} = 0mm$,$dX_{MEG} = 1mm$, $Z_{MEG} = 0mm$. In both examples, assume the positron target vertex is at (0,0,0). The photon momentum vector in the second example doesn't change, but the first example results in an angle change of 0.8 mrad. Therefore, we expect a mean angle change of $\phi_\gamma \sim 0.4$ mrad varying with Y_{LXe}. This is even less relevant for the θ_γ as Z_{MEG} only extends to $\sim \pm 25cm$. For the reasons mentioned, a global shift in dX_{MEG} of less than 1 mm shouldn't have a significant effect on $t_{e\gamma}$ or $\phi_\gamma/\theta_\gamma$ with a maximum error in ϕ_γ at the LXe edges in Y_{MEG} resulting in a systematic error of 0.8 mrad.

6.10.3 Y_{MEG}/Z_{MEG} Mis-Alignment

Next, we discuss global errors in either Y_{MEG}/Z_{MEG}. Assume there is a global shift of the LXe by 1 mm in Y_{MEG} i.e. to yield the photon direction, we add in quadrature (600, +1), (600, 500+1) at the center and the edge of the calorimeter respectively. In both cases, the effect on the timing is less than 2 ps. As for the effect on ϕ_γ, the effect is more significant. At the edge of the calorimeter, this produces a 0.98 mrad error, but at $Y_{MEG} = 0$, this produces a 1.7 mrad error.

The identical situation is true for a shift in Z_{MEG}, we produce a negligible effect on the photon timing, but we make a 1.7 mrad $\theta_{e\gamma}$ error at the center of the calorimeter. To be

clear, this is more relevant for $\phi_{e\gamma}$ (and therefore Y_{MEG} shifts of the LXe) as the MEG II $\phi_{e\gamma}$ resolution is better than its $\theta_{e\gamma}$ resolution. A similar effect is true for a rotation about X_{MEG} i.e. a mixing of Y_{MEG}/Z_{MEG}.

In summary, a precision in the relative position and orientation of the CDCH/LXe of $< 500\mu m$ would result in < 1 mrad biases in $\theta_{e\gamma}/\phi_{e\gamma}$.

6.10.4 CDCH LXe Cosmic Alignment

In the 2021 dataset, cosmic rays were used to measure the relative Z coordinate of the CDCH/LXe. No rotations have been estimated; as discussed above this is also relevant.

This is described in these two slides:Antoine Slides 1, Atsushi Slides 2. Details are also described in Shinji Ogawa's thesis page 91: Shinji Thesis

The cosmic ray alignment contained several issues. First, the cosmic rays must enter both the CDCH and the LXe and therefore only contains a very particular section of both detectors. Only cosmic ray tracks with $v_\gamma < 0$ were used in this analysis due to analysis complications. Therefore the analysis is not very sensitive to rotations.

The data is taken without a magnetic field and therefore the cosmic ray tracks in the CDCH have a small amount of hits, and therefore the quality is significantly degraded with respect to standard CDCH tracks. The track fitting/finding procedure has been investigated and improved, but not yet achieving resolutions of the standard analysis.

In addition, the LXe position reconstruction of cosmic rays relies on an assumption that the photon originated from the target. This is clearly not the case for the cosmic ray. Corrections are applied in order to attempt to suppress this bias. Different cosmic rays may create different shower lengths, etc., and thus the correction will likely vary with the energy,

particle, etc., which is not known.

In the 2021 cosmic ray alignment, a Z_{MEG} shift of 1.0 mm with respect to the previous alignment (independent LXe, CDCH alignments described above) was found. A single Gaussian fit resulted in an uncertainty of $\sim$ 0.8 mm, but of course this does not include systematic errors that could effect the relative alignment. Therefore, this alignment technique is consistent with the previous alignment from the optical survey and internal CDCH and LXe detectors. This is shown in Figure 6.7.

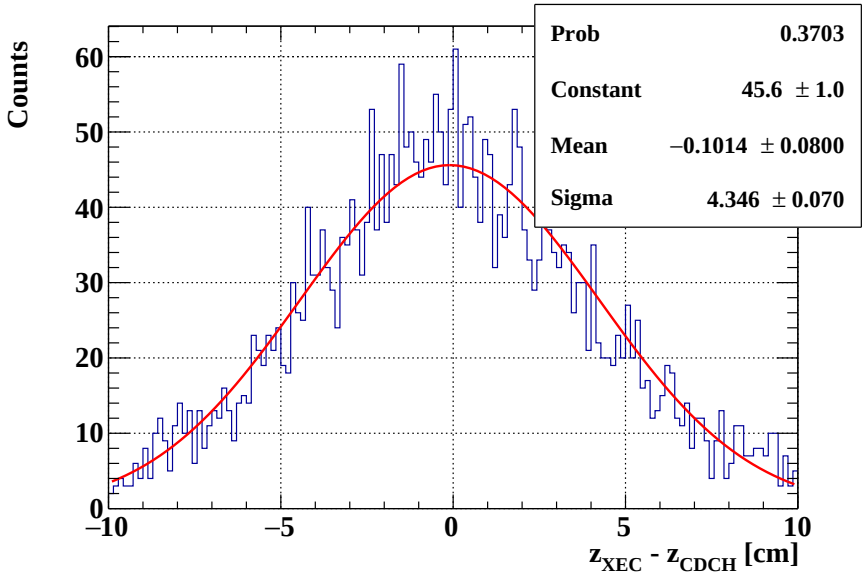


Figure 6.7: The relative CDCH-LXe alignment from cosmic rays (Figure made by Antoine Venturini)

Chapter 7

MEG II Calibrations

In this chapter we summarize the calibrations performed on the MEG II detectors. This generally consists of channel-by-channel time offsets and gain offsets required to achieve the optimal resolutions. The LXe detector also requires additional corrections to the energy scale and the reconstructed timing that are also discussed here.

7.1 CDCH

7.1.1 Wire-by-Wire Time Offsets

For the CDCH, the most critical calibration is verifying that there are no time offsets wire-by-wire; i.e. a hit arrival time on wire A and wire B with the same T_0 correspond to the same drift time. These unwanted time offsets create systematics in the distance of closest approach resolution and therefore degrade the kinematic resolution. Time offsets wire-by-wire are applied to remove this effect. The time offsets could be due to differences in cable lengths, delays between the DRS4 chips, etc.

This time offset wire-by-wire is calculated using a histogram of the drift time (hit arrival time - T_0 at the drift cell). An example is shown in Figure 7.1. As expected, the distribution starts to rise around a drift time of $\sim$ 0 ns. By fitting for the position of the rising edge for each wire, we calculate the relative offset required wire-by-wire. Note, negative drift times can be the result of instances when a high voltage noise peak is detected as the first ionization cluster. There are also rare instances where pileup is nearly on-top of the given hit; this could also result in a negative drift time.

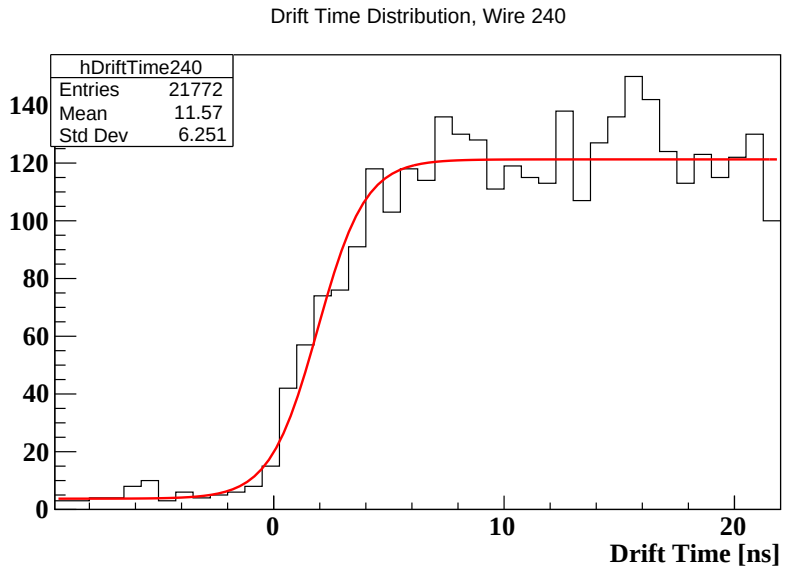


Figure 7.1: Drift time for hits on wire 240.

This also requires an additional global offset to align the timing on the CDCH wires with the time found in the SPX electronics. Or in other words, the rise of the drift time distribution is calculated wire-by-wire, but it's unclear where to set the T_0 (e.g. at the 20%, 80% of the rise of the drift time distribution). Since the relative wire time offsets have been calculated, this CDCH-SPX offset is a single timing offset for the entire CDCH. This is calculated by verifying that <(hit DOCA - track DOCA)> is near zero for hits with a track DOCA $\sim$ 2 mm (here we expect minimal bias due to ionization statistics). Effectively, this global CDCH-SPX time offset is calculated by the ML TXY. This bias is shown in Figure 5.17.

7.1.2 Wire-by-Wire Longitudinal Position Calibrations

In addition, the CDCH requires a calibration of the hit coordinates along the wire axis (local z). For z from charge division and z from time difference, we require the relative gain and time offsets end-to-end for each wire. Note, by incorporating the time offset end-to-end as $End_1 \rightarrow End_1 + t/2, End_2 \rightarrow End_2 - t/2$, the wire-by-wire time offset mentioned in the previous subsection is not affected. These calibrations are easily calculated by comparing the local z from the positron track (unbiased by the hit itself) to the local z measurement of the hit from charge division or time difference. A mean error over all hits on a wire implies a gain or time offset between the two ends. An example is shown in Figure 7.2, clearly one wire has a gain offset corresponding to a mean z error of ~ 25 cm. We could simply offset all z measurements, but instead we apply corrections to the relative time and gain of the two wire ends. A global signal propagation speed and wire resistivity are also required to convert the relative charge and time differences into z measurements; these are calculated by simply setting the mean z error to zero. It was also found that hit z measurements deviate non-linearly from the z from the positron track at large $|z|$ ($|z| \sim> 70$ cm); a polynomial fit was used to correct for this. This is a minor effect and has not been explained physically.

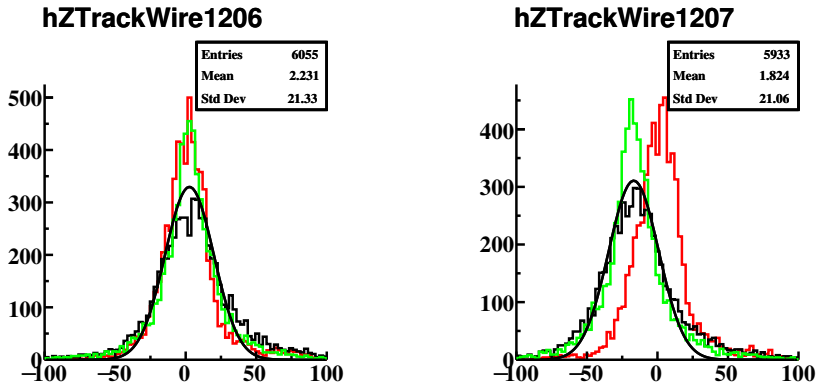


Figure 7.2: Measurement of z from charge division, z from the positron track, and the difference between the two are shown for two example wires in green, red, and black respectively in cm.

7.1.3 Wire Quality Check

We note that in order for a hit to be associated with an eventual fitted track, several wire-specific criteria must be met, this includes proper handling of noise and signals (or lack of signals) on the two ends, wire-by-wire time offsets, end-to-end time and gain offsets, wire alignment, TXY that is dependent on the cell size, etc. An issue with any of these criteria results in a lower rate of hits on quality fitted tracks. Therefore, while optimizing wire alignment, calibrations, etc. the number of hits on quality fitted tracks wire-by-wire is very informative to the status of each wire. This is shown in Figure 7.3. Some wires are shown to be completely dead (no signals), others have significant noise on a DRS4 chip, affecting 8 sequential wires and thus a lower hit rate. Here, the two sets of sequential wires (16 and 8 sequential wires in layers 3 and 8 respectively) have their high voltage disconnected and thus have no hits. In addition, there are O(20) wires, mainly in layers 8 and 9 with shorts. The general structure of the wire-by-wire hit rate (peaked near the central sectors) is similar to that of the MC and is the result of the trigger requirements. Note that the wire distribution suggests that there would be hits on wires not currently readout (e.g. plane 9 below wire number zero). We have verified that for signal positrons in the Monte Carlo this is a smaller effect, but nonetheless this could improve efficiency. For this reason, ~ 30 additional wires in the inner layers (layer 7,8,9) have been added to increase the ϕ wire coverage and thus signal efficiency.

7.2 SPX

In the SPX detector, the main calibration is calculating and removing the tile-by-tile time offsets. This is the relative timing of signals on the set of SiPMs on each individual tile. This is calculated by comparing the time estimated from a single tile to the average time measured from the cluster of intersected SPX tiles in a single positron trajectory. Ideally, the

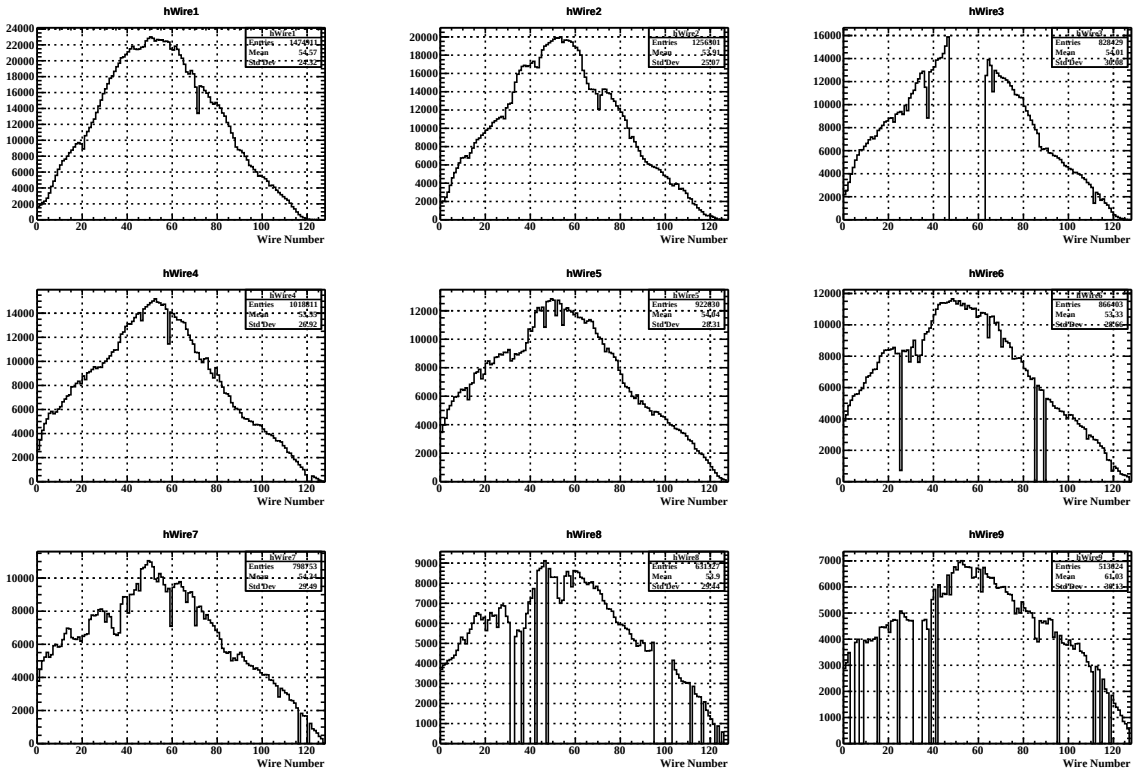


Figure 7.3: Number of hits on fitted quality tracks per wire.

SPX cluster time excludes information from the specific tile. Summing over a large amount of tiles, the relative timing offsets can be measured very precisely, contributing a negligible contribution to the timing resolution.

Note that this calibration is easy for nearby tiles, but a relative timing calibration for all tiles in the upstream/downstream modules (256 tiles each) also requires mixing of information on the edges of the timing counter modules. This requires higher order mixing as they will not be hit by the same positron track. A laser-based calibration is also used to aid in this process. In short, a fiber optic is connected to the individual timing counter tiles; by firing a laser into the fiber optic of many timing counter tiles, the signal time and thus the relative time offsets between the tiles can be estimated.

7.3 LXe

The LXe detector is the most difficult detector to calibrate and requires a large number of calibrations. These are summarized briefly in this section. In general, the experiment relies on a few main calibration techniques, most of which are used to monitor the energy scale, resolution, and the timing in the LXe.

The LXe detector requires calibrations to achieve the optimal performance. The main E_γ, t_γ calibration technique uses an alternate beam of π^- with an alternate liquid hydrogen target; some interact through the $\pi^- p \to \pi^0 n$; $\pi^0 \to \gamma\gamma$ channel. This calibration technique was also used in MEG I and is known as the charge exchange (CEX). The kinetic energy of the π^0 ($\beta \sim 0.2$) is the result of the π^0/π^- energy difference. By identifying back-to-back γ pairs, a set of quasi monochromatic γ enter the LXe detector: $E_\gamma = 0.5 \cdot m_{\pi^0}\gamma(1 \pm \beta cos(\theta_{rest}))$; $\theta_{rest} \sim 0$, $E_\gamma = 55/83$ MeV. Typically, the lower energy photon ($E_\gamma = 55$ MeV) is selected in the LXe detector to be closer to the signal γ energy (52.83 MeV).

The other photon enters a separate detector that consists of a 4 $\times$ 4 array of bismuth germanium oxide (BGO) crystals and a pre-shower counter for the photon timing (t_{ps}). By moving the BGO detector to be back-to-back with different LXe sections, we calibrate t_{LXe} and E_{LXe} as a function of position in the LXe detector.

The CEX calibration is performed infrequently and therefore cannot calibrate the time dependence of the detector's energy scale. This is estimated using a Cockcroft–Walton (CW) accelerator. The CW protons intersect an alternate $Li_2B_4O_7$ target. The process results in photons with 17.6 MeV. In contrast with the CEX, this is performed on a weekly basis, but is farther away from the signal energy. In addition, LED lights are used to illuminate the MPPCs to calculate the gain of each MPPC. Finally, an α source is used to estimate the quantum efficiency of the MPPCs.

7.3.1 Channel Time Offsets

The relative timing of all MPPCs and PMTs instrumented in the LXe detector is critical to achieve a precise LXe timing measurement.

This can be estimated using effectively any data by comparing the average photon time in the calorimeter (or the average signal time) to a given channel. This calibration involves estimating time walk effects and time offsets channel-by-channel, this was done using the CEX 55 MeV data[39].

