

Search for Low mass Higgs Portal Scalars at the MicroBooNE Detector in the NuMI Beam

Aditya Harish Bhanderi

Supervisor: Prof. Justin Evans

2022

A thesis submitted to the University of Manchester
for the degree of Doctor of Philosophy
in the Faculty of Science and Engineering



The University of Manchester

Department of Physics and Astronomy
School of Natural Sciences

Contents

Abstract	6
Declaration	7
Copyright Statement	8
Acknowledgements	9
1 Higgs Portal Scalar Search	13
1.1 Higgs Portal Model	13
1.1.1 Production	15
1.1.2 Decay	16
1.2 Higgs Portal Scalar Decay Length	18
1.3 Existing Limits on Higgs Portal Scalars	19
2 The MicroBooNE Experiment	24
2.1 Liquid Argon Time Projection Chamber	24
2.2 MicroBooNE Detector Coordinate System	26
2.3 MicroBooNE Time Projection Chamber	27
2.3.1 MicroBooNE Charge Collection System	29
2.3.2 MicroBooNE Light Collection System	30
2.4 MicroBooNE Cosmic Ray Tagger System	31
2.5 MicroBooNE Readout and Trigger System	32
2.5.1 MicroBooNE Beam Hardware Trigger	34
2.5.2 MicroBooNE Software Trigger	34
2.5.3 MicroBooNE External and Unbiased Trigger	35

2.6	MicroBooNE Event Display	35
3	The NuMI Beam and Flux at MicroBooNE	37
3.1	Neutrino Beam Production	38
3.1.1	Neutrino and Anti-neutrino Modes	41
3.1.2	MicroBooNE Proton Delivery Structure	41
3.2	The NuMI Beam Simulation	43
3.3	NuMI Neutrino Flux at MicroBooNE	44
3.3.1	NuMI Central-Value Flux Prediction	45
3.3.2	NuMI Central-Value Flux Prediction in Neutrino Energy and Angle	49
3.3.3	NuMI Central Value Flux Prediction by Neutrino Parent	50
3.4	Higgs Portal Scalar Flux at MicroBooNE	52
3.4.1	Flux prediction for KDAR from the NuMI Absorber	53
4	Particle Interactions in MicroBooNE	55
4.1	Neutrino Interactions in MicroBooNE	55
4.1.1	Quasi-Elastic Scattering	57
4.1.2	Meson Exchange Current	58
4.1.3	Resonant Scattering	59
4.1.4	Deep Inelastic Scattering	61
4.1.5	Coherent Scattering	61
4.2	Nuclear Effects	63
4.2.1	Initial State Effects	64
4.2.2	Multi-nucleon Correlation Effects	64
4.2.3	Final State Particle Interactions in the Argon	66
4.3	MicroBooNE Detector Response to the Passage of Particles	68
4.3.1	Heavy Charged Particle Interactions	68
4.3.2	Light Charged Particle Interactions	71
4.3.3	Neutral Particle Interactions	72
4.3.4	Electromagnetic Showers	73
4.3.5	Scintillation Light	74

4.4	Detector Effects	76
4.4.1	Diffusion	76
4.4.2	Space Charge Effects	78
4.4.3	Electron-Ion Recombination	78
4.4.4	Argon Purity	80
4.4.5	Rayleigh Scattering	81
4.4.6	Light Yield Attenuation	81
5	Simulation and Reconstruction	83
5.1	MicroBooNE Monte Carlo Generators	84
5.1.1	Higgs Portal Scalar Generator	84
5.1.2	Neutrino Generator	87
5.2	Particle Propagation in MicroBooNE	88
5.3	Cosmic Overlay	88
5.4	Optical Reconstruction	89
5.4.1	Optical Hits Reconstruction	89
5.4.2	Flash Reconstruction	90
5.5	TPC Signal Processing	91
5.5.1	Noise Filtering	92
5.5.2	2D Deconvolution	93
5.5.3	Region of Interest	93
5.5.4	Hit Finding	94
5.6	Pandora Reconstruction Framework	94
5.7	Calibration	97
5.7.1	Charge Calibration	98
5.7.2	Energy Calibration and Reconstruction	99
5.8	Samples	100
5.8.1	Beam-On Samples	101
5.8.2	Background samples	101
5.8.3	Background Sample Normalisation	104
5.8.4	Signal samples	111

6	Signal Pre-Selection	114
6.1	Decay Kinematics	114
6.1.1	Kaon Decay At Rest	115
6.1.2	Kaon Decay In Flight	116
6.1.3	Scalar Decay	117
6.1.4	Separation of Kaon Decay In Flight and Kaon Decay At Rest	122
6.2	Selection	123
6.2.1	Slice Identification	124
6.2.2	Cosmic Ray Tagging	125
6.2.3	Number of Contained Showers	126
6.2.4	Pandora Track-Shower Score	129
6.2.5	Pre-Selection Summary	129
6.3	Pre-Selection Efficiencies	131
7	Boosted Decision Trees	135
7.1	Boosted Decision Trees Definition	135
7.1.1	Selection of Variables in Decision Trees	137
7.1.2	Stopping Conditions of the Decision Trees	138
7.1.3	Boosting	139
7.2	Application to the Higgs Portal Scalar Search	140
7.2.1	Feature Importance	140
7.2.2	BDT Variables	142
7.3	BDT Performance	150
8	Systematic Uncertainties	155
8.1	Flux Uncertainties	157
8.1.1	Hadron Production Uncertainties	158
8.1.2	Beamline Geometry Uncertainties	158
8.1.3	Uncertainties due to Long-lived Kaons	159
8.1.4	Uncertainty due to KDAR from the NuMI Absorber	161
8.1.5	Total Flux Uncertainty on Signal and NuMI Overlay Background	161
8.2	Cross-section Uncertainties	163

8.2.1	Uncertainties due to CC and NC coherent pion production . .	164
8.2.2	Total Cross-section Uncertainty on the NuMI Overlay Back- ground	164
8.3	Re-interaction Uncertainties	165
8.4	Detector Uncertainties	167
8.4.1	Wire Modifications	167
8.4.2	Scintillation Light	169
8.4.3	Space Charge Effect and Recombination	170
8.4.4	Detector Variation Sample Statistics	170
8.4.5	Total Detector Uncertainty on the Signal and NuMI Overlay Background	171
8.5	Out-of-Cryostat Uncertainties	173
8.6	Statistical Uncertainties on the MC and Data Samples	173
8.7	Summary of Background and Signal Uncertainties	174
9	Limit Setting	178
9.1	Limit Extraction Procedure	178
9.1.1	The Statistical Test	179
9.1.2	Calculation of Confidence Levels	181
9.2	Results	183
9.2.1	BDT Score Selection	184
9.2.2	Application of the BDT Score Selection	185
9.2.3	Setting a Limit on the Higgs Portal Scalar Mixing Parameter	202
10	Conclusions	205

Abstract

The MicroBooNE experiment exploits Liquid Argon Time Projection Chamber (LArTPC) technology to detect neutrinos from two beams at Fermilab, Chicago: on-axis Booster Neutrino Beam (BNB) and the off-axis Neutrinos at the Main Injector (NuMI) beam. While the main aim of the experiment is to study the low-energy excess of electron neutrino events observed by the MicroBooNE experiment, which could be a signature of Beyond Standard Model (BSM) physics such as eV-scale neutrinos, there are many other models of BSM physics that the experiment can be exploited to probe. This thesis presents a BSM search for low-mass Higgs Portal Scalars at MicroBooNE in the NuMI beam using data collected during the Run 1 and Run 3 data-taking periods with a total exposure corresponding to 7.01×10^{20} protons on target (POT). The Higgs Portal Model is an extension to the Standard Model in which a dark-sector scalar, S , mixes with the Higgs boson with mixing angle θ and acquires a coupling to the Standard Model fermions via the Higgs Yukawa coupling. We search for low-mass Higgs Portal scalars, in the mass range $100 - 200$ MeV, at MicroBooNE via the production channel $K \rightarrow \pi + S$, where scalars are produced from kaons decaying at rest in the NuMI target and beam dump, as well as kaons decaying in flight in the decay pipe of the NuMI beam. In the MicroBooNE detector, we search for the decay channel $S \rightarrow e^+e^-$, which is the only decay channel available to the scalars in the mass range of our search. The results are expressed as limits, at the 95% confidence level, on the scalar-Higgs mixing angle for scalars in the mass range $100 - 200$ MeV.

Declaration

I hereby declare that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

1. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given the University of Manchester certain rights to use such Copyright, including for administrative purposes.
2. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.
3. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.
4. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see <http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=24420>), in any relevant Thesis restriction declarations deposited in the University Library, the University Library’s regulations (see <http://www.library.manchester.ac.uk/about/regulations/>) and in the University’s policy on Presentation of Theses.

Acknowledgements

Firstly, I would like to thank my supervisor, Prof. Justin Evans, for his academic guidance and constant support over the last four years and for his extensive help in reading through the drafts of this work. He is a great mentor and I'm grateful to have worked under his supervision. I would also like to thank my co-supervisor Stefan Söldner-Rembold for his advice throughout the PhD and for giving me an opportunity to do a PhD at Manchester.

My sincere thanks to Pawel Guzowski for his help with the Higgs portal scalar analysis. I can always count on him for quick help or guidance at any hour of the day. This PhD would have been difficult without his help and support. I'm very grateful to Davide Porzio, who is no longer with us, for mentoring me when I joined and helping me in understanding the LArSoft framework used at MicroBooNE and for amazing conversations at big data science workshop in Aachen Germany. I'm deeply grateful to Owen Goodwin for always quickly responding to all my queries related to the Higgs portal analysis and NuMI beam. I'm also thankful to Krishan Mistry for answering my questions related to NuMI beam and helping me produce samples used in the analysis. I would also like to thank Katrina Miller for her amazing explanation to all my questions related to the NuMI beam.

Thank you to Anyssa and Nicola for helping me with the presentations and providing me useful feedbacks. Thank you to David, Luis and Luciano for all the help with Pyhf software and for all the entertaining conversations. Thank you to Marina for all the help and for always being there during the course of the PhD. I would like to thank the entire Manchester neutrino group for making Manchester an enjoyable place to work.

I would like to also thank my family for all the selfless love, help and support in pursuing my ambition and for always being there for me. I would not have been where I am without your help.

Finally, I would like to thank Aditi for her unconditional support throughout the course of the PhD and for being around on any day of the week, at any hour of the day listening to me complaining about the bugs in coding or thesis writing.

Introduction

The matter antimatter asymmetry in the universe and the neutrino mass could be explained by CP violation in particle interactions and decays in more complex Beyond Standard Model (BSM) hidden or dark sector. The dark sector is defined as a collection of particles that do not interact directly with the Standard Model (SM) particles via strong, weak or electromagnetic forces. These particles are assumed to either interact gravitationally or via “portal” interactions that are constrained by the symmetries of the SM [1].

The MicroBooNE detector being a Liquid Argon Time Projection Chamber (LArTPC) has an excellent particle identification and reconstruction efficiencies and thus can be used to search for rare decays in dark sector. The experiment is exposed to neutrino flux from an on-axis Booster Neutrino Beam (BNB) and from the off-axis Neutrinos at the Main Injector (NuMI) beam.

The thesis describes a search for a new dark sector scalar boson S that decays to an electron-positron pair inside the MicroBooNE. The scalar S at MicroBooNE could be produced by both BNB and NuMI beams at different locations but this thesis focuses on the production of scalars from Kaons Decaying In Flight (KDIF) in the NuMI decay pipe, as well as Kaons Decaying At Rest (KDAR) in the NuMI target and hadronic absorber of the NuMI beam, via the decay channel $K \rightarrow \pi S$ with S consequently decaying into e^+e^- . For the energy ranges of the NuMI beam, scalars decay to e^+e^- with branching ratio equal to unity and therefore we search for scalars in the mass range of 100 – 200 MeV at MicroBooNE.

The hadronic absorber is located downstream of MicroBooNE at the end of the

NuMI decay pipe (described in section 3.1). Therefore, the decay products of the scalar produced at the beam dump would approach the detector in direction almost opposite to the neutrinos coming from the beam target. In addition, KDAR at target and at the hadronic absorber are two-body decays and therefore scalars produced will be monoenergetic. This results in KDAR to have unique kinematics which could be extremely useful to exclude the large neutrino background and therefore increasing the sensitivity of MicroBooNE to these Higgs Portal scalars.

The structure of this thesis is as follows: Chapter 1 describes the theoretical motivation and phenomenology of the Higgs Portal model. Chapter 2 describes the operational principle of LArTPCs and the MicroBooNE detector. Chapter 3 describes the NuMI neutrino beam and the flux of neutrinos (background of the study) and Higgs portal scalars (signal of the study) that MicroBooNE receives. Chapter 4 outlines the interactions of the particles produced in the MicroBooNE detector. Chapter 5 describes the methods and software used for simulation and subsequent reconstruction of an event at MicroBooNE. In addition this chapter also outlines the generation and normalisation of the simulated samples for neutrino interactions (background) and Higgs portal scalars (signal). Chapter 6 outlines the analysis workflow developed to discriminate a Higgs portal scalar decaying into an e^+e^- pair (signal) from the neutrino interaction (background). Chapter 7 describes multivariate machine learning technique called Boosted Decision Trees (BDTs) that is used to further discriminate the signal from background in higher-dimensionality variable space. Chapter 8 describes various sources of the systematic uncertainty considered in the analysis presented in this thesis. Finally chapter 9 presents the final results and exclusion region contours for Higgs portal model.

Chapter 1

Higgs Portal Scalar Search

This chapter describes the theoretical background of the Higgs Portal Model considered in this thesis. In addition, it discusses the searches and methods used to probe the Higgs Portal Model by various experiments along with their constraints on the $(\theta - M_S)$ parameter space that describes the model.

1.1 Higgs Portal Model

The SM of particle physics can explain the physics of familiar matter over a wide range of energy exceptionally well but it is unarguably incomplete as it cannot explain some of the outstanding puzzles in particle physics such as dark matter, neutrino mass, the hierarchy problem, inflation, and the baryon asymmetry in the Universe. All of these puzzles could be explained via an extension of the SM by introducing new weakly interacting particles that interact either gravitationally or via “portals” [1]. Numerous models and search strategies have been developed to look for these weakly interacting particles to try to explain these puzzles such as dark matter and inflation in Ref. [2, 3], baryon asymmetry in Ref. [4] and neutrino mass and the hierarchy problem in Ref. [5].

One such model is a dark sector that comprises a collection of particles that interact with the SM particles at mass scales comparable to familiar matter with very

weak interaction strengths via renormalizable “portals” such as the Higgs, vector or neutrino portals [6]. The study presented in this thesis considers the dark sector-SM interactions of particles by introducing a new dark matter candidate which is an electrically neutral real singlet scalar particle called a dark scalar that interacts with the SM particles via the Higgs Portal Model [7]. This dark scalar could possibly explain the dark matter and thus baryonic asymmetry observed in the Universe.

The Higgs Portal model is a minimal extension of the SM that has a new singlet scalar boson, S , that couples to the SM Higgs boson via the bilinear $H^\dagger H$ operator of the SM [8]. The Lagrangian, \mathcal{L} , of this model can be expressed as

$$\mathcal{L} = \mathcal{L}_{\text{SM}} + \mathcal{L}_{\text{DS}} + \mathcal{L}_{\text{HP}} \quad (1.1)$$

where

$$\mathcal{L}_{\text{HP}} = -(AS + \lambda S^2)H^\dagger H, \quad (1.2)$$

and \mathcal{L}_{SM} , \mathcal{L}_{DS} and \mathcal{L}_{HP} are the Lagrangians for the SM, dark sector and Higgs Portal, respectively. The dark sector Lagrangian may contain an interaction term between the Higgs portal scalar and the dark matter. The Lagrangian of the Higgs Portal contains two types of renormalisable dimensional and dimensionless portal couplings, A and λ , that couple the new scalar singlet to the SM Higgs doublet [8].

After diagonalisation and electro-weak symmetry breaking, we acquire two parameters that are relevant for our study, the physical mass of the dark sector scalar, M_S , and the mixing angle between the scalar and the Higgs, θ [6]. These parameters are related by the Lagrangian

$$\mathcal{L}_{\text{HP}} \supset \frac{1}{2}M_S^2 S^2 + S \sin\theta \left(\frac{2m_W^2}{v} W_\mu^+ W^{\mu+} + \frac{m_Z^2}{v} Z_\mu Z^\mu - \sum_f \frac{m_f}{v} \bar{f} f \right), \quad (1.3)$$

where f is a fermion with mass m_f , v is the vacuum expectation value of the Higgs field, and W and Z are gauge bosons with masses m_W and m_Z , respectively. We have assumed $\theta \ll 1$ due to tight experimental constraints on the mixing angle and

have also assumed that there are no new additional dark sector particles with mass lighter than half the Higgs portal scalar mass.

1.1.1 Production

The dominant channel for the production of the scalar at MicroBooNE is via the decay of kaons to pions that proceeds preferentially via the one-loop penguin decay shown in figure 1.1. Due to the mass dependence between the coupling of the scalar and a quark, S couples most strongly to heavy quarks and therefore the preferential mode of production is by a virtual top quark in the loop. The partial width for the kaon decaying into a pion and a scalar is given by [9]

$$\Gamma(K^\pm \rightarrow \pi^\pm S) \simeq \frac{\theta^2}{16\pi m_K} \left| \frac{3V_{td}^* V_{ts} m_t^2 m_K^2}{32\pi^2 v^3} \right|^2 \lambda^{1/2} \left(1, \frac{M_S^2}{m_K^2}, \frac{m_\pi^2}{m_K^2} \right), \quad (1.4)$$

where m_K is the mass of the kaon, V_{td} and V_{ts} are the CKM matrix elements, and λ is the Källén lambda function or the triangle function defined as

$$\lambda(x, y, z) \equiv x^2 + y^2 + z^2 - 2xy - 2yz - 2zx. \quad (1.5)$$

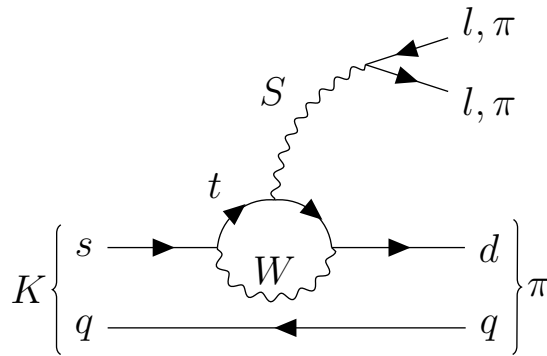


Figure 1.1: A kaon decaying to a pion and Higgs portal scalar via a penguin process. The scalar later decays into pairs of SM particles.

1.1.2 Decay

The kaon decays to the Higgs portal scalar with a probability proportional to square of the mixing term, $\sin^2 \theta$, and the decays of the Higgs portal scalar are also driven by the mixing term, $\sin \theta$. The scalar couples to the SM particles in the same fashion as the Higgs boson but with a strength reduced by a factor $\sin \theta$, and therefore the decay properties are the same as that of a Higgs boson. The maximum allowed mass for the Higgs portal scalar produced via kaon decay is $m_K - m_\pi \simeq 354$ MeV. The possible decay modes below this mass value are $S \rightarrow e^+e^-, \mu^+\mu^-, \pi^0\pi^0, \pi^+\pi^-$.

Using equation 1.3, the partial decay width for the scalar decaying to pions is calculated to be

$$\Gamma(S \rightarrow \pi\pi) = \theta^2 \frac{3|G_\pi(s)|^2}{32\pi v^2 M_S} \left(1 - \frac{4m_\pi^2}{M_S^2}\right)^{1/2}, \quad (1.6)$$

where m_π is the mass of the pion and

$$G_\pi(s) = \frac{2}{9}s + \frac{11}{9}m_\pi^2, \quad (1.7)$$

where $s = (p + p')^2$ for the four momenta, p and p' , of the final state pions. The complete derivation of the form factor is described in Ref. [10]. The total width for the scalar decaying to charged pions, $\pi^+\pi^-$, is twice that of equation 1.6, due to charge conjugation symmetry. The partial decay rate of the dark scalar decaying to charged leptons with mass m_ℓ can also be calculated using equation 1.3 and is given by

$$\Gamma(S \rightarrow \ell^+\ell^-) = \theta^2 \frac{m_\ell^2 M_S}{8\pi v^2} \left(1 - \frac{4m_\ell^2}{M_S^2}\right)^{3/2}. \quad (1.8)$$

Figure 1.2 shows the branching ratio of the various decay channels of the Higgs portal scalar as a function of its mass. The analysis presented in this thesis searches for the Higgs portal scalar in the mass range $100 \leq M_S \leq 200$ MeV where the scalar decays into an electron-positron pair. The MicroBooNE collaboration has

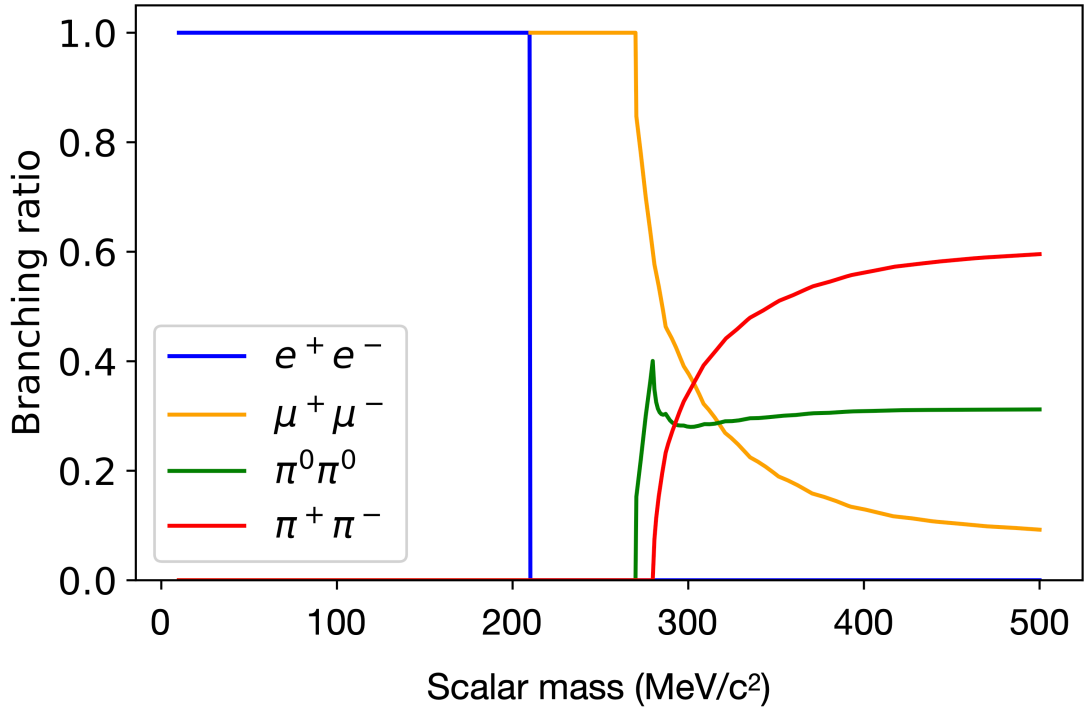


Figure 1.2: The branching ratio of the various decay channels of the Higgs portal scalars as a function of its mass [9].

already published searched for Higgs portal scalars decaying into e^+e^- pairs [11] as well as $\mu^+\mu^-$ pairs [12] for scalars produced from kaons decaying at rest in the NuMI beam absorber. This thesis searches for scalars decaying into an e^+e^- pair for scalars produced from Kaons Decaying At Rest (KDAR) in the NuMI target, Kaons Decaying In Flight (KDIF) in the NuMI decay pipe, and KDAR in the NuMI absorber. In addition, the analysis performed in this thesis takes into account approximately four times larger statistics compared to the previous search for the same decay mode and in the same mass range. Furthermore, this study exploits full calorimetric information to improve the performance of the BDT in distinguishing the signal from the background.

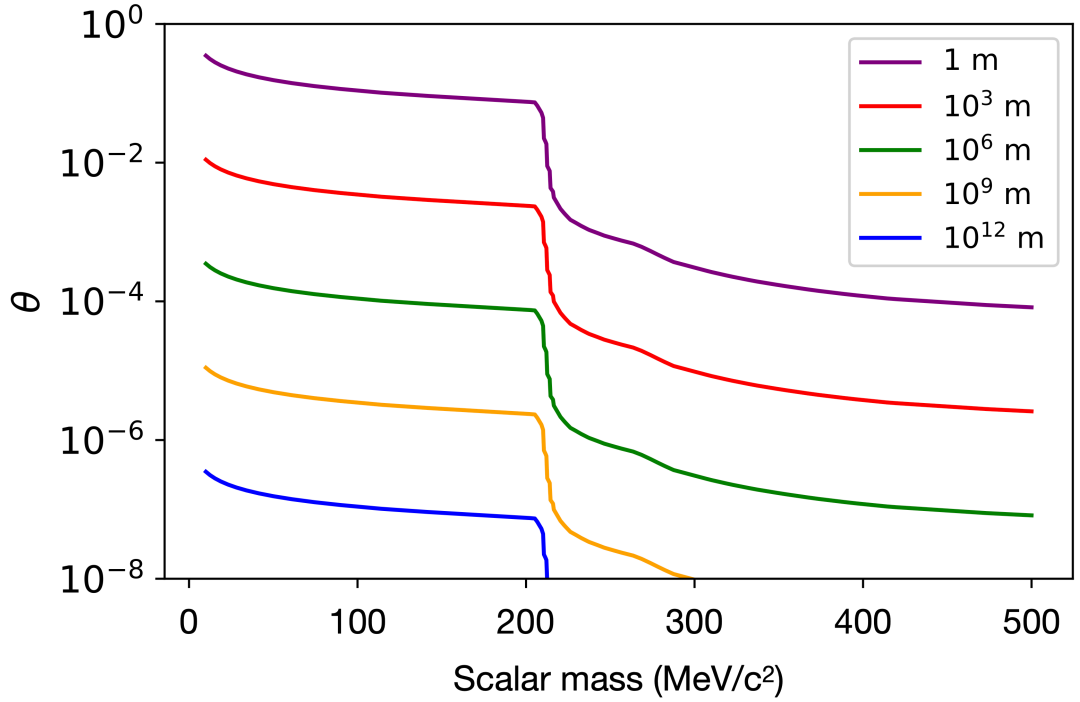


Figure 1.3: Isocontours of the Higgs portal scalar decay length in the $\theta - M_S$ parameter space [9].

1.2 Higgs Portal Scalar Decay Length

Figure 1.3 shows the isocontours of the scalar decay length in the $\theta - M_S$ parameter space. The scalar decay length, L , is given by

$$L_S = \gamma_S \beta_S \tau_S, \quad (1.9)$$

where $\tau_S = \hbar/\Gamma$ calculated using equations 1.6 and 1.8 for scalars decaying to pions and leptons respectively and γ_S is the Lorentz boost of the scalar defined as

$$\gamma_S = E_S/M_S, \quad (1.10)$$

where E_S is the energy of the scalar. For Higgs portal scalars produced from KDAR, E_S can be easily calculated using two-body decay kinematics,

$$E_S = \frac{m_K^2 + M_S^2 - m_\pi^2}{2m_K}, \quad (1.11)$$

where m_K and m_π are the masses of the parent kaon and daughter pion respectively. For scalars produced from KDIF, E_S is calculated by boosting daughter particles to the parent kaon's moving frame. The detailed descriptions of the two-body decay kinematics for KDAR and KDIF are given in section 6.1. The MicroBooNE detector samples scalars that propagate 800 m from the target and 100 m from the absorber. Therefore for masses of the scalar in the range $100 - 200 \text{ MeV}/c^2$, MicroBooNE can exclude values of mixing parameter below 10^{-2} . For masses of the scalar larger than $\sim 200 \text{ MeV}/c^2$, where $M_S > 2m_\mu$, the scalars can decay into different decay modes giving rise to the shape of the contours seen in the figure 1.3 as well as the shapes of the reinterpretations of the experiments PS191, CHARM and LSND described in the next section 1.3.

1.3 Existing Limits on Higgs Portal Scalars

For our region of interest with scalars in the mass range $100 - 200 \text{ MeV}$, the current experimental limits on the Higgs Portal scalar mixing angle, θ , for various experiments are listed below and summarised in figure 1.4.

- The **MicroBooNE** experiment has published a search for scalars produced from KDAR in the NuMI absorber for the decay channel $S \rightarrow e^+e^-$. For a mass range similar to the analysis presented in this thesis, but with a much smaller data sample corresponding to 1.93×10^{20} Protons On Target (POT), the observed limit at 95% Confidence Level (C.L.) on θ for this experiment was calculated to be in the range 10^{-3} to 3×10^{-4} [13], and is shown as the solid purple line in figure 1.4.
- The **MicroBooNE** collaboration has also published a search in the mass range $212 - 275 \text{ MeV}$ for scalars produced from KDAR in the NuMI absorber for the decay channel $S \rightarrow \mu^+\mu^-$. The data was taken in two different Forward Horn Current (FHC) and Reverse Horn Current (RHC) detector configurations (described in section 3.1 of chapter 3) corresponding to 7.01×10^{20} POT. The observed limit at 90% C.L. on the mixing parameter θ was calculated to be in

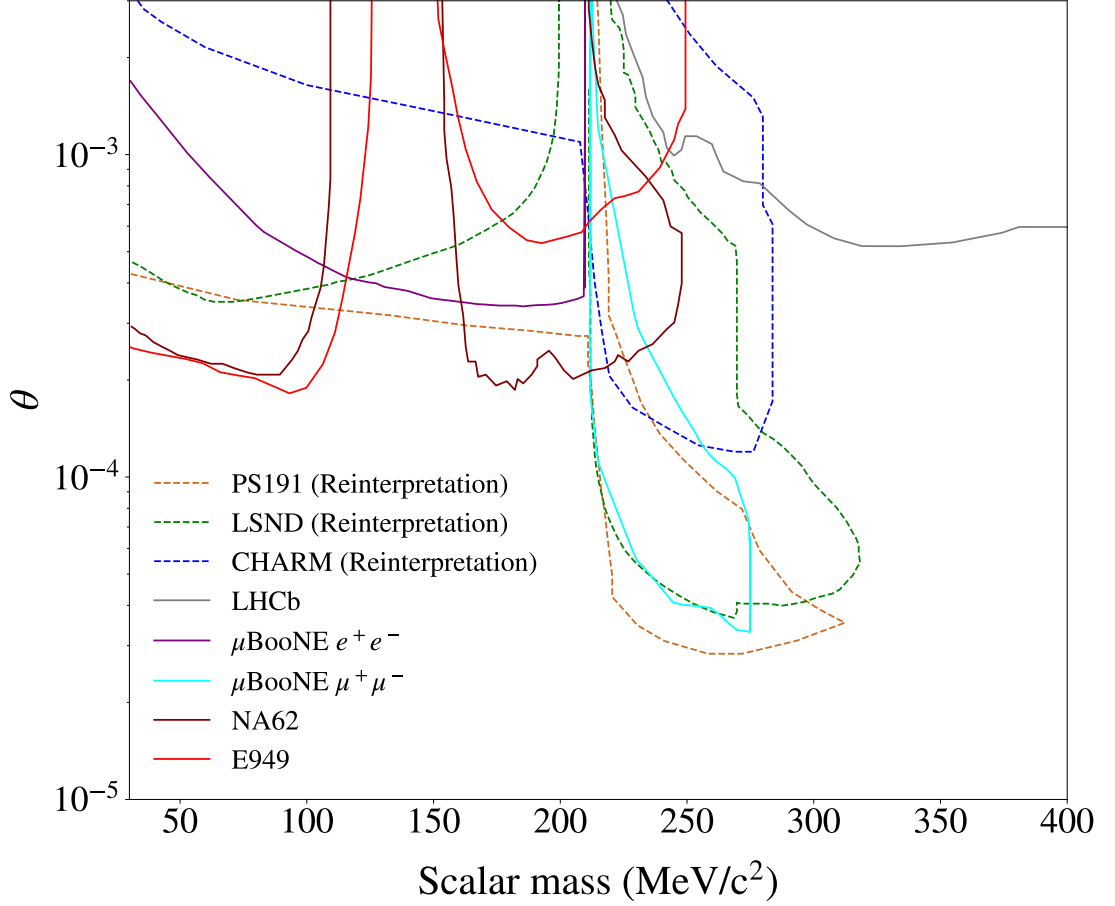


Figure 1.4: The existing experimental limits on the Higgs portal scalar's $(\theta - M_S)$ parameter space. The limits in dashed lines are reinterpretations of the historical experimental results whereas the limits in solid lines are dedicated searches. Micro-BooNE with decay channel $S \rightarrow e^+e^-$, LHCb and CHARM set a limit on the mixing parameter at 95% C.L. whereas the remaining experiments set the limit at 90% C.L.

the range 10^{-4} to 3×10^{-5} [12], and is shown as the solid cyan line in figure 1.4.

- The **E949** experiment has produced a limit on the branching fraction of the decay of $K^+ \rightarrow \pi^+ X$ where X is a massive non-interacting particle that exits the detector undetected [14]. Translating this limit into limits on the mixing angle θ using the width of the kaon decay to the Higgs Portal scalar in equation 1.4, θ is constrained to the level 2×10^{-4} to 10^{-3} at 90% C.L. for masses up to 250 MeV shown by the solid red line in figure 1.4. The gap in the sensitivity around the mass of the π^0 is due to the background from the decay $K^+ \rightarrow \pi^+ \pi^0$.

- The **NA62** experiment at CERN also produced limits on the branching fraction of the decay $K^+ \rightarrow \pi^+ + X$ where X is a long-lived feebly interacting particle that exits the detector undetected. The solid brown line in figure 1.4 was obtained through an interpretation of the $K^+ \rightarrow \pi^+ \nu \bar{\nu}$ decays in 2017 data collected by the NA62 experiment [15]. Translating this limit into limits on the mixing angle θ , limits are calculated in the range $2 \times 10^{-4} - 3 \times 10^{-4}$ at 90% C.L. for masses up to 250 MeV. The gap surrounding the mass of the π^0 is again due to the background from the decay $K^+ \rightarrow \pi^+ \pi^0$.
- The **LHCb** experiment at CERN has produced limits by searching for long-lived scalar particles in $B^+ \rightarrow K^+ \chi$ [16] and $B^0 \rightarrow K^{*0} \chi$ [17] decays with $\chi \rightarrow \mu^+ \mu^-$, in pp collision data corresponding to an integrated luminosity of 3.0 fb^{-1} . The 95% CL limit for these experiments on θ is of order $\sim 10^{-3}$ over the mass range 250 – 4700 MeV, and is shown by the solid grey line in figure 1.4.

The limits listed below are reinterpretations of experimental results calculated without using the raw experimental data and are usually carried out by researchers who are not part of the collaborations.

- The **CHARM** experiment at CERN performed a search for axion-like particles with a 400 GeV proton beam that is dumped on a thick copper target. In this experiment, decays of neutral and scalar particles to $\gamma\gamma$, e^+e^- and $\mu^+\mu^-$ were investigated. The production of the axion was simulated by a Monte Carlo program that is analogous to the simulation of π^0 production from a fixed copper target. In contrast with the Higgs Portal scalars where scalars are produced from kaons, the axion-like particles in the CHARM experiment are produced directly from the proton-copper interactions. In order to calculate a limit on θ , reinterpretation of the result relies on estimating the production rate of the kaons that decay to the Higgs Portal scalars. Therefore the limit on the mixing parameter θ for CHARM is taken from the most recent published results [18] in which a more accurate model is used in which absorption of the kaons in the thick copper target is taken into consideration. This is shown by

the dashed blue line in figure 1.4, and significantly reduces the axion production from kaons, thus weakening the limit compared to previous evaluations such as [19].

- The **LSND** experiment at Los Alamos was a liquid scintillator detector that collected over 10^{23} POT over its lifetime (1993-1998); this is one of the largest datasets collected by any fixed target experiment. The data was collected by impinging ~ 800 MeV protons from the linear accelerator on various targets such as a 20-cm water target in 1993, a 30-cm long water target followed by a water-cooled copper beam dump from 1993 to 1995, and finally a close-packed heavy metal target from 1996 to 1998 [20]. A recent publication [6] has estimated the production rate of Higgs Portal scalars at LSND with the dominant production mode being proton bremsstrahlung, $p + p \rightarrow X + S$, with X being an inclusive hadronic final state. The reinterpretation on the limits for the decay channel e^+e^- [21, 20] and $\mu^+\mu^-$ [22] in the mass region below ~ 200 MeV and 211 to 350 MeV were calculated to be $\sim 3 \times 10^{-4}$ and $\sim 3 \times 10^{-5}$ respectively. These limits are shown as the dashed green line in figure 1.4.
- The fixed target **PS191** experiment at CERN was originally built to search for sterile neutrinos from weak meson decays. The authors of the reinterpretation have produced limits on the Higgs portal scalars at 90% C.L. of $\theta \sim 3 \times 10^{-4} - 5 \times 10^{-4}$ for the scalar mass range 100 – 150 MeV and $\theta \sim 3 \times 10^{-5} - 1 \times 10^{-4}$ for the mass range 200 – 300 MeV [23]. These are shown as the dashed brown line in figure 1.4. The re-simulation of the beamline used does not take into account the PS191 magnetic system, which can have a significant impact on the flux of charged kaons and hence the subsequent decay product, the Higgs portal scalar, reaching the decay volume. The resultant limits are therefore expected to be conservative.

The analysis presented in this thesis sets the world’s best limit on the $(\theta - M_S)$ parameter space at 95% C.L. for the Higgs portal scalar decaying into e^+e^- pairs in the mass range 125 – 160 MeV. The current best limit in this mass range is the

reinterpretation of the PS191 experiment with limit on θ calculated to be $\sim 2 \times 10^{-4}$ at 90% C.L. We calculate a limit on θ which is better than this current best limit of 2×10^{-4} at 95% C.L. with full systematic treatment included in the calculation.

Chapter 2

The MicroBooNE Experiment

The Micro Booster Neutrino Experiment (MicroBooNE) is a liquid argon time projection chamber (LArTPC) designed to observe interactions of neutrinos produced by the on-axis Booster Neutrino Beam (BNB) and the off-axis Neutrinos at the Main Injector (NuMI) beam.

This chapter describes the technical aspects of the MicroBooNE experiment. Section 2.1 describes the working principle of the LArTPC. Section 2.2 describes the MicroBooNE coordinate system. Section 2.3 describes the MicroBooNE time projection chamber with subsections explaining the MicroBooNE charge and light collection systems in detail. Section 2.4 explains the MicroBooNE cosmic ray tagger system which is used to veto cosmogenic muons. Section 2.5 describes the MicroBooNE readout and trigger systems responsible for recording events in coincidence with the neutrino arrival time at MicroBooNE. Section 2.6 describes the MicroBooNE event display by taking an example of a neutrino interaction.

2.1 Liquid Argon Time Projection Chamber

The liquid argon time projection chamber is a technology that uses a large volume of liquid argon to create ionisation electrons and scintillation signals when traversed by charged particles. The operational principle of the LArTPC is shown in figure 2.1. An

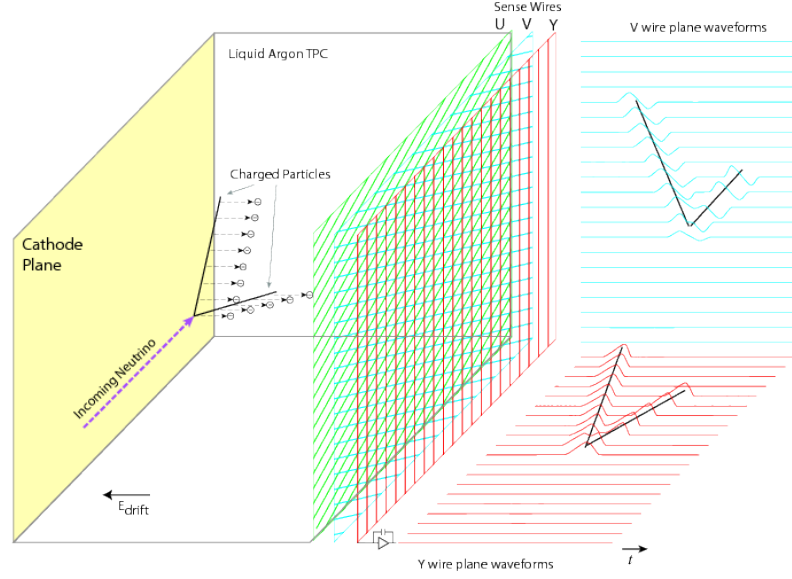


Figure 2.1: This schematic shows the operational principle of the LArTPC. The interaction of the neutrino with argon produces charged daughter particles that ionise and excite the argon. A uniform electric field drifts the ionisation electrons towards the sense wires to be read out. These wires are oriented at different angles to produce different projections of the event. These projections are combined to create a three-dimensional image of the event. Figure from Ref. [24].

incoming neutrino interacts with argon nuclei in the TPC’s active volume (described in section 2.3) and produces charged particles. These charged particles leave trails of ionisation electrons as they traverse through the argon. The electric field then drifts these ionisation electrons to the wire planes. These ionisation trails are then detected by the sense wires comprising two induction planes, “U” and “V”, with wires oriented $\pm 60^\circ$ relative to vertically aligned wires of the single collection plane, “Y”. The induction planes collect signals induced by drifting ionisation electrons, whereas the collection plane collects the ionisation electrons. The difference between the arrival time of the ionisation electrons at the sense wire planes and the time of the interaction is the electron drift time and is $\mathcal{O}(\text{ms})$ [24]. The detector also has a light collection system comprising 32 8-inch PhotoMultiplier Tubes (PMTs) located behind these wire planes to provide timing information from the prompt scintillation light produced by $\nu - \text{Ar}$ interactions [25] [26]. The detector has a Cosmic Ray Tagger (CRT) system consisting of scintillator panels surrounding the cryostat which was installed after MicroBooNE’s second data run to tag cosmic muons entering the

TPC.

Liquid Argon (LAr) is used as a detection medium because it has high density and is relatively cheap, enabling the detector to have a high neutrino interaction rate and significant ionisation per unit length of a charged particle track or shower. In addition, argon is a noble element and has relatively low ionisation potential which enables ionisation electrons to traverse large distances without being absorbed. This gives a LAr detector excellent calorimetry capability as well as spatial resolutions down to the scale of millimetres. The LAr detectors cannot directly observe chargeless particles such as neutrinos, photons, neutrons or neutral pions as they do not ionise or excite argon. Indirect observations of these neutral particles is possible if they produce charged particles upon their decay or their interaction with argon.

2.2 MicroBooNE Detector Coordinate System

The coordinate system of the MicroBooNE LArTPC is shown in figure 2.2. The direction of the BNB is shown by the green arrow and points along the z -axis of the MicroBooNE coordinate system. The x -axis points in the direction of the electric field from anode to cathode and the component y -axis points vertically upward perpendicular to the $x - z$ plane. The two angular quantities, the azimuthal angle ϕ and the zenith angle θ , are defined as

$$\begin{aligned}\theta &= \tan^{-1} \left(\frac{p_t}{p_z} \right); \quad p_t = \sqrt{p_x^2 + p_y^2}; \\ \phi &= \tan^{-1} \left(\frac{p_y}{p_x} \right),\end{aligned}\tag{2.1}$$

where p_x, p_y and p_z are the x, y and z components of particle's three-momentum, with p_t being the transverse momentum.

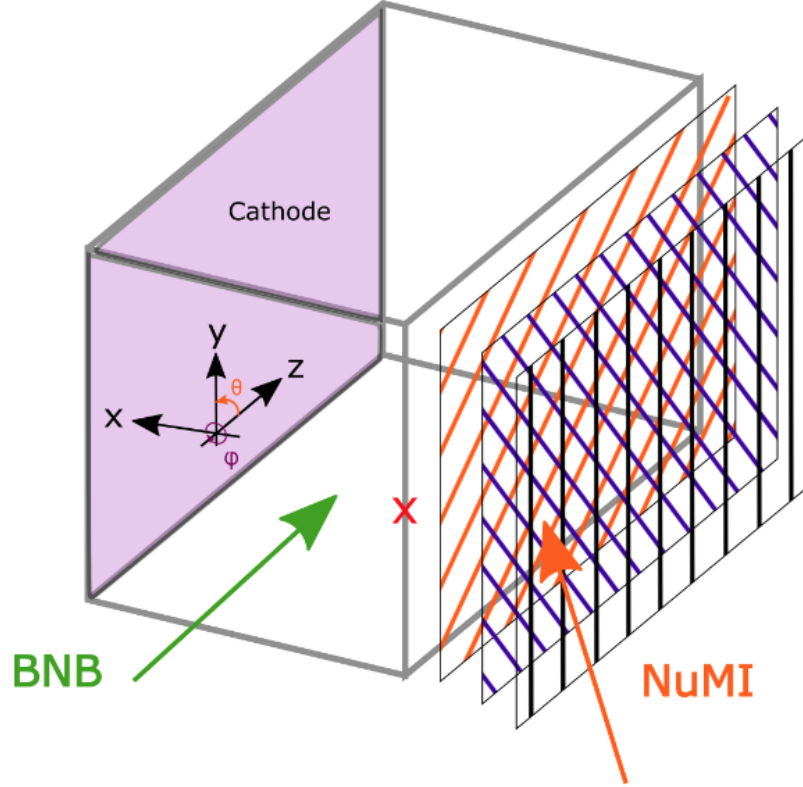


Figure 2.2: A schematic of the coordinate system used for the MicroBooNE LArTPC. The direction of the BNB is shown in green and the approximate direction of particles arriving from the NuMI target is shown in orange. The origin of the MicroBooNE coordinate system is marked by the red cross. The z -axis points in the direction co-linear with the BNB, the y -axis points vertically upwards and the x -axis points in the direction of the electric field from anode to cathode. The azimuthal angle, ϕ , is defined as the angle in the $x - y$ plane, whereas the polar angle, θ , is the angle relative to the z -axis. Figure from Ref. [27].

2.3 MicroBooNE Time Projection Chamber

The TPC in MicroBooNE is a cuboid composed of the anode plane assembly, the cathode and the field cage. It is housed inside a cylindrical cryostat containing 170 tonnes of liquid argon as shown in figure 2.3. The TPC has dimensions of 2.3 m (height) \times 2.6 m (width) \times 10.4 m (length). The anode plane assembly comprises three stainless steel sense wire planes that are coated with a thin layer of copper and silver. From inside out, the first two wire planes, “U” and “V”, are each made of 2400 wires that are oriented $\pm 60^\circ$ with respect to the vertically aligned wire plane “Y” made of 3456 wires. The distance between the wire planes is 3 mm and the wires within each plane are spaced with a 3 mm pitch. A very high negative voltage of

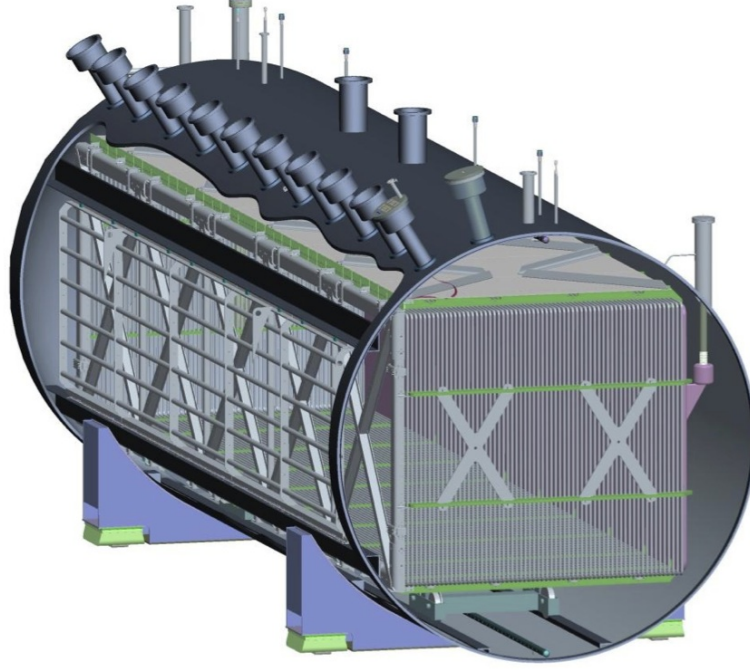


Figure 2.3: Schematic diagram of the MicroBooNE TPC housed inside the cylindrical cryostat (from Ref. [24]).

-70 kV is applied at the cathode to produce an electric field between the cathode and the anode.

The uniformity of the electric field between the cathode and the anode is created by the field cage which is a series of 64 thin-walled stainless steel pipes of diameter 2.54 cm that are mounted parallel to the cathode and the anode planes in a rectangular loop. These rectangular loops are electrically connected via a resistor divider chain to incrementally step the voltage down to maintain a uniform electric field of 273.9 V/cm between the cathode and the anode. This results in a drift time of 2.3 ms between the cathode and the anode and this time period defines a single readout frame in MicroBooNE. The volume enclosed by the field cage, the anode and the cathode is known as the “active” volume of the detector.

The fiducial volume is the region inside the TPC defined by the coordinates listed in table 2.1. These coordinates are chosen to be 10 cm from the outer edges of the TPC to exclude border effects (described in section 4.4.2) that often leads to mis-reconstruction of an event.

Coordinate	Requirement
x	$10 \text{ cm} < x < 246 \text{ cm}$
y	$-106 \text{ cm} < y < 106 \text{ cm}$
z	$10 \text{ cm} < z < 1026 \text{ cm}$

Table 2.1: The requirement for a region inside the TPC along x , y and z to exclude border effects in the reconstruction, also known as the fiducial volume.

2.3.1 MicroBooNE Charge Collection System

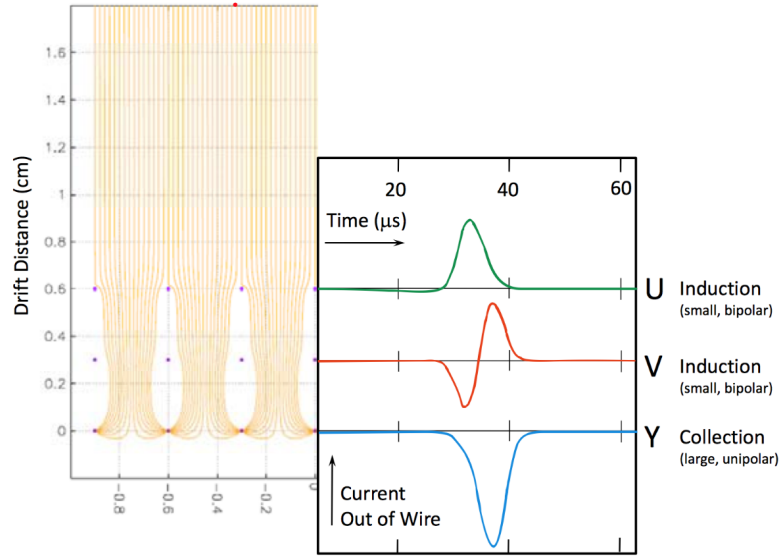


Figure 2.4: This diagram illustrates the signal response to ionisation electrons on the U, V and Y planes. The schematic on the left shows the $x-z$ plane of the TPC where the electric field lines along the electron drift direction are shown in orange and the wires of the U, V and Y planes are shown as purple dots. The schematic on the right shows the bipolar signal on the two induction planes U and V induced by the ionisation electron with final collection plane Y producing a unipolar signal from the collection of the electron charge. Figure from Ref. [25].

The MicroBooNE charge collection system comprises three sense wire planes, “U”, “V” and “Y”. The “U” and “V” planes have a voltage bias of -110 V and 0 V applied respectively such that drift electrons pass through these planes producing a bipolar signal as shown in figure 2.4. The “Y” plane has a voltage bias of 230 V applied to it to collect the drift-electron charge producing a unipolar signal.

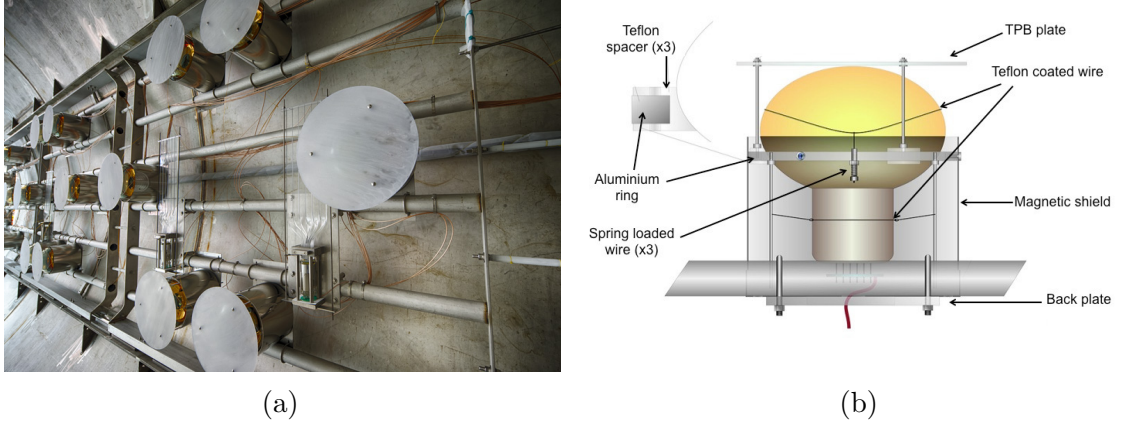


Figure 2.5: (a): A picture showing the PMTs mounted on the frame inside the cryostat. (b): Schematic showing a PMT with TPB plate to shift the wavelength of the scintillation light from Vacuum-UltraViolet (VUV) to the PMT’s sensitive region.

2.3.2 MicroBooNE Light Collection System

Liquid argon is an excellent scintillator and sampling the light produced from the interactions gives LArTPC detectors powerful capabilities and crucial information to complement the charge information recorded using the sense wires. The MicroBooNE light collection system comprises 32 Hamamatsu R5912-02mod cryogenic Photo-Multiplier Tubes (PMTs) each of 8-inch diameter. These PMTs are located inside the cryostat behind the sense-wire planes at the anode. A photograph of the PMTs mounted on the frame and a schematic of a PMT are shown in figure 2.5. These PMTs can detect light in the visible region with wavelengths in the range (300 – 650) nm. The efficiency¹ of these PMTs is about 20% for a wavelength of about 400 nm. The scintillation light produced in liquid argon is in the VUV region with a wavelength narrowly peaking at 128 nm (discussed in section 4.3.5 of chapter 4). The PMTs have effectively zero quantum efficiency at this wavelength, and therefore a plate coated with TetraPhenyl-Butadine (TPB) is placed in front of the PMT which absorbs the VUV light and re-emits it with a wavelength of about 425 nm. After the signals are split off from the baseline high voltage, they are divided into high-gain and low-gain channels with each carrying a fraction of 18% and 1.8% of the signal amplitude respectively. This extends the dynamic range of the Analog-to-Digital Converter

¹Percentage ratio of the number of detected photons to the number incident.

(ADC) readout of the PMT pulses [24]. Finally the signals recorded by each PMT are digitized at a rate of 64 MHz using an ADC.

The production and propagation of photons happens on microsecond-timescales and is a few orders of magnitude faster than the millisecond-timescale ionisation electron propagation (discussed in section 4.4.4), therefore scintillation light is exploited to provide timing information, which plays a crucial role in resolving the x -position of charge deposition along the drift direction, enabling three-dimensional reconstruction of the particle interaction. In addition, scintillation light plays an important role in the MicroBooNE trigger system (discussed in section 2.5), for example in excluding events with no neutrino interaction to save on resources.

2.4 MicroBooNE Cosmic Ray Tagger System

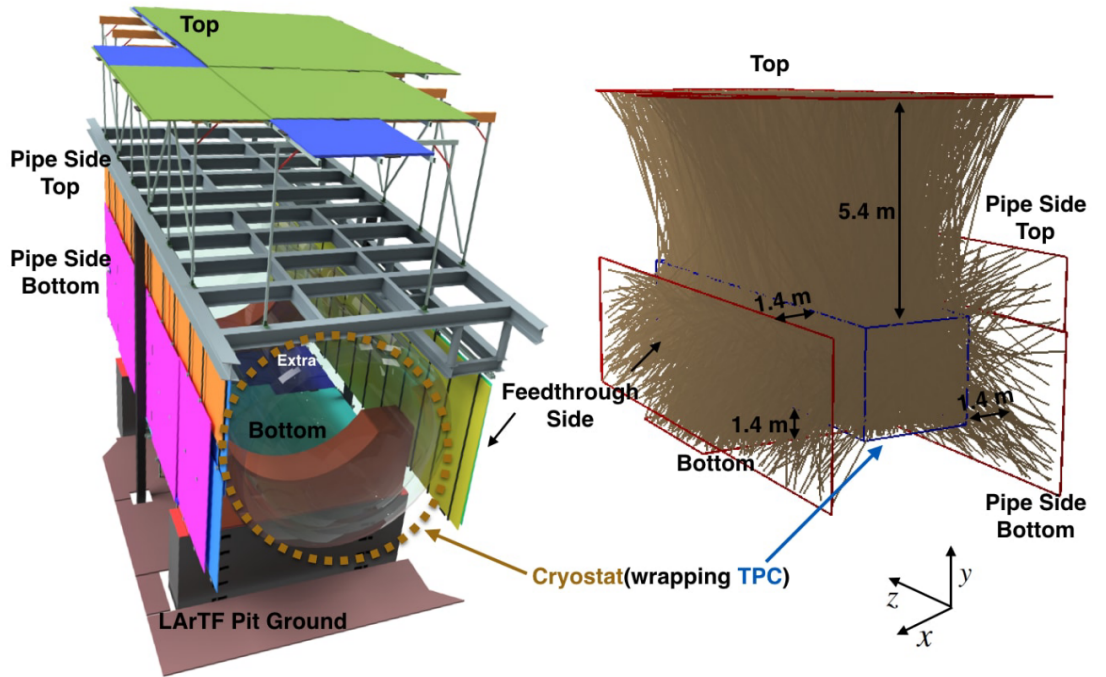


Figure 2.6: The schematic on left shows the positioning of four cosmic ray tagger planes with two planes above and below the detector and the other two planes on either side, encompassing the cylindrical walls of the MicroBooNE cryostat. The schematic on right shows the simulation of cosmic ray trajectories (brown lines) crossing the cosmic ray tagger system. Figure from Ref. [26].

The MicroBooNE detector, being located on the surface, experiences a high flux of

cosmic-ray particles traversing the detector. These particles can sometimes create signatures that resemble neutrino interactions or BSM decays. Therefore it is crucial to eliminate these cosmic muons using novel software techniques and detector technologies.

One such technology is the MicroBooNE Cosmic Ray Tagger (CRT) system, which is an external sub-detector that complements the LArTPC detector to identify and reject cosmic muons. The system was installed part-way through the detector operation in October 2017 and hence the data recorded before the installation relies only on software techniques to reject cosmic muons.

The CRT system comprises 73 scintillating modules made up of interleaved layers of scintillating strips located above, below and on two sides of the detector, along the neutrino beam direction as shown in figure 2.6 (left). Figure 2.6 (right) shows the cosmic muons traversing the MicroBooNE detector with a coverage of 85% estimated using the COsmic Ray SIMulations for KAScade (CORSIKA) program in conjunction with the GEometry ANd Tracking (GEANT) simulation model for the propagation of particles through matter [26]. A cosmic muon traversing the scintillating modules induces scintillation photons that are collected by silicon photomultipliers (SiPMs) which are then digitized and read out by a front end board. The recorded scintillator signal can be exploited to reconstruct geometric positions of the incoming or outgoing muons. The CRT system has a timing resolution of better than 100 ns which can be used to match CRT events with PMT and TPC signals to tag and reject cosmic muons. The extrapolated tracks reconstructed in the LArTPC coinciding with the track reconstructed by the CRT system will be tagged as background cosmogenic muons and are removed from the reconstructed event.

2.5 MicroBooNE Readout and Trigger System

This section covers the MicroBooNE trigger and readout systems used to record data from the NuMI beam. Various terminologies and time intervals used in understanding the trigger and readout system in the following subsections are illustrated in figure

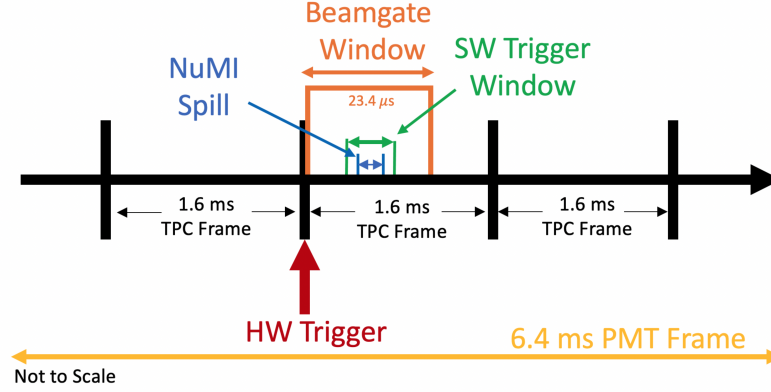


Figure 2.7: Drawing illustrates the trigger and readout systems of the MicroBooNE detector. The Fermilab accelerator division sends the hardware trigger signal to MicroBooNE to open a 6.4 ms time window, shown as the yellow interval, to record an event. This hardware trigger signal also opens a TPC readout time window of 4.8 ms, represented by the black arrow that is sliced into equal time intervals of 1.6 ms. Shown in orange, the beamgate window of $23.4 \mu\text{s}$ opens up 1.6 ms after the first TPC readout frame to record the NuMI beam spill in blue. The software trigger window in green opens just before the NuMI spill window to record an event. Figure from Ref. [27].

2.7. An event at MicroBooNE comprises a continuous readout of 6.4 ms of PMT and 4.8 ms of TPC data. The TPC readout window of 4.8 ms is split into three equal time frames of 1.6 ms. This frame size is chosen on the basis of the time taken by the ionisation electrons to drift from the cathode to the anode inside the MicroBooNE TPC under a nominal operating electric field of 500 V/cm. The current operating field of the MicroBooNE is 273.9 V/cm for which the drift time of the electrons is 2.3 ms.

Section 2.5.1 explains the MicroBooNE beam hardware trigger system that is responsible for initiating the readout of an event. Section 2.5.2 covers the MicroBooNE software trigger system that records an event on tape after rejecting possible cosmogenic events. Section 2.5.3 covers the MicroBooNE external and unbiased trigger system that collects a large sample of cosmogenic events to model and subsequently understand the MicroBooNE background.

2.5.1 MicroBooNE Beam Hardware Trigger

An event in MicroBooNE begins with the arrival of a signal called the beam hardware trigger from the Fermilab accelerator division. The signal can be from the accelerator clocks for the BNB and NuMI beams to record the neutrino interactions or from a function generator (pulser) producing signals at a set frequency to record beam-off events (with no neutrino interactions). This beam trigger causes the TPC readout window of 4.8 ms, as well as an unbiased window (no light information requirements) called the beam-gate window of $23.4\ \mu\text{s}$, to open starting 1.6 ms into the TPC readout frame.

2.5.2 MicroBooNE Software Trigger

With respect to the start of the beam-gate window, the software trigger (SW trigger) window opens from $4.69\ \mu\text{s}$ to $16.41\ \mu\text{s}$ to record neutrino interactions from the NuMI beam within the NuMI beam spill time window of range $(5.64 - 15.44)\ \mu\text{s}$. Due to the very tiny neutrino interaction cross-section, only approximately 1 in 50 of the NuMI spills result in a neutrino interaction inside the MicroBooNE detector. Recording an event on every hardware trigger is thus inefficient as most of these events will contain no neutrino interactions. To avoid reading out empty events, a threshold on the amount of scintillation light in the SW trigger is applied, which uses optical waveforms from the PMT system (described in section 5.4) and saves in real-time any event with a waveform above a threshold of 190 ADC counts which corresponds to scintillation light of 9.5 Photo-Electrons (PE). Approximately 14% of events from hardware triggers pass the SW trigger, which are then stored to tape for further analysis.

2.5.3 MicroBooNE External and Unbiased Trigger

Even after the application of the SW trigger, the majority of events do not contain any neutrino interaction. This could be due to cosmic activity producing scintillation light above the SW trigger threshold in time-coincidence with the beam spill window. Modelling these type of interactions is very crucial to better understand the MicroBooNE background and therefore MicroBooNE collects a very large sample of these events with the same run configuration and trigger condition as beam-on data. The events that pass through the SW trigger are referred to as external or beam-off events, whereas events that do not pass through the SW trigger are referred to as unbiased events at MicroBooNE. The same terminologies that are employed by MicroBooNE will be used throughout this thesis.

2.6 MicroBooNE Event Display

The LArTPC technology provides the MicroBooNE detector the capability to produce photographic quality images of neutrino interactions that have resemblance to images produced by bubble chambers. An example of an event display containing most of the possible SM particles detected in MicroBooNE is shown in figure 2.8. The induction planes, U and V, and the collection plane, Y, display different angular perspectives of a neutrino interaction with the collection plane being equivalent to looking top down and induction planes being equivalent to looking at a $\pm 60^\circ$ angle at the interaction. The colour scale represents the amount of charge deposited with red being high ionisation and green being low ionisation. Protons and muons produce a track-like trajectory and the interaction of these particles in liquid argon is described in section 4.3.1 of chapter 4. The interactions of the electrons and the photons with liquid argon are described in section 4.3.2 and 4.3.3 respectively and their resultant electromagnetic shower production is described in section 4.3.4 of chapter 4.

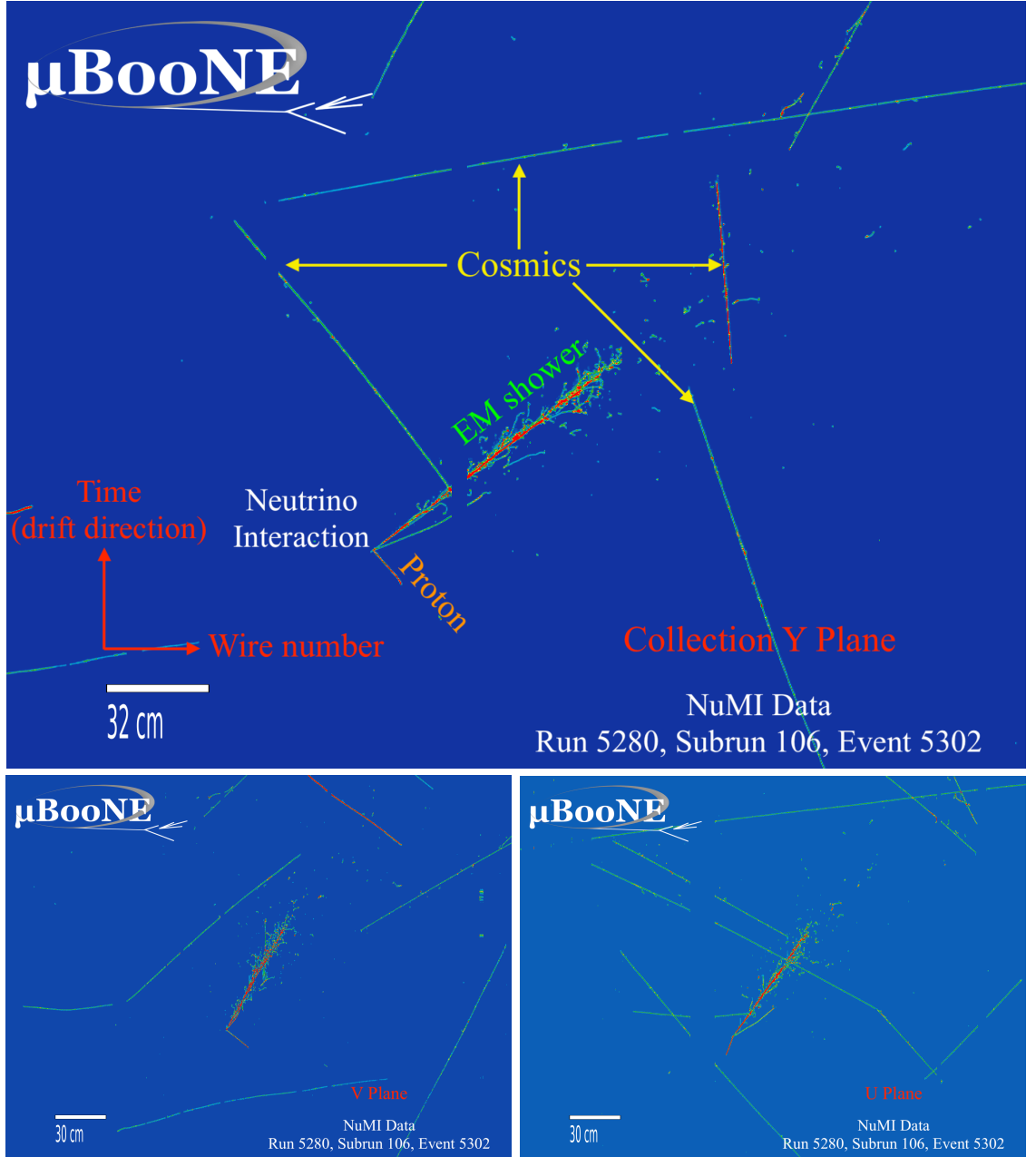


Figure 2.8: An example event display of a candidate neutrino interaction from the NuMI data recorded on the collection plane, Y (top), and the two induction planes, V (bottom left) and U (bottom right). The cosmic muons are indicated by yellow arrows, an electromagnetic shower by green text, a proton by orange text and the neutrino interaction vertex by white text. The x -axis represents the wire number and the y -axis represents the ionisation electron-drift direction. The colour scale represents the amount of charge deposited with red being high ionisation and green being low ionisation.

