

5-2023

Pion Detection for the MOLLER Parity-Violating Electron Scattering Experiment

Michael Tristan Hurst
William & Mary

Follow this and additional works at: <https://scholarworks.wm.edu/honorstheses>



Part of the [Other Physics Commons](#)

Recommended Citation

Hurst, Michael Tristan, "Pion Detection for the MOLLER Parity-Violating Electron Scattering Experiment" (2023). *Undergraduate Honors Theses*. William & Mary. Paper 1972.
<https://scholarworks.wm.edu/honorstheses/1972>

This Honors Thesis -- Open Access is brought to you for free and open access by the Theses, Dissertations, & Master Projects at W&M ScholarWorks. It has been accepted for inclusion in Undergraduate Honors Theses by an authorized administrator of W&M ScholarWorks. For more information, please contact scholarworks@wm.edu.

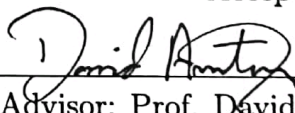
Pion Detection for the MOLLER Parity-Violating Electron Scattering Experiment

A thesis submitted in partial fulfillment of the requirement
for the degree of Bachelor of Science with Honors in
Physics from the College of William and Mary in Virginia,

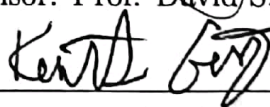
by

Michael Tristan Hurst

Accepted for Honors



Advisor: Prof. David S. Armstrong



Prof. Keith Griffioen



Prof. Todd Averett



Prof. Chad Vance

Williamsburg, Virginia
May 3, 2023

Contents

Acknowledgments	iii
List of Figures	iv
List of Tables	v
Abstract	v
1 Introduction	1
1.1 The Goal of the Experiment	1
1.2 Theory	2
1.3 Pion Production	5
2 MOLLER Experiment Detectors	7
2.1 Main Detector	7
2.2 Shower Max	8
2.3 Pion Detector	9
2.4 Expected Signal in Each Detector	10
3 Methods and Data Analysis	13
3.1 Detector Signals in Simulation	13
3.2 Binary Classification	14

4	Results and Conclusions	16
4.1	Results	16
4.1.1	Signal Separation	16
4.1.2	Pion Exit Scintillator	17
4.2	Classification	21
4.2.1	Data Structure	21
4.2.2	Performance of Classifiers	22
4.3	Plans for Future Work	27
4.4	Conclusion	28
A	Scripts for Data Analysis	29
	References	31

Acknowledgments

I would first like to thank my advisor, Dr. David Armstrong for his guidance and mentorship. I am grateful for the time he spent assisting me with editing, and the invaluable instruction he gave me throughout my research.

I'd like to thank Kate Evans for her time and all of the ROOT files she generated as I learned the ropes of the software.

I'd like to thank Zeke Wertz for his perspective on my work along the way.

Finally, I'd like to thank my friends and family for their support and encouragement throughout my work.

List of Figures

1.1	Tree Level Møller Scattering.	2
1.2	Right and Left Handedness.	2
1.3	Diagrams of Pion Production.	6
2.1	Main Detector System.	8
2.2	Distribution of Events on the Main Detector.	9
2.3	An Individual Shower Max Detector	10
2.4	View of Pion Detector	11
3.1	Random Forest Classifier.	15
4.1	Shower Max Signal.	18
4.2	Pion Detector Signal.	19
4.3	Pion Generated Shower Max and Pion Detector Coincidence.	21
4.4	Møller Generated Shower Max and Pion Detector Coincidence.	22
4.5	Width Optimization of Pion Exit Scintillator.	23
4.6	Optimizing Random Forest Classifier.	24
4.7	Confusion Matrix.	25
4.8	Random Forest Classifier	26
4.9	Random Forest Classifier with Rate Weighting	27
A.1	Classification of Entire Data Set.	30

List of Tables

4.1	Data Structure for Classification	22
-----	---	----

Abstract

The MOLLER Experiment at Jefferson Lab intends to make a precise measurement of the weak charge of the electron through parity-violating electron scattering. To achieve the level of precision required for the experiment, background rates of events other than electron-electron scattering must be known. Working with data from Monte-Carlo simulations created using a GEANT4 simulation package, I show that the combined signal from two existing detector subsystems of the MOLLER experiment allow for particle identification between electron and pion events. I worked to optimize an additional ‘Pion Exit Scintillator’ which improves the ability to distinguish particle identity at the cost of a large fraction of pion events. This identification ability is used to develop a machine learning based algorithm which is intended for use in the experimental determination of pion scattering asymmetry and the dilution of the electron signal in the main detector in counting mode measurements. The best trained classifier correctly classifies 95.1% of events.

Chapter 1

Introduction

The MOLLER Experiment (**M**easurement **O**f a **L**epton **L**epton **E**lectroweak **R**eaction) is an upcoming collaboration at Jefferson Lab Hall A using the upgraded 11 GeV electron beam to perform a high precision measurement of the asymmetry of longitudinally polarized electrons scattered by the electroweak interaction off of unpolarized electrons. The goal of my research is to perform background signal analysis of the asymmetry anticipated to be generated by one type of particle other than electrons, namely pions, that will need to be accounted for to achieve the unprecedented precision of the measurement.

1.1 The Goal of the Experiment

The MOLLER Experiment will make the most precise measurement of the weak charge of the electron that has ever been achieved. The last experiment that measured the weak charge of the electron through parity violating lepton-lepton scattering was the E158 experiment at the SLAC National Accelerator Laboratory. The E158 experiment made a measurement of the asymmetry of the scattered electrons with a precision of 17 ppb[1]. The anticipated asymmetry in the kinematic region MOLLER will operate is 33 ppb. MOLLER intends to improve upon the precision of the E158 by an order of magnitude, with an experimental goal of a precision of 0.8 ppb [2].

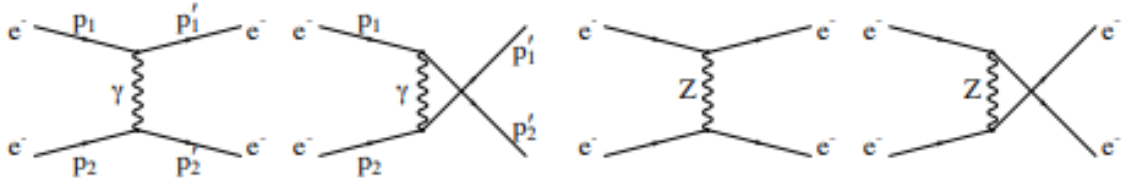


Figure 1.1: **Tree Level Møller Scattering Feynman Diagrams** The two diagrams on the left are of electromagnetic interaction. The two diagrams on the right are of weak nuclear interactions. Seen in [3].

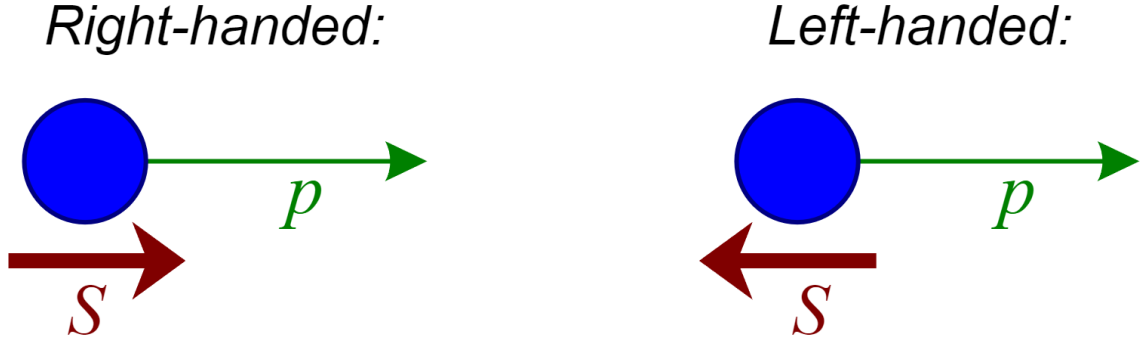


Figure 1.2: **Polarization of Particles** S is the spin of the particle. p is momentum of the particle.

1.2 Theory

The MOLLER Experiment will take place at Thomas Jefferson National Accelerator Facility (JLab) in Hall A. The experiment is a scattering experiment in which high energy, longitudinally polarized electrons are scattered off of unpolarized electrons in a liquid hydrogen target. The polarization of the electrons will be reversed very rapidly (\approx every millisecond) during the data taking. When the electrons are polarized along the direction of their momentum they are said to have a positive, or right-handed, helicity. When the polarization of the electrons is along the direction opposite of their momentum they are said to have a negative, or left-handed, helicity. These two cases can be seen in Fig. 1.2.

