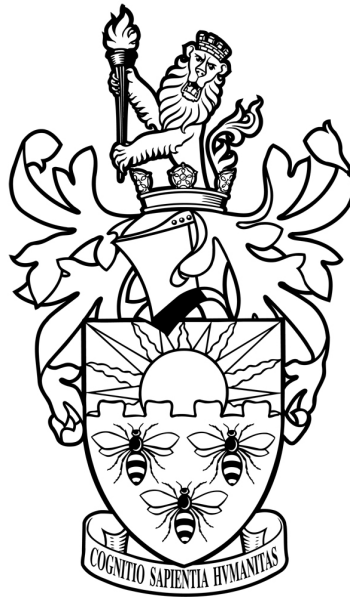


# SEARCHES FOR HEAVY NEUTRAL LEPTON DECAYS IN THE MICROBOONE DETECTOR

SALVATORE DAVIDE PORZIO

DEPARTMENT OF PHYSICS AND ASTRONOMY  
IN THE SCHOOL OF NATURAL SCIENCES



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

2019

# Contents

	Page
<b>1 Introduction</b>	<b>11</b>
<b>2 Heavy Neutral Leptons</b>	<b>14</b>
2.1 Right-Handed Neutral States . . . . .	15
2.2 Production . . . . .	17
2.3 Decay . . . . .	21
2.4 Previous Searches . . . . .	26
2.4.1 Peak Searches . . . . .	28
2.4.2 Decay Searches . . . . .	30
<b>3 The MicroBooNE Detector and the Beamline</b>	<b>32</b>
3.1 Energy Loss in Liquid Argon . . . . .	33
3.1.1 Tracks: Muons, Pions and Protons . . . . .	33
3.1.2 Showers: Electrons and Photons . . . . .	35
3.1.3 Scintillation Light . . . . .	36
3.2 Collecting and Recording the Signal . . . . .	37
3.2.1 The MicroBooNE Time Projection Chamber . . . . .	37
3.2.2 Cryogenics and Argon Purity . . . . .	40
3.2.3 TPC Readout and Signal Processing . . . . .	41
3.2.4 Light Collection System . . . . .	42
3.2.5 Cosmic Ray Tagger . . . . .	45
3.2.6 Data Acquisition . . . . .	46
3.3 Physics and Detector Effects . . . . .	48

## CONTENTS

---

3.3.1	Recombination . . . . .	49
3.3.2	Attenuation and Argon Purity . . . . .	50
3.3.3	Diffusion . . . . .	54
3.3.4	Space Charge . . . . .	56
3.4	The Booster Neutrino Beamline . . . . .	58
3.4.1	Proton Beam . . . . .	59
3.4.2	Beam Target and Focusing Horn . . . . .	61
3.4.3	Beam Composition . . . . .	63
<b>4</b>	<b>Triggers</b>	<b>65</b>
4.1	MicroBooNE Triggers . . . . .	66
4.1.1	Hardware Triggers . . . . .	66
4.1.2	Software Triggers . . . . .	67
4.2	HNL Trigger . . . . .	69
4.2.1	HNL Trigger Performance in Monte Carlo . . . . .	73
4.2.2	HNL Trigger Performance in Data . . . . .	75
4.3	Data Streams . . . . .	77
<b>5</b>	<b>Simulation and Data</b>	<b>79</b>
5.1	Software Framework . . . . .	80
5.2	Generation . . . . .	82
5.2.1	HNL Generator . . . . .	82
5.2.2	HNL Flux and Decay Spectrum . . . . .	85
5.2.3	HNL Cosmic Overlay . . . . .	91
5.2.4	HNL Sample Normalisation . . . . .	94
5.2.5	Other Monte Carlo Samples: BNB Neutrinos . . . . .	95
5.2.6	Other Monte Carlo Samples: Dirt . . . . .	97
5.3	Particle Propagation . . . . .	97
5.4	Detector Effects . . . . .	98
5.5	Data . . . . .	99
5.5.1	Data Quality . . . . .	100

## CONTENTS

---

5.5.2	Data Calibration . . . . .	104
5.6	Object Reconstruction and Identification . . . . .	109
5.6.1	Noise Filtering . . . . .	110
5.6.2	Deconvolution . . . . .	111
5.6.3	Hit Finding . . . . .	114
5.6.4	Clustering and 3D Object Reconstruction . . . . .	115
5.6.5	Optical Reconstruction . . . . .	118
5.7	Datasets . . . . .	119
5.7.1	Test Samples . . . . .	119
5.7.2	Control Samples . . . . .	122
5.7.3	Training Samples . . . . .	124
<b>6</b>	<b>Event Selection</b>	<b>128</b>
6.1	Signal Definition . . . . .	128
6.1.1	Candidate Identification . . . . .	129
6.1.2	High-level Variables . . . . .	135
6.2	Pre-selections . . . . .	140
6.2.1	Fiducial Volume . . . . .	141
6.2.2	Vertex-Track Distance . . . . .	143
6.2.3	Minimum Number of Hits . . . . .	144
6.2.4	Flash PE . . . . .	147
6.2.5	Vertex-Flash Distance . . . . .	150
6.2.6	Track Containment . . . . .	150
6.3	Kinematic Selections . . . . .	154
6.4	Pre-selection Efficiency Summary . . . . .	157
<b>7</b>	<b>Boosted Decision Trees</b>	<b>163</b>
7.1	BDT Definition . . . . .	163
7.2	Application to HNL Samples . . . . .	166
<b>8</b>	<b>Control Sample and Systematic Uncertainties</b>	<b>170</b>
8.1	Control Sample . . . . .	170



## CONTENTS

---

8.2	Uncertainty Determination . . . . .	173
8.2.1	Uni-sim Approach . . . . .	173
8.2.2	Multi-sim Approach . . . . .	174
8.3	Flux and POT Count Systematics . . . . .	175
8.4	Trigger Systematics . . . . .	176
8.5	Detector Systematics . . . . .	178
8.5.1	Calibration . . . . .	182
8.5.2	Space Charge Effect . . . . .	182
8.5.3	Dynamically Induced Charge . . . . .	182
8.5.4	Detector Systematics Summary . . . . .	183
8.6	Systematics Summary . . . . .	187
<b>9</b>	<b>Results</b>	<b>188</b>
9.1	Limit Extraction Procedure . . . . .	188
9.2	Mass-mixing Matrix Element Limits . . . . .	195
<b>10</b>	<b>Conclusions</b>	<b>198</b>
	<b>Glossary</b>	<b>201</b>
	<b>References</b>	<b>205</b>

# Abstract

This thesis presents new results on upper limits for the search of Heavy Neutral Leptons (HNL) in the mass range 260-385 MeV decaying to  $\mu\pi$  pairs using data collected with the MicroBooNE liquid argon time projection chamber operating at Fermilab. The measurement is the first of its kind ever performed in a liquid argon time projection chamber and it uses data collected in 2017 and 2018 with the Fermilab Booster Neutrino Beam, which has an average muon neutrino energy of 800 MeV, using an exposure corresponding to  $1.97 \times 10^{20}$  protons-on-target and a subset of data passing a specific trigger configured to look for possible delayed decays, occurring after the neutrino spill. The results show limits on Majorana Heavy Neutral Leptons in the mass region 260-385 MeV for  $|U_{\mu 4}|^2$  in the range  $[7 \times 10^{-8}, 5 \times 10^{-7}]$ , at the 90% confidence level.

# Declaration

I 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

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

First and foremost, I would like to thank my PhD supervisor, Stefan Söldner-Rembold for his constant support. No matter what time zone he would be in, I could always count on his swift advice a minute after I would ask, on any day of the week, at any hour of the day. He has been a great mentor and I am especially grateful for his guidance and his patience in reading through the drafts of this work. I would also like to thank my co-supervisor Andrzej Szalc for always being available for counsel, and for providing me with helpful advice on how to improve my research and my academic career.

I certainly would not have been able to bring this work to conclusion without the advice and support of Yun-Tse Tsai. She is a role model for what a working group convener should be. I hope one day to be able to match her curiosity, perseverance and kindness, and if not that, at least to be able to play badminton half as good as she does.

Special thanks must go to the people that have read through my thesis and have provided me with invaluable feedback, especially Andrew Furmanski and Nicola McConkey. I would also like to thank Pawel Guzowski for his help and support and Owen Goodwin for his thorough and excellent work on the [HNL](#) flux simulation and the related systematic uncertainties, which has provided the groundwork for the generation of the simulated datasets used in this research work.

The discovery of the espresso machine on the third floor of Wilson Hall, during my stay at Fermilab, has been a blessing for this research. Not because of the readily available stream of caffeine just a couple of floors below my office (that did help, though) but because of the wonderful opportunity it gave me to steal my Fermilab co-workers away from their desks and badger them with interesting problems to which they provided interesting solutions. Properly mentioning all of them would produce a document not too dissimilar in size to the average MicroBooNE publication's author list, but I am especially grateful to Colton Hill and Adam Lister, for making my stay in the United States not only a wonderful intellectual experience but also a great adventure.