Chapter 3

The NuMI Beam and Flux at MicroBooNE

The NuMI neutrino beam was originally built to provide neutrinos for the MINOS long-baseline oscillation experiment [28] at Fermilab. The NuMI beam has been running since 2005 and has provided the beam to a number of experiments including ArgoNeuT [29], MINERvA [30], MINOS(+) [31], NOvA [32] and PEANUT [33].

The MicroBooNE experiment was primarily designed to receive neutrino flux from the BNB. However, the experiment also receives a significant amount of off-axis neutrinos from the NuMI beam. Being exposed to both BNB and NuMI beams, the detector can be exploited to validate the results from BNB using NuMI beam. In addition, the NuMI beam can also be utilised to perform cross-section measurements as well as beyond the Standard Model searches such as the one presented in this thesis.

Section 3.1 outlines the NuMI neutrino and anti-neutrino beam production at Fermilab. Section 3.2 describes the NuMI beam simulation model used to simulate the hadron production and beamline geometry. Section 3.3 outlines the central-value flux predictions for the NuMI beam. Finally section 3.4 describes the flux predictions for the Higgs portal scalar signal at MicroBooNE.

3.1 Neutrino Beam Production

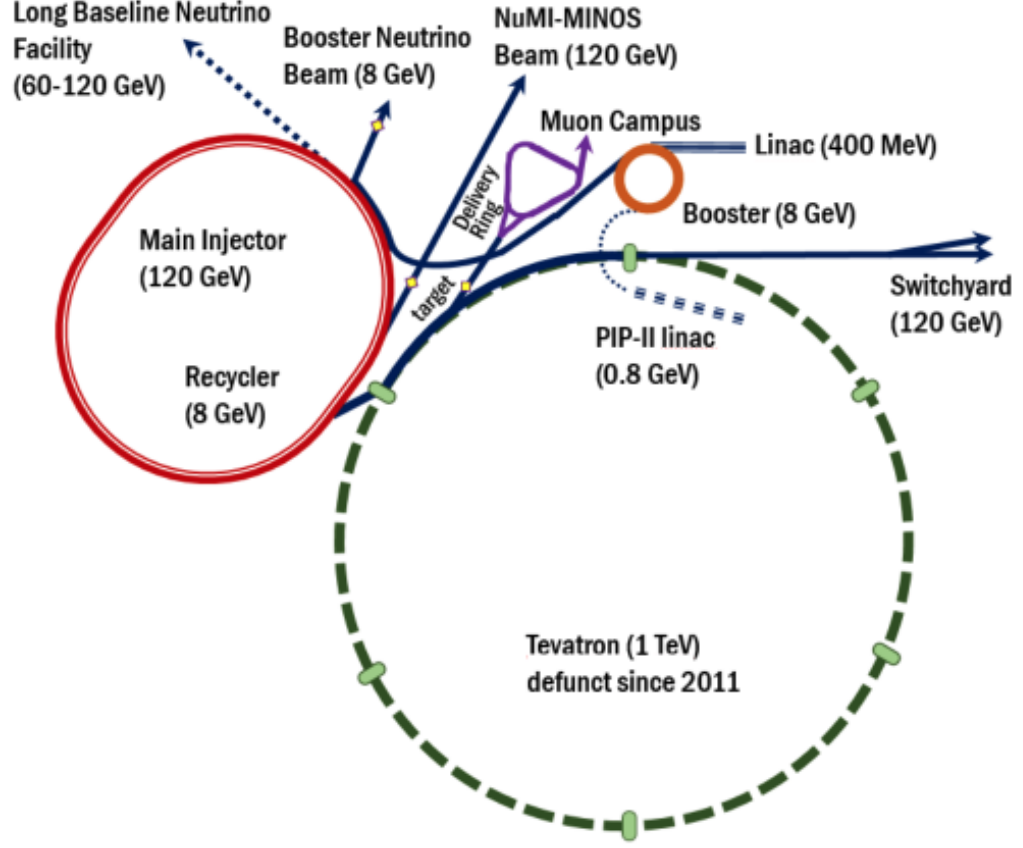


Figure 3.1: Schematic showing the Fermilab Accelerator complex with the Main Injector (in red), the Linac providing a 400 MeV proton beam for the Booster synchrotron (in orange) and the rest of the chain of accelerators as well as the BNB and NuMI beams. Figure from Ref. [34].

Figure 3.1 shows a schematic of the Fermilab accelerator complex outlining the chain of accelerators that are used in creating the BNB and NuMI neutrino beams. The neutrino beams at Fermilab are generated by impinging protons on a fixed target. The protons are produced by accelerating H^- ions to 400 MeV using a Linear accelerator (Linac) and passing them through a carbon foil that strips the two electrons from H^- ions to produce H^+ ions or protons. These protons are then accelerated to 8 GeV using the Booster synchrotron. At this stage, these protons can either be directed to collide with a beryllium target to produce the BNB or towards the Main Injector synchrotron to get further accelerated to 120 GeV. These highly accelerated protons are then steered to collide with a graphite target to produce the NuMI beam.

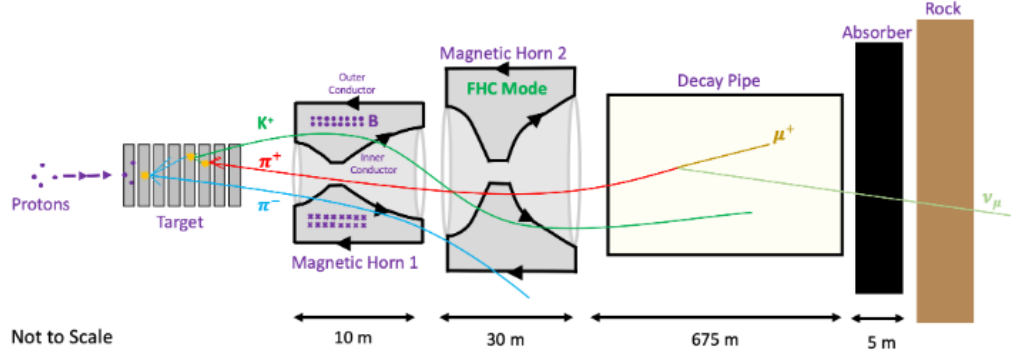


Figure 3.2: A diagram showing the production of a cascade of particles along the NuMI beamline, predominantly pions, kaons and muons, after the collision of protons with the graphite target. Depending on the detector configuration, positively or negatively charged particles are focused along the beamline by two magnetic focusing horns. These charged particles propagate through the decay pipe and decay to neutrinos. Any remaining hadrons and muons that did not decay are attenuated by the NuMI beam hadron absorber to produce a beam of neutrinos beyond this. The diagram is not to scale and the size of the target is exaggerated for clarity (from Ref. [27]).

Figure 3.2 shows the production of a cascade of particles following the collision of 120 GeV protons onto the graphite target. The interaction of the protons with the graphite produces a secondary beam of hadrons, primarily pions and kaons which are focused by two magnetic van der Meer horns. These horns allow the MicroBooNE detector to operate in two modes, neutrino mode, called Forward Horn Current (FHC) mode, to produce a beam of neutrinos, and anti-neutrino mode, called Reverse Horn Current (RHC) mode, to produce a beam of anti-neutrinos. In neutrino mode, a positive current (+200 kA) is applied to the horns to focus positively charged particles in the beamline direction and in anti-neutrino mode, a negative current (-200 kA) is applied to the horns to focus negatively charged particles. The focused beam of positively (negatively) charged particles then enters the decay pipe and propagates a distance of 675 m in a helium environment to decay to neutrinos (anti-neutrinos).

The beam is terminated by a hadron absorber as shown in figure 3.3, which is a large structure with dimensions 5.6 m (height) \times 5.5 m (width) \times 8.5 m (length) made of an aluminium and steel core that is shielded by steel and concrete blocks. According to the NuMI beam simulation, the beam reaching the NuMI absorber has

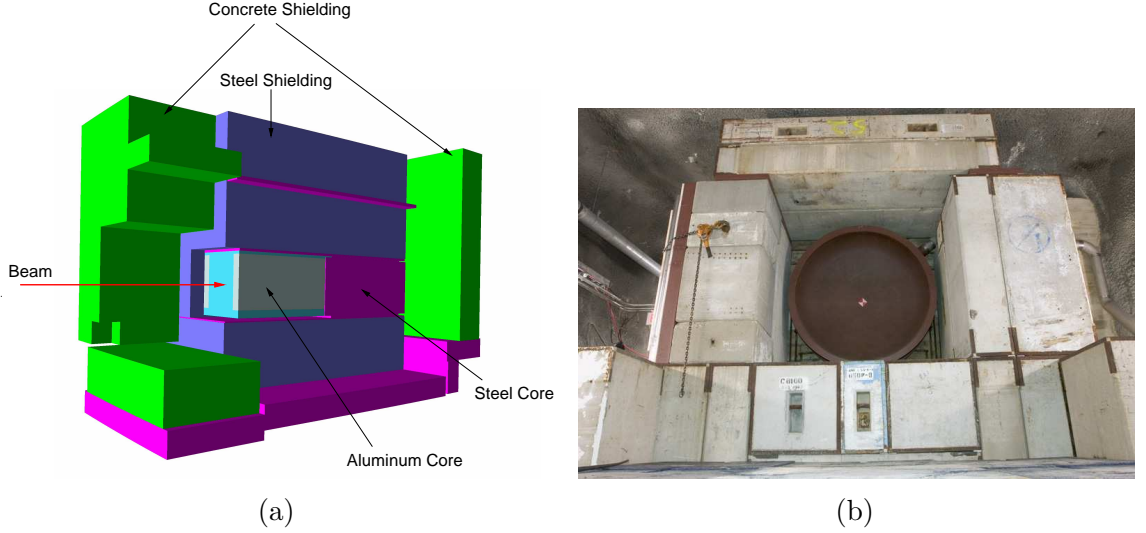


Figure 3.3: (a) A schematic of the NuMI beam hadron absorber with the direction of the beam represented by a red arrow pointing towards the aluminium core (in gray) and steel core (in pink) that is shielded by steel (in purple) and concrete blocks (in green). (b) A photograph of the NuMI beam hadron absorber before the full installation showing the concrete blocks shielding the downstream end of the decay pipe (from Ref. [28]).

$\sim 80\%$ protons that did not interact, $\sim 16\%$ mesons that did not decay in the decay pipe and $\sim 4\%$ electrons, neutrons and photons [35]. Following the large concrete structure is a muon shield, which is a 240 m solid dolomite rock between the absorber and the MINOS ND hall to range out the remaining muons in the NuMI beam to produce a beam of predominantly muon neutrinos [28].

The NuMI beam can be configured to operate in different energy configurations. During the course of its operation, it has run in two different energy modes: Low Energy (LE) mode from 2005 to 2012 and Medium Energy (ME) mode from 2012 until now. These modes are created by configuring the separation between the two horns, the magnitude and the polarity of the magnetic fields and the position of the target from the horns. The MicroBooNE experiment has collected data in the ME mode of the NuMI beam. In this mode, the target is positioned further upstream and the second magnetic horn is positioned further downstream relative to the first magnetic horn than in LE mode [28].

3.1.1 Neutrino and Anti-neutrino Modes

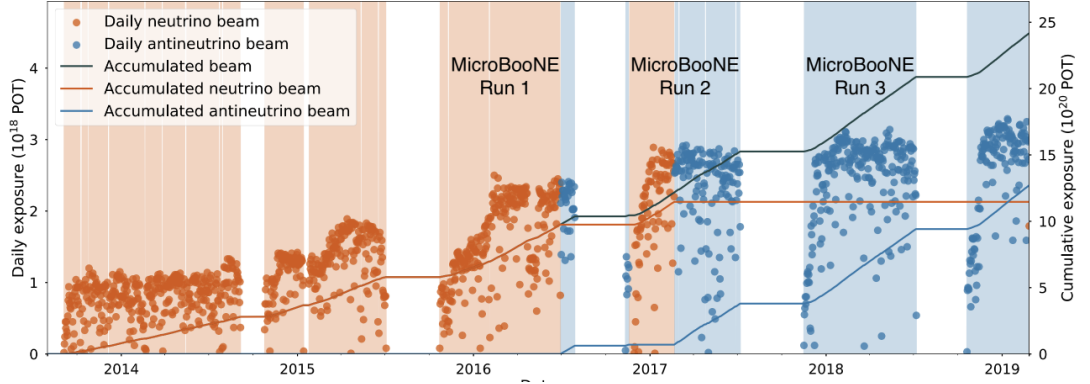


Figure 3.4: The daily and total cumulative POT delivered by the NuMI beam. The neutrino mode or the FHC detector configuration is represented by the orange colour, the anti-neutrino mode or the RHC detector configuration is represented by the blue colour and maintenance periods where the accelerator complex was shut down are represented by white spaces (from Ref. [36]).

As discussed in section 3.1, the NuMI beam operates in two modes, the neutrino, or FHC, mode and the anti-neutrino, or RHC, mode. Figure 3.4 shows the dates and durations that the NuMI beam operated in these modes along with the daily cumulative Protons On Target (POT) delivered by the NuMI beam. This thesis focuses on the FHC and RHC data collected during the Run 1 and Run 3 MicroBooNE data-taking periods and does not include Run 2 as it has not yet been processed by the collaboration.

3.1.2 MicroBooNE Proton Delivery Structure

As shown in figure 3.5, the beam of neutrinos, or the neutrino spill, detected at MicroBooNE is composed of six booster batches, each $1.6 \mu\text{s}$ long that are further subdivided into 84 buckets (or bunches) corresponding to the harmonic number¹ of the booster. The rise time of the booster extraction kicker² to extract the beam

¹The harmonic number is defined as the ratio of the frequency of the accelerating structure to the frequency of the revolution. In other words, the harmonic number is the maximum number of beam bunches a synchrotron can accelerate. A harmonic number of unity would mean that only once per revolution would the RF voltage be at the correct level to accelerate the protons.

²The extraction kickers are dipole magnets that are used to extract the entire particle beam and empty the synchrotron.

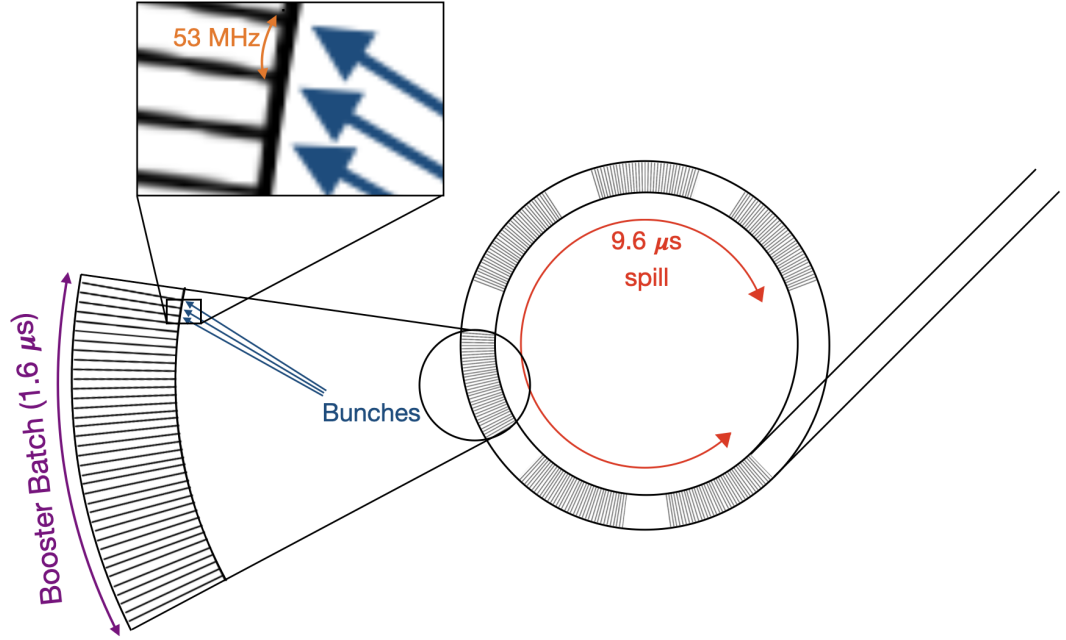


Figure 3.5: A schematic illustrating the meaning of bunch, batch and spill. The schematic is not to scale and the number of bunches and the spacing between the batches have been reduced for clarity.

from booster and inject it into the Main Injector is about 70 ns long and to reduce the losses at extraction, a gap is created in the beam structure by removing three bunches with a dedicated kicker [37]. A batch therefore comprises a total of 81 buckets occupied with protons forming the proton bunches containing 5.0×10^{12} protons.

The circumference of the Fermilab Main Injector is seven times larger than the booster and therefore the Main Injector has a capacity to store and accelerate seven booster batches. However, only six booster batches are injected into the Main Injector to allow the extraction kicker magnets to ramp up in the empty slot. The timing structure of the NuMI beam is therefore six times ($9.6 \mu\text{s}$) the BNB timing structure ($1.6 \mu\text{s}$) comprising a single booster batch.

To enhance the proton intensity per batch, the NuMI Injector later employed a technique called slip-stacking that enabled the synchrotron to combine two batches into one. If X batches are combined with the original 6 batches then the slip-stacking configuration is represented in the form $X + 6$. The NuMI beam was in the $4 + 6$

configuration during the Run 1 period of data taking and later in the same period (described in section 5.8.3) transitioned to the 6 + 6 configuration and remained in this configuration. The data used in this thesis was collected in both the 4 + 6 and 6 + 6 NuMI beam configurations with each spill corresponding to approximately 5.0×10^{13} POT and 6.0×10^{13} POT respectively.

3.2 The NuMI Beam Simulation

The NuMI simulation model has been frequently developed and upgraded by a number of experiments operating under the NuMI beamline. The model simulates collision of the protons on the target and the resulting decay products, hadrons and muons, as well as their subsequent decay to neutrinos. The propagation of these particles in the beamline geometry and their decays is modelled by the **GEANT4**³ software framework described in section 5.2.

In our analysis, we have used the **g4numi** beamline simulation to model the hadron production and beamline geometry. The **g4numi** simulation is derived from an earlier beamline simulation model, **g4numi_flugg**, wherein the **FLUKA** software framework was used to model the hadron production along with the **GEANT4** package to model the beamline geometry. The **g4numi_flugg** software was updated because of its incompatibility with the software framework used to constrain the flux prediction. The new **g4numi** beamline simulation has been employed by all the experiments using the NuMI beam including MicroBooNE, MINERvA and NOvA.

The output files produced from the **g4numi** beamline simulation are known as **dk2nu** files. These files contain information about the neutrino parent particle, which includes kinematics, location of the decay, and parent particle type, as well as the energy of the daughter neutrino in the centre of mass frame of its parent meson decay. The files are generic and could be used by any detector and volume to efficiently calculate the neutrino flux and flux systematics.

³**GEANT4** version **geant_4_2.p03** is used.

3.3 NuMI Neutrino Flux at MicroBooNE

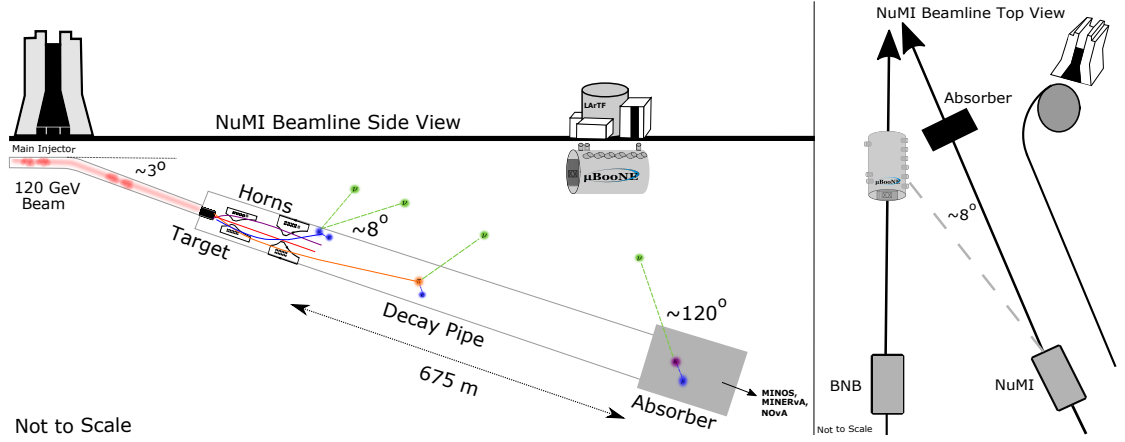


Figure 3.6: A schematic showing side and top views of the NuMI beamline and its orientation relative to the MicroBooNE detector. The beam of protons with energy of 120 GeV hit the target and produce hadrons and muons that either decay at the target or in flight along the decay pipe or at the absorber. The flux of neutrinos can approach the detector at angles ranging from 8° at target to 120° at the absorber relative to the direction of the NuMI beam. Figure from Ref. [27].

The flux at MicroBooNE is constrained using a package called Package to Predict Flux (PPFX). This package implements constraints on the hadron production and propagation using the data collected from the other experiments in the NuMI beam. In addition it takes into account the propagation of the uncertainties for the NuMI beamline simulation. The package was originally developed for the MINERvA experiment and has now been employed by other experiments using the NuMI beamline such as MicroBooNE, NOvA and MINOS+. The package was primarily developed from studies of two different experiments measuring meson and nucleon production and absorption cross-sections, NA49 with a thick (two interaction length) carbon target and MIPP with a thin (less than two interaction length) carbon target. The constrained flux predictions from these two measurements were then compared to the flux calculated using an *in situ* method from the study of neutrino interactions by the MINERvA experiment. The results showed a better agreement between the flux constrained using the thin target [38] compared to the thick target and therefore in this thesis we are using the flux constrained using the thin carbon target.

MicroBooNE is off-axis (vertically and horizontally) with respect to the NuMI beamline as shown in figure 3.6. The flux of neutrinos at MicroBooNE varies

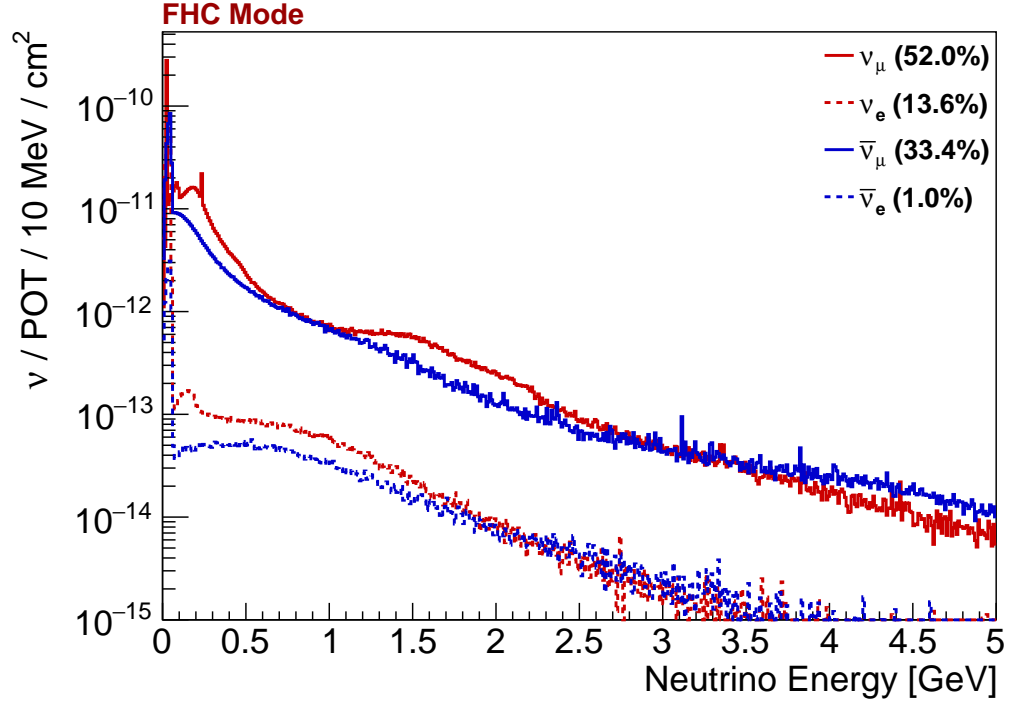
depending on the location and the angle at which neutrinos were produced along the NuMI beamline. The neutrinos produced at the target reach the detector at an angle of about 8° with respect to the beamline, whereas neutrinos from absorber reach the detector backwards at an angle 120° with respect to the beamline.

This section outlines the central-value flux prediction for the NuMI beam at MicroBooNE (described in section 3.3.1) and then breaks this flux prediction into its parent types for each neutrino flavour in the FHC and RHC beams configurations (described in section 3.3.3).

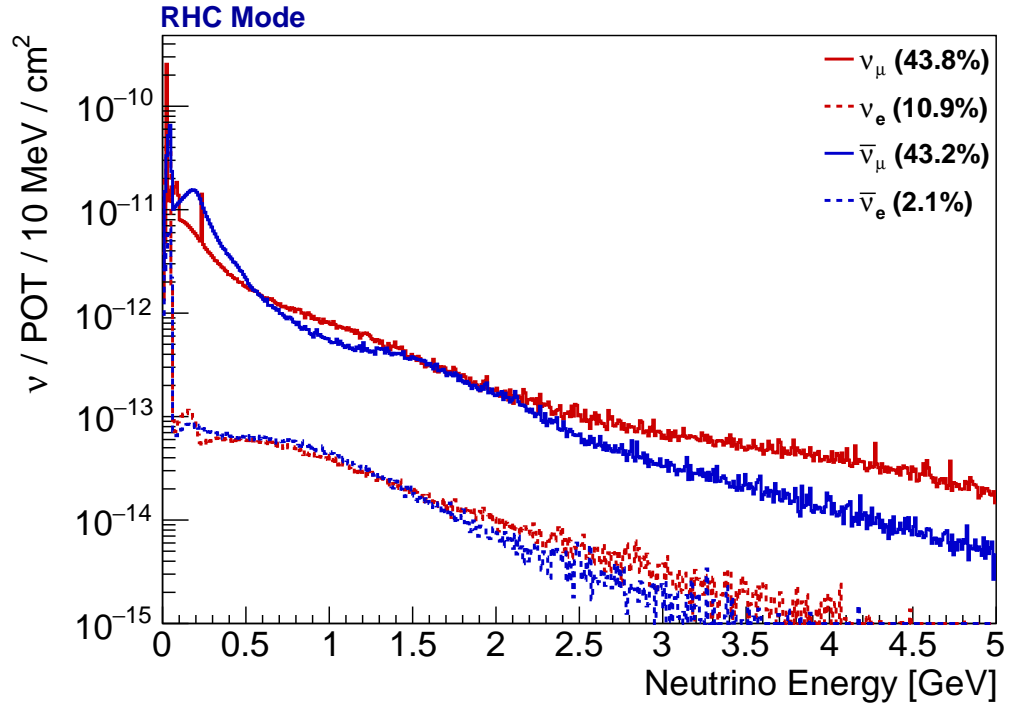
3.3.1 NuMI Central-Value Flux Prediction

Figures 3.7(a) and 3.7(b) show the central-value flux predictions for the NuMI beam at MicroBooNE constrained using the PPFX package in the `g4numi` beam simulation for the FHC and RHC beam configurations respectively. Across all neutrino energies, the majority of the flux is due to muon flavour neutrinos and anti-neutrinos for both the FHC and RHC configurations. The single-bin sharp peaks in the muon neutrino flux at $E_\nu = 29.8$ and $E_\nu = 236$ MeV are due to two-body decays at rest from parent pions and kaons respectively.

Figures 3.8(a) and 3.8(b) show the central-value flux predictions for the FHC and RHC configurations as a function of neutrino angle with respect to the NuMI beam direction. The contribution to the flux from neutrinos produced from the decays of mesons at the target and at the NuMI absorber is larger compared to the flux of neutrinos produced from the decays of mesons focused by the horns into the decay pipe. As a result, the neutrino and anti-neutrino flux predictions are similar for the two horn polarities in figure 3.7 with a slight enhancement of the intentionally favoured neutrino type. Ideally, one expects more neutrinos in FHC mode and anti-neutrinos in RHC mode as discussed in section 3.1.1 but due to the location of the MicroBooNE detector relative to the NuMI beamline, neutrinos from unfocused, undesired oppositely-charged mesons also contribute to the flux. This is also the reason for the low average energy of neutrinos observed at MicroBooNE

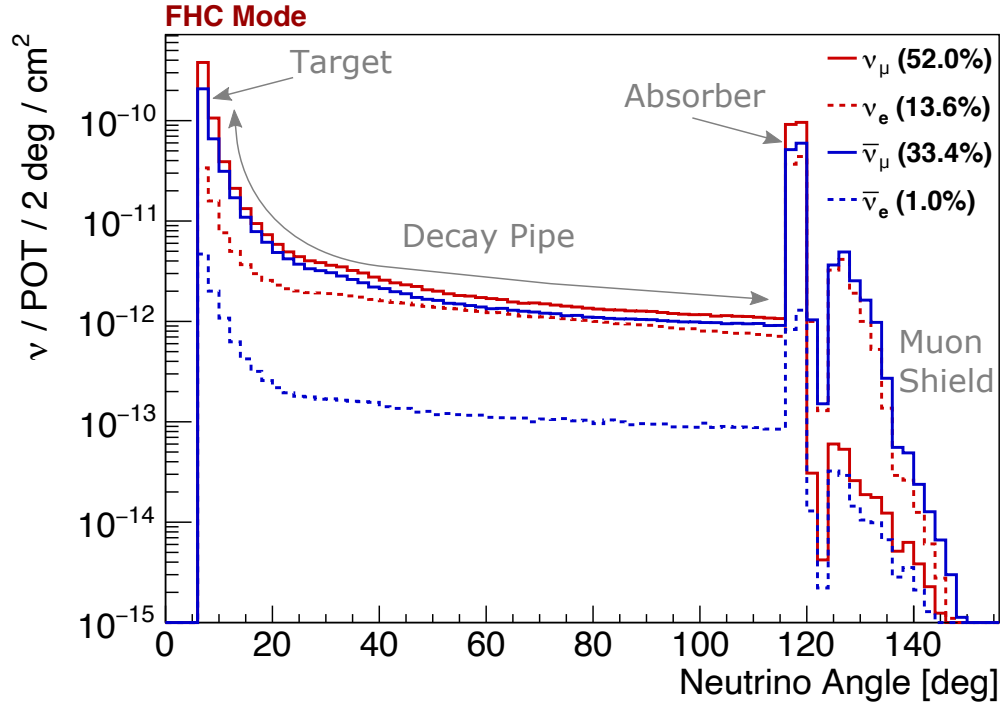


(a)

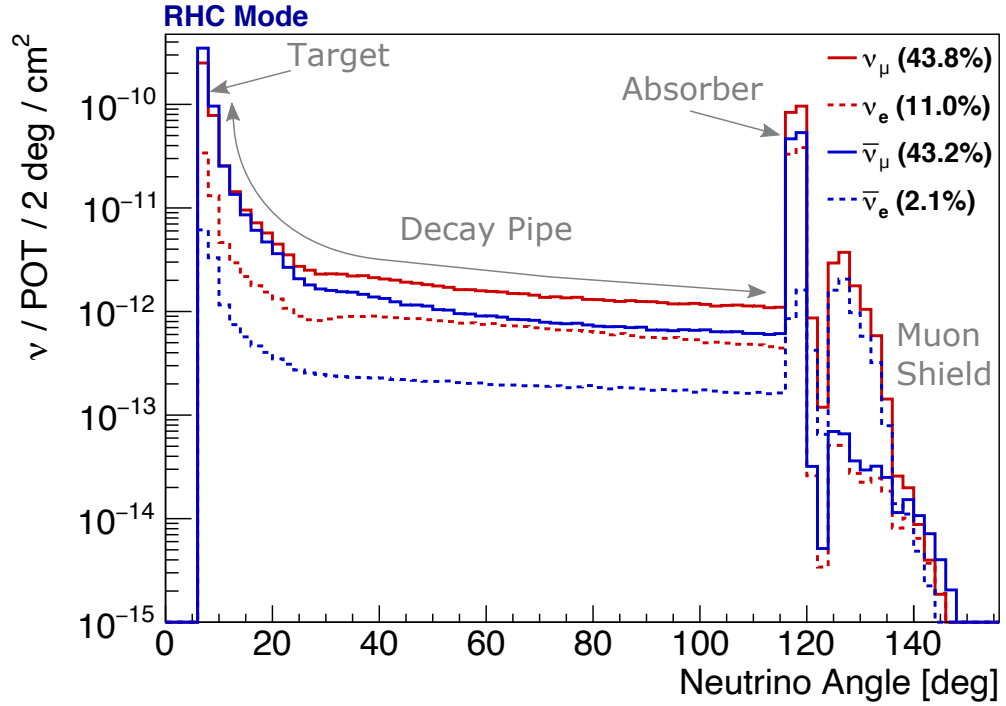


(b)

Figure 3.7: The central-value flux predictions for the NuMI beam at MicroBooNE constrained using PPFX package in the `g4numi` beam simulation for the FHC (a) and RHC (b) beam configurations. The legend shows the integrated proportions of the flux from each observed neutrino flavour (from Ref. [27]).



(a)



(b)

Figure 3.8: The central-value flux predictions for the NuMI beam at MicroBooNE constrained using PPFX package in the `g4numi` beam simulation for the FHC (a) and RHC (b) beam configurations as a function of neutrino angle. The legend shows the integrated proportions of the flux from each observed neutrino flavour (from Ref. [27]).

compared to other experiments along the NuMI beamline.

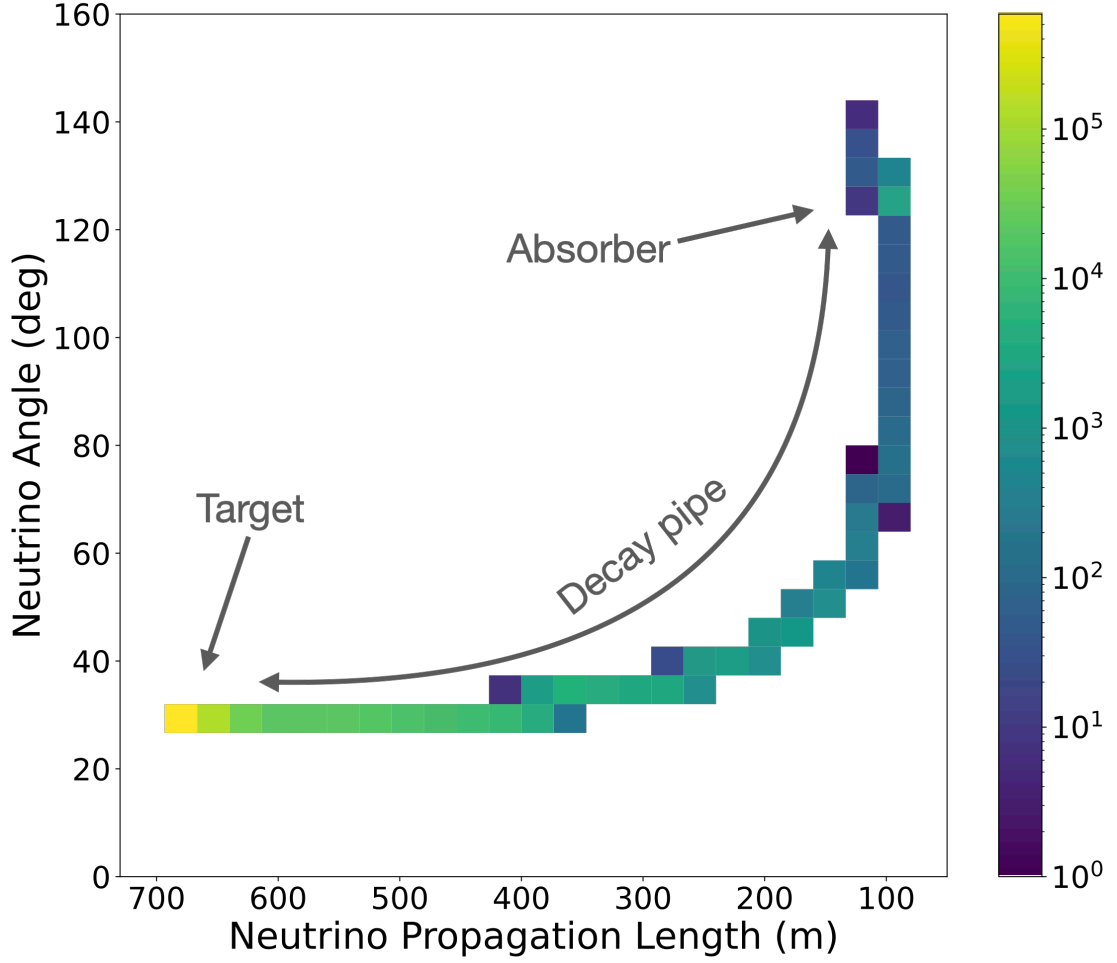


Figure 3.9: The relationship between the neutrino propagation length and the angle subtended by its trajectory with respect to the NuMI beamline. The fraction of neutrinos produced at the target is largest compared to the decay pipe and NuMI absorber. A neutrino produced at the target propagates a distance of ≈ 700 m before reaching the detector, whereas a neutrino produced at the NuMI absorber propagates a distance of only ≈ 100 m. The colour scale represents the number of entries in each bin. This figure is made for all available flavours of neutrinos and anti-neutrinos produced in the FHC configuration of the NuMI beam with a total of 2.33×10^{21} POT.

Figure 3.9 shows the relationship between the distance travelled by the neutrinos in FHC configuration from the point of production (decay position of the neutrino parents) along the NuMI beamline to the detector, and the angle subtended by its trajectory with respect to the NuMI beamline. As expected, the vast majority of neutrinos are produced from mesons decaying at target where neutrinos propagate a distance of around 700 m. The neutrinos produced at the NuMI absorber only travel a distance of around 100 m before reaching the detector at an angle of around 120° .

A wide angular range as a function of the distance travelled is observed for neutrinos produced in the decay pipe of length ~ 700 m (as described in section 3.1).

3.3.2 NuMI Central-Value Flux Prediction in Neutrino Energy and Angle

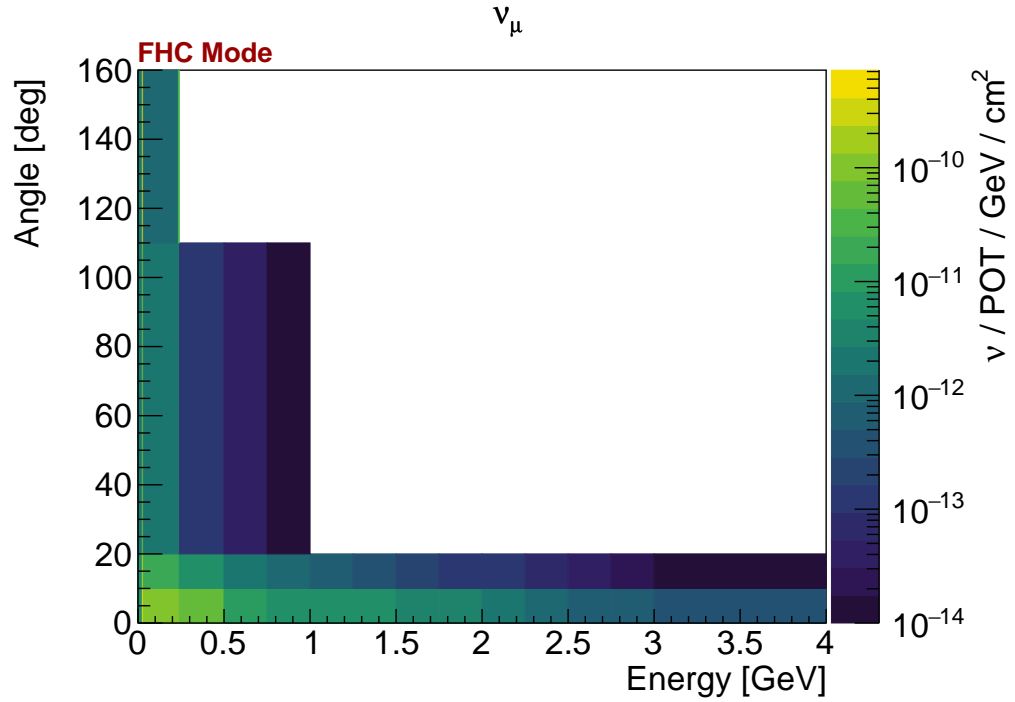


Figure 3.10: The central value ν_μ flux in the FHC beam configuration as a function of neutrino energy and angle with respect to the NuMI beamline (from Ref. [27]).

Figure 3.10 shows the central-value flux prediction of muon-neutrinos in the FHC beam configuration as a function of energy and angle with respect to the NuMI beamline. As can be seen, the two variables discussed in the section 3.3.1, the energy of the neutrino and its direction are highly correlated. The majority of high energy neutrinos (> 250 MeV) approach the detector at small angles relative to the NuMI beam direction, where the neutrinos are produced from the decays of mesons occurring at or just after the target. This is due to high energy neutrinos being more forward boosted compared to less forward boosted low energy neutrinos that approach the detector at larger angles.

The low energy neutrinos (< 250 MeV) approach the detector across a wide

range of angles with the vast majority being neutrinos produced around the target followed by the NuMI absorber and finally the decay pipe. The main source of low energy neutrino production at the NuMI absorber is due to kaons (produced from the collision of protons with the NuMI absorber as described in section 3.1) decaying at rest. Not only neutrinos but a large isotropic flux of potential BSM particles such as the signal of the analysis presented in this thesis, Higgs portal scalars, can also be produced at the NuMI absorber. The main source of low-energy neutrino production along the decay pipe is muons which can travel large distances before decaying to neutrinos. The position of the NuMI absorber and the decay pipe relative to the MicroBooNE detector can be seen in figure 3.6.

3.3.3 NuMI Central Value Flux Prediction by Neutrino Parent

Neutrino (Anti-neutrino) Production Channels	
Decay modes	Branching Ratio (%)
$\pi^\pm \rightarrow \mu^\pm + \nu_\mu(\bar{\nu}_\mu)$	99.9877
$\pi^\pm \rightarrow e^\pm + \nu_e(\bar{\nu}_e)$	0.0123
$K^\pm \rightarrow \mu^\pm + \nu_\mu(\bar{\nu}_\mu)$	63.55
$K^\pm \rightarrow \pi^0 + e^\pm + \nu_e(\bar{\nu}_e)$	5.07
$K^\pm \rightarrow \pi^0 + \mu^\pm + \nu_\mu(\bar{\nu}_\mu)$	3.353
$K_L^0 \rightarrow \pi^\mp + e^\pm + \nu_e(\bar{\nu}_e)$	40.55
$K_L^0 \rightarrow \pi^\mp + \mu^\pm + \nu_\mu(\bar{\nu}_\mu)$	27.04
$\mu^\pm \rightarrow e^\pm \nu_e(\bar{\nu}_e) + \bar{\nu}_\mu(\nu_\mu)$	100.0

Table 3.1: The branching ratios for various key decay modes responsible for the production of the neutrinos (anti-neutrinos) at MicroBooNE in the NuMI beam [39].

The neutrinos at MicroBooNE are mainly produced from the decays of muons, pions and kaons. Table 3.1 shows the branching ratios for various key decay modes responsible for the production of the neutrinos (anti-neutrinos) at MicroBooNE in the NuMI beam. Using the helicity suppression argument, the production of muon neutrinos are highly favoured over the production of electron neutrinos and therefore the beam predominantly contains muon neutrinos.

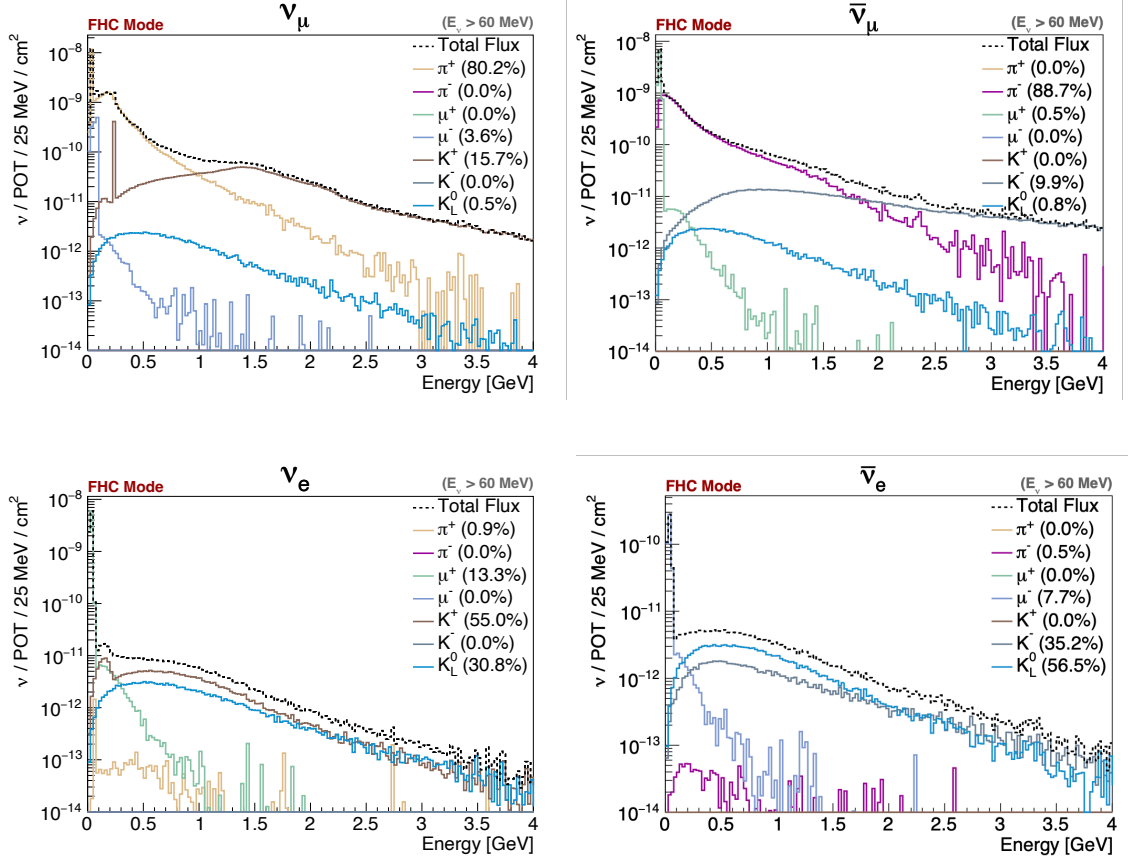


Figure 3.11: The central-value flux prediction with fractional contributions of the neutrino parents responsible for the production of neutrinos and anti-neutrinos in the FHC configuration of the NuMI beam. An energy threshold of $E_\nu > 60$ MeV is applied to exclude the flux due to low-energy muon decays (from Ref. [27]).

Figure 3.11 shows the fractional contributions of the neutrino parents to the total central-value flux prediction of neutrinos and anti-neutrinos in the FHC mode of the NuMI beam, for neutrinos with $E_\nu > 60$ MeV to exclude the flux due to low-energy muon decays. The vast majority of muon neutrinos (anti-neutrinos) are produced from the decays of positively (negatively) charged pions followed by the decays of positively (negatively) charged kaons. The contribution of the kaons begins to dominate at an energy of ~ 1 GeV in the case of the muon neutrino flux and ~ 2 GeV in the case of the muon anti-neutrino flux.

As already seen in table 3.1, the branching ratio of charged pion decays to electron neutrinos is significantly smaller than to muon neutrinos and therefore the contribution of the pion decays to the electron neutrino and anti-neutrino flux is negligible. The second and third most dominant decays responsible for the production of electron neutrinos and anti-neutrinos are kaon and muons decay respectively. The

contribution of the muons decaying at rest to the electron neutrino and anti-neutrino flux is significant up to ~ 60 MeV.

3.4 Higgs Portal Scalar Flux at MicroBooNE

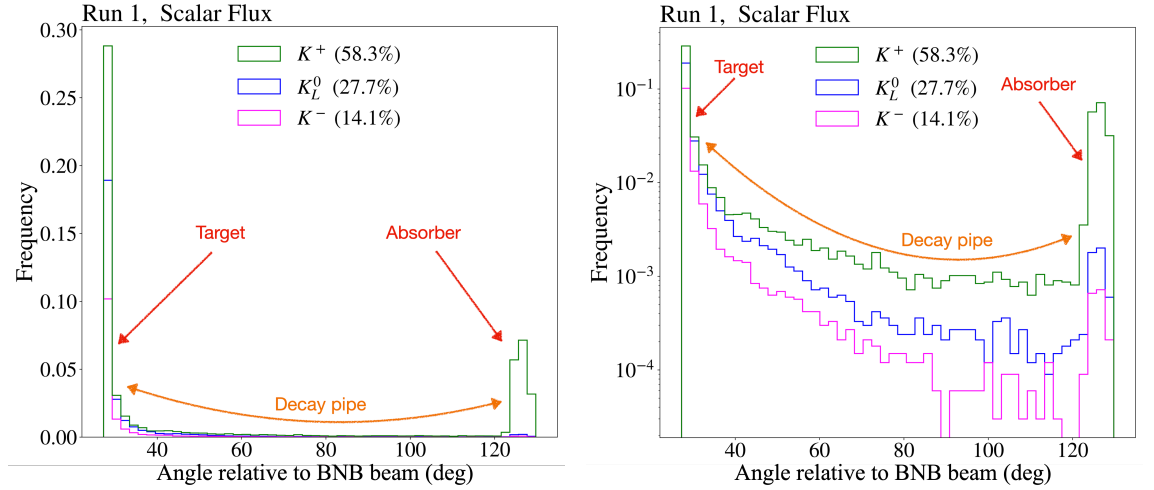


Figure 3.12: The probability density of the flux prediction with fractional contributions of the Higgs portal scalar parent responsible for the production of scalars in the FHC configuration of the NuMI beam. The distributions on left and right show the probability density in linear-scale and log-scale respectively.

The signal of Higgs portal scalars is produced from the decays of kaons at all locations along the NuMI beamline, including KDAR at the target, KDIF along the decay pipe and KDAR in the NuMI absorber. The generation and simulation of these scalars is described in section 5.1.1. Figure 3.12 shows the probability density of the flux prediction with fractional contributions of the Higgs portal scalar parent responsible for the production of scalars in the FHC configuration of the NuMI beam. The distributions on the left and right show the probability density in linear-scale and log-scale respectively. The dominant mode of scalar production along the entire NuMI beamline is from the decays of K^+ with a contribution of approximately 60% to the total scalar flux. It is also the dominant decay mode of scalar production at the NuMI absorber. The second most dominant mode of scalar production along the NuMI beamline is from the decays of K_L^0 with a contribution of about 28%, followed by K^- with contribution of about 14% to the total scalar flux.

Figure 3.13 shows the probability density of the flux prediction with fractional

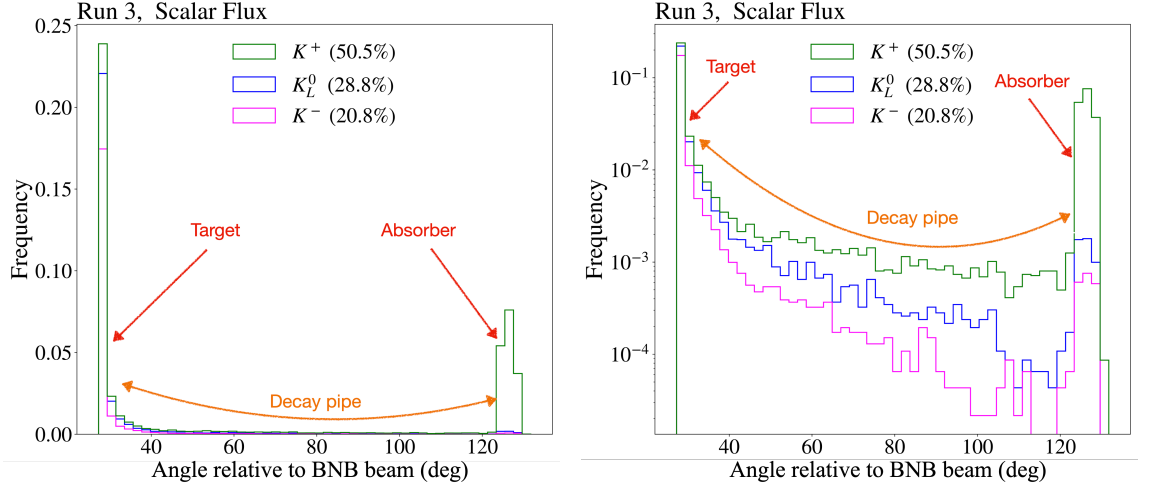


Figure 3.13: The probability density of the flux prediction with fractional contributions of the Higgs portal scalar parent responsible for the production of scalars in the RHC configuration of the NuMI beam. The distributions on left and right show the probability density in linear-scale and log-scale respectively.

contributions of the Higgs portal scalar parent responsible for the production of scalars in the RHC configuration of the NuMI beam. The distributions on the left and right show the probability density in linear-scale and log-scale respectively. As discussed in section 3.1.1, for Run 3, the beam operates in RHC mode and thus the contribution of the scalars produced from the decays of K^- is enhanced by about 50% compared to Run 1. Scalar production from the decays of K^+ is reduced with almost no change in the contribution of the charge-less K_L^0 . For both Run 1 and Run 3, the vast majority of the scalars are produced from the decays of kaons at target, followed by the NuMI absorber.

3.4.1 Flux prediction for KDAR from the NuMI Absorber

The **g4numi** beamline simulation discussed in section 3.2 has a default production rate for muon neutrinos produced from KDAR at the NuMI absorber of $0.011 \nu_\mu$ per POT in FHC mode and $0.0099 \nu_\mu$ per POT in RHC mode. However measurements from the MiniBooNE experiment [40] for the muon-neutrino flux from KDAR at the absorber does not agree with the MicroBooNE **g4numi** simulation production rate. The predictions from MiniBooNE comprise production rates estimated using a range of models including MARS [41], FLUKA [42] and GEANT4 [43, 44]. The rates derived

using these models vary from 0.06 to 0.12 KDAR ν_μ per POT. The final production rate accepted and employed by the MiniBooNE experiment was a rate derived using the GEANT4 simulation model that is equivalent to 0.085 KDAR ν_μ per POT.

The derived rate is significantly larger $\mathcal{O}(10)$ than the central value rate used in the MiniBooNE GEANT4 simulation and is well outside of the range of rates derived using different simulation models. If the production rate of `g4numi` was taken to be the true rate then final measured MiniBooNE cross-section would be an order of magnitude larger than the theoretical predictions. The cause of this discrepancy is unknown and is being investigated by the `g4numi` beamline simulation authors. By adopting the production rates derived by the MiniBooNE collaboration, the events with parent particles being kaons decaying at rest at the NuMI absorber are re-weighted by factors of 8.0 and 8.6 for the FHC and RHC modes respectively. In the analysis, this re-weighting is applied to events (associated to KDAR at the NuMI absorber) in signal as is done in the previous two MicroBooNE Higgs portal scalar searches [12, 13] as well as to background events.

Chapter 4

Particle Interactions in MicroBooNE

This chapter is devoted to understanding the interactions of the particles produced in the MicroBooNE detector. Section 4.1 covers scattering processes that contribute to the cross-sections of the neutrinos and anti-neutrinos across a wide neutrino energy range. Section 4.2 then covers various nuclear effects that complicate the neutrino interaction processes discussed in section 4.1. Section 4.3 describes the interactions of various particles produced in the MicroBooNE LArTPC such as pions, protons, neutrons, muons, electrons and photons. Finally section 4.4 outlines various detector effects that modify the cloud of ionisation electrons produced by the propagation of a charged particle in a LArTPC.

4.1 Neutrino Interactions in MicroBooNE

Although neutrino measurements such as neutrino oscillation and cross-section measurements are not the aim of the analysis presented in this thesis, these measurements are the principal goals of the Fermilab Short-Baseline Neutrino (SBN) program. These measurements encompass many essential aspects of detector design and operation that contribute significantly in search strategy employed in this thesis.

In addition, these measurements involve the understanding of the interactions of a neutrino with a target nucleus, nucleon or lepton via Charged Current (CC) or Neutral Current (NC) weak interactions. The CC interactions are mediated by a W^\pm boson whereas NC interactions are mediated by a Z^0 boson. The neutrino interactions are a primary source of background in the Higgs portal scalar search presented in this thesis. Therefore, this section outlines the physics of neutrino interactions with the argon atom relevant for the MicroBooNE LArTPC.

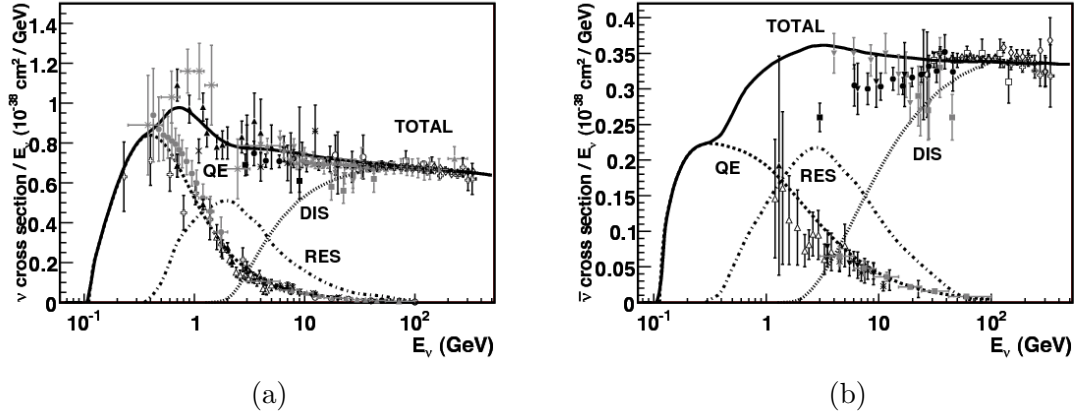


Figure 4.1: The total neutrino (a) and anti-neutrino (b) charged current cross-sections divided by energy of the neutrino, plotted as a function of this energy. The theoretical predictions provided by the NUANCE neutrino physics simulation [45] for QE, RES and DIS interactions are shown in dashed, dot-dashed and dotted lines respectively. The figures as well as the details of the data points from various experiments overlaid over the theoretical predictions can be found in Ref. [46].

Figure 4.1 shows various processes that contribute to the cross-section of neutrino interactions across a wide neutrino energy range of a few hundred MeV to several GeV. These processes include Quasi-Elastic (QE) scattering, Resonance production (RES) and Deep Inelastic Scattering (DIS). The cross-sections of these processes are functions of various different parameters including the energy of the incident neutrino. At low energies, from 100 MeV to 1.5 GeV, the QE process dominates, followed by RES processes up-to a few GeV, and at high energies DIS becomes the dominant process. The cross-section for the anti-neutrino is approximately a third of that of the neutrino due to helicity suppression.

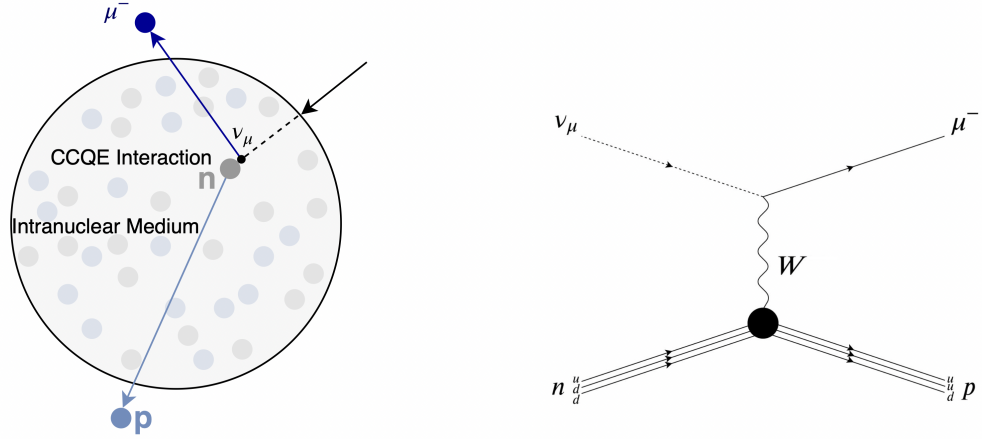


Figure 4.2: The diagram on left shows an example of a CCQE interaction: $\nu_\mu + n \rightarrow \mu^- + p$, where an incoming muon neutrino (black arrow) interacts with a neutron (gray) to produce a muon (dark blue arrow) and a proton (light blue arrow) exiting the nucleus without interacting further with the nucleons inside the nuclear medium. The diagram on the right shows the corresponding Feynman diagram for the process. Figure from Ref. [47].

4.1.1 Quasi-Elastic Scattering

In the low energy region from 0.1–1.5 GeV, QE scattering is the dominant neutrino interaction process. The QE scattering processes which are mediated by the charged current W^\pm are called Charged Current QE (CCQE) interactions. The CCQE interactions on a free nucleon are given by:

$$\begin{aligned} \nu_l + n &\rightarrow l^- + p, \\ \bar{\nu}_l + p &\rightarrow l^+ + n, \end{aligned} \tag{4.1}$$

where l is the lepton flavour. Figure 4.2 shows an example illustrating the CCQE interaction of a muon neutrino with a neutron to produce a proton and a lepton. The CCQE interaction is also known as a “1 particle 1 hole (1p-1h)¹” process as the neutrino (anti-neutrino) scatters off a neutron (proton) producing a lepton and a proton (neutron) in the final state.

¹Theoretically, the calculation involves construction of excited states 1p-1h by raising a particle with 4-momentum p above the Fermi level $p > k_F$ and the resultant hole with 4-momentum $h < k_F$ [48].

The QE scattering processes mediated by a neutral boson Z^0 are called Neutral Current QE (NCQE) interactions. The NCQE interactions on a free nucleon are given by:

$$\begin{aligned}\nu_l + p, n &\rightarrow \nu_l + p, n, \\ \bar{\nu}_l + p, n &\rightarrow \bar{\nu}_l + p, n.\end{aligned}\tag{4.2}$$

The final states for CCQE and NCQE interactions are different. The CCQE interaction produces a lepton and a nucleon whereas in an NCQE interaction, the final states are same as the initial states.

4.1.2 Meson Exchange Current

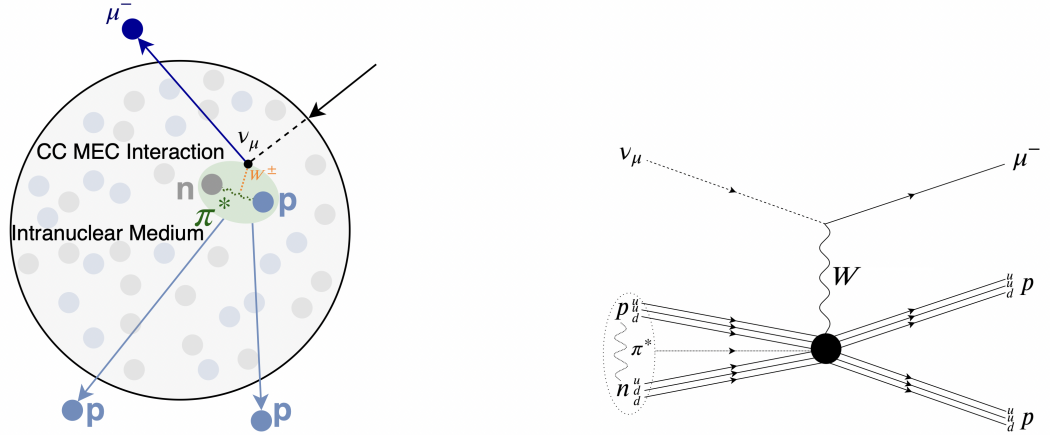


Figure 4.3: The diagram on left shows an example of a CC MEC interaction: $\nu_\mu + 1n1p \rightarrow \mu^- + 2p$, where an incoming muon neutrino (black arrow) interacts with a proton (blue) and a neutron (gray) bound by a virtual meson (green), to produce two protons (light blue arrows) exiting the nucleus. The diagram on right shows the corresponding Feynman diagram for the process. Figure from Ref. [47].

The Meson Exchange Current (MEC) interactions involve the exchange of a virtual meson between the bound states of two or more nucleons within the nucleus. Section 4.2.2 describes the two regions in the nuclear potential that are responsible for forming the bound state of nucleons via an exchange of 1 or 2 pions. Figure 4.3 shows an example diagram illustrating CC MEC interaction for the process $\nu_\mu + 1n1p \rightarrow \mu^- + 2p$ where an incoming muon neutrino interacts with a proton and a neutron bound by a

virtual meson to produce two protons. The significance of this process has only been recently pointed out to explain the enhancement observed in the double differential CCQE cross-section measurements from experiments such as MiniBooNE [49]. This discrepancy in the neutrino interaction cross-section by two-nucleon bound states is called the “2 particle 2 hole (2p-2h)” effect. This effect is important for neutrino scattering in the MicroBooNE energy range of $\mathcal{O}(1 \text{ GeV})$. MEC interactions resemble CCQE interactions but with the emission of two nucleons from the primary vertex, instead of one.

4.1.3 Resonant Scattering

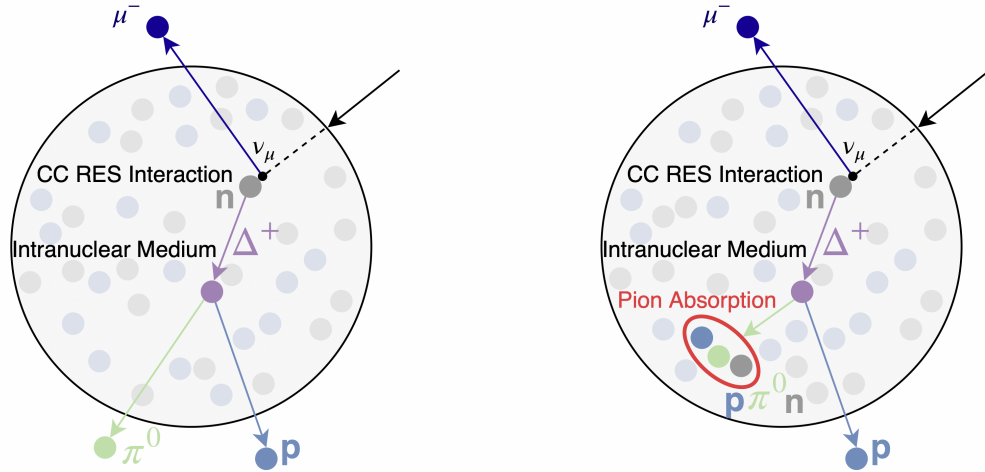


Figure 4.4: Diagrams showing examples of CC RES scattering processes: $\nu_\mu + n \rightarrow \mu^- + p + \pi^0$, where an incoming muon neutrino (black arrow) interacts with a neutron (gray) to produce a muon (dark blue) and a positively charged excited baryonic resonance (purple). The baryonic resonance quickly decays into a proton (light blue) and a pion (light green) inside the nucleus. The diagram on left shows these daughter particles exiting the nucleus without undergoing any further interactions whereas the diagram on right shows the pion being absorbed inside the nucleus producing only a proton in the final state. This type of RES interaction where a pion get absorbed could be misidentified as a CCQE interaction. Figure from Ref. [47].