7.3.2 Global Time Corrections

After the channel-by-channel time corrections were taken into account, the experiment observed offsets in the mean value of the photon time at the LXe when compared to that of the CEX pre-shower timing counter as a function of position in the calorimeter (i.e. $t_{LXe} - t_{PS}$ vs. $u_\gamma, v_\gamma, w_\gamma$). Of course, ideally this would be flat as a function of all measurables if the MPPC/PMT timing offsets were correct and the timing reconstruction algorithm was perfect. The need for these corrects suggests an underlying issue with either the timing offsets or the reconstruction algorithm; for the 2021 analysis, global position-dependent timing offsets were applied. Examples are shown in Figures 7.4 and 7.5.

7.3.3 Channel Gain Offsets

The gain offsets are measured as a function of time using multiple sets of data. In the calorimeter, the number of photons at an MPPC is equal to the number of incoming photons times the quantum efficiency (qe) times the gain (g).

The gain of the MPPCs is estimated using the LED data. Here, the LED intensity and

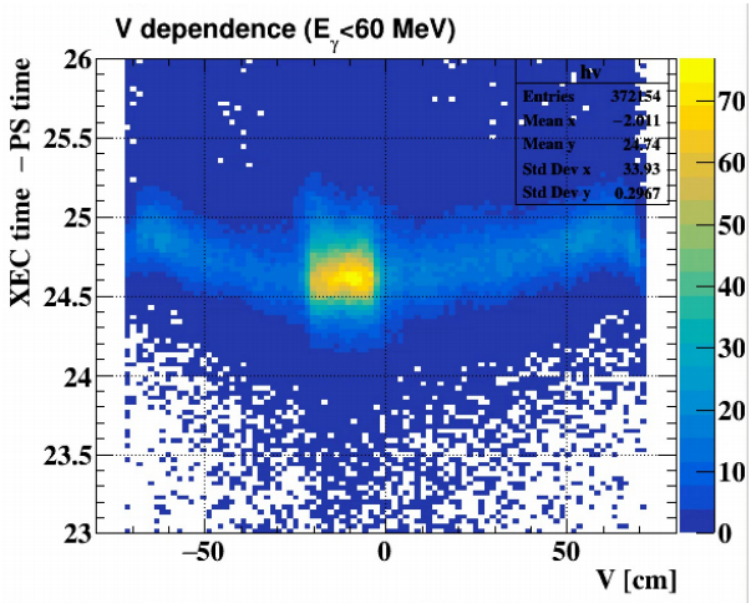


Figure 7.4: The difference in the photon time in the LXe and the CEX pre-shower counter as a function of v_{LXe} (Figure made by Atsushi Oya).

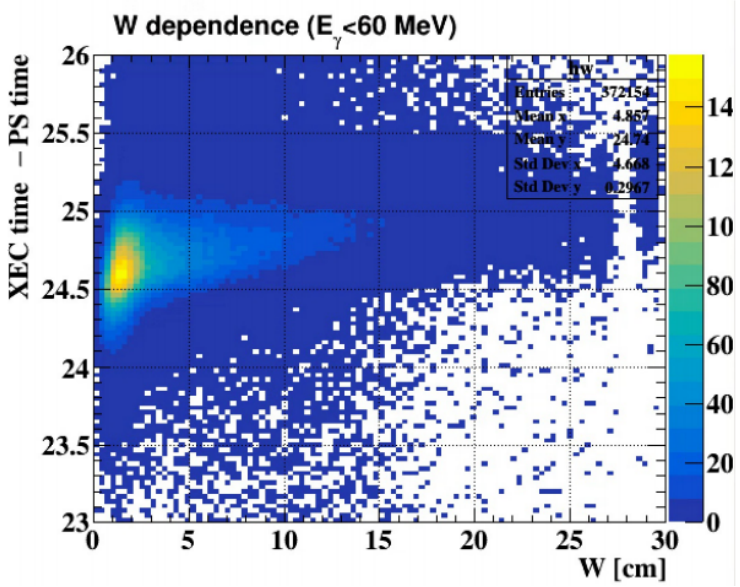


Figure 7.5: The difference in the photon time in the LXe and the CEX pre-shower counter as a function of w_{LXe} (Figure made by Atsushi Oya).

location is set such that the average number of photons reaching each MPPC is roughly 1. Therefore there are instances when 0,1,2, etc. photons (Poisson statistics) reach an MPPC and are amplified, each results in a peak in integrated charge near 0, X, Y, etc. The integrated charge at X directly yields the gain for each MPPC, g. This is done frequently to

monitor the gain throughout the physics run.

The quantum efficiency is measured using the α dataset, this is done weekly. In short, α events are simulated in the Monte Carlo; the charge is reconstructed in each MPPC in this MC α dataset, where the MC quantum efficiency is easily estimated. The relative reconstructed charge in the MC α and the data α datasets are used to correct the MC estimated quantum efficiency. Additionally, there is a single light yield parameter over all MPPCs. These gain techniques are discussed in detail in Satoru Kobayashi's thesis[39].

7.3.4 Relative and Absolute Energy Scale

Similar to the timing in the calorimeter, after calibrating away all channel-by-channel gain offsets, significant correlations between the average energy and the position in the calorimeter were observed. That is, the mean energy of a line (e.g. CW at 17.6 MeV, CEX at 55 MeV, or CEX at 83 MeV) was dependent on the position in the LXe detector.

This can of course be accounted for by simply correcting the average reconstructed energy as a function of position, but it does not address the underlying issue of why this position dependence exists. These corrections effectively achieve the relative energy scale at all positions in the calorimeter. This is likely the result of a miscalibration of the channel gain offsets or an artifact of the reconstructed energy algorithm. For example, the reconstruction could result in a higher/lower reconstructed energy near the edges of the calorimeter or in regions where more energy reaches the PMTs or MPPCs.

Any energy line yields the absolute energy scale. e.g. CEX line is known to have a peak at 55 MeV, thus it yields the relationship between integrated charge and the energy of the photon. This absolute energy scale can be checked by calculating the mode of the reconstructed energy for a different line (e.g. CW at 17.6 MeV or CEX at 83 MeV, etc.) than that used in

the calibration. Using this technique, discrepancies were found as a function of LXe position. That is, calibrating the absolute energy scale in LXe sections using the CEX line at 55 MeV does not result in a reconstructed energy mode at exactly 17.6 MeV in the CW line. Again, the underlying reason for this is not clear; likely this is a consequence of the reconstruction algorithm.

This discrepancy is shown in Figure 7.6; here the energy scale section-by-section is uniform up to $\sim$ 0.4% in the CEX 55 MeV peak, but clearly in other datasets the energy scale has variations of the order $\pm$2%. Further, the CW data has the additional complication of attempting to estimate the time-dependence on the spatial dependence of the energy scale. This is not accessible in the CEX as it is only done once a year. It was not exactly clear how to optimally estimate the energy scale given these discrepancies. The final 2021 calibration involves spatial and time dependent corrections using the CW data, and spatial corrections using the CEX data. This is described in Sataro Kobayashi's thesis, but additional corrections were made that are mentioned in recently slides shown by Kensuke Yamamoto.

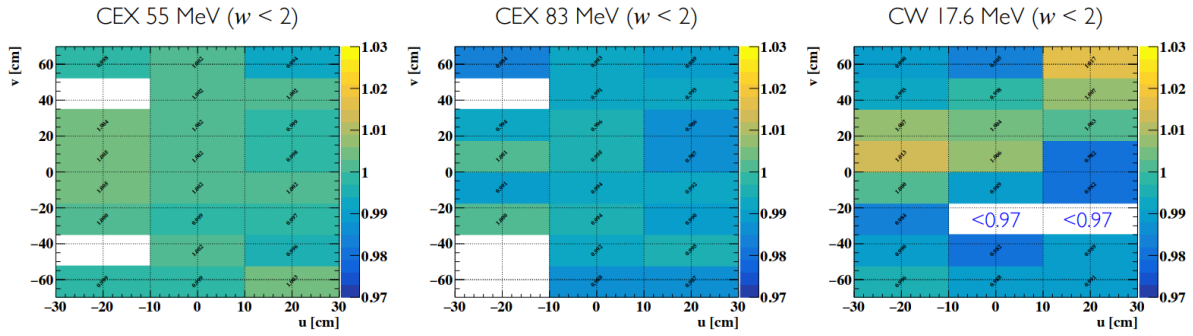


Figure 7.6: The normalized photon energy as a function of position in the calorimeter for different calibration lines (Figure made by Kensuke Yamamoto).

Chapter 8

Physics Analysis

In this chapter, we describe the physics analysis of the 2021 dataset. As a reminder, the MEG II search for $\mu \to e\gamma$ compares the reconstructed positron and photon kinematics $(t_{e\gamma}, \phi_{e\gamma}, \theta_{e\gamma}, E_\gamma, E_e)$ at the target to distinguish between the accidental background, the RMD background, and the signal. Here, $\theta_{e\gamma} = (\pi - \theta_e) - \theta_\gamma, \phi_{e\gamma} = (\pi + \phi_e) - \phi_\gamma$ such that our signal is at $\theta_{e\gamma}, \phi_{e\gamma} = 0$. Additionally, information from the RDC detector (t_{RDC}, E_{RDC}) is used to eliminate RMD accidental events.

First, we describe the positron and photon selection criteria, the blinded kinematic region definition, and the analysis region definition. We then discuss the probability density functions (PDFs) for accidental background (ACC), RMD background (RMD), and signal (SIG) events. We then discuss two approaches to the physics analysis: a cut and count analysis (CCA) and a maximum likelihood analysis (MLA). The CCA yields a higher limit with respect to that of the MLA, but the model has the advantage of being more understandable and clear. The MLA is described in significantly more detail here [11]. After outlining the analysis techniques, we discuss the normalization of the dataset. We then discuss the physics analysis results in the kinematic side band regions (discussed later), toy MC studies, and

the signal region.

The CCA and MLA implement a blind approach to the physics analysis. This has been become the standard for rare decay physics analyses since its implementation in a $K_L \to \mu^{\pm} e^{\mp}$ search at Brookhaven in the 1990s[57][58]. Several examples of how previous expectations biased physics results are discussed here [59].

8.1 Selection Criteria

In this section we list the particle selection criteria in the physics analysis.

8.1.1 Positron Selection

The positron selection was already described in Section 5.6. The selection criteria remove low quality positron tracks without a significant degradation in the signal efficiency.

8.1.2 Photon Selection

The reconstructed photon position in the calorimeter is restricted to the following region:

- $|u_\gamma| < 23.9$ cm
- $|v_\gamma| < 67.9$ cm
- $0 < w_\gamma < 42$ cm

Reconstructed photons are then selected based the pileup detection procedure. Below we list the pileup categories and whether or not they are included in the final analysis.

- No Pileup - Included

- Pileup detected, but properly deconvolved - Included
- Time coincidence pileup - Excluded
- Event reconstruction failed to converge - Excluded

The reconstructed photons then must pass a time reconstruction algorithm χ^2/DOF cut ($\chi^2_t/DOF < 1.8$) and a PMT selection criteria at the waveform analysis level ($\chi^2_{WF,PMT}/DOF < 8$). Finally, the reconstructed photon must pass cosmic ray selection criteria that is implemented to avoid cosmic rays contaminating the dataset. In short, this cosmic ray rejection makes a 2D cut on w_γ and the ratio of the charge collected on the MPPCs and the PMTs.

8.1.3 Pair Selection

All implemented physics analyses use one positron/photon pair per event, but the event reconstruction may result in many reconstructed positrons; we select the "best" reconstructed track.

As a reminder, there may be several reconstructed positron tracks that correspond to the same physical positron track (ghost track). The "best" ghost track is selected using ranking from the machine learning output, the number of turns, and whether or not the track is from the ML hit finding reprocessing. This is explained in detail at the end of Section 5.6. Further, there may be several physical positron tracks that exist in a single event. To be considered, first the positrons must satisfy the following:

- $|t_{e\gamma}| < 3$ ns
- $45 < E_e < 56$ MeV
- $\Theta_{e\gamma} < -0.95$ (solid angle)

After these selection, only $\sim 0.24\%$ events have multiple pairs at this stage[11]. The pair with the largest opening angle is selected (back-to-back with the photon).

8.2 Signal Kinematic Region

The following defines the signal kinematic region:

- $|t_{e\gamma}| < 0.5$ ns
- $52.2 < E_e < 53.5$ MeV
- $48 < E_\gamma < 58$ MeV
- $|\theta_{e\gamma}| < 40$ mrad
- $|\phi_{e\gamma}| < 40$ mrad

The $t_{e\gamma}$ region is defined ($\sigma_{t_{e\gamma}} \sim 100$ ps) such that all signal events should enter the analysis except events where the first positron turn is lost ($t_{e\gamma} \sim 2$ ns). A wider region containing $|t_{e\gamma}| < 1.0$ ns is blinded. As will be discussed, the momentum resolution is well fit to a core/tail resolution of 90/300 keV respectively, therefore the positron momentum analysis window size only removes the end of tail resolution. The photon energy window size is asymmetric towards low energy to maintain signal efficiency when several MeV is lost in material or significant scintillation light escapes through the inner-face. Finally the angular analysis window size is narrow in order to suppress accidental background (accidentals are roughly flat as a function of $\phi_{e\gamma}, \theta_{e\gamma}$); this choice also improves the positron energy resolution; the positron energy resolution tails are significantly narrower given $|\theta_{e\gamma}|, |\phi_{e\gamma}| < 40$ mrad. A few percent of signal efficiency is lost due to the narrow angular region choice.

8.3 Additional Physics Analysis Tools

In this section we briefly describe tools used to aid in the physics analysis. These are mainly used to study background events expected in the blinded region.

8.3.1 Time Sideband Analysis

Analysis in "sideband" kinematic regions is an extremely useful tool for physics analyses. In the MEG II physics analysis, the time side bands (all kinematics equal to the blinded region except $t_{e\gamma}$ is offset from 0) eliminates the possibility for events to be from true RMD events or signal and therefore can be used to study the accidental time-coincidences, which is our main source of background. We use these sidebands to estimate the number of accidentals in the signal region and build the probability density functions of accidental events (i.e. $E_\gamma, p_e, \theta_{e\gamma}, \phi_{e\gamma}, t_{e\gamma}$). This is discussed in the next several sections. As an example, the photon energy in the time sideband is shown in Figure 8.1. The photon energy for accidentals in the time sidebands should be equivalent to that in the signal region and therefore the photon energy in the time sideband is used to model the photon energy distribution for accidentals in the signal region.

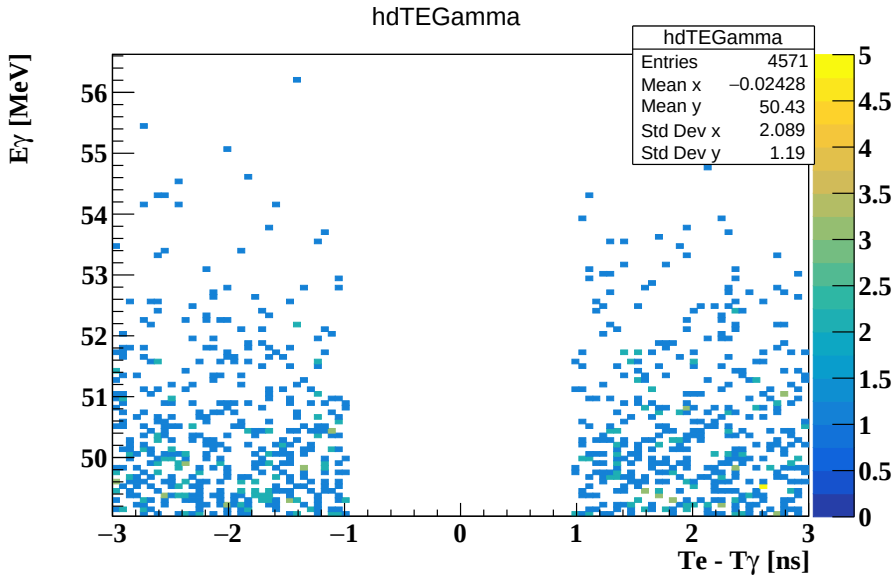


Figure 8.1: The distribution of events in the time sidebands.

Note, the signal signature is at the maximum allowed energy for an RMD photon or a Michel positron (52.83 MeV). For this reason, positrons and photons reconstructed above the signal energy in the time sideband are carefully inspected for systematic effects. Ideally

we identify features of these events, and eliminate them as they would otherwise contaminate the signal region. We found that photons in the sideband region had a significantly higher position χ^2/DOF in the calorimeter than that of a typical event ($\chi^2/DOF > 15$ with a typical $\chi^2/DOF \sim 1.5$). In addition, the events with a high χ^2/DOF were focused in a few regions of the calorimeter, this is shown in Figure 8.2. These regions corresponded to three miscablings (< 10 channels) that were fixed in the final analysis. Note, events with a large position χ^2/DOF due to pileup discrimination code failures still persist in the time sideband after the final selection. The photon energy PDF was still properly estimated using the time sideband, but ideally this will be resolved for future MEG II physics analyses.

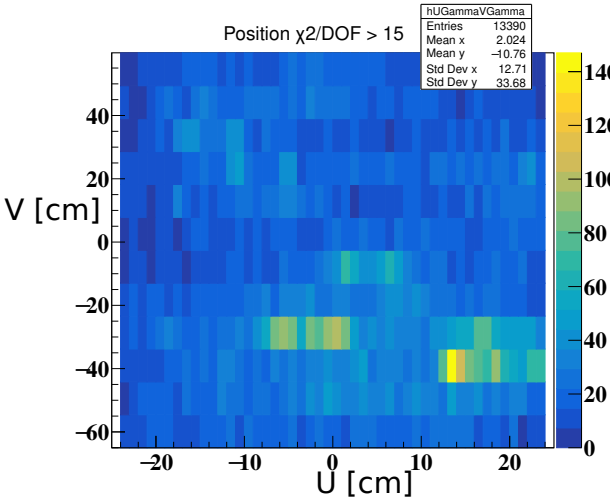


Figure 8.2: The distribution of events in u_γ, v_γ with a large position χ^2/DOF in the photon reconstruction.

8.3.2 Energy Sideband Analysis

The energy side bands (all kinematics equal to the blinded region except 45 MeV$< E_\gamma <$ 48 MeV) contain significantly more RMD events than the signal region and thus is used to estimate the number of RMD events in the signal region. Using this technique and the known kinematics of RMD events (more RMD events with lower photon energy), it is estimated that $N_{RMD} = 1.2 \pm 0.2$ in the signal region. Therefore, we expect a negligible amount of

N_{RMD} near the signal (in particular in the cut and count analysis where the analysis region is significantly narrower). As will be mentioned in the next section, we also use this sideband to estimate the timing resolution using the width of the RMD $t_{e\gamma}$ peak.