I am honoured to have been part of the Manchester High-Energy Physics group

and I would like to thank the people I have shared my time at the department with. Both the original cast, Chris Burr, Toby Nonnenmacher, and Jacob Rawling, to mention a few, and the new generation, Owen Goodwin, Vincent Basque, and Krishan Mistry, who have made Manchester such an enjoyable and productive place to work.

I would like to thank my parents for tirelessly helping me pursue my ambition, for always being there for me, and for at least having *tried* to wrap their heads around the time zone difference during my stay in the US. Although I did not realise it at the time, your 4am calls did brighten my successive day. Thank you for continuing being a daily part of my life despite the distance.

Finally, I need to thank Kasia for her unconditional support, for caring about this work as much as I do and for reminding that even scientists occasionally need to get out, stretch their legs and get for a bit in the sunlight. Much of this work would not have been possible without her constant encouragement and understanding.

# Introduction

This work is the first search for exotic physics with the MicroBooNE detector. Searches for Heavy Neutral Lepton (HNL) are performed in a liquid argon time projection chamber, looking for the delayed decays of HNLs to  $\mu\pi$  pairs. Sensitive searches for MeV-scale HNLs have started in 1981 with the PS191 experiment and further improvements have become possible in recent times on account of the increasing capabilities of modern detectors. Among them, liquid argon time projection chambers offer promising opportunities due to their high spatial resolution which enables detection and accurate reconstruction of key HNL decay kinematic features.

Right-handed HNLs are invoked in many models to provide a mass generation mechanism to the neutrino. The neutrino masses and the lack of a mechanism in the Standard Model (SM) that can describe them are one of the many tantalising clues that hints at new physics phenomena and could lead the way for physics beyond the Standard Model.

The largest currently operating liquid argon time projection chamber neutrino detector, MicroBooNE, has collected neutrino data since 2015. Using a novel online trigger, specifically designed to detect HNLs with a longer time-of-flight, MicroBooNE can boost its sensitivity to HNL decays by operating in a timing window which is free from SM neutrino interaction background.

This thesis describes the search for HNL decays, looking in particular at the  $\mu\pi$  decay channel, in a delayed timing window with no SM neutrino interaction background, covered by the specifically-designed online trigger. Section 2 provides an introduction to HNLs, their production and decay channels and the previous experimental searches

---

performed in the parameter space covered by MicroBooNE. Following this, the general physics properties and working principle of a liquid argon time projection chamber are outlined in [section 3](#). The same section also contains a description of the neutrino beam servicing the detector, with emphasis on the physical properties that affect the HNL production and flux features. [Section 4](#) describes the generic operation of a trigger in MicroBooNE and the development, implementation and monitoring of the HNL trigger, which plays a key role in this analysis. The analysis is based on the search of an excess over a distribution of reconstructed quantities from a background dataset. The modelling and simulation of the signal and the reconstruction of both data and simulated quantities is covered in [section 5](#). The section also contains a description of the software tools used and developed for this analysis and the final datasets that constitute the input for the main analysis framework. In [section 6](#) the analysis framework is presented. The section describes the techniques and methods to identify a HNL candidate and to discriminate it from background. [Section 7](#) describes the use of a Boosted Decision Tree for further discrimination of signal from background. [Section 8](#) presents the systematic uncertainties associated with the selection of HNL candidates and the results from a control sample used to validate the selection algorithm on data. In [section 9](#) are presented the final results and the methods used to obtain the exclusion region contours. A discussion of future outlooks is presented in the conclusion.

The author has been responsible for the development of the whole analysis chain from its inception to data simulation, trigger implementation, data calibration, analysis strategy development and limit extraction. This involved developing, testing and validating a new Monte Carlo generator for the simulation of HNL decays (HNLGen) in MicroBooNE which is now part of the standard modules of the MicroBooNE liquid argon software toolkit ([LArSoft](#)). A key feature of this analysis is the use of a novel trigger type, tuned specifically for the detection of delayed HNLs. The author has been responsible for the optimisation of the trigger configuration, its implementation and the execution of stability and performance tests. This work is one of the first analyses to make use of a new Monte Carlo generation technique called “Cosmic Overlay”, developed by the MicroBooNE collaboration, which uses real data events for the underlying noise and background of simulated samples. As one of the first collaborators to use this technique, the author has tested methods for the extraction of systematic uncertainties associated with this new type of Monte Carlo sample. Since



---

with these new methods, real data is used in conjunction with simulated events for the production of Monte Carlo samples, a correct calibration of the detector for the period of time covered by the underlying sample is essential. The author has also been the first MicroBooNE collaborator to use the data from Run 3 for a physics analysis, which involved performing a calibration of the detector, using techniques established by the calibration group. This involved deriving a calibration correction map of the MicroBooNE detector for its Run 3 period which is used for both the production of “Cosmic Overlay” samples and the determination of uncertainties related to the detector physics effect. The author has written and developed an analysis module ([HELEPHANTE](#)) used for the identification of [HNL](#) candidates in data and part of the analysis chain used for the discrimination of [HNL](#) decays from cosmic ray background which produces the limits shown in the final section of this thesis. The author has also performed an updated calculation of the uncertainties on the atmospheric muon-neutrino flux which arise from cosmic-ray primaries, using recent measurements of cosmic-ray primaries, which is not included in this thesis. These results were presented in a paper that was published by Physical Review D [1].

# Heavy Neutral Leptons

The Standard Model (SM) of fundamental particles is an incomplete theory. Perhaps the most obvious example is the evidence in support of the phenomenon of neutrino oscillation. Measured conclusively by the Super-Kamiokande [2] and SNO [3] collaborations, neutrino flavour oscillations can only be explained by accounting for non-null masses of at least two of the three neutrino types included in the SM.

In its current state, the canonical SM does not include any mechanism to provide masses to these neutrino states. The fermionic content of the electro-weak sector of the SM, shown in table 2.1, lists a left-handed lepton doublet  $\mathbb{L}_L^\alpha$  for the charged lepton and associated neutrino, and a right-handed singlet for the charged lepton state  $\alpha_R$ , but no right-handed chiral neutrino field. The absence of such a state means that no Dirac mass term can be built via the Yukawa coupling of the Higgs to the opposite chirality fields.

To address this deficiency of the SM with the introduction of the smallest number of particles that can fit in the current framework without adding any new physical principle (e.g., supersymmetry, extra-dimensions), the gap is filled with a neutrino right-handed singlet state [4]. This chapter will describe the consequence of adding a right-handed neutral lepton to the SM with emphasis on the experimental signatures that could be observed at MicroBooNE and the region in parameter space that can be explored.

## 2.1. RIGHT-HANDED NEUTRAL STATES

	Left-handed doublets	Right-handed singlets	
<b>Leptons</b>	$\mathbb{L}_L^\alpha = \begin{pmatrix} \nu_{\alpha L} \\ \alpha_L \end{pmatrix}$		$\alpha_R$
	$I, Y, Q = [\frac{1}{2}, -1, (\frac{0}{-1})]$		$I, Y, Q = [0, -1, -1]$
<b>Quarks</b>	$\mathbb{Q}_L^i = \begin{pmatrix} u_L^i \\ d_L^i \end{pmatrix}$	$u_R^i$	$d_R^i$
	$I, Y, Q = [\frac{1}{2}, \frac{1}{3}, (\frac{2/3}{-1/3})]$	$I, Y, Q = [0, 4/3, 2/3]$	$I, Y, Q = [0, -2/3, -1/3]$

**Table 2.1:** Fermionic content of the electroweak sector of the SM, divided across leptons and quarks, and left-handed and right-handed chiral fields. For each field the eigenvalues of the weak isospin ( $I$ ), hypercharge ( $Y$ ) and electric charge ( $Q$ ) are shown. The index  $\alpha$  runs over the three lepton families ( $\alpha = e, \mu, \tau$ ), while the index  $i$  runs over the three quark generations ( $i = 1, 2, 3$ ). No right-handed chiral component is available for the neutrino field  $\nu$ .

## 2.1 Right-Handed Neutral States

The existence of right-handed neutrino states allows for the construction of Dirac mass terms, following the same prescription for all the other SM particles. After electroweak symmetry breaking, a binomial mass term  $m_D(\overline{\nu}_L \nu_R + \text{h.c.})$  emerges, where  $m_D = Y\nu/\sqrt{2}$  is the Dirac mass resulting from the Yukawa coupling ( $Y$ ) of the Higgs doublet to the left and right handed components of the neutrino field and  $\nu$  is the Higgs vacuum expectation value.

Unlike for any other SM particles, an additional mass mechanism becomes available for neutrinos. While Dirac mass terms require both particle helicity states  $\psi_L + \psi_R = \psi$  to establish coupling with the Higgs, a massive neutrino can also be described using exclusively a single left-handed field under condition that:

$$\psi_R = C\overline{\psi}_L, \quad (2.1)$$

where  $C$  is the charge conjugation operator and  $\overline{\psi}$  denotes the adjoint spinor of  $\psi$ .