Electrons which are incident upon the liquid Hydrogen target will predominantly scatter through electromagnetic processes off of the protons and electrons of the Hydrogen atoms. The electromagnetic force acts the same for both left-handed and right-handed helicity particles, and thus the electromagnetic interactions should be identical for each electron beam polarization. This experiment is concerned with electrons that scatter off of other electrons, so called Møller scattering, named after the Danish physicist Christian Møller. Møller scattering is dominated by the electromagnetic process, seen in the two figures on the left of 1.1, but there is a very low probability of the two electrons scattering through the exchange of a Z^0 boson through the neutral current weak force. The weak force is the only fundamental force of nature that treats left and right handed helicity particles differently. Due to this different treatment, left-handed helicity electrons will scatter ever so slightly more than right-handed helicity electrons. The fractional value of the difference between the positive and negative helicity electrons is called the parity-violating asymmetry.

The asymmetry of the electroweak interaction between electrons is parameterized by the weak charge of the electron,

$$A_{PV} = \frac{\sigma_R - \sigma_L}{\sigma_R + \sigma_L} = CG_F Q_W^e \quad (1.1)$$

where A_{PV} is the parity-violating asymmetry of the electron interaction, σ_R is total cross section for scattering of right-handed helicity incident electrons, σ_L is the total cross section for scattering of left-handed helicity incident electrons, C is a factor dependent upon the kinematic and energetic constraints of the interaction, G_F is the Fermi Coupling constant, and Q_W^e is the weak charge of the electron [3]. Importantly, A_{PV} , which can be obtained through experiment, is directly proportional to Q_W^e .

The weak charge of the electron is a number which indicates how strongly the electron interacts with the vector part of the weak nuclear force. At tree level, assuming

the validity of the Standard Model of Particle Physics, Q_W^e is given by

$$Q_W^e = 1 - 4 \sin^2(\theta_W) \quad (1.2)$$

where θ_W is the electroweak mixing angle (equation from [3]).

The anticipated asymmetry is expected to be ≈ 33 parts per billion (ppb) [3]. To make the measurement of the asymmetry, data will be taken in two modes. The first phase will be in counting (or tracking) mode and the second will be in current (or integrating) mode. The counting mode measurements are made using very low beam current (as low as 10 pA, [2]) and enable characterization of the relative rates of signal and background events by identification of individual scattering events. Current mode will be the method of data taking from which the overall asymmetry will be calculated, and will run at 65 μA of beam current. In current mode, the helicity of the electrons will be rapidly flipped at a rate of 1.92 kHz [2]. Counting mode will be used to determine the fractional composition of particles at different detector systems. Counting mode can be thought of as a calibration mode that allows a measurement of the baseline amounts and kinematics of particles passing through detector systems.

The current mode measurements of the A_{PV} will produce approximately 134 billion Møller electron events per second [2]. Due to the enormous amount of events per second, determining Møller electron events from other background events in current mode is simply impossible. There will be an irreducible background signal from particles other than Møller electrons which could muddy the asymmetry of the scattered electrons. Background will be coming primarily from electron-proton elastic and inelastic scattering. The kinematics of these reactions are quite different and result in a distribution quite different from electron-electron events. One of the predominant particles which will be producing a background signal within the electron-electron Møller region is pions.

1.3 Pion Production

The incoming beam consists of electrons only from the Continuous Electron Beam Facility (CEBAF) at Jefferson Lab (JLab). Upon reaching the target, pions (π^+ , π^- , or π^0) may be produced through electron interactions with the nuclei of the Liquid Hydrogen or with the aluminum walls of target chamber. Pions will only be produced through electron-nuclei interaction. These interactions are very difficult to model. Most of the time, the electron will excite the nucleus to an excited state. Most of these nuclear excited states will decay through the strong nuclear force and potentially produce pions. The strong force is symmetric under parity reversal, and thus pions produced through strong interactions will not contribute to A_{PV} . Far fewer of these excited nuclear states will decay through the weak nuclear force emitting pions with an asymmetry which will be correlated to the polarization of the incoming electron. Some of these production methods can be seen in figure 1.3

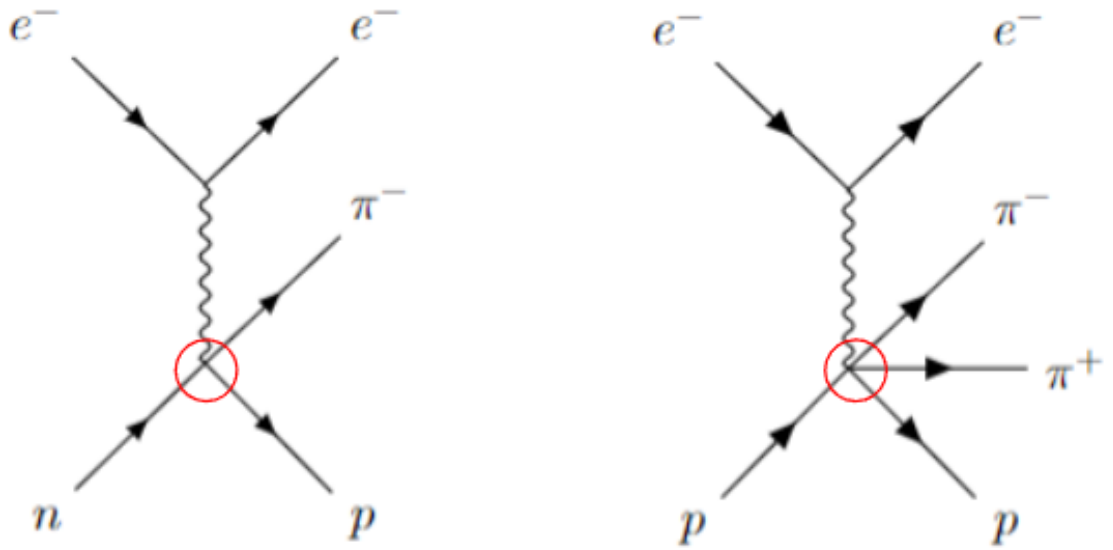


Figure 1.3: **Diagrams of Pion Production** These are some of the potential Feynman diagrams describing pion production. On the left is an event in which an incoming electron interacts with a neutron in the aluminum walls. On the right is an event in which an incoming electron interacts with a proton in either the hydrogen target or the aluminum walls. The red indicates a vertex that would include a weak decay of an excited nucleon state.

Chapter 2

MOLLER Experiment Detectors

2.1 Main Detector

The Main Detector system consists of six rings of fused silica (quartz) detectors. One section of the main detector can be seen on the right of figure 2.1, with ring one on the bottom right and ring six on the upper left. During data taking, scattered particles will be guided by magnets into the main detector. A charged particle that is incident upon the detector system generates Cerenkov light as the particles pass through at a velocity higher than the speed of light in the medium. Due to the extremely high rate that the measurement will occur at, signal from each of the six rings of the main detector will be integrated during each helicity window. The radial distribution of particles at the main detector can be seen in 2.2. This integrated rate yields σ_L , the signal for left handed scattering events, and σ_R , the signal for right handed scattering events. From these measurements of σ_L and σ_R , equation 1.1 is used to determine the A_{PV} of a given combination of right and left handed helicity windows.

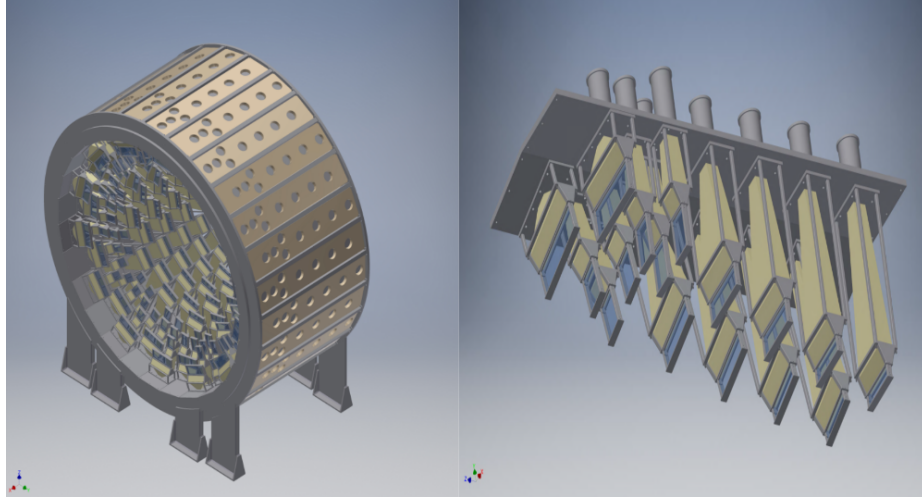


Figure 2.1: **CAD models of Main Detector System** Left: Model of the main detector system. The beam is incident from the bottom left. Right: Individual section of the main detector. Quartz detector faces are shown in blue. The light guides are shown in yellow. PMT housing is shown in gray. [2]

2.2 Shower Max

The Shower Max detector system is an electromagnetic calorimeter intended to measure electron flux. The SHMX detector system consists of 28 sections an entire circular region immediately behind the main detector Ring five. The system has full azimuthal acceptance and each segment consists of four interleaved layers of tungsten and fused silica quartz (see 2.3). Electrons undergo bremsstrahlung (braking radiation) losing energy as they pass through the dense tungsten. This creates a shower of lower energy gamma rays and positron electron pairs in the process. These lower energy particles can then undergo the process all over again. Pions incident upon the SHMX will lose very little energy and for the most part pass directly through the SHMX. When high energy particles pass through the quartz layers of the SHMX, they will release Cerenkov radiation. This radiation is guided into a Photo Multiplier Tube (PMT) for readout.