If the neutrino-nucleon center-of-mass energy exceeds the sum of the masses of a baryon, N^* , and the lepton, l , the neutrino can inelastically scatter off the target nucleon to produce a baryonic resonance, N^* . The resonance quickly decays to produce final state particles, predominantly a nucleon and a single pion, and this interaction process is known as resonant scattering, or resonant pion production

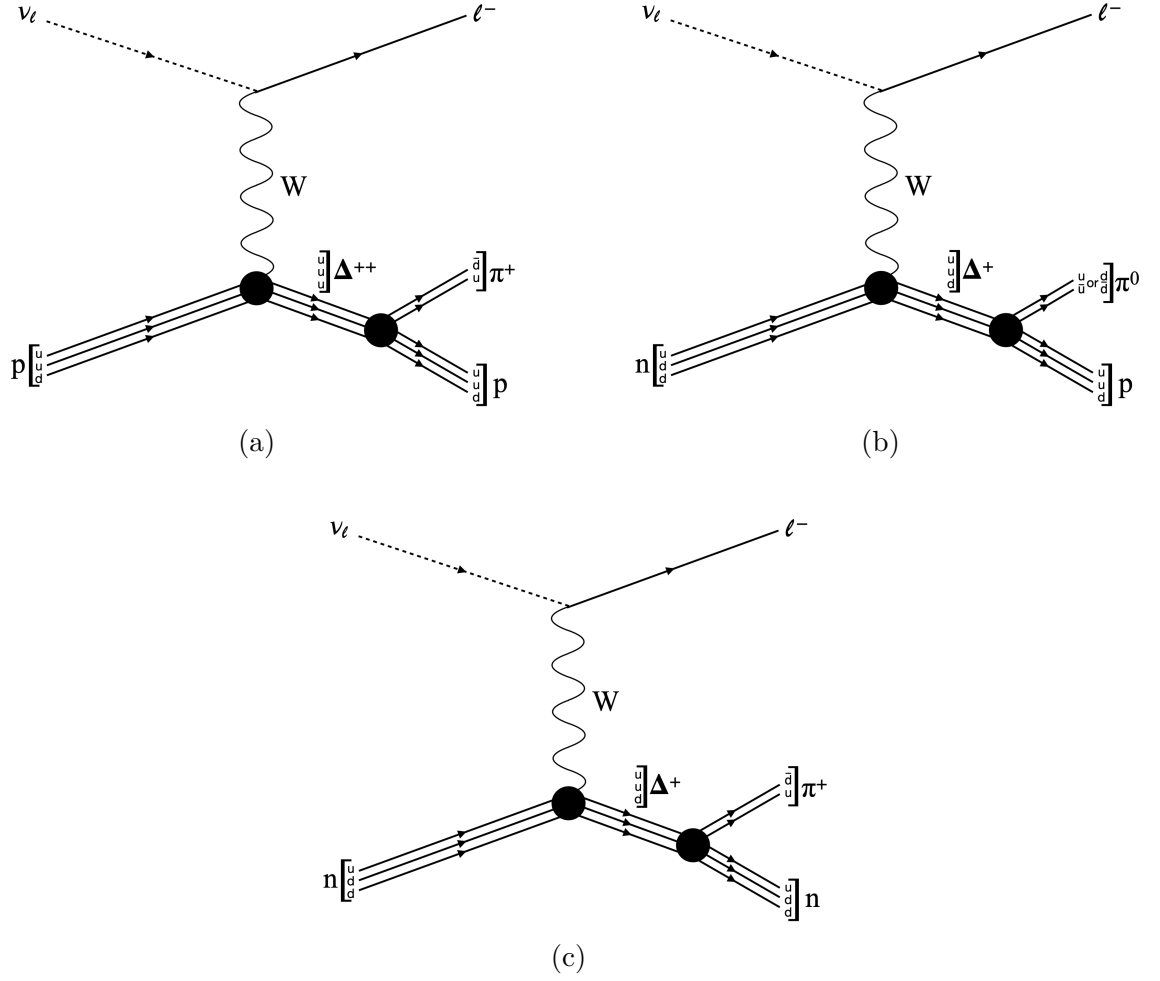


Figure 4.5: The Feynman diagrams for the most common CC RES scattering processes for neutrinos with (a) (b) and (c) showing the processes $\nu_l + p \rightarrow l^- + p + \pi^+$, $\nu_l + n \rightarrow l^- + p + \pi^0$ and $\nu_l + n \rightarrow l^- + n + \pi^+$ respectively.

or single pion production. Figure 4.4 shows an example illustrating the CC RES scattering process $\nu_\mu + n \rightarrow \mu^- + p + \pi^0$, where an incoming muon neutrino interacts with a neutron to produce a muon and a positively charged excited baryonic resonance, which quickly decays into a nucleon and a pion. The pion either exits the nucleus without undergoing any further interactions or is absorbed inside the nucleus. Examples of the most common CC RES scattering processes for neutrinos and anti-neutrinos are:

$$\begin{aligned}
 \nu_l + p &\rightarrow l^- + p + \pi^+, & \bar{\nu}_l + p &\rightarrow l^+ + p + \pi^-, \\
 \nu_l + n &\rightarrow l^- + p + \pi^0, & \bar{\nu}_l + p &\rightarrow l^+ + n + \pi^0, \\
 \nu_l + n &\rightarrow l^- + n + \pi^+, & \bar{\nu}_l + n &\rightarrow l^+ + n + \pi^-.
 \end{aligned} \tag{4.3}$$

The Feynman diagrams for the processes shown in equation 4.3 for neutrinos are shown in figure 4.5.

NC RES scattering processes occur if the neutrino-nucleon center-of-mass energy exceeds only the mass of the baryon. Examples of the most common NC RES scattering decay channels for neutrinos and anti-neutrinos are:

$$\begin{aligned}
 \nu_l + p &\rightarrow \nu_l + p + \pi^0, & \bar{\nu}_l + p &\rightarrow \bar{\nu}_l + p + \pi^0, \\
 \nu_l + p &\rightarrow \nu_l + n + \pi^+, & \bar{\nu}_l + n &\rightarrow \bar{\nu}_l + n + \pi^0, \\
 \nu_l + n &\rightarrow \nu_l + n + \pi^0, & \bar{\nu}_l + n &\rightarrow \bar{\nu}_l + n + \pi^0, \\
 \nu_l + n &\rightarrow \nu_l + p + \pi^-, & \bar{\nu}_l + n &\rightarrow \bar{\nu}_l + p + \pi^-.
 \end{aligned} \tag{4.4}$$

The Feynman diagrams for the processes shown in equation 4.4 for neutrinos are shown in figure 4.6.

4.1.4 Deep Inelastic Scattering

In DIS, a neutrino with energy above several GeV interacts with the sea or valence quarks of a nucleon, instead of a whole nucleon, via the exchange of weak boson. The high-energy neutrinos can resolve the internal structure of the nucleons and thus break the nucleons and produce hadronic jets, X . Figure 4.7 shows an example diagram illustrating the CCDIS process $\nu_l + q \rightarrow l^- + X$ where an incoming muon neutrino interacts with the up and/or down quarks of a nucleon to produce a muon and hadronic jets via the process of hadronisation and other final-state interactions.

4.1.5 Coherent Scattering

In Coherent (COH) scattering, a neutrino interacts coherently with the whole nucleus, A , resulting in the production of a pion. These interactions are only possible at low neutrino energy and therefore a negligible amount of energy is transferred to the nucleus. Examples of the CC and NC neutrino-nucleus COH interactions are:

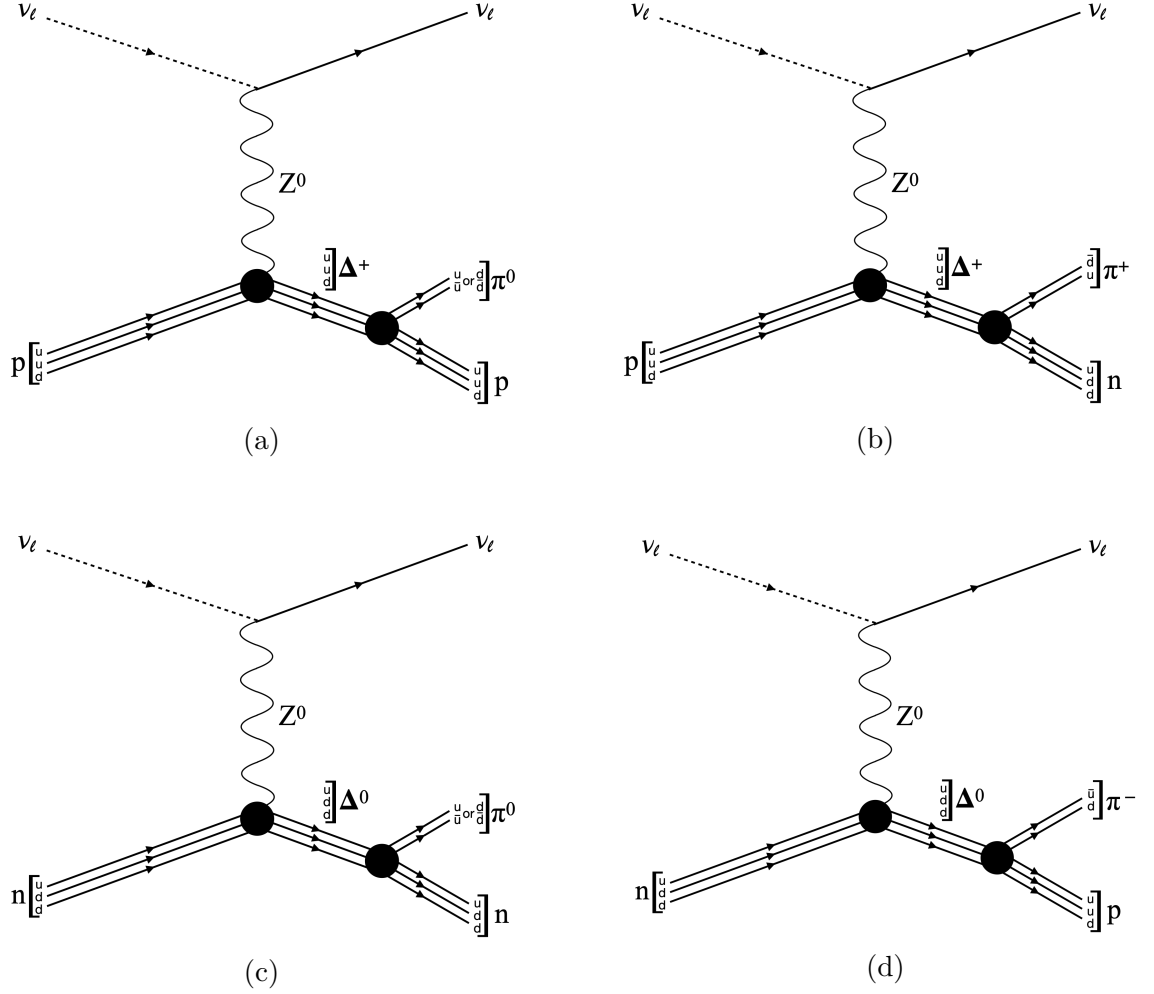


Figure 4.6: The Feynman diagrams for the most common NC RES scattering processes for neutrinos with (a), (b), (c) and (d) showing the processes $\nu_l + p \rightarrow \nu_l + p + \pi^0$, $\nu_l + p \rightarrow \nu_l + n + \pi^+$, $\nu_l + n \rightarrow \nu_l + n + \pi^0$ and $\nu_l + n \rightarrow \nu_l + p + \pi^-$ respectively.

$$\begin{aligned}
 \text{CC : } & \nu_l + A \rightarrow l^- + A + \pi^+, \quad \bar{\nu}_l + A \rightarrow l^+ + A + \pi^-. \\
 \text{NC : } & \nu_l + A \rightarrow \nu_l + A + \pi^0, \quad \bar{\nu}_l + A \rightarrow \bar{\nu}_l + A + \pi^0.
 \end{aligned}
 \tag{4.5}$$

The contribution of the cross-section from the COH scattering is sub-dominant compared to other scattering processes that produce pions such as RES [46]. The NC COH scattering process for the neutrino and anti-neutrino produces a neutral pion which can decay into a pair of photons and mimic the signal $S \rightarrow e^+e^-$ of the study presented in this thesis.

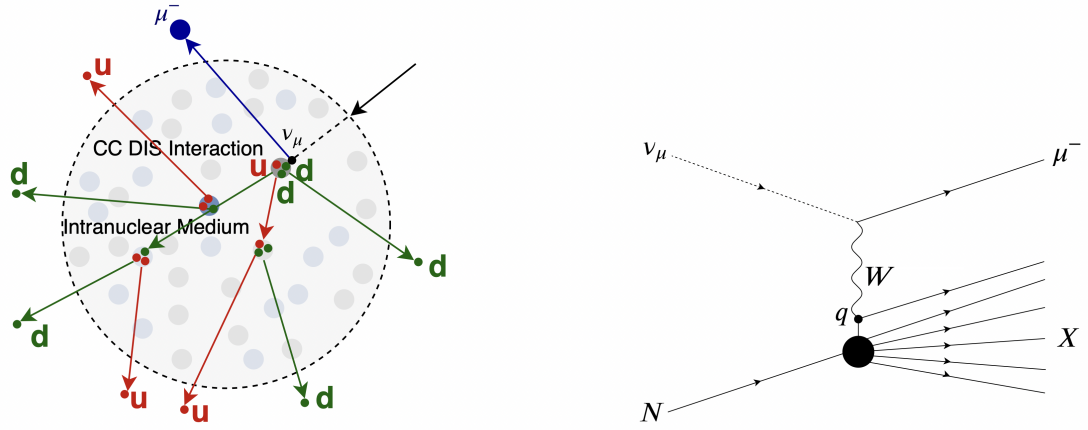


Figure 4.7: The diagram on the left shows an example of a CCDIS process: $\nu_\mu + q \rightarrow \mu^- + X$, where an incoming muon neutrino (black arrow) interacts with the up and/or down quarks of a nucleon (red and green) to produce a muon (dark blue arrow) and hadronic jets (red and green arrows) via the process of hadronisation and other final-state interactions. The diagram on right shows the corresponding Feynman diagram for the process. Figure from Ref. [47].

4.2 Nuclear Effects

The various interaction processes described in section 4.1 assume the neutrino interacts with “free” nucleons inside the argon nuclei. However, unlike hydrogen or deuterium, the nucleus of an argon has an aggregation of 22 neutrons and 18 protons that complicate the nucleus and introduces various nuclear physics processes, called nuclear effects, that break this assumption of “free” nucleons (described in section 4.2.2). The current knowledge of nuclear and particle physics cannot fully describe neutrino interactions in argon as the understanding and modelling of these nuclear effects is non-trivial. The consequence of this is large uncertainties in cross section measurements and in neutrino energy reconstruction due to these effects affecting the kinematics of the particles produced in the neutrino interactions.

In order to model the nuclear effects, the following two assumptions under the impulse approximation are made [47]:

1. The momentum of the incident neutrino is large enough to resolve the target nucleus as a collection of individual nucleons.

2. Propagation of the particles produced at the interaction vertex and the recoiling nucleon system are independent.

This section presents an overview of the two nuclear effects: the initial state effects outlined in section 4.2.1, and the multi-nucleon correlation effects outlined in section 4.2.2.

4.2.1 Initial State Effects

The initial state effects comprise the effects caused by:

- **Nuclear Binding Energy:** The mass difference between the sum of nucleons inside the nucleus and the nucleus itself which is also the minimum amount of energy required to disassemble the nucleus. The binding energy of the nucleus can accelerate and/or decelerate the hadrons produced from the neutrino interaction in the intranuclear medium.
- **Fermi Motion:** The nucleons are in motion inside the nucleus with a Fermi momentum of about 200 MeV [50]. The Fermi motion describes the initial-state momentum distribution of the nucleons before any neutrino-nucleon interactions. For heavy nuclei such as argon, it is extremely difficult to model.

4.2.2 Multi-nucleon Correlation Effects

The assumptions made in the impulse approximations cannot fully describe neutrino interactions with a nucleus as they do not take into account the multi-nucleon correlation effects. Heavy nuclei such as argon can be modelled using a mean-field approximation where a nucleon interacts independently with a potential form by the other nucleons. At short distances, a nucleon will repel another nucleon due to Pauli exclusion principle² [47]. This enables nucleons within the nucleus to propagate relatively large distances without colliding with the other nucleons. This large mean

²Identical fermions cannot occupy the same quantum state within a quantum system.

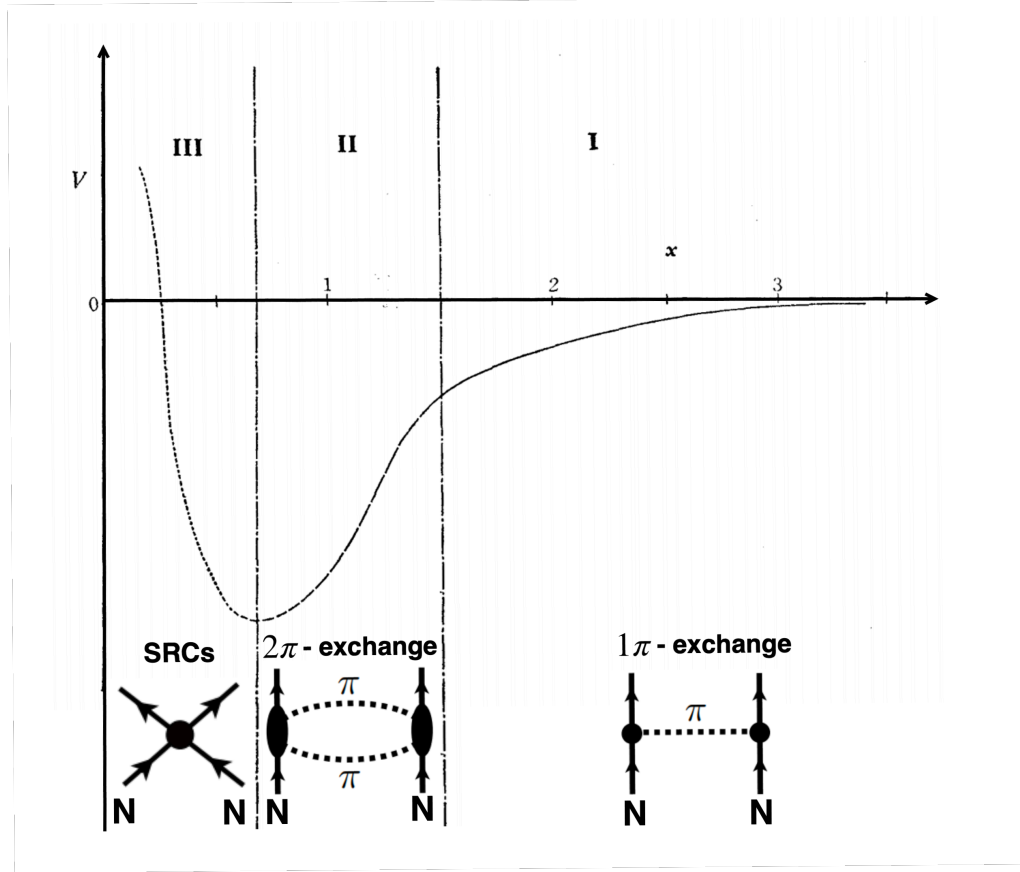


Figure 4.8: A diagram showing three regions of the multi-nucleon potential. The horizontal axis, x , represents the inter-nucleon distance in units of the pion Compton wavelength, $\kappa^{-1} \sim 1.4$ fm. The classical region (I) denotes a 1π -exchange potential. The dynamical region (II) denotes a 2π -exchange potential. Finally, the phenomenological region (III) denotes Short-Range Correlations (SRCs). Figure from Ref. [51].

free path of the nucleons allows one to model nucleons as freely moving particles in the intranuclear medium. This is referred to as the Shell or Independent Particle Model (IPM) which can describe various nuclear properties, such as the nuclear binding energy defined in section 4.2.1, but cannot explain long-range and short-range nucleon-nucleon correlations.

Figure 4.8 shows various scales in units of the pion Compton wavelength $\kappa^{-1} \sim 1.4$ fm that govern the nucleon-nucleon potential. The nuclear potential is split into three regions which are defined as follows:

- **Classical region (I):** For an inter-nucleon distance $x \geq 1.5\kappa^{-1}$, the nucleons exchange one pion and the interaction between two nucleons is attractive in nature [47].

- **Dynamical region (II):** For an inter-nucleon distance $0.7\kappa^{-1} \leq x \leq 1.5\kappa^{-1}$, the nucleons exchange two pions and the interaction between two nucleons is also attractive. The nucleons can also exchange other virtual mesons such as ω or ρ mesons [52].
- **Phenomenological region (III):** For an inter-nucleon distance $x \leq 0.7\kappa^{-1}$, the wave-functions of the two nucleons strongly overlap over short time scales and the mean-field approximation in the IPM (described in section 4.2.2) breaks down [52]. This is the least understood region of the nuclear physics [53].

4.2.3 Final State Particle Interactions in the Argon

Final State Interactions (FSIs) are the re-interactions of particles produced by the neutrino-nucleus scattering inside the nucleus. The re-interactions of these particles can produce other hadrons or knock out nucleons. Particles produced in a neutrino interaction can also be absorbed in the intranuclear medium. An example diagram of a CCQE FSI process is shown in figure 4.9, where an incoming muon neutrino interacts with a neutron (under the first assumption of the impulse approximation) to produce a muon that escapes the nucleus and a proton traversing through an intranuclear medium that is subjected to various independent FSIs (under the second assumption of the impulse approximation) such as elastic/inelastic scattering, hadron production/absorption and charge exchange. A brief description of these intranuclear processes that are responsible for altering the intermediate-state hadron kinematics is listed below.

- In an elastic scattering process, the direction of propagation of the final state hadron is modified with charge and the energy of incident and the final state hadrons being conserved.
- In an inelastic scattering process, both the energy and the direction of propagation for the incident and final state hadrons are modified. In addition, this process also modifies the energy state of the whole nucleus as particles leave

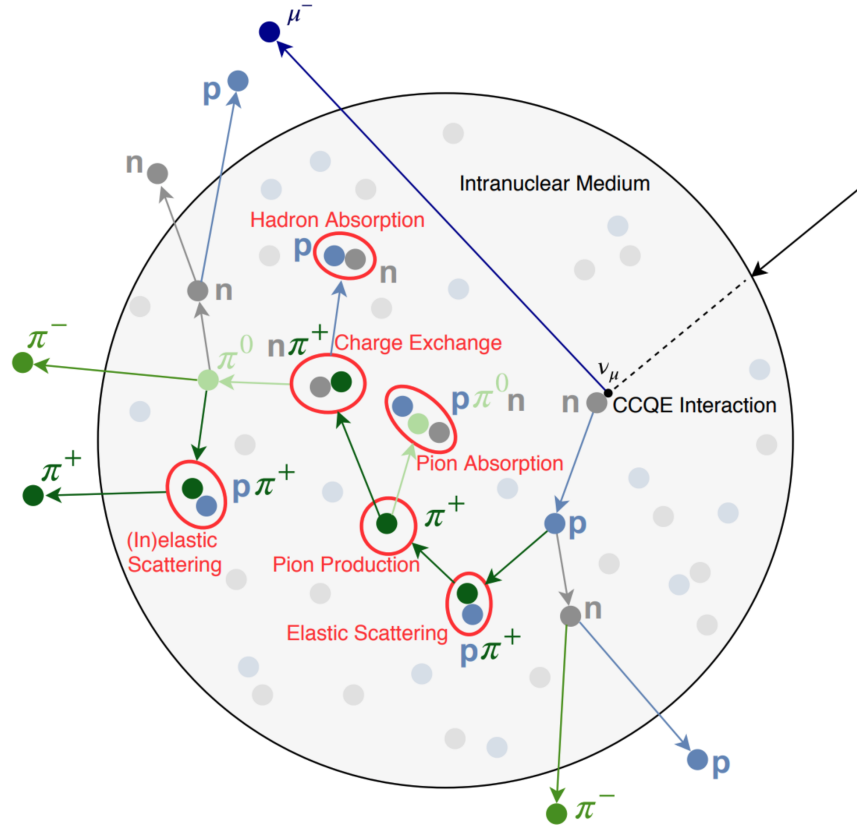


Figure 4.9: A diagram showing an example of a CCQE FSI process, where an incoming muon neutrino (black arrow) interacts with a neutron (gray) under the impulse approximation to produce a muon (dark blue) that escapes the nucleus and a proton that traverses through an intranuclear medium and is subjected to various independent FSIs (red circles) such as elastic/inelastic scattering, hadron production/absorption and charge exchange under the second assumption of impulse approximation. Figure from Ref. [47].

the nucleus.

- In a charge exchange process, hadrons exchange their charge as in the process $\pi^+ n \rightarrow \pi^0 p$ in figure 4.9.
- In a hadron absorption process, the hadrons undergo absorption inside the nucleus and the final state has no outgoing particles to measure.
- In a pion production, a hadron with sufficiently high energy produces a pion inside the nucleus.

These intranuclear processes are very challenging to model and can alter the intermediate-state hadron kinematics and thus the kinematics and multiplicities of the outgoing particles that are measured.

4.3 MicroBooNE Detector Response to the Passage of Particles

The most common final state particles in the $\mathcal{O}(1 \text{ GeV})$ neutrino interactions MicroBooNE is sensitive to comprise pions, protons, neutrons, muons, electrons and photons. This section outlines particle interaction in a LArTPC, including the interactions of heavy charged particles comprising muons, pions and protons, described in section 4.3.1, light charged particles comprising electrons and positrons interacting with lead described in section 4.3.2. The interactions responsible for the production of an electromagnetic shower by an electron and a photon relevant for the analysis performed in this thesis are described in section 4.3.4. Finally, scintillation light production in liquid argon is described in section 4.3.5.

4.3.1 Heavy Charged Particle Interactions

A charged particle propagating through a medium such as liquid argon ionises the medium, depositing energy along its path via the process of Coulomb scattering with electrons of the atoms of the medium. The mean energy, E , deposited per unit length propagated, x , for a charged particle with charge z in a medium with atomic number Z and atomic mass A is described by the Bethe-Bloch equation [39],

$$\left\langle -\frac{dE}{dx} \right\rangle = K z^2 \frac{Z}{A} \frac{1}{\beta^2} \left[\frac{1}{2} \ln \left(\frac{2m_e^2 c^2 \beta^2 \gamma^2 W_{\max}}{I^2} \right) - \beta^2 - \frac{\delta(\beta\gamma)}{2} \right], \quad (4.6)$$

where $\beta = v/c$ for a charged particle propagating with velocity v in the medium, $\gamma = (1 - \beta^2)^{-1/2}$, $m_e c^2$ is the rest mass energy of the electron, I is the mean ionisation potential or the mean excitation energy of the medium, W_{\max} is the maximum possible

energy transfer in a collision, $\delta(\beta\gamma)$ is an energy loss correction factor dependent on the density of the medium, and K is a constant defined as

$$K = 4\pi N_A r_e^2 m_e c^2, \quad (4.7)$$

where N_A is Avogadro's number and r_e is the classical electron radius.

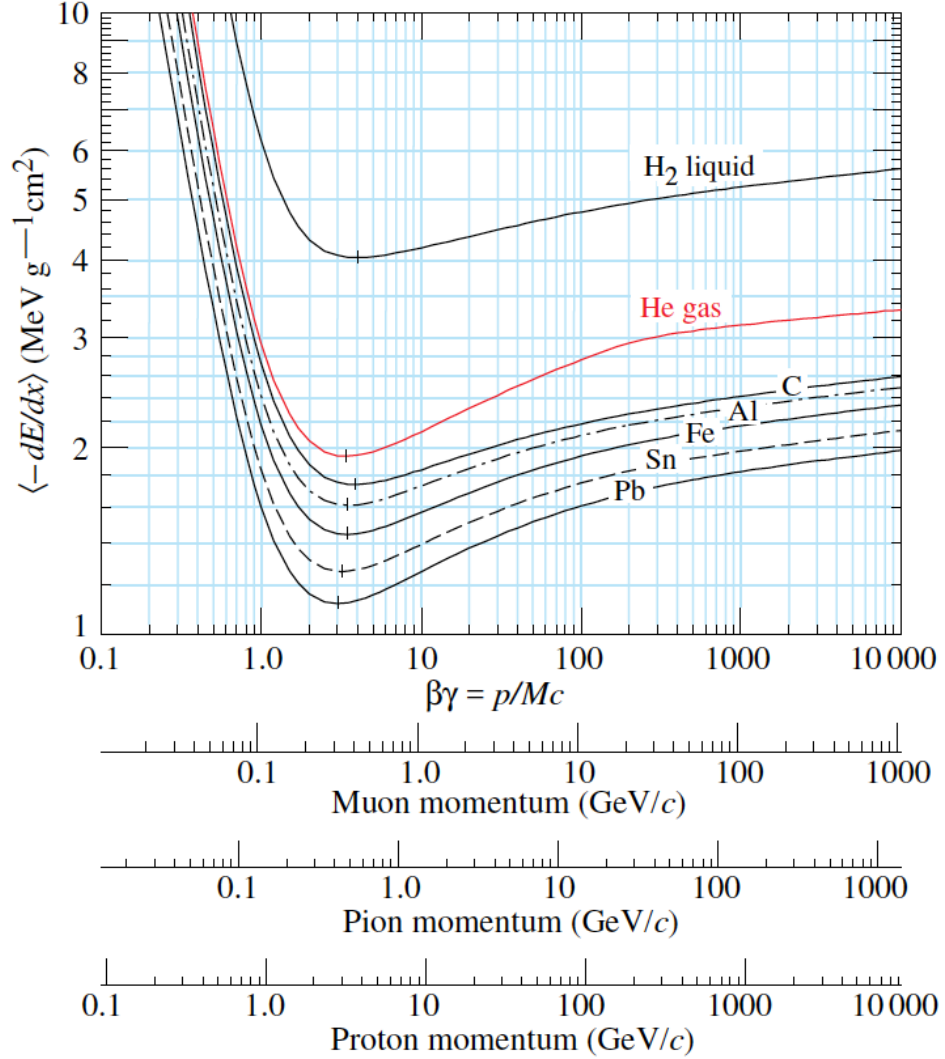


Figure 4.10: The Bethe-Bloch curve, or the stopping power or the mean energy loss of a charged particle per unit length as a function of particle momentum, $\beta\gamma$, for muons, pions and protons in several different mediums from the lightest, liquid hydrogen, H₂, to the heaviest, lead, Pb. Figure from Ref. [39].

Figure 4.10 shows the Bethe-Bloch curves defined using equation 4.6 for muons, pions and protons in several different mediums. Some important characteristics of the Bethe-Bloch curves are:

- For $\beta\gamma \sim 3$, the energy loss increases rapidly as β^{-2} as the speed of the particle decreases.
- For $\beta\gamma \sim 3 - 4$, the energy loss of the particles reaches a minimum and these particles are known as Minimum Ionising Particles (MIPs).
- For particle momenta above this MIP region, a relativistic regime is reached where $\beta \sim 1$ and the energy loss of the particles is influenced by the logarithmic term resulting in a rise in the energy loss.
- For very large particle momenta, the transverse component of the electric field the particles experience increases due to the Lorentz transformation $E \rightarrow \gamma E$. This increases the interaction strength. However, the medium in which the particle is propagating becomes polarised giving rise to a shielding effect which truncates the rise in the energy loss due to the logarithmic term resulting in a saturation for $\beta\gamma \gg 1$. This polarization is density dependent and this effect is encompassed in the density correction factor, δ .
- Across different mediums, the rate of energy loss per gram per square centimetre is greater for mediums with low atomic number than for those with high atomic number due to there being fewer electrons per gram in mediums of higher atomic number.
- The mean energy loss of a charged particle in matter is a stochastic process and the probability distribution for this process is described by a highly skewed Landau function.

In the MicroBooNE LArTPC, the particles we are interested in are produced in the energy range of tens of MeV to a few GeV. The masses of the particles propagating in a medium and their energy loss are highly correlated. Therefore a proton produced from a neutrino interaction at MicroBooNE is most likely to be produced in the region below $\beta\gamma \sim 3$ where there is a high energy loss. However, pions and muons are most likely to be produced in the MIP region of $\beta\gamma \sim 0.1 - 1$ GeV with a mean energy loss of around 2.2 MeV/cm in argon. This energy deposition of a particle

leads to a loss of momentum until the particle stops completely, depositing all of its energy rapidly at the end of its trajectory. This phenomenon of a rapid loss of energy at the end of the particle's trajectory is referred to as the Bragg peak. The energy transfer can sometimes produce a secondary electron with energy enough to produce further ionisation. These electrons are known as δ -rays.

4.3.2 Light Charged Particle Interactions

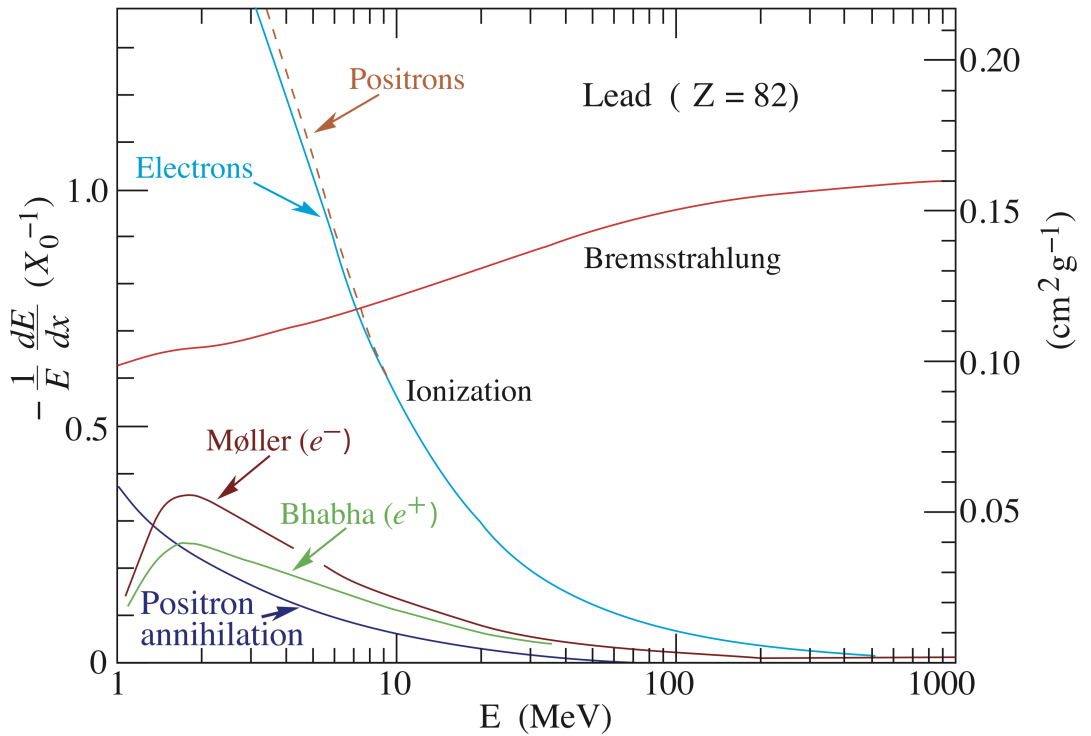


Figure 4.11: Fractional energy loss per radiation length in lead as a function of electron/positron energy. Figure from Ref. [39]

The other most common charged particles produced by the neutrino interactions at MicroBooNE are electrons and positrons. The fractional energy loss per radiation length³ as a function of electron/positron energy in lead is shown in figure 4.11. Some important characteristics of the electron/positron energy loss are,

- Energy loss of electrons/positrons in collision with electrons of the medium,

³A characteristic length of a material relating the energy loss of highly relativistic particles by electromagnetic interactions. It is usually measured in g cm^{-2} . For an electron, it is the mean distance over which an electron loses 67% of its energy by Bremsstrahlung. For a photon, it is 7/9 of the mean free path for pair production.

also referred to as ionisation, is a dominant process below $E \sim 7$ MeV.

- Above ~ 7 MeV, the energy loss by ionisation becomes sub-dominant and Bremsstrahlung becomes the dominant process, where an electron or a positron under the electric field of a nucleus decelerates and losses its energy by radiating photons.
- The energy at which the energy loss by ionisation is same as the energy loss by Bremsstrahlung is defined as the critical energy, E_c .
- For electrons, the final process responsible for the energy loss is Møller scattering, which is an electron-electron scattering similar to ionisation except the energy loss per collision is above 0.255 MeV [39].
- For positrons, the next most important process is positron-electron scattering also known as Bhabha scattering, which is again similar to ionisation except the energy loss per collision is above 0.255 MeV [39].
- For positrons, the final process responsible for the loss of energy in a medium is positron annihilation where a positron annihilates with an electron to produce two photons.

4.3.3 Neutral Particle Interactions

The most common neutral particles produced by neutrino interactions in argon are neutrons, neutral pions and photons. In the energy range of MicroBooNE, the dominant scattering processes for neutrons are elastic or inelastic scattering, followed by neutron capture. These scattering processes deposit very little energy in liquid argon and therefore neutrons are very challenging to reconstruct at MicroBooNE. The neutral pions almost immediately decay into two photons which can mimic our signal of two showers induced by the electron-positron pair from a Higgs portal scalar decay.

4.3.4 Electromagnetic Showers

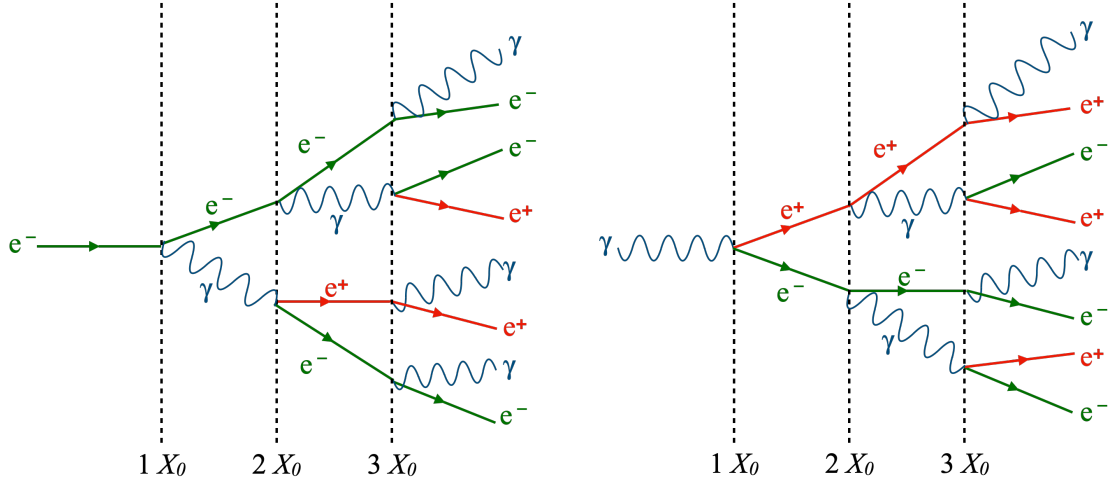


Figure 4.12: The diagrams on the left and right show electromagnetic shower production by a photon and an electron respectively.

Electrons and photons both produce electromagnetic showers in MicroBooNE as shown in figure 4.12. The left panel of the figure illustrates a high energy electron, $\gtrsim 100$ MeV, which, on interaction with the medium produces a photon via the process of Bremsstrahlung which in turn produces an e^+e^- pair via the process of pair production. The continuous process of Bremsstrahlung and pair production produces an electromagnetic cascade of photons, electrons and positrons in a dense medium. Photons initially do not produce any activity in the detector until they convert into e^+e^- pairs (as shown on the right of figure 4.12). At that point, an electromagnetic shower develops in the same way as for an electron. The development of an electromagnetic shower is a stochastic process comprising a number of discrete interactions. The signal in the analysis performed in this thesis is the Higgs portal scalar decaying into two showers and therefore it is crucial to understand the physics and the formation of an electromagnetic shower.

Figure 4.12 shows an example of an electromagnetic shower: a cascade of secondary particles produced by a photon or an electron where the number of particles approximately doubles after propagating the radiation length, X . The radiation length of argon is 14 cm [39]. The average energy of the particles after propagating x radiation lengths is

$$\langle E \rangle \approx \frac{E}{e^x}. \quad (4.8)$$

The production of the cascade of secondary particles continues until $\langle E \rangle$ falls below the critical energy, E_c (defined in section 4.3.2). Below E_c , electrons and positrons lose energy predominantly by ionisation. The maximum number of particles in an electromagnetic shower is reached at $\langle E \rangle = E_c$ in equation 4.8, allowing an x_{max} to be defined as

$$x_{max} = \frac{\ln(E/E_c)}{\ln 2}. \quad (4.9)$$

The transverse dimension of an electromagnetic shower is characterised by the characteristic constant of a material, called the Molière radius, R_M , which is defined as the radius of a cylinder containing 90% of the energy deposition of a shower. The Molière radius for a material with radiation length X_0 is defined as

$$R_M = X_0 \frac{E_s}{E_{crit}}, \quad (4.10)$$

where E_{crit} is the critical energy at which the electron energy is equal to the ionization loss per radiation length and $E_s = (4\pi/\alpha)^{1/2}m_e c^2$ is the scale energy with α being the fine structure constant [39].

4.3.5 Scintillation Light

The scintillation light in liquid argon is produced due to ionisation and excitation of the argon atoms. A charged particle exciting an argon atom gives rise to the process of self-trapped exciton luminescence while ionising an argon atom gives rise to the process of recombination luminescence as shown in figure 4.13. Both of these processes lead to the production of a short-lived dimeric molecule with another argon atom called a dimer or excimer. The production of the scintillation light in liquid argon is due to de-excitation of the two possible singlet and triplet excimer states. In

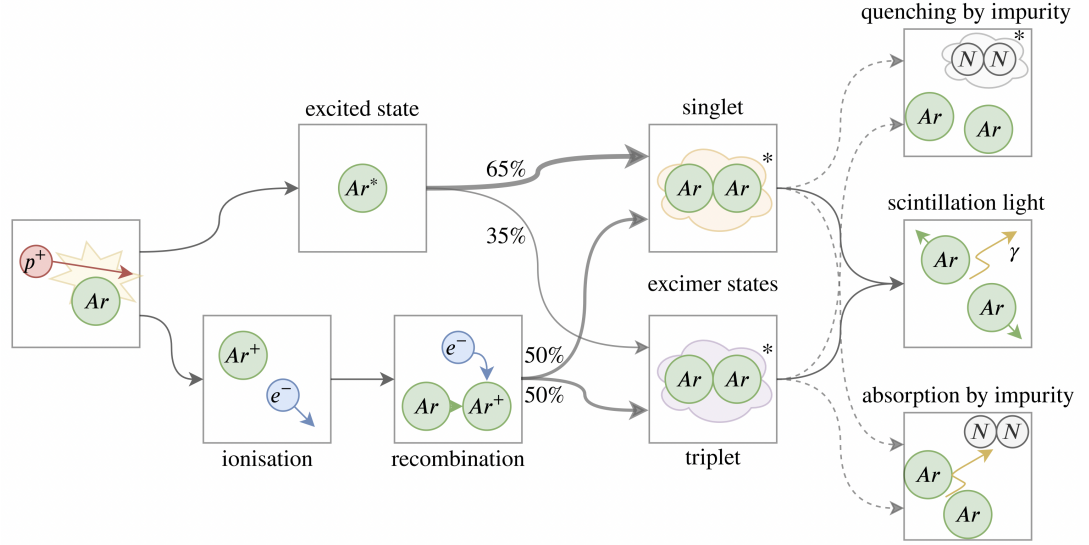


Figure 4.13: A flow chart showing the processes leading to scintillation light production in liquid argon. The upper sequence shows the process of self-trapped exciton luminescence due to excitation of the argon atom, whereas the lower sequence shows recombination luminescence due to ionisation of the argon atom. Both of these processes produce singlet and triplet excimer states in different proportions. The de-excitation of these two excimer states produces scintillation light in liquid argon. The light yield could also be suppressed by the presence of impurities which can either quench these excimer states or absorb the scintillation photons. Figure from [54].

self-trapped exciton luminescence about two-thirds of the excimers are in the singlet state with the rest in the triplet state, whereas in recombination luminescence both states are equally likely. The common impurities in liquid argon such as nitrogen, oxygen and water can suppress the amount of scintillation light by quenching or absorbing, which has important consequences in light detection, as discussed in section 4.4.6.

The de-excitation of the singlet state gives rise to a fast component of the scintillation light with a very short lifetime of ~ 6 ns, while the de-excitation of the triplet state gives rise to a slower component of the scintillation light with a relatively longer decay time of ~ 1.6 μ s. Both of these fast and slow components have their wavelengths narrowly peaking in the Vacuum Ultra-Violet (VUV) region at $\lambda \sim 128$ nm. Liquid argon can produce large amounts of light with as many as ~ 24000 photons per MeV of energy deposited under an electric field of 500 V/cm. At nominal pressure, the liquid argon is transparent to these photons.

4.4 Detector Effects

This section outlines various detector effects that modify the cloud of ionisation electrons produced by the propagation of a charged particle through a LArTPC. These effects include spreading of the ionisation cloud as a function of drift time, also known as diffusion, described in section 4.4.1; distortion in the electric field due to slow drifting Ar^+ ions, also known as the space charge effect, described in section 4.4.2; ionisation electrons recombining with argon ions to produce scintillation photons, also known as recombination, described in section 4.4.3; and finally contaminants in the argon medium attenuating the electron cloud as a function of drift time, described in section 4.4.4. These effects are crucial to understand as they play an important role in modeling the systematic uncertainties as well as calibrating the detector to provide accurate kinematic information such as position, energy and time of the particles produced in the LArTPC.

4.4.1 Diffusion

Diffusion causes the ionisation cloud propagating towards the anode to spread out as a function of drift time. This spreading is non-isotropic in nature due to the presence of the electric field. Therefore diffusion is characterised by two components, longitudinal and transverse, for spreading in the directions parallel and perpendicular to the direction of propagation respectively.

In longitudinal diffusion, the clouds of electrons produced near the cathode travel larger distances and are therefore affected the greatest by diffusion. This broadens the pulse recorded on a wire as a function of drift time as shown in figure 4.14. Pulses are narrower and taller for ionisation clouds produced near the anode where the drift time is small, and are wider and shorter for ionisation electrons produced near the cathode where the drift time is larger. The time-width of a signal pulse at a given time t , $\sigma_t(t)$, is given by

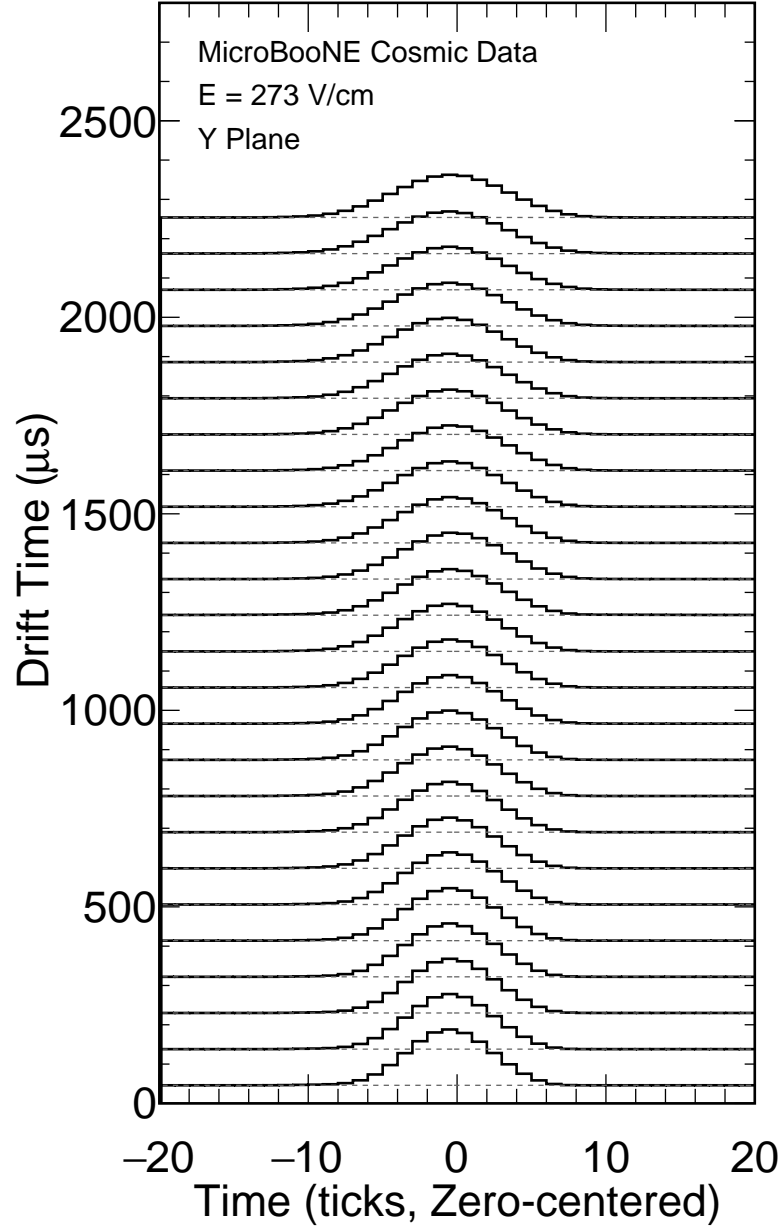


Figure 4.14: The visualisation of longitudinal diffusion on the charge waveforms recorded by the sense wires as a function of ionisation electron drift time. One bin correspond to one time tick which is equivalent to $0.5 \mu\text{s}$. Figure from [55].

$$\sigma_t^2(t) \simeq \sigma_t^2(0) + \left(\frac{2D_L}{v_d^2} \right) t, \quad (4.11)$$

where v_d is the drift velocity and D_L is an effective longitudinal diffusion constant calculated experimentally from the slope of the σ_t^2 vs t fit [55]. The constant D_L characterises the longitudinal diffusion.

Transverse diffusion is characterised by the transverse diffusion constant, D_T , defined as

$$D_T = \frac{\mu(E)\epsilon_e}{e}, \quad (4.12)$$

where ϵ_e is the energy of the electron, $\mu(E)$ is the electron mobility in an electric field of strength E , and e is the charge of an electron. In transverse diffusion, the smearing of the pulse causes the detection of a signal on multiple neighbouring wires resulting in a reduced position resolution of the detector.

4.4.2 Space Charge Effects

In the MicroBooNE detector, the electric field was designed to be uniform throughout the active volume of the TPC. However, ionisation in the detector, primarily due to cosmic rays, creates Ar^+ ions which slowly drift towards the cathode and accumulate, leading to a distortion in the uniformity of the electric field. This effect displaces the reconstructed position by distorting the trajectories of ionisation electron clouds as they propagate towards the anode. This effect is known as the Space Charge Effect (SCE). In addition, the SCE affects the amount of electron-ion recombination (discussed in section 4.4.3) and hence the amount of charged measured by the sense wires. An example of the SCE on reconstructed cosmic muon tracks is shown in figure 4.15. The SCE is more significant for an ionisation electron cloud produced at cathode compared to the anode due to build-up of slow-moving Ar^+ ions at cathode.

4.4.3 Electron-Ion Recombination

Electrons from ionisation can be thermalised by the interactions with the surrounding medium, and then recombine with the nearby Ar^+ ions rather than being drifted towards the anode by the electric field. The process of recombination introduces a non-linear relationship between the amount of charge measured per unit length, dQ/dx , at anode and the original amount of charge deposited per unit length, dE/dx . The conversion from dQ/dx to dE/dx is described by the inverse “Modified Box

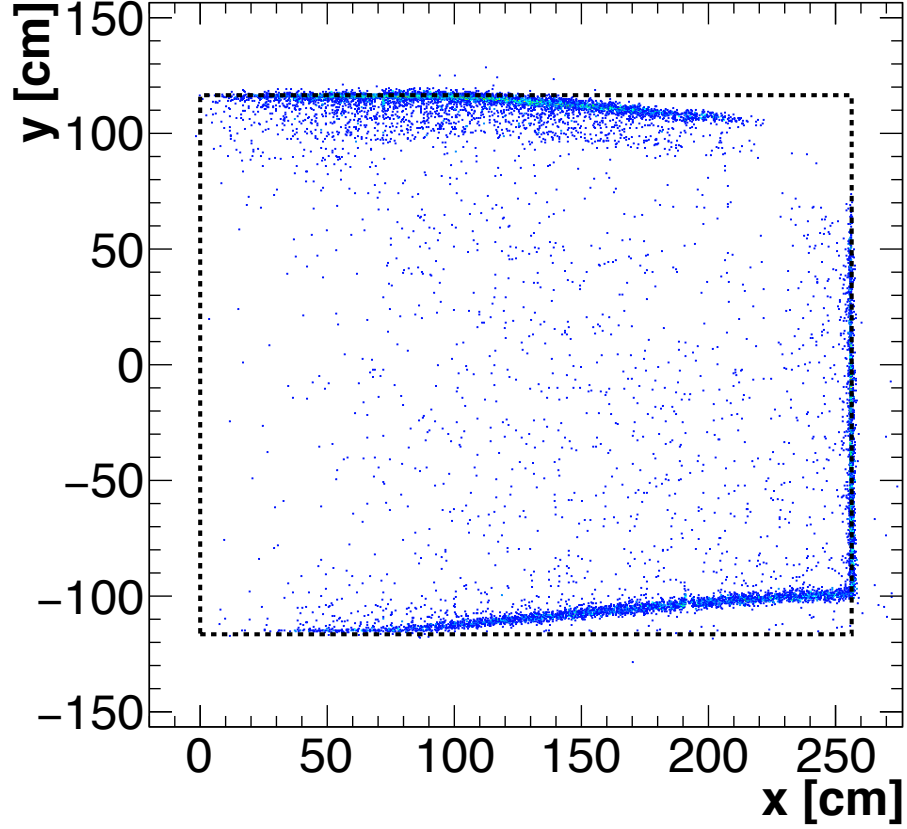


Figure 4.15: Entry/exit reconstructed points (blue) of cosmic muon tracks exhibiting an offset from TPC boundaries (black dashed lines) in the $x - y$ plane (perpendicular to the BNB beam) inside the MicroBooNE TPC. In the absence of a SCE, the reconstructed points should strictly lie at the TPC boundaries. The direction along x represents the ionisation drift direction with the cathode being located at $x = 256$ cm and the anode at $x = 0$ cm. The direction along y represents the top at $y = 116$ cm and the bottom at $y = -116$ cm of the TPC. The magnitude of the offset increases from anode to cathode due to distortion being larger for ionisation electrons originating at the cathode. Figure from Ref. [56].

Model” developed by the ArgoNeuT collaboration [57], which can be mathematically expressed as

$$\frac{dE}{dx} = \frac{\exp\left(\left(\frac{dQ}{dx}\right) \frac{\beta'}{\rho\mathcal{E}} W_{\text{ion}}\right) - \alpha}{\frac{\beta'}{\rho\mathcal{E}}}, \quad (4.13)$$

where ρ is the density of the argon, \mathcal{E} is the electric field strength, $W_{\text{ion}} = 23.6$ eV is the ionisation work function or the amount of energy required to ionise an argon atom and α and β' are parameters that are tuned using experimental data [58].

4.4.4 Argon Purity

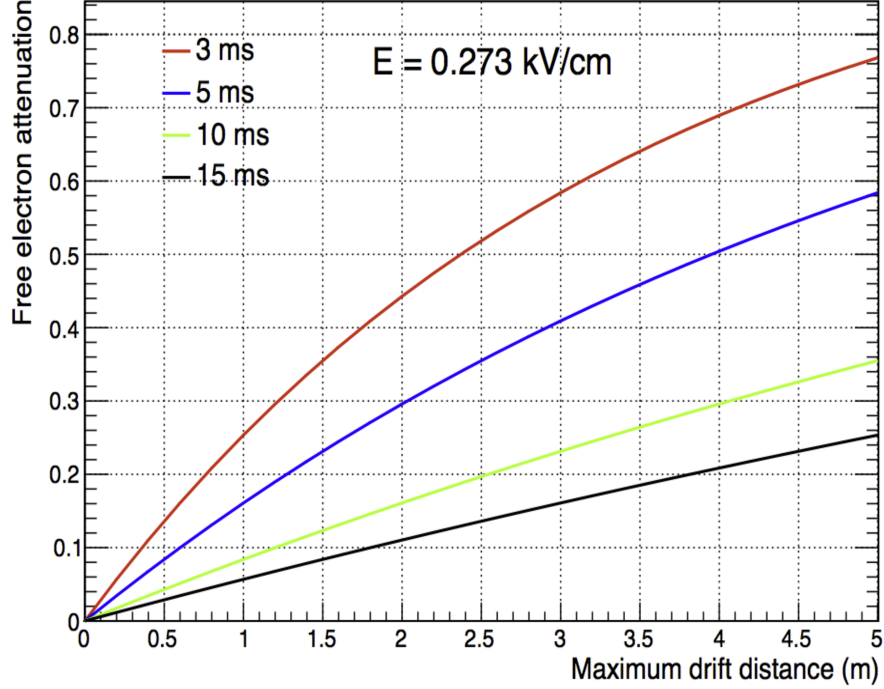


Figure 4.16: Fractional electron drift attenuation as a function of maximum drift distance for different values of drift-lifetimes of electrons propagating in a uniform electric field of 0.273 kV/cm. Figure from Ref. [59].

Impurities in liquid argon, specifically highly electro-negative contaminants such as oxygen, nitrogen or water, can capture the ionisation electrons and therefore reduce the amount of charge measured at the wire planes. The mean electron drift lifetime across the TPC is inversely proportional to the amount of contaminant present in the liquid argon as well as the strength of the electric field inside the TPC.

The amount of electro-negative impurity present in the liquid argon is measured by taking the fraction of the charge leaving the cathode, Q_C , that arrives at the anode, Q_A , after an electron drift time, t_{drift} . The electron drift-lifetime, τ_e , can be calculated using the relation

$$\frac{Q_A}{Q_C} = \exp\left(-\frac{t_{\text{drift}}}{\tau_e}\right). \quad (4.14)$$

At MicroBooNE, this fraction is frequently monitored to assess the purity of the liquid argon.

Figure 4.16 shows the fractional drift attenuation as a function of maximum drift distance for electrons with different values of drift-lifetimes propagating in a uniform electric field of 0.273 kV/cm. For a drift distance of 256 cm, an O₂ equivalent contamination of about 60 parts per trillion can result in $\sim 36\%$ of signal attenuation for an electron drift-lifetime of 5 ms [59]. The MicroBooNE detector was designed for an electron drift-lifetime of greater than or equal to 3 ms and a study performed using the data taken in August 2015 shows that the electron drift-lifetime is at least 6 ms which corresponds to an O₂ equivalent contamination of less than 50 parts per trillion for an electric field of 0.273 kV/cm [60].

4.4.5 Rayleigh Scattering

As discussed above, Rayleigh scattering is the elastic and coherent scattering of a photon with an atom and is the second most dominant mode of interaction for photons of energies below ~ 1 MeV. As discussed in section 4.3.5, the scintillation light has a wavelength peaking in the VUV region at $\lambda \sim 128$ nm where the energy of a photon is about 9.7 eV. At this energy, VUV scintillation photons have a high probability of undergoing Rayleigh scattering with the atomic nuclei. The scattering causes the photon to change its trajectory resulting in a smeared timing resolution of photon detection [61].

4.4.6 Light Yield Attenuation

As discussed in section 4.3.5, the scintillation light yield from excimer states can be attenuated by the contaminants present in the liquid argon via quenching or absorption. Quenching, also known as internal non-radiative relaxation, is the de-excitation of the excimers without the emission of a photon. In impurity quenching, a photon loses its energy via collisions with contaminants such as oxygen or nitrogen resulting in attenuation of the light yield. Impurity quenching is significant in the de-excitation of the triplet excimer state due to its significantly longer decay time

compared to the singlet excimer state. In impurity absorption, a photon is absorbed by the contaminants such as oxygen, water or nitrogen resulting, again, in a lower light yield at the PMTs. The problem of light yield attenuation was significantly reduced by monitoring the purity of the liquid argon using the method discussed in section [4.4.4](#).

Chapter 5

Simulation and Reconstruction

The simulation and subsequent reconstruction of an event at MicroBooNE is a multi-staged process that is handled through the Liquid Argon Software (**LArSoft**) framework [62]. Section 5.1 describes the first stage of this process, which, for the analysis presented in this thesis comprises generation of both Higgs portal scalar decay (signal) and neutrino interactions (background) and their daughter particles. Following generation, section 5.2 outlines propagation of these daughter particles and the resultant scintillation light and ionisation charge in the detector. Section 5.3 describes the procedure used to include cosmic rays into this MC simulation model. Sections 5.4 and 5.5 describe simulation and reconstruction of the PMT readouts and wire-plane readouts of the detector respectively. Section 5.6 outlines the **Pandora** multi-algorithm pattern recognition framework for identification and reconstruction of the cosmic muons and neutrino interactions in the MicroBooNE detector. Section 5.7 describes the charge calibration method used to convert charge measured per unit length to energy deposited per unit length in MC and data samples. Lastly section 5.8 outlines the signal and background samples for Run 1 and Run 3 used in the analysis presented in this thesis.

5.1 MicroBooNE Monte Carlo Generators

This section outlines the first stage of the event simulation process, which is generation of an event. For the analysis presented in this thesis, two different samples for the Higgs portal scalar signal and neutrino background are generated using Higgs portal scalar generator described in section 5.1.1 and the neutrino generator described in section 5.1.2.

5.1.1 Higgs Portal Scalar Generator

A “generic” Higgs portal scalar generator has been developed for the MicroBooNE experiment to simulate and study Higgs portal scalars produced from Kaons Decaying In Flight (KDIF) and Kaons Decaying At Rest (KDAR) at the NuMI target and at the NuMI absorber. The generator is capable of producing the Higgs Portal scalars in both the BNB as well as NuMI beams. The search for a low mass Higgs portal scalar at MicroBooNE in the NuMI beam described in this thesis uses this generic Higgs portal scalar generator for the production and simulation of the scalar. This generator has only two configurable parameters M_S and θ .

As discussed in section 3.2, the `g4numi` beamline simulation produces output files called `dk2nu` files. These files contain the full neutrino ancestry information from the proton-on-target interaction to the production of a neutrino. One major advantage of using these files is to calculate flux and flux systematics efficiently. Another advantage, discussed later in this section, is that `dk2nu` files contain the so called “importance weight” which makes generation fast and less resource intensive.

The primary parent particles of a neutrino comprise kaons and pions. The generator pre-selects the beamline neutrino parent kaons and decays them into Higgs portal scalars via $K \rightarrow \pi S$. Being a two-body decay, the momentum of the scalar produced from KDAR has a fixed magnitude, with the direction of the scalar being isotropic in the kaon’s rest frame. The momentum four-vector for the scalar is calculated by selecting random directions from the unit sphere. For KDIF, this

four-vector momentum is boosted to the lab frame using the momentum of the parent kaon. For KDAR, because these kaons are decaying at rest, no such boosting is required. To select KDAR at the target and at the NuMI beam absorber, the generator applies a momentum cut with kaon momentum $p_K < 1 \times 10^{-6}$ GeV/ c . To distinguish KDAR at the target from KDAR at the NuMI absorber, an additional cut of > 700 m on the kaon decay longitudinal position relative to the NuMI target is applied.

Following the cuts, the generator next checks if the scalar boson has a direction vector intercepting the active volume of the MicroBooNE TPC. Events not intercepting the detector are rejected and the process repeats with the next beamline kaon in the `dk2nu` file. For the events intercepting the detector, a decay vertex location along the trajectory of the scalar is selected inside the detector with the decay locations distributed according to an exponential decay. The lifetime of the scalar in the lab frame is calculated using equation 1.8 with time-dilation based on the scalar's Lorentz factor.

The generator then selects the decay mode of the scalar with probability based on the branching ratios of the possible scalar decay modes for a chosen Higgs portal scalar mass. The possible decay modes for the scalar, S , are $S \rightarrow e^+e^-$, $\mu^+\mu^-$, $\pi^0\pi^0$ and $\pi^+\pi^-$. These decay modes are all two-body decays and therefore the pairs produced are back to back in the Higgs portal scalar rest frame with direction of these pairs drawn randomly from a unit sphere. The momenta of these daughter particles are fixed and depend solely on the Higgs portal scalar mass. Finally, the momenta of the daughter particles are boosted to the lab frame using the four-vector momentum of the parent scalar. For the purposes of the analysis presented in this thesis, we are only interested in the scalar decay mode, $S \rightarrow e^+e^-$. The time of the scalar decay is calculated using the time of the parent kaon decay stored in the `dk2nu` files and the time of flight of the scalar. The proton-on-target time is randomly shifted to be in the NuMI beam spill time window discussed in section 2.5. The outputs, positions and momenta of the electron-positron pairs, are then used as inputs for the remainder of the simulation chain to produce reconstruction level information that is used in the analysis.

The normalisation factor also known as the total weight of an event, w_{event} is calculated by multiplying together the branching fraction of the scalar-producing decay, $\text{BR}(K^\pm \rightarrow \pi^\pm S)$, the decay position weight, P_{decay} , the inverse of the branching fraction of the neutrino-producing decay, $\text{BR}(K \rightarrow \nu + X)$, and the flux importance weighting, w_{imp} . The total weight therefore can be expressed as

$$w_{\text{event}} = \text{BR}(K^\pm \rightarrow \pi^\pm S) \times P_{\text{decay}} \times \frac{1}{\text{BR}(K \rightarrow \nu + X)} \times w_{\text{imp}}, \quad (5.1)$$

where the definitions of the four terms used are described as follows;

- **$\text{BR}(K^\pm \rightarrow \pi^\pm S)$** : The branching ratio of the scalar-producing decay, calculated using equation 1.4, which depends on θ , and M_S .
- **P_{decay}** : Probability of the scalar decaying inside the detector, mathematically expressed as $P_{\text{decay}} = e^{\frac{-L_1}{\gamma\beta c\tau}} - e^{\frac{-L_2}{\gamma\beta c\tau}}$, where L_1 and L_2 are the distances from the source of scalar production to the scalar's entry and exit points of the detector. The quantity $\gamma\beta c\tau$ in the exponent is the scalar decay length in the lab frame defined by the product of the velocity of the scalar, Lorentz factor, and scalar lifetime. The lifetime of the scalar is the inverse sum of the possible decay widths, described by equations 1.6 and 1.8, and is therefore again dependent on θ , and M_S .
- **$\text{BR}(K \rightarrow \nu + X)$** : The `dk2nu` files discussed in section 3.2 only contain kaon decays that produce neutrinos and not the number of kaons originally produced. To correct this, an event weight, w_{event} is calculated by inverting the summation of the branching fractions of all the possible neutrino-producing decays. For positively charged kaon decays in the `GEANT4` flux simulation, $\text{BR}(K^+ \rightarrow \mu^+ \nu_\mu) = 0.6339$, $\text{BR}(K^+ \rightarrow e^+ \nu_e \pi^0) = 0.0493$, $\text{BR}(K^+ \rightarrow \mu^+ \nu_\mu \pi^0) = 0.0330$. The combined sum of these branching fractions equates to 0.7458 and therefore the event weight is scaled up by a factor of $1/0.7458 = 1.40$.
- **w_{imp}** : The NuMI flux simulation does not simulate every single neutrino parent meson. Mesons with similar kinematics are suppressed and only a fraction of the mesons with unique kinematics are simulated to make generation less

resource intensive. The `dk2nu` files contain the so called “importance weight” which is the inverse of the fraction of these simulated mesons.

5.1.2 Neutrino Generator

Model element	Implementation
Nuclear model	Local Fermi Gas
Quasi-elastic scattering model	Nieves [63]
Meson exchange currents model	Nieves [64]
Resonant scattering model	KLN [65] and Berger-Sehgal [66]
Coherent scattering model	Berger-Sehgal [67]
Deep inelastic scattering model	Bodek-Yang [68]
Hadronic Final-state interactions model	INTRANUKE/hA 2018 [69]

Table 5.1: Table showing the models implemented for various neutrino-nucleon interaction model elements described in chapter 4 in the `GENIE` neutrino generator. This generator was used to produce the background MC samples for the analysis presented in this thesis.

The flux of neutrinos from the NuMI beamline simulation discussed in chapter 3 is passed to the `GENIE` [70] neutrino generator to simulate the neutrino interactions within and surrounding the MicroBooNE detector. The generator randomly samples neutrino energy, direction and flavour from the flux simulation files, and then steps the neutrino through the geometry of the MicroBooNE detector and its surrounding region. The generator takes into account the neutrino interaction cross-section for the target material being stepped through and the process of stepping is repeated until a neutrino interacts with the target material. Depending on the energy of the incident neutrino, the generator then selects its interaction mode and finally simulates the neutrino interaction and the subsequent final state interactions.

The background MC samples for the analysis presented in this thesis were generated using the `GENIE v3.0.6` generator and the neutrino-nucleon interaction models used in this version are listed in table 5.1. Additional details for these interaction models can be found in Ref. [71].

5.2 Particle Propagation in MicroBooNE

The output daughter particle list from the Higgs portal scalar generator and neutrino generator are passed into the **GEANT4** [43, 44] software framework to simulate particle propagation through the MicroBooNE detector. In **GEANT4**, the particles are independently propagated through the detector geometry until they either stop or escape the region of interest. The particles in **GEANT4** are incrementally stepped forward through the geometry of the detector where the length of each step is determined using the physics processes available to the particle. At each step, **GEANT4** calculates the energy deposited per unit length and converts it into scintillation photons and ionisation electrons that propagate through the detector volume in a uniform electric field.

The simulation for the propagation of all the ionisation electrons is resource intensive, and therefore the electrons are projected in bunches of 600 to the sense wire planes. At this stage, **GEANT4** also incorporates various detector effects described in section 4.4 such as diffusion, space charge effects, electron-ion recombination and electron drift lifetime.

Similarly for the simulation of photons, a voxel-based approach where a pre-generated library of 300,000 photons in a small volume subdivided from the larger detector volume is used to make the simulation less resource intensive. During simulation, the generator accesses this library to compute the probability of a photon reaching a specific unit of the light detection system depending on the voxel the photon originated from.

5.3 Cosmic Overlay

MicroBooNE, being a surface detector, experiences a continuous flux of cosmic rays with $\mathcal{O}(10)$ cosmic rays per event. The MicroBooNE collaboration has developed a technique to collect these real cosmic rays using an external unbiased trigger

(discussed in section 2.5.3 of chapter 2) and to superimpose, or overlay, the MC samples such as Higgs portal scalars or neutrino interactions on top of these cosmic ray data. This saves resources on simulating the cosmic rays as well as eliminating the uncertainties associated with modelling cosmic rays. This overlay technique furthermore incorporates real detector noise into the MC simulation.

5.4 Optical Reconstruction

The optical reconstruction framework combines the raw waveforms collected by a single PMT (discussed in section 2.3.2) to reconstruct “flashes”. These flashes represent optical activity in coincidence across several PMTs, most likely caused by a single interaction such as a neutrino or a cosmic ray interaction.

This section outlines the two major stages of the optical reconstruction framework, optical hit reconstruction described in section 5.4.1 and flash reconstruction described in section 5.4.2.