The  $C\overline{\psi}_L$  state is a right-handed state. Using the condition from [equation \(2.1\)](#) enables us to construct the particle field as:

$$\psi = \psi_L + \psi_R = \psi_L + C\overline{\psi}_L = \psi_L + \psi_L^C, \quad (2.2)$$

where  $\psi_L^C$  is used to indicate the charge-conjugated field. The function  $\psi$  defined from [equation \(2.2\)](#) is a *Majorana* field. Since the charge conjugate of the Majorana particle field written in this form is equivalent to the field itself ( $\psi^C = \psi$ ), the Majorana particle is its own anti-particle. A Majorana field constructed in this way patently violates charge conservation, and it is forbidden for any particle of the **SM** with the exception of neutrinos which allow the construction of Majorana states thanks to their null charges (see [table 2.1](#)). The formulation of a Majorana state allows for the construction of a mass term  $m_M/2(\overline{\nu}_R^C \nu_R)$ , where the factor of  $1/2$  accounts for the double counting caused by the hermitian conjugate terms being identical [\[5\]](#).

While allowing for a Majorana mass, the neutral nature of the right-handed state prevents it from interacting via any gauge boson except the Higgs, making detection potentially challenging. By construction all the **SM** charges of the right-handed neutrino state have to be zero. It is an electrically neutral (zero electric charge) weak isosinglet (zero weak isospin) state which consequently has zero hypercharge. Despite this, the right-handed neutrino can interact with the rest of the **SM** particles via the Yukawa coupling term with the lepton doublet and Higgs.

The Higgs Yukawa coupling offers a window of opportunity for the detection of the right-handed neutrino states. Weaker-than-weak interactions arise from mass mixing which allows them to be produced, and to decay, via **SM** interactions, albeit with a rate suppressed by the mass-mixing matrix element [\[6\]](#). Because of its extremely weak interaction with the **SM** particles the right-handed neutrino states are also referred to as sterile neutrinos. Since their mass is larger than the active neutrino masses the term *Heavy Neutral Lepton* (**HNL**) is also common, which will be used in this thesis.

We indicate the **HNL** with  $N$  and define it in terms of its parameters which are relevant to its experimental detection: its mass  $m_N$  and the additional elements of the extended Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix  $|U_{\alpha 4}|^2$  ( $\alpha = e, \mu, \tau$ ) which governs the mass-mixing. For a single additional **HNL** we write the flavour eigenstates of the left-handed neutrino as a linear combination of the mass eigenstates, including

the newly-added  $N$ , as:

$$\nu_\alpha = \sum_i U_{\alpha i} \nu_i + U_{\alpha 4} N, \quad (2.3)$$

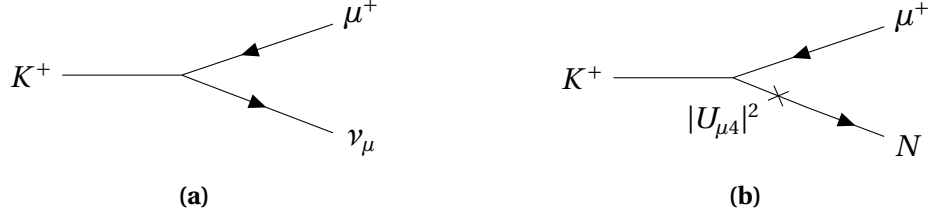
where  $U_{\alpha i}$  is the element of the PMNS matrix involving the mass eigenstate  $i$  ( $i = 1, 2, 3$ ) and lepton flavour  $\alpha$  ( $\alpha = e, \mu, \tau$ ). The index 4 is reserved for the HNL, the extension to a multi-generation case is done by summing the extra components of the additional HNLs.

A priori the number of HNLs is unconstrained and their mass and interaction scales can span many orders of magnitude. Their values can only be determined experimentally. While many models explore different combinations of these parameters (for a review, see Ref. [7]), in this thesis we will explore the existence of HNLs in a minimal way, assuming only the existence of a single HNL with a mass  $\mathcal{O}(100 \text{ MeV})$  which can be produced and detected in the MicroBooNE detector [6]. In the following sections we will describe the HNL production and decay mechanisms which enable, respectively, two different direct search methods: peak searches and beam-dumps (or decay) searches.

## 2.2 Production

From a phenomenological standpoint HNLs can be produced in place of neutrinos in any final state that is kinematically allowed and with a rate proportional to  $|U_{\alpha 4}|^2$ . This allows HNLs to be produced by the decay of mesons, with the most experimentally promising channel being the two-body decays of electrically charged mesons into a lepton and a neutrino (or HNL). Figure 2.1 shows the diagrams for  $K \rightarrow \mu^+ \nu_\mu (N)$ , although the same diagram can be built for any kinematically allowed combination of meson, charge and lepton flavour.

Assuming fixed parent meson momentum, two peaks are expected in the final state lepton energy spectrum: one for the active neutrinos and one ( $n$  in the  $n$ -generational non-degenerate case) for the HNL. The rate of HNLs produced scales with  $|U_{\alpha 4}|^2$  and in the absence of a signal, limits on  $U_{\alpha 4}$  can be placed as a function of HNL mass. Kaon factories, such as NA62 [8], produce large quantities of mesons with sufficient statistics to make these measurements possible (other experiments are described



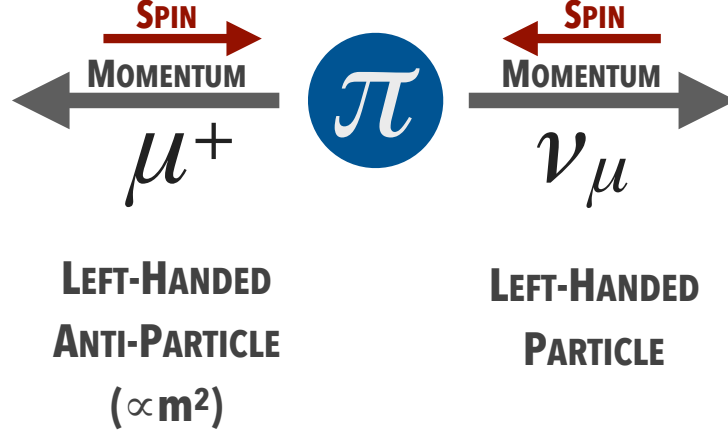
**Figure 2.1:** Diagrams for the production of active neutrinos (a) and of HNLs (b) mediated by the mixing element  $|U_{\mu 4}|^2$ .

in section 2.4). The energy of the initial proton beam which produces the charged mesons is directly responsible for the kinematic boundary on the highest value of HNL mass that can be produced and thus probed. The two-body leptonic decay of a meson  $M$  can create at most a HNL with a mass  $m_N = m_M - m_e \approx m_M$ , if assuming a non-zero value of  $|U_{e4}|^2$ . In kaon decay experiments, limits can thus be placed up to masses of  $\approx 490$  MeV. Higher energy proton beams can produce  $D$  mesons and increase the kinematic boundary to higher masses [9]. The Dirac or Majorana nature of the HNL cannot be probed in production mechanisms since it does not affect the production kinematics.

The charged lepton in the final state determines the mixing angle measured (e.g.,  $|U_{e4}|^2$  for  $K \rightarrow Ne$ , or  $|U_{\mu 4}|^2$  for  $K \rightarrow N\mu$ ). The large difference between the  $K \rightarrow \nu e$  and  $K \rightarrow \nu\mu$  decay rates due to helicity suppression, of the order of  $10^{-4}$ , suggests a similar disparity in the sensitivity that can be achieved on the  $|U_{e4}|^2$  and  $|U_{\mu 4}|^2$  matrix elements. Helicity un-suppression due to the larger HNL mass, however, causes the two leptonic decays to have a comparable rate.

### Helicity Un-suppression

Charged mesons decaying in flight are in a spin-0 state and their decay to a neutrino and charged lepton pair proceeds via the weak interaction. In the rest frame of the meson (shown in the diagram of figure 2.2) linear and angular momentum must be conserved, summing up to zero, and as a result the two quantities propagate in opposite direction with equal magnitude. This implies that the helicity of the two particles emitted in the decay must be identical. For the case of a  $\pi^+ \rightarrow \nu_\mu \mu^+$  decay, for example, the  $\nu_\mu$  must necessarily propagate in a left-helicity state, due to its low mass.



**Figure 2.2:** Helicity suppression mechanism for a  $\pi \rightarrow \nu\mu$  decay.

Since both particles share the same helicity, this implies that the  $\mu^+$  anti-particle must also propagate in a left-helicity state.

The weak interaction contribution to a left-helicity anti-particle introduces a  $m_l^2$  term in the decay width, where  $m_l$  is the mass of the lepton, so that the decay to a muon becomes much more likely than the decay to an electron, which has a significantly smaller mass. The helicity suppression mechanism reduces the decay rate to electrons by a factor of  $\sim (m_e/m_\mu)^2 \approx 10^{-4}$  compared to the decay to muons [10].