8.3.3 Toy MC Simulations

The probability density functions described below are used to generate toy MC simulations containing signal and accidental events in the blinded signal region. The first set of simulations contains randomly generated accidental events in the signal region based on the distribution of accidental events observed in the sidebands. We generate 1000 of these toy MC simulations. We apply the physics analysis software to the simulations to estimate the expected distribution in the upper-limit on the number of signal events in the absence of signal. In addition, using signal PDFs we generate signal toy MC simulations. This is used estimate signal efficiency i.e. the percentage of signal events that are inside the signal region.

8.4 Timing PDF

In this section we discuss the $t_{e\gamma}$ probability density functions (PDFs) for signal, RMD, and accidental events. As a reminder, the key upgrade in the timing resolution with respect to MEG I is the higher hit multiplicity of the pixelated timing counter (SPX). The resolution dependence on the number of hits in the timing counter ($\sigma_{t_e}/\sqrt{N_{SPX}}$) is included event-by-event. The timing resolution is best estimated using the $t_{e\gamma}$ peak due to RMD positron/photon pairs reconstructed in the magnetic spectrometer and the calorimeter respectively.

8.4.1 RMD Timing

The RMD timing PDF is estimated by fitting $t_{e\gamma}$ peak due to the true RMD e, γ pairs. This is done using data in the energy sideband to maximize the number of RMD events. We assume that the RMD timing has a resolution term from the positron and the photon timing where the positron timing takes the form, $\sigma_{t_e}/\sqrt{N_{SPX}}$, and the photon timing is assumed to be a single gaussian: σ_{t_γ}.

The positron timing, σ_{t_e}, is estimated using an even/odd analysis; that is, comparing the independently measured positron time from the even ordered tiles to the odd ordered tiles. This even/odd analysis results in a positron timing resolution of $\sigma_{t_e} = \frac{112}{\sqrt{N_{SPX}}}$ ps. Therefore, at the mean number of hits in the timing counter for signal, the positron timing resolution is 37.3 ps. In Figure 8.3, the width of the RMD peak as a function of the number of hits in the timing counter cluster is shown. Fitting this distribution to the mentioned RMD $t_{e\gamma}$ form, we measure $\sigma_{t_\gamma} = 71 \pm 3$ ps where σ_{t_e} is fixed from the even/odd analysis.

This timing resolution is verified by the CEX analysis; the CEX was described in Chapter 7. Fitting the CEX relative timing distribution $(t_{LXe} - t_{PS})$ results in an LXe timing core resolution of 68 ± 6 ps. Thus, this is consistent with the RMD peak fit result.

8.4.2 Signal Timing

The signal timing PDF is nearly identical to the result from the RMD timing fit. The signal PDF additionally includes a very minor correlation between the relative $t_{e\gamma}$ timing and the positron momentum due to the difference in the propagation length from the pixelated timing counter to the target. This correlation evaluated in the MC has a slope of 18 ± 0.5 ps/MeV. For any well-measured positron track, this effect is negligible ($3\sigma_{E_e,core} \sim 300$ keV; correction of 5 ps).

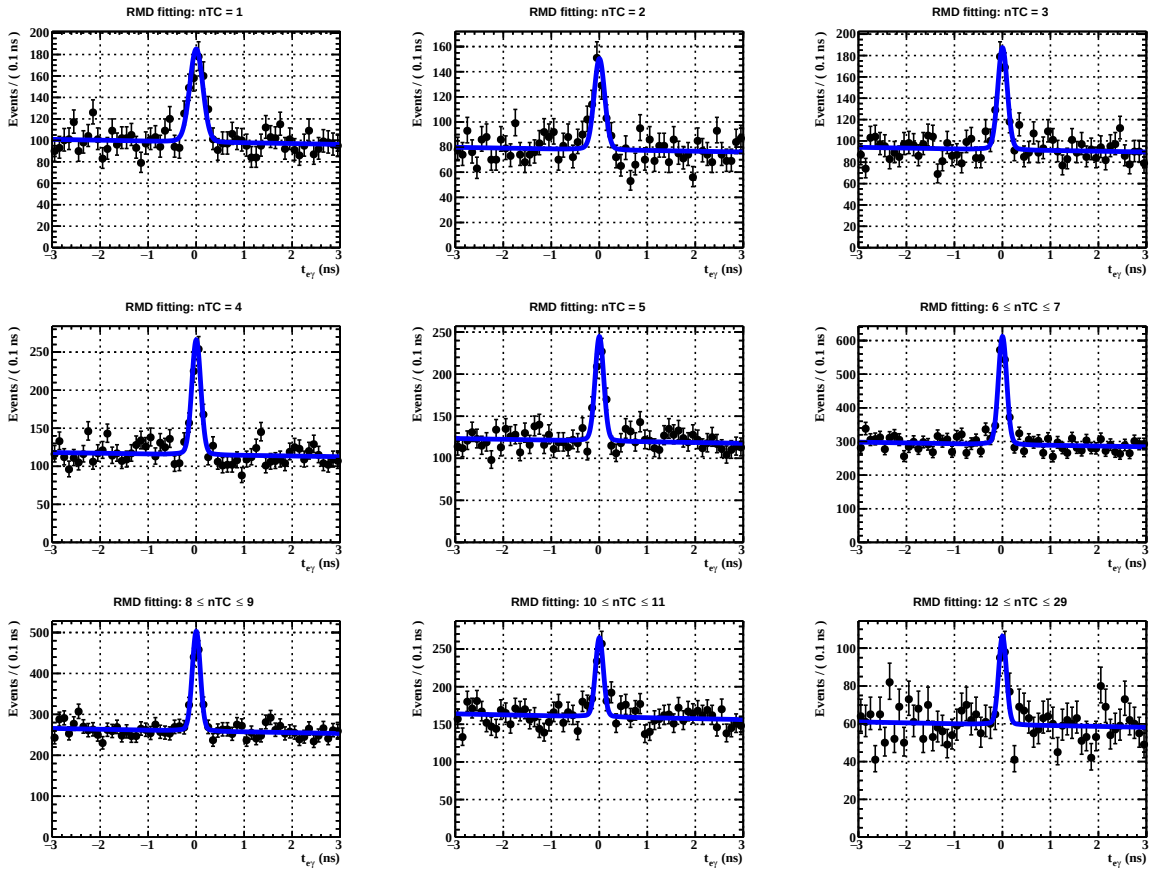


Figure 8.3: The relative positron photon time at the target as a function of the number of hits in the pixelated timing counter[11].

8.4.3 Accidental Timing

The accidental PDF is roughly flat (no value of $t_{e\gamma}$ is preferred for accidental events) and is extracted from the time sideband. The accidental timing PDF contains non-flatness due to trigger effects; i.e. a linear relationship between the number of entries and $t_{e\gamma}$. The number of entries was found to be correlated with the depth in the calorimeter; this dependence was also included.

8.5 Positron Energy PDF

The positron momentum resolution is best estimated using the Michel positron momentum distribution. No physical Michel positron can have a momentum larger than half the muon rest mass (52.83 MeV); this results in a sharp edge in the momentum distribution. In short, a more narrow momentum edge implies better the resolution. This fit also yields the energy scale, which could be offset from the true edge (52.83 MeV) due to an incorrect magnetic field map, misalignment of the CDCH and the magnetic field map, etc.

8.5.1 Accidental Energy

The reconstructed positron energy distribution in the time sideband exactly characterizes the momentum distribution for accidental events. The distribution is fit to the known Michel energy distribution, an acceptance function, and a resolution function. The acceptance function is required as lower momentum positrons have a lower efficiency. The resolution function is assumed to take the form of a sum of three gaussians. The first is a core of ~ 90 keV and the second is a tail of ~ 300 keV, the third is a very wide tail $\sim 1-2$ MeV, which is required to avoid poor fit results due to the fit attempting to accommodate wide tail entries. This is consistent with the double turn results described in Section 5.7.

To take into account the event-by-event information, the data is binned in ~ 7 bins of momentum uncertainty from the Kalman. This procedure has been verified to achieve the correct resolution in MC Michel data samples. An example fit to the momentum distribution is shown in Figure 8.4.

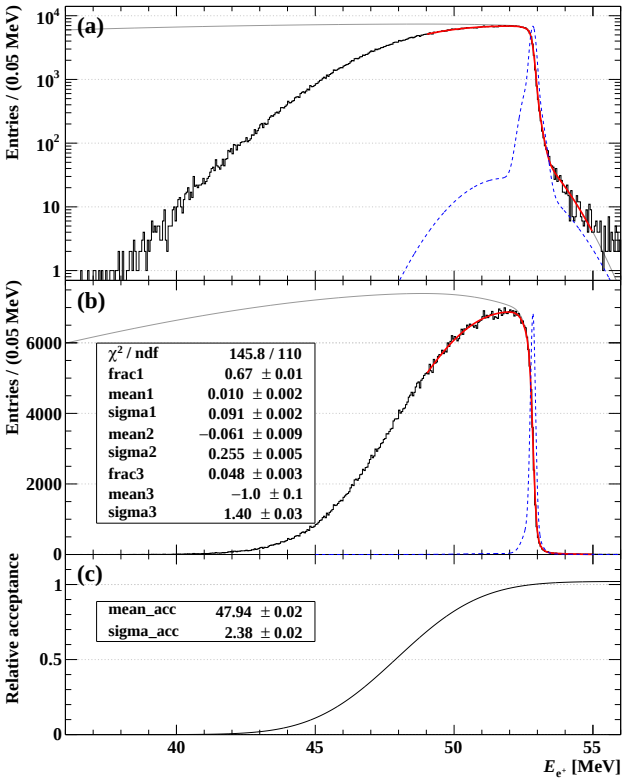


Figure 8.4: The Michel edge on a log and linear scale. The final subplot is the acceptance function.

8.5.2 Signal Energy

The signal momentum PDF is the resolution function described above around the signal energy ($m_\mu/2$). We apply small corrections based on the difference in resolution between the signal MC and Michel MC positron samples. The main difference is due to the fact that signal positron in the signal region must have $|\phi_{e\gamma}|, |\theta_{e\gamma}| < 40$ mrad.

8.5.3 RMD Energy

The positron RMD energy PDF is estimated using the known kinematic distributions for RMD positrons smeared with the resolution and acceptance from the Michel edge fit.

8.6 Angular PDF

The positron angular kinematics are estimated using ϕ_e, θ_e estimates from the Kalman filter. The photon angular kinematics, $\phi_\gamma, \theta_\gamma$, are built by propagating the photon from its reconstructed position in calorimeter $(u_\gamma, v_\gamma, w_\gamma)$ back to the positron vertex at the target (y_e, z_e). The angles are then be compared.

The positron vertex and angular kinematic resolutions are estimated from the double turn analysis. The photon position resolution is estimated using CW data described below.

8.6.1 Accidental Angular

The accidental angular PDF is extracted again using the time sideband data. The rate of accidentals is roughly uniform about $\theta_{e\gamma}, \phi_{e\gamma}$. We observe that the number of accidentals is dependent on $\theta_{e\gamma}, \phi_{e\gamma}$ and $\phi_{e\gamma}$ as a function of v_γ; this was taken into account in the PDF. We assume that these dependencies are due to trigger biases.

8.6.2 Signal Angular

The signal angular PDF requires the kinematic resolutions and the correlations between the relevant kinematic variables. On the positron side, this was mentioned in detail in Section 5.7. In short, the double turn analysis results in a measurement of the resolution for $z_e, y_e, \phi_e, \theta_e$; these are corrected using the double turn technique in Michel MC data, and measurements of the true MC resolutions (comparison with the MC) for Michel and signal positrons in the MC. The event-by-event Kalman covariance matrix is incorporated by estimating the pull in the kinematic comparison $((X_2 - X_1)/\sqrt{\sigma 1^2 + \sigma_2^2}$, $(X_{REC} - X_{MC})/\sigma_{REC}$, instead of the resolutions. Most of the correlations between the positron kinematic variables are extracted

from the double turn analysis, but some (e.g. $\delta z_e : \delta\theta_e$, rely on the MC.

On the photon side, the position resolution of the photon at the calorimeter is estimated using CW data with a collimator at inner face of the calorimeter; this is described in detail in Satoru Kobayashi's thesis[39]. A careful treatment of how to convert these correlations into the eventual $\theta_{e\gamma}, \phi_{e\gamma}$ PDFs is described here[11].

8.6.3 RMD Angular

Similar to the RMD positron energy PDF, the RMD angular PDFs are formed using the expected kinematic distributions for RMD events smeared based on the expected resolutions and acceptance.

8.7 Photon Energy PDF

The photon energy resolution and scale is evaluated using a variety of calibrations and tools that were described in detail in Chapter 7.

8.7.1 Signal Photon Energy

The photon energy resolution for signal is estimated from the CEX 55 MeV peak as a function of depth; this is shown in Figure 8.5. This distribution is fit to an "expgauss", i.e. the left side is an exponential and the right side is a gaussian (described in more detail in [11]). For the signal, the LXe energy resolution is estimated in sections of the calorimeter as shown in Figure 8.6.

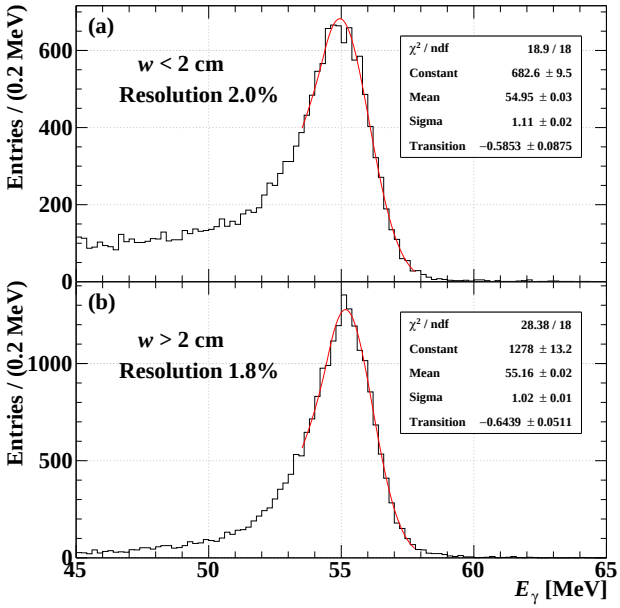


Figure 8.5: The LXe measured 55 MeV CEX peak at varying depth (w).

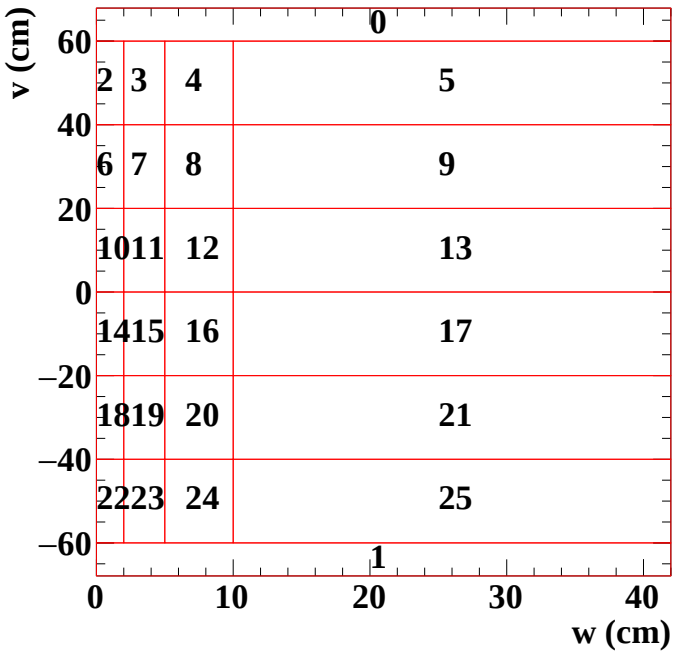


Figure 8.6: Sections of the calorimeter in the signal photon energy PDF.

8.7.2 RMD Photon Energy

The RMD photon energy distribution is calculated using a similar method to other RMD PDFs. The expected RMD photon energy distribution is smeared with the estimated resolution and acceptance, this is done section-by-section through the calorimeter.

8.7.3 Accidental Photon Energy

The accidental photon energy distribution is obtained from the photon energy distribution in the time sideband. For the accidental PDF, the LXe energy is separated into sections as shown in Figure 8.8. The photon energy distribution is fit to the Monte Carlo expected photon distribution (RMD and AIF photons) smeared by a resolution function, adjusted by the trigger, and offset by a energy scale term. There is also a cosmic ray component, estimated using cosmic ray data without beam. An example fit is shown in Figure 8.7.

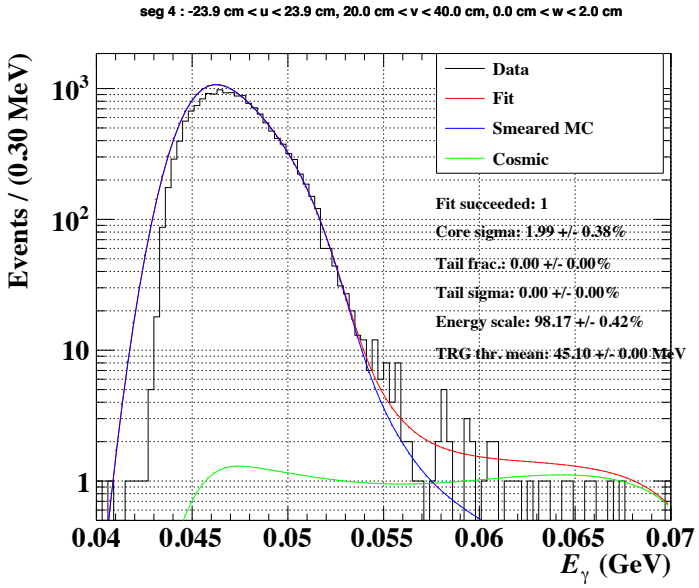


Figure 8.7: The background photon energy PDF in a section of the LXe.

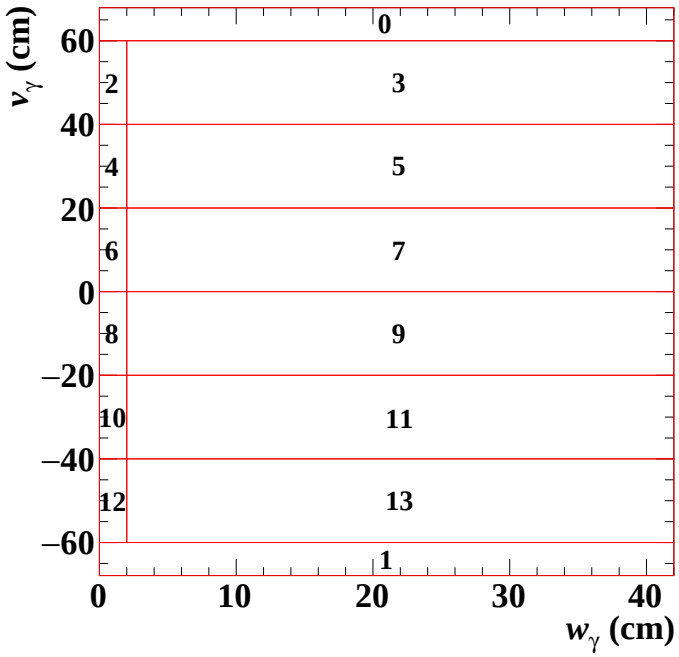


Figure 8.8: Sections of the calorimeter in the background photon energy PDF.