5.4.1 Optical Hits Reconstruction

The first stage of the optical reconstruction process is to estimate the baseline of the raw waveforms recorded by the PMTs. An example of the raw waveforms (blue) with amplitude in units of ADC is shown in figure 5.1. The green coloured line in this figure shows the baseline for these waveforms. This baseline is different for waveforms recorded within and outside the beamgate window described in section 2.5. For waveforms recorded outside the window (likely to be due to cosmic muons), a constant value equal to the first ADC value of the first recorded sample is chosen for the baseline. For waveforms recorded inside this window (likely to be due to neutrino interactions), an algorithm is used to estimate the baseline by interpolating the baseline at surrounding times [72]. After estimating the baseline, an independent algorithm then scans for the waveforms with an ADC above a certain threshold to find pulses. These pulses are known as “optical hits”.

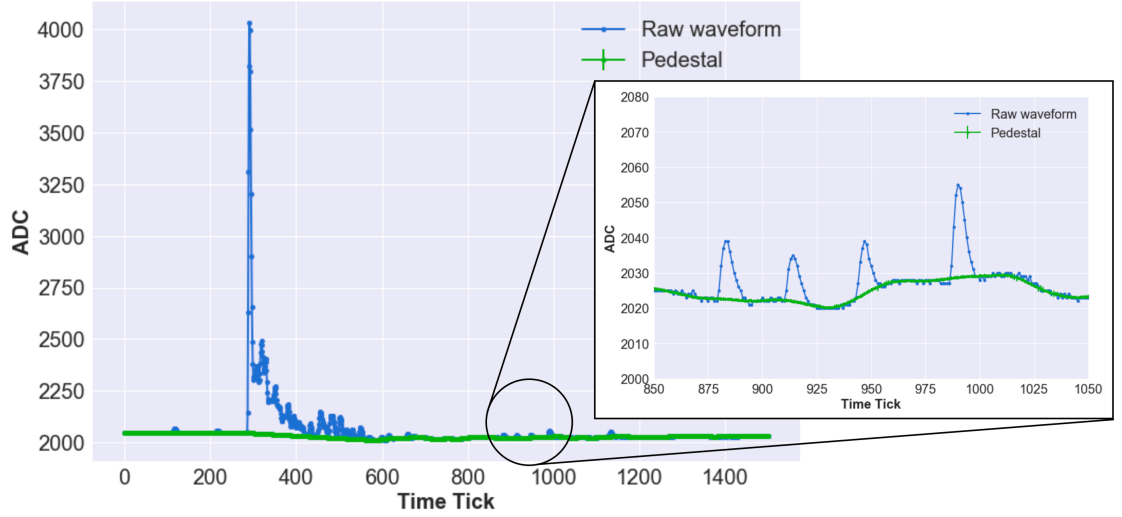


Figure 5.1: An example of the PMT raw waveforms (blue) inside the beamgate window. The y -axis shows the amplitude of these waveforms in units of ADC, whereas the x -axis shows the time-ticks with one time-tick corresponding to 15.6 ns. The baseline of these waveforms (in green) is calculated using an interpolation algorithm. The region between 850 and 1050 time-ticks is enlarged to show the individual photo-electron peaks recorded by the PMTs. Figure from [72].

5.4.2 Flash Reconstruction

The flash reconstruction algorithm clusters the optical hits associated to each PMT into configurable time ranges and searches for hits within these ranges that coincide with at least three PMTs. Once these coincident hits are found, an integration window of $8\ \mu\text{s}$ is applied to collect the late scintillation light associated with the interaction. An additional dead time window of $8\ \mu\text{s}$ is also applied to restrict any other flash in coincidence with late scintillation light to enter into the integration window. In the case of multiple candidate flashes within an $8\ \mu\text{s}$ time interval, the flash with the highest ADC or reconstructed photo-electrons is saved. Flashes occurring in the beam spill time window (described in 2.5) are of greater significance as these flashes are most likely to have been caused by a neutrino interaction.

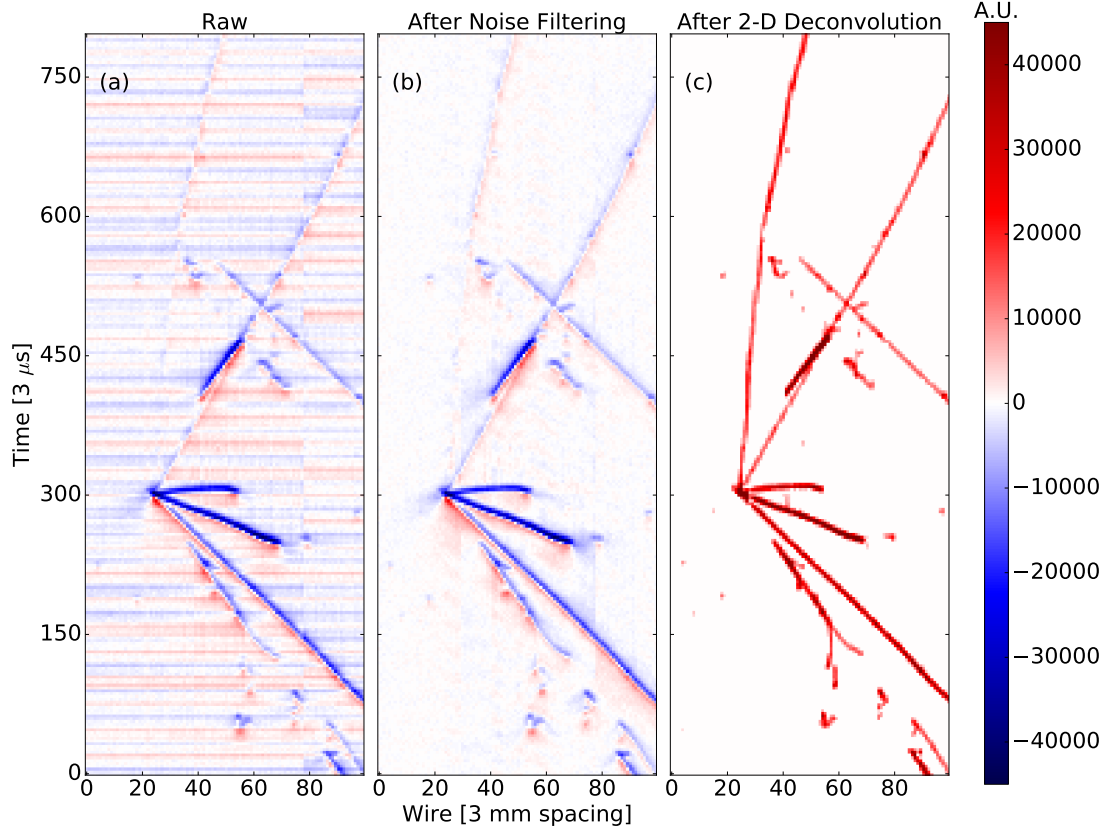


Figure 5.2: Event display of a neutrino interaction candidate from MicroBooNE data showing the first induction plane, U. Sub-figure (a) shows an unprocessed or raw waveform image of the interaction. Sub-figure (b) shows an image after filtering noise. Sub-figures (a) and (b) are in the units of average baseline (discussed in section 5.4.1) subtracted ADC scaled by 250 per $3 \mu\text{s}$. Sub-figure (c) shows the image of the charge spectrum in units of electrons per $3 \mu\text{s}$ after full 2D deconvolution. The difference in images before and after the deconvolution is significant with prolonged signals such as the track on the top left as well as the electromagnetic shower at the bottom recovered fully after 2D deconvolution. In addition, the image quality around the neutrino vertex has been enhanced which is expected to improve the pattern recognition and thus neutrino reconstruction efficiencies. Figure from Ref. [73].

5.5 TPC Signal Processing

The MicroBooNE sense wire planes described in section 2.3.1 record unprocessed or raw waveforms convolved with the field and the electronics response. A signal recorded by TPC wire planes is characterised by the following three elements:

1. The field response, which describes the induction and collection of charge on the TPC sense wires as ionisation electron drift through the wire planes [73].

This drift electron induces current on nearby sense wires which can distort the true waveform of the signal produced from ionisation charge passing by the wire planes [74].

2. The electronics response, which characterises the shaping and amplification of the current induced on the TPC wire in the cryogenic front-end electronics [73, 74].
3. The effects of diffusion and absorption due to propagation of ionisation charge distribution across the TPC [73].

The aim of the TPC signal processing is to accurately reconstruct particle interactions by eliminating these responses and effects via noise filtering and 2-D deconvolution and to extract the true number of ionisation electrons passing through a wire plane. An example event display of a neutrino interaction candidate on the first induction plane (U) undergoing noise-filtering and 2D deconvolution is shown in figure 5.2.

This section describes various steps carried out in the TPC signal processing to extract the ionisation charge spectrum from the digitised signal recorded by the TPC wire planes such as noise filtering (described in section 5.5.1), 2D deconvolution (described in section 5.5.2), region of interest finding (described in 5.5.3) and hit finding (described in section 5.5.4).

5.5.1 Noise Filtering

The first step of the TPC signal processing is filtering the noise from the raw unprocessed data recorded by the MicroBooNE sense wire planes. The MicroBooNE TPC experiences noise from various sources including noise inherent to the electronics and external excess TPC noises. The two largest sources of external excess noises are from the TPC drift voltage power supply and the low voltage regulators for the front-end electronics. MicroBooNE has developed various offline noise filtering techniques as well as upgrading several hardware systems to remove these noise

sources and retain excellent signal preservation [75].

5.5.2 2D Deconvolution

Following noise filtering, a mathematical technique is exploited to extract the original signal by deconvolving the detector responses, such as electronics or field responses, from the digitised signal measured by the wire planes [73]. In this method, the real field and electronics response functions are replaced by an effective software filter¹ response function during the extraction of the signal. This replaces the irregular bipolar waveform on the induction plane shown in figure 2.4 of chapter 2 with a more regular unipolar waveform using their respective response functions on the software filter. The two dimensions in deconvolution represent the analytical analysis for the waveforms done over the time dimension and wire dimension to account for long-range effects of the induction signals on neighbouring wires. These dimensions are encompassed within the software filter used to extract the ionisation electron distribution [74].

5.5.3 Region of Interest

The software filter discussed in section 5.5.2 to remove irregularities in the bipolar waveforms of the induction planes is still not optimal on its own due to these planes exhibiting suppressions at low frequency. During the process of deconvolution, the low frequency noise, mostly due to electronics, is amplified which results in undesirable uncertainties in the charge estimation [74]. This amplification of low-frequency noise could be suppressed by using low frequency noise filters in the deconvolution process, but a drawback of using these filters is alteration of the charge distribution in extended time ranges. Therefore to suppress the amplification of low-frequency noise, MicroBooNE has employed a technique called Region Of Interest (ROI) which limits the deconvolution to a small time window [77]. This limits the low frequency

¹MicroBooNE uses a Wiener filter [74, 76].

noise in the signal measured using the induction planes by identifying the signal and creating a ROI slightly bigger to cover this signal. This technique not only reduces the data size but also reduces the processing time of the deconvolution process [74].

5.5.4 Hit Finding

Following the identification of an ROI, an algorithm is used to further identify the signal waveforms within the ROI. These waveforms are fitted with one or more Gaussian functions to construct the objects called “hits”. The mean of the Gaussian corresponds to the time at which the signal was recorded on a wire and the integrated area of the hit corresponds to the amount of charge deposited on a wire. These “2D hits” form the basic inputs for the MicroBooNE **Pandora** reconstruction framework described in section 5.6.

5.6 Pandora Reconstruction Framework

The analysis presented in this thesis identifies and reconstructs the cosmic ray muons and the neutrino interactions in the MicroBooNE detector using the **Pandora** multi-algorithm pattern recognition framework [78]. The framework is also capable of reconstructing other particle interactions including the signal of our analysis, $S \rightarrow e^+e^-$, due to the generic topology approach employed by the framework.

The resultant “hit” (described in subsection 5.5.4) after carrying out various stages in TPC signal processing represents a signal detected on a specific wire at a definite drift time, t . Using the time of interaction, t_0 , and the drift time, t , a coordinate for the position of the charge deposited along x is calculated. The second coordinate is calculated using the position of the wire on which the charge is deposited (along z -direction). Using these two coordinates, **Pandora** constructs three 2D images of the events within the active volume of the TPC corresponding to planes U, V and Y. **Pandora** then correlates the features from these three images by exploiting the common x -coordinate and performs a 3D reconstruction.

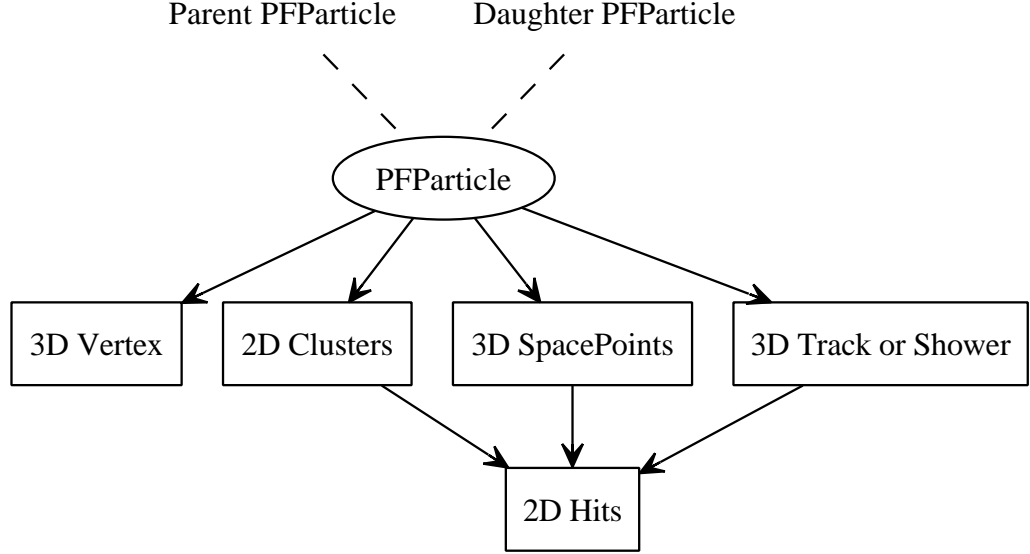


Figure 5.3: The hierarchy or structure chart showing various **Pandora** output data-products. The dashed lines show the reconstructed PFParticle hierarchies whereas the solid arrows show the objects associated with the PFParticles. Figure from Ref. [78].

Various output data-products of the **Pandora** pattern-recognition framework are shown in figure 5.3. The “PF” in PFParticle is an acronym for Particle Flow representing the **Pandora**'s hierarchical system of organising the particles with a flow from parent to daughter and each PFParticle corresponds to either a distinct track or a shower. The distinction between a track and a shower is estimated by **Pandora** by assigning a score to PFParticles between 0 and 1 with 0 being shower-like and 1 being track-like. The PFParticle is associated with a list of 2D clusters which group together relevant hits from each wire plane based on its geometrical proximity to construct geometrical shapes consistent with the expected 2D projection of tracks or showers to construct a track or a shower object. In addition, a PFParticle is also associated with 3D reconstructed hits (also known as Spacepoints²) and 3D vertices which are sets of reconstructed 3D positions and reconstructed vertex positions respectively. The vertex position is defined as the interaction point or the position of the first energy deposition in the active volume of the TPC.

A set of $\mathcal{O}(100)$ algorithms are used in identifying and reconstructing the particle interactions. These algorithms are shared between the following two reconstruction paths employed by the **Pandora** reconstruction framework:

²Spacepoints are constructed by combining 2D clusters of grouped hits that have been matched together across different planes in 3D.

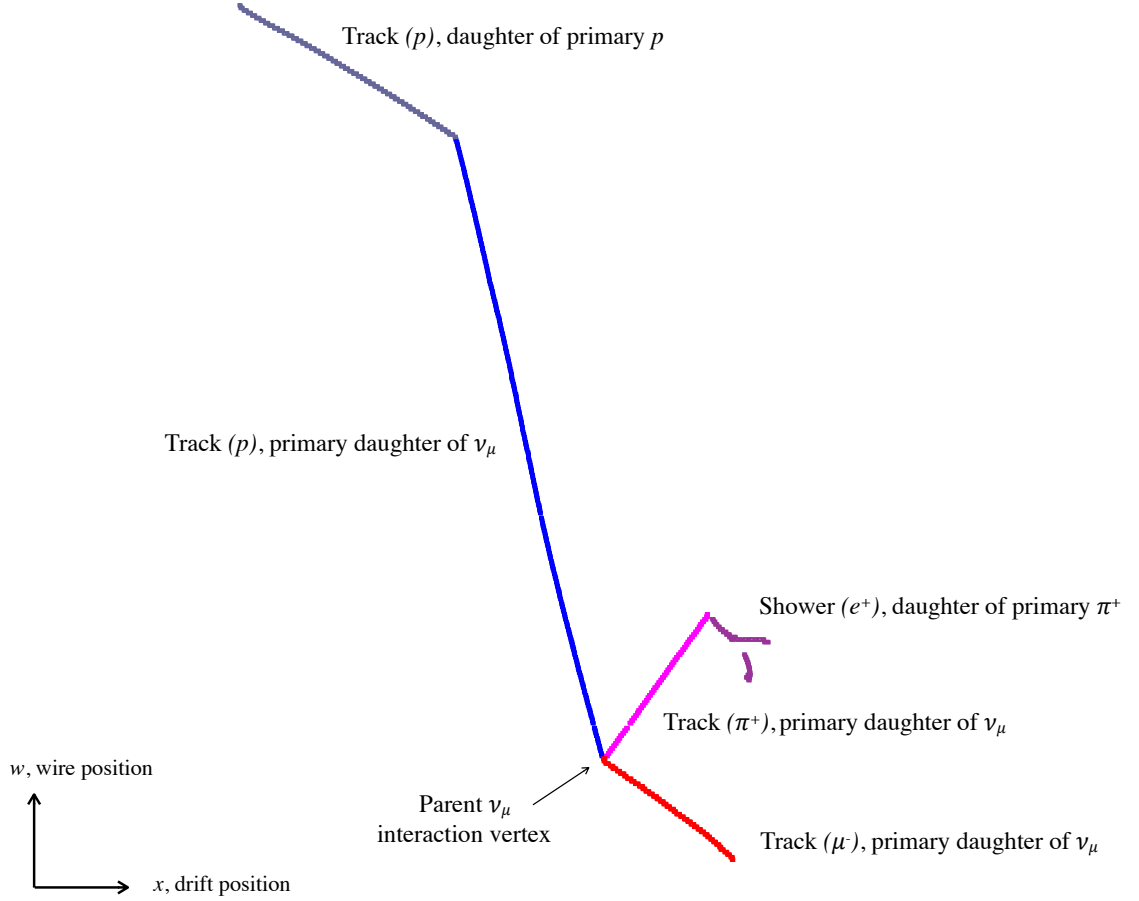


Figure 5.4: An example event display showing a hierarchy of particles for a simulated ν_μ event reconstructed by **Pandora** under the neutrino hypothesis. The reconstructed particles (indicated by a random colouring scheme) associated with the neutrino slice are added into the hierarchy of the neutrino vertex and are identified as either tracks or showers (indicated by text). Figure from Ref. [78].

- **PandoraCosmic**: Algorithms used in **PandoraCosmic** are track-oriented and therefore they are used in identifying cosmic ray muons. The algorithms assume a cluster of hits with a shower-like resemblance to be δ -rays (described in section 4.3.1 of chapter 4) produced by the cosmic ray muon that is at closest proximity to these hits. The vertex/start points of these reconstructed muons are assumed to be towards the top of the TPC with high y -coordinate value.
- **PandoraNu**: Algorithms used in **PandoraNu** identify the neutrino interaction vertex and reconstruct all the associated track-like and shower-like particles emerging from the vertex. The algorithms first create the parent neutrino followed by reconstructing visible particles which are then added as daughters into the hierarchy of the neutrino vertex as shown in figure 5.4.

The process of reconstruction begins by grouping all the reconstructed particles belonging to the the same interaction, be that a cosmic muon interaction or a neutrino interaction. A collection of these particles is known as a “slice” and it represents a physics interaction on a time scale of $\mathcal{O}(\mu\text{s})$. However, due to ionisation electron drift time, the corresponding charge collection time can be $\mathcal{O}(\text{ms})$, and therefore a typical event in MicroBooNE contains around four slices with the majority being from cosmic muons.

Pandora reconstructs a slice by first running the **PandoraCosmic** reconstruction path over all the hits recorded to construct muon tracks and associated δ -rays. The algorithms then tag and remove the obvious cosmic activity such as a muon going through the detector or a muon out of time in the trigger window (described in section 2.5). The remaining hits are then fed into the **PandoraNu** algorithms to reconstruct the slice under the neutrino hypothesis. The identification of the slices containing our Higgs portal scalar decay signal or background neutrino or cosmic muon interactions will be discussed in chapter 6 as part of the event selection.

5.7 Calibration

The analysis presented in this thesis uses calorimetric information to distinguish signal from background, and therefore the calibration of the calorimetric response of the ionisation signal is crucial to understand. The total charge extracted after noise filtering (described in section 5.5.1) and signal deconvolution (described in section 5.5.2) is not equal to the original charge produced from ionisation. Therefore after signal processing, a calibration procedure is applied to each event to estimate the true amount of energy deposited per unit length (dE/dx) from the charge measured per unit length (dQ/dx) by the anode wire planes. The x here is not an infinitesimal distance along the drift direction but rather a short segment of the track of a particle.

This section briefly goes over through this procedure by breaking it into two stages. Section 5.7.1 describes the process of charge calibration which corrects the position- and time-dependence of the detector response to deposited charge. Section

5.7.2 describes the process of energy calibration which estimates the true charge deposition from the calibrated charge deposition. The calibration was performed using the Bragg peaks from a sample of stopping muons produced from CC neutrino interactions. For highly ionising particles, a sample of stopping protons from neutrino interactions is used. The full details of the entire procedure can be found in [79].

5.7.1 Charge Calibration

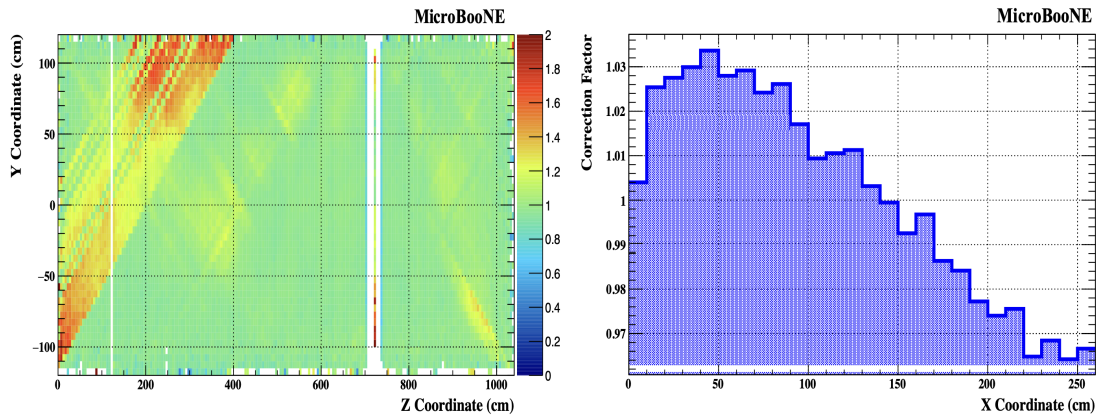


Figure 5.5: Left shows an example correction map for the correction factors in the $y - z$ plane applied to the collection plane for data collected between February to October 2016. The colour scale indicates the size of the correction factors required. Right shows an example correction map for the correction factor in the drift direction applied to the collection plane for data collected on February 25, 2016. Figure from Ref. [79].

The distortion in detector response due to diffusion (described in section 4.4.1), SCE (described in section 4.4.2), recombination (described in section 4.4.3), electron attachment due to argon impurity (described in section 4.4.4), and nonfunctional/unresponsive TPC channels causes the charge measured at the anode wire planes to differ from the total charge deposited. For the calibration, cosmic-ray muons with momentum in the range 4 – 5 GeV are used due to the characteristic peak at $dE/dx \sim 1.7$ MeV/cm. The calibration is carried out in three separate steps.

1. **Calibration in the $y - z$ plane:** The aim of this step is to remove the effects of space charge, transverse diffusion and unresponsive TPC channels. An example correction map for the correction factors in the $y - z$ plane applied

to the collection plane for data collected between February to October 2016 is shown in figure 5.5 (left).

2. **Calibration in the drift direction:** The aim of this step is to remove effects of space charge, longitudinal diffusion and electron attachment due to argon impurity. This calibration is performed for the data derived for each day. An example correction map for the correction factor in the drift direction applied to the collection plane for data collected on February 25, 2016 is shown in figure 5.5 (right).
3. **Calibration in time:** The aim of this step is to remove the temporal variations in the detector response, which are the changes in detector response over longer timescale due to effects such as changes in temperature or changes in argon purity.

5.7.2 Energy Calibration and Reconstruction

The mathematical expression, equation 4.13, for the inverse “modified box model” described in section 4.4.3 is used to calculate the absolute energy loss, dE/dx , from the calibrated $(dQ/dx)_{\text{calib}}$ described in section 5.7.1. An additional parameter, called a calibration constant, C_{cal} , to convert “ADC/cm” to “(number of electrons)/cm” is included in the exponent of equation 4.13. The new equation can therefore be expressed as

$$\frac{dE}{dx} = \frac{\exp\left(\frac{(dQ/dx)_{\text{calib}}}{C_{\text{cal}}} \frac{\beta'}{\rho\mathcal{E}} W_{\text{ion}}\right) - \alpha}{\frac{\beta'}{\rho\mathcal{E}}}. \quad (5.2)$$

The aim is to precisely determine C_{cal} to calculate the absolute energy loss. The amount of charge measured after removing the electronics and field response is described in section 5.5.3. For the amount of energy deposited per unit length, a sample of stopping muons from neutrino interactions with a well-understood dE/dx in the MIP region (discussed in section 4.3.1) is used. For muons this MIP region is

in the range 250 – 450 MeV and is theoretically well understood to better than 1%. The C_{cal} is calculated from the χ^2 minimum of the difference between the measured and the theoretical value of the energy loss in this MIP region [79].

5.8 Samples

Sample	POT count	Hardware Triggers	Event count
Run 1			
MC Neutrino Overlay	23.36×10^{20}	-	914729
MC Dirt Overlay	16.73×10^{20}	-	569506
Beam-Off	-	9199232.74	904362
Beam-On	2.002×10^{20}	5268051.0	610496
Run 3			
MC Neutrino Overlay	19.89×10^{20}	-	746098
MC Dirt Overlay	10.32×10^{20}	-	386248
Beam-Off	-	32878305.25	3237058
Beam-On	5.009×10^{20}	10363728.0	1104276

Table 5.2: Data (beam-on) and background samples (beam-off + neutrino overlay + dirt overlay) with their corresponding POT count, hardware triggers and the number of events used in the analysis. The event count is the number of hardware triggers that pass the software triggers.

This section describes the various data and MC samples that are produced and used in the analysis presented in this thesis, namely the data, referred to as “beam-on” in this thesis is described in section 5.8.1, and the samples used to generate the background prediction, which comprise the beam-off data, the NuMI MC neutrino overlay samples, and the NuMI MC out-of-cryostat neutrino overlay (dirt) samples is described in 5.8.2. Table 5.2 shows a summary of these data and background samples collected during the Run 1 and Run 3 data-taking periods. The processed NuMI data samples are from Run 1 and Run 3 with POTs equivalent to 2.002×10^{20} and 5.009×10^{20} , respectively. As discussed in sections 2.4 and 3.1, the detector and beam conditions for the two runs are different and therefore in the analysis, these are

analysed separately, each with their own sets of MC overlay background and signal samples.

5.8.1 Beam-On Samples

The sample of data collected by the MicroBooNE when exposed to the neutrinos coming from the NuMI beam is referred to as the “beam-on” sample and is shown in table 5.2. Each sample has a total number of POT and corresponding total numbers of hardware triggers (described in section 2.5.1) and reconstructed events associated with it. The number of reconstructed events shown in the table is equal to the number of hardware triggers which pass the software triggers described in section 2.5.2. A number of beam and detector related criteria are applied to the sample before qualifying it to be suitable for the analysis. The beam related criteria involve ensuring the intensity and position of the beam, and the current in the focusing horns are within the expected values. The detector related criteria involve ensuring suitable detector conditions such as the high voltage system being at the nominal value for collecting data.

5.8.2 Background samples

The prediction for the “beam-on” data sample discussed in section 5.8.1 is equivalent to the combination of MC neutrino overlay samples and “beam-off” sample containing cosmic muons recorded using external unbiased trigger described in section 2.5.3. This combination represents the MicroBooNE SM event prediction and is therefore used as the background for the study presented in this thesis.

Beam-Off

As discussed in section 2.5.2, only approximately 1 in 50 of the NuMI spills result in a neutrino interaction inside the MicroBooNE detector. This is due to the very

tiny neutrino interaction cross-sections. Separating an event comprising only cosmic muons from a neutrino interaction is difficult at the trigger level. Therefore to model these cosmogenic events, the MicroBooNE collaboration has developed a technique to record a sample of data when the beam is off called the “beam-off” sample that contains only cosmic muons. The hardware triggers used to collect the “beam-off” sample are described in section 2.5.1 and the number of hardware triggers used in the analysis is shown in table 5.2.

MC In-the-Cryostat Neutrino Overlay

The SM neutrino background for the search presented in this thesis contains neutrino interactions generated inside the MicroBooNE cryostat using the **GENIE** neutrino generator discussed in section 5.1.2. These neutrino interactions are overlaid on top of the data cosmic rays containing cosmic muons via the process described in section 5.3. These samples are produced to model a wide range of NuMI neutrino interactions received by MicroBooNE. An event in a sample is generated using the NuMI flux spectrum from the NuMI beam simulation described in section 3.2. A spill in the simulation has a simulated POT intensity of 5×10^{13} and the neutrinos are produced with a uniform time distribution across this spill window. The standard MC neutrino overlay samples used in the analysis are shown in table 5.2.

MC Out-Of-Cryostat Neutrino Overlay

The MC neutrino overlay sample contains only interactions of neutrinos with the liquid argon. However, to realistically model the “beam-on” data sample, interactions of neutrinos outside the cryostat or the external steel structure of the detector also need to be simulated as these interactions can sometimes produce daughter particles which travel inside the cryostat and produce enough scintillation light to pass the software trigger described in section 2.5.2. These interactions are called “dirt” or “out-of-cryostat” interactions and the sample of neutrino interactions generated outside the MicroBooNE cryostat is called the “dirt” sample or “out-of-cryostat” sample. The samples are produced in similar fashion to the MC neutrino overlay

sample with the neutrinos interacting outside the cryostat instead of in the liquid argon in the cryostat. The out-of-cryostat neutrino interactions are also overlaid with data cosmic rays containing cosmic muons via the process described in section 5.3. The MC “out-of-cryostat” neutrino overlay samples with their simulated POT and the numbers of reconstructed events used in the analysis are shown in table 5.2.

Event Classification

In the entire thesis, all histograms are shown with the events classified into the following truth-level categories:

- **DIRT**: Out-of-cryostat neutrino interactions from the dirt sample.
- **EXT**: Cosmic ray interactions from the beam-off sample.
- ν_e **NC**: All ν_e and $\bar{\nu}_e$ neutral current interactions from the MC neutrino overlay sample.
- ν_e **CC**: All ν_e and $\bar{\nu}_e$ charged current interactions from the MC neutrino overlay sample.
- ν_μ **NC**: ν_μ and $\bar{\nu}_\mu$ neutral current interactions with zero π^0 from the MC neutrino overlay sample.
- ν_μ **NC N π^0** : ν_μ and $\bar{\nu}_\mu$ neutral current interactions with N π^0 s from the MC neutrino overlay sample.
- ν_μ **CC**: ν_μ and $\bar{\nu}_\mu$ charged current interactions with zero π^0 from the MC neutrino overlay sample.
- ν_μ **CC N π^0** : ν_μ and $\bar{\nu}_\mu$ charged current interactions with N π^0 s from the MC neutrino overlay sample.
- **Data**: All interactions from the beam-on sample.

5.8.3 Background Sample Normalisation

The SM prediction from the MC comprises a combination of the “beam-off”, MC neutrino overlay and MC “out-of-cryostat” neutrino overlay samples described in section 5.8.2. In order to compare this SM prediction to the “beam-on” data sample described in section 5.8.1, a scaling is applied to normalise the SM prediction to the “beam-on” data sample. The MC “in-cryostat” and “out-of-cryostat” neutrino overlay samples are scaled by a factor

$$x_{\text{MC}} = \frac{\text{POT}_{\text{beam-on}}}{\text{POT}_{\text{MC}}}, \quad (5.3)$$

where $\text{POT}_{\text{beam-on}}$ is the POT collected when the beam is “on” and POT_{MC} is the simulated POT.

The “beam-off” samples are scaled using the number of hardware triggers for the “beam-on” sample, $\text{HW}_{\text{beam-on}}$, and the “beam-off” sample, $\text{HW}_{\text{beam-off}}$, by a factor

$$x_{\text{beam-off}} = 0.98 \times \frac{\text{HW}_{\text{beam-on}}}{\text{HW}_{\text{beam-off}}}. \quad (5.4)$$

The scaling factor assumes that none of the hardware triggers contain neutrino interaction in the “beam-on” sample. However as described in section 2.5.2, approximately 1 in 50 or 2% of the NuMI spills result in a neutrino interaction and therefore a normalisation factor of 0.98 in the scaling is used to account for this.

Flash Timing Distribution

The SM prediction with the background normalisation procedure discussed in section 5.8.3 is validated using the distribution of the largest flash time in an event prior to applying any selection. These distributions are shown in figure 5.6 (left) and 5.7 (left) for Run 1 and Run 3 respectively. The contribution of the out-of-cryostat (dirt) in these distributions is scaled down by factors derived using the procedure discussed in the next section. The calculation of the systematic uncertainty used

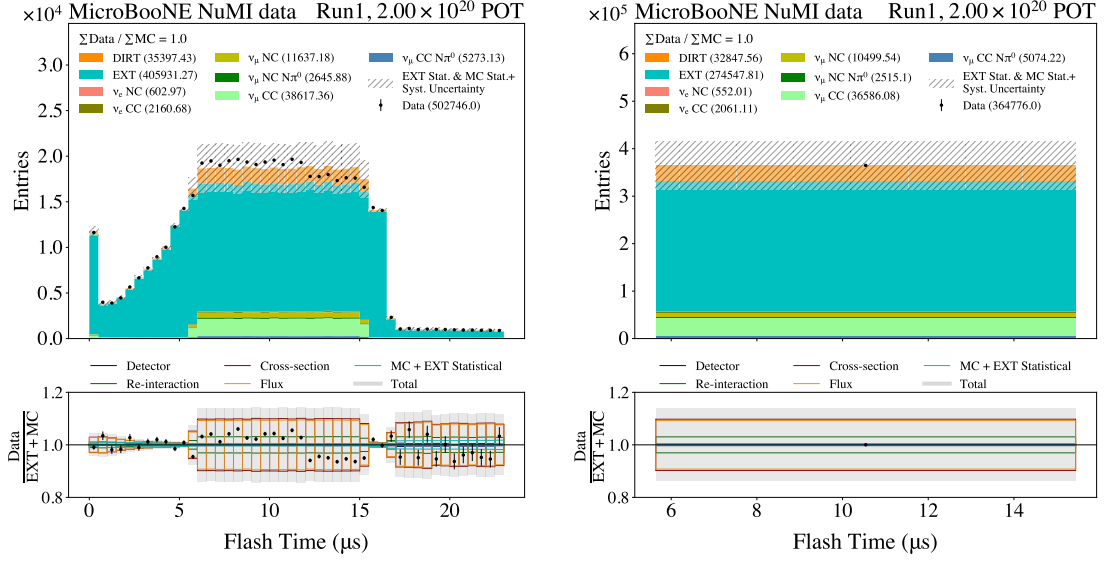


Figure 5.6: (Left) A comparison of the stacked prediction (EXT+MC) to the data (beam-on) for the Run 1 flash time distribution before any selection is applied. (Right) A single-bin flash time distribution for Run 1 in the NuMI beam spill time window. In these distributions, beam-off and out-of-cryostat samples are scaled using their respective normalisation factors.

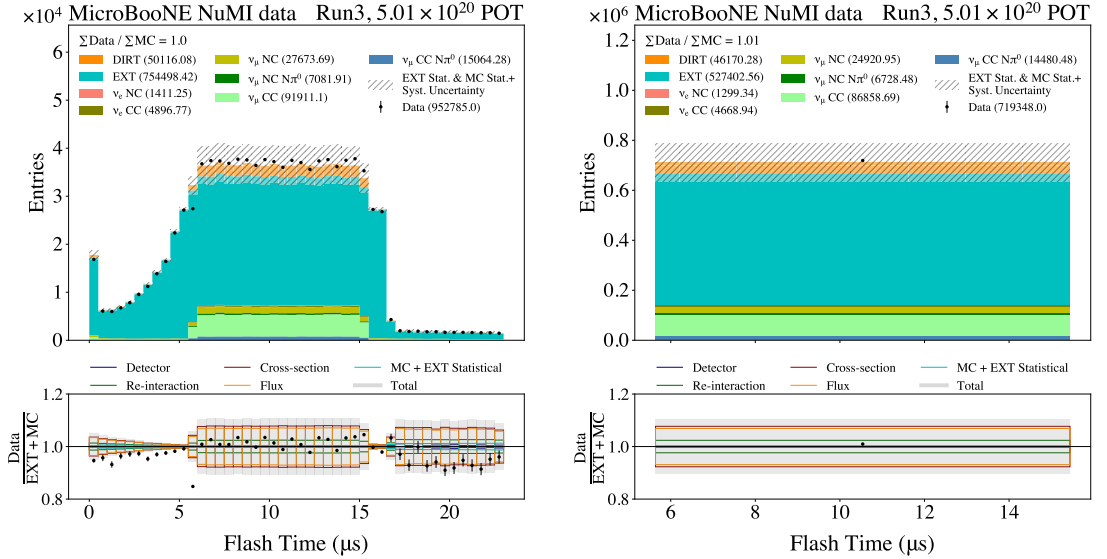


Figure 5.7: (Left) A comparison of the stacked prediction (EXT+MC) to the data (beam-on) for the Run 3 flash time distribution after applying the software trigger. (Right) A single-bin flash time distribution for Run 3 in the NuMI beam spill time window. In these distributions, beam-off and out-of-cryostat samples are scaled using their respective normalisation factors.

in these distributions is discussed in chapter 8. The beam gate window discussed in section 2.5 starts at $0 \mu\text{s}$ and the peak in event rate at this time is due to late scintillation light of the events occurring before the beam gate window. Following

this peak, the slow rise until around the start of the NuMI beam spill time window ($\sim 6 \mu\text{s}$) is caused by flashes occurring before the software trigger window with its late-light arriving inside the software trigger window. The closer the flash time is to the beam spill window, more likely it is to trigger the software trigger window resulting in an exponential rise in event rate from $0 \mu\text{s}$ to $\sim 6 \mu\text{s}$. The “shoulder-like” characteristic of the distributions is due to the software trigger window being wider than the beam spill time window discussed in section 2.5.2.

For Run 1, the “step-down” characteristic of the distribution midway through the beam window is due to slip-stacking in the proton delivery structure discussed in section 3.1.2. About 70% of the NuMI Run 1 data was recorded with the beam in a 4+6 slip-stacking configuration. The last two un-stacked batches will therefore have lower proton intensity and hence lower number of neutrinos produced than the first four stacked batches. The remaining 30% of the NuMI Run 1 data was recorded with the beam in a 6+6 slip-stacking configuration. For Run 3, the entirety of the data was recorded with the beam in a 6+6 slip-stacking configuration and therefore no “step-down” like shape is seen in figure 5.7 (left).

The neutrinos in the simulation are generated uniformly over the time window with the number of protons on target per spill comparable to the 4+6 slip-stacking configuration. As a result, the correct way of validating the SM prediction is by comparing the flash time distribution of the prediction to the data over the entire beam spill time window of $(5.64 - 15.44 \mu\text{s})$ as shown in figure 5.6 right for Run 1 and 5.7 right for Run 3. As can be seen the MC predictions agree well for both Run 1 and Run 3 NuMI data-sets. The normalisation for the analysis shown in this thesis is performed using the total POT delivered rather than POT delivered per spill and therefore the proton time distribution has no impact on the selection performance.

Out-Of-Cryostat Normalisation Factor

The out-of-cryostat interactions in the flash time distributions for the Run 1 and Run 3 shown in the previous section are tuned using factors derived by comparing the MC predictions for the flash time distribution in the beam spill time window to

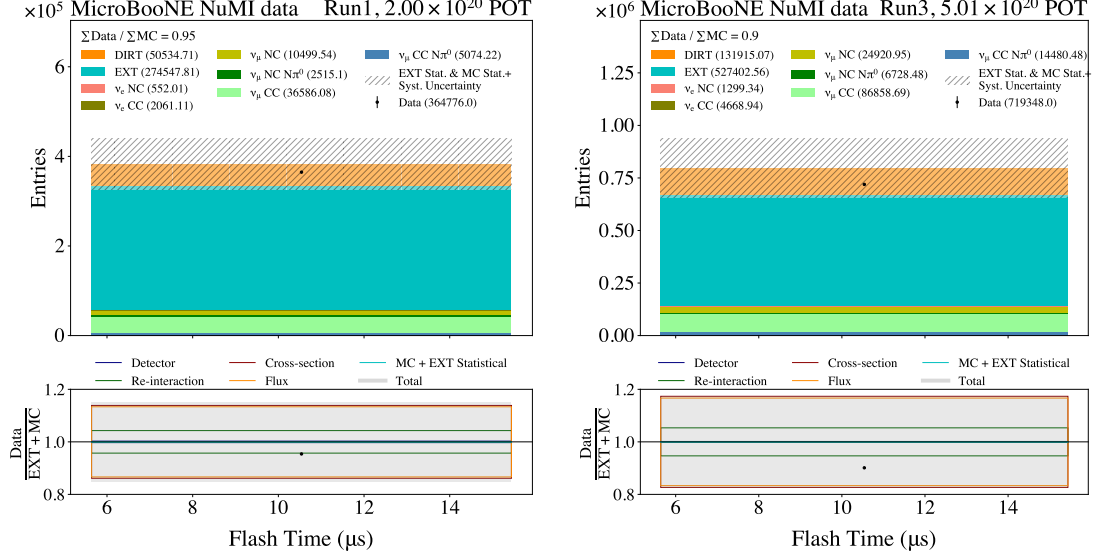


Figure 5.8: A comparison of the stacked prediction (EXT+MC) to the data (beam-on) for the Run 1 (left) and Run 3 (right) single-bin flash time distribution in the NuMI beam spill time window after applying the software trigger. In these distributions, only the beam-off sample is scaled and the out-of-cryostat sample is not scaled.

the data. Figure 5.8 shows the single-bin flash time distributions for Run 1 (left) and Run 3 (right) in the beam spill time window prior to tuning out-of-cryostat interactions. At MicroBooNE, the dirt sample is the most uncertain sample due to the presence of large sets of unknowns in modelling the out-of-cryostat interactions (described in section 8.5). Therefore the analysis presented in this thesis assigns a 100% uncertainty to the dirt sample resulting in larger uncertainty band in the subplots of figure 5.8 for Run 1 and Run 3. To mop up the residual data-MC disagreement, a down-scaling of 35% and 65% to the out-of-cryostat Run 1 and Run 3 interactions are calculated purely by tuning the MC predictions to data. The contribution of these out-of-cryostat interactions to the final selected background events is negligible and therefore has almost no impact on the final results even with the inclusion of 100% uncertainty to these interactions.

Neutral pion Normalisation Factor

The major source of background for the analysis performed in this thesis is $\pi^0 \rightarrow \gamma\gamma$ for two-shower like events. The MC expectations for the BNB dataset show significant tension with measurements of MicroBooNE data rich in π^0 events. Therefore the

MicroBooNE collaboration has performed a detailed study comparing the π^0 rich BNB data and MC for the following three different cases [80]:

- **No scaling** is applied.
- A **data-driven scaling** factor is applied whereby an energy-dependent weight is applied to the true π^0 energy. This weight is equal to $1 - 0.4 \times E_\pi(\text{GeV})$ for $E_\pi < 0.6$ GeV and $1 - 0.4 \times 0.6$ for $E_\pi \geq 0.6$ GeV.
- A **π^0 normalisation scaling** factor equivalent to 0.759 is applied to CC and NC π^0 events. This factor was calculated based on the normalization differences found in the π^0 mass distributions.

A similar study is here carried out for the NuMI dataset to validate the π^0 rich MC predictions. The selections listed below are applied to select a π^0 rich sample from the MC predictions and data.

- At least one reconstructed neutrino slice (described in section 5.6) in an event.
- Both shower candidates must have a Pandora track score (described in section 5.6) less than 0.8 to select shower-like events.
- Shower direction aligned with the trajectory connecting reconstructed vertex and shower start direction.
- Conversion distance greater than 3 cm to reject showers produced due to electrons.
- Cosine of the opening angle between the two showers to be less than 0.94 to reject overlapping showers where the Pandora reconstruction framework performs poorly.
- Leading and sub-leading shower collection-plane energies to be greater than 50 MeV and 20 MeV respectively below which the Pandora reconstruction framework performs poorly.

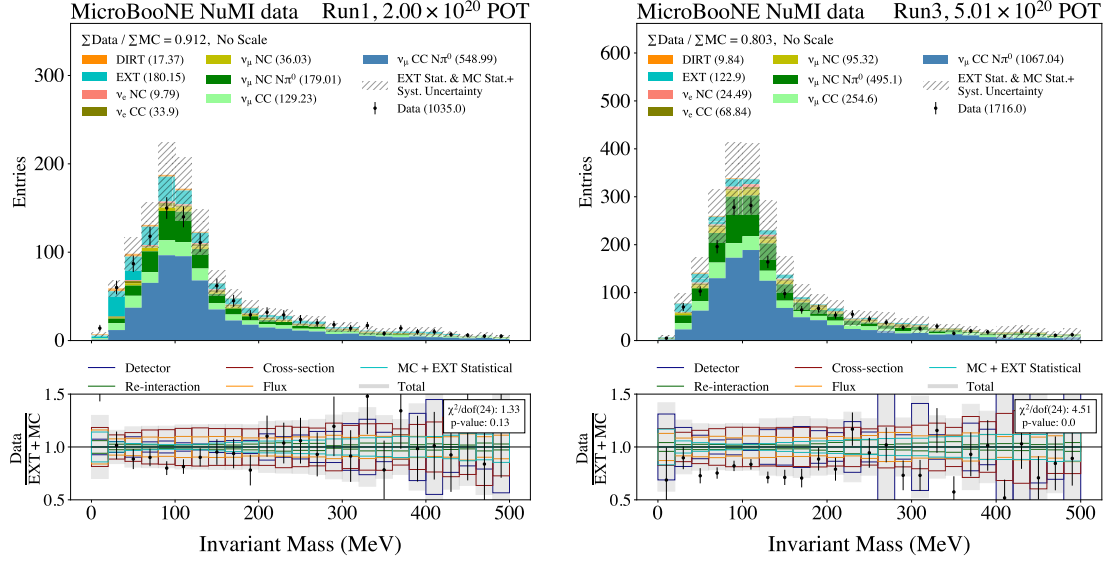


Figure 5.9: A comparison of the stacked prediction (EXT+MC) to the data (beam-on) of the invariant mass variable for Run 1 (left) and Run 3 (right) samples rich in π^0 . No scaling is applied to these distributions. The ratios of χ^2 to degrees of freedom and the corresponding p -value is shown in the text boxes in the lower panel.

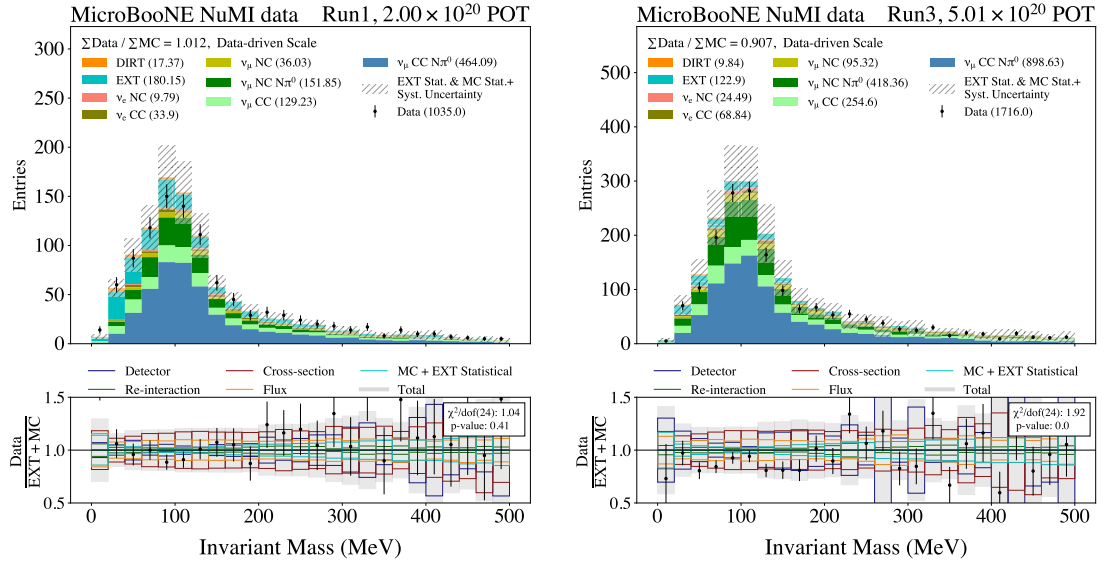


Figure 5.10: A comparison of the stacked prediction (EXT+MC) to the data (beam-on) of the invariant mass variable for Run 1 (left) and Run 3 (right) samples rich in π^0 . A data-driven scaling is applied to these distributions. The ratios of χ^2 to degrees of freedom and the corresponding p -value is shown in the text boxes in the lower panel.

Post-selection, the distributions of the discriminating variable used in setting the limit for two-shower like events, the invariant mass, are shown in figures 5.9–5.11 for three different cases: π^0 rich MC predictions with no scaling applied in figure 5.9; a data-driven scaling applied in figure 5.10; and a π^0 normalisation scaling applied in

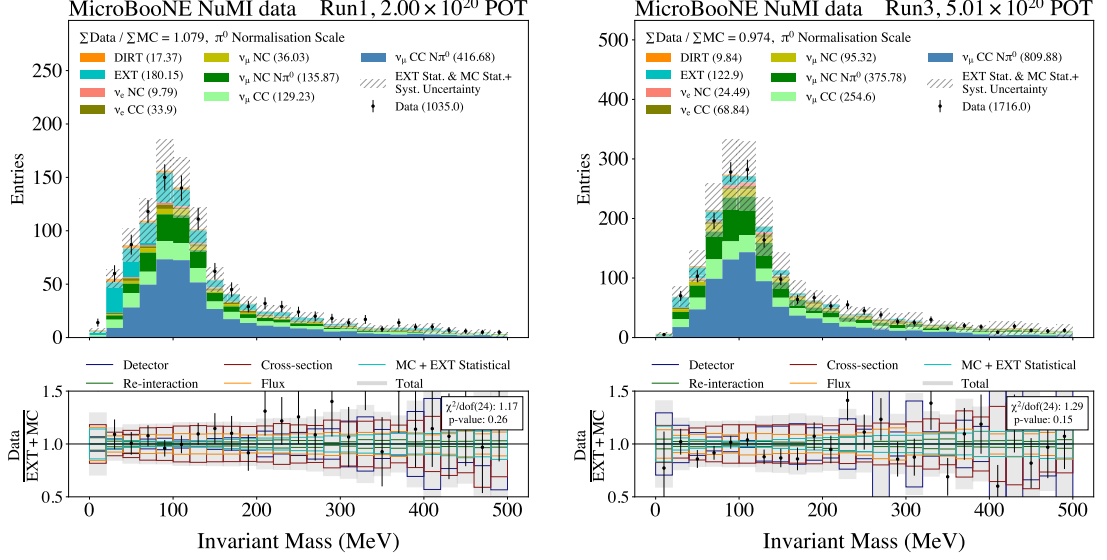


Figure 5.11: A comparison of the stacked prediction (EXT+MC) to the data (beam-on) of the invariant mass variable for Run 1 (left) and Run 3 (right) samples rich in π^0 . A π^0 normalisation scaling is applied to these distributions. The ratios of χ^2 to degrees of freedom and the corresponding p -value is shown in the text boxes in the lower panel.

figure 5.11. The calculation of the systematic uncertainty used in these distributions is discussed in chapter 8. In the case where no scaling is applied, the MC predictions are seen to be overestimated and as a result the p -value is calculated to be zero. In the case where data-driven scaling is applied, Run 1 shows a good data-MC agreement with larger p -value compared to Run 3 where the MC predictions are still slightly overestimated with a p -value zero. Finally, in the case where the π^0 normalisation scaling is applied, Run 1 and Run 3 both show good data-MC agreement with their corresponding p -values calculated to be 0.26 and 0.15 respectively.

Similarly to the BNB data, the agreement for the NuMI data is better in the case when a flat scaling is applied to the MC events containing a CC or NC interaction containing a π^0 . The NuMI and BNB predictions for the fluxes at MicroBooNE are completely independent and therefore a similar MC excess of π^0 enriched events in both the samples suggests that the cause of this data-MC disagreement is most likely due to a cross-section mis-modeling.

5.8.4 Signal samples

Higgs Portal Scalar Mass (MeV)	Run 1 Events (POT)	Run 3 Events (POT)
100	37877 (1.6×10^{23})	46327 (2.1×10^{23})
125	30865 (7.2×10^{22})	47387 (1.2×10^{23})
130	35899 (7.5×10^{22})	47861 (1.1×10^{23})
135	34947 (6.5×10^{22})	46368 (9.6×10^{22})
140	33868 (5.7×10^{22})	47109 (8.9×10^{22})
145	34053 (5.2×10^{22})	46659 (7.8×10^{22})
150	33404 (4.6×10^{22})	46262 (7.0×10^{22})
200	43725 (2.0×10^{22})	46350 (2.4×10^{22})

Table 5.3: The MC Higgs portal scalar samples for different mass values along with their corresponding number of events and POT used in the analysis.

The MC Higgs portal scalar generator described in section 5.1.1 is used to produce the signal samples across a range of targeted mass values. These signal samples along with their corresponding numbers of events and POT used in the analysis are listed in table 5.3. The scalar mass values range from 100 MeV to 200 MeV with additional samples generated between 100 MeV and 150 MeV where the sensitivity is expected to change rapidly due to background $\pi^0 \rightarrow \gamma\gamma$ decays for two-shower like signal events. Furthermore, in this mass range, the NA62 and E949 experiments are insensitive as discussed in section 1.3. This analysis is not expected to have competitive sensitivity beyond current experimental limits for scalar masses below 100 MeV or above 200 MeV, so samples for these masses were not generated or analysed.

Similar to MC neutrino overlay background, the signal samples too are overlaid on top of the “beam-off” sample containing cosmic muons using the process described in section 5.3.

Scalar Mass (MeV) $\theta(10^{-4})$	100	125	130	135	140	145	150	200
	5.36	5.50	5.54	5.58	5.63	5.68	5.73	6.58

Table 5.4: The values of θ corresponding to the KOTO central-value for various Higgs portal scalar mass values considered in this thesis.

Signal sample Normalisation

The signal normalisation is dependent on the mixing parameter θ , the parameter on which the limit is set in the analysis presented in this thesis. The Higgs portal scalar generator described in section 5.1.1 uses the values of θ corresponding to the value calculated by phenomenologists [81], that are required to explain the preliminary anomalous excess of neutral kaon decays reported by the KOTO collaboration in 2019 [82], which is now less significant [83]. These values of θ are listed in table 5.4 and are referred to as “KOTO central-values” in this thesis, although they are not directly provided by the KOTO collaboration.

Before setting the limit, the relationship between the signal event normalisation and θ must be evaluated. The number of Higgs portal scalar decays in the detector, N_{decay} , is a function of θ and M_S , and can be expressed as

$$N_{\text{decay}}(\theta, M_S) = N_{\text{POT}} \times N_{K-\text{decay}} \times K_{BR}(\theta, M_S) \times P_{SA} \times P_{\text{decay}}(\theta, M_S) \quad (5.5)$$

where $N_{K-\text{decay}}$ is the number of kaons decaying at rest at the target or in flight in the decay pipe or at the NuMI absorber per POT, and P_{SA} the probability that the trajectory of a Higgs portal scalar intersects with the detector. The terms $K_{BR}(\theta, M_S)$ and $P_{\text{decay}}(\theta, M_S)$ represent the probability of a kaon decaying to a Higgs portal scalar and the probability of a Higgs portal scalar decaying inside the detector, respectively. Both of these are functions of the square of the mixing parameter.

The branching ratio for the kaon decay to Higgs portal scalars is given by equation 1.4. The mixing parameter can be factorised out and can be used as a proportionality factor to the number of kaon decays to Higgs portal scalars as

$$K_{BR}(\theta, M_S) = \theta^2 \times K_{BR}(M_S). \quad (5.6)$$

The probability of the Higgs portal scalar decaying inside the detector can be expressed as

$$P_{\text{decay}}(\theta, M_S) = \left(e^{-\frac{\Gamma_T L_1}{\gamma\beta}} - e^{-\frac{\Gamma_T L_2}{\gamma\beta}} \right) \frac{\Gamma_{\text{signal}}}{\Gamma_T}, \quad (5.7)$$

where L_1 and L_2 are the distances from the source of scalar production to the scalar's entry and exit points of the detector. Γ_{signal} and Γ_T are the decay width to the signal channel and total decay width respectively. The total decay length ($\Gamma_T/\gamma\beta$) in the exponent is dependent on θ as discussed in section 1.2. For the mass range chosen in the analysis, the decay length is much longer than the distance to the detector (~ 700 m from the target and ~ 100 m from the absorber as discussed in section 3.3.1). Therefore assuming $\Gamma_T/\gamma\beta \gg L_1, L_2$ equation 5.7 can be expressed as

$$P_{\text{decay}}(\theta, M_S) \approx (L_1 - L_2) \left(\frac{\Gamma_T}{\gamma\beta} \right) \left(\frac{\Gamma_{\text{sig}}}{\Gamma_T} \right). \quad (5.8)$$

Factoring out the mixing parameter gives

$$P_{\text{decay}}(\theta, M_S) \approx \theta^2 \times P_{\text{decay}}(M_S). \quad (5.9)$$

Substituting this decay probability into equation 5.5 gives an event rate

$$N_{\text{decay}}(\theta, M_S) \approx \theta^4 \times N_{\text{decay}}(M_S), \quad (5.10)$$

which is directly proportional to θ^4 . When setting the limit (described in chapter 9), the KOTO central-values of θ used in the generator is normalised using this relationship to find the value of θ which corresponds to the number of events required for exclusion.

Chapter 6

Signal Pre-Selection

This chapter outlines the analysis workflow developed to discriminate a Higgs portal scalar decaying into an e^+e^- pair from the NuMI background comprising NuMI neutrino MC interactions, out-of-cryostat (dirt) and beam-off events. Section 6.1 outlines the decay kinematics of the Higgs portal scalars produced from KDAR and KDIF and their resultant decay products (e^+e^-). Section 6.2 outlines the selections that are applied to discriminate the two-shower signal topology from the background. Finally section 6.3 shows the cumulative selection efficiencies for each sample throughout the selection.

6.1 Decay Kinematics

The scalars produced from KDAR at the NuMI target and at the NuMI absorber are mono-energetic and have low momentum compared to scalars produced from KDIF which have energies ranging across a wide spectrum (depending on the angle at which the scalars are produced and the energy of the parent kaon). The decay of the scalar $S \rightarrow e^+e^-$ produces a two-shower like signature emerging from a common vertex in the detector volume. The decay kinematics are highly dependent on M_S . This section provides a brief overview of the true distributions of the decay features for the Higgs portal scalar signal.

6.1.1 Kaon Decay At Rest

For KDAR at the NuMI target and at the NuMI absorber, the energy and therefore the momentum of the scalar is calculated using two-body decay dynamics. For the process $K \rightarrow \pi S$ in the kaon rest frame,

$$q_K = q_\pi + q_S \quad \text{or} \quad q_\pi = q_K - q_S, \quad (6.1)$$

where q_K , q_π and q_S are the 4-momenta for the parent kaon and its daughter particles, the pion and scalar, respectively. Squaring the expression gives

$$q_\pi^2 = (q_K - q_S)^2 \quad (6.2)$$

$$= q_K^2 - 2q_K \cdot q_S + q_S^2. \quad (6.3)$$

Since the kaon is decaying at rest, $q_K = (E_K, \vec{p}_K) = (m_K, \vec{0})$ with $E_K = m_K$ and $\vec{p}_K = \vec{0}$ being the rest mass energy and momentum of the kaon. For the decay products, the pion has energy E_π and momentum \vec{p}_π and 4-momentum $q_\pi = (E_\pi, \vec{p}_\pi)$, and the scalar has energy E_S and momentum \vec{p}_S and 4-momentum $q_S = (E_S, \vec{p}_S)$. Substituting these 4-momenta into the above expression yields

$$m_\pi^2 = m_K^2 - 2(E_K E_S - \vec{p}_K \cdot \vec{p}_S) + M_S^2 \quad (6.4)$$

$$= m_K^2 - 2(m_K E_S) + M_S^2. \quad (6.5)$$

Re-arranging the above expression gives

$$E_\pi = \frac{m_K^2 + m_\pi^2 - M_S^2}{2m_K}, \quad (6.6)$$

$$E_S = \frac{m_K^2 + M_S^2 - m_\pi^2}{2m_K}. \quad (6.7)$$

Using the energy-momentum relation $p = \sqrt{E^2 - m^2}$, the momenta of the pion and scalar are given by

$$\begin{aligned} p_\pi &= \frac{\sqrt{(m_K^2 + m_\pi^2 - M_S^2)^2 - 4m_K^2 m_\pi^2}}{2m_K}, \\ p_S &= \frac{\sqrt{(m_K^2 + M_S^2 - m_\pi^2)^2 - 4m_K^2 M_S^2}}{2m_K}. \end{aligned} \quad (6.8)$$

An immediate consequence of this expression is that the kaon can only decay if the mass of the decay products is less than or equal to mass of the kaon. The energy and momentum of the scalar is entirely determined by the masses of the parent kaon and its decay products.

6.1.2 Kaon Decay In Flight

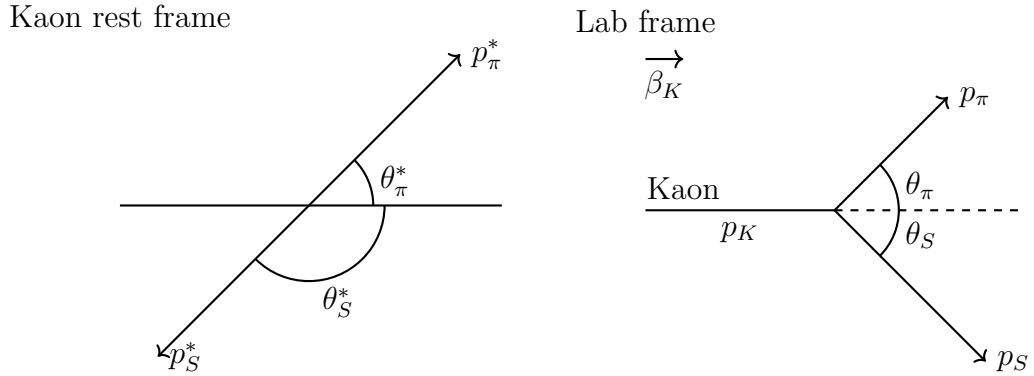


Figure 6.1: Schematic showing the kinematics of the kaon and its daughter pion and scalar in the kaon rest frame (left) and lab frame (right). In the kaon rest frame, the daughter particles emerge back-to-back with $\theta_\pi^* + \theta_S^* = 180^\circ$, whereas in the lab frame these angles do not sum up to 180° since $p_K \neq 0$.

Figure 6.1 shows the difference in kinematics of the daughter particles in the rest frame of the kaon and in the lab frame. In the kaon rest frame, the daughter particles emerge back-to-back with $\theta_\pi^* + \theta_S^* = 180^\circ$, whereas in the lab frame these angles do not sum up to 180° since $p_K \neq 0$ and consequently $p_\pi^* \neq p_\pi$ and $p_S^* \neq p_S$. For KDIF, the dot-product between the kaon and daughter particle 3-momenta in equation 6.7

is non-vanishing with the energy of the daughter particles defined as

$$E_\pi = \frac{m_K^2 + m_\pi^2 - M_S^2 + 2\vec{p}_K \cdot \vec{p}_\pi}{2m_K}, \quad (6.9)$$

$$E_S = \frac{m_K^2 + M_S^2 - m_\pi^2 + 2\vec{p}_K \cdot \vec{p}_S}{2m_K}. \quad (6.10)$$

The final-state momenta of the daughter particles are derived by boosting the p_π and p_S into the lab frame (or using energy-momentum conservation), and can be expressed as

$$\begin{aligned} p_\pi &= \frac{(m_K^2 + m_\pi^2 - M_S^2) p_K \cos \theta_\pi^* \pm 2E_K \sqrt{m_K^2 p_\pi^{*2} - m_\pi^2 p_K^2 \sin^2 \theta_\pi^*}}{2(m_K^2 + p_K^2 \sin^2 \theta_\pi^*)}, \\ p_S &= \frac{(m_K^2 + M_S^2 - m_\pi^2) p_K \cos \theta_S^* \pm 2E_K \sqrt{m_K^2 p_S^{*2} - M_S^2 p_K^2 \sin^2 \theta_S^*}}{2(m_K^2 + p_K^2 \sin^2 \theta_S^*)}. \end{aligned} \quad (6.11)$$

The angles θ_π and θ_S of the daughter particles with respect to the parent kaon direction are defined as

$$\begin{aligned} \tan \theta_\pi &= \frac{\sin \theta_\pi^*}{\gamma(\beta/\beta_\pi^* + \cos \theta_\pi^*)}, \\ \tan \theta_S &= \frac{\sin \theta_S^*}{\gamma(\beta/\beta_S^* + \cos \theta_S^*)}, \end{aligned} \quad (6.12)$$

where β is the velocity of the kaon and β_π^* and β_S^* are the velocity of the pion and the scalar in the kaon rest frame respectively. The opening angle between these two particles, $\alpha_{\pi S}$, can therefore be defined as $\alpha_{\pi S} = \theta_\pi + \theta_S$.

6.1.3 Scalar Decay

The scalar produced from the kaon itself decays to two products via the process $S \rightarrow e^+e^-$. Therefore, the same procedure used in calculating the decay kinematics for the KDIF process $K \rightarrow \pi S$ can be used to describe the decay kinematics of a scalar decay, $S \rightarrow e^+e^-$, in the MicroBooNE detector. This section provides a brief overview of the decay kinematics of the final state particles, e^+e^- , produced from a scalar originating at different locations along the NuMI beamline, including KDAR at the target, KDIF along the decay pipe and KDAR in the NuMI absorber.

Pair Production from a Scalar Produced from KDAR at the NuMI Target

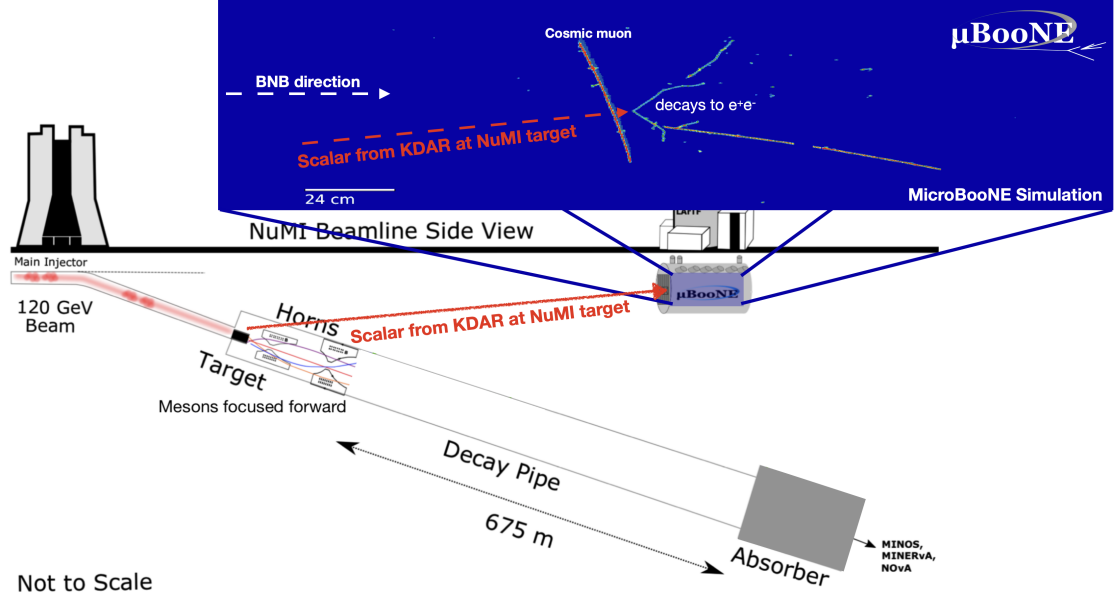


Figure 6.2: Event display of a simulated Higgs portal scalar signal produced from KDAR at the NuMI target. The scalar being produced at the NuMI target approaches the detector from the bottom left as indicated by the red arrow before subsequently decaying into an electron-positron pair inside the MicroBooNE detector. The nominal BNB direction is indicated by the horizontal white arrow on the left and the cosmic muon that entered from the surface is annotated with white text.