For HNLs, the helicity suppression is minimised by the larger mass of the HNL produced with the final-state lepton. The un-suppression mechanism causes the HNL production channel to obtain an enhancement factor, with respect to the equivalent SM neutrino production channels, which in the case of  $\pi \rightarrow \nu e$  can be as large as  $10^5$ . We write the kinematic enhancement factor as [11, 12]:

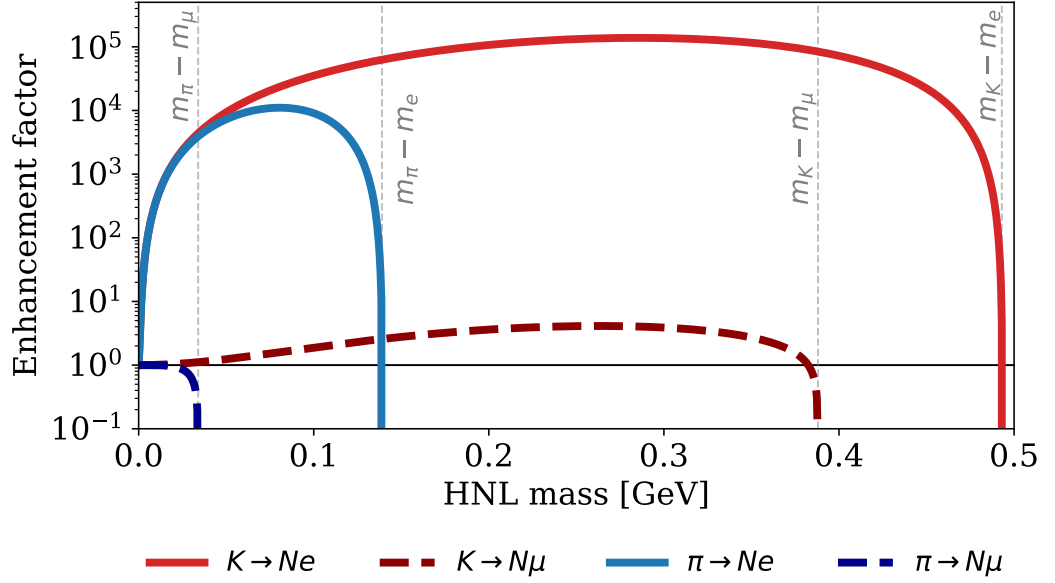
$$K_{l,M,N} = \frac{\rho(\delta_M^l, \delta_M^N)}{\delta_M^l(1 - \delta_M^l)^2}, \quad (2.4)$$

where  $\delta_b^a = (m_a/m_b)^2$  represents the squared mass ratio of the particles  $a$  and  $b$ ,  $l$  is the flavour of the final state lepton,  $N$  indicates the HNL, and  $M$  the parent meson that decays [12]. The kinematic factor  $\rho$  consists of a term  $\lambda$  proportional to the two-

## 2.2. PRODUCTION

body phase space factor and a term  $\mathcal{F}_M$  proportional to the matrix element written as follows:

$$\begin{aligned}\rho(a, b) &= \mathcal{F}_M(a, b) \sqrt{\lambda(1, a, b)}, \\ \mathcal{F}_M(a, b) &= a + b - (a - b)^2, \\ \lambda(x, y, z) &= x^2 + y^2 + z^2 - 2(xy + yz + xz).\end{aligned}\tag{2.5}$$



**Figure 2.3:** Kinematic enhancement factor of the HNL production rates from meson decays with respect to the analogous SM neutrino production channels, as a function of HNL mass, given by equation (2.4). Red lines indicate enhancement factors associated with production from kaons, blue lines are used for production from pions. Solid lines are used for production channels governed by the  $|U_{e4}|^2$  mixing element while dashed lines are used for  $|U_{\mu 4}|^2$  governed production channels. The vertical dashed gray lines indicate the kinematic limits where the production channel is not kinematically accessible, due to the mass of the HNL becoming larger than the available energy.

Figure 2.3 shows the kinematic enhancement factor from equation (2.4) for the kaon and pion decays to a HNL. The effect of the kinematic factor  $K_{l,M,N}$  is two-fold. It suppresses HNL production when its mass reaches the kinematic limit for a specific channel to be feasible, thus providing a threshold effect and it takes into



account the helicity un-suppression of channels that would normally be suppressed in a conventional neutrino beam [6].

Unlike the lighter sterile states with masses in the eV range, often invoked to explain the short-baseline oscillation anomalies observed by the LSND [13] and Mini-BooNE [14–16] collaborations, the  $\mathcal{O}(100 \text{ MeV})$  HNLs produced by meson decay do not oscillate, since the mass is sufficiently large to prevent oscillatory effects with the SM neutrinos due to coherence loss (see Ref. [17]). Such HNLs can instead propagate for long distances before decaying due to mass mixing.

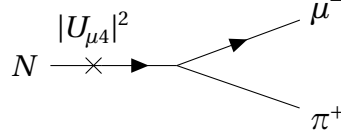
## 2.3 Decay

In addition to kaon factories, meson decays occur also abundantly in neutrino accelerator experiments where focused neutrino beams are produced by letting mesons collimated by strong magnetic fields decay to a lepton and neutrino pair. The neutrino resulting from the decay gets a Lorentz boost in the lab frame and thus inherits the momentum direction of the parent meson, travelling towards a down-stream detector which records its possible interaction with the detector medium. The neutrino beam generation process is described in more details in [section 3.4](#).

A consequence of replacing the SM neutrino production diagram ([figure 2.1a](#)) with the HNL production diagram ([figure 2.1b](#)) is that an analogous flux of HNLs can be produced that would be equally directed towards the neutrino detector. Due to the larger mass, kinematic effects cause the flux to be more collimated than the SM neutrino flux but to first approximation we will consider in this section the HNL flux to be proportional to  $|U_{l4}|^2 K_{l,M,N}$  (details of the flux simulation, which takes into account the full kinematic effects, will be discussed in [section 5.2](#)).

HNLs are unstable particles. Albeit possibly long-lived, they eventually decay via the same mass-mixing Yukawa coupling responsible for their production which is also proportional to  $|U_{l4}|^2$ . [Figure 2.4](#) shows a decay diagram for a  $|U_{\mu 4}|^2$  mediated  $N \rightarrow \mu\pi$  decay channel.

Provided the combined  $|U_{l4}|^2$  mass-mixing elements are small enough, HNLs would have a sufficiently large half-life to travel to detectors located down-stream of the



**Figure 2.4:** Diagram for the decay of a HNL to a  $\mu\pi$  pairs.

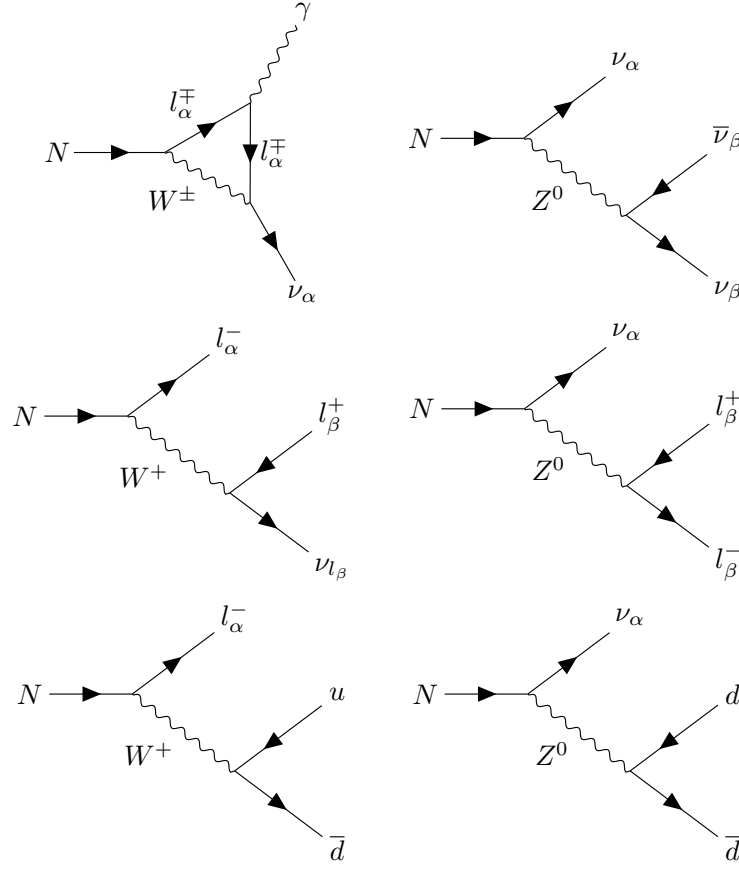
beamline and decay to observable particles. Searches for these type of signatures are called beam-dump or decay searches. A beam-dump minimises the background by absorbing all the other secondary products of the parent meson decay, such as muons, so that only HNLs and neutrinos reach the detector. In colliders both searches are possible, since production and decay occur in the same volume.