8.8 RDC PDF

In this section we briefly discuss how the RDC PDFs are built. For more details, see [11].

8.8.1 RDC Signal and RMD

For the signal, the RDC PDF is effectively a uniform distribution in $t_{LXe-RDC}$ i.e. there is some chance of intersecting the RDC at any time while there is a signal event in the LXe-CDCH-SPX detectors. This is identical in the RMD PDF case. To be clear, the RMD PDF contains events where an RMD photon intersects the LXe and a RMD positron intersects the CDCH (not the RDC). The expected energy (E_{RDC}) and time ($t_{RDC-LXe}$) distributions can be estimated using random triggers.

8.8.2 RDC Accidental

The timing, $t_{RDC-LXe}$, is estimated simply from the observed peak in MEG triggers. The energy, E_{RDC}, is correlated with the photon energy E_γ as we expect lower energy positrons with higher energy photons.

8.9 PDF Summary

The average probability density functions for signal and accidentals are shown in Figure 8.9.

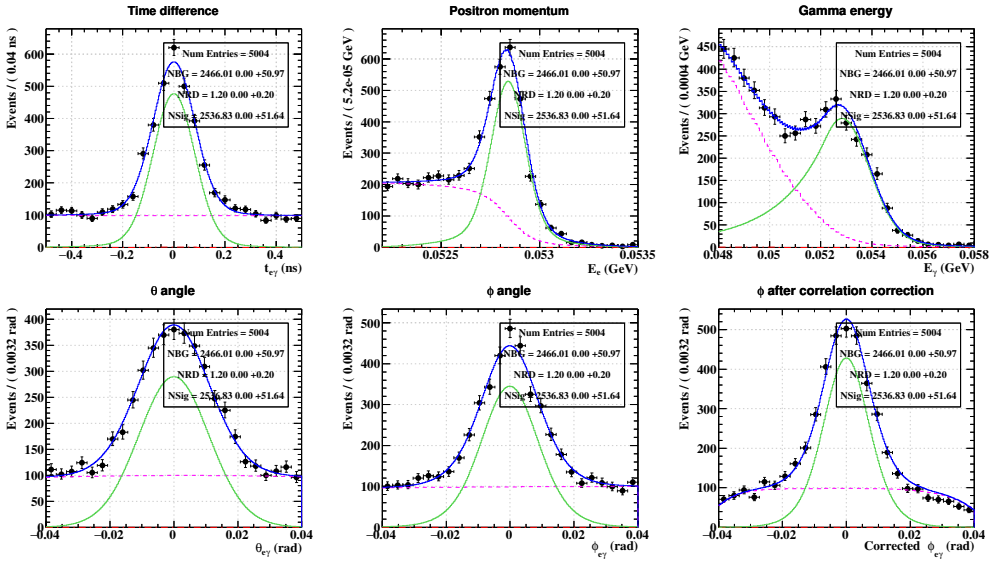


Figure 8.9: The average probability density functions for signal and accidentals.

8.10 Normalization

The Cut and Count Analysis (CCA) results in the number of events in the signal region and in the absence of a discovery, an upper-limit on the number of signal events. The Maximum Likelihood Analysis (MLA) results in the best fit number of signal events and an upper-limit

on the number of signal events. To convert the upper limit on the number of signal events into an upper limit on the branching fraction of $\mu^+ \to e^+\gamma$, we require a normalization.

We can write the branching fraction of $\mu^+ \to e^+\gamma$ as $\mathcal{B}_{e\gamma} = \frac{N_{e\gamma}}{N_\mu}$; we can then set $N_{e\gamma} = 1$, this is known as the single event sensitivity (SES), which is defined as the signal branching fraction such that we expect to observe one signal event in the full dataset. The normalization N_μ is equal to 1/SES. In MEG II, the normalization is calculated using the same two techniques used in MEG I; in short, the normalization estimates rely on the measured number of Michel positrons intersecting the timing counter and the measured number of e^+, γ RMD pairs. Both techniques require several correction factors estimated from a variety of techniques to convert these measurements into normalization estimates.

8.10.1 Michel Normalization

In this subsection, we describe the normalization measurement using the number of Michel decays in an independent trigger, triggering on single pixelated timing counter hits (no calorimeter information). Using these events, we measure the number of positrons with a momentum greater than 50 MeV. This number is then scaled by the trigger prescale factor over the full dataset. This normalization is estimated using two terms.

The first, $N_{e\nu\bar{\nu}}$ is the number of detected positrons in the single pixelated timing counter trigger that pass our selection criteria (114739). This is equal to the number of stopped muons, $N^{\mu}_{e\nu\bar{\nu}}$ times the branching fraction for Michel positrons ($\mathcal{B}_{e\nu\bar{\nu}}$) times the fraction of Michel positrons with energy above 50 MeV ($f_{e\nu\bar{\nu},*}$) divided by the prescale factor ($P_{e\nu\bar{\nu}}$) times the trigger efficiency ($\epsilon^{e\nu\bar{\nu}}_{TRG}$) times the positron efficiency ($\epsilon^{e\nu\bar{\nu}}_{e}$).

$$N_{e\nu\bar{\nu}} = \frac{N^{\mu}_{e\nu\bar{\nu}}}{P_{e\nu\bar{\nu}}} \cdot \mathcal{B}_{e\nu\bar{\nu}} \cdot f_{e\nu\bar{\nu},*} \cdot T_{e\nu\bar{\nu}} \cdot \epsilon^{e\nu\bar{\nu}}_{TRG} \cdot \epsilon^{e\nu\bar{\nu}}_{e} \tag{8.1}$$

The second, $N_{e\gamma}$ is the number of detected signal events. This is equal to the number of stopped muons, $N^{\mu}_{e\gamma}$ times the branching fraction for signal positrons ($\mathcal{B}_{e\gamma}$) divided by the prescale factor ($P_{e\gamma}$) times the trigger efficiency ($\epsilon^{e\gamma}_{TRG}$) times the positron efficiency ($\epsilon^{e\gamma}_{e}$) times the acceptance for photons in the calorimeter given a positron ($A^{\gamma}_{e\gamma}$) times the photon reconstruction efficiency ($\epsilon^{e\gamma}_{\gamma}$) times the selection criteria given both a positron and a photon are reconstructed ($\epsilon^{e\gamma}_{SEL}$).

$$N_{e\gamma} = \frac{N^{\mu}_{e\gamma}}{P_{e\gamma}} \cdot \mathcal{B}_{e\gamma} \cdot T_{e\gamma} \cdot \epsilon^{e\gamma}_{TRG} \cdot \epsilon^{e\gamma}_{e} \cdot A^{\gamma}_{e\gamma} \cdot \epsilon^{e\gamma}_{\gamma} \cdot \epsilon^{e\gamma}_{SEL} \tag{8.2}$$

We then rearrange the two equations to be equal to $\mathcal{B}_{e\nu\bar{\nu}}$ and $\mathcal{B}_{e\gamma}$ respectively and then divide the two equations: $\mathcal{B}_{e\gamma}/\mathcal{B}_{e\nu\bar{\nu}}$. Note, $T_{e\nu\bar{\nu}}/T_{e\gamma} = 1$ and $N^{\mu}_{e\gamma}/N^{\mu}_{e\nu\bar{\nu}} = 1$ as the triggers are taken over the same time period with the same beam rate. In addition, the signal prescale ($P_{e\gamma}$) is equal to one for a vast majority of the 2021 run; we set it equal to one in the below equation.

$$\frac{\mathcal{B}_{e\gamma}}{\mathcal{B}_{e\nu\bar{\nu}}} = \frac{N_{e\gamma}}{N_{e\nu\bar{\nu}}} \cdot \frac{f_{e\nu\bar{\nu},*}}{P_{e\nu\bar{\nu}}} \cdot \frac{\epsilon^{e\nu\bar{\nu}}_{TRG}}{\epsilon^{e\gamma}_{TRG}} \cdot \frac{\epsilon^{e\nu\bar{\nu}}_{e}}{\epsilon^{e\gamma}_{e}} \cdot \frac{1}{A^{\gamma}_{e\gamma}} \cdot \frac{1}{\epsilon^{e\gamma}_{\gamma}} \cdot \frac{1}{\epsilon^{e\gamma}_{SEL}} \tag{8.3}$$

As mentioned in Chapter 1, $\mathcal{B}_{e\nu\bar{\nu}} \sim 1$, including this change and setting $N_{e\gamma} = 1$ recovers the single event sensitivity (SES):

$$SES = \frac{1}{N_{e\nu\bar{\nu}}} \cdot \frac{f_{e\nu\bar{\nu},*}}{P_{e\nu\bar{\nu}}} \cdot \frac{\epsilon^{e\nu\bar{\nu}}_{TRG}}{\epsilon^{e\gamma}_{TRG}} \cdot \frac{\epsilon^{e\nu\bar{\nu}}_{e}}{\epsilon^{e\gamma}_{e}} \cdot \frac{1}{A^{\gamma}_{e\gamma}} \cdot \frac{1}{\epsilon^{e\gamma}_{\gamma}} \cdot \frac{1}{\epsilon^{e\gamma}_{SEL}} \tag{8.4}$$

or

$$N_{\mu} = N_{e\nu\bar{\nu}} \cdot \frac{P_{e\nu\bar{\nu}}}{f_{e\nu\bar{\nu},*}} \cdot \frac{\epsilon^{e\gamma}_{TRG}}{\epsilon^{e\nu\bar{\nu}}_{TRG}} \cdot \frac{\epsilon^{e\gamma}_{e}}{\epsilon^{e\nu\bar{\nu}}_{e}} \cdot A^{\gamma}_{e\gamma} \cdot \epsilon^{e\gamma}_{\gamma} \cdot \epsilon^{e\gamma}_{SEL} \tag{8.5}$$

Here, $N_{e\nu\bar{\nu}}$, the number of detected positrons that pass our selection criteria (114739), $f_{e\nu\bar{\nu},*}$ is the branching fraction for Michel positrons in this kinematic range (0.101), $P_{e\nu\bar{\nu}}$ is the trigger prescale factor ($2-7\cdot10^6$), the $\frac{\epsilon^{e\gamma}_{TRG}}{\epsilon^{e\nu\bar{\nu}}_{TRG}}$ numerator includes the inefficiencies in the physics trigger such as spatial match failures, energy reconstructed online at too low of energy, etc. (~ 0.78 [13][43]) and the denominator includes trigger inefficiencies due to dead time in the single timing counter trigger, which was negligible. The term $\frac{\epsilon^{e\gamma}_{e}}{\epsilon^{e\nu\bar{\nu}}_{e}} \sim 1.09$ is the relative positron tracking efficiency for signal and Michel positrons (in the selected momentum region), which is calculated using the acceptance function from the Michel edge fit. The term $A^{\gamma}_{e\gamma} \sim 0.97$ is the signal acceptance given the positron is already in the spectrometer (optimized in the design of the experiment), $\epsilon^{e\gamma}_{\gamma}$ is the photon analysis efficiency due to photons intersecting the COBRA magnet and other material prior to reaching the LXe (~ 0.67[39]) and another contribution from pileup (~ 0.93) that is taken from the MC, finally $\epsilon^{e\gamma}_{SEL} \sim 0.93$ is the percentage of signal positron/photon pairs that are detected but outside of the analysis window (e.g. missing first turn and thus the relative positron photon timing is out of the analysis region or cuts due to the narrow $\theta_{e\gamma}, \phi_{e\gamma}$ region).

The uncertainty here is dominated by a few of the efficiency terms that are more difficult to estimate. For example, $\epsilon^{e\gamma}_{SEL}, \epsilon^{e\gamma}_{\gamma}$ is taken from the Monte Carlo, we assign each of them

$\sim 3\%$ uncertainty. Combined with other effects, we assign a total uncertainty of $\sim 4.5\%$. This results in a normalization estimate of $N_\mu = 2.55 \pm 0.13 \cdot 10^{12}$.

8.10.2 RMD Normalization

In this calculation we follow a similar procedure, but instead use the number of RMD decays in the physics trigger dataset to measure the normalization.

$$N_\mu = \frac{N_{e\nu\bar{\nu}\gamma}}{\mathcal{B}_{e\nu\bar{\nu}\gamma,*}} \cdot \frac{\epsilon_{TRG}^{e\gamma}}{\epsilon_{TRG}^{e\nu\bar{\nu}\gamma}} \cdot \frac{\epsilon_e^{e\gamma}}{\epsilon_e^{e\nu\bar{\nu}\gamma}} \cdot \frac{\epsilon_\gamma^{e\gamma}}{\epsilon_\gamma^{e\nu\bar{\nu}\gamma}} \cdot \frac{\epsilon_{SEL}^{e\gamma}}{\epsilon_{SEL}^{e\nu\bar{\nu}\gamma}} \tag{8.6}$$

Below we show the equation used to estimate the number of effective muon stops where $\mathcal{B}_{e\nu\bar{\nu}\gamma,*}$ is the branching fraction of RMD events in the energy sideband region. The relative positron efficiency ($\frac{\epsilon_e^{e\gamma}}{\epsilon_e^{e\nu\bar{\nu}\gamma}}$) factor is due to the differences in momentum and is estimated using the Michel edge fitted acceptance. The relative photon efficiency ($\frac{\epsilon_\gamma^{e\gamma}}{\epsilon_\gamma^{e\nu\bar{\nu}\gamma}}$) is evaluated using the expected RMD photon spectrum smeared by the resolution; this is a minor effect. For the relative selection efficiency ($\frac{\epsilon_{SEL}^{e\gamma}}{\epsilon_{SEL}^{e\nu\bar{\nu}\gamma}}$), both RMD and signal events share most inefficiencies due to selection, but RMD positrons have a significantly wider distribution in $\theta_{e\gamma}, \phi_{e\gamma}$ and therefore are not as effected by the narrow angular analysis region. Finally, the relative trigger efficiency ($\frac{\epsilon_{TRG}^{e\gamma}}{\epsilon_{TRG}^{e\nu\bar{\nu}\gamma}}$) is estimated using the time sidebands, where the difference in the efficiency as a function of $\theta_{e\gamma}, \phi_{e\gamma}$ are the result of the spatial match in the physics trigger that is optimized for $\mu \to e\gamma$ events. This calculation results in a normalization of $3.1 \pm 0.3 \cdot 10^{12}$ where the uncertainty is dominated by the estimate of $\frac{\epsilon_\gamma^{e\gamma}}{\epsilon_\gamma^{e\nu\bar{\nu}\gamma}}$.

8.10.3 Combined Normalization

The two normalization estimates are combined using their weighted uncertainty; the combined normalization is equal to $2.64 \pm 0.1 \cdot 10^{12}$. This is the normalization for the MLA, the CCA normalization is slightly lower due to a lower signal efficiency. The MLA SES is equal to $3.79 \cdot 10^{-13}$, which is the signal branching fraction such that we'd expect to see one signal event in the 2021 dataset.

The CCA results in a number of events in the signal region and the MLA results in a best fit number of signal events, $N^*_{e\gamma}$. In the absence of many events (not enough to claim a discovery), we estimate the upper-limit on the number of signal events. Dividing this by the normalization yields the upper-limit on the branching fraction of $\mu \to e\gamma$.

In the likelihood analysis, it is standard to quote the upper limit on the number of signal events at the 90% confidence level which is defined as $N^*_{e\gamma} + 1.645\sigma$. The $\pm 1.645\sigma$ is selected as this is the half width that 90% of the normal distribution.

8.11 Cut and Count analysis

8.11.1 Overview

Traditionally, physics analysis searches for rare decays were done using a blind cut and count analysis strategy[57]. In this section, we describe a blind cut and count analysis approach to the MEG II 2021 physics data. In this strategy, we define an analysis region around the signal (centered at $t_{e\gamma} = 0, \phi_{e\gamma} = 0, \theta_{e\gamma} = 0$, $E_\gamma = 52.83$ MeV,$E_e = 52.83$ MeV) and then simply count the number of events in the signal region.

8.11.2 Analysis Region

Our signal region is defined as a hypersphere in $t_{e\gamma} \cdot \phi_{e\gamma} \cdot \theta_{e\gamma} \cdot p_e \cdot E_\gamma$ space around the signal signature with a hyperradius, $r_H < 3.45$. The hyperradius cutoff optimization is described below. The hyperradius is defined below:

$$r_H = \sqrt{(\frac{t_{e\gamma}}{\sigma_{t_{e\gamma}}})^2 + (\frac{\phi_{e\gamma}}{\sigma_{\phi_{e\gamma}}})^2 + (\frac{\theta_{e\gamma}}{\sigma_{\theta_{e\gamma}}})^2 + (\frac{(\frac{m_\mu}{2} - \alpha) - E_e}{\sigma_{p_e}})^2 + (\frac{(\frac{m_\mu}{2} - \beta) - E_\gamma}{\sigma_{E_\gamma}})^2} \tag{8.7}$$

The σs represent the estimated resolution event-by-event. The σs and therefore the analysis window size is dependent on the estimated kinematic resolution from the Kalman modified by the pull distribution as mentioned in Section 5.7. We use an average pull of $1.17, 1.95, 1.23, 1.19$ for $y_e, z_e, \phi_e, \theta_e$ respectively (multiplying the covariance diagonal by this quantity). Approximate values for the resolutions are: $\sigma_{y_e} \sim 0.7$ mm,$\sigma_{z_e} \sim 2.0$ mm, $\sigma_{\theta_e} \sim 7.2$ mrad, $\sigma_{\phi_e} \sim \sqrt{5.8^2 + 5.3^2 \cdot tan(\phi_e)}$ mrad, $\sigma_{u_\gamma} = 2.5$ mm, $\sigma_{v_\gamma} = 2.5$ mm, $\sigma_{t_{e\gamma}} = \sqrt{71^2 + 112^2/N_{SPX}}$ ps, $\sigma_{p_e} = 0.1$ MeV,$\sigma_{E_\gamma} = 1.1$ MeV.

None of the γ estimates use event-dependent information, and we use a flat 0.1 MeV resolution for all positrons. The event-dependent positron momentum resolution was not incorporated. The α, β terms represent a shift with respect to the signal signature (0.05, 0.3 MeV for the positron and the photon energy respectively). These shifts are incorporated as the region greater than the signal energy (> 52.83 MeV) contain significantly less background events, but still contains signal efficiency.