Figure 6.2 shows an event display of a Higgs portal scalar produced from KDAR at the NuMI target and its subsequent decay into an electron-positron pair inside the MicroBooNE detector. The distributions of momenta of these final-state e^+e^- pairs for three different scalar masses, 100, 150 and 200 MeV, are shown in the left hand panel of figure 6.3. The momenta of the two daughter particles are anti-correlated and the degree of asymmetry is determined by the kinematics of these daughter particles, which is entirely dependent on the mass of the scalar and the angle of the daughter particles with respect to the parent scalar direction.

The opening angles between the two showers, $\alpha_{e^+e^-} = \theta_{e^+} + \theta_{e^-}$, calculated by replacing π and S with e^+ and e^- in equation 6.12, for three different scalar masses, 100, 150 and 200 MeV, are shown in the right hand panel of figure 6.3. The distributions for the opening angles are highly peaked for the mono-energetic scalar with the most probable opening angle highly dependent on the scalar mass. As the scalar mass increases, β_S relative to β_{e^+} and β_{e^-} decreases and therefore the opening

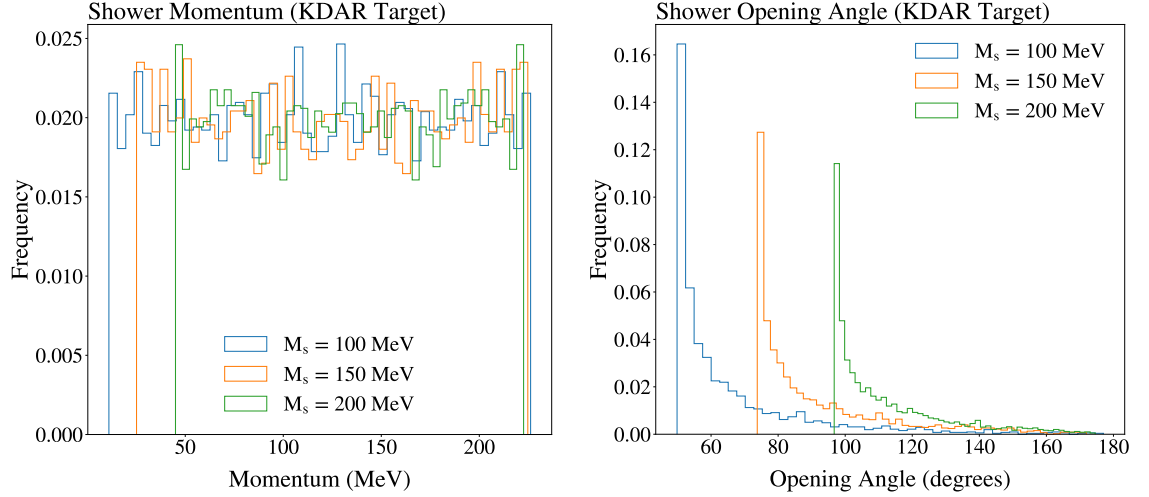


Figure 6.3: Left shows the true momentum distributions of the decay products (e^+e^-) of the Higgs portal scalars produced from KDAR at the target for three different scalar masses: 100, 150 and 200 MeV. Right shows the opening angle between these decay products for the same mass range.

angle increases.

Pair Production from Scalars Produced from KDAR at the NuMI Absorber

Figure 6.4 shows an event display of a Higgs portal scalar produced from KDAR at the NuMI absorber and its subsequent decay into an e^+e^- pair inside the MicroBooNE detector. As the scalar is produced from KDAR, the distributions of the final state showers' momenta and the opening angles between the showers are similar to those of the scalars produced from KDAR at the target. Therefore, the analysis presented in this thesis combines the scalars produced from KDAR at the target and KDAR at the NuMI absorber instead of treating the two separately.

Pair Production from Scalars Produced from KDIF

Figure 6.5 shows an example event display of a Higgs portal scalar produced from KDIF and its subsequent decay into an e^+e^- pair inside the MicroBooNE detector. The scalar can approach the detector across a wide angular range as described in section 3.4 before decaying into an e^+e^- pair. The energy and thus the momentum

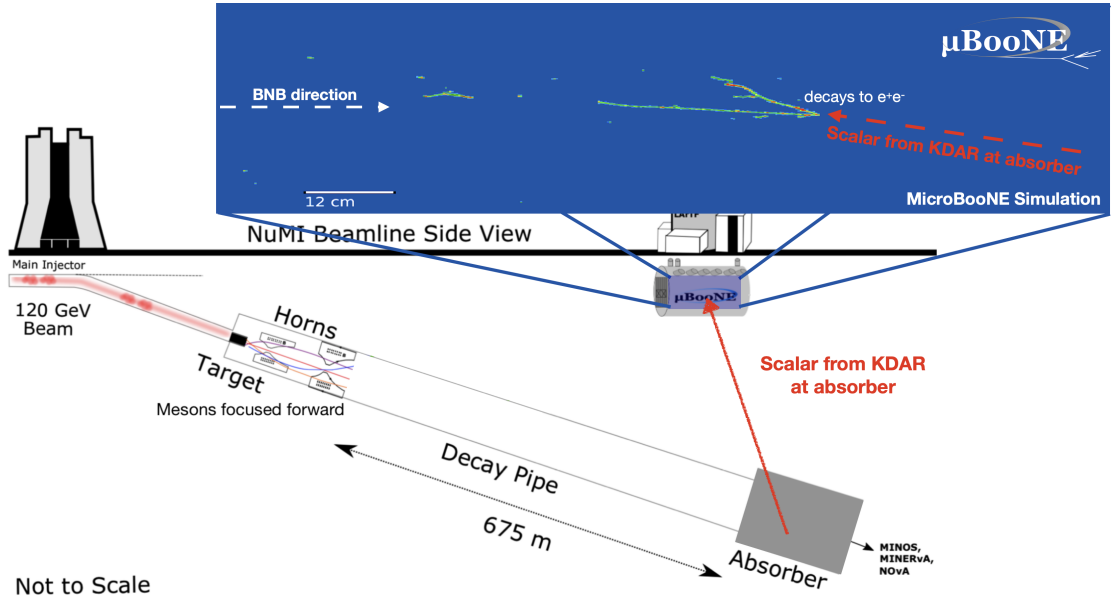


Figure 6.4: Event display of a simulated Higgs portal scalar signal produced from KDAR at the NuMI absorber. Being produced at the NuMI absorber, the scalar approaches the detector from bottom right as indicated by the red arrow before subsequently decaying into an electron-positron pair inside the MicroBooNE detector. The sideview of the NuMI beamline is not to scale and therefore the angle of the red arrow with respect to the NuMI beamline does not match with the one shown in the event display. The BNB direction is indicated by the horizontal white arrow on the left.

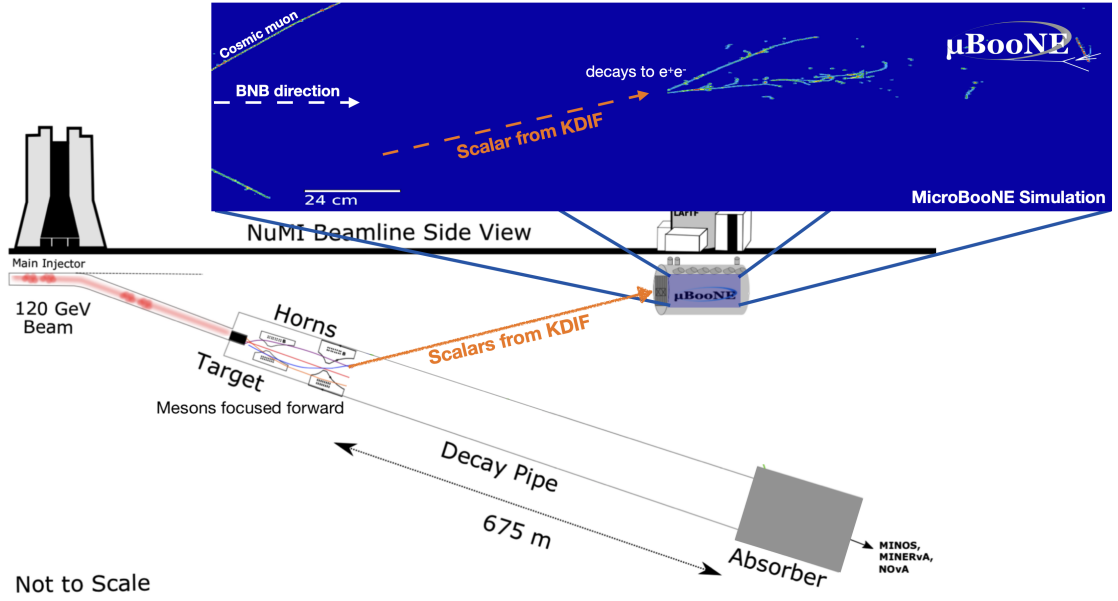


Figure 6.5: An example event display of a simulated Higgs portal scalar produced from KDIF. The scalar in this example is approaching the detector from bottom left (indicated by an orange arrow) with the kaon decaying close to the NuMI target. In general, the kaon can decay at any location along the decay pipe and therefore the scalar can approach the detector across a wide angular range described in section 3.4 before subsequently decaying into an electron-positron pair inside the MicroBooNE detector. The BNB direction is indicated by the horizontal white arrow on the left and the cosmic muon entering from the surface is annotated with the white text on the top left.

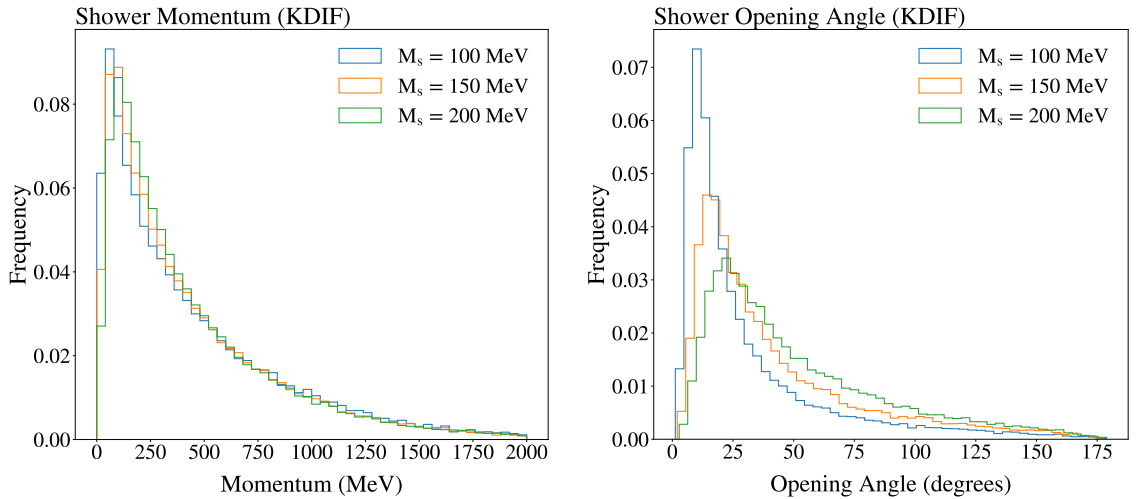


Figure 6.6: Left shows the true momentum distributions of the decay products (e^+e^-) of the Higgs portal scalars produced from KDIF for three different scalar masses: 100, 150 and 200 MeV. Right shows the opening angle between these decay products for the same mass range.

of the scalar is dependent on this angle due to a similar argument as used for the neutrino angle and energy relationship discussed in section 3.3.2. Figure 6.6 (left) shows the distributions of momenta of the final state particles of the scalars produced from KDIF for various mass values considered in the analysis, whereas figure 6.6 (right) shows the opening angle between the scalar's daughter particles. The most probable opening angle is dependent on the scalar mass and the spread in the distribution is due to the varied values of β_S for scalars produced from KDIF.

6.1.4 Separation of Kaon Decay In Flight and Kaon Decay At Rest

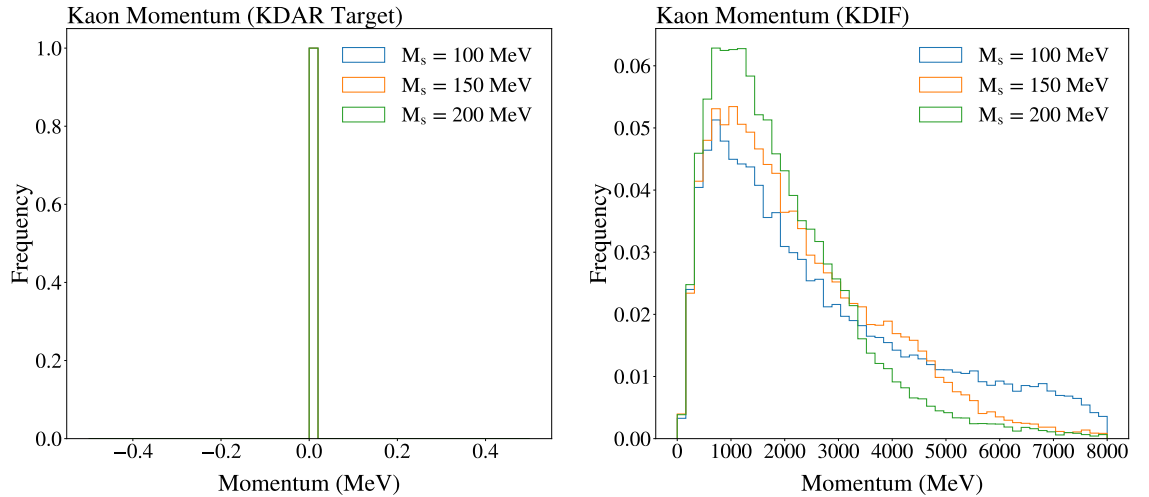


Figure 6.7: Truth level distributions of the momentum of the kaon decaying at rest at the target (left) and the momentum of the kaon decaying in flight (right). The distributions for different masses of the scalar overlap in the case of KDAR at target. The momentum distributions for KDAR at the NuMI absorber are not shown as they are similar to KDAR at the target.

As discussed in section 5.1.1, the Higgs portal scalars are generated using a “generic” generator where scalars are produced from KDAR at the target and at the NuMI absorber as well as from KDIF. Also discussed in the previous section, the scalars produced from KDAR at the target and at the NuMI absorber are combined. This leads to the formation of two types of signal: scalars produced from KDAR and scalars produced from KDIF. The analysis presented in this thesis sets a limit using these two types of signal.

Figure 6.7 shows the true momentum of the kaon decaying at rest at the target (left) and the kaon decaying in flight (right). The scalars produced from KDAR at the target and at the NuMI absorber have their parent kaon momenta equal to zero, whereas a scalar produced from KDIF has a non-zero parent kaon momenta. To separate KDAR and KDIF in the MC, this truth level information for the momenta of the scalar's parent kaon is used. The selection for these two types of signal is similar but the BDT models trained to separate these signals from background are different (as discussed in chapter 7).

6.2 Selection

This section outlines the following pre-selections that are applied to discriminate the two-shower signal topology produced from scalar from KDIF and KDAR shown in the event displays in the previous section from the background samples.

- **Pandora** slice identification to select the event that **Pandora** classifies as a clear neutrino interaction is described in section 6.2.1.
- Cosmic ray tagging to veto the cosmic rays in the Run 3 data-set is described in section 6.2.2.
- The number of reconstructed showers with interaction vertex contained within the fiducial volume (described in section 2.3) of the MicroBooNE detector is used to select events with exactly two showers, since our signal is showers induced by electron-positron pairs. In addition, one-shower like events are also selected to enhance the final sensitivity. This selection is described in section 6.2.3.
- **Pandora** track-shower score selection to exclude the events that **Pandora** classifies as track-like is described in section 6.2.4.

In addition, this section also compares stacked predictions (beam-off + MC) to data (beam-on) for Run 1 and Run 3 distributions of the variables used in the selection.

These distributions are classified into various categories described in section 5.8.2. The normalisation procedure described in section 5.8.3 is followed to normalise these distribution. The systematic uncertainties on these distributions are evaluated using the procedure described in chapter 8. These distributions are overlaid on top of the area normalised event distributions for the two types of signal: scalars produced from KDAR and KDIF discussed in the previous section. An example scalar mass $M_S = 150$ MeV is considered for these two types of signal. The systematic uncertainties on the signal (also described in chapter 8) are not shown on these distributions for clarity. The events in these distributions are the interactions that have passed the software trigger described in section 2.5.2.

6.2.1 Slice Identification

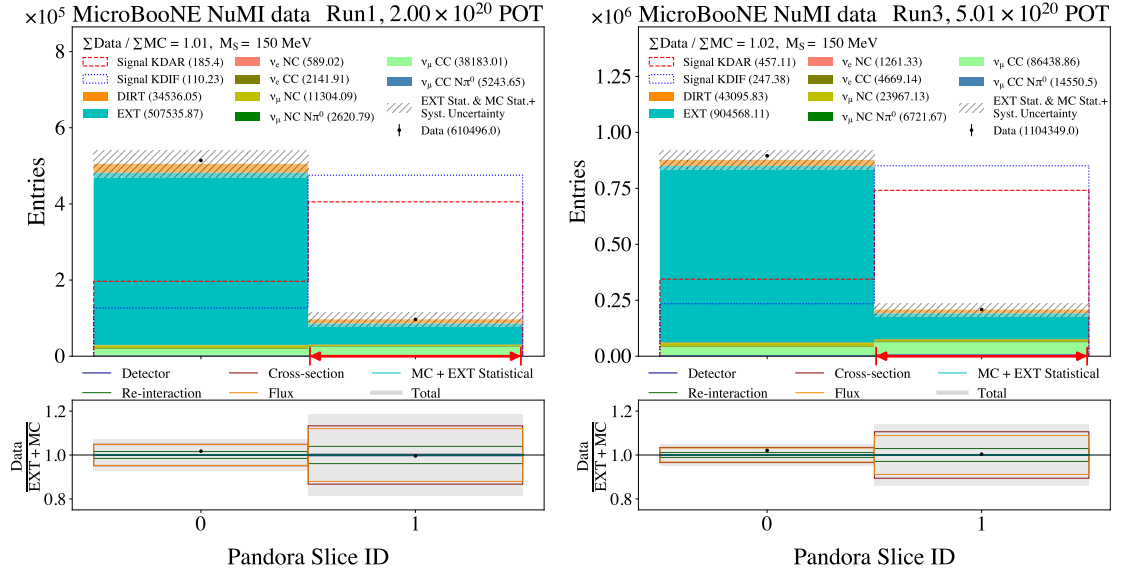


Figure 6.8: A comparison of the stacked prediction (beam-off+MC) to the data (beam-on) for the **Pandora Slice ID** variable. The distribution is overlaid on top of the area-normalised Higgs portal scalar distribution for a scalar with mass $M_S = 150$ MeV produced from KDAR (red dashed line) and KDIF (blue dashed line). The distributions on the left and right correspond to samples for Run 1 and Run 3 respectively. The selected region is indicated by the red interval on the x -axis.

As discussed in section 5.6, the first stage of the **Pandora** reconstruction framework is to group all the reconstructed particles belonging to the same interaction. The obvious cosmic activity is removed and the remaining hit information is used to

construct slices under the two beam or cosmic hypotheses. **Pandora** assigns a value equal to 0 to a slice that favours the cosmic hypothesis, whereas a value equal to 1 is assigned if a slice favours the beam hypothesis.

Figure 6.8 shows the distribution of the **Pandora** slice ID variable for Run 1 (left) and Run 3 (right) overlaid on top of the area-normalised Higgs portal scalar distribution for a scalar with mass $M_S = 150$ MeV produced from KDAR and KDIF. Events favouring the cosmic hypothesis with a value equal to 0 are rejected and events favouring the beam hypothesis, indicated by the red interval, are accepted. The Higgs portal scalar signal mostly favours beam hypothesis compared to cosmic hypothesis. Post-selection, a significant fraction of cosmic activity and out-of-cryostat interactions are rejected.

6.2.2 Cosmic Ray Tagging

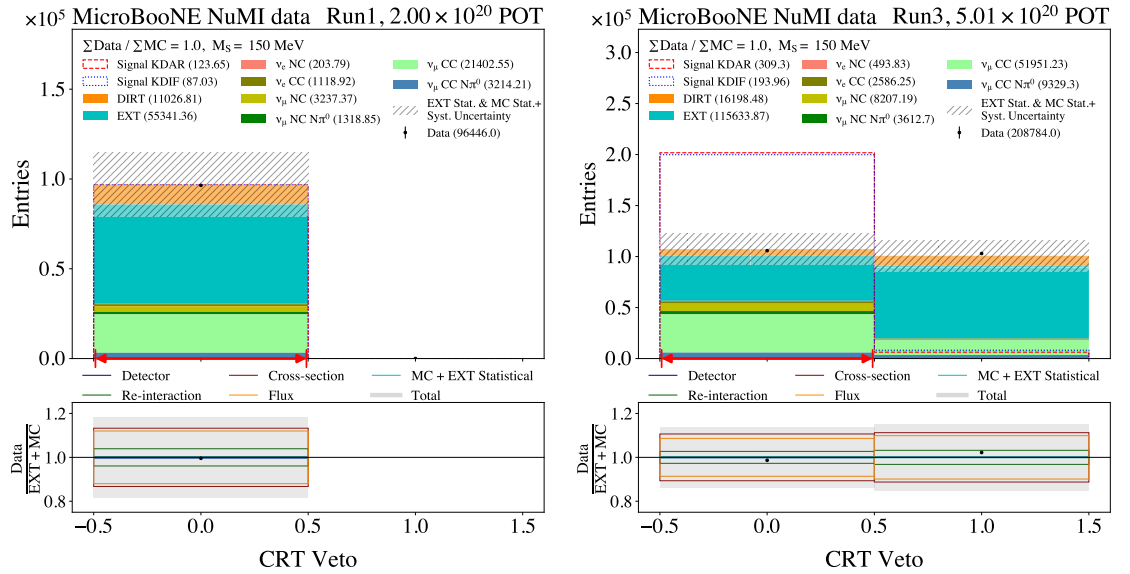


Figure 6.9: A comparison of the stacked prediction (beam-off+MC) to the data (beam-on) for the **CRT Veto** variable. The distribution is overlaid on top of the area-normalised Higgs portal scalar distribution for a scalar with mass $M_S = 150$ MeV produced from KDAR (red dashed line) and KDIF (blue dashed line). The distributions on the left and right correspond to samples for Run 1 and Run 3 respectively. The CRT system was not installed when samples for Run 1 were being recorded and therefore the distribution on left is single-binned with no cosmic tagging via the CRT. The selected region is indicated by the red interval on the x -axis.

Further cosmic activity is removed using the MicroBooNE Cosmic Ray Tagger (CRT)

system discussed in section 2.4. The CRT system was installed part-way through the detector operation and therefore this selection can only be applied to data taken in the RHC configuration of the beam, which is Run 3. The CRT system vetoes events of cosmic origin by searching for a time coincidence between the hits recorded by the CRT panels and the optical flash within the NuMI beam spill time window (described in section 2.5). A value of 1 is assigned if a CRT hit is within $1\ \mu\text{s}$ of the beam flash, 0 otherwise.

Figure 6.9 shows the distribution of the CRT veto variable for Run 1 (left) and Run 3 (right) overlaid on top of the area-normalised Higgs portal scalar distribution for a scalar with mass $M_S = 150\ \text{MeV}$ produced from KDAR (red dashed line) and KDIF (blue dashed line). A very tiny fraction of the signal events are tagged as cosmics. Post-selection (region indicated by the red bar on the x -axis), a significant fraction of beam-off and out-of-cryostat events are rejected in the Run 3 data-set.

6.2.3 Number of Contained Showers

Figure 6.10 shows the distribution of the number of reconstructed showers with interaction vertex contained inside the fiducial volume of the MicroBooNE detector for Run 1 (left) and Run 3 (right) overlaid on top of the area-normalised Higgs portal scalar distribution for a scalar with mass $M_S = 150\ \text{MeV}$ produced from KDAR (red dashed line) and KDIF (blue dashed line).

The event rate for the desired signal of two-shower like events (shown by the red interval on the x -axis) is more than a quarter of the total signal events with a significant fraction of the remaining signal events being reconstructed with either zero or one contained shower. This is due to limited reconstruction and shower identification efficiencies of the **Pandora** software framework, where showers in an event are either mis-reconstructed or are mis-identified as tracks. A small opening angle between the two showers (described in section 6.1.3) causes showers to overlap on top of each other and this can lead to mis-reconstruction of the showers with **Pandora** reconstructing two separate showers as one shower. Therefore, the analysis

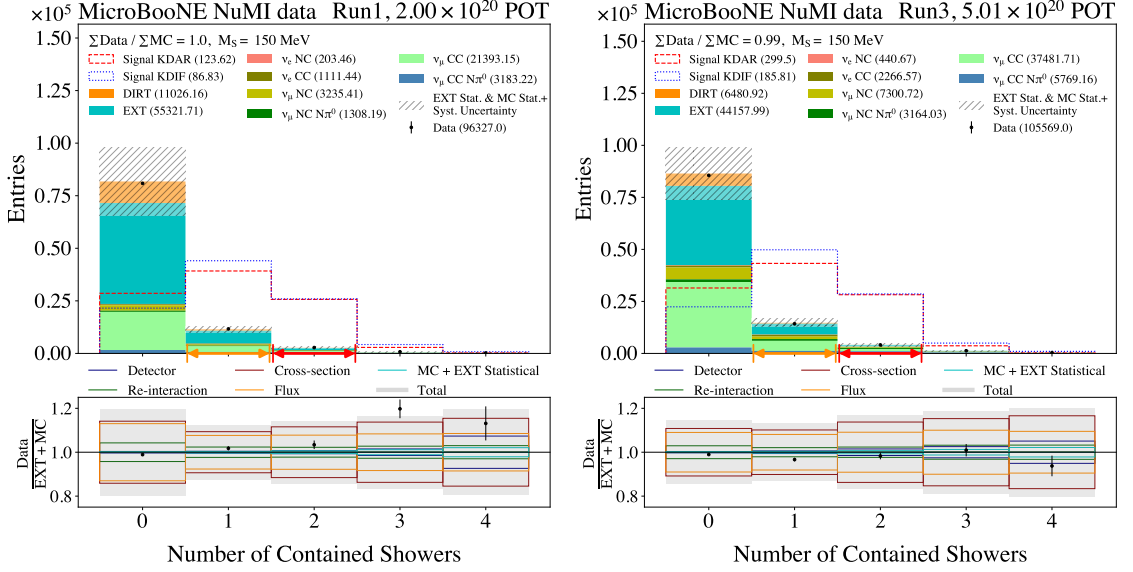


Figure 6.10: A comparison of the stacked prediction (beam-off+MC) to the data (beam-on) for the number of showers with interaction vertex contained inside the fiducial volume (described in section 2.3) of the MicroBooNE detector. The distribution is overlaid on top of the area-normalised Higgs portal scalar distribution for a scalar with mass $M_S = 150$ MeV produced from KDAR (red dashed line) and KDIF (blue dashed line). The distributions on the left and right correspond to samples for Run 1 and Run 3 respectively. Selected regions are indicated by the interval in red on the x -axis for events with two showers reconstructed and the orange region on the x -axis for events with only one shower reconstructed.

presented in this thesis sets a limit (described in chapter 9) using events with two showers reconstructed as well as one shower (shown by the orange interval on the x -axis) reconstructed by the **Pandora** reconstruction framework.

The two reconstructed showers are then classified as leading and sub-leading showers. A leading shower in an event is the one that has more reconstructed hits compared to the other, sub-leading shower. In case of an event where both the showers have an equal number of hits, the one with larger energy of the two is selected as the leading shower. The reason for this shower classification is to make the training set for the BDT (discussed in chapter 7) more effective because the leading shower has more hits and therefore has relatively better kinematic information.

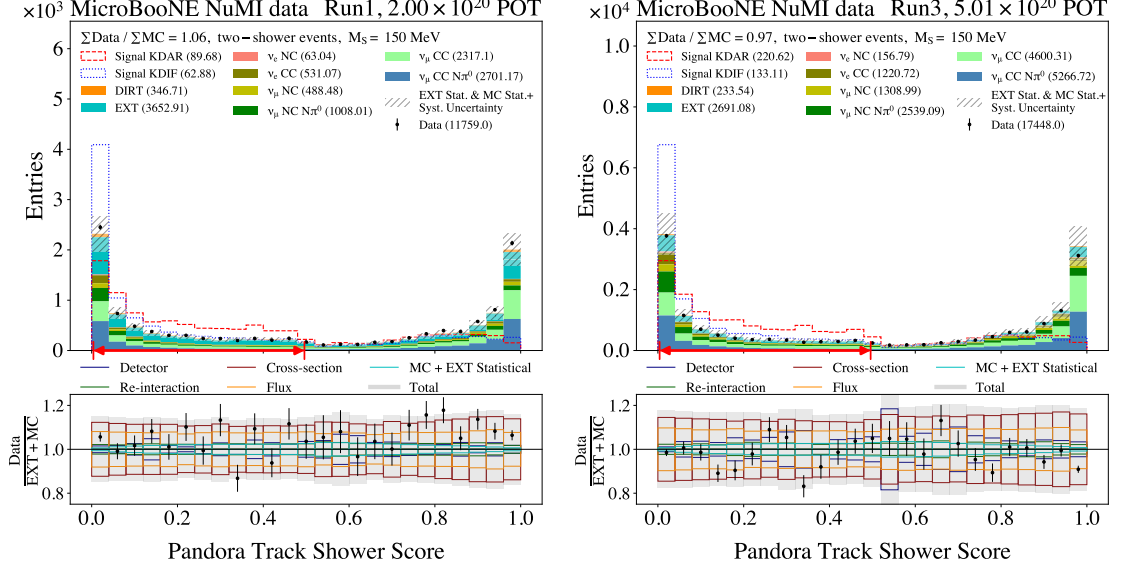


Figure 6.11: A comparison of the stacked prediction (beam-off+MC) to the data (beam-on) for the **Pandora Track-Shower Score** variable after the two-shower selection. The distribution is overlaid on top of the area-normalised Higgs portal scalar distribution for a scalar with mass $M_S = 150$ MeV produced from KDAR (red dashed line) and KDIF (blue dashed line). The entries in each bin correspond to the number of reconstructed candidate objects in an event and not the event rate. The distributions on the left and right correspond to samples for Run 1 and Run 3 respectively. The selected region is indicated by the interval in red on the x -axis.

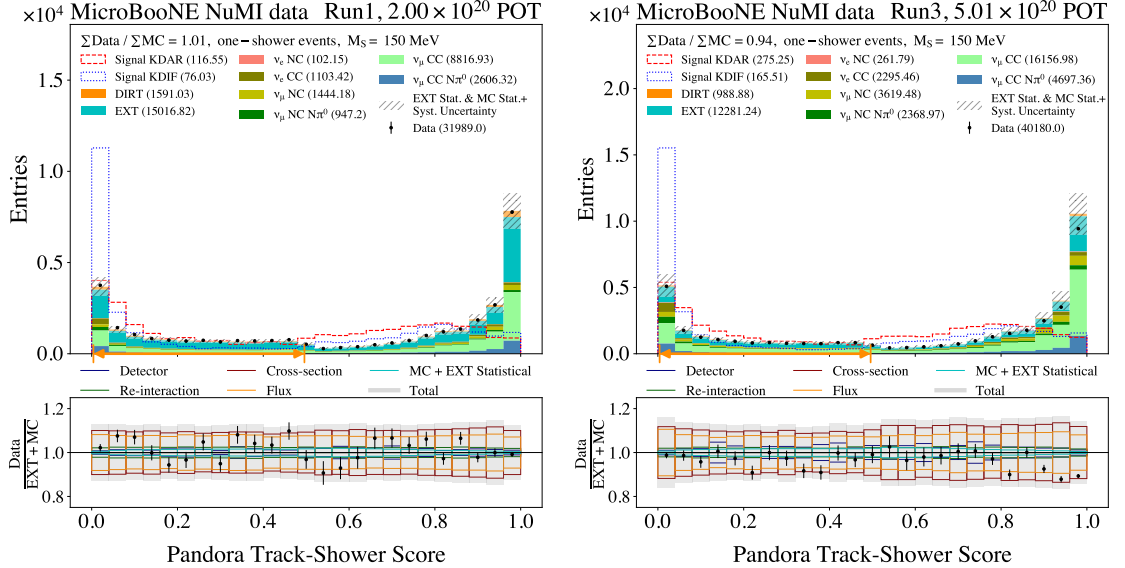


Figure 6.12: A comparison of the stacked prediction (beam-off+MC) to the data (beam-on) for the **Pandora Track-Shower Score** variable after the one-shower selection. The distribution is overlaid on top of the area-normalised Higgs portal scalar distribution for a scalar with mass $M_S = 150$ MeV produced from KDAR (red dashed line) and KDIF (blue dashed line). The entries in each bin correspond to the number of reconstructed candidate objects in an event and not the event rate. The distributions on the left and right correspond to samples for Run 1 and Run 3 respectively. The selected region is indicated by the interval in orange on the x -axis.

6.2.4 Pandora Track-Shower Score

As discussed in section 5.6, **Pandora** classifies the reconstructed objects in an event by assigning a score between 0 and 1 with 0 being an object that is completely shower-like and 1 being object that is completely track-like. Figures 6.11 and 6.12 show the distributions of the **Pandora** track-shower score for reconstructed objects in an event after the two-shower and one-shower selections respectively. The panels on the left show the distributions for Run 1 samples and the panels on the right show the distributions for Run 3 samples. These distributions are overlaid on top of the area-normalised Higgs portal scalar distribution for a scalar with mass $M_S = 150$ MeV produced from KDAR (red dashed line) and KDIF (blue dashed line). In the analysis presented in this thesis, reconstructed objects with scores in range 0 to 0.5 are selected as shower-like objects in an event. These ranges are indicated by the intervals on the x -axes in red for the two-shower selection and orange for the one-shower selection. Post-selection, a significant fraction of the track-like candidates are rejected in both one-shower and two-shower selected samples in Run 1 and Run 3.

6.2.5 Pre-Selection Summary

Figures 6.13 and 6.14 show the distributions of the shower energy and invariant mass for the one-shower and two-shower pre-selections, respectively. In these figures, the distributions on the left and right correspond to samples for Run 1 and Run 3 respectively. These distributions are overlaid on top of the area-normalised Higgs portal scalar distributions for a scalar with mass $M_S = 150$ MeV produced from KDAR (red dashed line) and KDIF (blue dashed line). In the analysis presented in this thesis, these distributions are eventually used as discriminatory variables in the limit setting softwares to set the final limit (described in chapter 9) after the further BDT selection, described in chapter 7, has been applied.

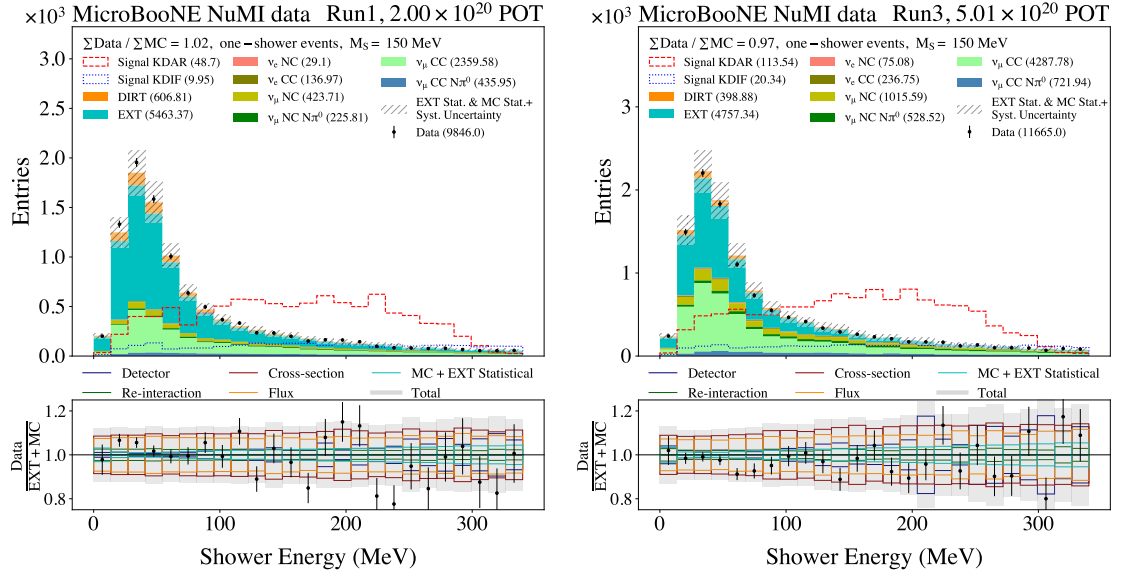


Figure 6.13: A comparison of the stacked prediction (beam-off+MC) to the data (beam-on) for shower energy after the one-shower selection. The distribution is overlaid on top of the area-normalised Higgs portal scalar distribution for a scalar with mass $M_S = 150$ MeV produced from KDAR (red dashed line) and KDIF (blue dashed line). The distributions on the left and right correspond to samples for Run 1 and Run 3 respectively.

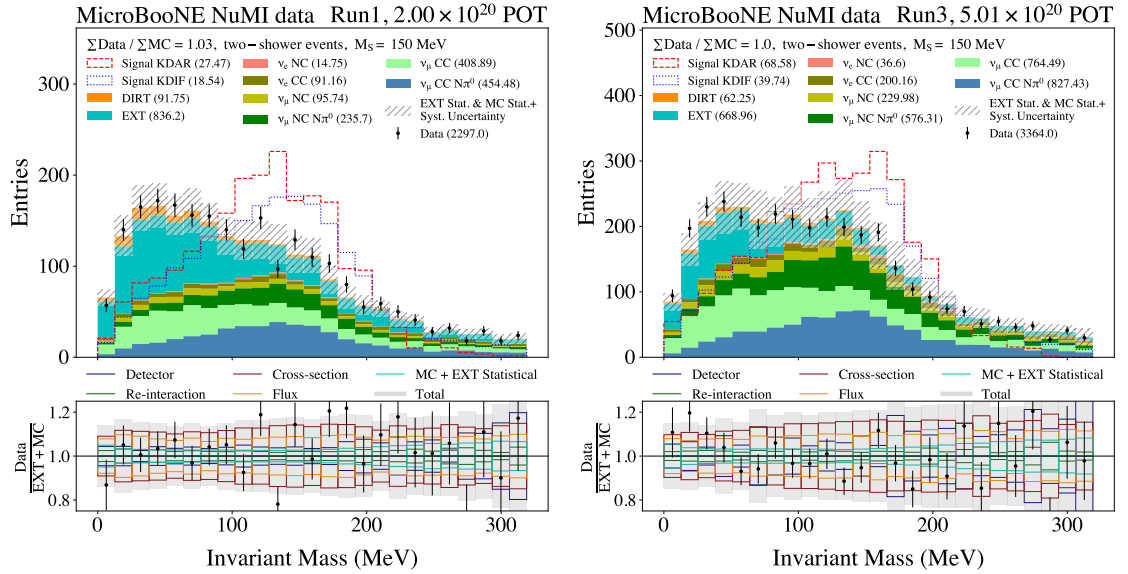


Figure 6.14: A comparison of the stacked prediction (beam-off+MC) to the data (beam-on) for the invariant mass after the two-shower selection. The distribution is overlaid on top of the area-normalised Higgs portal scalar distribution for a scalar with mass $M_S = 150$ MeV produced from KDAR (red dashed line) and KDIF (blue dashed line). The distributions on the left and right correspond to samples for Run 1 and Run 3 respectively.

6.3 Pre-Selection Efficiencies

The selection efficiency is defined as the number of events passing the selection divided by the total number of events. For the selection of the analysis presented in this thesis, the efficiencies are calculated by taking the ratio of the number of reconstructed events passing the selection to the total number of reconstructed events that have passed the software trigger described in section 2.5.2. The pre-selection in the analysis separates the two different types of signal samples, scalars from KDAR and KDIF, from the common NuMI background sample. Therefore the efficiencies for background and NuMI data are similar for these two different type of signals.

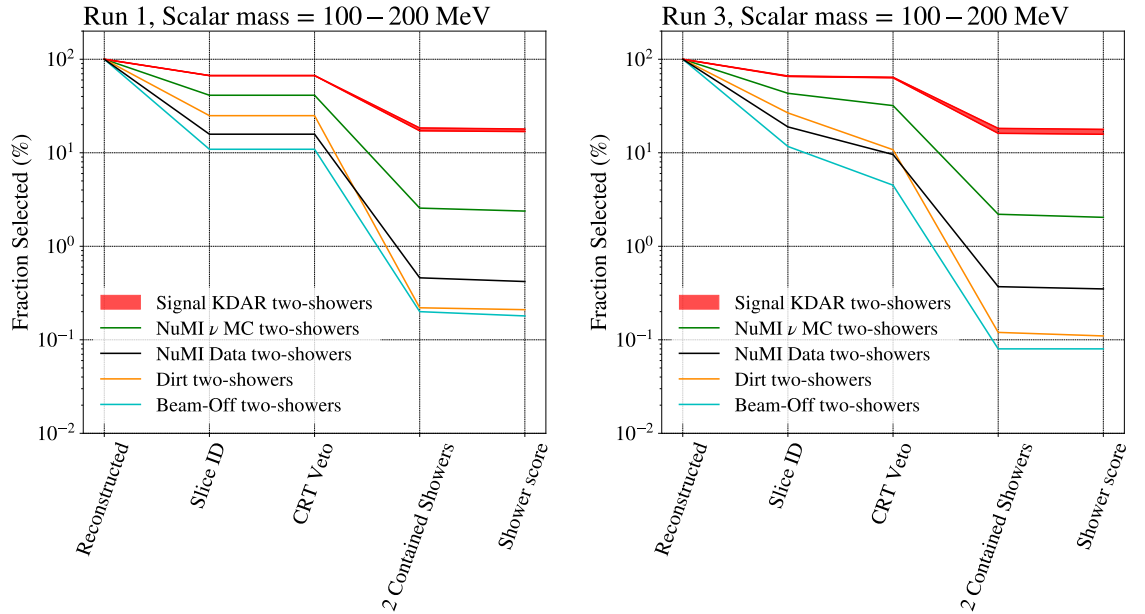


Figure 6.15: Run 1 (left) and Run 3 (right) cumulative pre-selection efficiencies for the two shower selection. The band in red shows the signal for scalars produced from KDAR with masses in the range ($M_S = 100 - 200$ MeV). Lines in green, orange and cyan show the cumulative pre-selection efficiencies of the background prediction comprising NuMI in-the-cryostat ν MC overlay (NuMI ν MC), out-of-cryostat ν MC overlay (Dirt) and beam-off samples, whereas the line in black shows the cumulative selection efficiency of the NuMI beam-on (data) sample.

Figure 6.15 and 6.16 show the cumulative selection efficiencies for the two-shower selection for the KDAR and KDIF signal (red band for scalars with mass in the range $100 - 200$ MeV), background (green, orange and cyan lines for NuMI in-the-cryostat ν MC overlay, out-of-cryostat ν MC overlay and beam-off) and data (black line).

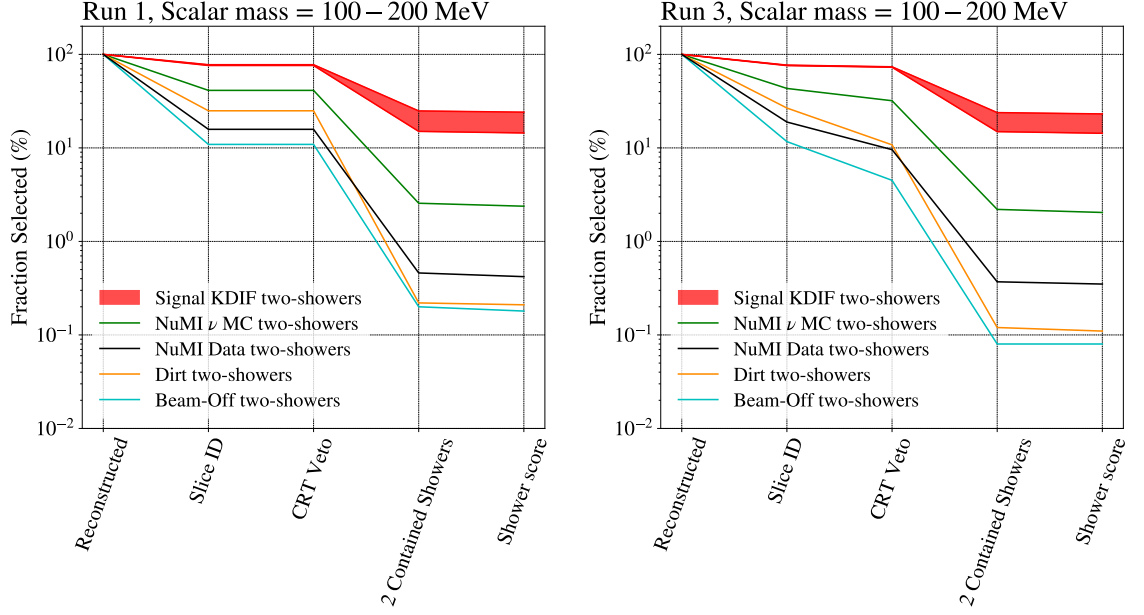


Figure 6.16: Run 1 (left) and Run 3 (right) cumulative pre-selection efficiencies for the two shower selection. The band in red shows the signal for scalars produced from KDIF with masses in the range ($M_S = 100 - 200$ MeV). Lines in green, orange and cyan show the cumulative pre-selection efficiencies of the background prediction comprising NuMI in-the-cryostat ν MC overlay (NuMI ν MC), out-of-cryostat ν MC overlay (Dirt) and beam-off samples, whereas the line in black shows the cumulative selection efficiency of the NuMI beam-on (data) sample.

The left and right panels show the selection efficiencies for the Run 1 and Run 3 samples respectively. For both KDAR and KDIF, the out-of-cryostat ν MC overlay and beam-off samples are reduced to $\sim 0.2\%$ for Run 1 and $\sim 0.1\%$ for Run 3. In-the-cryostat ν MC samples for these cases are reduced to about 2% for Run 1 with a slight improvement in the case of Run 3 due to the CRT Veto. The NuMI data samples for these cases are reduced to about 0.3 – 0.4% for Run 1 and Run 3. Approximately 20% of the KDAR two-shower and around 15 – 25% of the KDIF two-shower signal is selected for Run 1 and Run 3 samples in these two cases.

Figures 6.17 and 6.18 show the cumulative selection efficiencies for the one-shower selection for the KDAR and KDIF signal (red band for scalars with mass in the range 100 – 200 MeV), background (green, orange and cyan lines for NuMI in-the-cryostat ν MC overlay, out-of-cryostat ν MC overlay and beam-off) and data (black line). The left and right panels show the selection efficiencies for Run 1 and Run 3 samples respectively. For both KDAR and KDIF, the out-of-cryostat ν MC overlay and beam-off samples are reduced to $\sim 1\%$ for Run 1 and $\sim 0.4 - 0.7\%$ for Run 3. NuMI

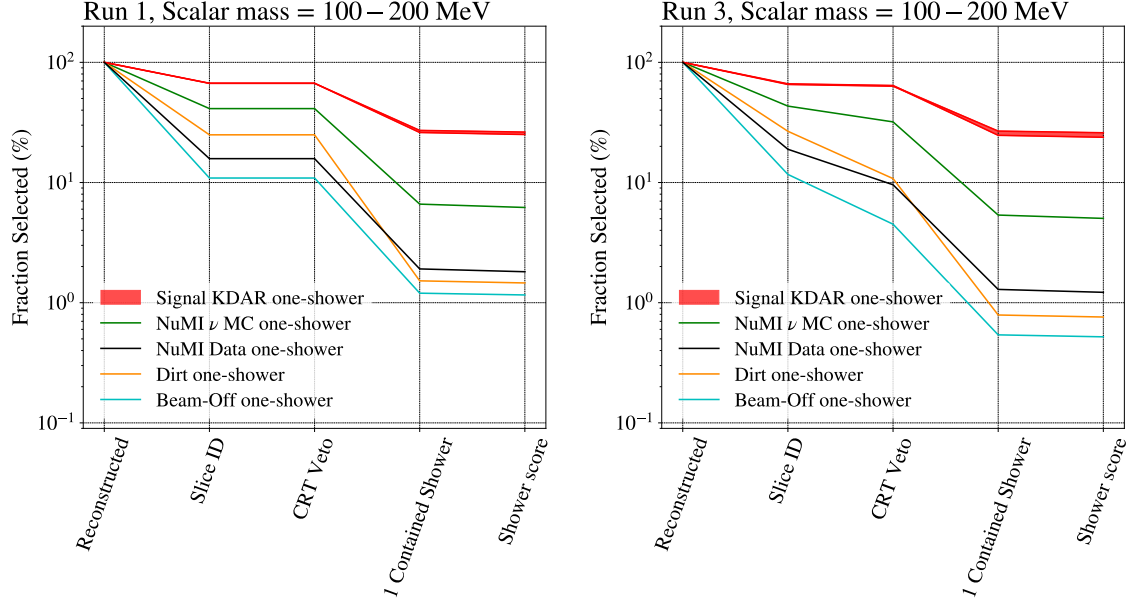


Figure 6.17: Run 1 (left) and Run 3 (right) cumulative pre-selection efficiencies for the one shower selection. The band in red shows the signal for scalars produced from KDAR with masses in the range ($M_S = 100 - 200$ MeV). Lines in green, orange and cyan show the cumulative pre-selection efficiencies of the background prediction comprising NuMI in-the-cryostat ν MC overlay (NuMI ν MC), out-of-cryostat ν MC overlay (Dirt) and beam-off samples, whereas the line in black shows the cumulative selection efficiency of the NuMI beam-on (data) sample.

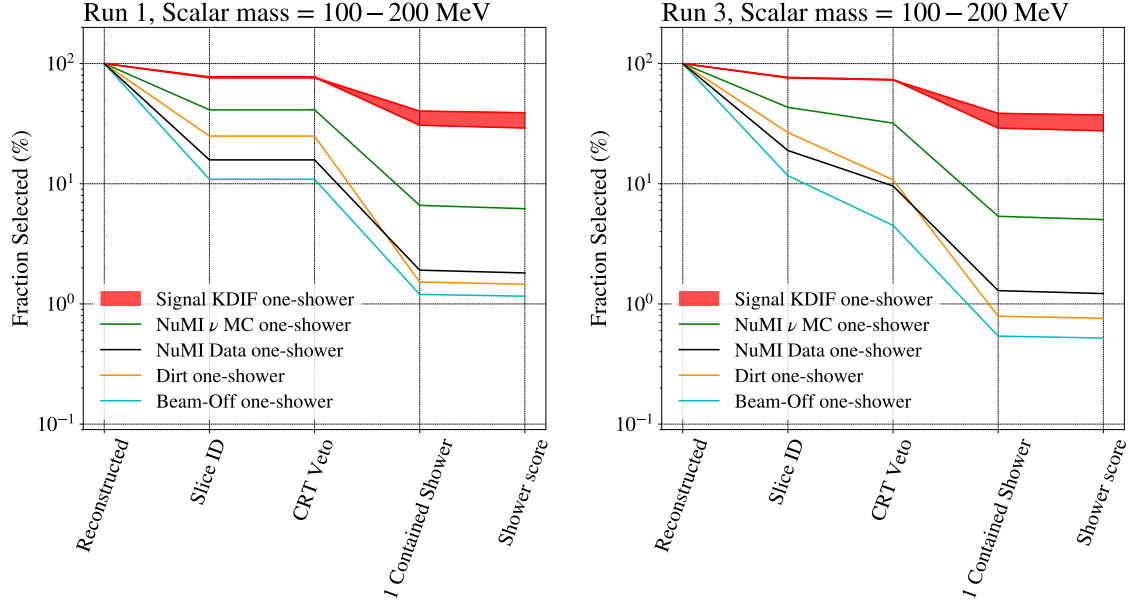


Figure 6.18: Run 1 (left) and Run 3 (right) cumulative pre-selection efficiencies for the one shower selection. The band in red shows the signal for scalars produced from KDIF with masses in the range ($M_S = 100 - 200$ MeV). Lines in green, orange and cyan show the cumulative pre-selection efficiencies of the background prediction comprising NuMI in-the-cryostat ν MC overlay (NuMI ν MC), out-of-cryostat ν MC overlay (Dirt) and beam-off samples, whereas the line in black shows the cumulative selection efficiency of the NuMI beam-on (data) sample.

in-the-cryostat ν MC overlay samples for these cases are reduced to about 6% for Run 1 and 5% for Run 3. A slight improvement in the background rejection for Run 3 is due to the additional CRT veto selection. The NuMI data samples for these cases are reduced to about 2% for Run 1 and slightly above 1% for Run 3. Approximately 30% of the KDAR one-shower and around 30 – 40% of the KDIF one-shower signal is selected for both Run 1 and Run 3. The CRT veto selection in Run 3 has almost no impact on the signal selection for both these cases.

Chapter 7

Boosted Decision Trees

The pre-selections described in chapter 6 have limited performance in classification of the two signal and background classes as these selections are rectangular with limited dimensionality in variable space. To separate these two classes in higher-dimensionality, a supervised multivariate machine learning technique called Boosted Decision Trees (BDTs) is used. The operational principle of this machine learning technique (BDT) is described in section 7.1, followed by its application to the analysis of the thesis, the Higgs portal scalar search, which is described in section 7.2. Lastly, the performance of the BDTs trained to separate the two classes (signal and background) in the analysis is described in section 7.3.

7.1 Boosted Decision Trees Definition

The term “decision tree” is attributed to the tree-structured supervised learning approach, where the root or the base node in a decision tree represents an entire dataset, internal nodes represent the variables (or features) of a dataset, branches represent the decisions, and finally leaves represent the outcomes of the decisions as shown in figure 7.1. The leaf is the final node of a decision tree which cannot be segregated further and therefore does not contain any further branches. The basic operational principle of this algorithm is to create a decision tree by recursively

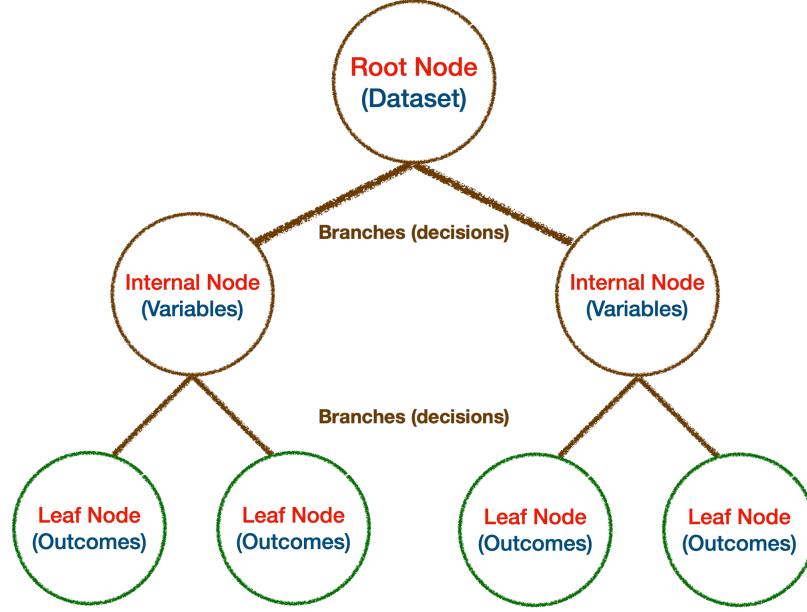


Figure 7.1: Schematic showing the components of a decision tree where the root node represents an entire dataset, internal nodes represent the variables of a dataset, branches (brown) represent the decisions and finally leaves (green) represent the outcomes of the decisions.

asking questions to partition data by making decisions on the basis of the variable (or feature) value until some stopping condition is reached. The decision trees mimic the human level thinking and are thus easy to interpret and understand. Among numerous other usages, decision trees are primarily used in classification and regression, but for the purposes of the analysis presented in this thesis, we will use them to classify our binary classes: signal and background. The algorithm for classification using decision trees works as follows:

1. Begin the tree with the root node by taking the complete dataset, S , comprising all signal and background events.
2. Sort all the events in S according to each variable.
3. Select the variable and splitting value that provides the best separation and split the node into two branches with one containing events that satisfy the criterion and the other containing events that fail it (described in section 7.1.1).
4. For each node, recursively apply from steps 2 and 3 until a stopping condition is reached (described in section 7.1.2).

All variables can be considered at each node regardless of the number of times they have been used in previous iterations to enable the algorithm to find and select intervals of interest within each variable.

7.1.1 Selection of Variables in Decision Trees

Selecting variables to split nodes is a complicated process and it is the core of the decision tree algorithm. In order to select a variable, the following two criteria are used: entropy and information gain.

Entropy

Entropy is defined as the measure of the randomness in a system. In the case of decision tree, the system is “information” that is being processed and the entropy is a measure of randomness or uncertainty of a variable (or feature). Ideally, a branch with zero entropy is known as a leaf node and a branch with non-zero entropy is split into two new nodes using a feature until it reaches a leaf node.

In the case of classification of two classes with $i = 0$ for background and $i = 1$ for signal, the entropy for the root node with complete dataset S or some internal node with subset s can be written as

$$E(S \text{ or } s) = - \sum_{i=0}^1 p_i \log_2(p_i) \quad (7.1)$$

where p_i is the fraction of dataset S or subset s which is in class i at a root node or internal node, respectively. The objective of the decision tree is to minimise this entropy¹. To measure the change in entropy at each node before and after the application of a feature, the algorithm uses information gain.

¹Example: Applying feature 1 at node 1 discriminates 5 signal events and 2 background events, whereas applying feature 2 at node 2 discriminates 3 signal events and 2 background events. The entropy at node 1 is calculated to be $E_1 = -(5/7) \log_2(5/7) - (2/7) \log_2(2/7) \approx 0.86$. Similarly $E_2 \approx 0.97$. Therefore as $E_1 < E_2$, the algorithm chooses feature 1 with less entropy to discriminate the two classes.

Information Gain

Information gain in decision tree is a statistical property that calculates how well a given variable separates the two classes in a training dataset according to their target classification by measuring the changes in entropy before and after splitting a dataset based on a variable, X . In terms of entropy, E , the information gain, IG , can be mathematically expressed as

$$IG = E(s) - E(s, X), \quad (7.2)$$

where $E(s, X)$ is the entropy of the dataset given some feature X . In addition to minimising entropy, it is also crucial for a decision tree algorithm to find a variable that yields the highest information gain. The algorithm then chooses the most important feature with the highest gain to be the root or base node.

7.1.2 Stopping Conditions of the Decision Trees

The stopping conditions of the decision trees to stop splitting and declare final nodes as leaves are as follows:

- The size of the sample at the node has reached its statistical limit.
- The decision tree has fully separated the two classes and all the events now belong to the same class in the final node.
- The splitting values for each variable cannot further improve the separation.
- A user defined tree depth (number of iterative nodes) has been reached. In the analysis presented in this thesis, the maximum depth for a tree was set to 6.

7.1.3 Boosting

Boosting is a general technique used in machine learning such as neural networks or decision trees to construct a strong learning model (or a classifier) by combining multiple weak learning models (or classifiers). A weak classifier is a model that has low prediction accuracy, marginally better than random guessing. A weak classifier is prone to over-fitting as it cannot classify data that exhibits large fluctuations from the training data-set.

The Boosting method combines such weak classifiers into a more strong and stable classifier with a smaller error rate and better performance in classification. Some of the most common types of boosting are Adaptive Boosting or **AdaBoost**, Gradient Boosting and Extreme Gradient Boosting or **XGBoost**.

Adaptive Boosting

One of the earliest boosting models developed that adapts and self-corrects by assigning a larger weight to the mis-classified events after every decision tree produced in the training process. The decision trees are recursively produced until the difference between the predicted and actual values (also known as residuals) falls below an acceptable threshold [84]. This boosting method is not suitable for a large data-set or when the features in a data-sets are correlated.

Gradient Boosting

Unlike AdaBoost, Gradient Boosting does not assign weights to the mis-classified events. Instead, it optimises an objective function, defined as the sum of the training loss function and regularization function. The training loss function measures residuals, whereas the regularization function controls the complexity of the model. A common choice of loss function for classification is logistic regression and since the analysis classifies two classes (signal and background), the BDTs in the analysis are also trained using logistic regression. The regularization function adds penalty as

model complexity increases by decreasing the importance given to higher order terms so that the model can be described by a less complex equation. This regularization is especially important when the dataset has large number of features. Gradient Boosting employs two type of regularisation to prevent the model from over-fitting and to address the feature selection: L1 or Least absolute shrinkage and selection operator (*Lasso*) Regression and L2 or *Ridge Regression*. The detailed explanation for these two regularisation method is described in Ref. [84]. In the analysis presented in this thesis, **XGBoost** was configured to use L2 or *Ridge Regression* to prevent the model from over-fitting.

Extreme Gradient Boosting

Extreme gradient boosting enhances the gradient boosting discussed in the previous section for computational speed and scale by using multiple CPU cores in parallel during the training. This boosting algorithm can handle large datasets and is thus an optimal choice for big data applications.

7.2 Application to the Higgs Portal Scalar Search

In the analysis presented in this thesis, showers produced from scalars from KDAR and KDIF are separated from background using the **XGBoost** algorithm. About 60% of the simulated signal and background samples discussed in section 5.8.2 are used in training the BDT with the remaining 40% used in assessing the sensitivity.

7.2.1 Feature Importance

As discussed in the previous section, **XGBoost** was configured to use L2 or *Ridge Regression* to prevent the model from over-fitting. In this method of regularisation, the model prevents over-fitting by assigning a small non-zero weight to the feature that is of least importance in separating the signal from background as opposed to L1

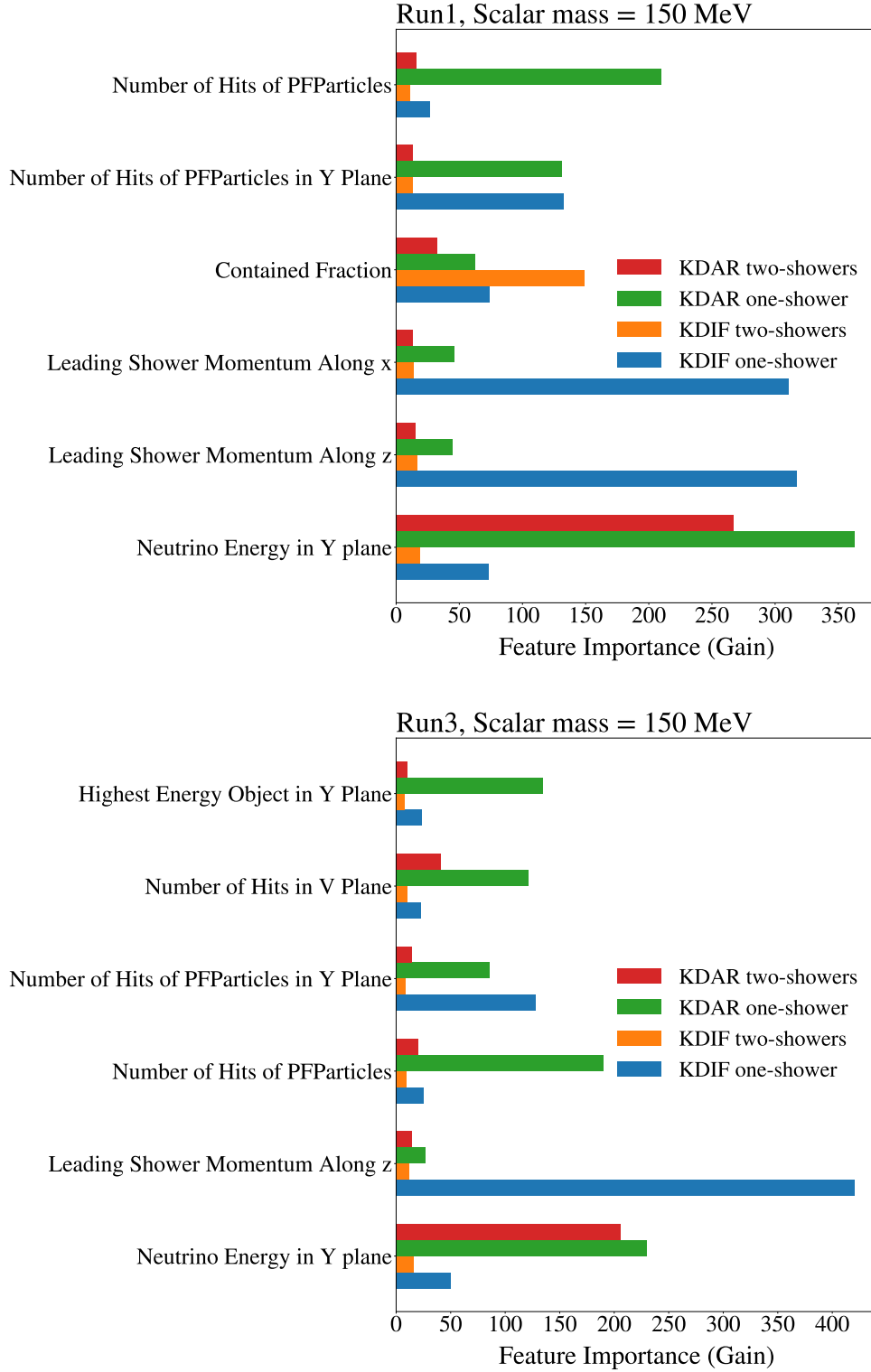


Figure 7.2: Importance of the BDT variables in terms of gain for two types of signal after the one-shower and two-shower pre-selections have been applied to the Run 1 (top) Run 3 (bottom) samples. Gains of the variables for scalars produced from KDAR with two-shower and one-shower pre-selections applied are shown in red and green respectively, whereas the gains for scalars produced from KDIF with two-shower and one-shower pre-selections applied are shown in orange and blue respectively. The gains shown are for BDTs trained using a scalar of mass 150 MeV.

or *Lasso* Regression, where an absolute zero weight is assigned to the least important features. In the analysis presented in this thesis, about 200 reconstructed variables are fed into `XGBoost` to select variables that are of high importance in discriminating the signal from background. From these ~ 200 variables, the 28 most important variables for eight different BDT models designed to identify the signal produced from KDAR and KDIF for the one-shower and two-shower selections in Run 1 and Run 3 are narrowed down. Figure 7.2 shows the feature importance in terms of gain (described in section 7.1.1) for six of these most important variables for an example scalar mass of 150 MeV for Run 1 and Run 3. The remaining 22 variables with small gains are excluded from these figures for clarity. For Run 3 KDIF two-showers, the variables with a significant discriminating power are in this list of 22 variables.

7.2.2 BDT Variables

This section compares the MC stacked prediction to data for variables with reasonable gain in discriminating the four different types of signal from background in Run 1 and Run 3. The calculation of the systematic uncertainties on the distributions of the BDT variables is described in chapter 8.

Energy deposited in the Y plane

Figures 7.3 and 7.4 show the distributions of the energy deposited in the collection plane (Y) after the one-shower and two-shower pre-selections respectively. The data and background distributions are overlaid on top of the area-normalised Higgs portal scalar distributions for a scalar of mass 150 MeV produced from KDAR (red dashed line) and KDIF (blue dashed line). The distributions on the left and right correspond to samples for Run 1 and Run 3 respectively. In both of these variables, the data agrees within the uncertainty of the MC predictions for Run 1 and Run 3. As discussed in section 6.1.1, the energy of the scalar produced from KDAR has a characteristic peak with its energy given by equation 6.7. Therefore the BDT has assigned the largest gain to this variable in identifying scalars produced from KDAR

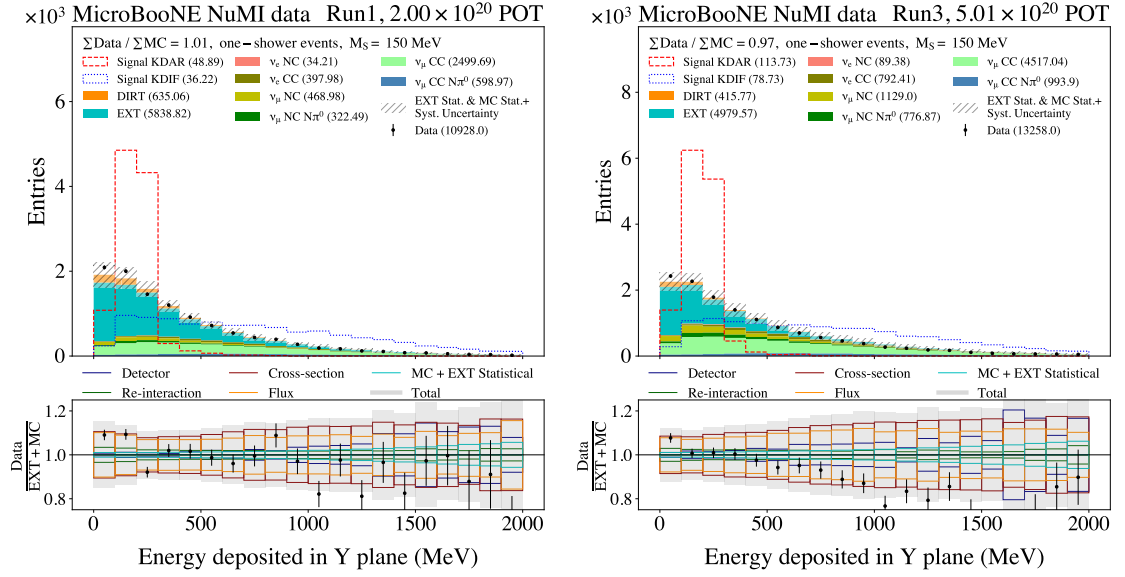


Figure 7.3: A comparison of the stacked prediction (EXT+MC) to the data (beam-on) after the one-shower pre-selections for energy deposited in the Y plane. The distribution is overlaid on top of the area-normalised Higgs portal scalar distribution for a scalar of mass 150 MeV produced from KDAR (red dashed line) and KDIF (blue dashed line). The distributions on the left and right correspond to samples for Run 1 and Run 3 respectively.