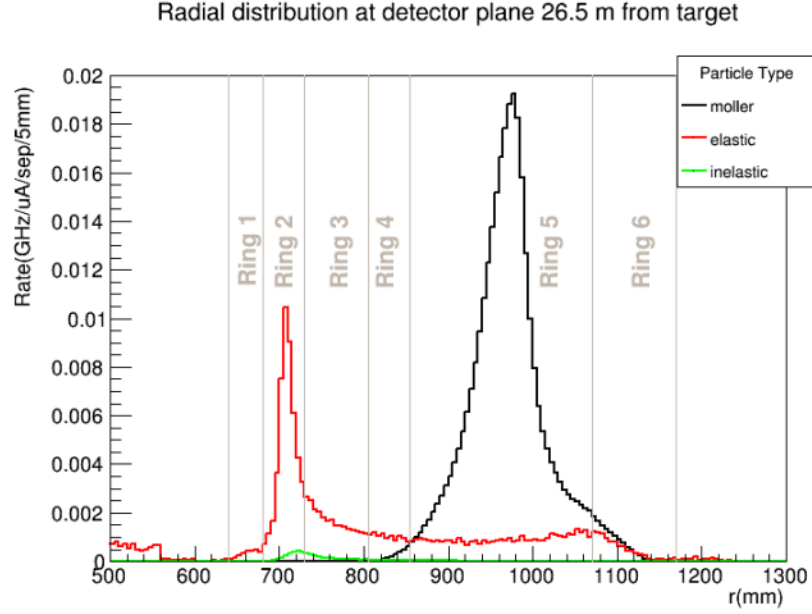


Figure 2.2: **Simulated distribution of events on the Main Detector System** Radial extent of Møller scattered events (black), elastic electron-proton events (red), and inelastic electron-proton events (green). [2]

2.3 Pion Detector

The Pion detector system lies downstream of the main detector and consists of 28 acrylic Cerenkov radiation bars situated inside of a thick lead torus. The radiator bars are 2.54 cm x 23 cm x 7 cm. The lead torus is an annulus with 90 cm inner radius, 115 cm outer radius, and 15 cm in length [2]. The lead torus is intended to shield the pion detectors from the Møller scattered electrons thus increasing the relative amount of signal generated in the pion detectors from pions and pion decay products such as muons (99 % of the time pions will decay into muons and neutrinos [4]).

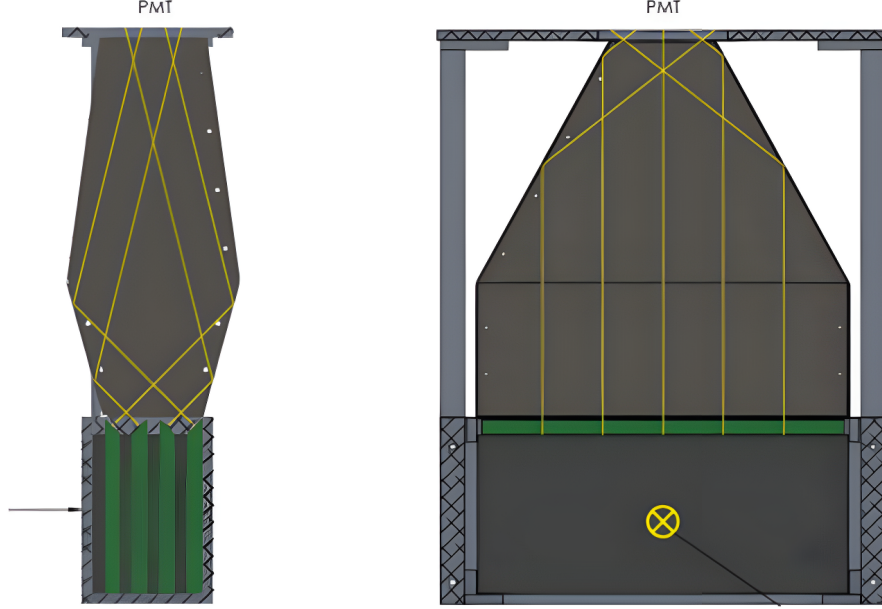


Figure 2.3: **CAD models of a individual Shower Max Detector.** Left: A side on view of a Shower Max detector, with the four layers of tungsten in gray and the four layers of quartz in green. The beam is coming from the left. Right: A front on view of a Shower Max detector, with the beam going into the page. [2, pg. 61]

2.4 Expected Signal in Each Detector

Because the particle detectors are at different locations along the beamline, the fractional composition of particles which produce a signal in each detector will differ.

The following equations are a simplified description of the amount of signal that will be seen by each of the detector systems

$$A_{PV}^{SHMX} = A_{PV}^{Mol} f_{Mol}^{SHMX} + A_{PV}^{\pi} f_{\pi}^{SHMX} \quad (2.1)$$

$$A_{PV}^{PD} = A_{PV}^{Mol} f_{Mol}^{PD} + A_{PV}^{\pi} f_{\pi}^{PD} \quad (2.2)$$

where $A_{PV}^{SHMX/PD}$ is the total asymmetry seen by the SHMX detectors in the main detector and Pion Detectors respectively , $A_{PV}^{Mol/\pi}$ is the asymmetry of the Møller

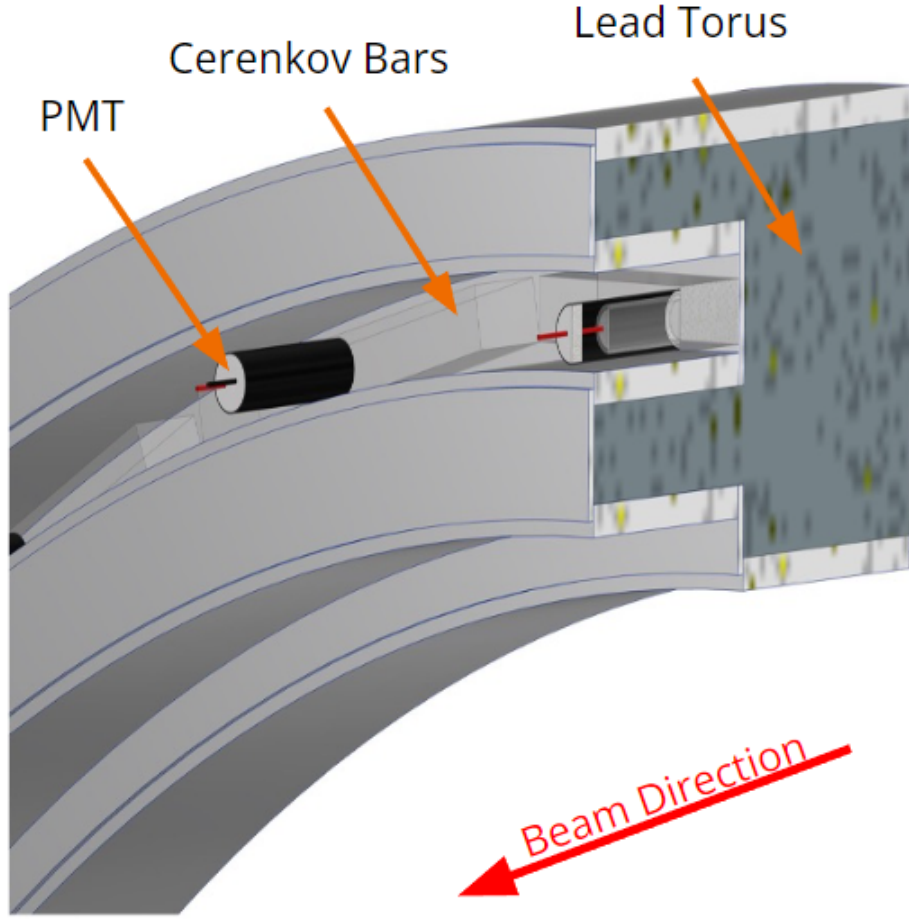


Figure 2.4: **CAD model of Pion Detectors.** A slice of the Pion Detectors situated within the lead torus. The long Cerenkov Bars and the PMTs which are attached to each bar are housed inside of and shielded by the lead. The beam is coming out of the page from the right.

electrons and pions respectively, $f_{Mol/\pi}^{SHMX}$ is the fraction of the signal in SHMX that is generated by Møller electrons and pions respectively, and $f_{Mol/\pi}^{PD}$ is the fraction of signal in the Pion Detector that is generated by Møller electrons and pions respectively.