We implement the correlation between $d\theta_{e\gamma}/d\phi_{e\gamma}$ due to the correlation between $d\phi_e/dz_e$; where $\frac{d\phi_{e\gamma}}{d\theta_{e\gamma}} \sim -0.4$. This is included by shifting the center of $\phi_{e\gamma}$ based on the value of $\theta_{e\gamma}$.

8.11.3 Expected Number of Background Events

The cut and count analysis requires an estimate on the number of background events. This is critical for optimizing the signal region, and for determining the upper-limit on the number of signal events in the Feldman-Cousins approach[60]. We aim for less than 1 expected background event in the signal region such that not many events are required to claim a signal. First, we discuss our estimates of the number of background events using the toy MC simulations containing only accidentals. The number of expected background events per simulation had a mean of 0.576 (576 events in 1000 toy MC simulations). The distribution in the number of background events per simulation is shown in Figure 8.10.

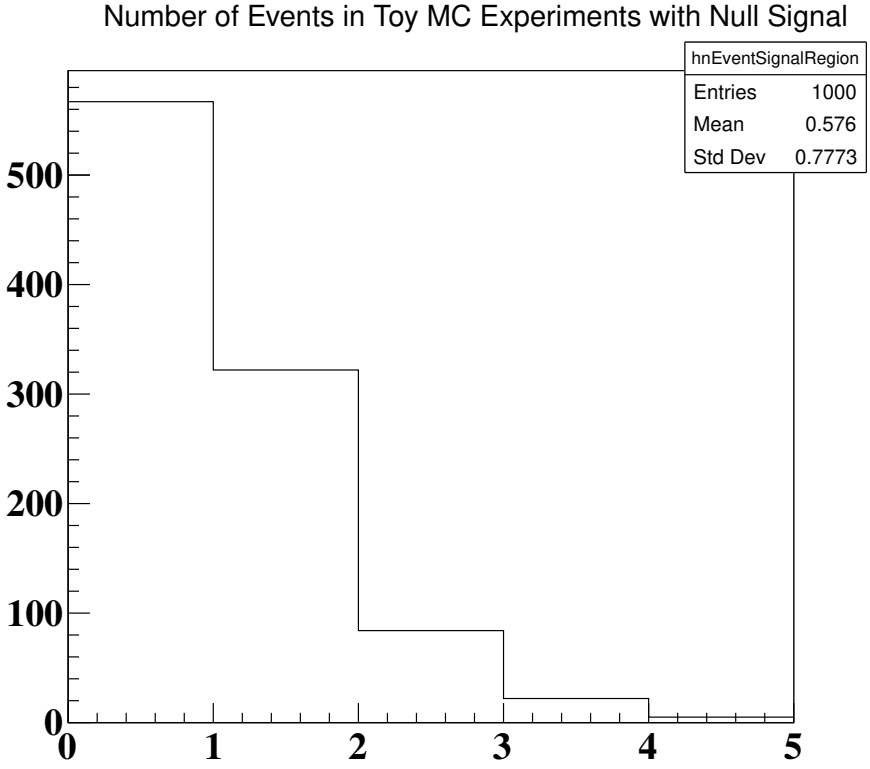


Figure 8.10: The number of background events in the signal region in 1000 toy MC simulations.

In the toy MC study, $\sim 57\%/32\%$ of experiments observe 0,1 events respectively. Observing ~ 5 event in the signal region without there being any true signal be very unlikely ($\sim 0.5\%$

of toy MC experiments).

The expected number of accidental background events in the signal region is also determined using the time sideband regions. By using a large time sideband region ($-7 < t_{e\gamma} < -1$ ns and $1 < t_{e\gamma} < 5$), with a signal region of roughly $|t_{e\gamma}| < 3.45 \cdot 80$ ps, we can estimate N_{ACC} with $\sim$ 20x more statistics than the signal region. Here, we estimate the number of events in a hypersphere using all kinematics except $t_{e\gamma}$; we then divide the number of events observed in the full time sideband by the relative size of the time sideband and a nominal signal region (e.g. $\pm 3.45 \cdot 80$ ps). This significantly overestimates the true number of events in the signal region as this would accept events with $t_{e\gamma} = 3.45\sigma_{t_{e\gamma}}$ and $p_e = 3.45\sigma_{p_e}$ (which would be excluded in the actual analysis hypersphere). For this reason, we use a correction factor from the toy MC simulation to estimate the true expected number of events i.e. the ratio of the number of events in a hypersphere using all kinematics, and the number of events in a hypersphere including all kinematics except $t_{e\gamma}$ and requiring $|t_{e\gamma}| < 3.45 \cdot 80$ ps; this correction factor is 0.524.

We observe 22 events in the full sideband (using a hypersphere with all kinematics except $t_{e\gamma}$) in a time sideband size 18.12x larger than the nominal signal region (10 ns/ $\pm\, 3.45 \cdot 80$ ps). Using our correction factor, we expect $N_{ACC} = 22 \cdot 0.524/18.12 = 0.636 \pm 0.14$ events. This is consistent with the toy MC study results of $N_{ACC} = 0.576$ within 1σ. In Figure 8.11, we show N_{ACC} for a given r_H in the full time sideband.

As an additional check of the sideband data, we analyzed the time sideband with the full hypersphere (including the $t_{e\gamma}$ variable), but offset the mean of $t_{e\gamma}$.

Using an offset of $[-5.0, -4.5, -4.0, -3.5, -3.0, -2.5, -2.0, -1.5, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0]$ results in $[0, 0, 0, 1, 0, 1, 0, 0, 0, 1, 2, 1, 0, 0, 1, 0]$ events respectively or an average of 0.438 accidental events (± 0.17), this is consistent with the previous two results.

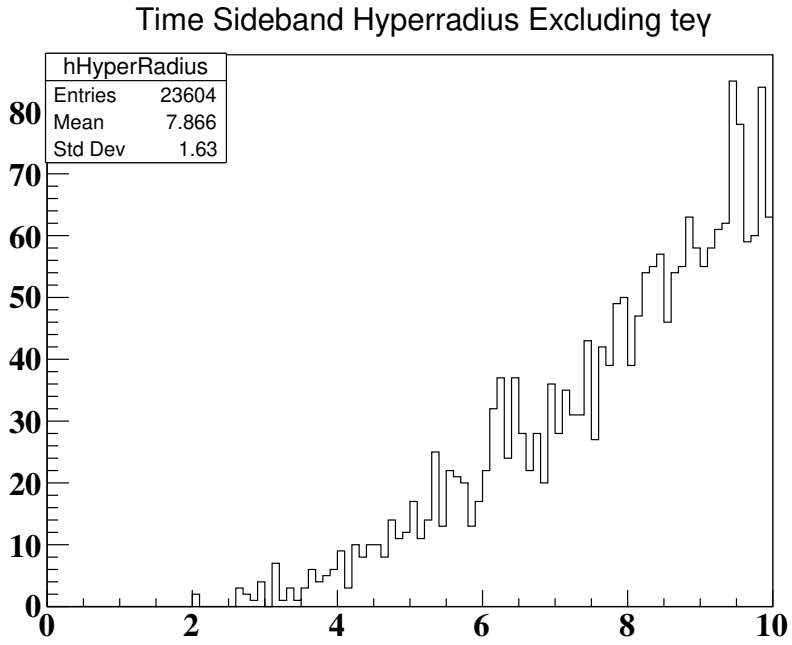


Figure 8.11: The number of events in the time sidebands ($-7 < t_{e\gamma} < -1$ ns and $1 < t_{e\gamma} < 5$) for a given hyperradius.

8.11.4 Relative Normalization

To calculate the normalization of the cut and count analysis, we simply take the normalization of the maximum likelihood analysis mentioned in the previous section and multiply it by the relative efficiency of signal in the cut and count analysis region to the maximum likelihood analysis region: ϵ_{CCA}.

We estimate the relative efficiency using both the full MC simulation and the toy MC simulations both containing only signal. Note that the signal MC E_γ resolution is significantly narrower than that of the true resolution estimated from the CEX data. This is reflected in the toy MC, but not the full MC. Therefore, in the full MC we replace the E_γ resolution with a toy MC estimate.

The relative efficiency is measured to $\epsilon_{CCA} = 0.724/0.777$ in the full MC and toy MC respectively. Both MC simulations have their own shortcomings. The full MC differs from

the data in terms of signal/noise on the waveforms, wire alignment, etc. The toy MC does not use the true particle reconstruction software, but the resolutions contain corrections due to the differences in the resolutions between data and MC. We choose to use the average efficiency or $\epsilon_{CCA} = (0.724 + 0.777)/2 = 0.751$.

8.11.5 Single Event Sensitivity

In this cut and count analysis, $N_{\mu,CCA} = \epsilon_{CCA} \cdot 2.64 \cdot 10^{12} = 1.98 \cdot 10^{12}$ or a SES of $5.04 \cdot 10^{-13}$. The upper-limit at the 90% confidence level in the absence of signal using Bayesian statistics is $2.3 \cdot SES = 11.60 \cdot 10^{-13}$.

The Feldman Cousins approach inputs the number of expected background or N_{ACC} into the upper-limit, here we use the average of the toy MC experiments and that of the time sideband ($N_{ACC} = 0.636 + 0.576)/2 = 0.606$). We use the average as the time sideband is limited by statistics whereas the toy MC could contain systematic deviations from the data.

We can also use Poisson statistics and the Feldman Cousins approach to calculate an estimate of the average upper-limit we should expect. Here, the probability of seeing N_{OBS} events is determined by Poisson statistics ($Pois(N_{OBS}, N_{ACC})$). For N_{OBS} we then calculate the upper limit using the Feldman Cousins approach ($UL_{FC}(N_{OBS}, N_{ACC})$). Therefore the average upper-limit we expect is given by the following equation:

$$< UL_{FC} >= SES \cdot \sum_{i=0}^{4} Pois(N_{OBS}, N_{ACC}) \cdot UL_{FC}(N_{OBS}, N_{ACC}) \tag{8.8}$$

This yields a mean upper-limit $< UL_{FC} >= 2.947 \cdot SES = 14.86 \cdot 10^{-13}$. We set our hyperradius cutoff to 3.45 by minimizing this mean upper-limit. Note, this value is skewed

by tail toy MC experiments with a large number of events in the signal region.

In the absence of any events, we achieve an upper-limit of $1.84 \cdot SES = 9.28 \cdot 10^{-13}$, this is the median toy MC experiment upper-limit, which is typically called the sensitivity. Based on the toy MC simulation results, $\sim$ 57% of experiments should result in the absence of events and $\sim$ 32% of experiments should have one event (sum of 89%).

8.11.6 Signal Region

The blinded box analysis detected zero events inside the signal region ($r_H < 3.45$). We opened up the hyperradius such that we detected 10 events; this is achieved by setting $r_H < 5.35$. The kinematics for these events are tabulated below. The event which is closest to the signal signature contains $\theta_{e\gamma} \sim -3.3\sigma_{\theta_{e\gamma}}$ and therefore it is not inside the signal region when added in quadrature with the other values kinematics.

Run #	Event #	$\theta_{e\gamma}$ [mrad]	$\phi_{e\gamma}$ [mrad]	E_e [MeV]	E_γ [MeV]	$t_{e\gamma}$[ns]	r_H	MLA Rank
402458	22	1.3	3.4	52.69	49.51	0.14	4.2	2
401563	1286	-27.7	-2.2	52.97	51.95	-0.11	4.2	1
403059	2406	-13.3	-1.4	52.74	52.01	-0.29	4.3	3
405800	1663	5.2	-31.1	52.67	49.63	-0.06	4.6	-
401603	2718	24.0	-22.8	52.77	49.19	-0.10	4.9	5
405442	9	-30.3	9.9	52.77	49.72	-0.04	4.9	4
401221	892	-14.7	8.7	52.74	49.27	-0.25	5.1	11
401611	2589	-13.5	-0.1	52.74	48.77	0.21	5.1	9
406530	570	9.6	19.6	52.79	49.98	0.21	5.1	13
402692	2734	32.0	-21.8	52.53	51.75	-0.11	5.3	7

This analysis results in an upper-limit on the branching fraction of $\mu^+ \to e^+\gamma$ of $9.28 \cdot 10^{-12}$ at the 90% confidence level using the Feldman Cousin's approach. This is an upper-limit equivalent to the median experiment in the toy MC study. The limit is roughly a factor of two larger than the limit set of MEG I. Note that this result only used data from the short 2021 physics run (roughly 1/3 of the amount of data collected in 2022), whereas the MEG I result used 5 years of data.

8.12 Maximum Likelihood Analysis

8.12.1 Overview

In the cut and count analysis, one of the key shortcomings is that the analysis window needs to be small enough such that the background is significantly suppressed; this forces a reduction in the signal efficiency that degrades the overall sensitivity of the analysis. Recovering this signal efficiency would improve the upper-limit of the cut and count analysis by a factor of 1.3. This is improved in the maximum likelihood analysis. The maximum likelihood analysis inputs the dataset and the set of probability density functions to determine the best fit number of signal, accidental and RMD events in the signal region. The probability density functions describe the expected density of signal, accidental, and RMD events in the $t_{e\gamma}, \theta_{e\gamma}, \phi_{e\gamma}, E_e, E_\gamma$ space.

Here, the analysis window is typically chosen to be significantly wider than that of a blind box to encompass a vast majority of the signal and thus maximize the sensitivity.

8.12.2 Formalism

Over the full dataset we fit for $N, N_{ACC}, N_{RMD}, N_{SIG}$ which represents the number of events, the number of accidentals, the number of RMDs and the number of signal events respectively with the restriction that $N = N_{ACC} + N_{RMD} + N_{SIG}$. We refer to the fit parameters as $\vec{\theta}$. The dataset consists of n measurements each with a set of measurables $\vec{x_i}$. The likelihood function is written below:

$$L(\vec{\theta}) = \prod_{i=1}^{n} p(\vec{x_i}; \vec{\theta}) \tag{8.9}$$

The fit parameters ($\vec{\theta}$) are all Poisson random variables and therefore the product can be instead written as:

$$L(\vec{\theta}) = Pois(n|N)\prod_{i=1}^{n} p(\vec{x_i};\vec{\theta}) = \frac{N^n e^{-N}}{n!}\prod_{i=1}^{n} p(\vec{x_i};\vec{\theta}) = \frac{e^{-N}N^n}{n!}\prod_{i=1}^{n} p(\vec{x_i};\vec{\theta}) \tag{8.10}$$

In our case, the probability density function, $p(\vec{x_i};\vec{\theta})$, can be decomposed into three components using the law of total probability. The decomposition consists of the probability of being signal, accidental, or RMD (e.g. N_{SIG}/N) multiplied by the probability density functions of the signal, accidental, or RMD events. This can be written as shown below where $S(\vec{x_i}), A(\vec{x_i}), R(\vec{x_i})$ are the PDFs of the signal, accidental, and RMD events respectively e.g. $p(\vec{x_i}|SIG) = S(\vec{x_i})$. More details on the breakdown of the PDFs are given in[11]. The probability for a given set of measurables is given by:

$$p(\vec{x_i};\vec{\theta}) = \frac{N_{SIG}\cdot S(\vec{x_i}) + N_{ACC}\cdot A(\vec{x_i}) + N_{RMD}\cdot R(\vec{x_i})}{N} \tag{8.11}$$

We can then rewrite our likelihood as:

$$L(\vec{x_i}|N_{ACC}, N_{RMD}, N_{SIG}) = \frac{N^n e^{-N}}{n!}\prod_{i=1}^{n} \frac{N_{SIG}\cdot S(\vec{x_i}) + N_{ACC}\cdot A(\vec{x_i}) + N_{RMD}\cdot R(\vec{x_i})}{N} \tag{8.12}$$

Note that the N^n terms cancel. This is the maximum likelihood function, i.e. there is a set of parameters $N, N_{SIG}, N_{ACC}, N_{RMD}$ such that this function is maximized.

In addition, we apply three nuisance parameters to the maximum likelihood fit. The first two restrict the parameter limits based on the expected number of accidentals ($< N_{ACC} >$) and

the expected number of RMD events ($< N_{RMD} >$) that were observed in the sidebands. The maximum likelihood is multiplied by a gaussian term $N(< N_{ACC} >, \sigma_{N_{ACC}})$ such that the maximum likelihood is lowered if the fitted value of N_{ACC} deviates from the time sideband expectation; a similar term is added for the RMD events. Finally, the maximum likelihood is modified to allow for a shift of the target in X_{MEG}. The target position, X_{TGT}, shifts the mean value of $\phi_{e\gamma}$ as a function of ϕ_e. The hole analysis resulted in a final uncertainty in the target position in X_{MEG} of $\sim 100\mu m$. The shift of the target is limited by adding another gaussian term $N(0, \sigma_{X_{TGT}})$. This will be discussed more in the conclusions chapter, but in short its incorporation seems unnecessary; adding the parameter implies that we would be estimating the relative position of the target and the CDCH using a few potential signal events.

It is standard to then the apply the negative natural log to both sides of the equation. This is done as optimization routines typically minimize functions rather than maximize them (e.g. minimizing a χ^2). The natural log is applied as it is typically easier to work with sums rather than products. This is shown below. Note, we can remove the $ln(n!)$ term as it is independent of the fit parameters.

$$\begin{aligned} -ln(L(\vec{x_i}|\vec{\theta})) = (\frac{< N_{ACC} > - N_{ACC}}{2\sigma_{N_{ACC}}})^2 + (\frac{< N_{RMD} > - N_{RMD}}{2\sigma_{N_{RMD}}})^2 + (\frac{X_{TGT}}{2\sigma_{X_{TGT}}})^2 \\ + N + ln(n!) - \sum_{i=1}^{n} ln(N_{SIG} \cdot S(\vec{x_i}) + N_{ACC} \cdot A(\vec{x_i}) + N_{RMD} \cdot R(\vec{x_i})) \end{aligned} \tag{8.13}$$

Here, the measurables, $\vec{x_i}$, can be split into two categories. The measurables $\theta_{e\gamma}, \phi_{e\gamma}$, $t_{e\gamma}, p_e, E_\gamma, t_{RDC}, E_{RDC}, n_{SPX}$ are used directly to distinguish between signal, accidental and RMD events. We label these measurables as $\vec{m}_i$. The second set of measurables are known as "conditional observables" i.e. $v_\gamma, w_\gamma, cov_{e,Kalman}, \phi_e$, we label these as $\vec{c_i}$. Our measurables $\vec{m}_i$ may be a function of these conditionals, for example it was observed that the accidental

$t_{e\gamma}$ distribution was dependent on w_γ. The list of conditional observables excludes θ_e, u_γ in order to form a BTSM model-independent likelihood analysis; this is done as different models anticipate different polarization effects and therefore models may be dependent on the value of the signal θ_e or θ_γ.