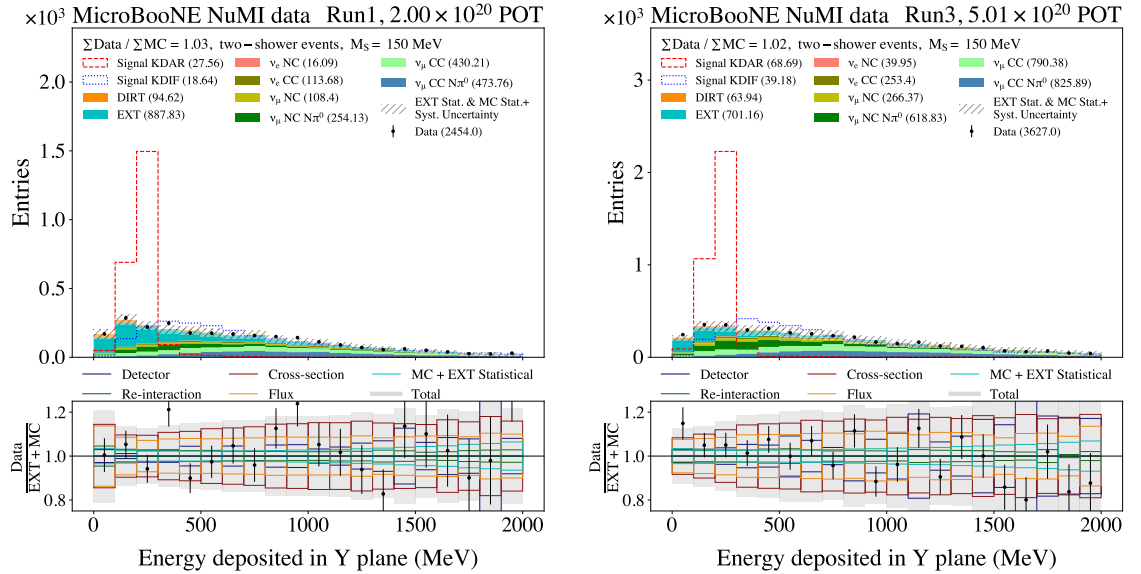


Figure 7.4: A comparison of the stacked prediction (EXT+MC) to the data (beam-on) after the two-shower pre-selections for neutrino or scalar energy measured using the charge collected on the Y plane. The distribution is overlaid on top of the area-normalised Higgs portal scalar distribution for a scalar of mass 150 MeV produced from KDAR (red dashed line) and KDIF (blue dashed line). The distributions on the left and right correspond to samples for Run 1 and Run 3 respectively.

(as shown in figure 7.2).

Leading Shower Momentum Along x and z

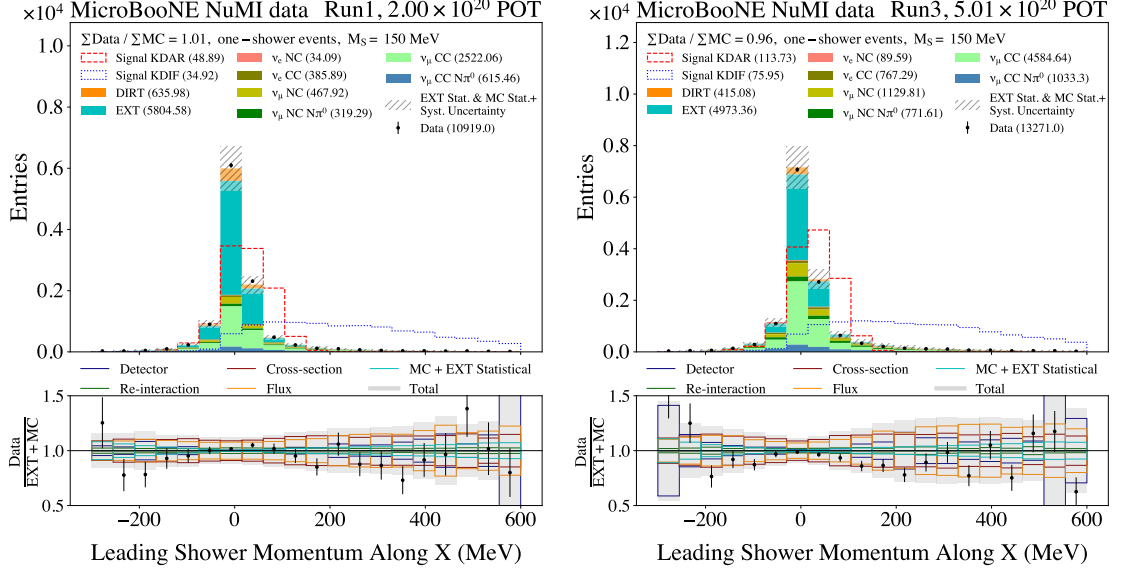


Figure 7.5: A comparison of the stacked prediction (EXT+MC) to the data (beam-on) after the one-shower pre-selection for the x component of the leading shower momentum vector. The distribution is overlaid on top of the area-normalised Higgs portal scalar distribution for a scalar of mass 150 MeV produced from KDAR (red dashed line) and KDIF (blue dashed line). The distributions on the left and right correspond to samples for Run 1 and Run 3 respectively.

Figures 7.5 and 7.6 show the x and z components of the leading shower momentum vector (described in section 6.2.3) after the one-shower pre-selection. The data and background distributions are overlaid on top of the area-normalised Higgs portal scalar distributions for a scalar of mass 150 MeV produced from KDAR (red dashed line) and KDIF (blue dashed line). The distributions on the left and right correspond to samples for Run 1 and Run 3 respectively. In both of these variables, the MC predictions agree well with the data for Run 1 and Run 3. These variables are crucial for showers from scalars produced from KDIF (as shown in figure 7.2) due to the decay kinematics of these showers described in section 6.1.3. Therefore the BDT has assigned a very large gain to this variable in identifying scalars produced from KDIF in one-shower pre-selected samples.

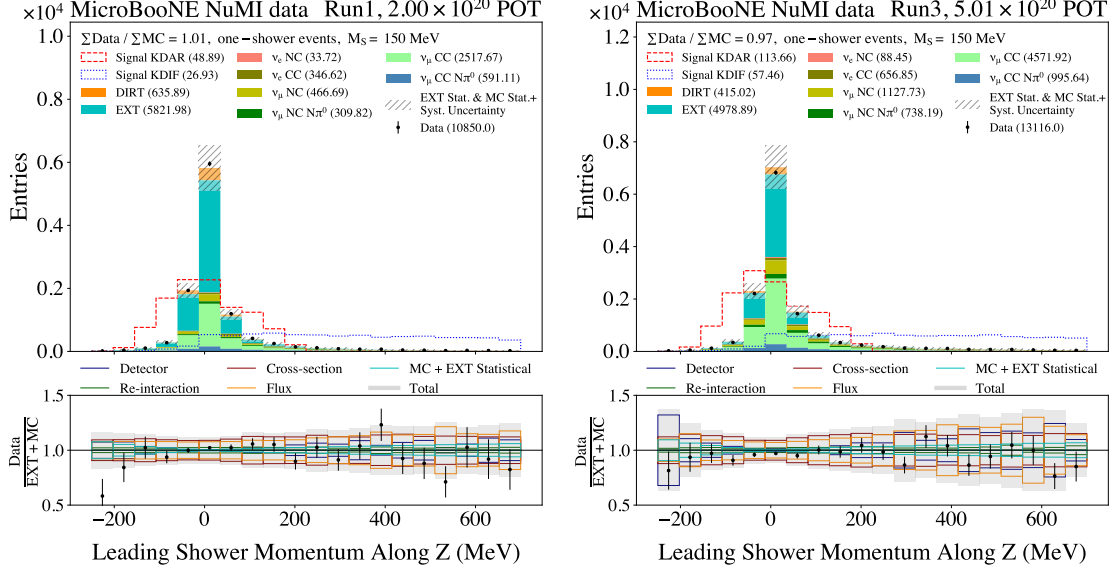


Figure 7.6: A comparison of the stacked prediction (EXT+MC) to the data (beam-on) after the one-shower pre-selection for the z component of the leading shower momentum vector. The distribution is overlaid on top of the area-normalised Higgs portal scalar distribution for a scalar of mass 150 MeV produced from KDAR (red dashed line) and KDIF (blue dashed line). The distributions on the left and right correspond to samples for Run 1 and Run 3 respectively.

Number of Hits of PFParticles

Figures 7.7 and 7.8 show the distributions of the numbers of hits in clusters associated to Pandora PFParticles (described in section 5.6) in the three wire-planes (described in section 2.3) and in just the collection plane (Y), respectively. The distributions on the left and right correspond to one-shower like samples for Run 1 and Run 3 respectively. In both of these variables, the MC predictions agree well with the data for both Run 1 and Run 3. These variables are crucial for identifying showers from scalars produced from KDAR (as shown in figure 7.2) due to the unique decay kinematics of the mono-energetic scalars produced from KDAR. Additionally, the number of hits measured using charge collected in the collection plane is important for identifying scalars produced from KDIF in one-shower samples.

Number of hits in V Plane

Figure 7.9 shows the distribution of the total number of hits measured in the second induction plane, V (described in section 2.3.1). The data and background distributions

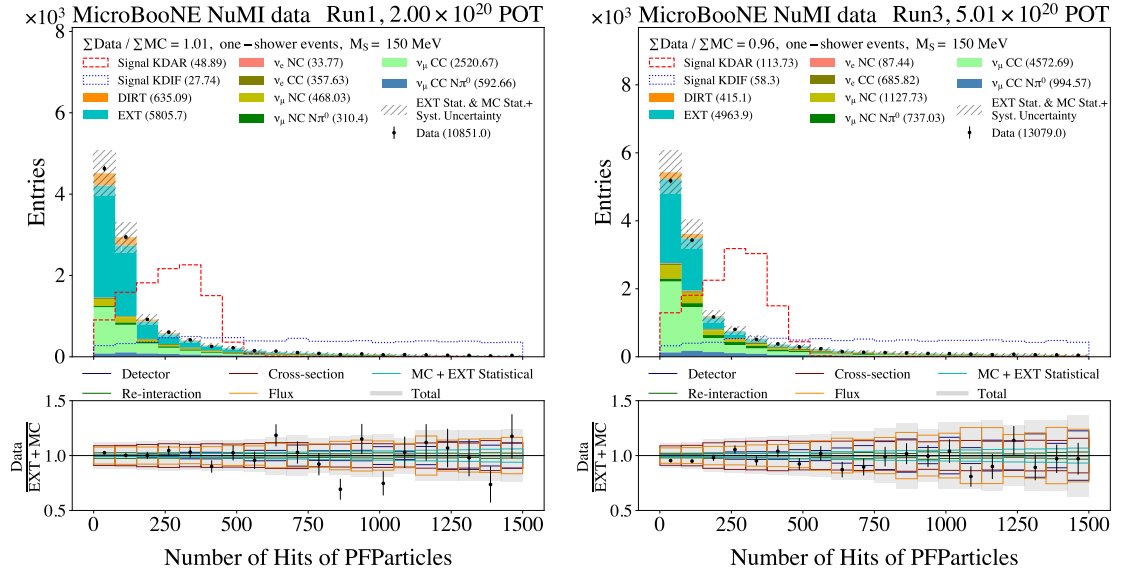


Figure 7.7: A comparison of the stacked prediction (EXT+MC) to the data (beam-on) after the one-shower pre-selection for the number of hits in clusters associated to Pandora PFParticles in the three wire-planes. The distribution is overlaid on top of the area-normalised Higgs portal scalar distribution for a scalar of mass 150 MeV produced from KDAR (red dashed line) and KDIF (blue dashed line). The distributions on the left and right correspond to samples for Run 1 and Run 3 respectively.

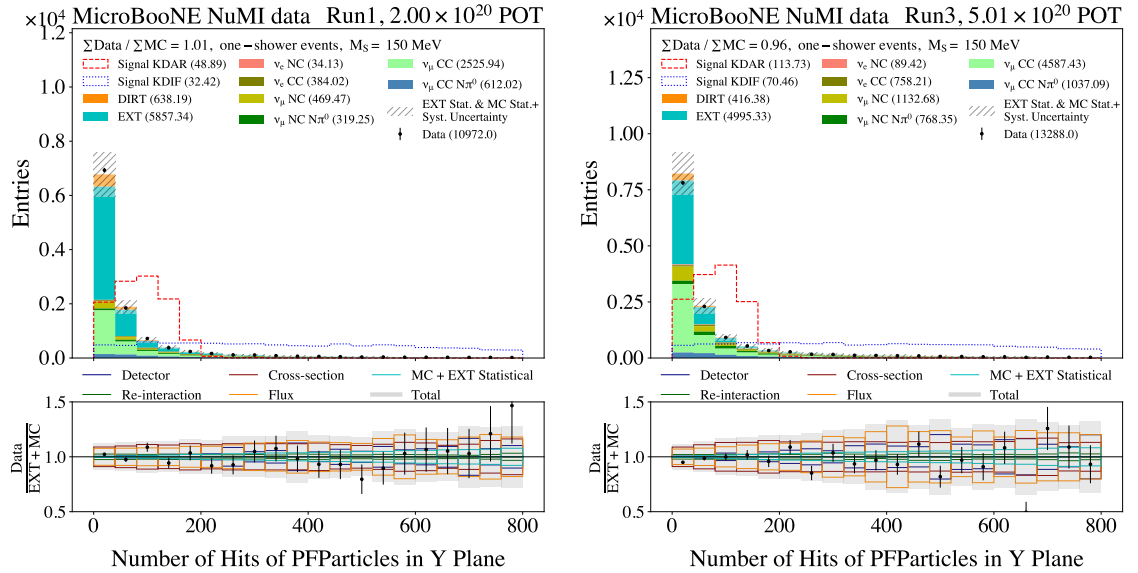


Figure 7.8: A comparison of the stacked prediction (EXT+MC) to the data (beam-on) after the one-shower pre-selection for the number of hits in clusters associated to Pandora PFParticles in the collection plane (Y). The distribution is overlaid on top of the area-normalised Higgs portal scalar distribution for a scalar of mass 150 MeV produced from KDAR (red dashed line) and KDIF (blue dashed line). The distributions on the left and right correspond to samples for Run 1 and Run 3 respectively.

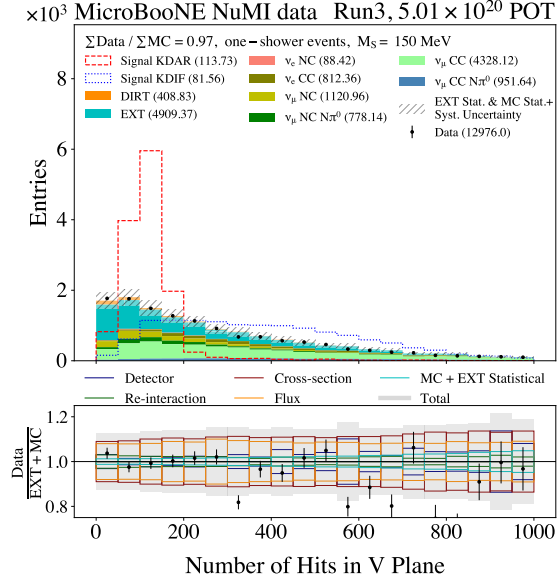


Figure 7.9: A comparison of the stacked prediction (EXT+MC) to the data (beam-on) after the one-shower pre-selection for the total number of hits measured in the second induction plane (V). The distribution is overlaid on top of the area-normalised Higgs portal scalar distribution for a scalar of mass 150 MeV produced from KDAR (red dashed line) and KDIF (blue dashed line). This distribution is produced using signal and background samples for Run 3.

are overlaid on top of the area-normalised Higgs portal scalar distributions for a scalar of mass 150 MeV produced from KDAR (red dashed line) and KDIF (blue dashed line). This distribution is produced using one-shower like signal and background samples for Run 3. As can be seen in the figure, the MC predictions agree well with the data. This variable is crucial for identifying showers from scalars produced from KDAR (as shown in figure 7.2) due to its unique decay kinematics (described in section 6.1.3). This variable is only important for Run 3 possibly due to the difference in hits induced by the objects in Run 1 and Run 3. In Run 3, the Pandora slices with cosmic muons are rejected due to additional CRT veto pre-selection and therefore Run 3 has cleaner samples with less cosmic-ray background than in Run 1.

Contained Fraction

In an event with a total number of hits due to shower(s) h_{shower} , due to track(s) h_{track} , and a total number of hits due to charge deposited outside the fiducial volume

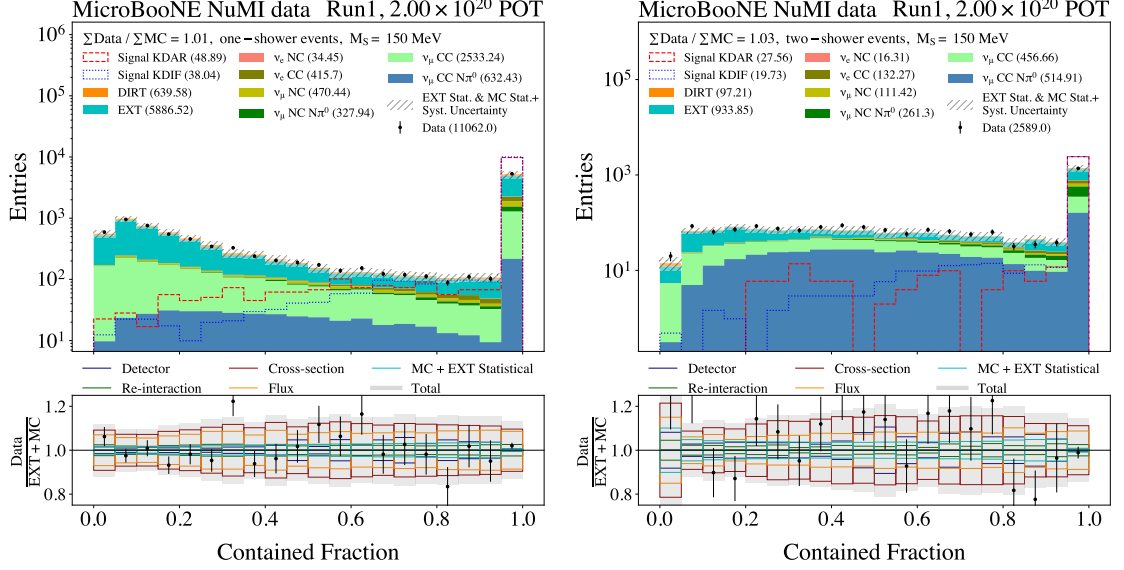


Figure 7.10: A comparison of the stacked prediction (EXT+MC) to the data (beam-on) after the one-shower (left) and two-shower (right) pre-selections for the fraction of hits of the **Pandora** PFParticle contained inside the fiducial volume of the MicroBooNE TPC. Both of these distributions show the log-scaled event rate for Run 1 samples. The distribution is overlaid on top of the area-normalised Higgs portal scalar distribution for a scalar of mass 150 MeV produced from KDAR (red dashed line) and KDIF (blue dashed line).

h_{outFV} , the contained fraction, cf, is defined as

$$\text{cf} = \frac{h_{\text{shower}} + h_{\text{track}}}{h_{\text{shower}} + h_{\text{track}} + h_{\text{outFV}}}. \quad (7.3)$$

Figure 7.10 shows the distribution of the fraction of hits of the PFParticle (described in section 5.6) contained inside the fiducial volume (described in section 6.2.3) of the MicroBooNE TPC. The distributions on the left and right correspond to one-shower and two-shower like Run 1 signal and background samples. The data and background distributions are overlaid on top of the area-normalised Higgs portal scalar distributions for a scalar of mass 150 MeV produced from KDAR (red dashed line) and KDIF (blue dashed line). From the figure, it can be seen that the contained fraction is close to 1 for both types of KDAR and KDIF signal in one-shower and two-shower selected samples, which also explains the gain of this variable in Run 1 on figure 7.2. The tracks of cosmic muons are generally not fully contained and therefore this variable is only important for Run 1 where cosmics are not rejected due to the absence of CRT veto pre-selection. It can be noticed in the figure that

the MC predictions agree well with the data for both these distributions.

Highest Energy Object in the Y Plane

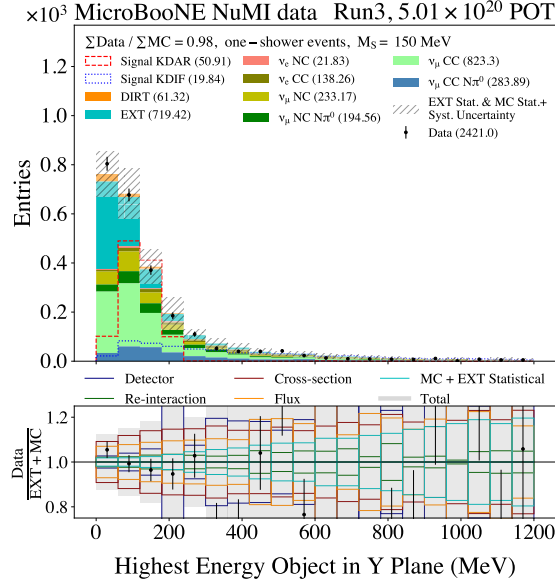


Figure 7.11: A comparison of the stacked prediction (EXT+MC) to the data (beam-on) after the one-shower pre-selection for the object (most likely a shower due to the pre-selection) that has the highest energy compared to all other objects in a slice. This energy is calculated using the charge collected in the collection plane (Y). The distribution is overlaid on top of the area-normalised Higgs portal scalar distribution for a scalar of mass 150 MeV produced from KDAR (red dashed line) and KDIF (blue dashed line). This distribution is produced using signal and background samples for Run 3.

Figure 7.11 shows the distribution of the energy of the object that has the highest energy of all the other objects in a slice. This energy is calculated using the charge collected on the collection plane (Y). Due to the pre-selection applied in the Run 3 signal and background samples used in this distribution, this object is most likely a shower. The data and background distributions are overlaid on top of the area-normalised Higgs portal scalar distributions for a scalar of mass 150 MeV produced from KDAR (red dashed line) and KDIF (blue dashed line). As it can be seen in the figure, the MC predictions agree well with data. This variable is crucial for identifying showers from scalars produced from KDAR in one-shower Run 3 samples as can be seen in figure 7.2.

7.3 BDT Performance

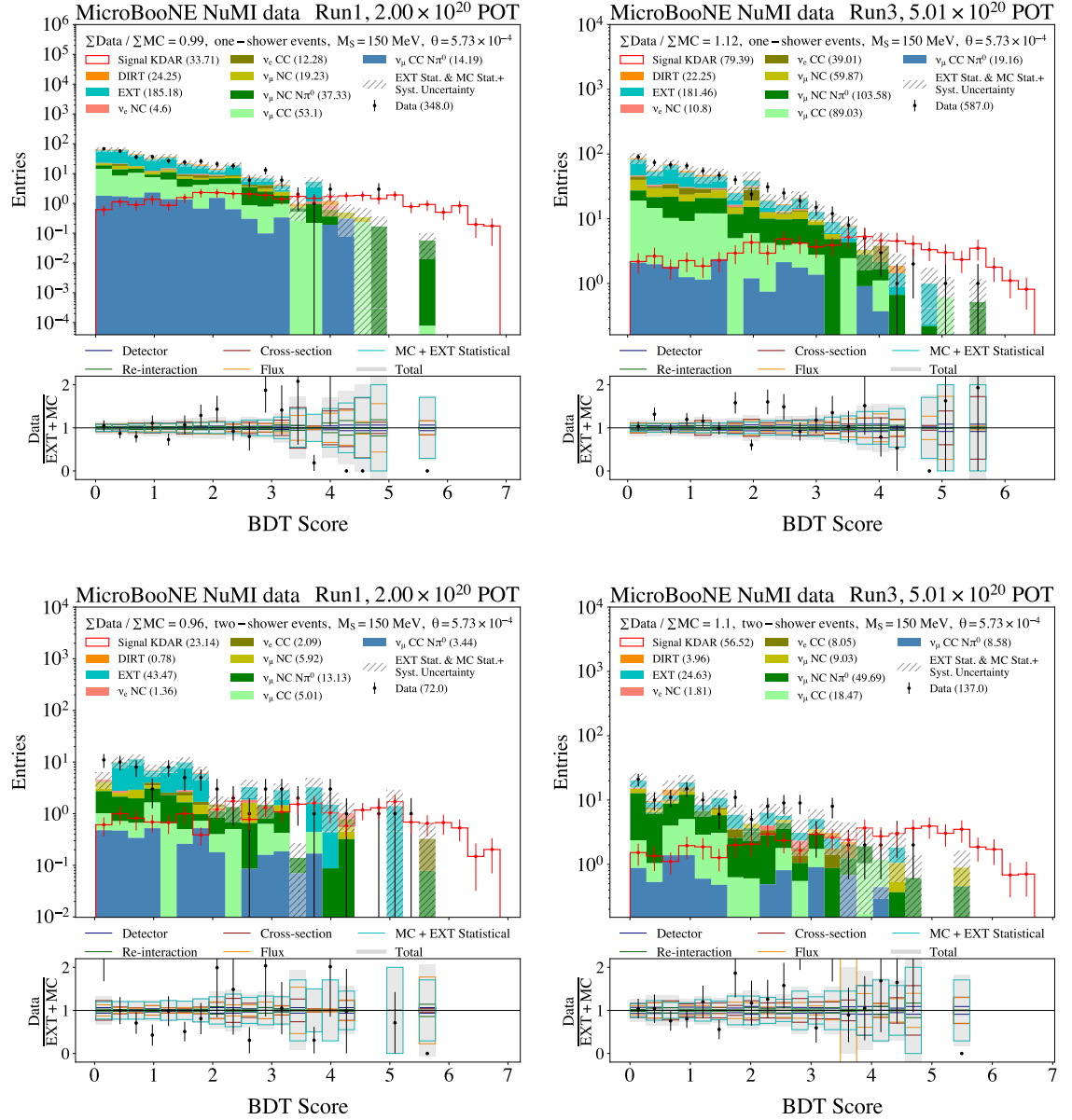


Figure 7.12: A comparison of the stacked prediction (EXT+MC) to the data (beam-on) after the one-shower (top panel) and two-shower (bottom panel) pre-selections as a function of the BDT score. The data and background distributions are overlaid on top of the Higgs portal scalar distribution normalised using the value of θ corresponding to the KOTO central-value (described in section 5.8.4) for a scalar of mass 150 MeV produced from KDAR. The distributions in the left and right hand panels correspond to samples for Run 1 and Run 3 respectively. The event rate corresponds to the number of events in the test samples of the respective BDTs.

Figures 7.12 and 7.13 show the distributions of the BDT scores for a scalar of mass 150 MeV produced from KDAR and KDIF (red) after the one-shower (top panel) and two-shower (bottom panel) pre-selection. The distributions on the left and right

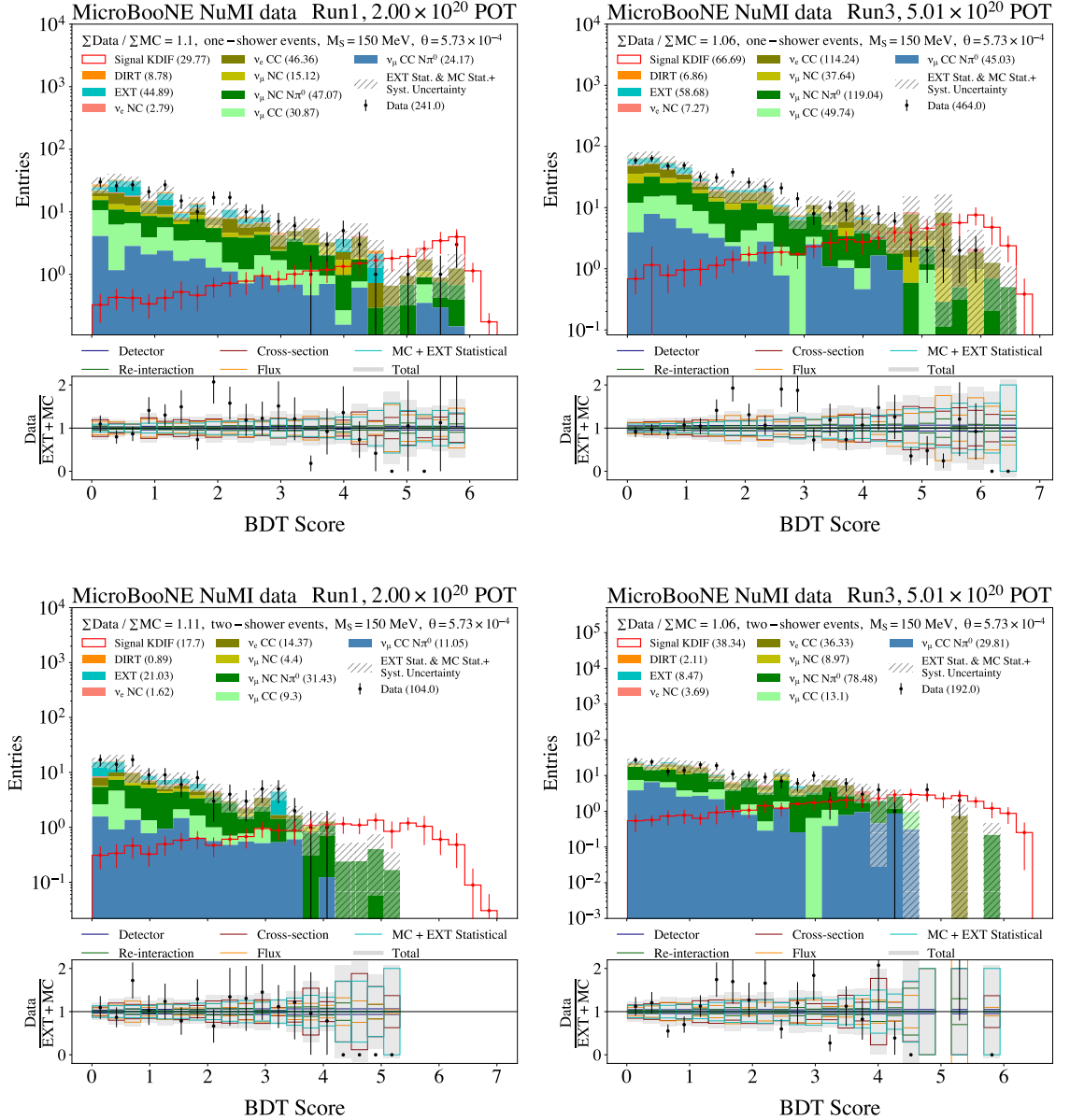


Figure 7.13: A comparison of the stacked prediction (EXT+MC) to the data (beam-on) after the one-shower (top panel) and two-shower (bottom panel) pre-selections as a function of the BDT score. The data and background distributions are overlaid on top of the Higgs portal scalar distribution normalised using the value of θ corresponding to the KOTO central-value (described in section 5.8.4) for a scalar of mass 150 MeV produced from KDIF. The distributions in the left and right hand panels correspond to samples for Run 1 and Run 3 respectively. The event rate corresponds to the number of events in the test samples of the respective BDTs.

hand panels correspond to samples for Run 1 and Run 3 respectively. The event rate corresponds to the number of events in the test samples of the respective BDTs. For signal, the event rate corresponds to the value of θ for the KOTO central-value described in section 5.8.4. The output of the BDT is a probability value x that quantifies how signal-like the event is, with 0 being the most background-like and 1 the most signal-like. In this analysis, we transform these probabilities from the unit interval to the real line using the *logit* transformation defined as:

$$\text{logit}(x) = \ln \frac{x}{1-x}, \quad (7.4)$$

where x is the BDT score in the range $(0, 1)$ and $\text{logit}(x = 0.5) = 0.0$.

A common figure of merit used to evaluate the performance of a BDT is the Receiver Operating Characteristic (ROC) curve, which shows the relationship between the true positive rate or signal efficiency (fraction of correctly identified signal events) and the true negative rate or background rejection (fraction of correctly identified background events) across different thresholds on the BDT score. A quantity of interest in this curve is integral Area Under the Curve (AUC), which quantifies the ability of the BDT in discriminating the two classes. It can be interpreted as the probability of a randomly chosen signal event getting a higher BDT score than a randomly chosen background event.

Figure 7.14 shows the ROC curves and their corresponding AUC values for the four different BDT models trained to discriminate scalars produced from KDAR and KDIF with one-shower and two-shower pre-selections for Run 1 samples. The curves for Run 3 are similar to Run 1 with the additional CRT Veto selection in Run 3 samples having no impact on the BDT performance. Therefore the ROC curves for Run 3 samples are not shown here. The diagonal red lines correspond to classifying events at random with probability of 0.5. The further the ROC curve is from the diagonal, the better the performance of the BDT, resulting in better efficiency (described in section 6.3) and purity².

The BDTs for the four different types of signal show excellent performance with

²Ratio of the number of signal events selected to the total number of events selected.

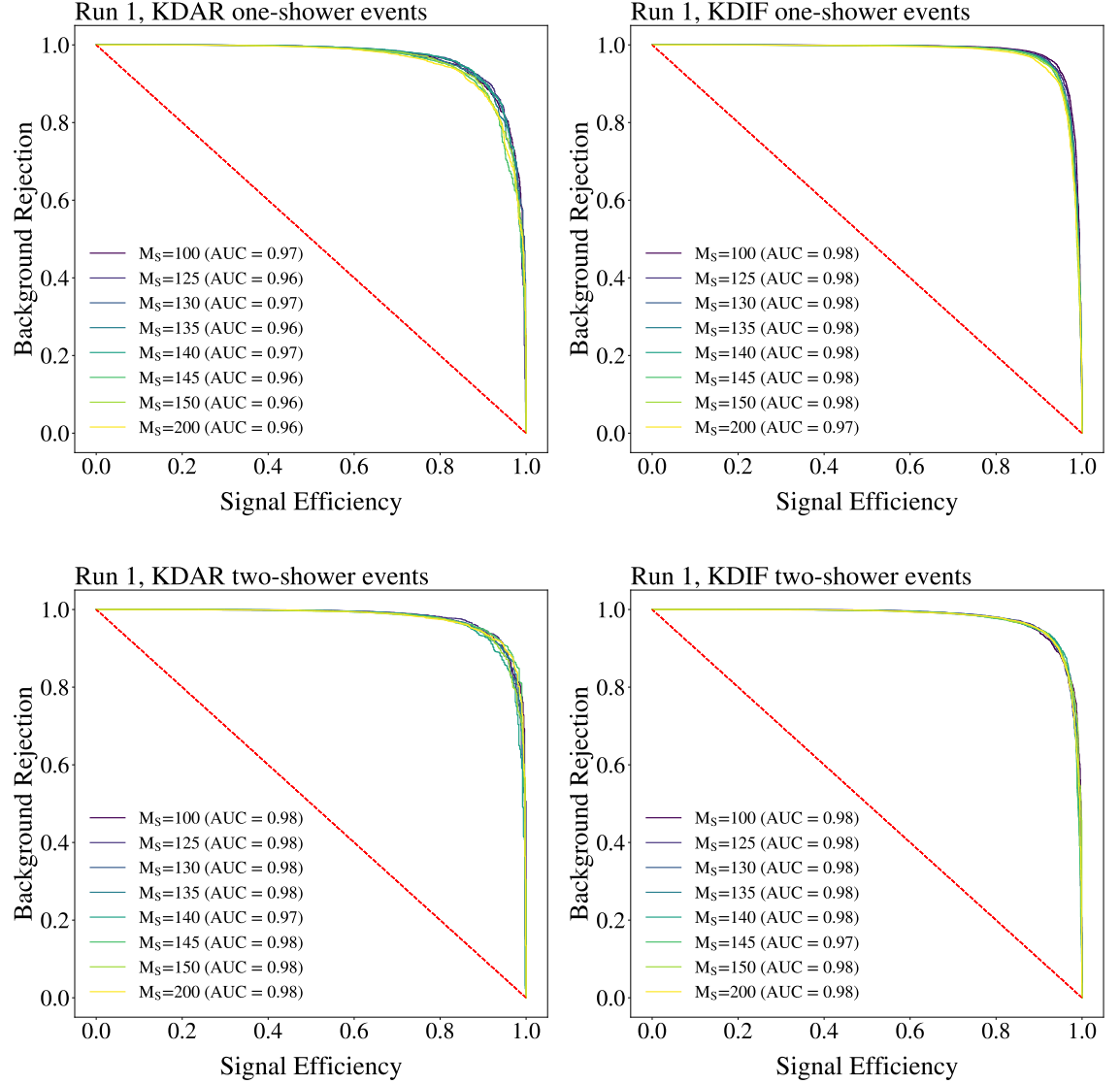


Figure 7.14: ROC curves of the BDTs trained for each scalar mass value (from purple to yellow). The performance of the BDTs across different mass values for each signal type are similar with an average AUC value of ~ 0.96 for scalars produced from KDAR in the one-shower pre-selected sample and ~ 0.98 for the others. The diagonal red line corresponds to classifying events at random with probability of 0.5.

average values of AUC around 0.98 across different masses of the scalars produced from KDIF with one-shower and two-shower pre-selections and KDAR with two-shower pre-selection. For scalars produced from KDAR with one-shower selection, the performance of the BDT is slightly poorer with an average value of AUC around 0.96 across different masses of the scalars.

Chapter 8

Systematic Uncertainties

In the analysis presented in this thesis, a wide range of sources of systematic uncertainty are taken into consideration. In background, the beam-off sample is real data recorded using an external unbiased trigger described in section 2.5.3 and therefore the systematic uncertainty is due to only the statistical uncertainty of this dataset. For the MC background, the following main sources of uncertainties are considered

- **Flux:** Uncertainties associated with modelling of the NuMI beamline and hadron production described in section 8.1.
- **Cross-section:** Uncertainties associated with modelling of neutrino interactions from the GENIE neutrino generator described in section 8.2
- **Re-interaction:** Uncertainties associated with modelling the re-interactions of daughter hadron particles in the liquid argon described in section 8.3.
- **Detector:** Uncertainties associated with modelling of the detector response described in section 8.4.
- **Out-of-Cryostat:** Uncertainties associated with modelling of the out-of-cryostat interactions described in section 8.5.
- **Statistical Uncertainty:** Uncertainties associated with the available statistics

of the MC samples used in the analysis described in section 8.6.

For signal, the uncertainties associated with modelling of neutrino interactions are not considered as the signal in our study is a scalar decay rather than an interaction. In addition, uncertainties associated with modelling the re-interactions of daughter hadron particles is also not considered as the decay product is an electron-positron pair with no hadron production. The chapter considers an example scalar mass of 150 MeV to study the systematic uncertainties. For the remaining seven masses of the scalar, a similar procedure to that described in this chapter is followed.

To estimate impact of the the systematic uncertainties on the MC background sample, the following two methods are used:

- **Event Re-weighting:** The models used in the MC simulation are described by sets of parameters and these parameters have uncertainties associated to them. To account for these uncertainties in the MC simulation, an event weight is calculated by re-sampling the parameters used in the simulation within their uncertainties. This weight is applied to the nominal simulation which uses nominal values of the parameters. The sample produced using the nominal simulation is called the Central Value (CV) sample and a sample produced after applying the weights resulting from a single re-sampling of all simulation parameters is known as a systematic “universe”. This method is well suited to assigning the impact of uncertainties in aspects of the simulation that happen at resource intensive stages of the MC generation procedure. In the analysis presented in this thesis, this method is used to evaluate the impact of flux, cross-section and re-interaction uncertainties.
- **Sample Re-simulation:** In this method, a new MC simulation is carried out to generate a sample with a single parameter varied by $\pm 1\sigma$ with respect to its CV. This is a comprehensive approach to estimating uncertainties and provides a realistic depiction of the uncertainty associated with a particular effect. This method is computationally more resource intensive as a new sample is generated for every parameter tweaked and thus in the analysis, it is only used to evaluate detector uncertainties.

One of the key advantages of using weights is that it allows multi-parameter re-sampling where more than one parameter is simultaneously sampled within its uncertainty to evaluate weights. For a given distribution, the uncertainty is described by constructing a covariance matrix correlating the total number of events in two bins with indices i and j . In the systematically shifted universes the number of events in these bins are N_i^k and N_j^k and in the CV universes the number of events are N_i^{cv} and N_j^{cv} . After sampling over a number of universes N_{uni} the entries in the ij th bin of the covariance matrix are

$$E_{ij} = \frac{1}{N_{\text{uni}}} \sum_k^{N_{\text{uni}}} (N_i^k - N_i^{\text{cv}}) (N_j^k - N_j^{\text{cv}}). \quad (8.1)$$

This technique of multi-parameter re-sampling is also known as multi-sim method and is capable of describing the variation in parameters on the final physics distributions accurately without being computationally too resource intensive. In the alternative case of the uni-sim method, a single parameter is re-sampled only once and therefore $N_{\text{uni}} = 1$. Assuming the uncertainty to be Gaussian for both the uni-sim and multi-sim methods, the diagonal elements of this covariance matrix equate to the variances on the individual bins and therefore the systematic uncertainty on each bin i , σ_i , is given by

$$\sigma_i = \sqrt{E_{ii}}. \quad (8.2)$$

When plotting systematic uncertainty bands on distributions in this thesis, it is these σ_i that are shown.

8.1 Flux Uncertainties

This section outlines the method used to calculate the uncertainties for the NuMI flux. The flux uncertainties can be broken down into the following two main categories: hadron production uncertainties described in section 8.1.1 and beamline geometry uncertainties described in section 8.1.2.

8.1.1 Hadron Production Uncertainties

Type	Name	Description
Thin target	$pC \rightarrow \pi X$	Constraints on pion production from pC collisions.
	$pC \rightarrow KX$	Constraints on kaon production from pC collisions.
	$pC \rightarrow \text{nucleon}X$	Constraints on proton/neutron production from pC interactions.
	$nC \rightarrow \pi X$	Constraints on pion production from nC collisions.
	nucleon-A	Nucleons interacting in a material that is not Carbon.
	Meson Incident	Mesons that interact on any material in the beamline.
Attenuation	Others	Interactions not covered by thin target data.
Absorption	Absorption	Corrections to the absorption cross sections.

Table 8.1: The constraints in various hadron production channels used in the PPFX flux prediction (from Ref. [27])

To calculate the flux uncertainties associated with hadron production, parameters used to constrain NuMI flux prediction in the PPFX package (described in section 3.2) are re-sampled simultaneously to calculate event weights. The list of hadron production channels constrained by the PPFX package is tabulated in table 8.1. In total 600 systematic “universes” corresponding to 600 different flux predictions were taken into account to calculate the flux uncertainty in signal and background (shown in section 8.1.5) for the analysis presented in this thesis.

Figure 8.1 shows the fractional uncertainty on the FHC (left) and RHC (right) neutrino energy spectra for the largest contributing neutrino flavour to the background of the study presented in this thesis, ν_μ . The largest contribution to the uncertainty is from mesons re-interacting with the material in the beamline, **Meson Incident**, and nucleons interacting with the materials that are not carbon, **Nucleon-A**, due to statistically limited data for these production modes. The uncertainties across different neutrino flavours and horn current configurations are of similar magnitude [27].

8.1.2 Beamline Geometry Uncertainties

To evaluate the systematic uncertainty on the NuMI beamline geometry, the **g4numi** simulation package described in section 3.2 is recursively run by changing the parameters tabulated in table 8.2 within their estimated $\pm 1\sigma$ uncertainty. The fractional

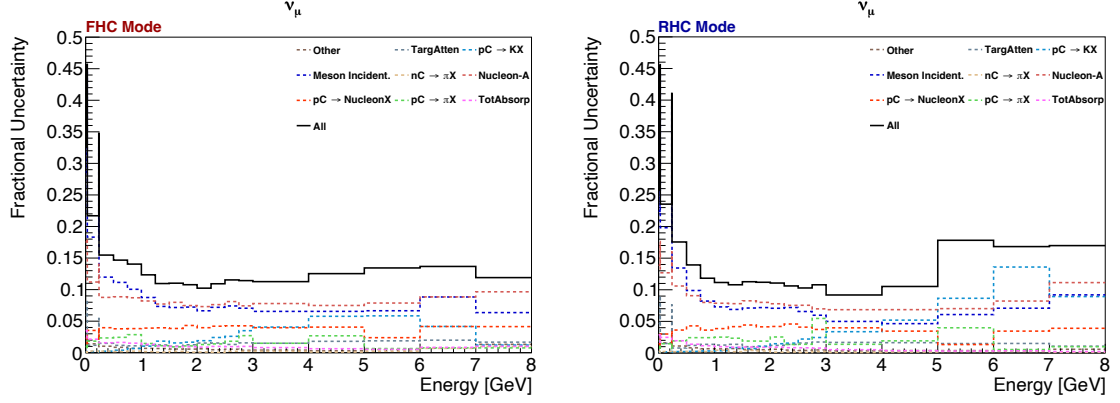


Figure 8.1: The fractional uncertainty on the FHC (Run 1) and RHC (Run 3) ν_μ energy spectra. Figure from Ref. [27].

Variation	CV	1σ Uncertainty
Horn Current	200 kA	± 2 kA
Horn 1 x Position	0 mm	± 3 mm
Horn 1 y Position	0 mm	± 3 mm
Beam Spot Size	1.3 mm	± 0.2 mm
Horn 2 x Position	0 mm	± 3 mm
Horn 2 y Position	0 mm	± 3 mm
Water on Horns	1 mm	± 1 mm
Beam x Position	0 mm	± 1 mm
Beam y Position	0 mm	± 1 mm
Target z Position	1433 mm	± 7 mm

Table 8.2: The $\pm 1\sigma$ beamline variations used in the simulation. There are in total 20 variations with two variations from the CV for each parameter. Particle attenuation and absorption for the materials used in the NuMI beamline are considered in hadron production uncertainties described in previous section (from Ref. [27]).

uncertainty on these variations are smaller compared to the hadron production uncertainties described in previous section with an exception for low energy decays at rest where uncertainties can change the flux prediction by up to $\sim 10\%$ [27].

8.1.3 Uncertainties due to Long-lived Kaons

As discussed in section 3.3.3, a significant fraction of the electron neutrino (anti-neutrino) flux is from K^+ (K^-) and K_L^0 decays for $E_\nu > 60$ MeV. The contributions of K_L^0 decays in the production of electron neutrinos and electron anti-neutrinos are 31% and 57% respectively. The modelling of these decays is therefore crucial for

the SM prediction of the background as well as the prediction of the BSM signal in searches such as the one presented in this thesis where a significant fraction of the Higgs portal scalars are produced from these K_L^0 decays.

The PPFx simulation package (described in section 3.3) constrains the uncertainties on kaons via the following two production modes:

- **Direct production:** Production of kaons from the collision of highly accelerated 120 GeV protons with a graphite target. This is the second hadron production channel where uncertainties are constrained using thin target data, $pC \rightarrow KX$ listed in table 8.1.
- **Production via hadron secondary interactions:** Production of kaons from nucleons interacting with the material that is not carbon and from mesons interacting with the material in the beamline. These are the final two channels where uncertainties are constrained using thin target data, **nucleon-A** and **Meson Incident** in table 8.1.

Due to statistical limitations in the available data, a fixed 40% uncertainty is assigned to kaons (as well as pions) produced via hadron secondary interactions [27]. For direct production of kaons, data for K_L^0 is also statistically limited. To estimate the uncertainty on K_L^0 , PPFx constrains the K_L^0 yield by constraining the K^+ and K^- yield with hadron production data followed by propagating the uncertainty using the Gatignon/Wachsmuth formula,

$$N(K_{L(S)}^0) = \frac{N(K^+) + 3N(K^-)}{4}, \quad (8.3)$$

where $N(K_{L(S)}^0)$, $N(K^+)$ and $N(K^-)$ are the numbers of $K_{L(S)}^0$, K^+ and K^- produced in the $p - p$ collision respectively. The formula assumes isospin symmetry in the quark-parton model where nuclear forces for the protons and neutrons are treated equally [85]. The parton model describing the production of kaons in $p - p$ collisions is also not yet validated. Furthermore, the formula assumes $p - p$ collisions but carbon atoms within the graphite target also contain neutrons, and $p - n$ collisions are not taken into account within PPFx. A more realistic approach demands a more

accurate description of the pC interaction and therefore in the analysis presented in this thesis, a 100% uncertainty on the events in signal and background produced from a K_L^0 parent meson is used.

8.1.4 Uncertainty due to KDAR from the NuMI Absorber

As discussed in section 3.4.1, the production rate for the KDAR ν_μ per POT at MicroBooNE calculated using the `g4numi` beamline simulation does not agree with any of the measurements performed by the MiniBooNE experiment. The rates derived by the MiniBooNE experiment using different models range from 0.06 to 0.12 with an accepted central value rate equivalent to 0.085 estimated using the `GEANT4` simulation model. Considering these upper and lower bounds around the central value, a fixed 30% uncertainty on the muon-neutrinos produced from KDAR at the NuMI absorber is used for the signal. The final background distributions used in setting the limit have a negligible number of these events and therefore no uncertainty on the background overlay sample is used.

8.1.5 Total Flux Uncertainty on Signal and NuMI Overlay Background

Figures 8.2 and 8.3 show the distributions of the shower energy and invariant mass, and the impact of the PPFX flux variations, for the one-shower and two-shower like events for signals produced from scalars from KDAR (left) and NuMI overlay background (right). The distributions in these figures are produced using Run 1 samples. For signal, the fractional flux uncertainties for both shower energy and invariant mass are in the range (26 – 28%). For background, the fractional flux uncertainty for shower energy is relatively larger (20 – 23%) compared to that for the invariant mass (11 – 15%). The flux uncertainties for the distributions of the shower energy and invariant mass for Run 3 signal (scalars produced from KDAR) and background samples are similar to Run 1 and are therefore not shown here.

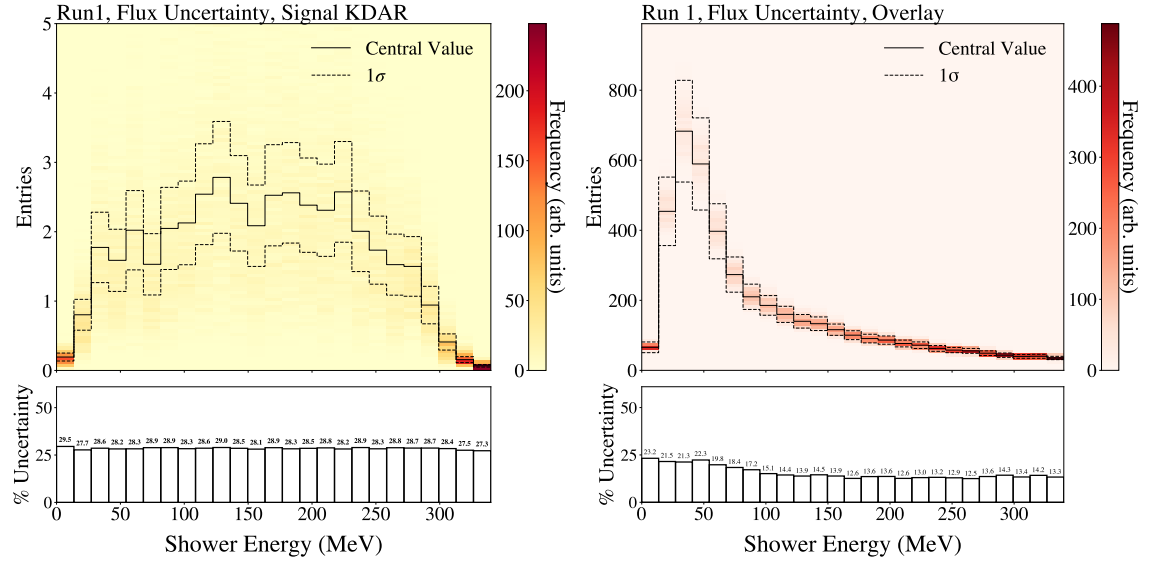


Figure 8.2: The distribution of the shower energy for the PPFX flux variations (universes) on the one-shower like events for signal produced from scalars from KDAR (left) and NuMI overlay background (right). The colour scale shows the density of the universes and the central black line represents the central value with the black dashed lines showing the standard deviation of all these universes. The distributions shown are produced using Run 1 samples.

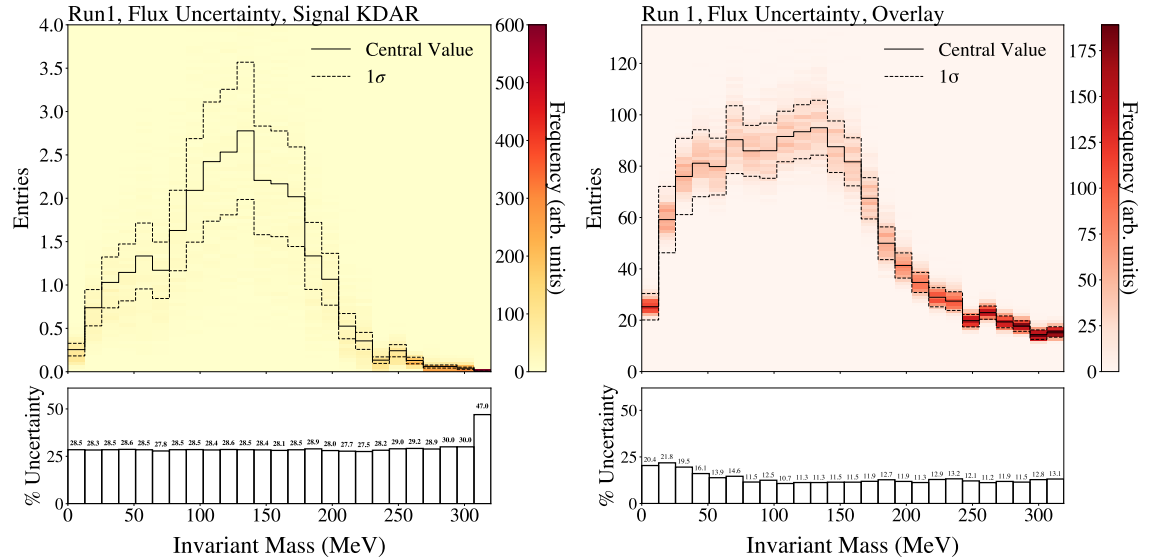


Figure 8.3: The distribution of the invariant mass for the PPFX flux variations (universes) on the two-shower like events for signal produced from scalars from KDAR (left) and NuMI overlay background (right). The colour scale shows the density of the universes and the central black line represents the central value with the black dashed lines showing the standard deviation of all these universes. The distributions shown are produced using Run 1 samples.

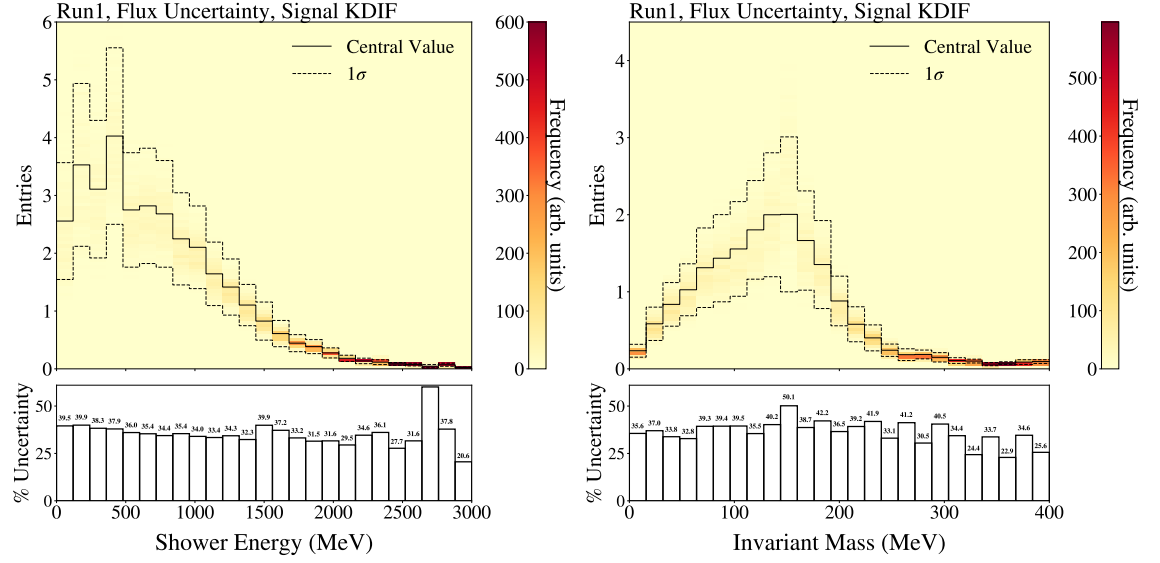


Figure 8.4: The distribution of the shower energy (left) and invariant mass (right) for the PPFX flux variations (universes) on the one-shower (left) and two-shower (right) like signal events produced from scalars from KDIF. The colour scale shows the density of the universes and the central black line represents the central value with the black dashed lines showing the standard deviation of all these universes. The distributions shown are produced using Run 1 samples.

Figure 8.4 shows the distributions of the shower energy and invariant mass, and the impact of the PPFX flux variations, for the one-shower (left) and two-shower (right) like signals produced from scalars from KDIF. The distributions in this figure are produced using Run 1 samples. The flux uncertainties for their corresponding background distributions will be same as the ones shown in figures 8.2 (right) and 8.3 (right) due to the similar selection applied for scalars from KDAR and KDIF (as discussed in section 6.1.4). The fractional flux uncertainties for both of these signals is in the range (35 – 40%), which is larger compared to that for the one-shower and two-shower signals produced from scalars from KDAR.

8.2 Cross-section Uncertainties

The neutrino interaction cross-section uncertainties are evaluated by varying the parameters used in the GENIE neutrino generator (described in section 5.1.2). In total 44 model parameters are considered in the new GENIE¹ model tune for MicroBooNE.

¹GENIE v3.0.6

The list of these parameters with their CV and $\pm 1\sigma$ uncertainties along with their usage in the **GENIE** neutrino generator is covered in Ref. [86].

These parameters are simultaneously re-sampled within $\pm 1\sigma$ uncertainties to calculate event weights. In total 600 systematic “universes” are taken into account to calculate the cross-section uncertainty on the background. The signal for the analysis presented in this thesis is a decay and therefore has no cross-section uncertainty associated to it.

8.2.1 Uncertainties due to CC and NC coherent pion production

The existing event re-weighting calculators for the coherent pion production (COH) cross-section in the MicroBooNE **GENIE** neutrino generator are incompatible with the Berger-Sehgal model [67] chosen for the CV. The incompatibility is due to technical limitations in the implementation of the Berger-Sehgal COH model in the MicroBooNE **GENIE** neutrino generator. To address this problem, new parameters called **NormCCCOH** and **NormCCCOH** are introduced to scale the CC COH and NC COH cross-sections by constant factors. Therefore in the analysis presented in this thesis, a 100% cross-section uncertainty in background is used for events associated to CC COH and NC COH to cover the best-fit result from MiniBooNE CC and NC π^0 data [86].

8.2.2 Total Cross-section Uncertainty on the NuMI Overlay Background

Figure 8.5 shows the distribution of the shower energy (left) and invariant mass (right), and the impact of the **GENIE** cross-section variations (universes), for the one-shower (left) and two-shower (right) like events in NuMI overlay background. The distributions in this figure are produced using Run 1 samples. The fractional

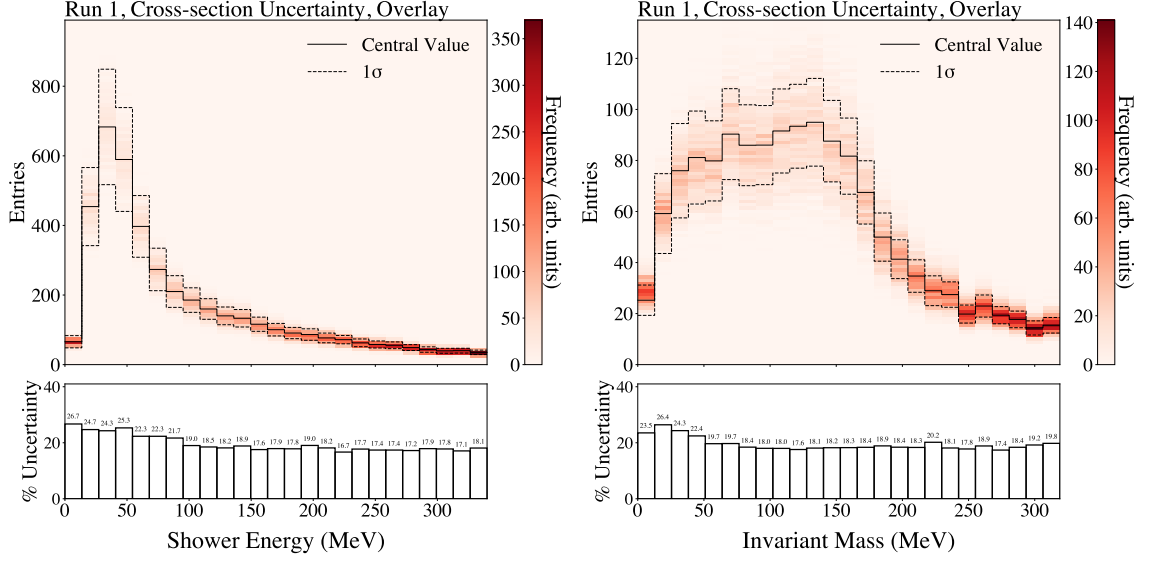


Figure 8.5: The distribution of the shower energy (left) and invariant mass (right) showing the impact of the **GENIE** cross-section variations (universes) on the one-shower (left) and two-shower (right) like events in NuMI overlay background. The colour scale shows the density of the universes and the central black line represents the central value with the black dashed lines showing the standard deviation of all these universes. The distributions shown are produced using Run 1 samples.

cross-section uncertainty for the distribution of shower energy is slightly larger (22 – 25%) compared to that for invariant mass (17 – 19%) in the regions of higher event rate. The cross-section uncertainties for the distributions of the shower energy and invariant mass for Run 3 background samples are similar to those for Run 1 and are therefore not shown here.

8.3 Re-interaction Uncertainties

Channel	Definition
Elastic	$\pi^\pm + N \rightarrow \pi^\pm + N$
Inelastic/Quasielastic	$\pi^\pm + N \rightarrow \pi'^\pm + N'$
Absorption	$\pi^\pm + N \rightarrow N'$
Single Charge Exchange	$\pi^\pm + N \rightarrow \pi^0 + N'$
Double Charge Exchange	$\pi^\pm + N \rightarrow \pi^\mp + N'$
Pion Production	$\pi^\pm + N \rightarrow n\pi + N'$

Table 8.3: The cross-section modes for pion-nucleus interactions considered in the **GEANT4** re-interaction uncertainty, where N represents an Argon nucleus.

The interaction of a neutrino with argon produces hadrons comprising protons and

pions that re-interact with the argon as they propagate in the LArTPC. These re-interactions affect the event reconstruction and therefore have direct impact on the identification of a neutrino in the background. The uncertainty due to hadrons re-interacting with the argon is evaluated using the open-source **GEANT4Reweight** framework [87]. For charged pions, the channels listed in table 8.3 are varied and for protons, the total elastic and inelastic cross-sections are varied. For signal, the Higgs portal scalar decays $S \rightarrow e^+e^-$ have no hadrons in the final state and therefore the re-interaction uncertainty will be zero.

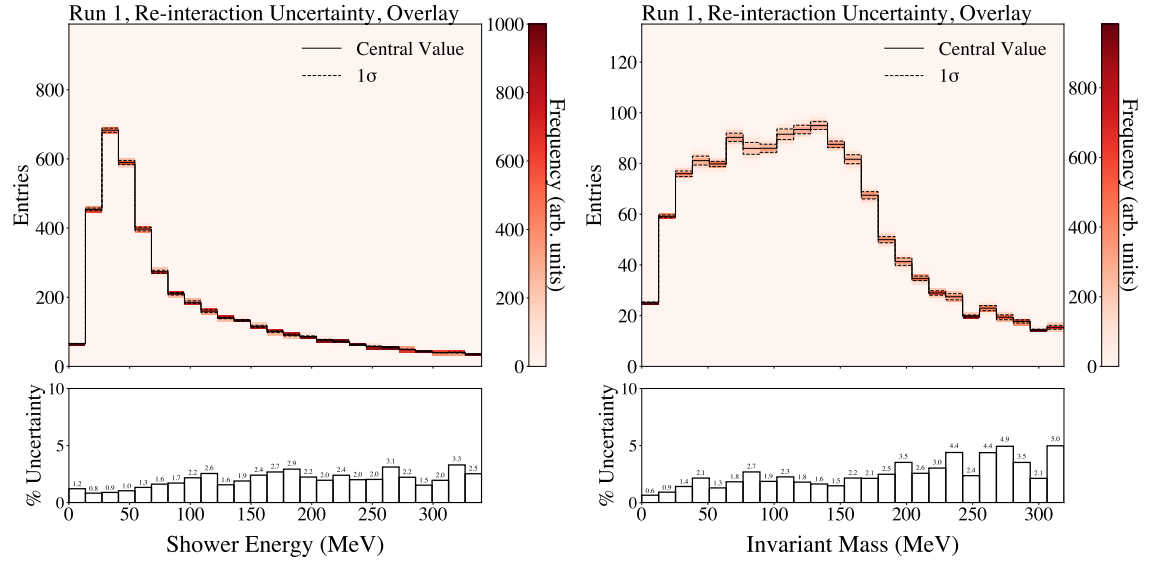


Figure 8.6: The distribution of the shower energy (left) and invariant mass (right) showing the impact of the **GEANT4** hadron-argon interaction variations (universes) on the one-shower (left) and two-shower (right) like events in NuMI overlay background. The colour scale shows the density of the universes and the central black line represents the central value with the black dashed lines showing the standard deviation of all these universes. The distributions shown are produced using Run 1 samples.

In total 1000 systematic “universes” are taken into account to calculate the re-interaction uncertainty in the NuMI overlay background. Figure 8.6 shows the distribution of the shower energy (left) and invariant mass (right) and shows the impact of the **GEANT4** hadron-argon interaction variations (universes), for the one-shower (left) and two-shower (right) like events in NuMI overlay background. The distributions in this figure are produced using Run 1 samples. The fractional hadron re-interaction uncertainties for both of these distributions are very small (1 – 3%) in the bins with larger event rates due to the pre-selections that preferentially select showers induced by electrons, positrons and photons. The re-interaction uncertainties

for the distributions of the shower energy and invariant mass for Run 3 background samples are similar to those for Run 1 and are therefore not shown here.

8.4 Detector Uncertainties

The systematic uncertainties associated with modelling of the detector response are evaluated using the sample re-simulation method, whereby a new MC sample is generated by changing a specific parameter in the nominal simulation. The list of nine variations accounting for several effects used in the analysis presented in this thesis is shown in table 8.4. This section provides a brief explanation of these variations.

Parameter	Summary
Wire Modifications	
Wire Mod X	Modifications to waveforms as a function of x
Wire Mod YZ	Modifications to waveforms in the $y - z$ plane
Wire Mod θ_{XZ}	Modifications to waveforms as a function of θ_{XZ}
Wire Mod θ_{YZ}	Modifications to waveforms as a function of θ_{YZ}
Scintillation Light	
Light Yield Down	A 25% reduction in the light yield
Light Rayleigh Length	Modification of the Rayleigh Scattering length to 90 cm (from 60 cm)
Light Yield Attenuation	A modification of the light attenuation length to 10m
Other re-simulations	
SCE	Implement an alternate data-driven SCE map to cover the possible electric field variations.
Recombination	Reduce the value of parameter β' in the Modified Box model by 13%

Table 8.4: A list of detector variations accounting for several effects considered in the analysis.

8.4.1 Wire Modifications

The modifications to simulation waveforms based on a parameterization of observed differences between simulation and the actual detector response are used to quantify the systematic uncertainties. A detector response comprises measurements of two hit properties: amplitude and width of the ionization charge signals from the TPC. To determine the detector response, a sample of Anode/Cathode Piercing Tracks

(ACPT) of cosmic muons in beam-off data is compared with the simulation to derive a ratio. A spline fit² to the measured ratio is performed to obtain a smooth function that describes the discrepancy between the data and simulation. This spline fit provides the simulation modification factor which can be used to modify the wire waveforms [88]. This section outlines various modifications to simulation waveforms implemented in the analysis including modifications to waveforms along the drift direction x , in the $y-z$ plane, and as functions of drift (θ_{XZ}) and wire (θ_{YZ}) direction.

Wire Mod X

Modifications to waveforms as a function of x encompass variations in charge response as a function of the drift direction x , where drift-dependent effects such as electron diffusion (described in section 4.4.1) and attenuation (described in section 4.4.6) are embodied.