The A_{PV} in equation 1.1 is only the asymmetry from Møller electrons. In counting mode measurements, the Pion Detector system is used to determine the amount of

pions present in the beam which is diluting the Møller signal in the main detector. Because the pion detector is situated within the lead, electrons will be suppressed by a factor of more than 10^3 within the Pion detector signal. However, the initial π^-/e^- ratio is expected to be 10^{-3} [2]. Thus, an approximately equal amount of pions and Møller electrons will reach the Pion detectors. Individual scattering events need to be classified as either Møller scattered electrons or pions to be able to obtain the fractional components of each detector system's signal. To determine the identity of individual scattering events in counting mode, I have examined the effect of an additional particle detector located behind the Pion Detectors.

In current mode, the measured signal from the detector systems will be the left side of equations 2.1 and 2.2. If the fractional components of each signal are known from counting mode, and the left hand side of each equation is known from current mode, and we utilize the A_{PV}^{Mol} from the main detector, then the pion detectors can be used to determine the A_{PV}^{π} in current mode. This allows for an experimental determination of the asymmetry of the pions, which can then be used to remove the asymmetry in the main detector caused by pions to isolate the parity-violating asymmetry of Møller scattered electrons. This hinges on the ability to distinguish between pions and electrons in counting mode.

Chapter 3

Methods and Data Analysis

3.1 Detector Signals in Simulation

As the detector systems have not yet been built for the MOLLER experiment, I have analyzed GEANT-4 Monte-Carlo simulations of Møller scattered electrons, “moll events”, and scattered pions, “pion events”, generated in the MOLLER experiments implementation titled Remoll [5]. This data analysis is done with the Root data analysis framework by CERN [6]. The moll events were generated from 500,000 simulated Møller scattered electrons and the pion events were generated from 2.5 million scattered pions.

The two detector systems of concern to me are the Shower Max and the Pion Detectors. At the beginning of my analysis, the Shower Max PMT system was not fully implemented into the geometry of the simulation. To approximate the Shower Max’s PMT signal for any given event, I added up all of the charged particles that made contact with a layer of quartz inside the Shower Max layers. Each charged particle that passes through the quartz at higher than the speed of light in the material will give off Cerenkov light, so the number of charged particles passing through the

layers of quartz we assume is proportional to the Shower Max PMT output.

The Pion Detector PMT subsystem is implemented in the simulation, so to read out the signal of the detector from the simulation I added up all the optical photons that contact the front of the Pion Detector PMTs.

An important note is that every event generated by the simulation includes a number called the rate. This is a number in $\frac{Hz}{\mu A}$ which means the number of this event per second per μA of beam current. It is effectively a measure of how likely a given event is to occur.

3.2 Binary Classification

To determine the fractional rate of the pion flux and the moller electron flux in 2.1 and 2.2, it is necessary to somehow label a given event as being generated by a pion or an electron. When working with simulations of the experiment, the type of particle which generated a given output from a detector system is known. In the experiment however, the only information that is accessible is the detector system outputs. To accomplish classification then, we must decide what combination of detector signals are indicative of a Møller or a Pion event. This kind of sorting into two categories is a well known problem in computer science known as a Binary Classification.

The classifier which I have investigated is a Random Forest Classifier from SciKit Learn [7]. It offered a wide enough versatility and ability to avoid overfitting to the training set which was attractive for a proof of concept classification.

A Random Forest Classifier is a type of machine learning classifier which utilizes the entire data set which it is given to train multiple decision trees. Each decision tree takes input data and then gives a classification based on the input data. The Random Forest model then averages the answers of all the decision trees to return

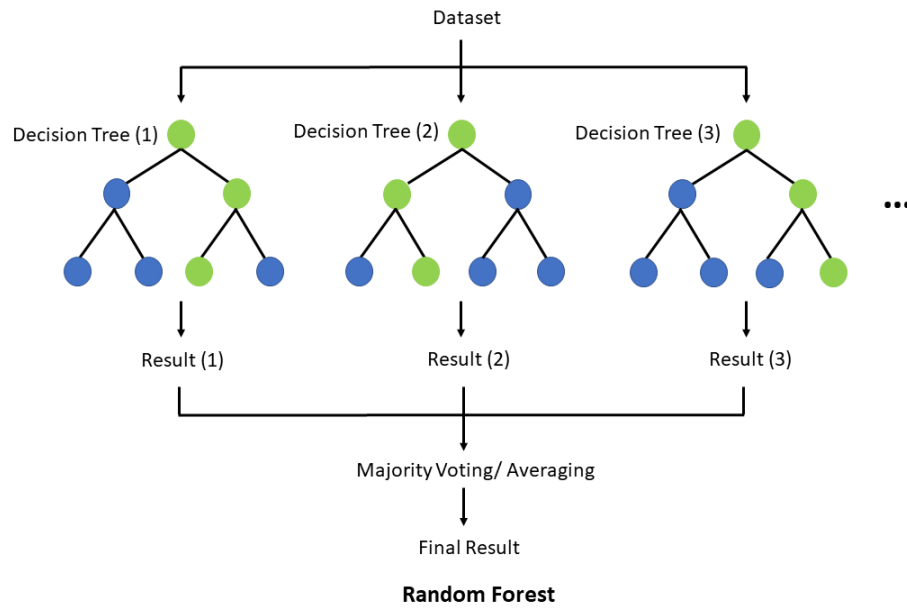


Figure 3.1: **Random Forest Classifier** A Random Forest Classifier creates many decision trees. Each tree may classify the events differently, but the Random Forest Classifier averages the results of every decision tree to output a percentage likelihood of an input event to belong to a certain category.

a probability of the event belonging to either of the possible binary outputs. When created, the Random Forest Classifier has random cutoffs in its decision trees. The classifier must be fitted to a particular data set. When it is fitted, a function is used to find optimal parameters for the cutoff values of all the decision trees so that the maximum number of events are correctly classified into their true class. For more information see the documentation at [7].

A common rudimentary analysis metric is the **score** of a given classifier. This is defined as the percentage of input data that the classifier correctly assigns to its true class.

Chapter 4

Results and Conclusions

4.1 Results

4.1.1 Signal Separation

The initial cuts I made to simulated data where to only include simulated events in which there was at least one charged particle in the SHMX and there was at least one optical photon on the pion PMT.

We hypothesized that the Shower Max will generate a large signal when Møller electrons pass through and generate a small signal when Pions pass through. Figure 4.1 shows that this is the case. Pions predominantly generate less than 20 charged particles in the SHMX while Møller events generate above 60 charged particles in the SHMX.

We hypothesized that the Pion Detector will generate a large signal when Pions pass through and generate a small signal when Møller electrons pass through. Figure 4.2 shows that this is the case. Pion events send around 500-600 optical photons in the pion PMT while the majority of moll events send fewer than 100.

These two signals alone already provide a great deal of separation between the

two types of particles. There is however a little overlap in the sensitivity of the two detectors, which means there are some areas in which the signals may be ambiguous. One potential way to remedy this overlapping region would be with a Pion Exit Scintillator. This is a hypothetical detector that would sit behind the Pion Detectors and the lead torus that would generate a signal any time a charged particle passed through it. The motivation behind this detector is that the pions that make a signal in the Pion Detector are expected lose very little energy in the lead torus, therefore they will produce light in the Pion Detectors then have enough energy left to generate signal in the Pion Exit Scintillator. Electrons which generate signal in the Pion Detectors will lose most of their energy to the lead torus and we expect these electrons to not make it to the Pion Exit Scintillator most of the time.

4.1.2 Pion Exit Scintillator

In simulation, I analyzed a sensitive detector region immediately behind the Pion Detectors. This was to approximate the effect a Pion Exit Scintillator could have on improving the ability of the SHMX and Pion Detectors to separate pion events from Møller events. This Pion Exit Scintillator is an annulus that is 40 mm thick in radius, centered directly behind the Pion Detector acrylic bars.