The probability density functions $S(\vec{x_i})$, $A(\vec{x_i}), R(\vec{x_i})$ are formally written as $S(\vec{m_i}|\vec{c_i})p(\vec{c_i})$, $A(\vec{m_i}|\vec{c_i})p(\vec{c_i})$, $R(\vec{m_i}|\vec{c_i})p(\vec{c_i})$ i.e. the probability of the conditionals times the probability of the measurables given the conditionals. For the actual minimization, we can omit the $p(\vec{c_i})$ terms as these are independent of the fit parameters. Rewriting the minimization function:

$$\begin{aligned} -ln(L(\vec{x_i}|\vec{\theta})) = (\frac{< N_{ACC} > - N_{ACC}}{2\sigma_{N_{ACC}}})^2 + (\frac{< N_{RMD} > - N_{RMD}}{2\sigma_{N_{RMD}}})^2 + (\frac{X_{TGT}}{2\sigma_{X_{TGT}}})^2 \\ N - \sum_{i=1}^{n} ln(N_{SIG} \cdot S(\vec{m_i}|\vec{c_i}) + N_{ACC} \cdot A(\vec{m_i}|\vec{c_i}) + N_{RMD} \cdot R(\vec{m_i}|\vec{c_i})) \end{aligned} \tag{8.14}$$

We use MINUIT[61] to perform the minimization, yielding the optimal values of N, N_{ACC}, N_{RMD}, N_{SIG}, $XTGT$.

8.12.3 Time Sidebands

As mentioned previously, the time sidebands are used to estimate the expected number of accidental events and the distributions for accidentals. In the full analysis region, we expect $N_{ACC} = 68 \pm 3.5$. In addition, time sidebands are used to estimate the best fit number of signal events in the absence of any signal. Several time sidebands were investigated yielding upper-limits at the 90% confidence level of $N_{SIG} < 1.91, 2.17, 3.4, 2.49, 2.61$. However, one time sideband results in a best fit $N_{SIG} = 1.22$ and an upper limit of $N_{SIG} < 5.26$.

In summary, in the absence of signal the typical upper-limit on the number of signal events

in the time sidebands is close to 2, but there is a chance of a significantly higher upper-limit due to accidentals contaminating the signal region. This is a consequence of our current resolutions.

8.12.4 Toy MC Studies

We also use the toy MC simulations containing the expected distribution of accidentals in the signal region to estimate the best fit number of signal events and the upper-limit on the number of signal events in the absence of signal. The upper-limit on the number of signal events in each toy MC is shown in Figure 8.12. This result is consistent with the time sideband studies, the typical upper-limit is ~ 2 with a long tail. The median toy MC experiment results in an upper-limit on N_{SIG} of 2.31 at the 90% confidence level or a median upper limit on the branching fraction of $\mu^+ \to e^+\gamma$ of $8.8 \cdot 10^{-13}$ at the 90% confidence level.

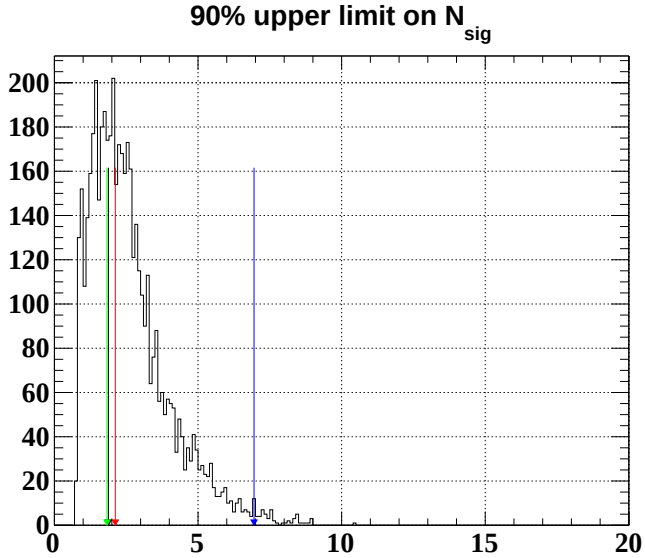


Figure 8.12: The upper-limit on the number of signal events at the 90% confidence level in the 1000 toy MC experiments.

Finally, in Figure 8.13 we show the minimized negative log likelihood for the toy MC experiments.

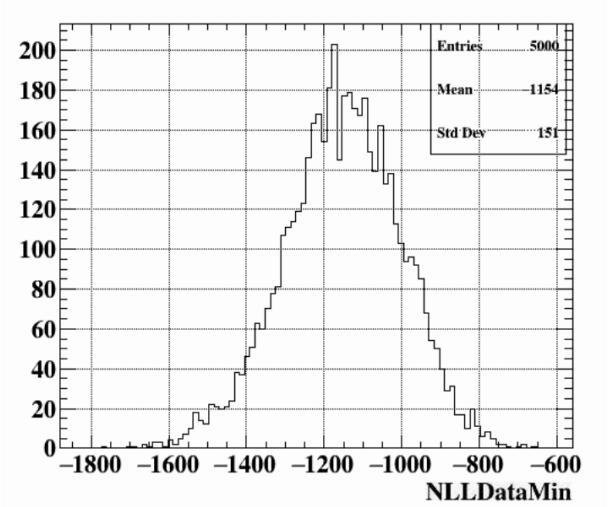


Figure 8.13: The minimized negative log likelihood for each toy MC experiment.

8.12.5 Signal Region

The maximum likelihood analysis applied to the signal region resulted in an upper-limit on the number of signal events of 1.98. This corresponds to an upper-limit on the branching fraction of $\mu^+ \to e^+\gamma$ of $7.5 \cdot 10^{-13}$ at the 90% confidence level ($N_{SIG} < 1.98$). This is very close to the median experiment as shown in Figure 8.12.

The 1D kinematic distributions of the events in the signal region are shown in Figure 8.14. The event distributions are overlaid with the PDF for signal and background each multiplied by the number of fitted events. Visually, the distributions are well-fit. We note that the minimized negative log likelihood is equal to -1100 and therefore consistent with the median toy MC experiment.

The ranking of the events with the highest signal probability in the maximum likelihood analysis is included in the above table (MLA Rank). Note that event 1663 in run 405800 contains an in-time RDC hit ($t_{RDC-LXE}$ consistent with an RMD decay); this information is only used in the maximum likelihood analysis thus this event is only highly ranked in the cut and count analysis. The other top 9 events show a strong agreement between the

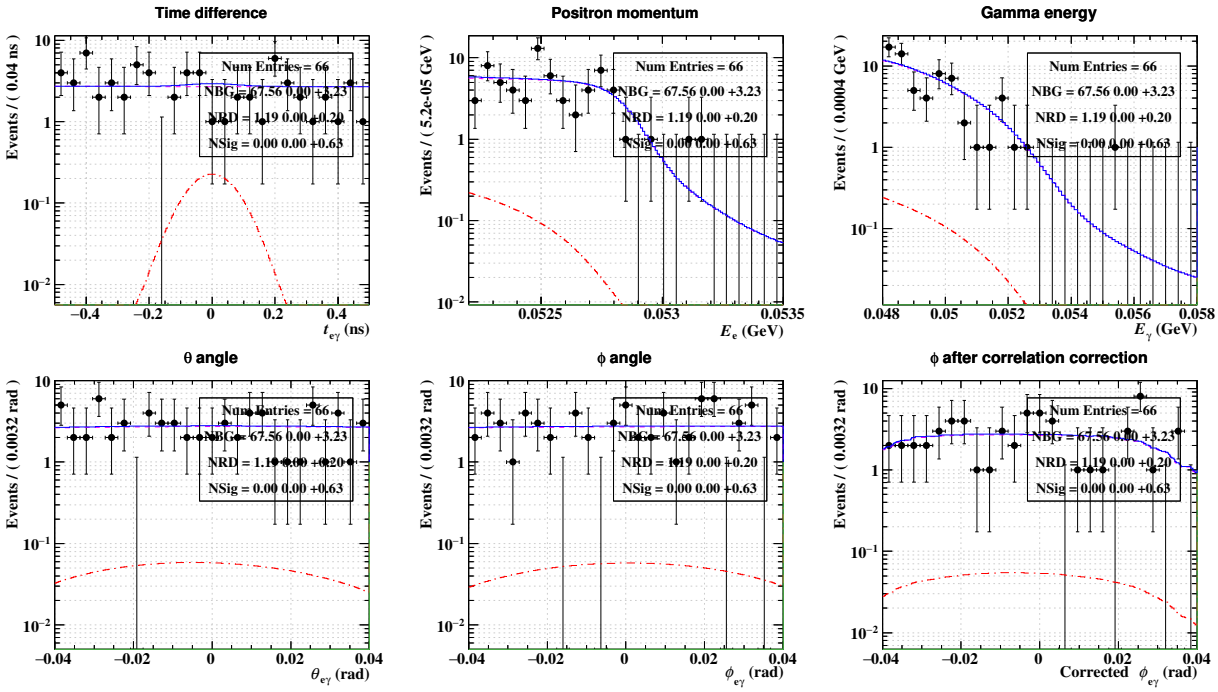


Figure 8.14: 1D histograms of the kinematic distributions in the signal region overlaid with the signal and background PDFs scaled by the number of signal and background events respectively.

two physics analyses; these top 9 events ranked based on the CCA all contain a maximum likelihood analysis ranking less than or equal to 13.

In general, the maximum likelihood analysis is more likely to assign higher signal probability to low energy photon events as the analysis includes the full E_γ PDF. Additional differences can be expected as only the maximum likelihood analysis uses the photon energy PDF that is dependent on the position in the calorimeter.

Chapter 9

Conclusions

9.1 Target Analysis

In Chapter 4, we describe the procedure for determining the position, rotation, and shape of the stopping target with respect to the drift chamber. This analysis relies on a camera analysis that monitors the position, rotation, and shape of the target with respect to the camera effectively continuously by imaging dots printed on the target surface. The camera analysis results in a precision of $\sim 10\mu m$ normal to the target's surface. This is significantly better than that required to suppress potential ϕ_e biases. We also describe how the optical survey of the target, optical survey of the drift chamber and a CT scan of the target were incorporated in the target analysis. This analysis was checked by a technique that measures the relative position of the target and the drift chamber by measuring the position of holes cut in the target surface using the positron vertex distribution.

9.2 Positron Analysis

The positron analysis was described in detail in Chapter 5. In summary, the analysis results in hit distance of closest approach measurements on-par with expectations of the design of the experiment with a core resolution of $\sim 110\mu m$ and a tail resolution of $\sim 240\mu m$ ($A_2/A_1 \sim 0.59$). At the kinematic level, the core resolutions at $4 \cdot 10^7 \mu/s$ beam rate are $0.74, 2.00$ mm for y_e, z_e respectively, $4.07, 7.23$ mrad for ϕ_e, θ_e respectively, and 103 keV for p_e; these are the core resolutions when fit to a double gaussian where $A_{tail}/A_{core} \sim 0.1$. These resolutions include correlations between the kinematic variables. Overall, these are comparable to that of the design of the experiment where σ_{p_e} is significantly improved with respect to the design ($\sim 30\%$) and σ_{θ_e} is degraded by $\sim 36\%$ with respect to the design of the experiment.

9.3 2021 Physics Result

The cut and count analysis (Section 8.11) resulted in a null signal and thus an upper-limit on the branching fraction of $\mu^+ \to e^+\gamma$ of $9.28 \cdot 10^{-13}$ at the 90% confidence level. The maximum likelihood analysis (Section 8.12) resulted in an improved upper-limit of $7.5 \cdot 10^{-13}$ at the 90% confidence level due to the higher signal efficiency than that of the CCA. The events with the highest probability of being signal in the blinded region show strong agreement between the two analyses. Both results are consistent with the sensitivity measurements from the toy MC simulations in the absence of signal ($8.8 \cdot 10^{-13}$ and $9.28 \cdot 10^{-13}$ for the MLA and the CCA respectively).

9.4 Expected Sensitivity of the Full Dataset

The effective number of muons observed in the 2021,2022,2023 physics runs are shown in Figure 9.1. The 2021 physics run was a brief run consisting of roughly 1/7 of the data taken in 2022,2023 and the 2023 run is still ongoing. The amount of data already taken is expected to reach a sensitivity of $\sim 9 \cdot 10^{-14}$ at the 90% confidence level.

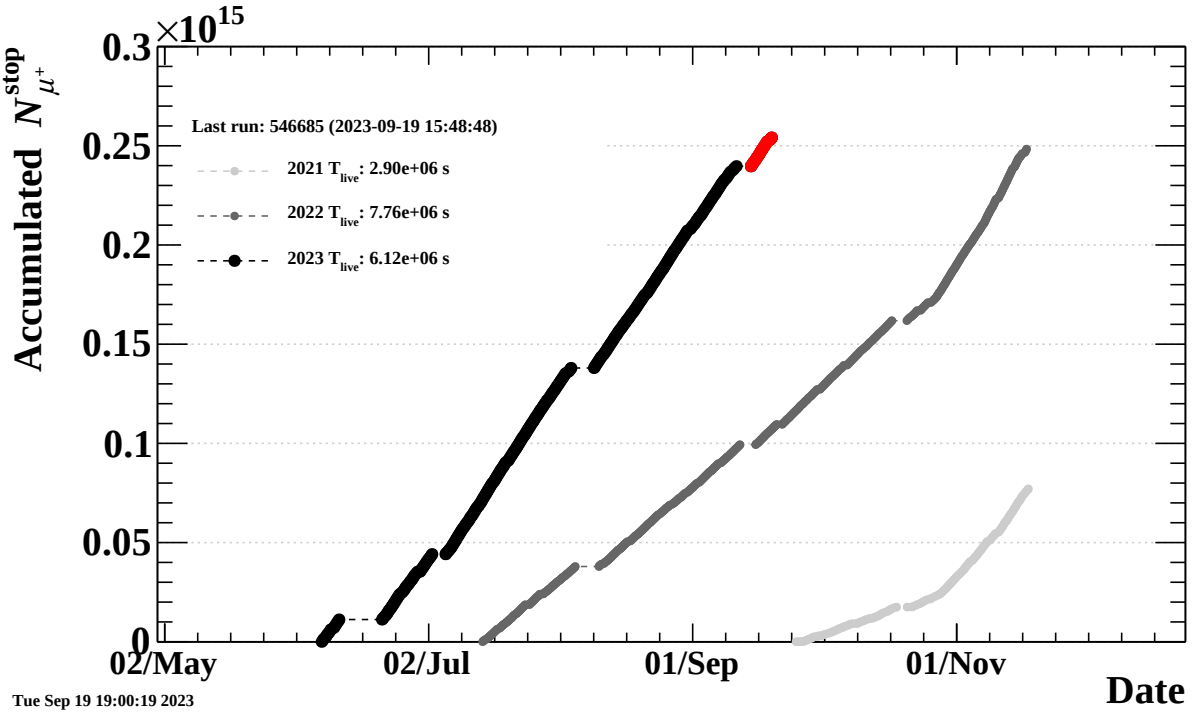


Figure 9.1: The effective number of muons observed in 2021, 2022, 2023 as of September 19th 2023.

Using the data taken in 2022 and 2023 and an estimate of the DAQ weeks for future runs, we estimate the MEG II sensitivity over the lifespan of the experiment. This is shown in Figure 9.2 assuming 20 DAQ weeks per year. Here, the median expected sensitivity is shown as a black dot whereas the green and yellow bars represent the 68,90 percentile respectively. The current likelihood analysis expects that the MEG II design goal of an upper-limit on the branching fraction of $6 \cdot 10^{-14}$ at the 90% confidence level could be achieved in roughly 65 DAQ weeks.

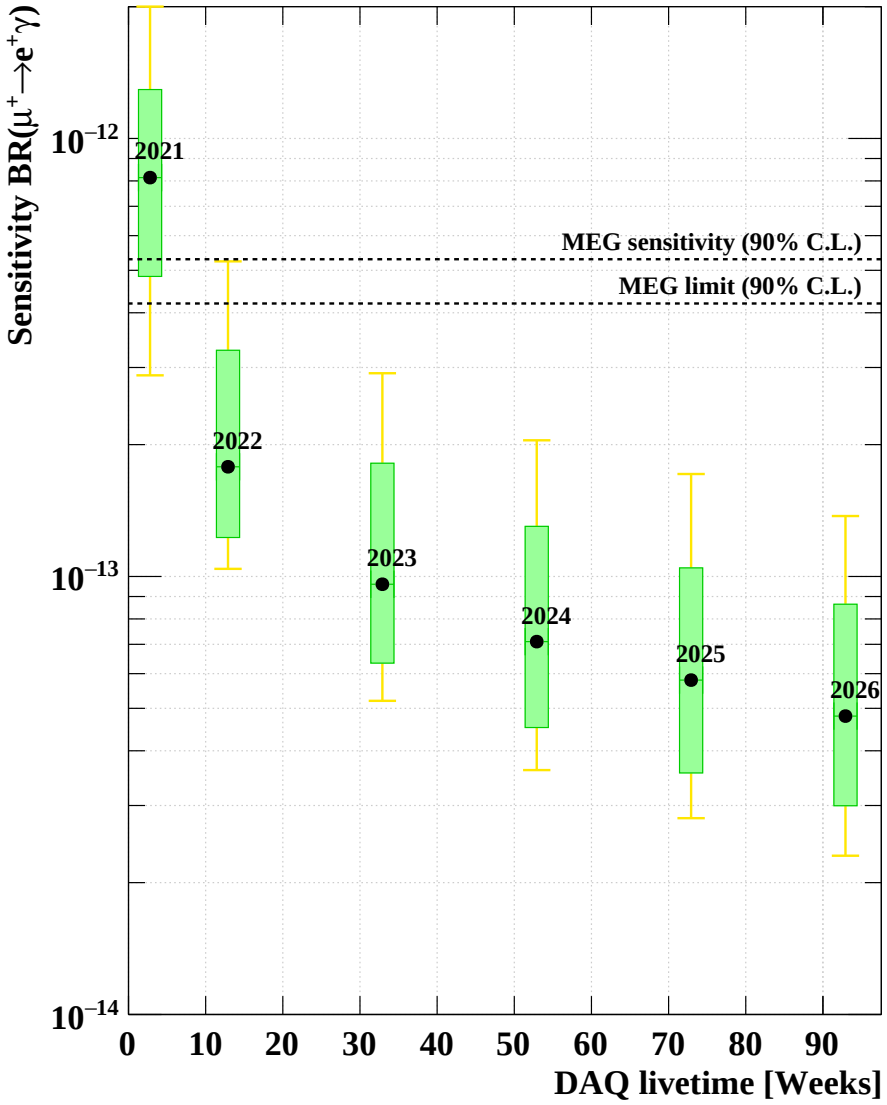


Figure 9.2: The sensitivity of MEG II over the lifetime of the experiment.