Unlike in production searches, for decay searches the number of events scales with  $|U_{l4}|^4$ , since the mass-mixing matrix element is involved at both production and decay. The signature is a vertex with no hadronic recoil activity (unlike neutrino interactions) and with a total invariant mass peaking at the HNL mass.

Several decay channels are available (a full list and phenomenological treatment is available in Ref. [18, 19]). In figure 2.5 are shown the Feynman diagrams for the possible decays. The possible final states available to HNLs that can be produced by the neutrino beam facility at Fermilab are:

$$\begin{aligned}
 N &\rightarrow \gamma \nu, \quad N \rightarrow \nu \nu \nu, \\
 N &\rightarrow e^- e^+ \nu, \quad N \rightarrow e^\mp \mu^\pm \nu, \quad N \rightarrow \mu^- \mu^+ \nu, \\
 N &\rightarrow e^\mp \pi^\pm, \quad N \rightarrow \mu^\mp \pi^\pm, \quad N \rightarrow \nu \pi^0
 \end{aligned} \tag{2.6}$$

All the decay channels listed above, with the exception of the channels containing a charged meson, are sensitive only to *combinations* of mass-mixing matrix elements (i.e., they cannot probe *single* elements of the mass-mixing matrix). The decay channel  $\nu \nu \nu$  and  $\nu \pi^0$  are possible only via neutral-current diagrams with the exchange of a  $Z$  boson. They are thus sensitive to a combination of  $|U_{e4}|^2$ ,  $|U_{\mu4}|^2$  and  $|U_{\tau4}|^2$ . The final states  $e e \nu$  and  $\mu \mu \nu$  are produced via charged-current and neutral-current interactions. The channel  $\gamma \nu$  has only a charged-current diagram, but since it occurs at loop level, all three possible flavours can be involved. The same occurs with the

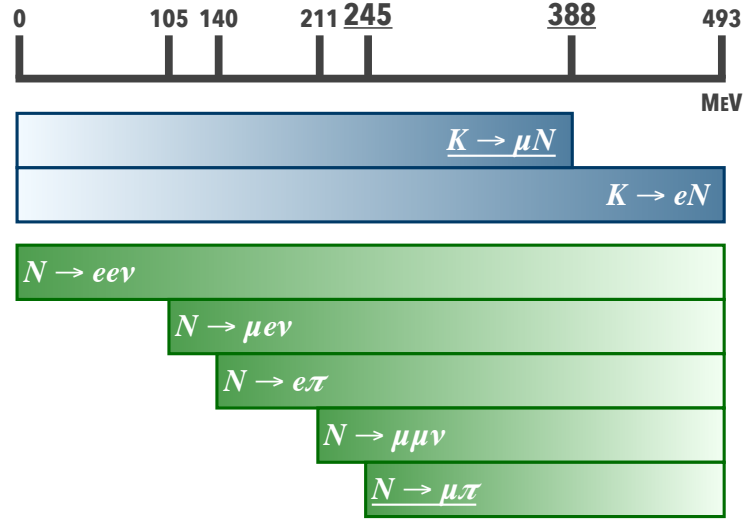


**Figure 2.5:** Feynman diagrams for the HNL decays. *First row:* The loop level photonic decay with the radiative emission of a photon is shown to the left. This decay channel is significantly suppressed by the loop factor, leading to an extremely low branching fraction. On the right is shown the HNL decay channel to three neutrinos which despite the high branching fraction is virtually unobservable. *Second row:* Both diagram show a final state with two leptons and a neutrino which can be CC-mediated (left) or NC-mediated (right). For final states with same flavour lepton (e.g.,  $e e \nu$ ) both diagrams can contribute, albeit with different magnitudes depending on the values of the mass-mixing matrix elements. *Third row:* Final states including a meson-lepton pair in the decay products, both charged for CC-mediated diagrams (left) or both neutral for NC-mediated (right).

$e \mu \nu$  channel, where only a charged-current diagram is involved, but the same final state can be obtained by swapping the electron with the muon and consequently the mass-mixing element involved (the state is thus sensitive to a combination of  $|U_{e4}|^2$

and  $|U_{\mu 4}|^2$ ). The final two channels,  $e\pi$  and  $\mu\pi$ , are direct probes of their mass-mixing matrix elements, since they are dependent, respectively, only on  $|U_{e4}|^2$  and  $|U_{\mu 4}|^2$ .

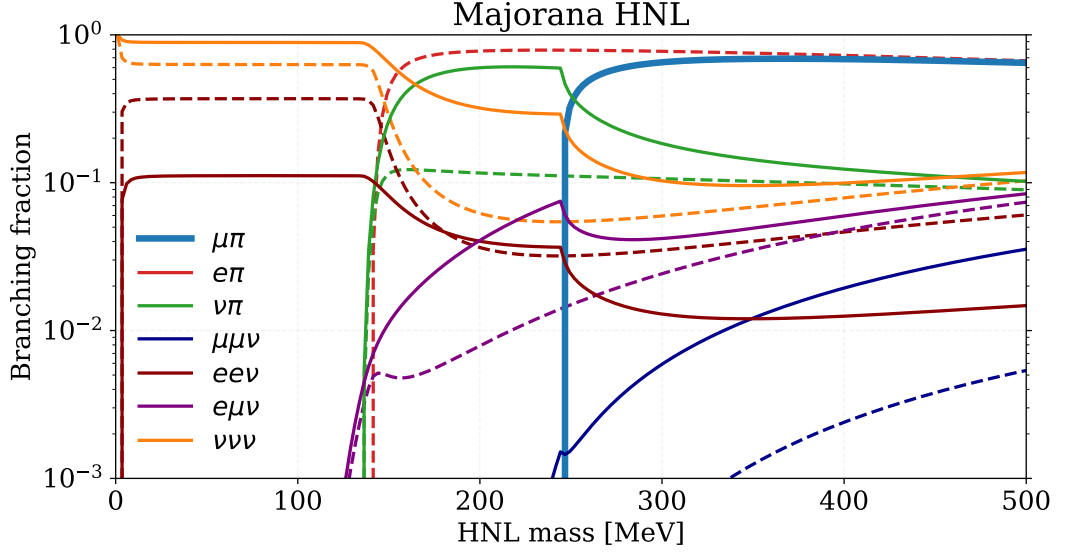
Each of these decay channels becomes accessible when the mass of the HNL is high enough to decay to the final state particles. Figure 2.6 shows the kinematic boundaries for the production and decay channels.



**Figure 2.6:** Boundaries on HNL mass for kinematically allowed production (blue) and decay (green) channels. Production channels limit the maximum value of HNL mass that can be probed. Decay channels are responsible for boundaries on the minimum mass value. The production and decay channels considered in this thesis and the mass values boundaries that can be probed are underlined.

By assuming different values for the mass-mixing matrix elements involved, different production and decay channels become accessible. Figure 2.7 shows the branching fractions for the decay channels listed above. Since the  $e\pi$  and  $\mu\pi$  decays have only a single charge-current diagram sensitive to the respective mass-mixing matrix element these channels exist only if the electron-coupling mass-mixing element and the muon-coupling mass-mixing element are not zero, respectively.

In this thesis we place limits on  $|U_{\mu 4}|^2$  assuming the other mass-mixing matrix elements to vanish (see muon-only coupling in figure 2.7). The most promising decay channel candidate is the  $N \rightarrow \mu\pi$  decay, since it is the dominating decay channel



**Figure 2.7:** Branching fractions for the HNL decay channels (decay width are for a Majorana HNL and are taken from Ref.[18]). The branching fractions vary for different combinations of the mass-mixing matrix elements. We assume a hierarchical scenario, showing with solid lines the muon-only coupling case ( $[U_{e4} : U_{\mu 4} : U_{\tau 4}] = [0 : 1 : 0]$ ) and with dashed lines the electron-only coupling case ( $[U_{e4} : U_{\mu 4} : U_{\tau 4}] = [1 : 0 : 0]$ ). Here we probe  $|U_{\mu 4}|^2$  and work in the former assumption (solid lines). In light blue and thicker line width is shown the decay channel studied here ( $\mu\pi$ ).

for all mass values kinematically allowed, while the production channel chosen is  $K \rightarrow \mu N$ . This restricts the mass range that can be probed to  $[245, 388]$  MeV.