Wire Mod YZ

Modifications to waveforms in the $y-z$ plane encompasses variation in detector response due to shorted or cross-connected TPC channels that results in distortion in electric field between the wire planes and therefore non-uniformities in the charge response in the $y-z$ plane.

Wire Mod θ_{XZ} and θ_{YZ}

Modifications to waveforms as functions of θ_{XZ} and θ_{YZ} encompasses variation in detector response due to orientation of the particle trajectory defined by the angular variables θ_{XZ} and θ_{YZ} to account for the effects related to long-range induced charge signals on the wires and signal processing [88].

²Spline fitting is a form of interpolation where instead of fitting a single high-degree polynomial to a set of values, a low-degree polynomial is fit to a subset of values.

8.4.2 Scintillation Light

The MicroBooNE scintillation light model has known limitations. In addition, MicroBooNE's scintillation light show significant time-dependent variations. The true sources for these mis-modellings and variations are unknown. Therefore it is crucial to account for the systematic variations due to light yield decline and light mis-modelling. This section outlines three variations for scintillation light considered in the analysis.

Light Yield Down

The Light Yield Down variation accounts for the fact that a 25% lower light yield was measured, compared to the simulations, for events producing light at the anode and at the cathode. Therefore an alternate sample with 25% reduction in the light yield is generated to account for this effect.

Rayleigh Scattering Length

The Rayleigh scattering length used by the current MicroBooNE simulation is 60 cm. However, several measurements are showing significant tension with this value, with measurements ranging from 60 cm to 90 cm [89, 90, 91, 92]. Therefore an alternate sample is generated using 90 cm to account for the range in measured values.

Light Yield Attenuation

A significant decline in light yield between the Run 1 and Run 3 data-taking periods of MicroBooNE was observed. This decline was seen to be dependent on the position of the interaction along the drift direction indicating attenuation caused by quenching and absorption. The effect of quenching and absorption on light yield attenuation is described in section 4.4.6. The detector uncertainties associated with light yield attenuation are only relevant for Run 3 as this data was recorded after the light yield

had declined.

8.4.3 Space Charge Effect and Recombination

The space charge effect and recombination are described in section 4.4.2 and section 4.4.3 respectively. The SCE and the recombination are interrelated as SCE can affect the amount of electron-ion recombination and thus the amount of charge measured by the wires planes. To assess the uncertainty from the space charge effect, an alternate data-driven electric-field map is implemented to account for possible electric field distortions. For recombination, the value of the parameter β' in equation 4.13 for the Modified Box model is reduced by 13%. This reduction is motivated by studies using MicroBooNE data [79].

8.4.4 Detector Variation Sample Statistics

To evaluate detector uncertainties, a sample re-simulation method was employed whereby samples for each of the detector variations (described in previous sections) along with the central-value are produced and a standard deviation across these detector variation samples with respect to the central-value sample is calculated. The samples used contain exactly the same sets of simulated events, to ensure there is no statistical variation between the different universes.

Due to the re-simulation method being computationally resource intensive, the re-simulated sample sizes are $\mathcal{O}(10)$ smaller compared to the neutrino overlay sample with average POT for all detector variation samples equivalent to 3.67×10^{20} for Run 1 and 3.22×10^{20} for Run 3. This POT is reduced slightly further due to sample production inefficiencies, whereby a variation sample might not have events existing in other variations due to error(s) in sample production. The impact of this statistical limitation increases significantly for very large values of the BDT scores where the vast majority of the background events are rejected. Therefore in the analysis presented in this thesis, detector uncertainties are calculated using a

single-binned histogram (after the BDT selection).

For signal, the detector variation samples for the mass points 125, 130, 135, 140 and 145 MeV were not produced and in the analysis, variation samples for the mass point 150 MeV are used to evaluate the detector uncertainties.

8.4.5 Total Detector Uncertainty on the Signal and NuMI Overlay Background

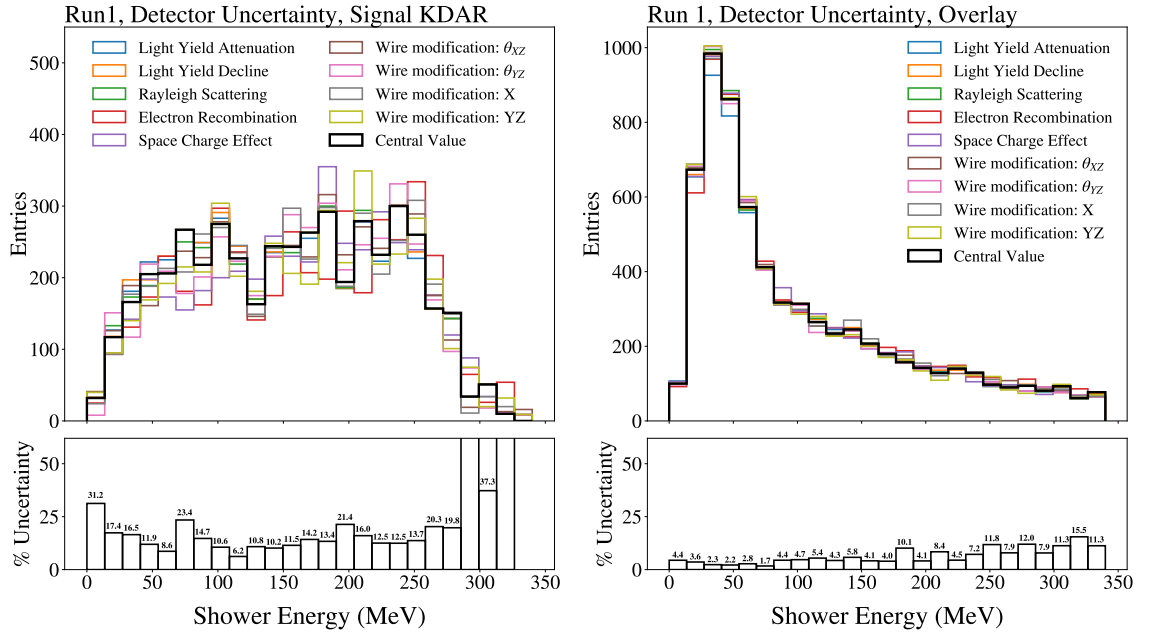


Figure 8.7: The distribution of the shower energy showing the impact of the detector variations on the one-shower like events for signal produced from scalars from KDAR (left) and NuMI overlay background (right). The black line represents the central value with different coloured lines showing the detector variations. The distributions shown are produced using Run 1 samples.

Figures 8.7 and 8.8 show the distributions of the shower energy and invariant mass, and the impact of the detector variations, for the one-shower and two-shower like events for signal produced from scalars from KDAR (left) and NuMI overlay background (right). The distributions in these figures are produced using Run 1 samples. For signal, the fractional detector uncertainties for both shower energy and invariant mass are in the range (10 – 20%) for bins with larger statistics. For background, the fractional detector uncertainty for shower energy is slightly smaller

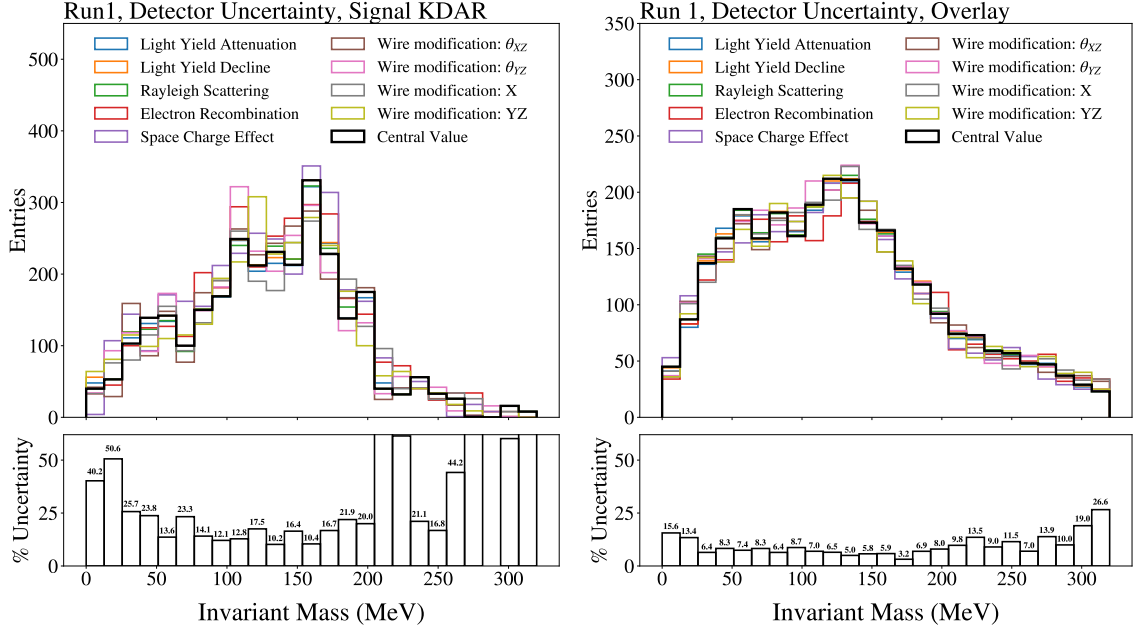


Figure 8.8: The distribution of the invariant mass showing the impact of the detector variations on the two-shower like events for signal produced from scalars from KDAR (left) and NuMI overlay background (right). The black line represents the central value with different coloured lines showing the detector variations. The distributions shown are produced using Run 1 samples.

(2 – 5%) compared to that for invariant mass (2 – 7%) in bins with larger statistics. The detector uncertainties for the distributions of the shower energy and invariant mass for Run 3 signal (scalars produced from KDAR) and background samples are similar to Run 1 samples and are therefore not shown here.

Figure 8.9 shows the distributions of the shower energy and invariant mass, and the impact of the detector variations, for the one-shower (left) and two-shower (right) like signals produced from scalars from KDIF. The distributions in this figure are produced using Run 1 samples. The fractional detector uncertainties for their corresponding background distributions will be same as the ones shown in figures 8.7 (right) and 8.8 (right) due to the similar selection applied for scalars from KDAR and KDIF (as discussed in section 6.1.4). The fractional detector uncertainties for shower energy (3 – 7%) are slightly smaller than those for invariant mass (5 – 10%) in bins of the distributions with relatively larger statistics. The detector uncertainties for the distributions of the shower energy and invariant mass for the Run 3 signal sample are similar to those for the Run 1 signal (scalar produced from KDIF) sample and are therefore not shown here.

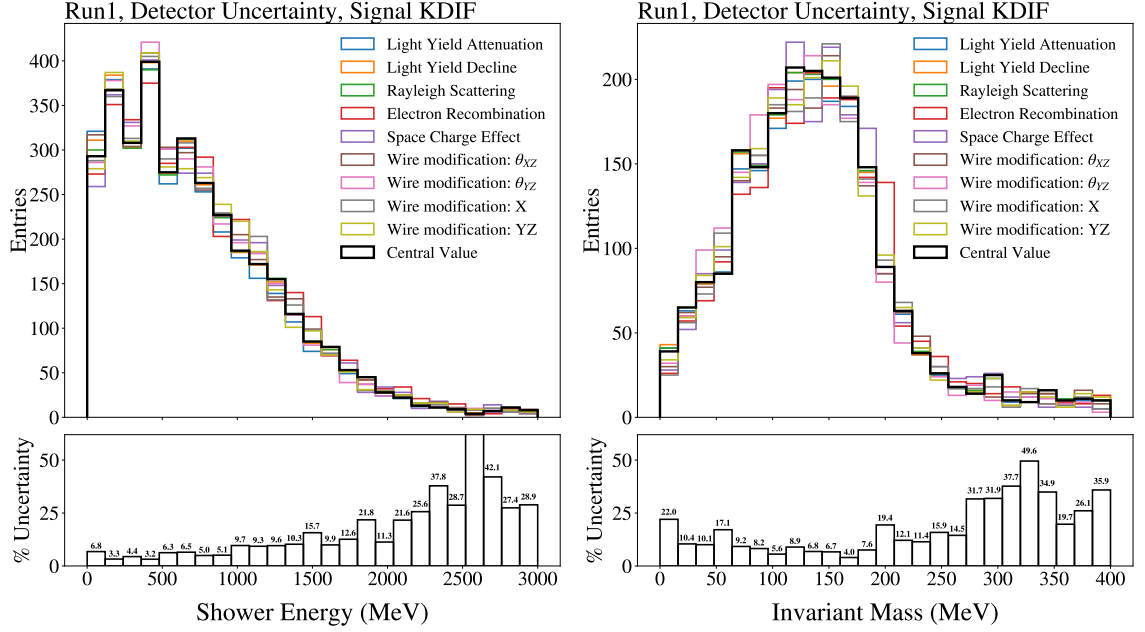


Figure 8.9: The distribution of the shower energy (left) and invariant mass (right) showing the impact of the detector variations on the one-shower (left) and two-shower (right) like signal events produced from scalars from KDIF. The black line represents the central value with different coloured lines showing the detector variations. The distributions shown are produced using Run 1 samples.

8.5 Out-of-Cryostat Uncertainties

The simulation of out-of-cryostat interactions relies on modelling the geometry of the building surrounding MicroBooNE as well as modelling the interactions of the neutrinos in materials of different density around the building. Due to the presence of this large set of unknowns, a 100% uncertainty on the out-of-cryostat interactions is used in the analysis presented in this thesis.

8.6 Statistical Uncertainties on the MC and Data Samples

The MC neutrino background sample is weighted by the **GENIE** and **PPFX** weights on an event by event basis. The statistical uncertainty on the MC neutrino background sample, $\sigma_{\text{MC Stat.}}$, on each bin, is given by

$$\sigma_{\text{MC Stat.}} = \sqrt{\sum_i w_i^2}, \quad (8.4)$$

where w_i is the weight of event i .

For beam-off data recorded using the MicroBooNE external and unbiased trigger discussed in section 2.5.3, and for beam-on data, the uncertainties are assumed to be Gaussian with

$$\sigma_{\text{Beam-off Stat.}}, \sigma_{\text{Beam-on Stat.}} = \sqrt{N}, \quad (8.5)$$

where N is the number of events in a bin.

8.7 Summary of Background and Signal Uncertainties

This section describes the background systematic and statistical uncertainties shown in the lower panel of the shower energy distribution in figure 6.13 and invariant mass distribution in figure 6.14 for Run 1 (left panels in these figures). The fractional uncertainty in this section is therefore calculated by accounting for the uncertainty in the total background prediction, which comprises the statistical uncertainty in the beam-off (EXT) sample (described in section 8.6), statistical and systematic uncertainty in the MC neutrino overlay sample, and a 100% uncertainty in the out-of-cryostat MC neutrino overlay sample (described in section 8.5). The systematic and statistical uncertainty on the signal is not shown in any of the stacked distributions in this thesis for clarity. However, this section summarises these uncertainties for four different types of signals for Run 1. The colour scheme for each type of uncertainty is similar to the one used in every stacked distributions shown in this thesis.

Figure 8.10 shows the percentage fractional uncertainty broken down by the source of uncertainty for background. The left hand panel shows the fractional uncertainty for the one-shower pre-selection where the shower energy distribution is

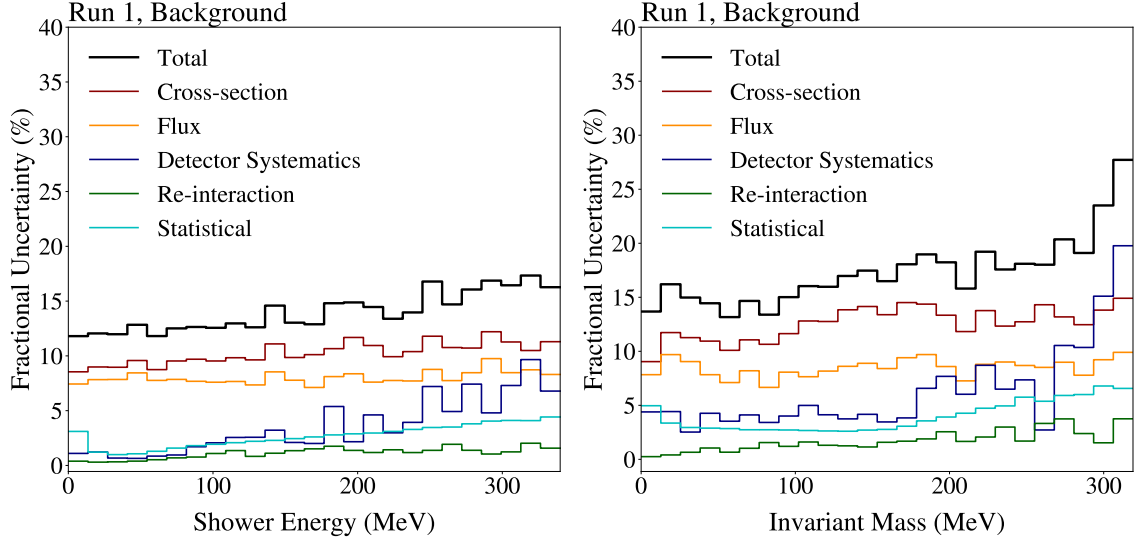


Figure 8.10: The percentage fractional uncertainty broken down by the source of uncertainty for background. The distributions on left and right correspond to samples for one-shower and two-shower pre-selection.

used to set the limit, and the right hand panel shows the fractional uncertainty for the two-shower pre-selection where the invariant mass distribution is used to set the limit. The dominant source of uncertainty in these distributions is due to cross-sections, contributing $\sim 10\%$ in the one-shower sample and $\sim 13\%$ in the two-shower sample. The second most dominant source of uncertainty is from the flux with a contribution of $\sim 8\%$ for both the one-shower and two-shower pre-selected samples. In the case of shower energy, the contributions of the detector and statistical uncertainties are smaller ($\sim 2\%$) compared to those for invariant mass ($\sim 5\%$) for bins with larger statistics. In both shower energy and invariant mass, the hadron re-interaction uncertainty is the least dominant source of uncertainty with a contribution of around or below $\sim 1\%$. These uncertainties are of similar order for Run 3 samples and are therefore not shown here.

Figures 8.11 and 8.12 show the percentage fractional uncertainty broken down by the source of uncertainty for the signal of scalars produced from KDAR and KDIF respectively. The left hand panels correspond to samples with a one-shower pre-selection where the shower energy distribution is used to set the limit and the right hand panels correspond to samples with a two-shower pre-selection where the invariant mass distribution is used to set the limit. The dominant source of uncertainty for each signal type is due to the flux, contributing $\sim 30\%$ for signal

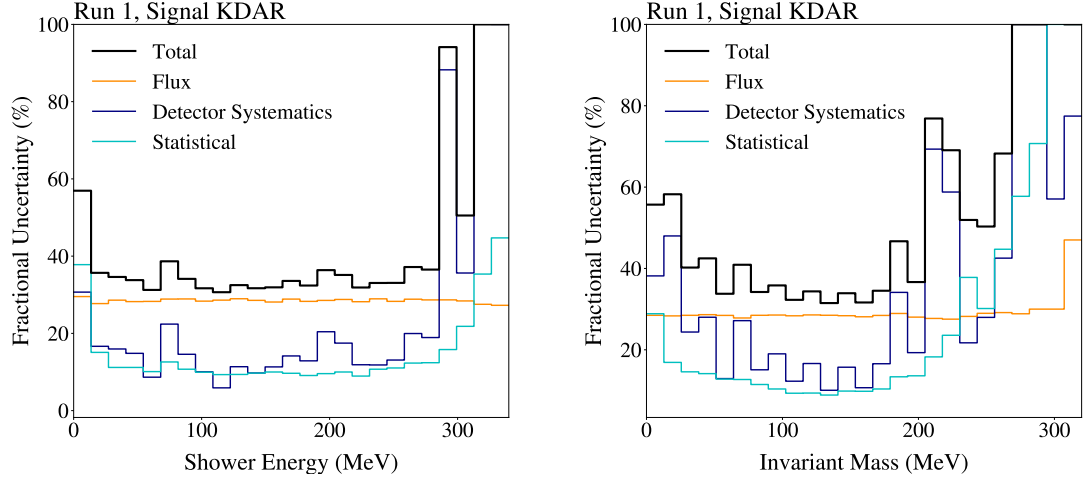


Figure 8.11: The percentage fractional uncertainty broken down by the source of uncertainty for the signal of scalars produced from KDAR. The distributions on the left and right correspond to samples for one-shower and two-shower pre-selections.

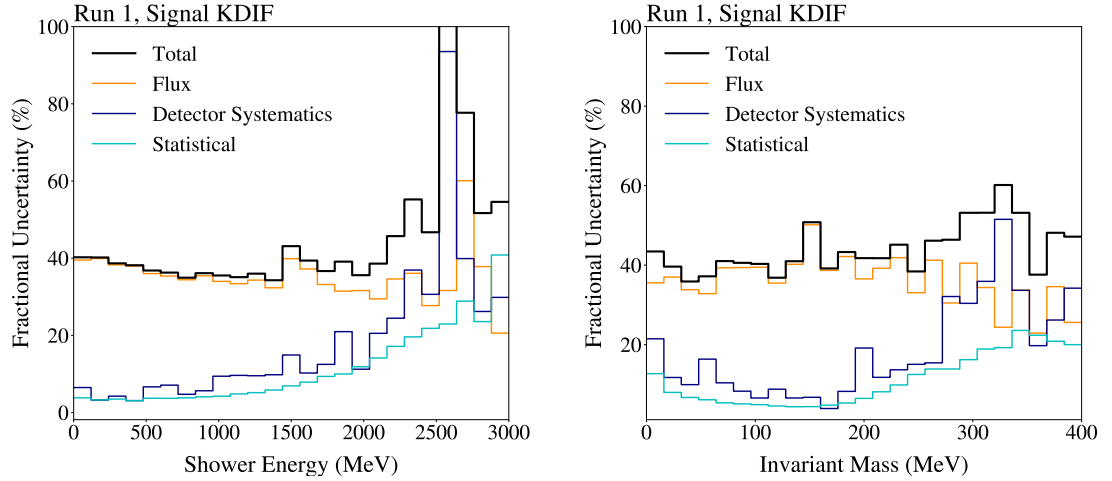


Figure 8.12: The percentage fractional uncertainty broken down by the source of uncertainty for the signal of scalars produced from KDIF. The distributions on the left and right correspond to samples for one-shower and two-shower pre-selections.

produced from KDAR and $\sim 40\%$ for signal produced from KDIF in bins with larger statistics. The second and third most dominant contributions in these bins are due to detector and statistical uncertainties. Large fluctuations in the detector uncertainty are due to low detector variation sample statistics. These uncertainties are of similar order for Run 3 samples and are therefore not shown here.

Chapter 9

Limit Setting

The predictions of the BDT models trained on the MC signal and SM background predictions can be used to calculate the Higgs portal scalar limit on the mixing parameter. Section 9.1 describes the statistical method used to set the upper limit on the mixing parameter, followed by section 9.2 showing this limit and the associated distributions used to calculate this limit.

9.1 Limit Extraction Procedure

A BSM or new physics search is ultimately described by a discriminating variable(s) chosen to be sensitive to a parameter of the search system. The result of the search analysis comprises distributions of this discriminating variable(s) for the BSM process, SM background process(es) and the observation from data. These discriminating variable distributions are used as inputs to perform a statistical analysis that derives statistical quantities by testing the observed data with the following two distinct hypothesis:

- **NULL hypothesis:** Also referred to as the “background-only” hypothesis is a model that is used as the basis of a test. In this model, no new or BSM physics (or signal) is present and in the analysis presented in this thesis, it represents the SM prediction for the NuMI background (described in section 5.8.2).

- **TEST hypothesis:** Also referred to as the “signal + background” hypothesis is an alternative model that is used for testing against the NULL hypothesis. This model is intended to describe new physics that has a distinguishable effect in addition to the SM background prediction which comprises the majority of the observed data. In the analysis presented in this thesis, the model represents the combined sum of the Higgs portal scalar signal (described in section 5.8.4) and the SM prediction for the NuMI background (described in section 5.8.2).

It is common to introduce a “signal strength” parameter μ to quantify the level of signal contribution in the test hypothesis, such that $\mu = 0$ corresponds to the null hypothesis and $\mu = 1$ corresponds to the test hypothesis. The objective is to evaluate this parameter, μ , the parameter of interest. In our study, this depends on the Higgs-scalar mixing parameter θ . The hypothesis for each Higgs portal scalar mass is constructed from a reference signal normalisation, which is then multiplied by the parameter of interest μ . The reference value and scaling parameter can be translated into the Higgs-scalar mixing parameter θ using the procedure discussed in section 5.8.4.

A hypothesis is defined by a set of parameters that are intrinsic to the MC model used. The parameters, excluding the parameter of interest θ are known as nuisance parameters. The uncertainties in these nuisance parameter values corresponds to the systematic uncertainties described in chapter 8.

9.1.1 The Statistical Test

A test static quantifies the degree to which the observations are consistent with the test and null hypotheses. For searches with low statistics, the nominal test statistic with the largest separation power is the likelihood ratio [93], where the ratio of Probability Distribution Functions (PDFs) for the test and null hypotheses is evaluated.

For Poisson PDFs with predicted number of background events b , signal events s , and observed data events d in a bin, the likelihoods for the two simple null L^{NULL}

and test L^{TEST} hypotheses can be expressed as [94]

$$\begin{aligned} L^{\text{NULL}}(s, b, d) &= \frac{e^{-b} b^d}{d!}, \\ L^{\text{TEST}}(s, b, d, \mu) &= \frac{e^{-(b+\mu s)} (b + \mu s)^d}{d!}, \end{aligned} \quad (9.1)$$

with Poisson likelihood ratio, Q , expressed as

$$Q(s, b, d, \mu) = \frac{L^{\text{TEST}}(s, b, d, \mu)}{L^{\text{NULL}}(s, b, d)} = \frac{\frac{e^{-(b+\mu s)} (b + \mu s)^d}{d!}}{\frac{e^{-b} b^d}{d!}}. \quad (9.2)$$

For multiple bins and/or multiple channels, the total likelihood is evaluated by taking the product of these likelihoods as

$$\begin{aligned} Q &= \prod_{i=1}^{N_{\text{channels}}} \prod_{j=1}^{N_{\text{bins}}} \frac{e^{-(b_{ij} + \mu s_{ij})} (b_{ij} + \mu s_{ij})^{d_{ij}}}{d_{ij}!} \cdot \frac{e^{-b_{ij}} b_{ij}^{d_{ij}}}{d_{ij}!}, \\ &= \prod_{i=1}^{N_{\text{channels}}} \prod_{j=1}^{N_{\text{bins}}} e^{-\mu s_{ij}} \left(\frac{b_{ij} + \mu s_{ij}}{b_{ij}} \right)^{d_{ij}}, \end{aligned} \quad (9.3)$$

where i and j are the indices that run over the number of channels N_{channels} and number of bins in a channel N_{bins} respectively. Maximising the likelihood for multiple bins is computationally very resource intensive. Therefore a Negative Log Likelihood Ratio (NLLR) test statistic Γ_{NLLR} is evaluated so that product of likelihoods in Q becomes a sum over likelihoods as

$$\Gamma_{\text{NLLR}} = -2 \ln(Q) = 2 \sum_{i=1}^{N_{\text{channels}}} \sum_{j=1}^{N_{\text{bins}}} \left(\mu s_{ij} - d_{ij} \ln \left(1 + \frac{\mu s_{ij}}{b_{ij}} \right) \right). \quad (9.4)$$

Calculating this test statistic is now significantly less resource intensive compared to the Poisson likelihood ratio in equation 9.3. To account for the uncertainties in the nuisance parameters, the predicted numbers of signal (s) and background (b) events become function of nuisance parameters η , $s(\eta)$, $b(\eta)$, in the equation 9.4. Inclusion of uncertainties in nuisance parameters broadens the PDFs and therefore degrades the separation power of the two hypotheses. Detailed explanation for the inclusion of these uncertainties in the calculation is described in Ref. [95].

A high value of Γ_{NLLR} indicates the dataset to have been more likely to have resulted from the distribution of the null hypothesis, whereas a low value of Γ_{NLLR} indicates dataset to have been more likely to have resulted from the distribution of the test hypothesis. The values of Γ_{NLLR} where the two test and null hypotheses overlap quantifies the sensitivity of the experiments. A large set of toy MC pseudo-data is generated to construct the PDFs for a NLLR test statistic of null and test hypotheses. In practice, the two distributions for null and test hypotheses generally overlap indicating no preference of any one hypothesis over the other in that overlapping region. An example for a Γ_{NLLR} is shown in figure 9.1, and the definitions of the terms such as CL_B and CL_{S+B} in legend is described in the following section.

9.1.2 Calculation of Confidence Levels

After producing the pseudo-data for the null and test hypotheses, the confidence levels or p-values for the two hypotheses are calculated. These confidence levels quantify the probability to observe an outcome in the pseudo-data sets that is less like the test hypothesis (less signal-like) than observed in data. For a reference data set, the confidence levels, CL, and p-values, PV, for the null (B) and test (S + B) hypotheses can be constructed using the semi-infinite integral of the PDFs of the two hypotheses given by $\frac{\partial\phi_B}{\partial\Gamma}$ and $\frac{\partial\phi_{S+B}}{\partial\Gamma}$, respectively, as

$$\begin{aligned}\text{CL}_B &= 1 - \text{PV}_B = \phi_B(\Gamma \geq \Gamma_{\text{ref}}) = \int_{\Gamma_{\text{ref}}}^{\infty} \frac{\partial\phi_B}{\partial\Gamma} d\Gamma, \\ \text{CL}_{S+B} &= \text{PV}_{S+B} = \phi_{S+B}(\Gamma \geq \Gamma_{\text{ref}}) = \int_{\Gamma_{\text{ref}}}^{\infty} \frac{\partial\phi_{S+B}}{\partial\Gamma} d\Gamma.\end{aligned}\tag{9.5}$$

The values for CL_B and CL_{S+B} correspond to the probability for the null and test hypotheses to produce an outcome more background-like than observed in the data. The values $1 - \text{CL}_B$ and CL_{S+B} correspond to the p-values for the null hypothesis PV_B and test hypothesis PV_{S+B} respectively. These confidence levels are functions of the signal strength μ that depends on the mixing parameter θ . Therefore a range of CL_B and CL_{S+B} values for different values of μ are evaluated. Figure 9.1 shows the corresponding test statistic PDFs with these values for an arbitrary value of the

mixing parameter at an example scalar mass of 150 MeV. To evaluate the confidence levels, the distributions are integrated from the observed NLLR outcomes. The agreement between the data and the background model is calculated by comparing the relationship of the null, test and observed test statistics.

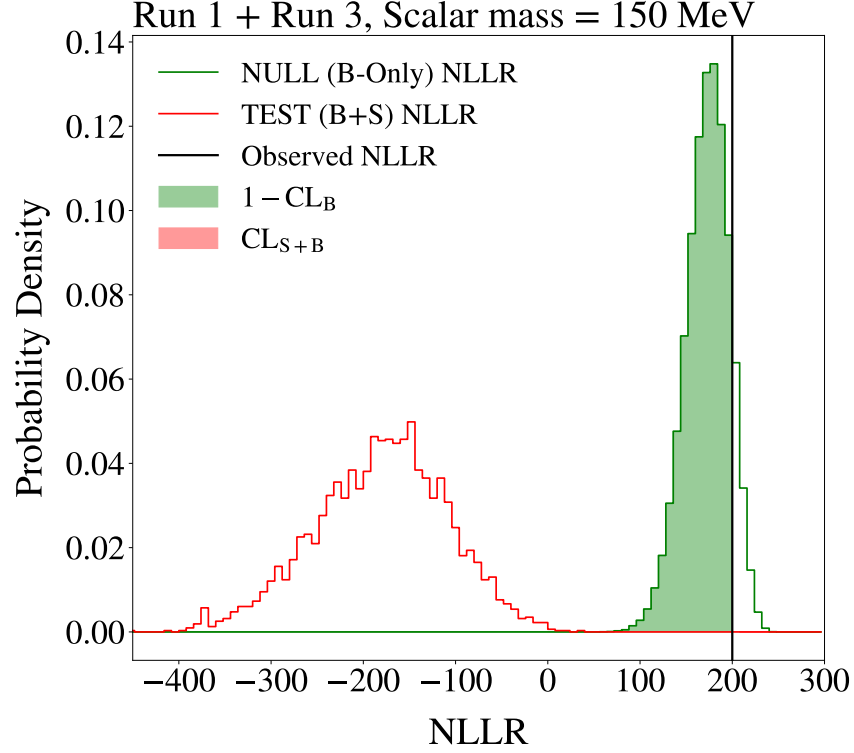


Figure 9.1: Probability density distributions for a NLLR test statistic calculated for null (green) and test (red) hypotheses. The shaded region in green corresponds to the p-value for the null hypothesis ($PV_B = 1 - CL_B$). The shaded region in red corresponding to the p-value for the test hypothesis ($PV_{S+B} = CL_{S+B}$) is extremely small and is therefore not visible in this figure. None of the model parameters in any of the scalar masses have overlapping regions of the two hypotheses and therefore an example scalar mass of 150 MeV for an arbitrary value of the model parameter is shown here.

A pure and traditional frequentist hypothesis test simply calculates CL_{S+B} to evaluate the exclusion limits for the model parameters. However, these limits are not reliable in the case where the data fluctuates significantly below the background predictions¹. This would result in a very strong constraint on the test hypothesis despite the null hypothesis poorly modelling the data. A modified background model or larger statistics will therefore be incapable of reproducing the exclusion limits.

¹Example: A counting experiment where the number of data events observed is significantly lower than that predicted by the null hypothesis. As a result, the NLLR will be much more positive than the median NLLR of the null hypothesis.

To overcome this, `Pyhf` [96] limit setting software (described in section 9.2) uses a modified-Frequentist statistic, CL_S , defined as

$$CL_S(\mu) = \frac{CL_{S+B}(\mu)}{CL_B(\mu)} = \frac{PV_{S+B}(\mu)}{1 - PV_B(\mu)}. \quad (9.6)$$

For exclusion this quantity is calculated as a function of the model parameter until $CL_S(\mu_{\text{limit}}) < \alpha$ where α is the fractional confidence level specified by the user to exclude the signal at a desired confidence level of $1 - \alpha$ and μ_{limit} is the value of the model parameter required to meet this condition.

In the analysis presented in this thesis, an exclusion is performed at the 95% CL which corresponds to $\alpha = 0.05$. The value of μ_{limit} corresponding to this CL is translated into the mixing parameter θ using the procedure outlined in section 5.8.4.

9.2 Results

`Pyhf`, a CLs method (described in previous section) software toolkit, is used to set the final limit on the model parameter of the analysis, the mixing parameter θ [96]. For one-shower and two-shower pre-selected samples, the energy of the shower and invariant mass distributions are chosen to be the discriminating variables sensitive to this model parameter. The uncertainties in these distributions are evaluated using the procedure outlined in chapter 8. A total of 8 different distributions corresponding to four different types of signal for Run 1 and Run 3 are combined in `Pyhf` to estimate the final limit for each scalar mass value. Section 9.2.1 outlines the procedure followed to evaluate the most optimal BDT score that yields the best limit in these 8 channels. Section 9.2.2 shows the distributions of the discriminating variable for these 8 channels after applying the BDT score selection. The uncertainties associated to each bin in these distributions are shown in the lower panel of these distributions. Finally section 9.2.3 shows the derived 95% CL Higgs portal scalar median expected and observed limit on the mixing parameter θ .

9.2.1 BDT Score Selection

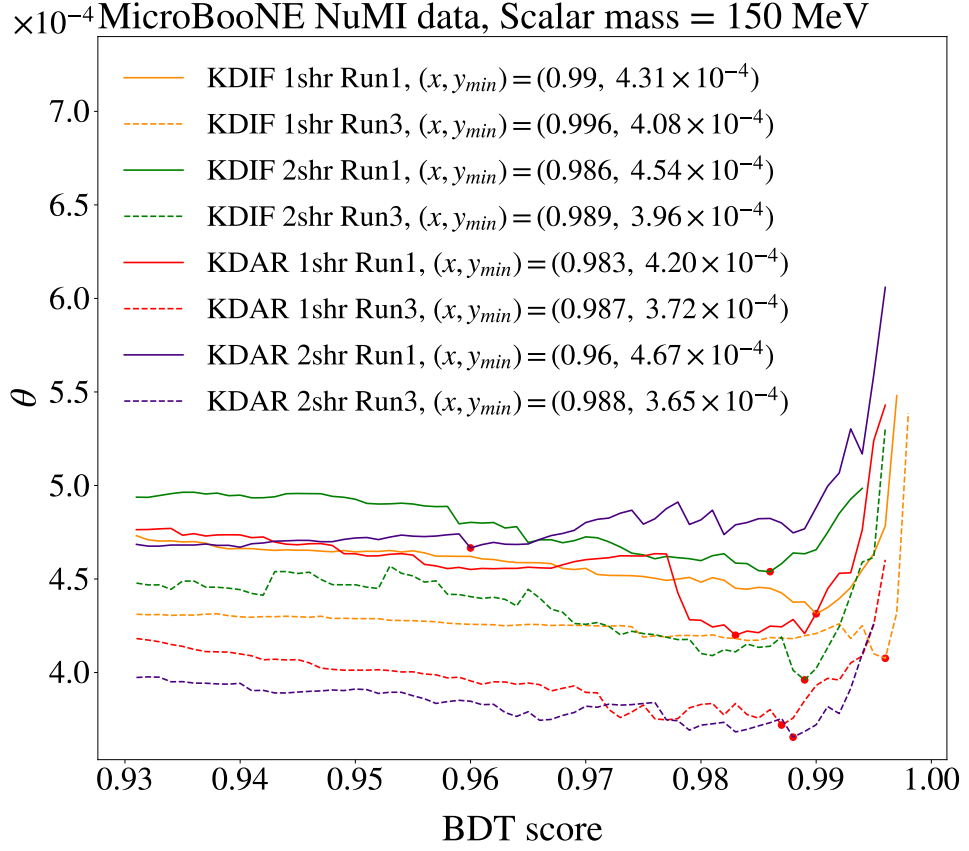


Figure 9.2: The median expected limit on the mixing parameter as a function of the BDT score for four different types of signal (indicated by 4 different colours) produced from scalars of mass 150 MeV for Run 1 (solid line) and Run 3 (dashed line) samples. The coordinates of the minima for each curve (red marker) are shown in the legend.

Figure 9.2 shows the median expected limit on the mixing parameter as a function of BDT score for four different types of signals (channels) produced from scalar of mass 150 MeV for Run 1 (solid line) and Run 3 (dashed line) samples. The minima (red marker on the curves with its corresponding value shown in legend) for each of the total 8 channels corresponds to the value of the BDT score that yields the best limit. Due to larger statistics of Run 3, the median expected limits for different channels in Run 3 are better than in Run 1 with the two most important channels being scalars produced from KDAR in samples with one-shower and two-shower pre-selections applied. These 8 channels represent one scalar mass value and the associated distributions of invariant mass (for two-shower like events) and energy (for one-shower like events) for these channels are eventually combined into the `Pyhf`

limit setting software to calculate the final limit for this mass point. This process is followed for all scalar mass values to evaluate the 95% CL observed and median expected limits described in the next section.

The background MC sample statistics for large values of BDT score is limited and therefore the minima for each of these curves is driven by the statistical fluctuations. The choice of optimising the BDT score using this method is therefore open to question and a better approach such as curve fitting to the data points is required to interpolate the values of the minima. The analysis presented in this thesis calculates the minima without fitting the curve to these data points assuming that the method will have insignificant impact on the final limit.

9.2.2 Application of the BDT Score Selection

Figures 9.3 – 9.18 show the distributions of the discriminating variables, the invariant mass for two-shower like events and the shower energy for one-shower like events, after applying the optimal BDT score selections described in previous section. The distributions correspond to the 8 different channels described in the previous section for each scalar mass value used in setting the final limit. The background is sparsely distributed in these discriminating variables and to avoid overestimating the sensitivity in the limit setting. These distributions are flattened into a single-binned histogram.

Scalar Mass (MeV)	Observed	Expected	Total Uncertainty	Excess (σ)
100	36	21.72	6.38	2.24
125	44	34.75	14.29	0.65
130	22	17.12	6.88	0.71
135	34	42.57	18.35	-0.47
140	39	24.57	8.62	1.67
145	27	24.61	10.79	0.22
150	37	25.25	7.47	1.57
200	36	23.19	8.21	1.56

Table 9.1: The total number of events observed in data, events expected in the background prediction (MC+EXT) and the total (statistical + systematic) uncertainty associated with the MC prediction.

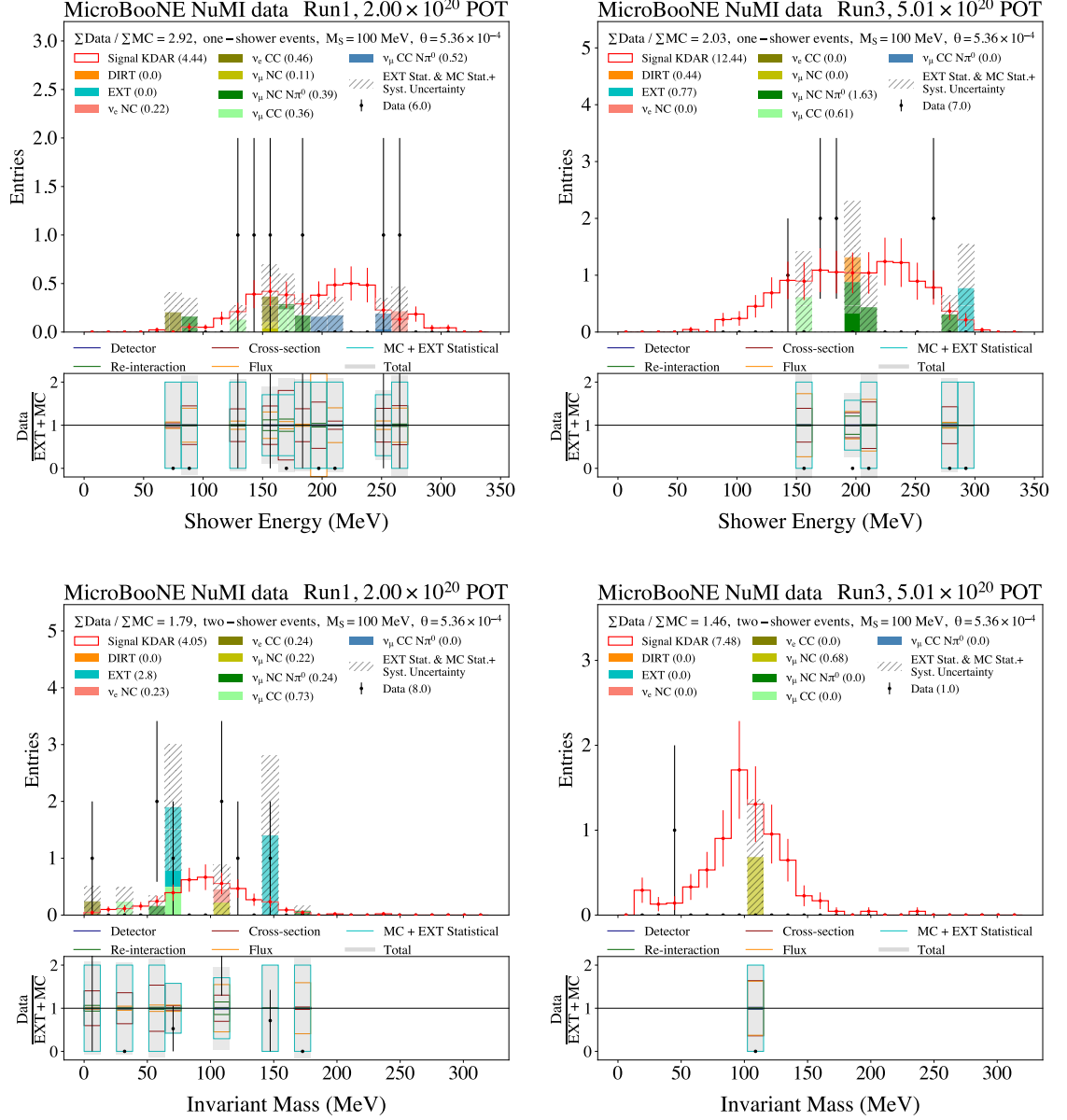


Figure 9.3: A comparison of the stacked prediction (EXT+MC) to the data (beam-on) for the distributions of shower energy (top panels) and invariant mass (bottom panels) after the one-shower and two-shower pre-selections respectively. These distributions are overlaid on top of the Higgs portal scalar distributions normalised using the value of θ corresponding to the KOTO central-value (described in section 5.8.4) for a scalar of mass 100 MeV produced from KDAR. The distributions on the left and right correspond to samples for Run 1 and Run 3 respectively.

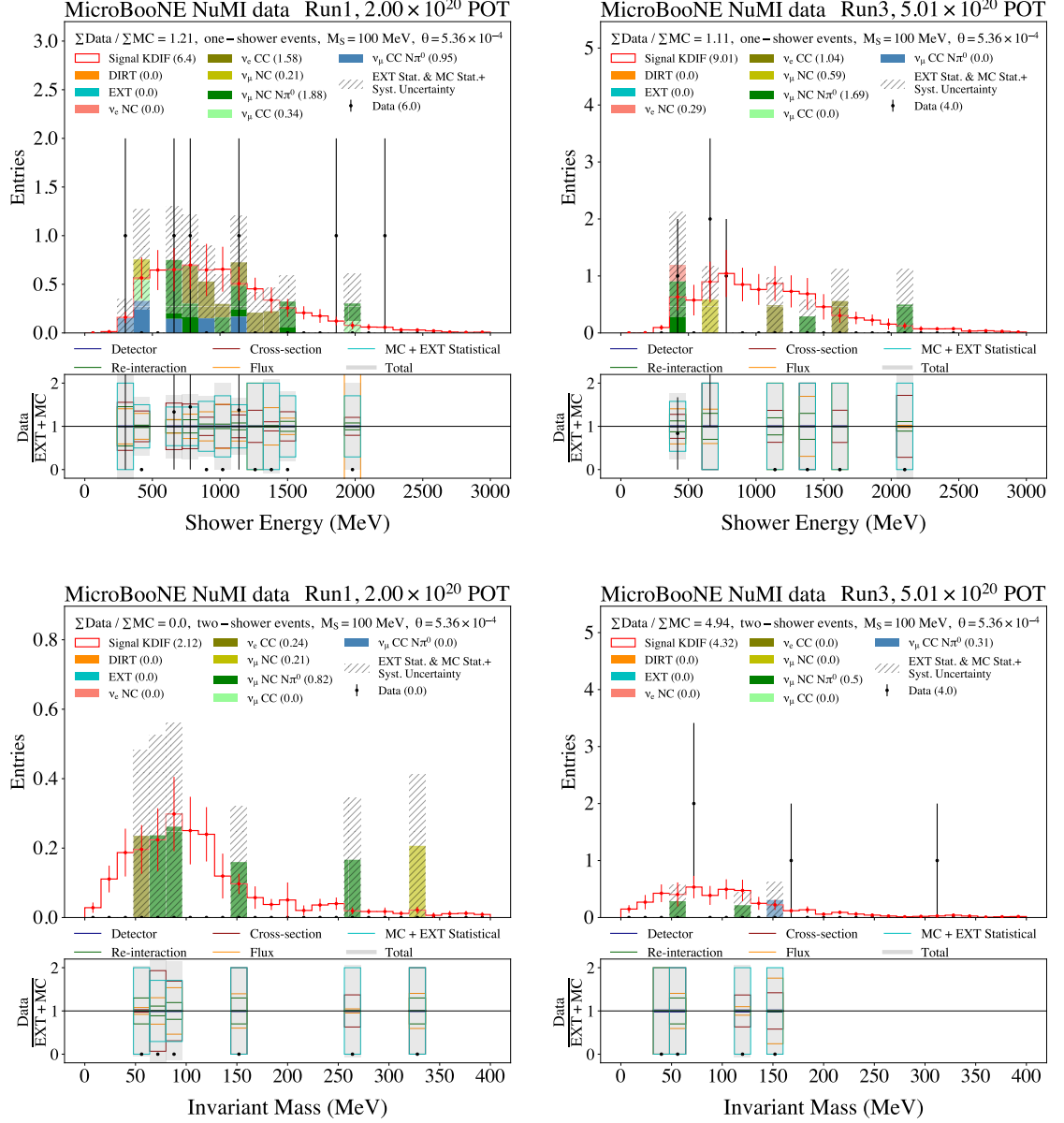


Figure 9.4: A comparison of the stacked prediction (EXT+MC) to the data (beam-on) for the distributions of shower energy (top panels) and invariant mass (bottom panels) after the one-shower and two-shower pre-selections respectively. These distributions are overlaid on top of the Higgs portal scalar distributions normalised using the value of θ corresponding to the KOTO central-value (described in section 5.8.4) for a scalar of mass 100 MeV produced from KDIF. The distributions on the left and right correspond to samples for Run 1 and Run 3 respectively.

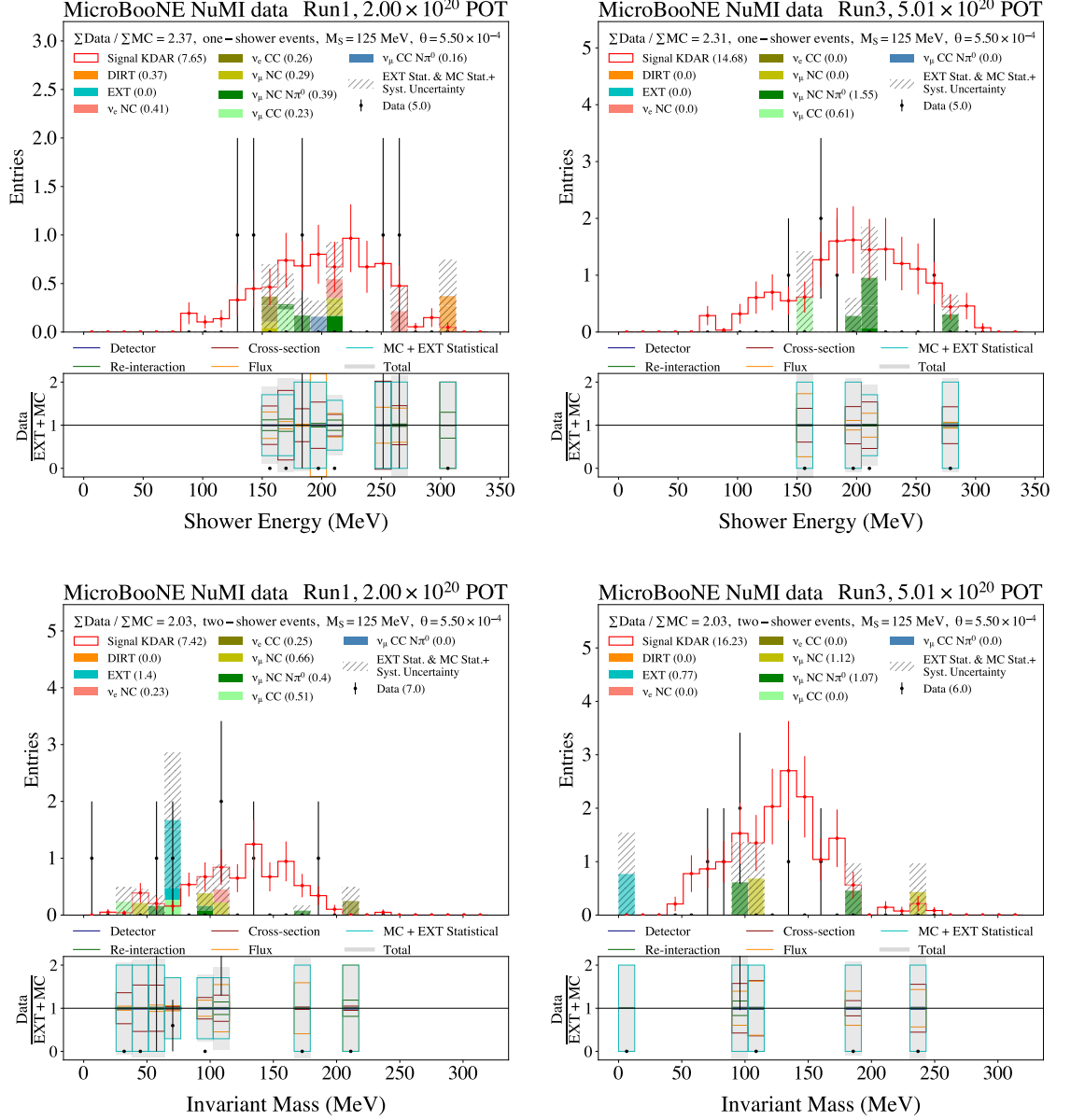


Figure 9.5: A comparison of the stacked prediction (EXT+MC) to the data (beam-on) for the distributions of shower energy (top panels) and invariant mass (bottom panels) after the one-shower and two-shower pre-selections respectively. These distributions are overlaid on top of the Higgs portal scalar distributions normalised using the value of θ corresponding to the KOTO central-value (described in section 5.8.4) for a scalar of mass 125 MeV produced from KДАР. The distributions on the left and right correspond to samples for Run 1 and Run 3 respectively.

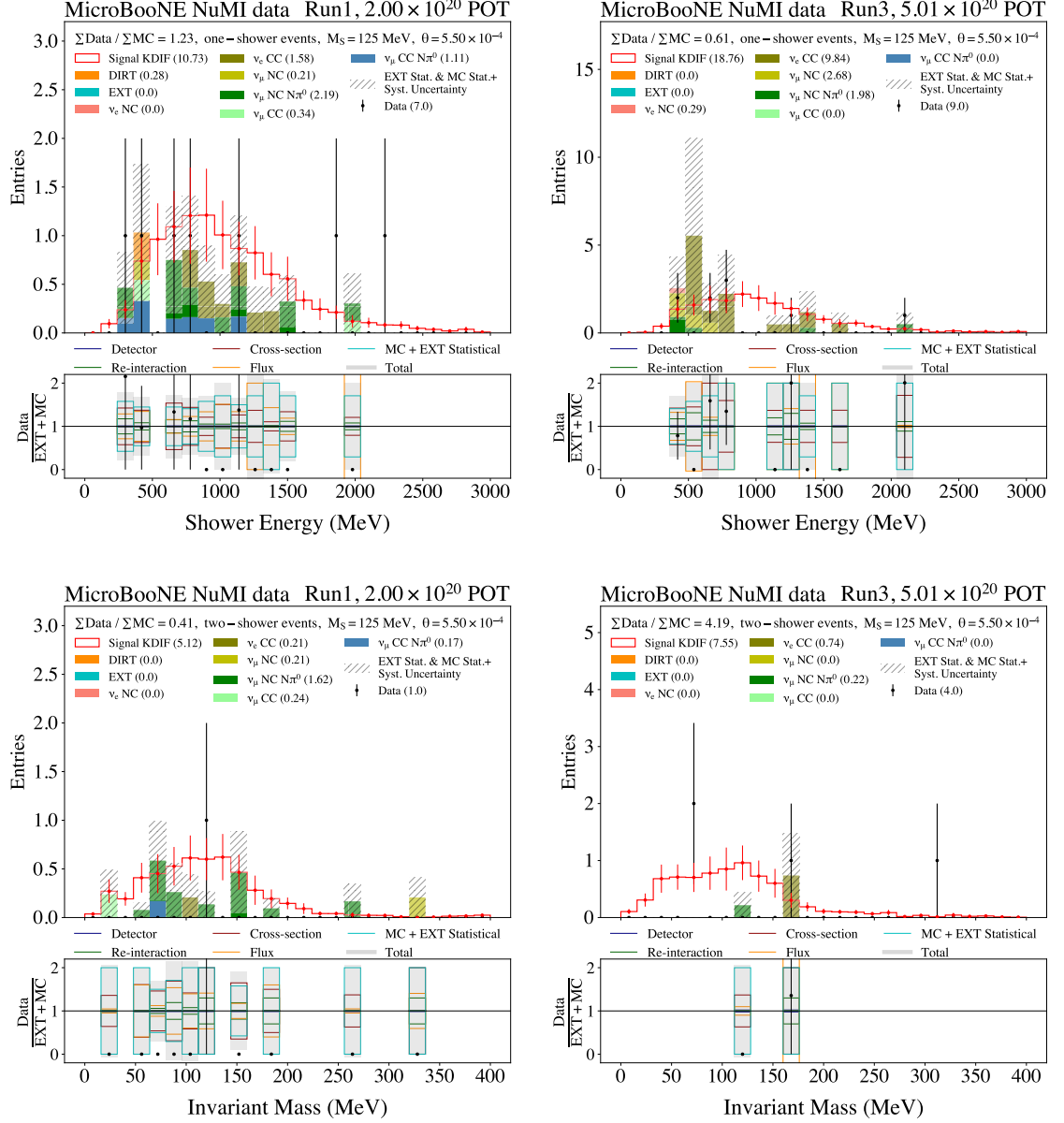


Figure 9.6: A comparison of the stacked prediction (EXT+MC) to the data (beam-on) for the distributions of shower energy (top panels) and invariant mass (bottom panels) after the one-shower and two-shower pre-selections respectively. These distributions are overlaid on top of the Higgs portal scalar distributions normalised using the value of θ corresponding to the KOTO central-value (described in section 5.8.4) for a scalar of mass 125 MeV produced from KDIF. The distributions on the left and right correspond to samples for Run 1 and Run 3 respectively.

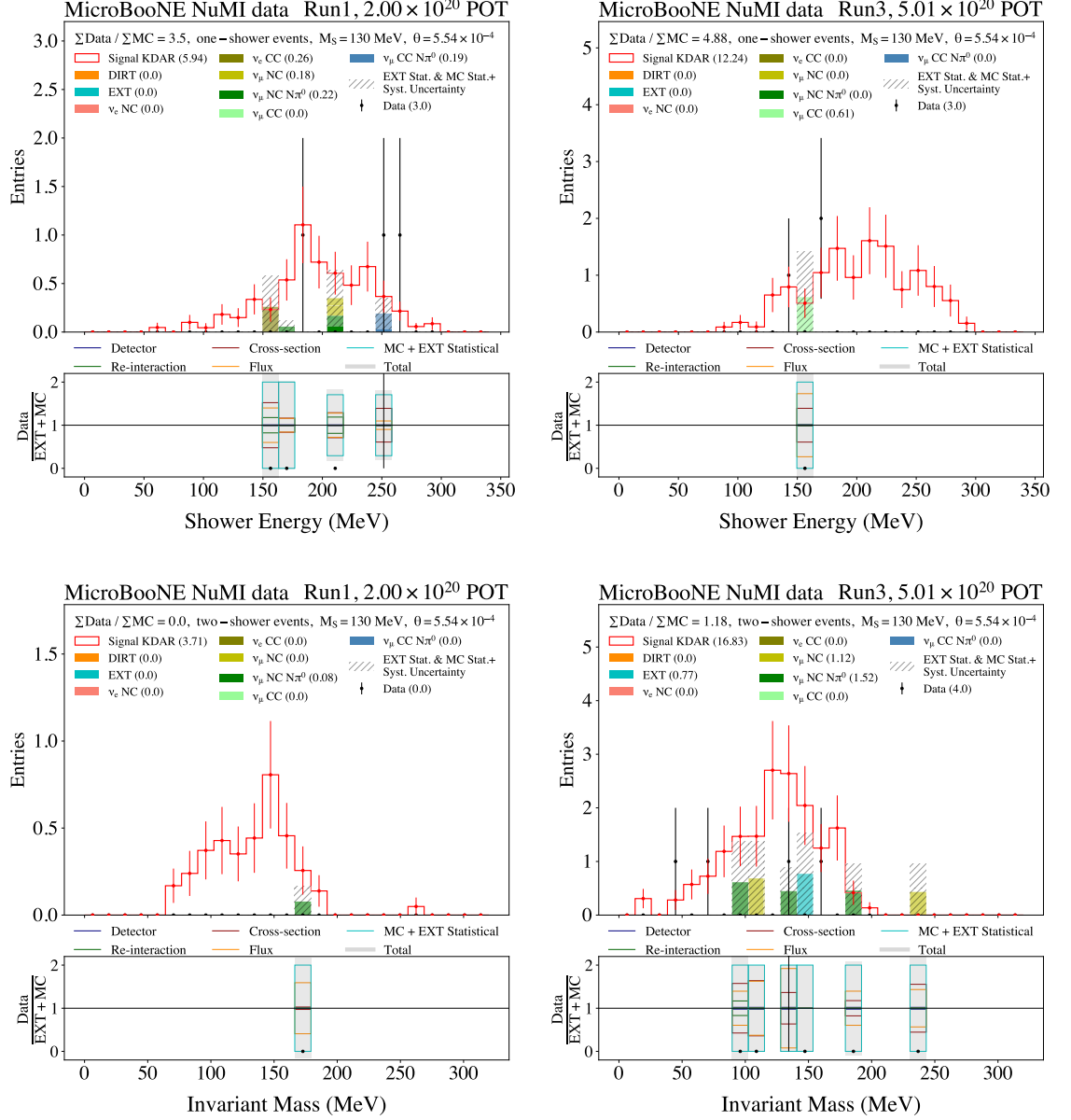


Figure 9.7: A comparison of the stacked prediction (EXT+MC) to the data (beam-on) for the distributions of shower energy (top panels) and invariant mass (bottom panels) after the one-shower and two-shower pre-selections respectively. These distributions are overlaid on top of the Higgs portal scalar distributions normalised using the value of θ corresponding to the KOTO central-value (described in section 5.8.4) for a scalar of mass 130 MeV produced from KDAR. The distributions on the left and right correspond to samples for Run 1 and Run 3 respectively.

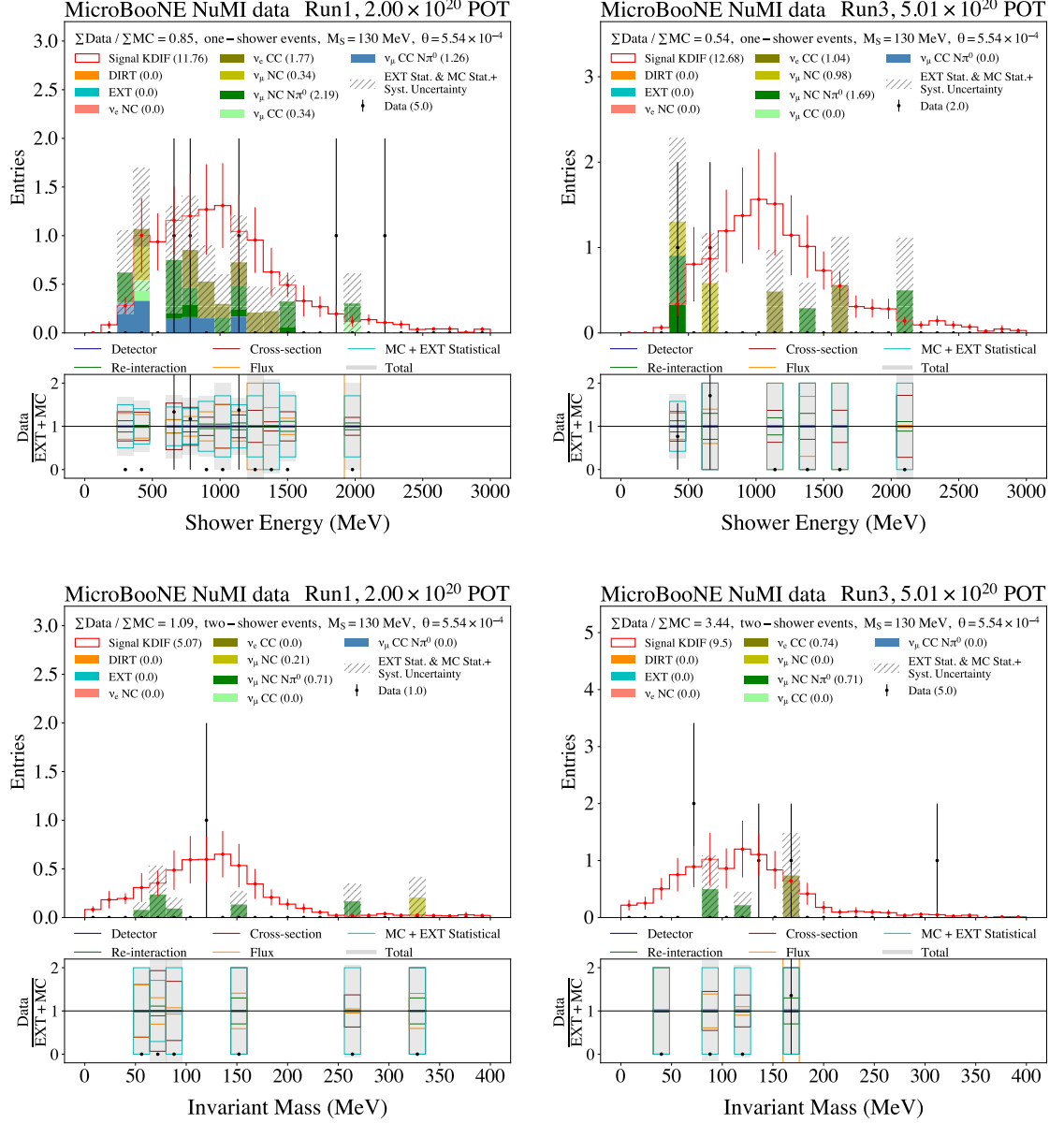


Figure 9.8: A comparison of the stacked prediction (EXT+MC) to the data (beam-on) for the distributions of shower energy (top panels) and invariant mass (bottom panels) after the one-shower and two-shower pre-selections respectively. These distributions are overlaid on top of the Higgs portal scalar distributions normalised using the value of θ corresponding to the KOTO central-value (described in section 5.8.4) for a scalar of mass 130 MeV produced from KDIF. The distributions on the left and right correspond to samples for Run 1 and Run 3 respectively.