To quantify the effectiveness of the Pion Exit Scintillator in reducing the amount of signal from electrons observed by the pion detector system, I plotted the amount of charged particle hits on the SHMX detector layers vs the amount of optical photon hits on the pion detector photo-multiplier tube (PMT) as seen in Figs. 4.3 and 4.4. This is the top images in these figures (which is plotting one generator at a time from Figs. 4.1 and 4.2), while the bottom is the same with the added condition that the Pion Exit Scintillator had a charged particle pass through it. Notably, even with no additional detector, distinction between the two types of particles is possible with the

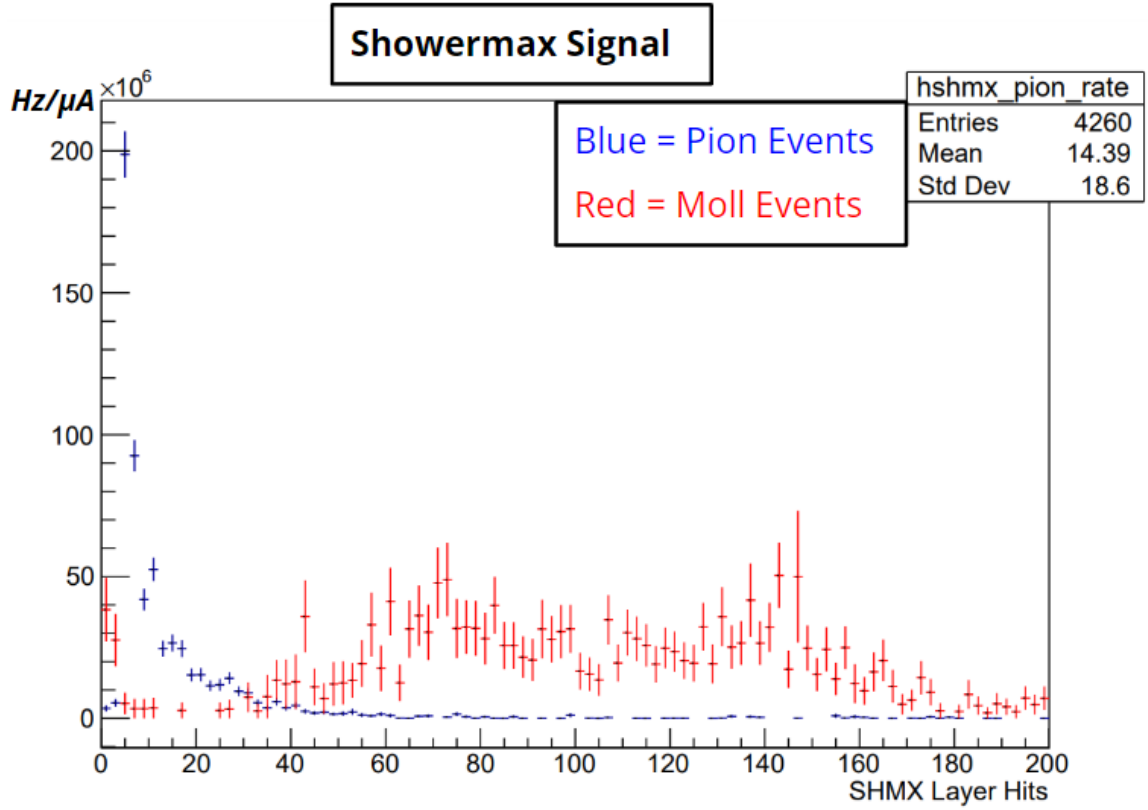


Figure 4.1: **Simulated Shower Max Signal** A histogram of the Shower Max signal in simulation for both Pion and Moll events. The x-axis is the number of charged particles that were seen in the layers of the SHMX. The y-axis is the rate in $\frac{\text{Hz}}{\mu\text{A}}$. Pion events are shown in blue, and moll events are shown in red.

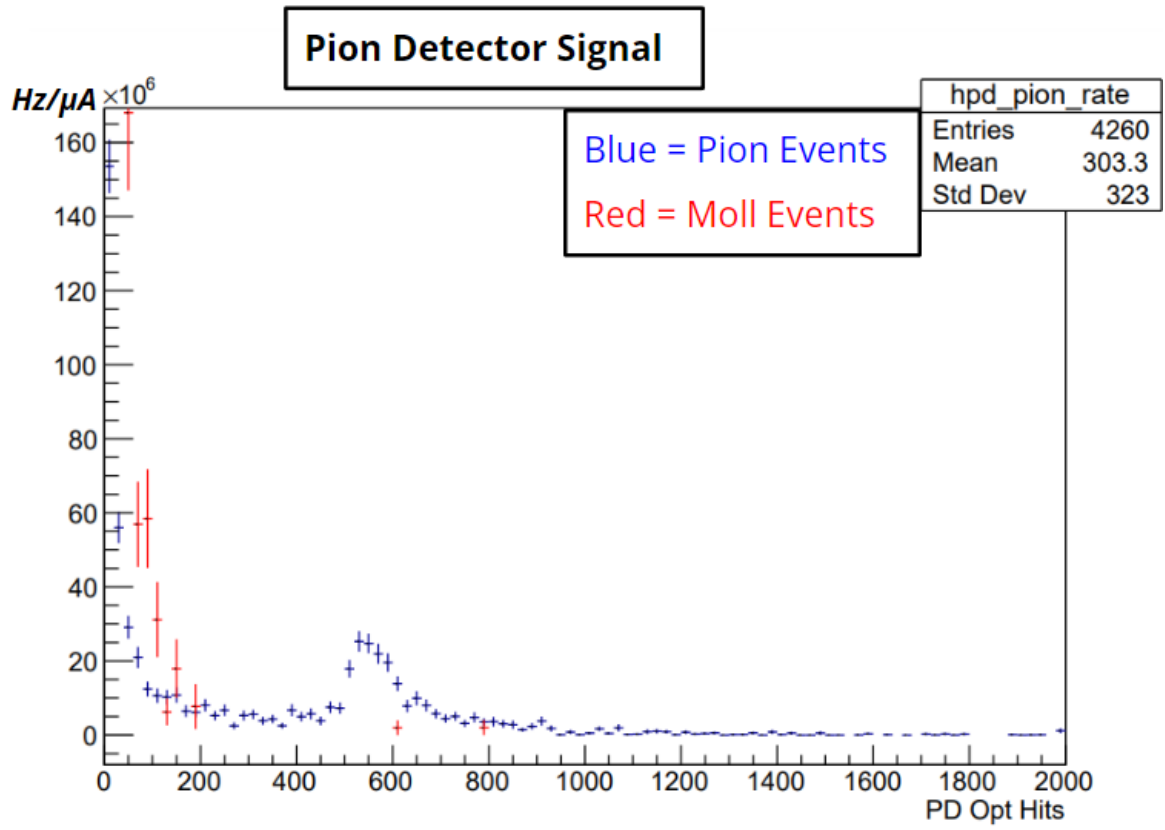


Figure 4.2: **Simulated Pion Detector Signal** A histogram of the Pion Detector signal in simulation for both Pion and Moll events. The x-axis is the number of optical photons that contacted the front of the Pion Detector PMT. The y-axis is the rate in $\frac{\text{Hz}}{\mu\text{A}}$. Pion events are shown in blue, and moll events are shown in red.

combined signal from both detector systems. There is however still some ambiguity and overlap in phase space between the Pion generated events in the upper plot in 4.3 and the Møller generated events in the upper plot in 4.4.

This additional exit scintillator triggering condition enhances the ability of the pion detector to separate electron signals from pion signals. Most pion events produce very little signal in the SHMX and a large amount signal in the Pion PMT. These results are promising as the additional requirement of the Pion Exit Scintillator firing removes some of the overlap between the signals seen by the SHMX and Pion PMT. As can be seen from the shapes of the signal distributions in figures 4.4 and 4.3, there is a combined signal response that is sensitive to pions and a region that is sensitive to Møller electrons. The Pion sensitive region is 0 to 30 charged particles seen by the SHMX and 0 to 500 optical photons seen by the Pion Detector PMT. The Moll sensitive region is 30 to 200 charged particles in the SHMX and 0 to 400 optical photons seen by the Pion Detector PMT. I used these regions to optimize a width for the Pion Exit Scintillator as seen in Fig. 4.5. I determined that a trigger scintillator of ≈ 40 mm was optimal to improve the Pion Exit Scintillator's enhancement of pion event classification.

One drawback of the Pion Exit Scintillator is the drop in statistics. If we were to utilize only those events in which the SHMX saw a signal, the Pion detector PMT saw light, AND the Pion Exit Scintillator saw a signal, we keep only 22% of Pion events (737 events/3343 events, 4.3). This is a very big drop in statistics for pion events, and ultimately means that the Pion Exit Scintillator likely will not be utilized for the small amount of additional distinguishing power they add is not worth the drop in statistics.