9.5 Potential Improvements

In this section, shortcomings and potential methods for improvements are discussed when applicable.

9.5.1 Magnetic Field Map

In Chapter 6, we mentioned that the currently implemented magnetic field map is that theoretically calculated using the estimated position of the coils. This map yielded the best resolutions as seen in the double turn technique and the Michel edge fit. This map was used despite three attempts to measure the magnetic field map. Each has their own issues, e.g. the first magnetic field map is expected to have bowing in the translation stage in which the hall probes were moved, the final field map requires calibrations of the hall probes.

In the attached document, a MC analysis is discussed where the MC data was analyzed with a shifted magnetic field or a different field map (e.g. analyzed using a measured field and generated with the field calculated using the coil positions). This resulted in $\sim 10\%$ degradations in the momentum resolution and small degradations in the other kinematics.

It's unclear how optimized the current magnetic field map is, we can only state that the mean value of the Michel edge is centered at zero as a function of θ_e, ϕ_e and that the double turn distribution biases are suppressed.

For this reason, a more thorough investigation of the magnetic field map could yield an improvement in the momentum resolution of the order $\sim 10\%$. For example, attempting to rotate the magnetic field map with respect to the CDCH. In addition, one could attempt to calibrate the final magnetic field map measurement. Finally, one could try applying restrictions from Maxwell's equations to all magnetic field maps; this was only tried with the first magnetic field map and yielded a significant improvement with respect to the true measured field map (still inferior to the calculated field).

9.5.2 Track Finder

The track finder is discussed in Chapter 5.4 and is also discussed in [14]. The track finder was not optimized on data, but based on the Monte Carlo simulation. It's possible that through parameter optimization, the track finding efficiency could be improved.

By opening up the track finder to significantly more hits (using the ML hit finding algorithm developed by Yusuke Uchiyama), a significantly higher tracking efficiency can be achieved even through it is expected that a large fraction of the CDCH hits are already reconstructed properly. This indicates that there may be some inefficiencies of the track finder.

One could imagine training a neural network in the Monte Carlo where the inputs are the CDCH hit arrival times and CDCH hit z (e.g. time difference on the two ends), SPX cluster times/positions, and the network trains to cluster the hits/clusters into tracks. For example, graph neural networks have been used to train for this purpose[62] and could be explored in MEG II.

9.5.3 Photon Energy

Integrating the photon energy resolution over the full calorimeter (resolution is dependent on the depth), the resolution is $\sim 1.9\%$. Unfortunately, this is significantly degraded with respect to the design of the MEG II experiment ($\sim 1.05\%$); this should result in significantly more accidentals in the signal region than expected.

This could be due to a number of reasons such as issues with the reconstruction algorithm, pileup discrimination, and calibration. As mentioned, even after calibrating the MPPC/PMT gains (Chapter 7), corrections to the energy as a function of position and time had to be applied to the photon energy reconstruction. In theory, when using a perfect reconstruction algorithm, pileup discrimination and gain calibration, these corrections should not be

necessary.

We note that when checking the time sideband data, we observed a significant amount of events that clearly contained pileup, but were not flagged by the pileup code. These events remained in the final analysis although they were properly handled by the PDFs. Alternatives to the pileup discrimination algorithm should be explored to eliminate these events. As mentioned in the last section, graph neural networks have also been applied to reconstruct calorimeter energy[62]. Applications like this could aid, for example, in correctly assigning the energy to a given incoming photon (avoiding pileup).

It's also possible that this is due to MPPCs/PMTs calibration issues. The UCI group has discussed alternative techniques to relative MPPC/PMT gain/time calibrations. This involves finding all pairs of MPPCs that are the same distance away from shower centers; these pairs should have the same amount of charge, deviations from this suggest a miscalibration. Using this pair matrix, one can deduce the gain/timing calibration.

9.5.4 Target Hole Alignment

The target alignment was achieved using a variety of techniques. The relative position/rotation of the target with respect to the CDCH is estimated by the survey and then verified by the hole analysis. The hole analysis results in some systematics that have not been understood. For example, the CT scan measured the relative position of the holes very precisely. The distance between the holes was then cross-checked in the hole analysis; it was found that these do not agree. This implies that there are significant systematics in this measurement that are not fully understood. This measurement is difficult as it requires finding the position of the holes in the vertex distribution on a falling slope (significantly less beam intensity on one side of the hole than the other). Alternative methods to fit for the hole positions should be explored.

9.5.5 Optical Survey

The optical survey alignment of the detectors (CDCH, LXe, SPX) is quoted to < 1 mm. This unfortunately can potentially lead to large systematic errors in $\theta_{e\gamma}, \phi_{e\gamma}$ ($\sim 1-2$ mrad). Alternative techniques to metrology of the detectors should be explored. This was discussed in Chapter 5.5.

Bibliography

[1] Francesca Borzumati and Antonio Masiero. Large muon- and electron-number nonconservation in supergravity theories. *Phys. Rev. Lett.*, 57:961–964, Aug 1986.

[2] Andre de Gouvea and Petr Vogel. Lepton Flavor and Number Conservation, and Physics Beyond the Standard Model. *Prog. Part. Nucl. Phys.*, 71:75–92, 2013.

[3] L. Bartoszek et al. Mu2e Technical Design Report. Technical report, Fermi National Accelerator Lab., 10 2014.

[4] A. Blondel et al. Research Proposal for an Experiment to Search for the Decay $\mu \to eee$. Technical report, Paul Scherrer Institute, 1 2013.

[5] A. M. Baldini et al. Search for the lepton flavour violating decay $\mu^+ \to e^+\gamma$ with the full dataset of the MEG experiment. *Eur. Phys. J. C*, 76(8):434, 2016.

[6] Walter Blum, Werner Riegler, and Luigi Rolandi. *Particle detection with drift chambers; 2nd ed.* Springer, Berlin, 2008.

[7] A. M. Baldini et al. The design of the MEG II experiment. *Eur. Phys. J. C*, 78(5):380, 2018.

[8] Hajime Nishiguchi. The meg positron spectrometer. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 581(1):538–541, 2007. VCI 2007.

[9] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.

[10] T. Libeiro, W. Kyle, S. Kobayashi, M. Francesconi, L. Galli, K. Ieki, T. Iwamoto, W. Molzon, T. Mori, M. Nakao, D. Nicolò, S. Ogawa, W. Ootani, and D. Palo. Novel x-ray scanning technique for in-situ alignment of photo-detectors in the megii calorimeter. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 1048:167901, 2023.

[11] et. al Kobayashi .S, Oya A. Likelihood analysis for 2021 dataset. Technical report, University of Tokyo, 2023.

[12] Alessandro M. Baldini et al. The Search for $\mu+ \to e+\gamma$ with 10–14 Sensitivity: The Upgrade of the MEG Experiment. *Symmetry*, 13(9):1591, 2021.

[13] K. Afanaciev et al. Operation and performance of MEG II detector. *Submitted*, 10 2023.

[14] A. M. Baldini et al. Performances of a new generation tracking detector: the MEG II cylindrical drfit chamber. *Submitted*, 10 2023.

[15] K. Afanaciev et al. A search for $\mu^+ \to e^+\gamma$ with the first dataset of the MEG II experiment. *Submitted*, 10 2023.

[16] R. L. Workman and Others. Review of Particle Physics. *PTEP*, 2022:083C01, 2022.

[17] R. R. Crittenden, W. D. Walker, and J. Ballam. Radiative decay modes of the muon. *Phys. Rev.*, 121:1823–1832, Mar 1961.

[18] Y. Fukuda et al. Evidence for oscillation of atmospheric neutrinos. *Phys. Rev. Lett.*, 81:1562–1567, 1998.

[19] Ivan Esteban, M.C. Gonzalez-Garcia, Michele Maltoni, Thomas Schwetz, and Albert Zhou. The fate of hints: updated global analysis of three-flavor neutrino oscillations. *Journal of High Energy Physics*, 2020(9), sep 2020.

[20] T. P. Cheng and Ling-Fong Li. $\mu \to e\gamma$ in Theories With Dirac and Majorana Neutrino Mass Terms. *Phys. Rev. Lett.*, 45:1908, 1980.

[21] L.J. Hall, V.A. Kostelecky, and S. Raby. New flavor violations in supergravity models. *Nuclear Physics B*, 267(2):415–432, 1986.

[22] F. Abusalma et al. Expression of Interest for Evolution of the Mu2e Experiment. Technical report, Fermi National Accelerator Lab., 2 2018.

[23] Yongsoo Jho, Simon Knapen, and Diego Redigolo. Lepton-flavor violating axions at MEG II. *Journal of High Energy Physics*, 2022(10), oct 2022.

[24] Yongsoo Jho, Simon Knapen, and Diego Redigolo. Lepton-flavor violating axions at MEG II. *Journal of High Energy Physics*, 2022(10), oct 2022.

[25] K. Hayasaka et al. Search for Lepton Flavor Violating Tau Decays into Three Leptons with 719 Million Produced Tau+Tau- Pairs. *Phys. Lett. B*, 687:139–143, 2010.

[26] Bernard Aubert et al. Searches for Lepton Flavor Violation in the Decays tau+- —> e+- gamma and tau+- —> mu+- gamma. *Phys. Rev. Lett.*, 104:021802, 2010.

[27] William J. Marciano, Toshinori Mori, and J. Michael Roney. Charged Lepton Flavor Violation Experiments. *Ann. Rev. Nucl. Part. Sci.*, 58:315–341, 2008.

[28] Wilhelm H. Bertl et al. A Search for muon to electron conversion in muonic gold. *Eur. Phys. J. C*, 47:337–346, 2006.

[29] R. Abramishvili et al. COMET Phase-I Technical Design Report. *PTEP*, 2020(3):033C01, 2020.

[30] Pierre Billoir. Track Fitting With Multiple Scattering: A New Method. *Nucl. Instrum. Meth. A*, 225:352–366, 1984.

[31] Moritz Nadler and Rudolf Fruhwirth. Robust track fitting in the Belle II inner tracking detector. *J. Phys. Conf. Ser.*, 396:022037, 2012.

[32] M. Ahmed, J. F. Amann, D. Barlow, K. Black, R. D. Bolton, M. L. Brooks, S. Carius, Y. K. Chen, A. Chernyshev, H. M. Concannon, M. D. Cooper, P. S. Cooper, J. Crocker, J. R. Dittmann, M. Dzemidzic, A. Empl, R. J. Fisk, E. Fleet, W. Foreman, C. A. Gagliardi, D. Haim, A. Hallin, C. M. Hoffman, G. E. Hogan, E. B. Hughes, E. V. Hungerford, C. C. H. Jui, G. J. Kim, J. E. Knott, D. D. Koetke, T. Kozlowski, M. A. Kroupa, A. R. Kunselman, K. A. Lan, V. Laptev, D. Lee, F. Liu, R. W. Manweiler, R. Marshall, B. W. Mayes, R. E. Mischke, B. M. K. Nefkens, L. M. Nickerson, P. M. Nord, M. A. Oothoudt, J. N. Otis, R. Phelps, L. E. Piilonen, C. Pillai, L. Pinsky, M. W. Ritter, C. Smith, T. D. S. Stanislaus, K. M. Stantz, J. J. Szymanski, L. Tang, W. B. Tippens, R. E. Tribble, X. L. Tu, L. A. Van Ausdeln, W. H. von Witch, D. Whitehouse, C. Wilkinson, B. Wright, S. C. Wright, Y. Zhang, and K. O. H. Ziock. Search for the lepton-family-number nonconserving decay $\mu^+ \to e^+\gamma$. *Phys. Rev. D*, 65:112002, Jun 2002.

[33] D. Palo, M. Hildebrandt, A. Hofer, W. Kyle, D. Lad, T. Libeiro, and W. Molzon. Precise Photographic Monitoring of MEG II Thin-film Muon Stopping Target Position and Shape. *Nucl. Instrum. Meth. A*, 944:162511, 2019.

[34] G. Cavoto, G. Chiarello, M. Hildebrandt, A. Hofer, K. Ieki, M. Meucci, S. Milana, V. Pettinacci, F. Renga, and C. Voena. A photogrammetric method for target monitoring inside the MEG II detector. *Review of Scientific Instruments*, 92(4):043707, 04 2021.

[35] A.M. Baldini, H. Benmansour, G. Cavoto, F. Cei, M. Chiappini, G. Chiarello, C. Chiri, G. Cocciolo, A. Corvaglia, F. Cuna, M. Francesconi, L. Galli, F. Grancagnolo, M. Grassi, M. Meucci, A. Miccoli, D. Nicoló, M. Panareo, A. Papa, C. Pinto, F. Raffaelli, F. Renga, G. Signorelli, G.F. Tassielli, A. Venturini, B. Vitali, and C. Voena. The measuring systems of the wire tension for the meg ii drift chamber by means of the resonant frequency technique. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 1045:167534, 2023.

[36] Paolo W. Cattaneo, Matteo De Gerone, Flavio Gatti, Miki Nishimura, Wataru Ootani, Massimo Rossella, and Yusuke Uchiyama. Development of High Precision Timing Counter Based on Plastic Scintillator with SiPM Readout. *IEEE Trans. Nucl. Sci.*, 61(5):2657–2666, 2014.

[37] Nishimura M. *Positron Timing Measurement to Search for Lepton Flavor Violating Decay in MEG II.* PhD thesis, University of Tokyo, 2018.

[38] Ayaka Matsushita, Sei Ban, Hicham Benmansour, Giovanni Dal Maso, Marco Francesconi, Luca Galli, Fumihito Ikeda, Toshiyuki Iwamoto, Satoru Kobayashi, Toshinori Mori, Donato Nicolo, Rina Onda, Wataru Ootani, Atsushi Oya, Shinji Ogawa,

Angela Papa, Patrick Schwendimann, Yusuke Uchiyama, Bastiano Vitali, Kensuke Yamamoto, Taku Yonemoto, and Keisuke Yoshida. Commissioning of liquid xenon gamma-ray detector for meg ii experiment. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 1047:167877, 2023.

[39] Kobayashi S. *Full Commissioning of Liquid Xenon Scintillation Detector to Search for $\mu+ \to e+\gamma$ with the Highest Sensitivity in MEG II Experiment*. PhD thesis, University of Tokyo, 2022.

[40] Ogawa S. *Development of liquid xenon gamma-ray detector for MEG II experiment with highly granular readout realized by VUV sensitive large area MPPC*. PhD thesis, University of Tokyo, 2016.

[41] Onda R. *Suppresion of γ-ray backgrounds for the highest sensitivity of $\mu+ \to e+\gamma$ search in MEG II experiment*. PhD thesis, University of Tokyo, 2021.

[42] Stefan Ritt, Roberto Dinapoli, and Ueli Hartmann. Application of the drs chip for fast waveform digitizing. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 623(1):486–488, 2010. 1st International Conference on Technology and Instrumentation in Particle Physics.

[43] Marco Francesconi, Alessandro Baldini, Hicham Benmansour, Fabrizio Cei, Marco Chiappini, Gianluigi Chiarello, Luca Galli, Marco Grassi, Ueli Hartmann, Fabio Morsani, Donato Nicoló, Angela Papa, Stefan Ritt, Elmar Schmid, Giovanni Signorelli, and Bastiano Vitali. The wavedaq integrated trigger and data acquisition system for the meg ii experiment. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 1045:167542, 2023.

[44] Marco Francesconi et al. The trigger system for the MEG II experiment. *Nucl. Instrum. Meth. A*, 1046:167736, 2023.

[45] The Imaging Source. Dmk 33uj003 usb 3.0 monochrome industrial camera, 2020.

[46] Spectral Instruments. Rvt100 radiation/vacuum tolerant video imaging camera, 2022.

[47] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.

[48] Johannes Rauch and Tobias Schlüter. Genfit — a generic track-fitting toolkit. *Journal of Physics: Conference Series*, 608(1):012042, apr 2015.

[49] R. Brun and F. Rademakers. ROOT: An object oriented data analysis framework. *Nucl. Instrum. Meth. A*, 389:81–86, 1997.

[50] R. Veenhof. GARFIELD, recent developments. *Nucl. Instrum. Meth. A*, 419:726–730, 1998.

[51] S. Agostinelli et al. GEANT4–a simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250–303, 2003.

[52] M. Cascella, F. Grancagnolo, and G. Tassielli. Cluster counting/timing techniques for drift chambers. *Nuclear Physics B - Proceedings Supplements*, 248-250:127–130, 2014. 1st Conference on Charged Lepton Flavor Violation.

[53] G. Cataldi, F. Grancagnolo, and S. Spagnolo. Cluster counting in helium based gas mixtures. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 386(2):458–469, 1997.

[54] François Chollet et al. Keras. `https://keras.io`, 2015.

[55] Baracchini E. Renga F. Positron variables correlations and resolutions and how to extract them from data. Technical report, INFN, KEK Respectively, 2011.

[56] et. al Fujii Y., Iwamoto T. Likelihood analysis on meg data 2009–2011. Technical report, University of Tokyo, INFN, 2013.

[57] K. Arisaka, L. B. Auerbach, S. Axelrod, J. Belz, K. A. Biery, P. Buchholz, M. D. Chapman, R. D. Cousins, M. V. Diwan, M. Eckhause, J. F. Ginkel, C. Guss, A. D. Hancock, A. P. Heinson, V. L. Highland, G. W. Hoffmann, J. Horvath, G. M. Irwin, D. Joyce, T. Kaarsberg, J. R. Kane, C. J. Kenney, S. H. Kettell, W. W. Kinnison, P. Knibbe, J. Konigsberg, Y. Kuang, K. Lang, D. M. Lee, J. Margulies, C. Mathiazhagan, W. K. McFarlane, R. J. McKee, P. Melese, E. C. Milner, W. R. Molzon, D. A. Ouimette, P. J. Riley, J. L. Ritchie, P. Rubin, G. H. Sanders, A. J. Schwartz, M. Sivertz, W. E. Slater, J. Urheim, W. F. Vulcan, D. L. Wagner, R. E. Welsh, R. J. Whyley, R. G. Winter, M. T. Witkowski, S. G. Wojcicki, A. Yamashita, and H. J. Ziock. Improved upper limit on the branching ratio b($k_l^0 \to \mu^\pm e^\mp$). *Phys. Rev. Lett.*, 70:1049–1052, Feb 1993.

[58] D. Ambrose et al. New limit on muon and electron lepton number violation from K0(L) —> mu+- e-+ decay. *Phys. Rev. Lett.*, 81:5734–5737, 1998.