The decay width for the  $\mu\pi$  channel is given by [18]:

$$\Gamma(N \rightarrow \mu\pi) = 2 \frac{G_F^2}{16\pi} f_\pi^2 |V_{ud}|^2 |U_{\mu 4}|^2 m_N^3 I_1(\delta_N^\mu, \delta_N^\pi), \quad (2.7)$$

$$I_1(x, y) = [(1+x-y)(1+x)-4x] \sqrt{\lambda(1, x, y)},$$

where  $G_F$  is the Fermi coupling constant,  $f_\pi$  is the pion decay constant,  $|V_{ud}|$  is the CKM matrix element,  $\delta_b^a = (m_a/m_b)^2$  represents the squared mass ratio,  $I_1$  is an integral over phase-space (such that  $I_1(0,0,0) = 1$ ), and  $\lambda$  has been defined in equation (2.5) as the two-body decay phase-space factor. The decay width grows with the third power of the mass of the HNL. The factor of two comes from considering a Majorana decay

which sums over the two Dirac decay channels. The angular distribution of the decay products for Dirac HNLs depends on the polarisation of the HNL with a dependency proportional to  $A \pm B \cos \theta$ , where  $\theta$  is the angle of the daughter lepton direction with respect to HNL parent meson direction in the HNL rest frame and the sign depending on the HNL helicity state. For Majorana neutrinos, an isotropy in the decay occurs if the charges of the final state are not separated, with the HNL and anti-HNL decays occurring in equal number, cancelling the  $\cos \theta$  dependency. Since MicroBooNE is a charge insensitive detector (e.g., it cannot distinguish a  $\pi^+$  from a  $\pi^-$ ), the sum of the two decay channels  $\mu^-\pi^+$  and  $\mu^+\pi^-$  is considered [18, 19].

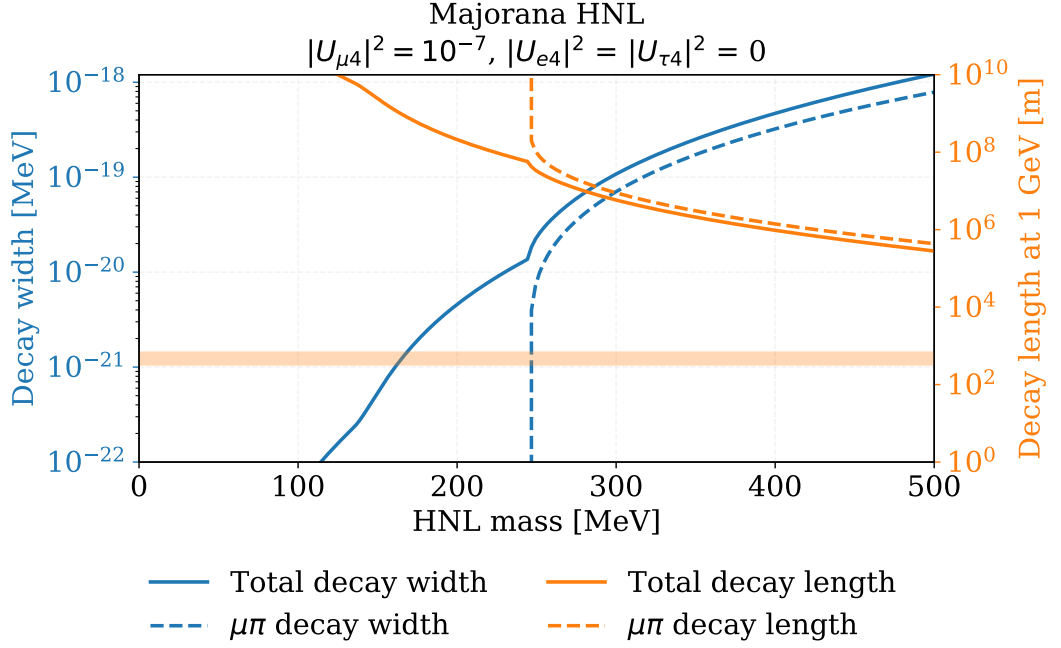
Due to the very small decay width, the decay lengths are extremely large. If the total decay length  $\Lambda_N$  is much larger than the distance to the detector ( $\Lambda_N \gg 470$  m), we can approximate the number of decays occurring in the detector by:

$$N_{\text{decays}} \simeq \int_{m_N}^{\infty} \phi_N(E_N) \frac{V}{\lambda_{\mu\pi}} dE_N, \quad (2.8)$$

where  $\phi_N(E_N)$  is the HNL flux,  $V$  is the detector volume and  $\lambda_{\mu\pi}$  is the partial decay length for the  $\mu\pi$  decay. Figure 2.8 shows that for the mass-mixing matrix element values studied in this work ( $|U_{\mu 4}|^2 = 10^{-7}$ ), the total decay length is more than three orders of magnitude larger than the distance from source to detector (470 m) which allows us to neglect the number of neutrinos decaying before reaching the detector [20].

## 2.4 Previous Searches

The number of fermion families, and thus the number of active SM neutrinos coupling to the weak bosons, has been measured by the LEP collaborations [21]. The invisible decay width of the  $Z$  boson, produced in  $e^+e^-$  collisions, can be measured by subtracting the visible decays to quarks and charged leptons from the total  $Z$  width. Assuming the invisible width to be due to the decay of  $N_\nu$  light neutrino species, the ratio of the invisible decay width to the leptonic one returns a value of  $N_\nu = 2.984 \pm 0.008$ . This value sets the number of active light neutrinos to three, for neutrinos with a mass below  $m_Z/2$ , where  $m_Z$  is the  $Z$  boson mass [22]. The number of right-handed *sterile* neutrinos, their masses and coupling to the SM neutrinos is however unconstrained.



**Figure 2.8:** Decay width (blue) and decay length (orange) for the sum of all decay channels (solid) and the  $\mu\pi$  decay channel (dashed). Decay width and decay length are calculated for a Majorana HNL with coupling set to  $|U_{\mu 4}|^2 = 10^{-7}$ . Decay length is calculated for a HNL with energy of 1 GeV. The horizontal orange line shows the distance between the neutrino source and the MicroBooNE detector ( $\approx 470$  m). The figure illustrates how the decay length is orders of magnitude larger than the distance travelled before decaying in the detector.

Aside from anomalies that might be interpreted in terms of models containing a potential HNL (e.g., the short-baseline oscillation anomalies observed by LSND [13] and MiniBooNE [14–16]), no experimental evidence for their existence has been found to date.

Experimental searches for HNLs have been performed at several mass scales, with different phenomenological effects manifesting at each scale. Oscillation experiments and measurements of  $\beta$  decays can probe keV scale neutrinos while collider experiments have access to mass ranges on the order of GeV and above (for a general overview that encompasses a larger mass range, from eV to TeV, see Ref. [23]). The LArTPC detectors of the SBN and DUNE programs are suited for these type of searches and studies on the expected sensitivity have been performed [6, 19]. We will focus on

MeV-scale, with HNLs coupling to  $|U_{\mu 4}|^2$  that can be probed with two different search methods: peak searches and beam-dump searches.

### 2.4.1 Peak Searches

Peak search experiments look for monochromatic lines in the lepton energy spectrum in the final state of a charged meson decay. The leptonic two-body decay of an electrically charged meson ( $P \rightarrow l \nu$ ) produces a lepton and a neutrino with well defined momenta in the meson rest frame, due to energy and momentum conservation constraints. For a fixed parent meson momentum, in the meson rest frame, we expect the charged lepton spectrum to be monochromatic with a line at:

$$p_l = \frac{m_p^2 + m_l^2}{2m_p}, \quad (2.9)$$

where  $m_p$  and  $m_l$  are the masses of the parent meson and final-state lepton, respectively [9]. In the presence of an additional HNL with mass  $m_N$ , an additional line can be found at:

$$p'_l = \frac{\sqrt{((m_p - m_N)^2 + m_l^2)((m_p + m_N)^2 + m_l^2)}}{2m_p}. \quad (2.10)$$

The flavour of the lepton used for the search determines the mass-mixing element probed. The height of the additional line can be used to measure the mass-mixing matrix element value and to place upper limits on the mass-mixing matrix element. Limits placed by peak search experiments are independent of the Dirac or Majorana nature of the HNL since it does not affect the kinematics of the production mechanism [9]. For  $|U_{\mu 4}|^2$ , the most competitive limits have been obtained by the experiments described in the following sections. A more complete review of the experimental constraints for the other masses ranges, and including  $|U_{e 4}|^2$ , can be found in Refs. [9, 23].

### Pion Decay Spectrum Peak Searches

- **SIN** (Swiss Institute for Nuclear Research) has performed a search for admixtures of massive neutrino in the pion decay  $\pi^+ \rightarrow \mu^+ \nu$  by stopping positive pions in a



scintillator in 1981 and a germanium detector in 1987. The use of pions allows for the study of low-mass HNLs with masses  $\mathcal{O}(10 \text{ MeV})$ . Upper limits on the mass-mixing matrix element  $|U_{\mu 4}|^2$  were set at  $10^{-5}$  in the 1-30 MeV mass range [24–26].

- The **PIENU** collaboration at TRIUMF has performed analogous measurements of the  $\pi^+ \rightarrow \mu^+ \nu$  decay channel, with a recent improvement in 2019. Pions are stopped using a thick plastic scintillator. A set of wire chambers, silicon strip detectors and calorimeters allow for a level of precision sufficient to place upper limits of  $|U_{\mu 4}|^2 < 3 \times 10^{-6}$  in the range 15-30 MeV and  $|U_{\mu 4}|^2 < 10^{-5}$  in the range 28-32 MeV [27].