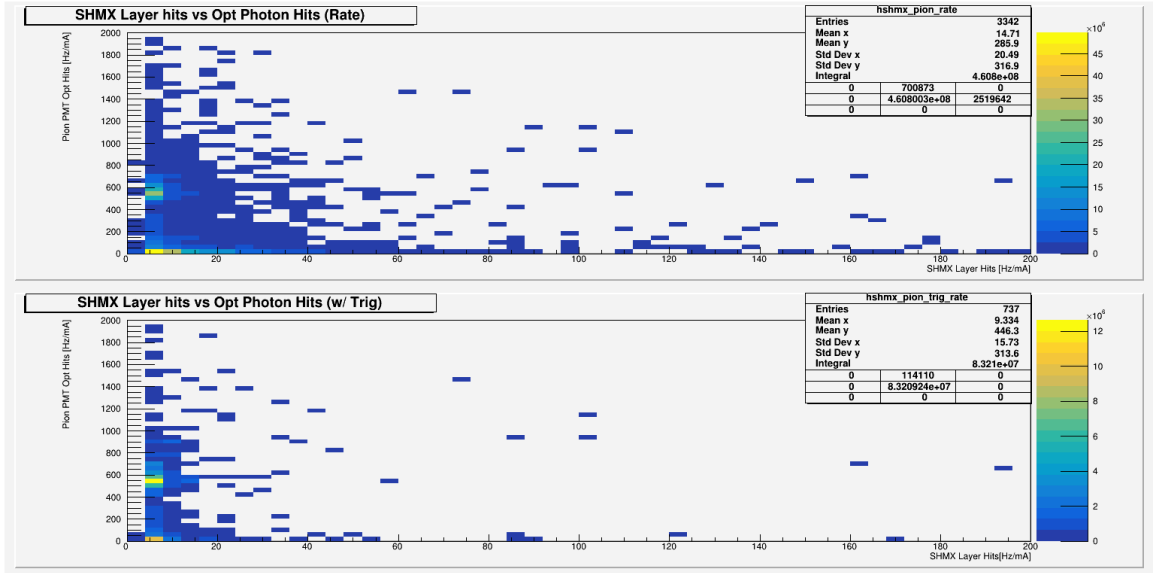


Figure 4.3: **Pion Scattering Events** The x-axis is the number of charged particle hits on the SHMX detector. The y-axis is the number of optical photons on the Pion detector PMT. The upper plot is for any scattered pion event in which there was a signal in the SHMX and Pion PMT detectors. The bottom plot is the same data with the condition that a charged particle hit the downstream trigger scintillator.

4.2 Classification

4.2.1 Data Structure

To make the simulation data intelligible to the SciKit Learn classification algorithms, I formatted a combination of simulation outputs into CSV format seen in Table 4.1. The first column is the label for a given event. The following columns have relevant observables that would be accessible during an experimental run.

To train the classifiers, simulation data CSV's were split into a training set and a testing set. The split chosen was 80% of data into training set with 20% reserved for testing.

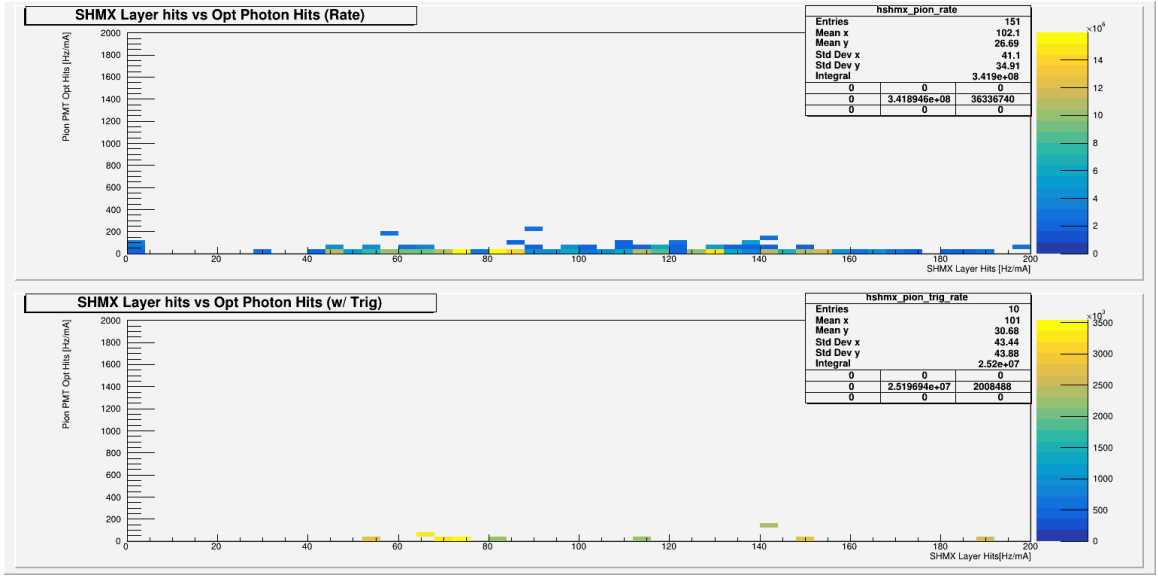


Figure 4.4: **Moll Scattering Events** Same style plots as in 4.3, only change is the all the scattering events are Moller Electron scattering. Far fewer electron events make it to the Pion detector and therefore the amount of hits seen in this plot is an order of magnitude less than in 4.3. As the events are from electrons, there is a high signal in the SHMX and a low signal in the Pion PMT.

Generator Type	SHMX Signal	PD Signal	Pion Detector Trigger	Rate [$\frac{Hz}{\mu A}$]
0 (Moller)	Integer	Integer	Boolean	Float
1 (Pion)	Integer	Integer	Boolean	Float

Table 4.1: Format for CSV files used for classifier training. Generator Type is 0 for Møller events and 1 for Pion events. SHMX signal is the x-axis of Fig.4.3 . PD Signal is the y-axis of Fig. 4.3. Pion Detector Trigger is a boolean for whether or not the Pion Exit Scintillator was triggered. The rate is the amount of times per second per μA of beam current that the particular event will occur.

4.2.2 Performance of Classifiers

One thing to consider when building a Random Forest Classifier is the number of trees in the forest. With fewer decision trees, training takes a shorter amount of time, but the classifiers performance suffers. From Fig. 4.6, I determined that 20 trees was a good balance between computational speed and the effectiveness of classification.

A fitted random forest classifier will take an input of SHMX signal and Pion Dete-

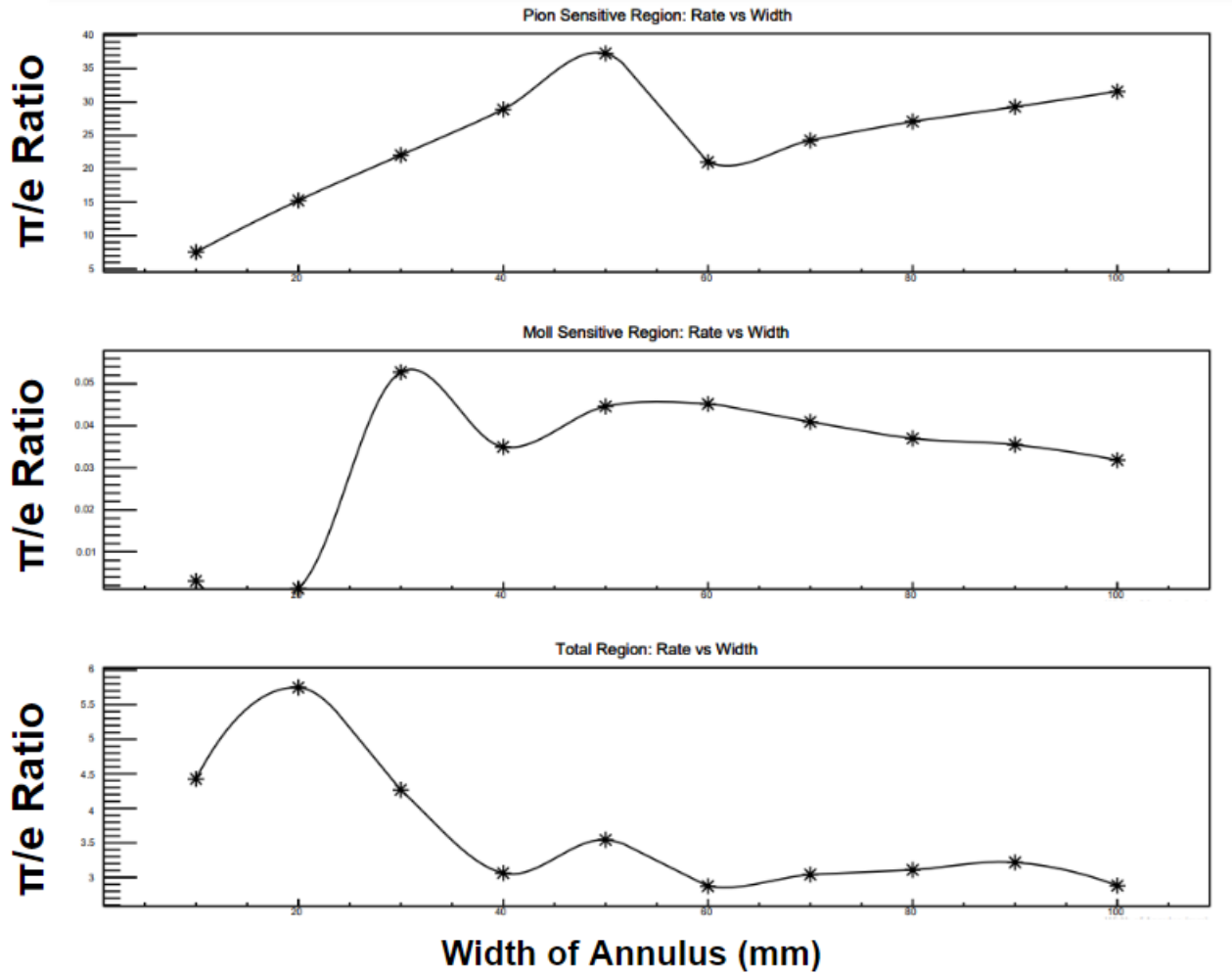


Figure 4.5: **Width Optimization of the Pion Exit Scintillator** Plots of the ratio of pion rate to Møller rate in three regions of the bottom plots from figures 4.3 and 4.4. The x-axis is the width of the annulus of the Pion Exit Scintillator. The π rate is the integrated rate in a given region of Fig. 4.3. The e rate is the integrated rate in a given region of 4.4. The y-axis is the π rate divided by the e rate. The bigger this ratio is the more the more signal the combined detectors get from pions.