[59] Joshua R Klein and Aaron Roodman. Blind analysis in nuclear and particle physics. *Annual Review of Nuclear and Particle Science*, 55(1):141–163, 2005.

[60] Gary J. Feldman and Robert D. Cousins. A Unified approach to the classical statistical analysis of small signals. *Phys. Rev. D*, 57:3873–3889, 1998.

[61] F. James. MINUIT Function Minimization and Error Analysis: Reference Manual Version 94.1. Technical report, CERN, 1994.

[62] Jonathan Shlomi, Peter Battaglia, and Jean-Roch Vlimant. Graph neural networks in particle physics. *Machine Learning: Science and Technology*, 2(2):021001, dec 2020.

Appendix A

Camera Analysis Appendices

A.1 Absolute Target Position Using Spectrometer Tracking

Here, we describe the technique used to check the position and orientation of the target with respect to the magnetic spectrometer independent of an optical survey. The technique is based on imaging small holes in the target using Michel positron trajectories measured in the magnetic spectrometer; it was first developed for MEG[5]. This is an important part of the alignment procedure since it provides a check for potential errors in the optical survey of both the target and the tracking chambers in the magnetic spectrometer, in the magnetic field measurements, and in the particle tracking. The technique provides a measurement with very limited time dependence due to limited statistics, and can only measure the position of a limited number of holes and only near the target center where the muon stopping rate is high. The photographic technique provides an effectively continuous monitoring of changes in the target position and shape over the full surface of the target to allow maximum use of the limited tracking statistics.

The procedure is as follows. A set of small holes ($\sim$ 6 mm diameter) will be made in the target film (6 were used in MEG). The positions of the holes will be imaged in 3 dimensions by detecting the deficit of positrons originating from the hole locations. The coordinate normal to the target surface is determined by measuring the apparent position of each hole in the target plane as a function of the angle of the positron as it intercepts the assumed target plane. We determine the position and direction of e^+ at the target by projecting the helical trajectory measured in the spectrometer back to the plane of the target[5]. Target deformation and translation normal to the target surface are correlated with the linear first order dependence of each hole's apparent y_M position on $tan\phi$, where ϕ is the e^+ angle at the target plane. The optimal target plane position is found when the apparent hole position is independent of angle.

The MEG experiment used this correlation to measure each hole's position normal to the target surface with an uncertainty σ, which varied by year, ranging from 0.3-0.5 mm. This produced a systematic uncertainty in the signal acceptance that reduced the total sensitivity by 13%[5]. The uncertainty was largely the result of the lack of statistics available to measure the time dependence of the position. The photographic technique will allow time dependent corrections to the target position and allow the full tracking statistics to produce a high statistics check of the spectrometer-target relative alignment. This will allow full exploitation of the improved angular resolution of MEG II.

A.2 Threshold Parameter

Here we confirm that the photographic technique is independent of the threshold parameter (described in Section 4.4). First, we calculated the dispersion in individual dot positions in a single image while varying the threshold parameter from 50 to 100 (out of full scale value of 255). The dispersion in the dot's position, $\sigma = 0.10$ pixels, is comparable to the dispersion

in sequential images taken with the same threshold parameter ($\sigma = 0.12$ pixels).

Additionally, we analyzed 20 sequential images with varying threshold parameters in the range 50-90 and calculated the dispersion in the values of the fit parameters for each image. The dispersion in the parameters for a given image with varying threshold parameter is comparable to the dispersion from sequential images with a constant threshold parameter, implying there are no systematic effects larger than the dispersion in the value of fitted parameters from sequential images. Both dispersion measurements are shown in Table B1.

Table B1: Top: The dispersion in the value of the fitted parameters for a given image with a varying threshold parameter (5 values ranging 50-90, out of full scale value of 255). Bottom: The dispersion in the value of the fitted parameters from sequential images with a constant threshold parameter is shown for reference.

ϕ[mrad]	θ[mrad]	ψ[mrad]	x^C[μm]	y^C[μm]	z^C[μm]	Bow[μm]
0.02	0.01	0.03	0.96	0.76	34.03	1.86
0.02	0.01	0.03	0.71	0.61	39.13	1.00

A.3 Determination of Camera's Effective Focal Length

Here, we discuss the determination of the camera's effective focal length (*efl*), which is related to the value of I discussed in Section 4.4.4 by the focal length approximation for an in-focus object. The manufacturer gives a nominal value of 50 mm for our complex lens.

The *efl* is measured by exploiting the fact that an incorrect value for the *efl* creates characteristic systematic residuals for the dot positions in an object with significant depth of field. As an example, two residual plots with different fixed values of I are shown in Figure A.1.

By including I as an additional parameter in the fit, we reduce the magnitude of these residuals. We minimized the χ^2 defined in Section 4.1 (now with the additional parameter) on 20 sequential images to calculate the average optimal I. The optimal value of I is found to be 51.61 ± 0.05 mm, corresponding to a best fit value of $efl = 49.47$ mm, close to the nominal

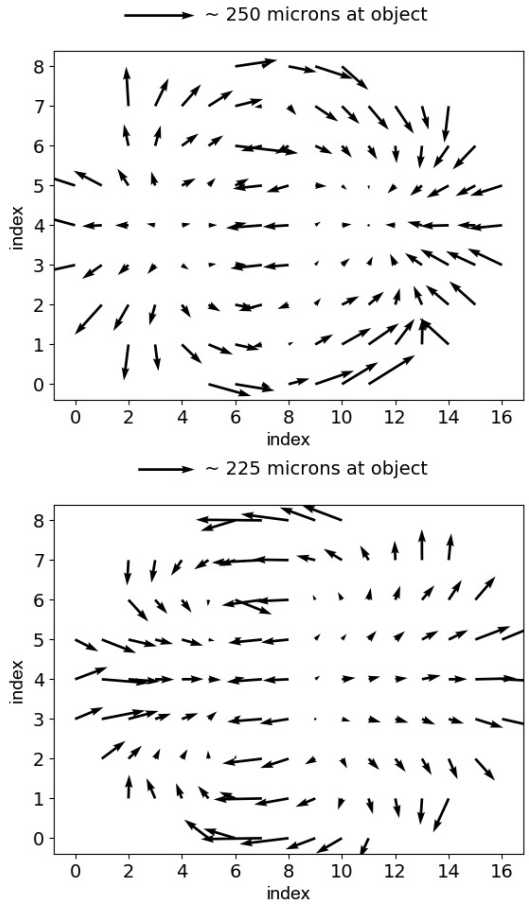


Figure A.1: The residuals are plotted as a function of their row and column indices. The residuals on the top and bottom are from a minimization with a fixed I of 48 mm and 54 mm respectively.

50 mm focal length given by the manufacturer. I is fixed to this value for all analyses.

Further, we verified that using a value of I (and therefore the effective focal length) different than the best fit value does not affect the *change* in measured target position, orientation, and shape. We analyzed 20 sequential images with fixed I values ranging from 49-53 mm. For each image in the set, the dispersion in the fit parameters as the value of I is changed is significantly lower that of the dispersion from sequential images.

Appendix B

Neural Network Applications to Improve Drift Chamber Track Position Measurements Appendices

B.1 Positron Measurements Using The Same σ_{DOCA}

As a check, we verify that the double turn measurements are still improved in the neural network cases when using the same Kalman filter σ_{DOCA} with all DOCA estimators. Here, the resolution when using the conventional TXY function has effectively stayed the same, and in some kinematics (e.g. momentum comparison) the resolution is degraded. Here, we adjusted the track selection on the number of hits per track to achieve a comparable tracking efficiency.

B.2 Positron Kinematic Measurements When Suppressing the Bias in the Conventional TXY function

The conventional analysis results in a ¡DOCA bias¿ of $\sim 50\mu m$; a possible way to correct for this is a simple correction to the drift time to remove this bias. The conventional TXY function was modified, transforming $t_{drift} \rightarrow t_{drift} - 2ns$. In Figure B.2, the hit resolutions are compared again in the three cases.

In Figure B.3, we compare the positron double turn measurement resolutions with the modified conventional TXY function. We note that just suppressing this ¡DOCA bias¿ in the conventional approach does indeed yield a slight resolution improvement; the convolutional neural network case still results in the optimal resolution.

Finally, using the same σ_{DOCA} as the neural network approach and applying the $t_{drift} \rightarrow t_{drift} - 2ns$ transformation also results in comparable measurements to that shown in this appendix.

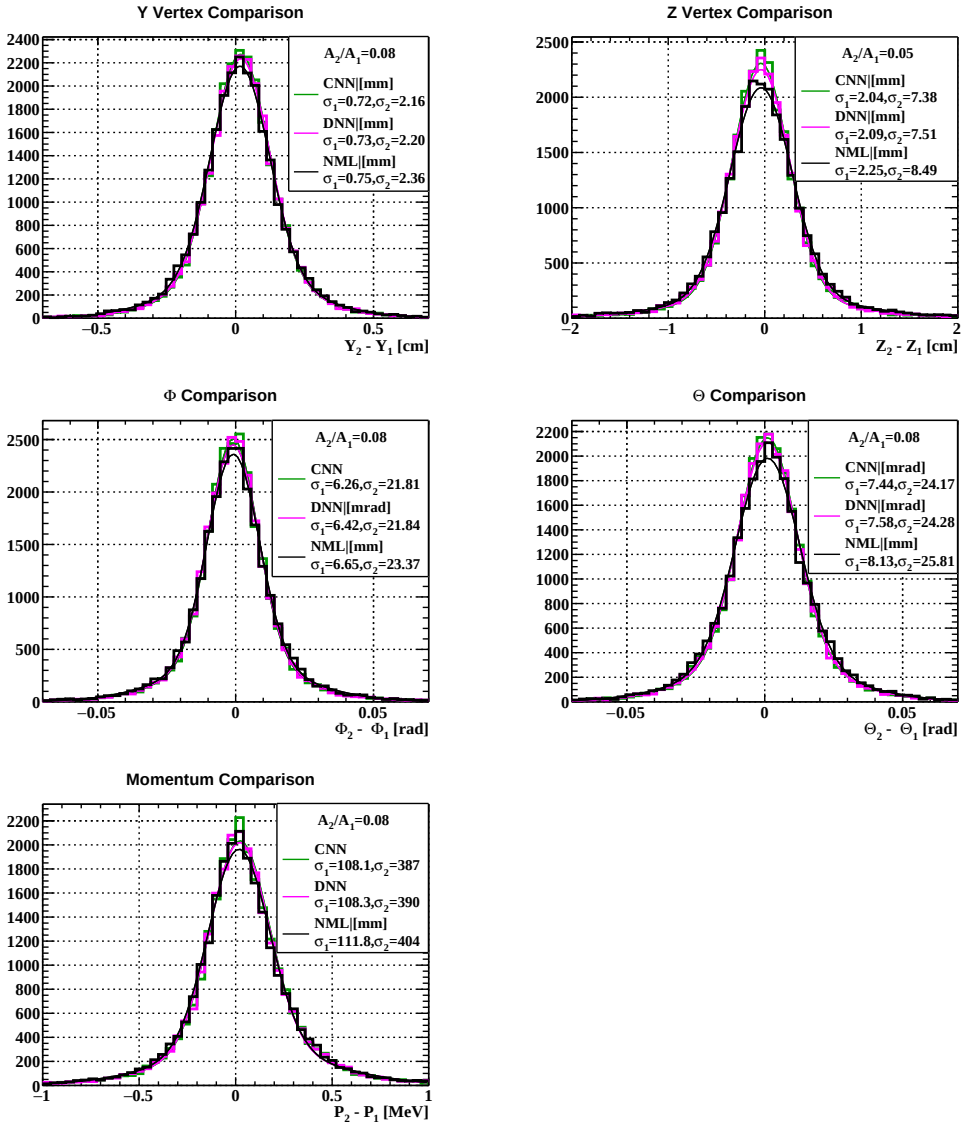


Figure B.1: Here, we show the kinematic comparison at a common plane between two independently fit/measured turns of a single two-turn track. In this plot, all fits use the same Kalman filter σ_{DOCA}. The NML case represents the standard or conventional DOCA estimation, DNN represents the dense neural network, and CNN represents the convolutional neural network that inputs waveform voltages.

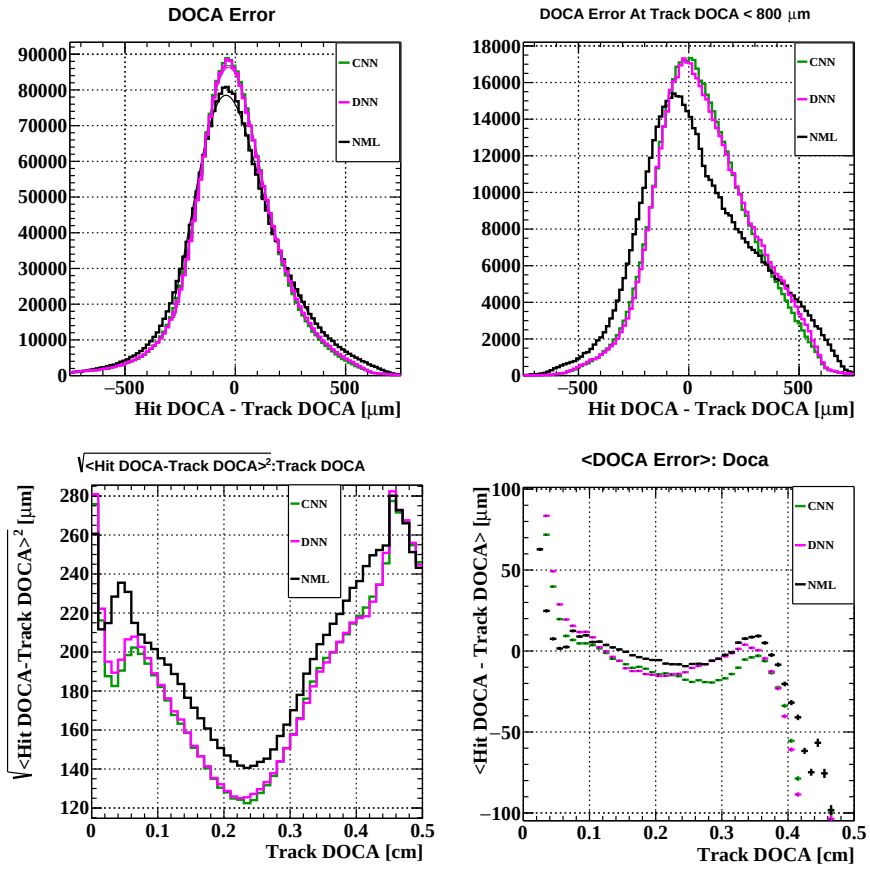


Figure B.2: The hit level results are compared using the three DOCA estimators. In this plot, the conventional TXY function has been modified to suppress the DOCA bias; transforming $t_{drift} \rightarrow t_{drift} - 2ns$. NML represents the standard or conventional DOCA estimation, DNN represents the dense neural network, and CNN represents the convolutional neural network that includes the waveform voltages.

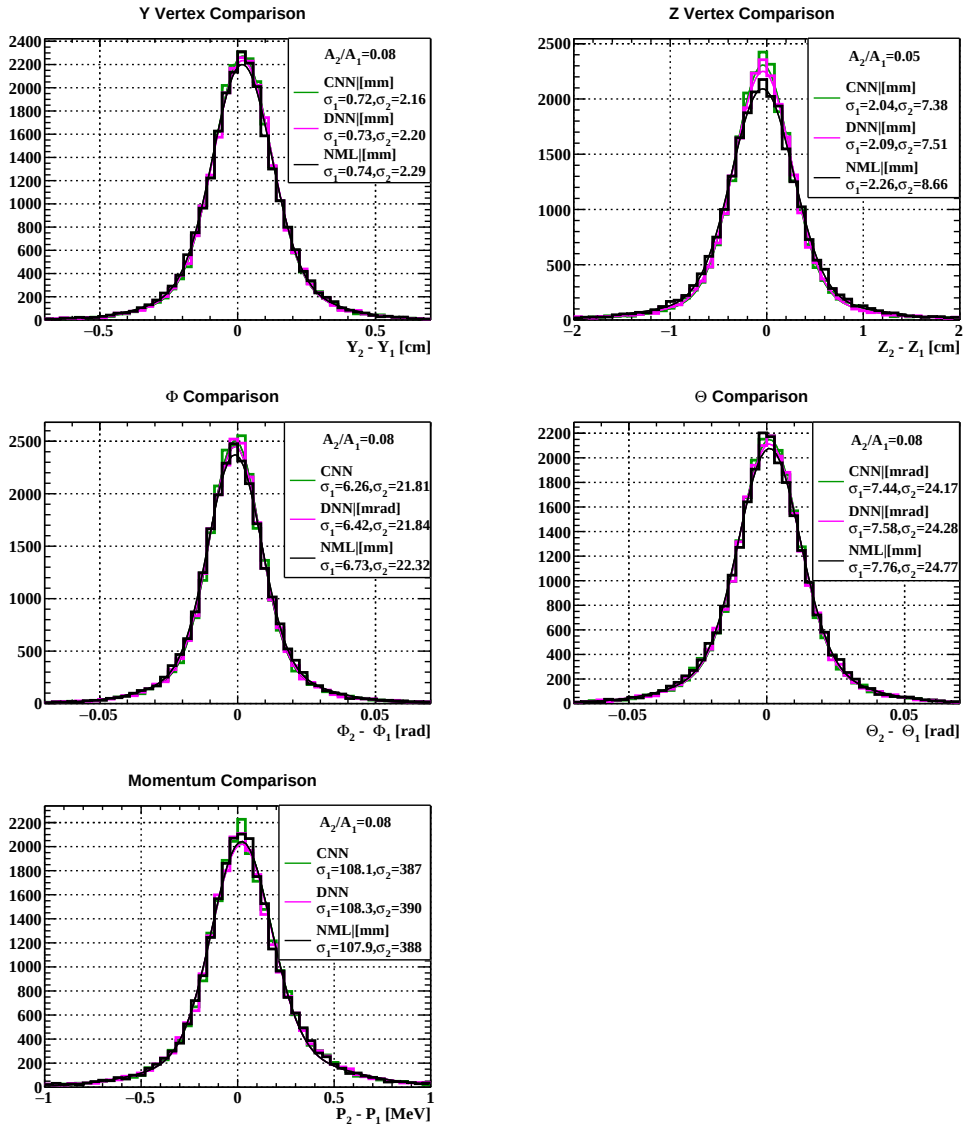


Figure B.3: Here, we show the kinematic comparison at a common plane between two independently fit/measured turns of a single two-turn track. In this plot, the conventional TXY function has been modified to suppress the DOCA bias; transforming $t_{drift} \rightarrow t_{drift} - 2ns$. The NML case represents the standard or conventional DOCA estimation, DNN represents the dense neural network, and CNN represents the convolutional neural network that inputs waveform voltages.