### Kaon Decay Spectrum Peak Searches

- The **KEK** collaboration has performed measurements of the muon range spectrum in the stopped  $K^+$  decay in 1982, using a 0.5 GeV proton beam and a magnetic spectrograph, studying different decay modes. Using  $K$  mesons allows for the study of a higher HNL mass range with respect to pion decays. The KEK collaboration analysed  $10^6$  muons from kaon decay. Limits were placed in the  $|U_{\mu 4}|^2$  range  $10^{-4}$ - $10^{-6}$  in the mass range 70-300 MeV [28, 29].
- The **E949** collaboration at Brookhaven National Laboratory has observed  $10^{12}$  stopped kaons from 21.5 GeV energy protons in 2002. Upper limits on the mass-mixing matrix element were set at the level of  $10^{-7}$  to  $10^{-9}$  for HNLs in the mass region 175 to 300 MeV, yielding the most competitive limits from peak searches in this mass range [30].
- The **NA62** collaboration at CERN analysed  $10^8$  stopped kaons from 400 GeV protons extracted from the CERN SPS in 2015. Upper limits were placed at the  $|U_{\mu 4}|^2 < 10^{-7}$  level in the mass range 250-373 MeV, extending the peak searches mass range above 340 MeV [8].

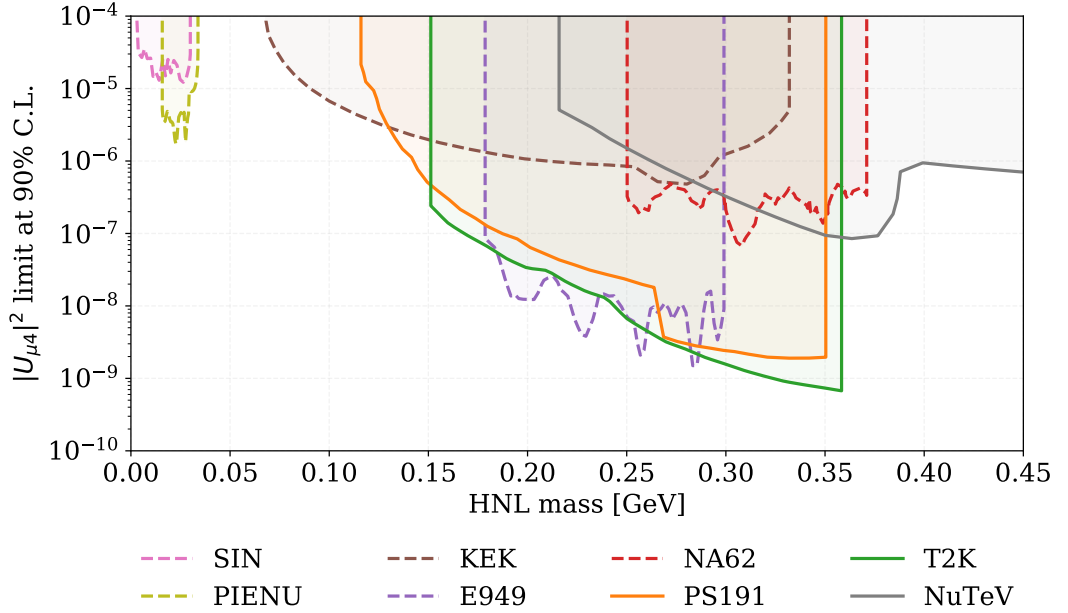
### 2.4.2 Decay Searches

Decay searches look for decay products of the **HNL**, which can propagate from their site of production and decay in a detector. Different combinations of production and decay channel offer sensitivities to different mass-mixing matrix elements. Historically these searches have been performed in beam-dump experiments designed to suppress **SM** interactions that can generate background events. Despite being built for different physics, with higher **SM** neutrino interaction background rate, modern neutrino oscillation detectors share sufficient similarities with beam-dump experiments and have sufficient high resolution to compete with custom-built detectors. For  $|U_{\mu 4}|^2$ , the most competitive limits have been placed by the following experiments:

- The CERN **PS191** experiment was performed in 1984, with an exposure of 19.2 GeV protons on a beryllium target resulting in  $\approx 10^{19}$  protons on target. The detector was located 128 m from the target and *off-axis*, with a  $2.3^\circ$  angle with respect to the beamline. The PS191 detector is an example of beam-dump type search, consisting of a detector specifically designed to look for **HNL** decays. PS191 operated in 1984 and still holds very competitive limits. For a fixed flux of **HNLs**, the number of decays increases with detector volume and with proximity to the flux source, while the number of background events (**SM** neutrino interactions) increases with the target atomic number (higher cross section). PS191 maximises signal and minimises background by using a large 216 m<sup>3</sup> flash chamber filled with helium, which provides a minimal  $\nu$ -medium cross section. Limits were placed in the mass range 120-350 MeV for  $|U_{\mu 4}|^2$  in the range  $10^{-5}$ - $10^{-9}$  [31, 32].
- The **T2K** collaboration performed a recent search for **HNLs** using data collected between 2010 and 2017 by its off-axis near detector ND280 for a total exposure of  $\approx 2 \times 10^{21}$  protons on target. ND280 measures neutrinos produced by 30 GeV protons impacting on a graphite target and it is located 280 m from the target with an off-axis angle of  $2.04^\circ$ . The analysis combines different production and decay channels and places the most stringent limits to date in the mass range 150-360 MeV for  $|U_{\mu 4}|^2$  in the range  $10^{-7}$ - $10^{-9}$  [33].
- The **NuTeV** collaboration at Fermilab performed a search for **HNL** decays in 1996,

produced by the secondary mesons resulting from the interaction of 800 GeV protons. The high energy of the beamline enabled the production of  $D$  mesons and thus access to a larger mass range for the HNLs, allowing the production of HNLs with masses up to 2 GeV. The detector was located 1400 m downstream of the production target and used a dataset consisting of  $\approx 10^{18}$  protons on target. Limits were placed in the mass range 225-2000 MeV for  $|U_{\mu 4}|^2$  in the range  $10^{-6}$ - $10^{-7}$  [34].

Figure 2.9 shows the limits previously set in the  $|U_{\mu 4}|^2$ - $m_N$  space.



**Figure 2.9:** Previous searches for  $|U_{\mu 4}|^2$  coupling HNL. Dashed lines show the limits of peak search experiments, solid lines indicate the limit set by decay search experiments. Limits are for Dirac neutrinos at 90% confidence level.

# The MicroBooNE Detector and the Beamline

The detector used in the Micro Booster Neutrino Experiment (MicroBooNE) consists of a large 170-ton Liquid Argon Time Projection Chamber (LArTPC) designed for the precision measurement of neutrino interactions in the Booster Neutrino Beam (BNB) at Fermilab. It is part of a series of detectors employing liquefied noble gases as a medium for the detection of neutrino interactions, combining calorimetry measurements and extremely high spatial resolution [35]. Due to their excellent performance, LArTPCs have become the technology of choice for the development of Fermilab's neutrino programme which includes the three Short Baseline Neutrino (SBN) detectors [36] (SBND, MicroBooNE and ICARUS) and the upcoming Deep Underground Neutrino Experiment (DUNE) experiment [37].

LArTPCs are versatile detectors and their physics applications include oscillation analysis at both short (SBN [36]) and long (DUNE [37]) baselines and detailed neutrino-argon cross-section measurements. Their high spatial resolution makes them also ideal candidates for searches in astroparticle and exotic physics. MicroBooNE [35] is the first of the family of SBN/DUNE detectors operating at Fermilab, collecting data since 2015, and its purpose is thus two-fold: as an independent experiment producing physics results and as an R&D project, paving the way towards large scale LArTPC detectors.

This section presents an overview of the working principle of the Liquid Argon Time Projection Chamber technology, its operation and the neutrino beam servicing

it, focusing in particular on the design details of the MicroBooNE LArTPC. It will address the mechanism by which electrons and photons are produced in the detector via energy loss in [section 3.1](#), the methods employed by the MicroBooNE detector to collect the generated signals for digitisation and successive analysis in [section 3.2](#), and the physics involved in the transition from signal production to signal collection, responsible for some of the sources of systematic uncertainty, in [section 3.3](#). In [section 3.4](#) the principles of the operation of the BNB will be outlined. References for the most part of the content of this section can be found in Ref.[[35](#)].