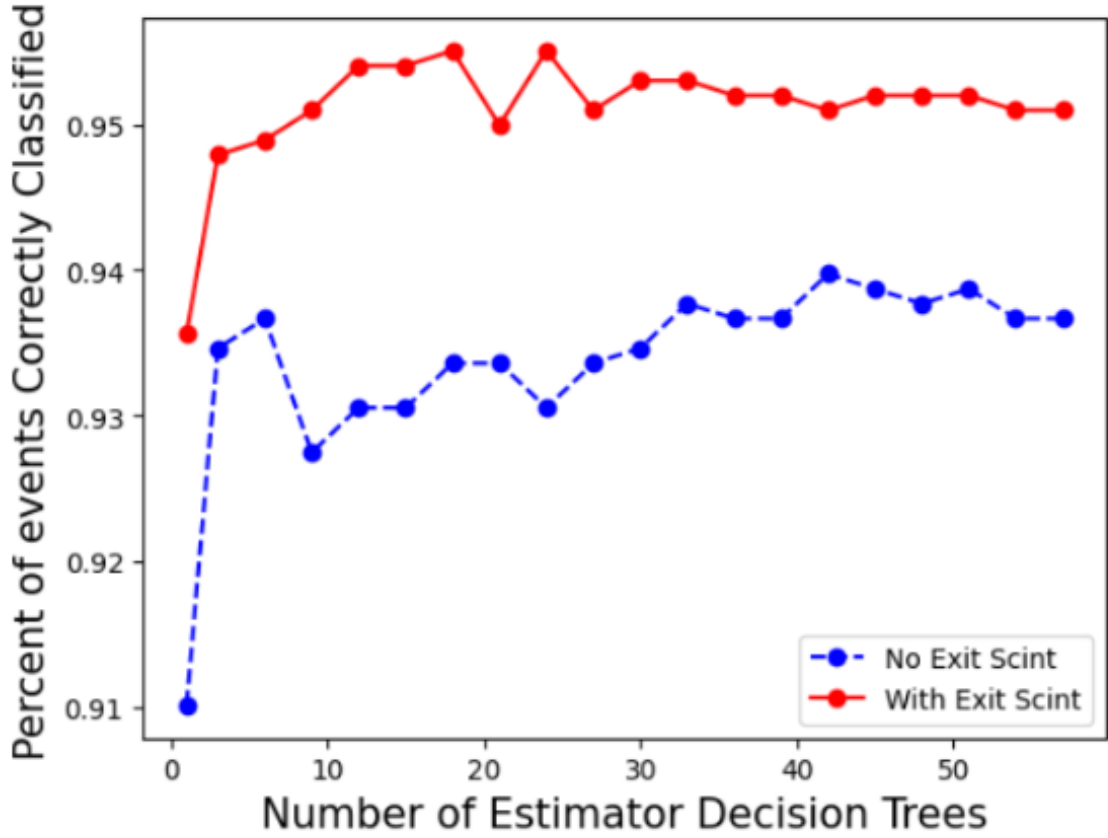


Figure 4.6: **Impact of Number of Trees on Classifier Performance.** Plot of number of decision trees in our given forest classifier versus the percentage of correctly classified events.

cotor signal for a given event and return the probability of the signals being generated by a Møller scattered electron and the probability of the signal being generated by a scattered pion. These probabilities plotted against all the events in the training and testing set can be seen in Figure. A.1.

A Confusion Matrix is a useful way of analyzing the performance of a particular classifier. The on-diagonal elements are correct predictions, while the off-diagonal elements are incorrect predictions 4.7. This type of plot is useful for determining the true positive, false positive, true negative, and false negative rate of a given classifier.

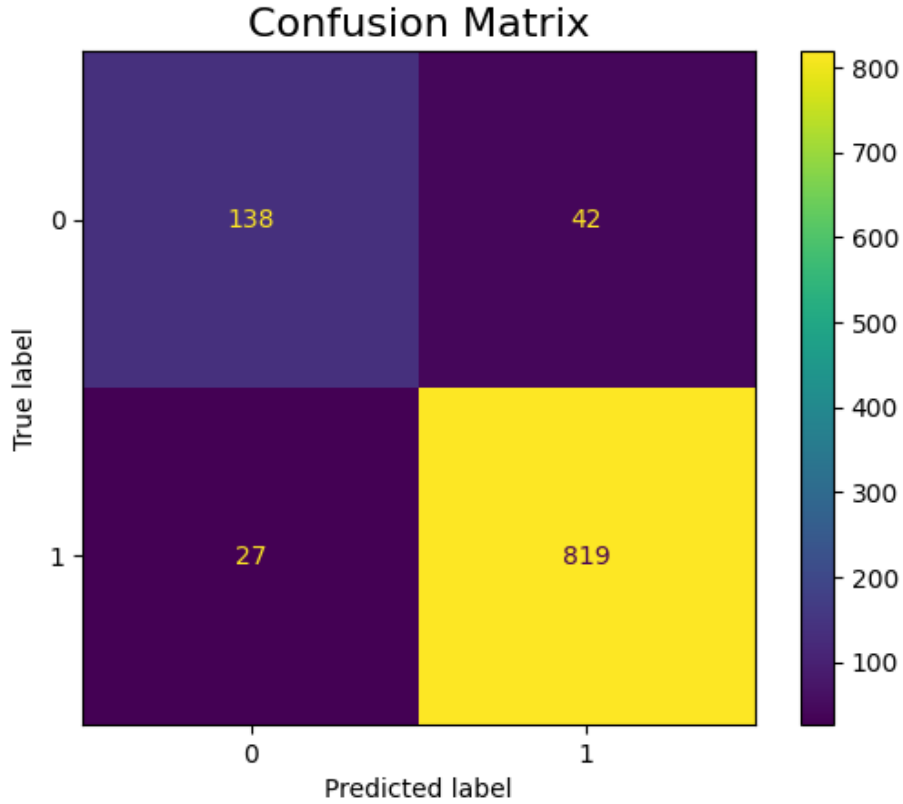


Figure 4.7: **Confusion Matrix for a Trained Classifier** The x-axis shows what label the trained model predicted for a certain event. The y-axis shows the true label for each event. 0 corresponds to Møller event and 1 corresponds to pion event. This is the confusion matrix for the classifier used in Fig. 4.8.

To look beyond just the numerical performance of any given classifier, it is useful to look at the boundaries that the classifier has developed between different regions of confidence. I have plotted the same data as in figures 4.3 and 4.4 and I have overlaid on top of those data points the classifiers confidence of classification in that region. The darker the colored region, the more confident the classifier is in that region. In figure 4.8, there is a single Møller event in the upper left that causes vast regions of the previously dark blue region to become much lighter. It is likely that this is representative of an over-fitting to the training data that the classifier was trained on. These are outlier events that are not as likely to occur.

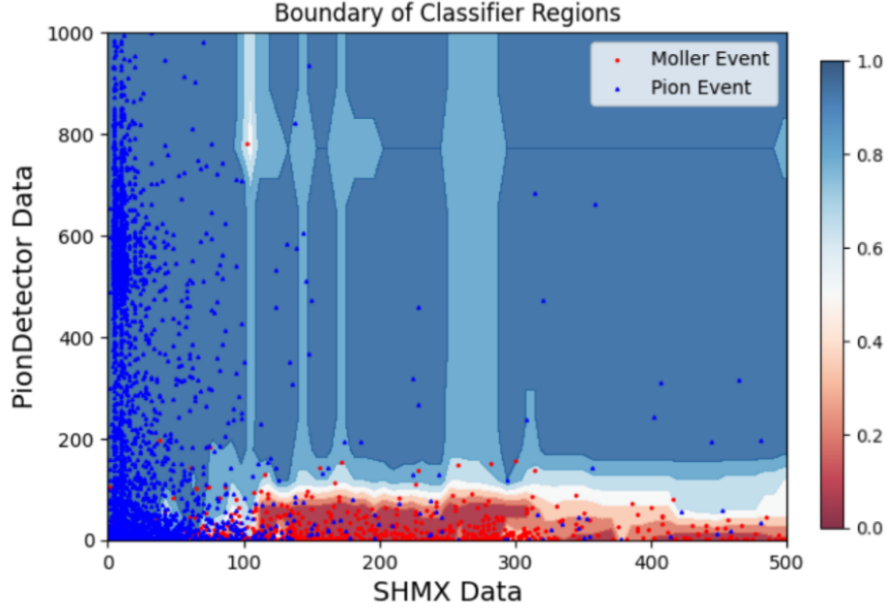


Figure 4.8: **Random Forest Classifier.** The gradient represents the model's confidence of prediction. Dark blue represents an area the model has high confidence that events are pion generated. Dark red represents an area the model has high confidence that events are Møller generated. Lighter colored regions are areas where the model is less confident. White areas are those where the model classifies events as equally likely to be a pion as an electron. Score = 93.2%

To remove some of this over-fitting behavior, it is possible to fit the model with the relative rate of each event occurring. This is the final column in my data structure (see Table 4.1). When the rate was accounted for in training of the classifier, the performance on the testing set was moderately improved (93.2% correct to 95.1% correct), but the area in which the model was less confident was decreased considerably.