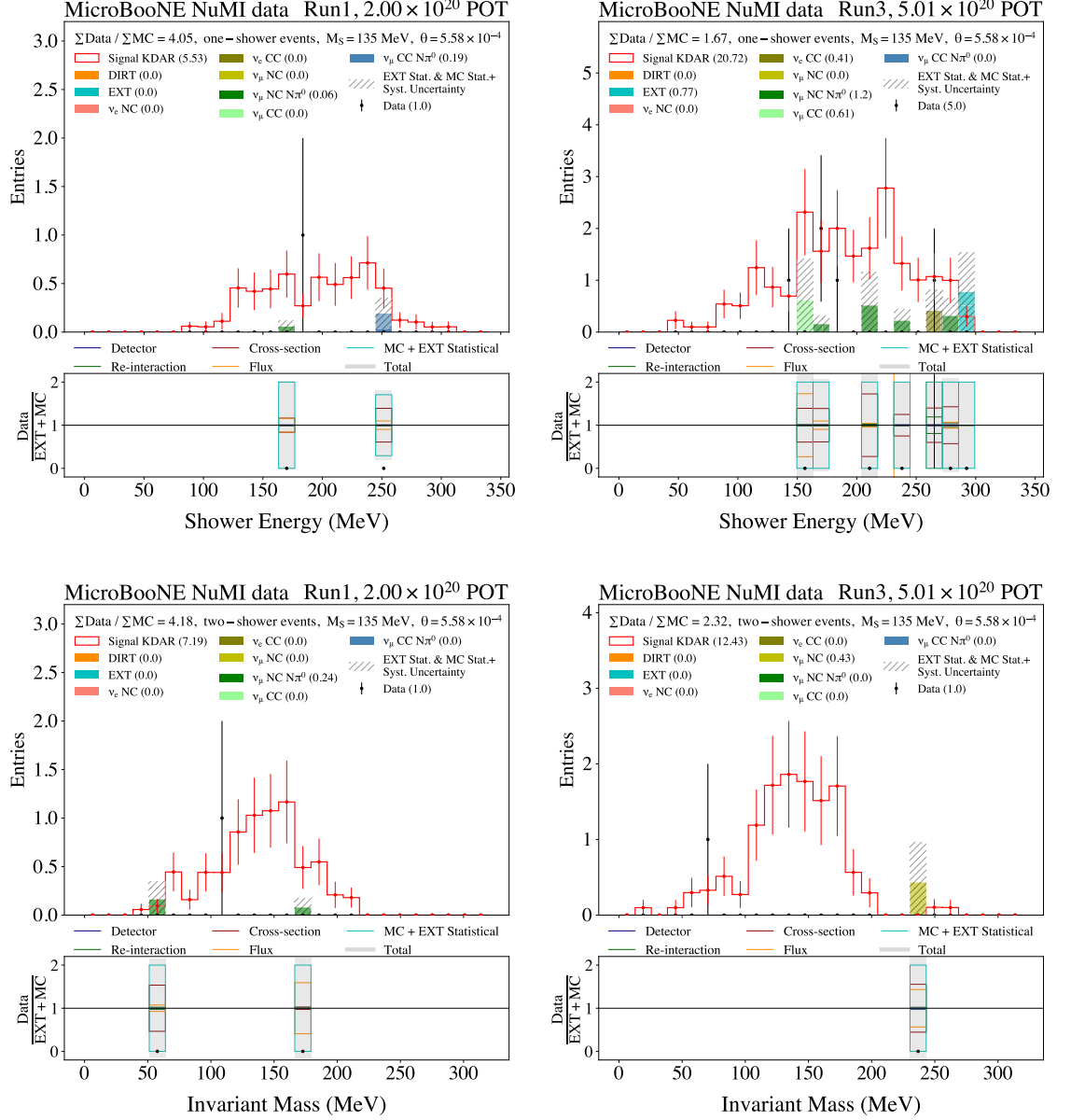
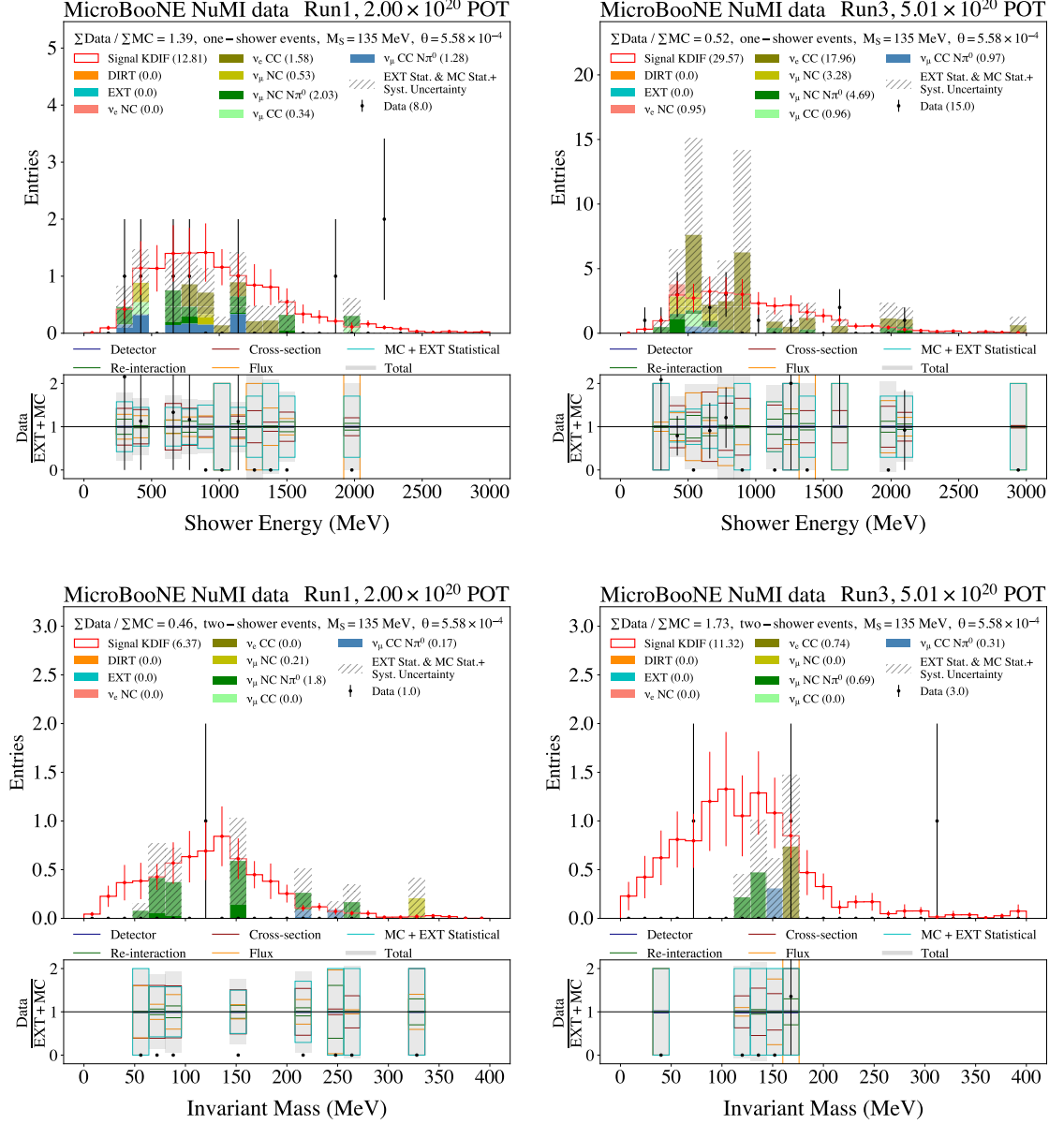


Figure 9.9: A comparison of the stacked prediction (EXT+MC) to the data (beam-on) for the distributions of shower energy (top panels) and invariant mass (bottom panels) after the one-shower and two-shower pre-selections respectively. These distributions are overlaid on top of the Higgs portal scalar distributions normalised using the value of θ corresponding to the KOTO central-value (described in section 5.8.4) for a scalar of mass 135 MeV produced from KDAR. The distributions on the left and right correspond to samples for Run 1 and Run 3 respectively.



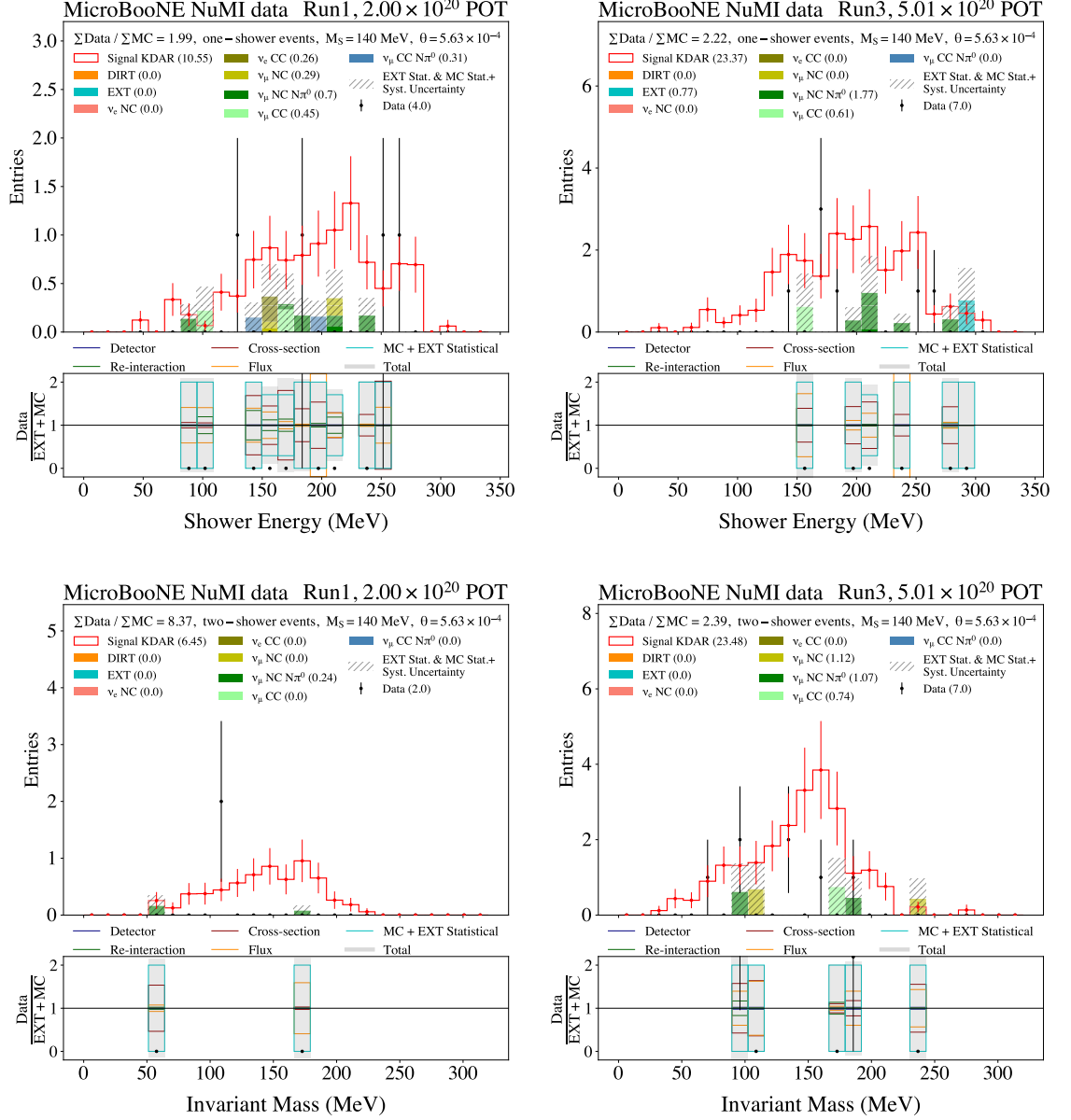


Figure 9.11: A comparison of the stacked prediction (EXT+MC) to the data (beam-on) for the distributions of shower energy (top panels) and invariant mass (bottom panels) after the one-shower and two-shower pre-selections respectively. These distributions are overlaid on top of the Higgs portal scalar distributions normalised using the value of θ corresponding to the KOTO central-value (described in section 5.8.4) for a scalar of mass 140 MeV produced from KDAR. The distributions on the left and right correspond to samples for Run 1 and Run 3 respectively.

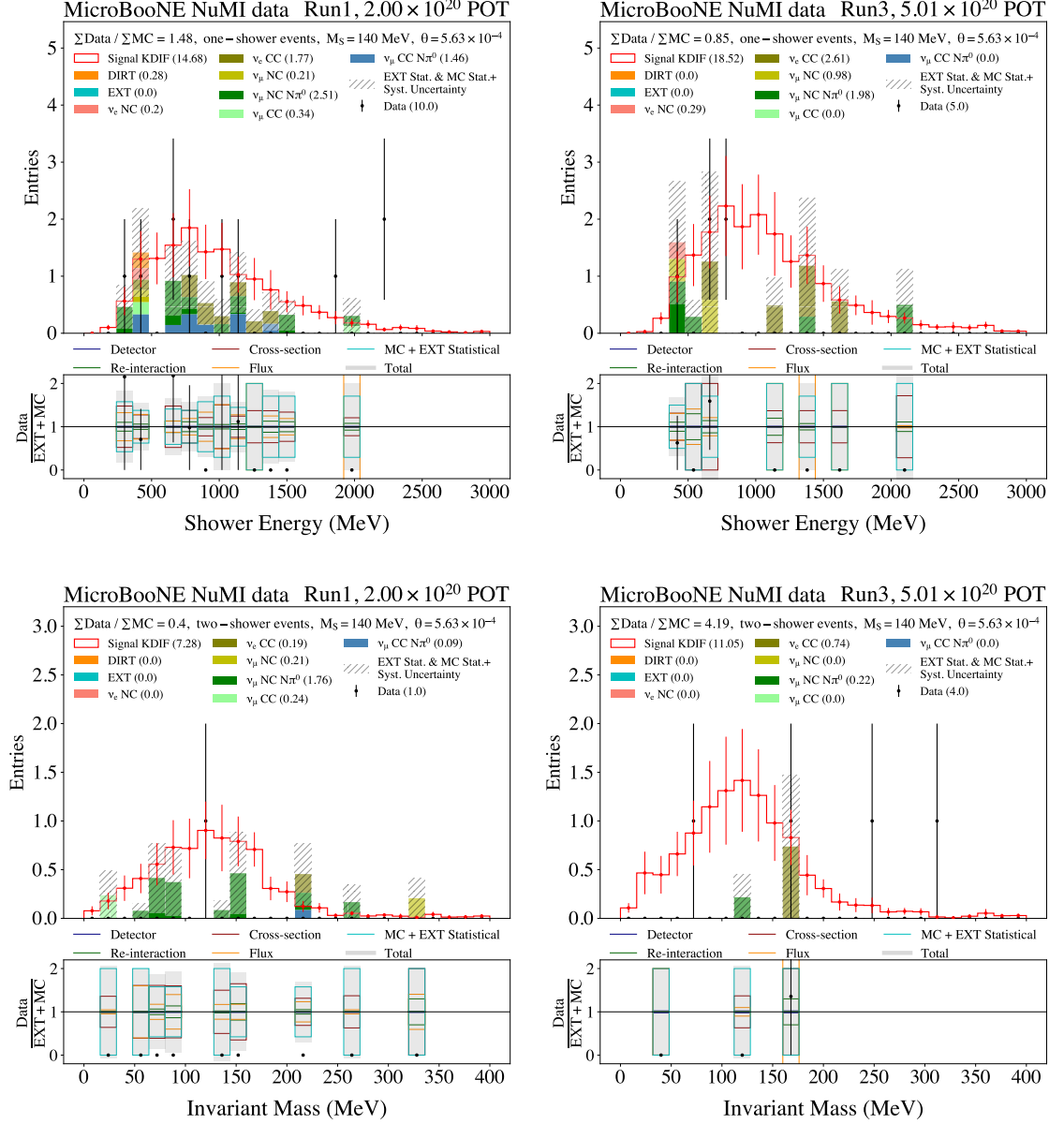


Figure 9.12: A comparison of the stacked prediction (EXT+MC) to the data (beam-on) for the distributions of shower energy (top panels) and invariant mass (bottom panels) after the one-shower and two-shower pre-selections respectively. These distributions are overlaid on top of the Higgs portal scalar distributions normalised using the value of θ corresponding to the KOTO central-value (described in section 5.8.4) for a scalar of mass 140 MeV produced from KDIF. The distributions on the left and right correspond to samples for Run 1 and Run 3 respectively.

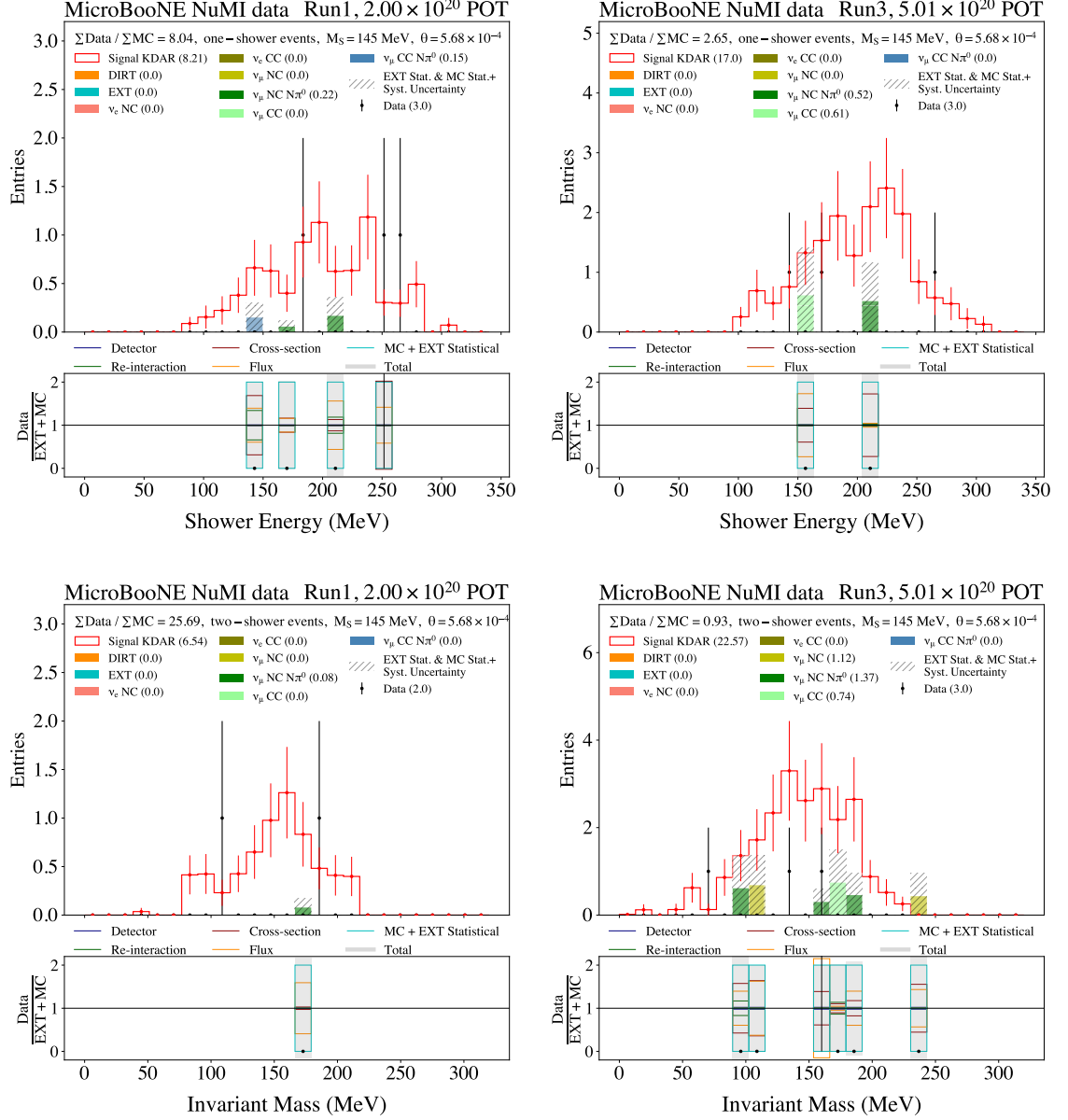


Figure 9.13: A comparison of the stacked prediction (EXT+MC) to the data (beam-on) for the distributions of shower energy (top panels) and invariant mass (bottom panels) after the one-shower and two-shower pre-selections respectively. These distributions are overlaid on top of the Higgs portal scalar distributions normalised using the value of θ corresponding to the KOTO central-value (described in section 5.8.4) for a scalar of mass 145 MeV produced from KDAR. The distributions on the left and right correspond to samples for Run 1 and Run 3 respectively.

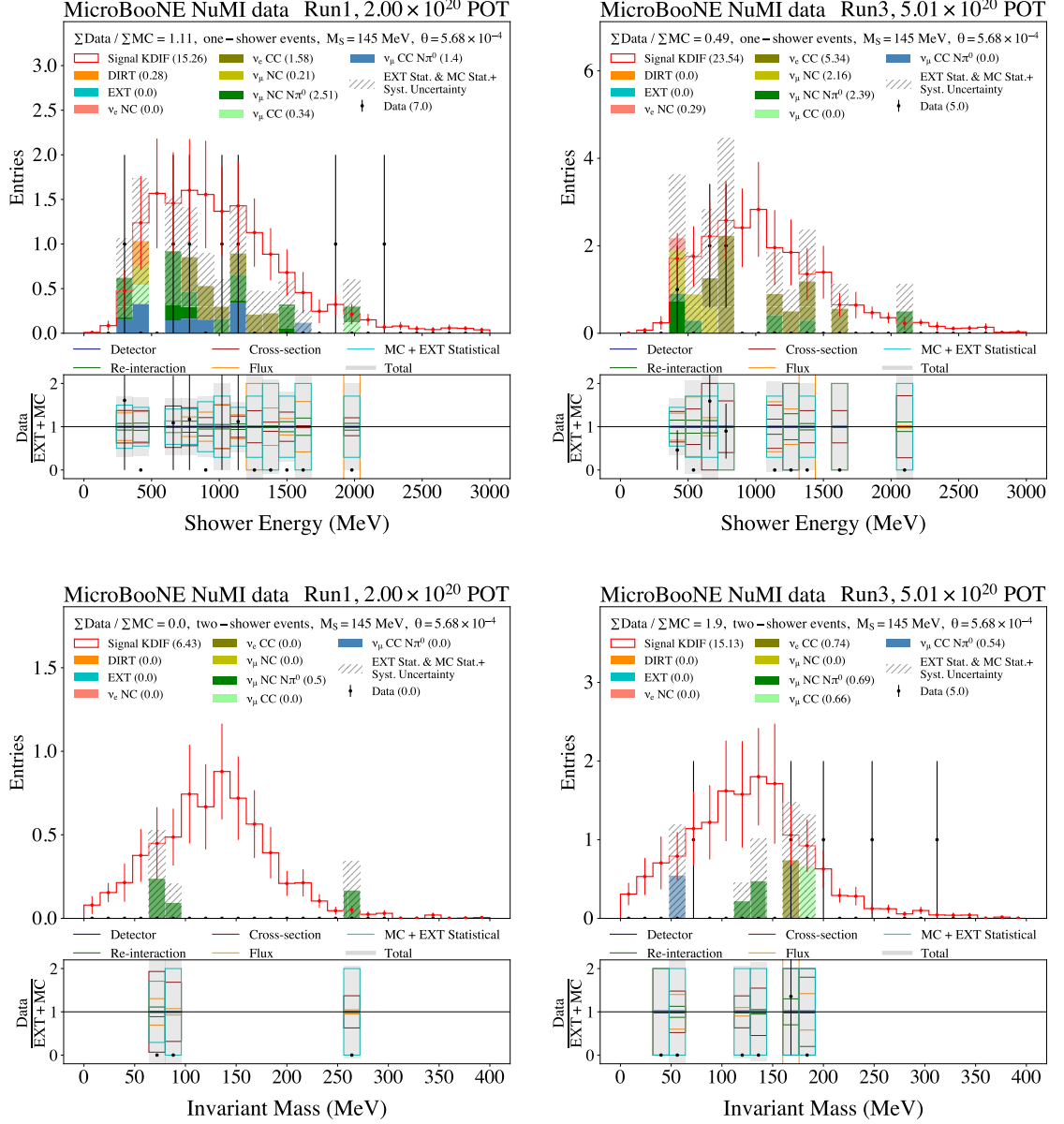


Figure 9.14: A comparison of the stacked prediction (EXT+MC) to the data (beam-on) for the distributions of shower energy (top panels) and invariant mass (bottom panels) after the one-shower and two-shower pre-selections respectively. These distributions are overlaid on top of the Higgs portal scalar distributions normalised using the value of θ corresponding to the KOTO central-value (described in section 5.8.4) for a scalar of mass 145 MeV produced from KDIF. The distributions on the left and right correspond to samples for Run 1 and Run 3 respectively.

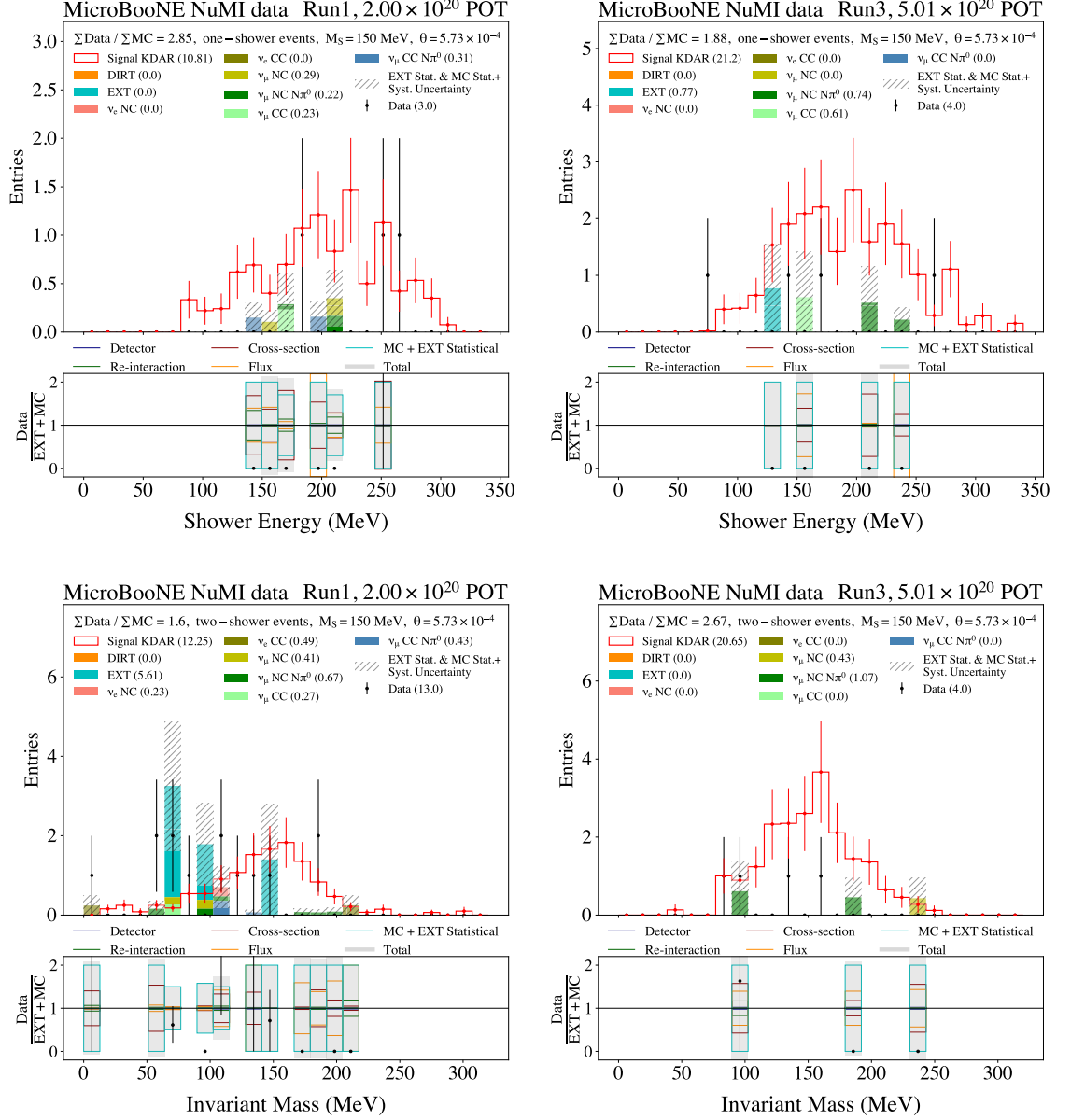


Figure 9.15: A comparison of the stacked prediction (EXT+MC) to the data (beam-on) for the distributions of shower energy (top panels) and invariant mass (bottom panels) after the one-shower and two-shower pre-selections respectively. These distributions are overlaid on top of the Higgs portal scalar distributions normalised using the value of θ corresponding to the KOTO central-value (described in section 5.8.4) for a scalar of mass 150 MeV produced from KDAR. The distributions on the left and right correspond to samples for Run 1 and Run 3 respectively.

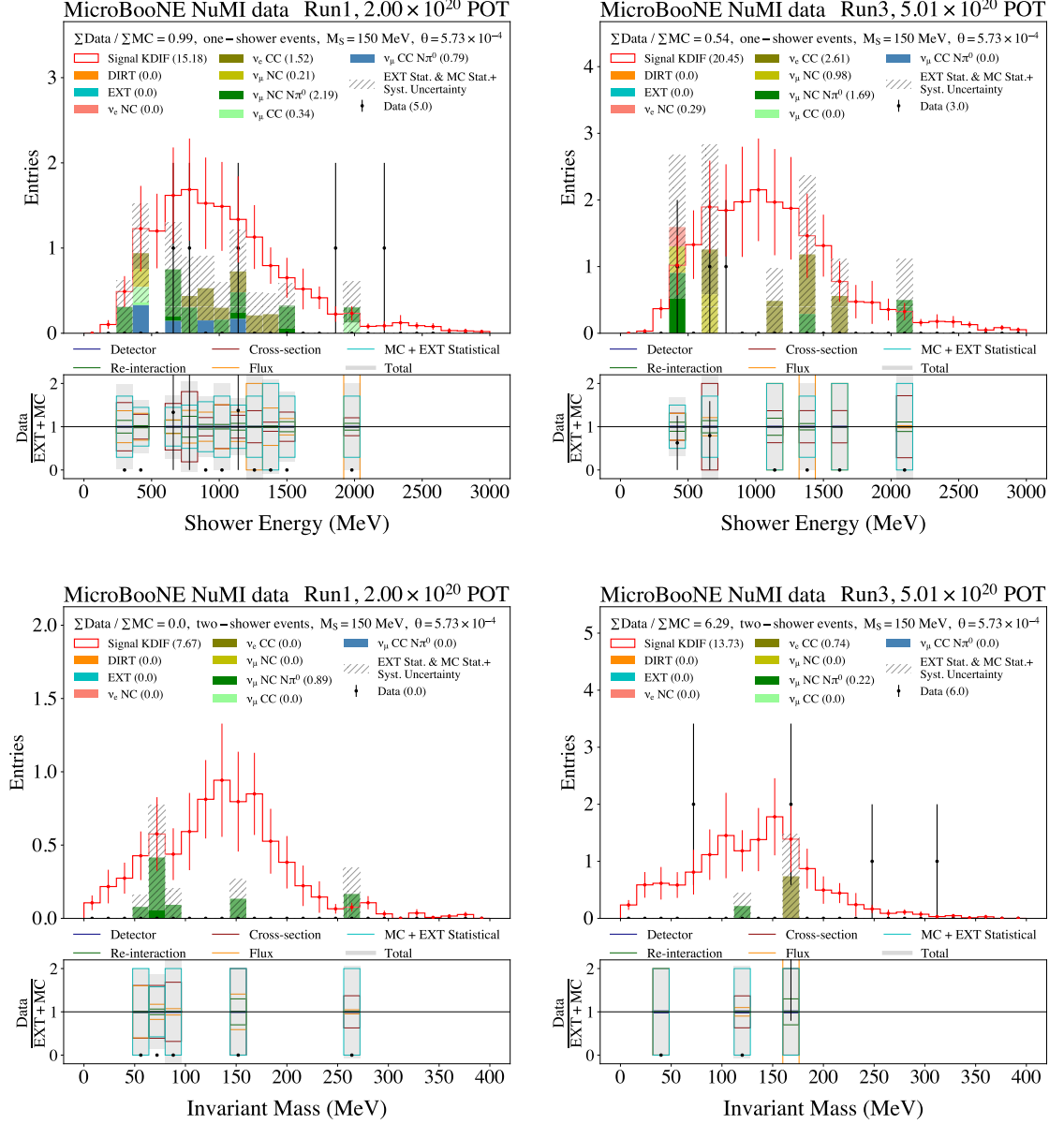
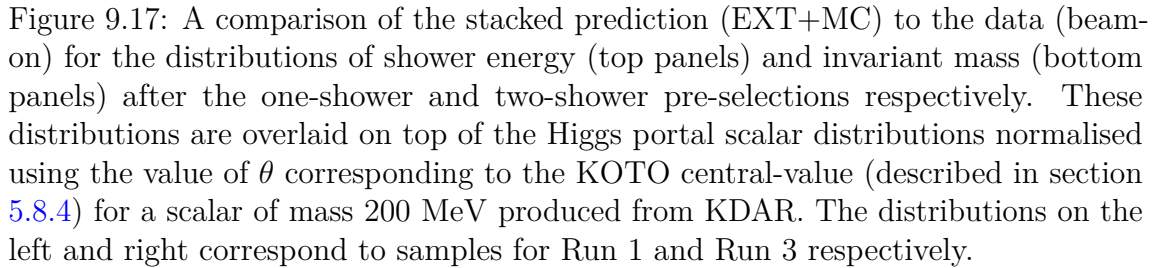


Figure 9.16: A comparison of the stacked prediction (EXT+MC) to the data (beam-on) for the distributions of shower energy (top panels) and invariant mass (bottom panels) after the one-shower and two-shower pre-selections respectively. These distributions are overlaid on top of the Higgs portal scalar distributions normalised using the value of θ corresponding to the KOTO central-value (described in section 5.8.4) for a scalar of mass 150 MeV produced from KDIF. The distributions on the left and right correspond to samples for Run 1 and Run 3 respectively.



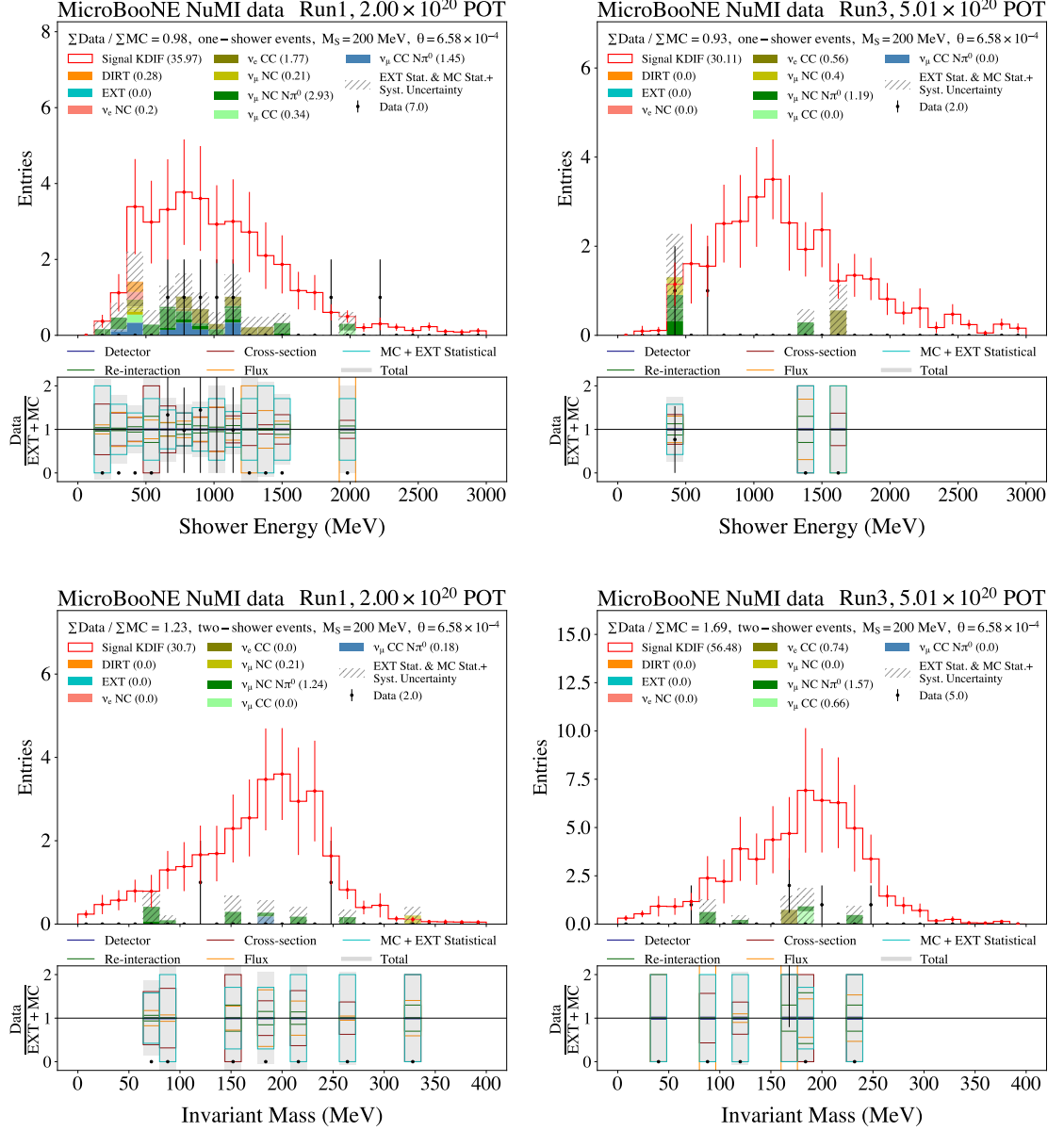


Figure 9.18: A comparison of the stacked prediction (EXT+MC) to the data (beam-on) for the distributions of shower energy (top panels) and invariant mass (bottom panels) after the one-shower and two-shower pre-selections respectively. These distributions are overlaid on top of the Higgs portal scalar distributions normalised using the value of θ corresponding to the KOTO central-value (described in section 5.8.4) for a scalar of mass 200 MeV produced from KDIF. The distributions on the left and right correspond to samples for Run 1 and Run 3 respectively.

Table 9.1 shows the total number of events observed in data, events expected in the background prediction (MC+EXT) and the total (statistical + systematic) uncertainty associated with the background prediction. No significant excess is observed for any of the scalar mass points. The dominant source of uncertainty in background predictions for each scalar mass value is the flux and cross-section uncertainty as discussed in section 8.7 except for $M_S = 135$ MeV where MC statistical uncertainty also contributes and produces the largest fractional uncertainty ($\sim 40\%$) compared to other scalar mass values.

9.2.3 Setting a Limit on the Higgs Portal Scalar Mixing Parameter

M_S (MeV)	MicroBooNE Limits on θ ($\times 10^{-4}$)			
	Observed	Median Expected	Exp. $\pm 1\sigma$	Exp. $\pm 2\sigma$
100	3.08	3.26	2.94 - 3.63	2.69 - 3.97
125	2.93	3.06	2.77 - 3.39	2.55 - 3.73
130	2.88	2.94	2.65 - 3.29	2.42 - 3.65
135	3.21	2.87	2.58 - 3.21	2.35 - 3.57
140	2.79	2.85	2.58 - 3.17	2.37 - 3.49
145	2.47	2.79	2.51 - 3.12	2.30 - 3.45
150	2.9	2.83	2.56 - 3.15	2.35 - 3.48
200	2.31	2.41	2.17 - 2.71	2.03 - 3.00

Table 9.2: The 95% CL observed and median expected limits on the Higgs portal scalar mixing parameter θ for each individual scalar mass value. These limits are produced using NuMI data (beam-on) corresponding to an exposure of combined Run 1 and Run 3 with 7.01×10^{20} POT. The numbers for observed, median expected and their corresponding $\pm 1\sigma$ and $\pm 2\sigma$ limits in this table are all multiplied by a factor of 10^4 .

Table 9.2 and figure 9.19 show the 95% CL observed and median expected limits on the mixing parameter of the Higgs portal scalar model, calculated by combining the corresponding 8 distributions shown in the previous section for each scalar mass value. The bands in green and yellow show 1σ (68%) and 2σ (95%) interval bands on the expected limit respectively. The observed limit for each scalar mass point is contained within the 1σ interval with an exception of $M_S = 145$ MeV which is marginally outside the 1σ interval.

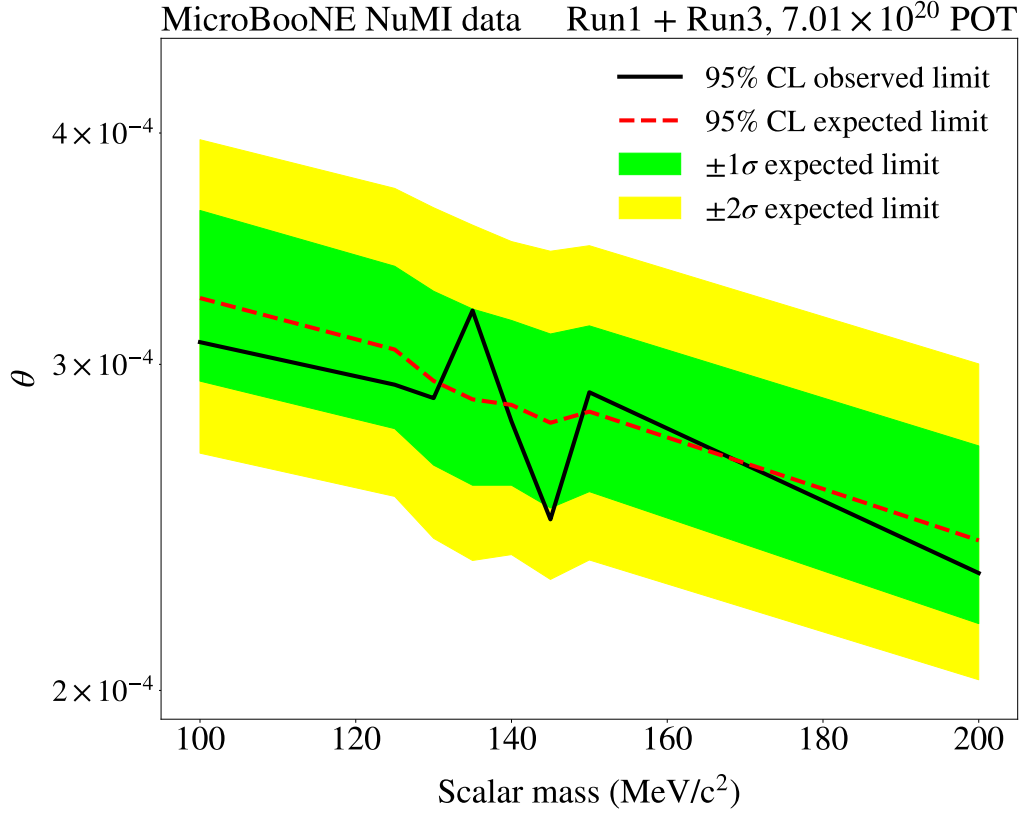


Figure 9.19: The 95% CL median expected and observed limit on the Higgs portal scalar mixing parameter θ . The bands in green and yellow show 1σ and 2σ uncertainty on the expected limit respectively.

At $M_S = 135$ MeV the observed limit is worse than the expected even though there is an overall deficit at this mass value as shown in table 9.1. This is because the two channels that contribute the most to this limit, the KDAR one-shower and two-shower Run 3 distributions, both have an excess of data over the background predictions, whereas the deficit is mostly in the KDIF one-shower in the Run 3 channel, which contributes less to the combined expected and observed limits. At $M_S = 145$ MeV, the observed limit is better than expected even though there is an overall excess of data as shown in table 9.1. This is because the KDIF one-shower Run 3 channel has a deficit of the observed data with respect to the background expectation and this channel contributes significantly to the combined observed limit.

Figure 9.20 shows the region of $(\theta - M_S)$ parameter space that corresponds to the observed exclusion (solid black line) calculated in the analysis presented in this thesis. The existing experimental limits for dedicated Higgs portal scalar searches are shown in solid lines, whereas reinterpretations of historical experimental data are shown in

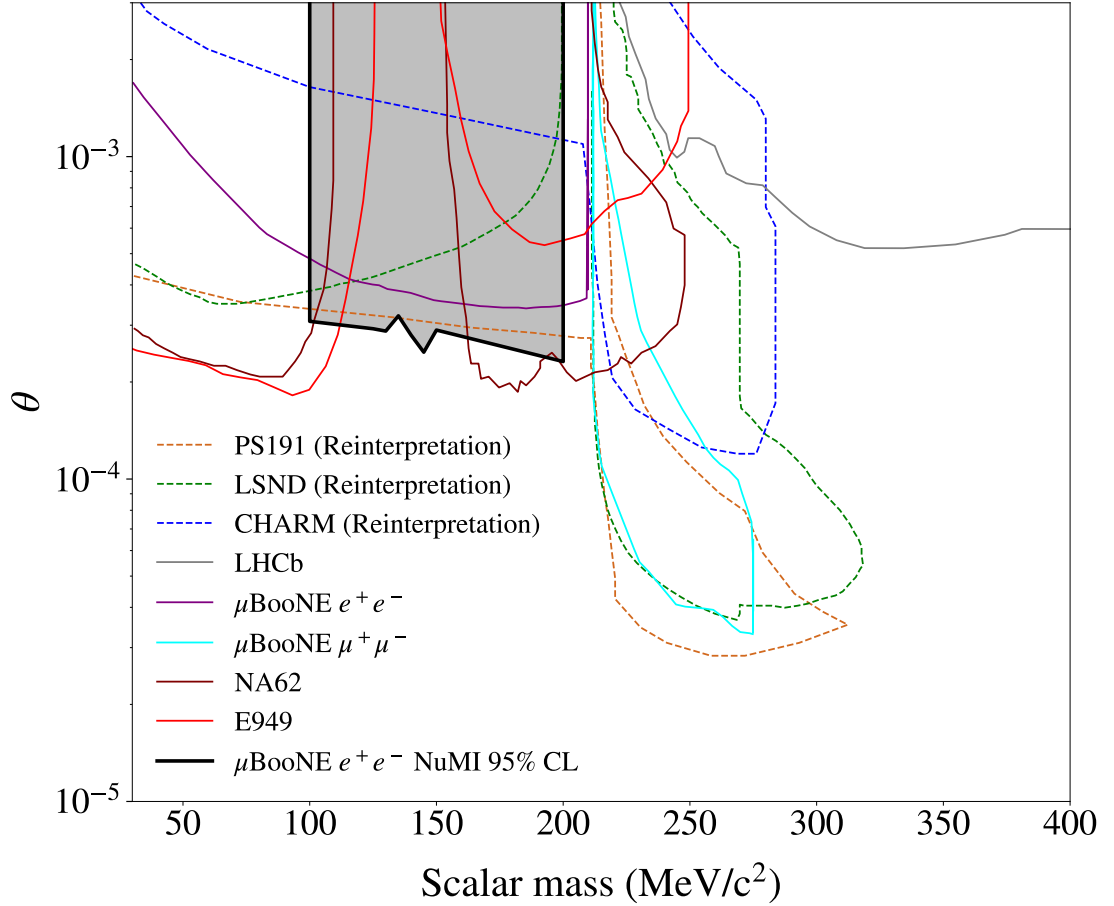


Figure 9.20: The existing experimental limits on the $(\theta - M_S)$ parameter space. The observed limit calculated in the analysis presented in this thesis is shown in solid black. Limits in dashed lines are reinterpretations of the historical experimental results whereas the limits in solid lines are dedicated Higgs portal scalar searches.

dashed lines. We set the world’s best limit in the mass range (125 – 160 MeV), where experiments such as E949 [14] and NA62 [15] are insensitive due to the background from the decay $K^+ \rightarrow \pi^+ \pi^0$. As described in section 1.3, the previous strongest limits in this region are a reinterpretation of PS191 [23] and a dedicated search of Higgs portal scalars decaying into e^+e^- by the MicroBooNE collaboration [11]. This previous analysis used data corresponding to only 1.93×10^{20} POT, and only scalars produced in the NuMI absorber were considered. The analysis presented in this thesis takes into account the signal of Higgs portal scalars produced from the decays of kaons at all locations along the NuMI beamline, including KDAR at the target, KDIF along the decay pipe and KDAR in the NuMI absorber with a total exposure of 7.01×10^{20} POT.

Chapter 10

Conclusions

This thesis presents a search for a new electrically neutral dark sector scalar boson decaying into an e^+e^- pair, which is a dominant decay channel for a scalar with mass in the range $(100 - 200 \text{ MeV})$. The search focuses on scalars originating at three different locations along the NuMI beamline, including scalars produced from KDAR at the target, from KDIF along the decay pipe and from KDAR in the NuMI absorber. In this search, data collected during the MicroBooNE Run 1 and Run 3 periods corresponding to 7.01×10^{20} POT is used.

In total, eight scalar mass values in the range $(100 - 200 \text{ MeV})$ are considered, with five amongst these in the range $(125 - 145 \text{ MeV})$ considered to study the impact of the dominant background, the decay of a neutral pion ($\pi^0 \rightarrow \gamma\gamma$) with $M_{\pi^0} = 134.97 \text{ MeV}$ mimicking the two-shower topology of the scalar decay ($S \rightarrow e^+e^-$) signal. For each scalar mass value, scalars produced from KDAR at the target and in the NuMI absorber are combined due to the similar decay kinematics of these scalars. This leads to the formation of two types of signal: scalars produced from KDAR and scalars produced from KDIF. The identification and reconstruction of these scalars decaying into e^+e^- as well as neutrino interactions in background is performed using the **Pandora** reconstruction framework.

To separate these e^+e^- signals from the neutrino interaction background, a pre-selection is applied, which comprises selection of **Pandora** slices containing isolated

scalar or neutrino interactions that produce shower-like objects (**Pandora** track-shower score less than 0.5) with an interaction vertex contained inside the fiducial volume of the MicroBooNE TPC. An additional CRT Veto pre-selection to veto cosmic ray muons is applied to Run 3 samples. The **Pandora** software framework, like any other software framework, has limitations in reconstruction with two showers in an event often mis-reconstructed as one shower due to overlapping caused by a small opening angle between the showers or due to the difficulty of identifying showers or discriminating the showers from tracks. The latter two are also reasons why there might be zero reconstructed showers in an event. The analysis therefore uses two sets of samples to set the final limit on θ with first containing events where only one shower is reconstructed and the second containing events where two showers are reconstructed. The efficiency of the pre-selection on the signal is $\sim 15 - 25\%$ for the two-shower samples and $\sim 30 - 40\%$ for the one-shower samples.

To further discriminate signal from background in multi-dimensional variable space, a multi-variate machine learning algorithm called a BDT is trained for each of the 8 scalar mass values with four different types of signal in Run 1 and Run 3. The BDTs show an excellent performance in separating these four different types of signal with an average value of AUC around 0.98 across different masses of the scalars produced from KDIF with one-shower and two-shower pre-selections and KDAR with a two-shower pre-selection. For scalars produced from KDAR with a one-shower pre-selection, the performance of the BDT is slightly poorer with an average value of AUC around 0.96 across different scalar mass points.

For the MC background prediction, various sources of systematic uncertainty comprising flux, cross-section, detector and hadron re-interaction uncertainties evaluated using both event-reweighting and sample re-simulation methods are considered. For signal, only flux and detector uncertainties are considered due to the signal being a decay of scalar with non-hadronic final state particles.

The discriminating variables used to set the limit on the mixing parameter for the one-shower and two-shower samples are distributions for shower energy and invariant mass respectively. The optimal BDT scores yielding the best expected median limit for 4 different types of signals corresponding to scalars produced from KDAR and

KDIF with one-shower and two-shower pre-selections in Run 1 and Run 3 samples are calculated from the distributions of the expected median limit against BDT score. No significant excess in these signal-like distributions of discriminating variables is observed for any of the scalar mass point. The dominant source of uncertainty for each scalar mass value is due to statistical uncertainty arising from the limited sample size for the signal and background samples after the optimal BDT score selection.

Using these discriminating variables, a 95% observed upper limit on the scalar-Higgs mixing parameter in the range $(2.3 \times 10^{-4} - 3.2 \times 10^{-4})$ for scalar masses in the range $(100 - 200 \text{ MeV})$ is set. The calculated limit is world's best limit in the mass range $(125 - 160 \text{ MeV})$. The sensitivity of the Higgs portal scalar model can further be improved by including additional NuMI data with an exposure of 1.5×10^{21} POT, which at the time of this work is unprocessed. The fluctuations in the final limit for $M_S = 135 \text{ MeV}$ and $M_S = 145 \text{ MeV}$ could be due to using a poor BDT score selection method (discussed in section 9.2.1) where the minima for each curve is driven by the statistical fluctuations caused by limited background and signal MC sample statistics. The choice of optimising the BDT score using this method is therefore argumentative and a better approach such as curve fitting to the data points could significantly improve the final limit. Targeted improvements to the algorithms such as **Pandora** reconstruction can improve identification and thus separation of the Higgs portal scalar signal from the neutrino background. The analysis methods used in this work could be used in future Higgs portal scalar searches in the LArTPCs of the DUNE and SBN experiments to further improve the sensitivity to the Higgs portal scalar model.

References

- [1] J. Alexander *et al.*, “Dark Sectors 2016 Workshop: Community Report,” 2016.
- [2] V. V. Khoze, “Inflation and dark matter in the Higgs portal of classically scale invariant Standard Model,” *Journal of High Energy Physics*, vol. 2013, no. 11, 2013.
- [3] K. Enqvist, S. Nurmi, T. Tenkanen, and K. Tuominen, “Standard Model with a real singlet scalar and inflation,” *Journal of Cosmology and Astroparticle Physics*, vol. 2014, no. 08, pp. 035–035, 2014.
- [4] E. Hall, R. McGehee, H. Murayama, and B. Suter, “Asymmetric dark matter may not be light,” *Phys. Rev. D*, vol. 106, p. 075008, 2022.
- [5] I. Hierro, S. King, and S. Rigolin, “Higgs portal dark matter and neutrino mass and mixing with a doubly charged scalar,” *Physics Letters B*, vol. 769, pp. 121–128, 2017.
- [6] S. Foroughi-Abari and A. Ritz, “LSND constraints on the Higgs portal,” *Phys. Rev. D*, vol. 102, no. 3, 2020.
- [7] B. Patt and F. Wilczek, “Higgs-field Portal into Hidden Sectors,” 2006.
- [8] G. Lanfranchi, M. Pospelov, and P. Schuster, “The Search for Feebly Interacting Particles,” *Annual Review of Nuclear and Particle Science*, vol. 71, no. 1, pp. 279–313, 2021.
- [9] B. Batell, J. Berger, and A. Ismail, “Probing the Higgs portal at the Fermilab short-baseline neutrino experiments,” *Phys. Rev. D*, vol. 100, no. 11, 2019.

- [10] J. F. Donoghue, J. Gasser, and H. Leutwyler, “The Decay of a Light Higgs Boson,” *Nucl. Phys. B*, vol. 343, pp. 341–368, 1990.
- [11] P. Abratenko *et al.*, “Search for a Higgs Portal Scalar Decaying to Electron-Positron Pairs in the MicroBooNE Detector,” *Phys. Rev. Lett.*, vol. 127, no. 15, p. 151803, 2021.
- [12] P. Abratenko *et al.*, “Search for long-lived heavy neutral leptons and higgs portal scalars decaying in the microboone detector,” *Phys. Rev. D*, vol. 106, p. 092006, Nov 2022.
- [13] P. Abratenko *et al.*, “Search for a Higgs Portal Scalar Decaying to Electron-Positron Pairs in the MicroBooNE Detector,” *Phys. Rev. Lett.*, vol. 127, no. 15, 2021.
- [14] A. V. Artamonov *et al.*, “Study of the decay $K^+ \rightarrow \pi^+ \nu \bar{\nu}$ in the momentum region $140 < P_\pi < 199$ MeV/c,” *Phys. Rev. D*, vol. 79, p. 092004, 2009.
- [15] E. Cortina Gil *et al.*, “Search for a feebly interacting particle X in the decay $K^+ \rightarrow \pi^+ X$,” *J. High Energy Phys.*, vol. 03, p. 058, 2021.
- [16] R. Aaij *et al.*, “Search for long-lived scalar particles in $B^+ \rightarrow K^+ \chi(\mu^+ \mu^-)$ decays,” *Phys. Rev. D*, vol. 95, no. 7, p. 071101, 2017.
- [17] R. Aaij *et al.*, “Search for hidden-sector bosons in $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decays,” *Phys. Rev. Lett.*, vol. 115, no. 16, p. 161802, 2015.
- [18] M. W. Winkler, “Decay and detection of a light scalar boson mixing with the higgs boson,” *Phys. Rev. D*, vol. 99, no. 1, 2019.
- [19] J. D. Clarke, R. Foot, and R. R. Volkas, “Phenomenology of a very light scalar ($100 \text{ MeV} < m_h < 10 \text{ GeV}$) mixing with the SM Higgs,” *J. High Energy Phys.*, vol. 02, p. 123, 2014.
- [20] A. Aguilar *et al.*, “Evidence for neutrino oscillations from the observation of $\bar{\nu}_e$ appearance in a $\bar{\nu}_\mu$ beam,” *Phys. Rev. D*, vol. 64, p. 112007, 2001.
- [21] C. Athanassopoulos *et al.*, “Evidence for $\nu_\mu \rightarrow \nu_e$ oscillations from pion decay in flight neutrinos,” *Phys. Rev. C*, vol. 58, pp. 2489–2511, 1998.

- [22] L. B. Auerbach *et al.*, “Search for $\pi^0 \rightarrow \nu_\mu \bar{\nu}_\mu$ decay in LSND,” *Phys. Rev. Lett.*, vol. 92, p. 091801, 2004.
- [23] D. Gorbunov, I. Krasnov, and S. Suворov, “Constraints on light scalars from PS191 results,” *Phys. Lett. B*, vol. 820, p. 136524, 2021.
- [24] R. Acciarri *et al.*, “Design and Construction of the MicroBooNE Detector,” *JINST*, vol. 12, no. 02, p. P02017, 2017.
- [25] J. Joshi and X. Qian, “Signal Processing in the MicroBooNE LArTPC,” 2015.
- [26] C. Adams *et al.*, “Design and construction of the MicroBooNE Cosmic Ray Tagger system,” *JINST*, vol. 14, no. 04, p. P04004, 2019.
- [27] K. Mistry, *Measurement of the Flux-Averaged Differential Charged-Current Electron Neutrino and Antineutrino Cross Section on Argon with the MicroBooNE Detector*. PhD thesis, University of Manchester, 2021.
- [28] P. Adamson *et al.*, “The NuMI Neutrino Beam,” *Nucl. Instrum. Meth. A*, vol. 806, pp. 279–306, 2016.
- [29] C. Anderson *et al.*, “The ArgoNeuT Detector in the NuMI Low-Energy beam line at Fermilab,” *JINST*, vol. 7, p. P10019, 2012.
- [30] L. Aliaga *et al.*, “Design, Calibration, and Performance of the MINERvA Detector,” *Nucl. Instrum. Meth. A*, vol. 743, pp. 130–159, 2014.
- [31] D. G. Michael *et al.*, “The Magnetized steel and scintillator calorimeters of the MINOS experiment,” *Nucl. Instrum. Meth. A*, vol. 596, pp. 190–228, 2008.
- [32] A. Habig, “The NOvA Experiment,” *Nucl. Phys. B Proc. Suppl.*, vol. 229-232, pp. 460–460, 2012.
- [33] A. Russo, “The PEANUT experiment in the NuMI beam at Fermilab,” *AIP Conf. Proc.*, vol. 1222, no. 1, pp. 131–134, 2010.
- [34] V. Shiltsev, “Fermilab Proton Accelerator Complex Status and Improvement Plans,” *Mod. Phys. Lett. A*, vol. 32, no. 16, p. 1730012, 2017.

- [35] A. Abramov, P. Galkin, N. Galyaev, and V. Garkusha, “Advanced Conceptual Design of the NuMI Hadron Beam Absorber Core,” 2000. Internal NuMI-B-652.
- [36] L. Cremonesi, “Cross-section measurements in the NOvA Near Detector,” 2020. NOvA: Neutrino 2020.
- [37] J. Eldred, V. Lebedev, K. Seiya, and V. Shiltsev, “Beam intensity effects in Fermilab Booster synchrotron,” *Phys. Rev. Accel. Beams*, vol. 24, p. 044001, 2021.
- [38] L. Aliaga *et al.*, “Neutrino flux predictions for the NuMI beam,” *Phys. Rev. D*, vol. 94, no. 9, 2016.
- [39] P. A. Zyla *et al.*, “Review of Particle Physics,” *Progress of Theoretical and Experimental Physics*, vol. 2020, no. 8, 2020. 083C01.
- [40] A. A. Aguilar-Arevalo *et al.*, “First Measurement of Monoenergetic Muon Neutrino Charged Current Interactions,” *Phys. Rev. Lett.*, vol. 120, no. 14, p. 141802, 2018.
- [41] N. V. Mokhov, “Recent MARS15 Developments: Nuclide Inventory, DPA and Gas Production,” in *46th ICFA Advanced Beam Dynamics Workshop on High-Intensity and High-Brightness Hadron Beams*, 2010.
- [42] T. T. Böhlen, F. Cerutti, M. P. W. Chin, A. Fassò, A. Ferrari, P. G. Ortega, A. Mairani, P. R. Sala, G. Smirnov, and V. Vlachoudis, “The FLUKA Code: Developments and Challenges for High Energy and Medical Applications,” *Nucl. Data Sheets*, vol. 120, pp. 211–214, 2014.
- [43] S. Agostinelli *et al.*, “GEANT4: A simulation toolkit,” *Nucl. Instrum. Meth.*, vol. A506, pp. 250–303, 2003. We use `GEANT4 v4_10_1_p03d`.
- [44] J. Allison *et al.*, “Recent developments in Geant4,” *Nucl. Instrum. Meth.*, vol. A835, pp. 186–225, 2016.
- [45] D. Casper, “The nuance neutrino physics simulation, and the future,” *Nuclear Physics B - Proceedings Supplements*, vol. 112, no. 1-3, pp. 161–170, 2002.

- [46] J. A. Formaggio and G. P. Zeller, “From eV to EeV: Neutrino cross sections across energy scales,” *Reviews of Modern Physics*, vol. 84, no. 3, pp. 1307–1341, 2012.
- [47] L. H. Bathe-Peters, “Studies of Single-Transverse Kinematic Variables for Neutrino Interactions on Argon,” 2020.
- [48] I. Ruiz Simo, J. Amaro, M. Barbaro, J. Caballero, G. Megias, and T. Donnelly, “Two-nucleon emission in neutrino and electron scattering from nuclei: The modified convolution approximation,” *Annals of Physics*, vol. 388, pp. 323–349, 2018.
- [49] A. A. Aguilar-Arevalo *et al.*, “First measurement of the muon neutrino charged current quasielastic double differential cross section,” *Phys. Rev. D*, vol. 81, p. 092005, 2010.
- [50] B. Martin, *Nuclear and Particle Physics An Introduction*. Wiley, 2008.
- [51] M. Taketani, “Introduction. Nuclear Forces and Meson Theory,” *Progress of Theoretical Physics Supplement*, vol. 3, pp. 1–12, 1956.
- [52] T. V. Cuyck, “Short-range correlations and meson-exchange currents in neutrino-nucleus scattering,” 2017.
- [53] W. Weise, “Yukawa's Pion, Low-Energy QCD and Nuclear Chiral Dynamics,” *Progress of Theoretical Physics Supplement*, vol. 170, pp. 161–184, 2007.
- [54] W. Van De Pontseele, *Search for Electron Neutrino Anomalies with the Micro-BooNE Detector*. PhD thesis, University of Oxford, 2020.
- [55] P. Abratenko *et al.*, “Measurement of the longitudinal diffusion of ionization electrons in the MicroBooNE detector,” *JINST*, vol. 16, no. 09, p. P09025, 2021.
- [56] P. Abratenko *et al.*, “Measurement of space charge effects in the MicroBooNE LArTPC using cosmic muons,” *JINST*, vol. 15, no. 12, p. P12037, 2020.
- [57] R. Acciarri *et al.*, “A Study of Electron Recombination Using Highly Ionizing Particles in the ArgoNeuT Liquid Argon TPC,” *JINST*, vol. 8, p. P08005, 2013.

- [58] T. Yang, “Calibration of calorimetric measurement in a liquid argon time projection chamber,” *Instruments*, vol. 5, no. 1, p. 2, 2020.
- [59] “A Measurement of the Attenuation of Drifting Electrons in the MicroBooNE LArTPC,” 2017.
- [60] “Measurement of the Electronegative Contaminants and Drift Electron Lifetime in the MicroBooNE Experiment,” 2016.
- [61] D. Garcia-Gamez, P. Green, and A. M. Szelc, “Predicting Transport Effects of Scintillation Light Signals in Large-Scale Liquid Argon Detectors,” *Eur. Phys. J. C*, vol. 81, no. 4, p. 349, 2021.
- [62] E. L. Snider and G. Petrillo, “LArSoft: Toolkit for Simulation, Reconstruction and Analysis of Liquid Argon TPC Neutrino Detectors,” *J. Phys. Conf. Ser.*, vol. 898, no. 4, p. 042057, 2017.
- [63] J. Nieves, J. E. Amaro, and M. Valverde, “Inclusive quasi-elastic neutrino reactions,” *Phys. Rev. C*, vol. 70, p. 055503, 2004. [Erratum: *Phys.Rev.C* 72, 019902 (2005)].
- [64] J. Schwehr, D. Cherdack, and R. Gran, “GENIE implementation of IFIC Valencia model for QE-like 2p2h neutrino-nucleus cross section,” *arXiv: hep-ph*, 2016.
- [65] K. S. Kuzmin, V. V. Lyubushkin, and V. A. Naumov, “Lepton polarization in neutrino nucleon interactions,” *Mod. Phys. Lett. A*, vol. 19, pp. 2815–2829, 2004.
- [66] C. Berger and L. M. Sehgal, “Lepton mass effects in single pion production by neutrinos,” *Phys. Rev. D*, vol. 76, p. 113004, 2007.
- [67] C. Berger and L. M. Sehgal, “Partially conserved axial vector current and coherent pion production by low energy neutrinos,” *Phys. Rev. D*, vol. 79, p. 053003, 2009.
- [68] A. Bodek and U. K. Yang, “Modeling deep inelastic cross-sections in the few GeV region,” *Nucl. Phys. B Proc. Suppl.*, vol. 112, pp. 70–76, 2002.

- [69] S. Dytman, Y. Hayato, R. Raboanary, J. T. Sobczyk, J. Tena Vidal, and N. Vololoniaina, “Comparison of validation methods of simulations for final state interactions in hadron production experiments,” *Phys. Rev. D*, vol. 104, no. 5, p. 053006, 2021.
- [70] C. Andreopoulos *et al.*, “The GENIE Neutrino Monte Carlo Generator,” *Nucl. Instrum. Meth. A*, vol. 614, pp. 87–104, 2010.
- [71] J. Tena-Vidal *et al.*, “Neutrino-nucleon cross-section model tuning in GENIE v3,” *Phys. Rev. D*, vol. 104, no. 7, p. 072009, 2021.
- [72] M. Del Tutto, *First measurements of inclusive muon neutrino charged current differential cross sections on argon at 0.8 GeV average neutrino energy with the MicroBooNE detector*. PhD thesis, University of Oxford, 2019.
- [73] C. Adams *et al.*, “Ionization electron signal processing in single phase LArTPCs. Part II. Data/simulation comparison and performance in MicroBooNE,” *JINST*, vol. 13, no. 07, pp. P07007–P07007, 2018.
- [74] C. Adams *et al.*, “Ionization electron signal processing in single phase LArTPCs. Part I. Algorithm Description and quantitative evaluation with MicroBooNE simulation,” *JINST*, vol. 13, no. 07, pp. P07006–P07006, 2018.
- [75] R. Acciarri *et al.*, “Noise Characterization and Filtering in the MicroBooNE Liquid Argon TPC,” *JINST*, vol. 12, no. 08, pp. P08003–P08003, 2017.
- [76] W. Tang, X. Li, X. Qian, H. Wei, and C. Zhang, “Data Unfolding with Wiener-SVD Method,” *JINST*, vol. 12, no. 10, pp. P10002–P10002, 2017.
- [77] B. Baller, “Liquid argon TPC signal formation, signal processing and reconstruction techniques,” *JINST*, vol. 12, no. 07, pp. P07010–P07010, 2017.
- [78] R. Acciarri *et al.*, “The Pandora multi-algorithm approach to automated pattern recognition of cosmic-ray muon and neutrino events in the MicroBooNE detector,” *Eur. Phys. J.*, vol. C78, no. 1, p. 82, 2018.
- [79] C. Adams *et al.*, “Calibration of the charge and energy loss per unit length of the MicroBooNE liquid argon time projection chamber using muons and protons,” *JINST*, vol. 15, no. 03, p. P03022, 2020.

- [80] “PeLEE Tech-Note Review Material, docdb 32933,” 2020.
- [81] D. Egana-Ugrinovic, S. Homiller, and P. Meade, “Light Scalars and the Koto Anomaly,” *Phys. Rev. Lett.*, vol. 124, no. 19, p. 191801, 2020.
- [82] S. Shinohara, “Search for the rare decay $K_L \rightarrow \pi^0 \nu \bar{\nu}$ at J-PARC KOTO experiment,” *Journal of Physics: Conference Series*, vol. 1526, p. 012002, 2020.
- [83] J. K. Ahn *et al.*, “Study of the $K_L \rightarrow \pi^0 \nu \bar{\nu}$ Decay at the J-PARC KOTO Experiment,” *Phys. Rev. Lett.*, vol. 126, p. 121801, 2021.
- [84] Y. Coadou, “Boosted Decision Trees and Applications,” *EPJ Web Conf.*, vol. 55, p. 02004, 2013.
- [85] M. Bonesini, A. Marchionni, F. Pietropaolo, and T. T. de Fatis, “On particle production for high energy neutrino beams,” *The European Physical Journal C*, vol. 20, no. 1, pp. 13–27, 2001.
- [86] P. Abratenko *et al.*, “New Theory-driven GENIE Tune for MicroBooNE,” *submitted to Phys. Rev. D*, 2021.
- [87] J. Calcutt, C. Thorpe, K. Mahn, and L. Fields, “Geant4Reweight: a framework for evaluating and propagating hadronic interaction uncertainties in Geant4,” *JINST*, vol. 16, no. 08, p. P08042, 2021.
- [88] P. Abratenko *et al.*, “Novel approach for evaluating detector-related uncertainties in a LArTPC using MicroBooNE data,” *Eur. Phys. J. C*, vol. 82, no. 5, p. 454, 2022.
- [89] N. Ishida *et al.*, “Attenuation length measurements of scintillation light in liquid rare gases and their mixtures using an improved reflection suppresser,” *Nucl. Instrum. Meth. A*, vol. 384, no. 2, pp. 380–386, 1997.
- [90] E. Grace *et al.*, “Index of refraction, Rayleigh scattering length, and Sellmeier coefficients in solid and liquid argon and xenon,” *Nucl. Instrum. Meth. A*, vol. 867, pp. 204–208, 2017.
- [91] A. Neumeier *et al.*, “Attenuation measurements of vacuum ultraviolet light in liquid argon revisited,” *Nucl. Instrum. Meth. A*, vol. 800, pp. 70–81, 2015.

- [92] J. Calvo *et al.*, “Measurement of the attenuation length of argon scintillation light in the ArDM LAr TPC,” *Astroparticle Physics*, vol. 97, pp. 186–196, 2018.
- [93] J. Neyman and E. S. Pearson, “On the Problem of the Most Efficient Tests of Statistical Hypotheses,” *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, vol. 231, pp. 289–337, 1933.
- [94] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, “Asymptotic formulae for likelihood-based tests of new physics,” *Eur. Phys. J. C*, vol. 71, p. 1554, 2011. [Erratum: *Eur.Phys.J.C* 73, 2501 (2013)].
- [95] ATLAS, CMS, LHC Higgs Combination Group, “Procedure for the LHC Higgs boson search combination in Summer 2011,” *CMS-NOTE-2011-005, ATL-PHYS-PUB-2011*, 2011.
- [96] M. Feickert, L. Heinrich, and G. Stark, “pyhf: pure-python implementation of histfactory with tensors and automatic differentiation,” 2022.