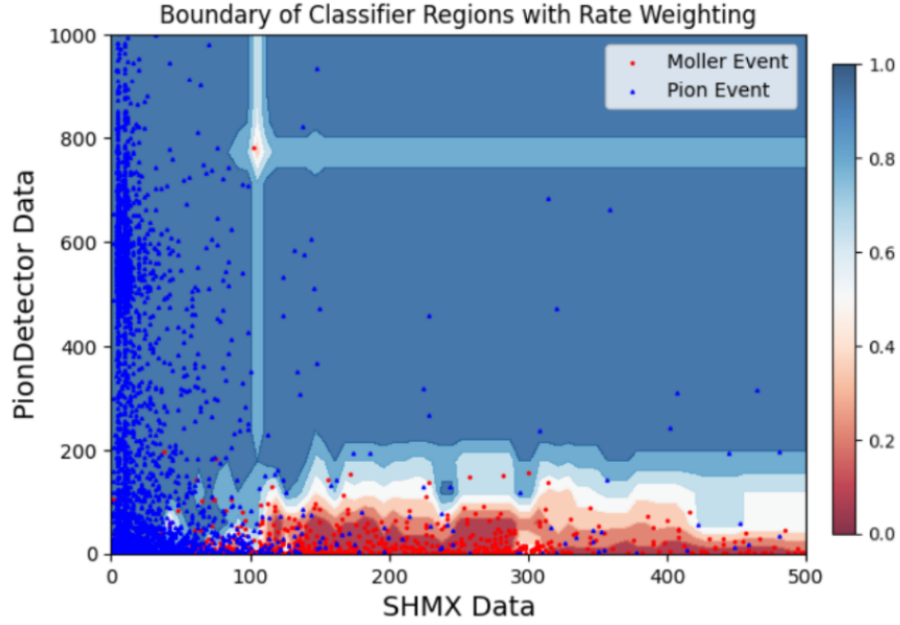


Figure 4.9: **Random Forest Classifier with Rate Weighting.** The gradient represents the model's confidence of prediction. Dark blue represents an area the model has high confidence that events are pion generated. Dark red represents an area the model has high confidence that events are Møller generated. Lighter colored regions are areas where the model is less confident. White areas are those where the model classifies events as equally likely to be a pion as an electron. Score = 95.1%

4.3 Plans for Future Work

In the future, it would be ideal to add more observables to see if there is a further improvement in the performance of classifier of these kinds. Further studies in this area could consist of fitting more kinds of classifiers with this data to see if these other classifiers could improve the classification performance along different metrics, such as purity (reducing false positives) or efficiency (reducing false negatives).

4.4 Conclusion

The Shower Max electromagnetic calorimeter and the Pion Detector subsystem signals can be combined to differentiate between pion generated events and Møller electron generated events. An additional Pion Exit Scintillator improves the distinguishing capabilities but unacceptably reduces the rate of pion events to be worth using. Machine Learning classification of Pion and Møller electron events is possible and would be beneficial for the MOLLER experimental run in determining the fractional components of the scattered beam composition during counting mode runs.

Appendix A

Scripts for Data Analysis

The Python classification algorithms and the ROOT data analysis can both be found at my Github page at <https://github.com/mt-hurst/HonorsThesis>. An important note is that in my Python code, the data is split randomly when executing the script, so the exact values of the performance of the classifiers can shift depending on the random splitting of the data.

Figures 4.1 and 4.2 were produced using *shmx_pion_separate.c*.

Figures 4.3 and 4.4 were produced using *shmx_pion_coinc.c*.

Figures 4.6, 4.7, 4.8, 4.9, and A.1 were created using *RateClassification.py*.

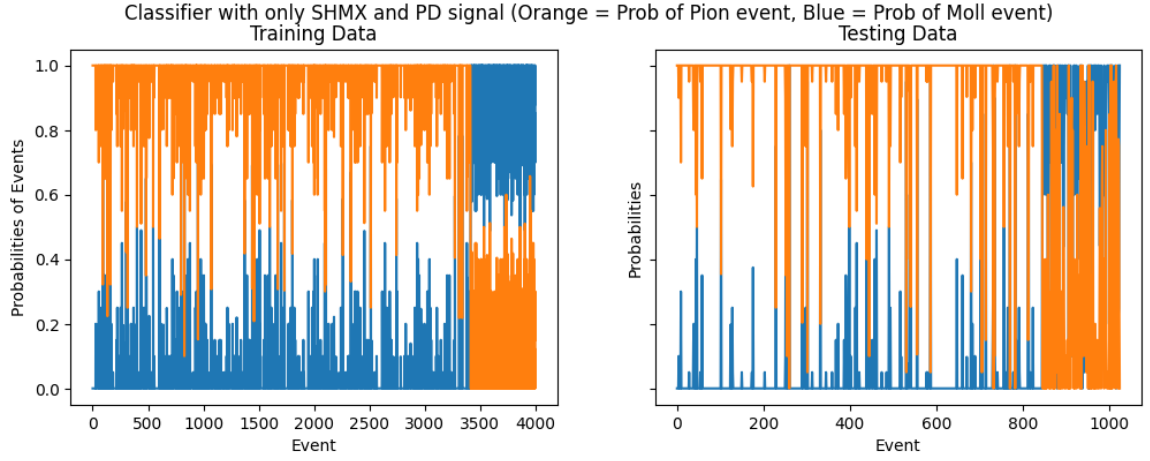


Figure A.1: **Classification of Full Data Set.** The left is the classifier's probabilistic output for each event on the training data. The right is the classifier's probabilistic output for each event on the testing data. For every event an orange and a blue point is plotted. The orange is the probability assigned to the event to be a Pion produced event and the blue is the probability assigned to be a Moll produced event. For these data sets, the first bunch of data are all pion events, and the tail end of the data is the moll events.

References

- [1] P. L. Anthony et al. “Precision Measurement of the Weak Mixing Angle in Møller Scattering”. In: *Physical Review Letters* 95.8 (Aug. 2005). DOI: [10 . 1103 / physrevlett . 95 . 081601](https://doi.org/10.1103/physrevlett.95.081601). URL: [https : // doi . org / 10 . 1103 % 5 C % 2 F physrevlett . 95 . 081601](https://doi.org/10.1103%5C%2Fphysrevlett.95.081601).
- [2] MOLLER Collaboration. *Conceptual Design Report*. 2019 (Revised 2020). URL: [https : // moller . jlab . org / DocDB / 0006 / 000630 / 002 / MOLLER _ CDR _ final _ aug _ 4 _ 2020 . pdf](https://moller.jlab.org/DocDB/0006/000630/002/MOLLER_CDR_final_aug_4_2020.pdf).
- [3] MOLLER Collaboration et al. *The MOLLER Experiment: An Ultra-Precise Measurement of the Weak Mixing Angle Using Møller Scattering*. 2014. arXiv: [1411 . 4088 \[nucl-ex\]](https://arxiv.org/abs/1411.4088).
- [4] R. L. Workman et al. “Review of Particle Physics”. In: *PTEP* 2022 (2022), p. 083C01. DOI: [10 . 1093 / ptep / ptac097](https://doi.org/10.1093/ptep/ptac097).
- [5] *The MOLLER Collaboration’s implementaion of GEANT-4*. URL: [https : // github . com / JeffersonLab / remoll](https://github.com/JeffersonLab/remoll).
- [6] *ROOT Reference Documentation*. URL: [https : // root . cern / doc / v626 /](https://root.cern/doc/v626/).
- [7] *ScikitLearn Documentation: Random Forest Classifier*. URL: [https : // scikit - learn . org / stable / modules / generated / sklearn . ensemble . RandomForestClassifier . html # sklearn . ensemble . RandomForestClassifier](https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html#sklearn.ensemble.RandomForestClassifier).