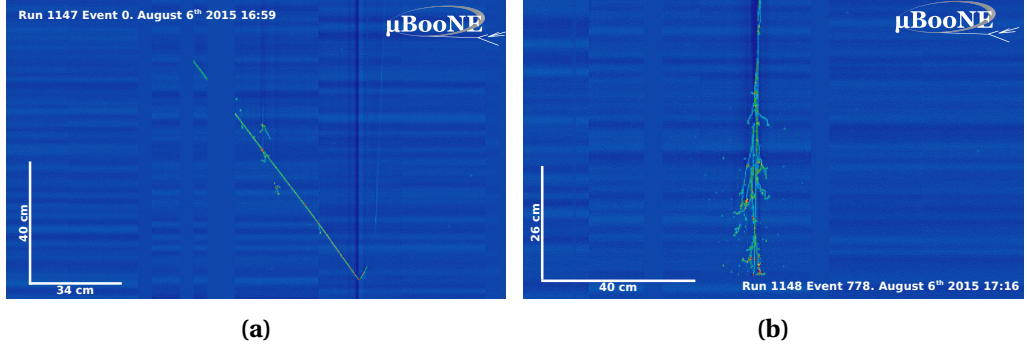
## 3.1 Energy Loss in Liquid Argon

Particles traversing a medium, like the liquid argon of a LArTPC, experience energy loss and produce or excite secondary particles, such as scintillation photons and ionisation electrons, along or at the end of their trajectory. These secondaries can then be collected by the detector instrumentation. This process, which is ultimately responsible for the detection of such primary particles, is fundamental to the understanding and interpretation of the signal obtained.

At the energy ranges relevant in LArTPCs, muons, pions and protons lose energy primarily via ionisation. The continuous trail of excited electrons along the primary particle trajectory gives rise to elongated, segment-like structures. This topology is referred to as *track* or *track-like* objects. For photons and electrons on the other hand, radiative processes dominate over ionisation losses. These processes are discrete and stochastic in nature and give rise to electromagnetic cascades, referred in LArTPCs as *showers* or *shower-like* objects. [Figure 3.1](#) shows the differences between track and shower objects in liquid argon.

### 3.1.1 Tracks: Muons, Pions and Protons

As muons, pions and protons traverse liquid argon, they interact electromagnetically with its atoms, predominantly via Coulomb scattering, freeing electrons along their trajectory. The distribution of mean charge freed, via energy deposition, per unit



**Figure 3.1:** Comparison of a track (a) and shower (b) object in liquid argon.

distance is described by the Bethe-Bloch formula [22]:

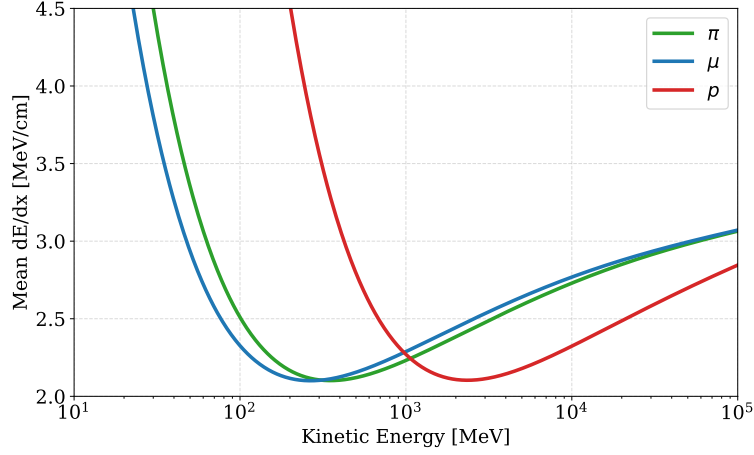
$$-\left\langle \frac{dE}{dx} \right\rangle = K z^2 \frac{Z}{A} \frac{1}{\beta^2} \left[ \frac{1}{2} \ln \frac{2m_e c^2 \beta^2 \gamma^2 T_{\max}}{I^2} - \beta^2 - \frac{\delta(\beta\gamma)}{2} \right], \quad (3.1)$$

where  $Z$  and  $A$  represent the atomic charge and mass of the medium traversed by the particle,  $I$  the mean excitation energy,  $\delta(\beta\gamma)$  a density correction to the energy loss, and  $K$  is a constant, set to  $4\pi N_A r_e^2 m_e$ . The energy  $T_{\max}$  is the maximum possible energy transfer from a single collision:

$$T_{\max} = \frac{2m_e \beta^2 \gamma^2}{1 + (2\gamma m_e)/M + (m_e/M)^2}, \quad (3.2)$$

where  $M$  is the mass of the incident particle. More information on the value of the constants used and on parameterisations used for the density effect correction can be found in Chapter 33 of Ref. [22].

Figure 3.2 shows the mean  $dE/dx$  in liquid argon for muons, protons and charged pions, as derived from equation (3.1). From the figure it is possible to extract two features regarding the passage of muon, pions and protons in liquid argon. At the range of energies relevant to LArTPC for muon and pions in liquid argon [100 MeV - 1 GeV], the mean  $dE/dx$  is around 2.2 MeV/cm which corresponds to the minimum of the Bethe-Bloch equation. Particles in this regime, such as cosmic-ray (CR) muons, can travel large distances in the detector before their energy loss  $dE/dx$  increases exponentially towards the end of their ionising trail (see figure 3.1a). These particles are referred to as Minimum-Ionising Particles (MIP). Because of their ability to travel for large distances in the detector and the constant  $\langle dE/dx \rangle$  profile, MIPs are particularly



**Figure 3.2:** Mean collision energy loss per unit length ( $dE/dx$ ) in liquid argon for muons, protons and charged pions as a function of their kinetic energy.

useful for the calibration of the detector. Due to their very similar mass values, pions and muons share a very similar  $\langle dE/dx \rangle$  profile. Protons, however, have significantly more mass, and  $dE/dx$  can thus be used to discriminate protons from muons and pions [22].

#### 3.1.2 Showers: Electrons and Photons

While this analysis concerns itself mainly with track objects (the  $\mu$  and  $\pi$  produced by the HNL decay are responsible for a *two-track* signature), we will touch briefly on the electron and photon energy losses to understand some of the effects related to track reconstruction.

For electrons, radiative photon production from bremsstrahlung is the dominant contribution to energy loss. Ionisation energy loss scales logarithmically with the energy of the electron, while energy losses due to bremsstrahlung scale linearly. Photons with energies beyond a few MeV lose energy predominantly via  $e^+e^-$  pair production. The electrons (and positrons) produced by pair production can then generate additional photons which subsequently produce a further  $e^+e^-$  pair. A consequence of this process is the generation of a cascade of particles, with increasing multiplicity and decreasing energy per particle, until a critical energy  $E_c$  is reached, for both electrons

and photons. Low-energy electrons eventually do not have enough energy for emitting bremsstrahlung. Low-energy photons interact predominantly via photoelectric effect and Compton scattering once their energy for pair production is below threshold. Such cascades are referred to as electromagnetic showers (see [figure 3.1b](#)), due to their branching features [22].

#### 3.1.3 Scintillation Light

Along with ionisation electrons, energy is released by the passage of the ionising particles across the argon in the form of *scintillation light*. Two important mechanisms, induced by the action of an electric field, are responsible for scintillation: recombination of the released electron with the argon ion and exciton self-trapping, in which excited argon atoms can combine with other argon atoms in the ground state. In both cases the result is the formation of an excited dimer (excimer) molecule where an  $\text{Ar}_2^*$  molecule consists of two argon atoms sharing a bound electron.

The excimer molecule is short-lived and decays to two ground-state argon atoms, isotropically releasing an ultraviolet photon (narrowly peaked at 128 nm) with a timing constant that depends on the excitation state. For a singlet state, which has a shorter mean lifetime, the decay constant is  $\tau_S \approx 6$  ns, while for the triplet state, which has a longer mean lifetime, the decay constant is  $\tau_T \approx 1.5$   $\mu$ s. We denote the amplitudes for the two decays by  $A_S$  and  $A_T$ , so that the time dependence of the scintillation light emission in pure liquid argon can be represented by a probability distribution function (PDF):

$$I(t) = \frac{A_S}{\tau_S} \exp\left(-\frac{t}{\tau_S}\right) + \frac{A_T}{\tau_T} \exp\left(-\frac{t}{\tau_T}\right). \quad (3.3)$$

The ratio of single-fast to triplet-slow states is  $\approx 1:3$  for MIPs, i.e.,  $A_S \approx 0.25$  and  $A_T \approx 0.75$  (in pure argon  $A_S + A_T = 1$ ) [38].

Detecting ultraviolet photons is a challenging experimental task since the glass window of conventional optical photodetectors is not transparent to UV light. Wavelength shifters are employed to allow detection of UV photons as described in [section 3.2.4](#). The first electronic excited state of argon is at a higher energy than either the singlet or triplet excited states. The 128 nm scintillation light can thus not be reabsorbed, making liquid argon transparent to its own scintillation light and allowing for high



detection efficiencies across significant photon travel distances. Furthermore, argon is an extremely bright scintillator, yielding approximately 4,000 photons per MeV of deposited energy in a 273 V/cm electric field.

The timescales of 6 ns and 1.6  $\mu$ s are *prompt* compared to the time it takes for the electrons to move towards the detector instrumentation ( $\mathcal{O}(\text{ms})$ ), making scintillation photons an extremely useful tool for triggering and time-tagging events [35].

## 3.2 Collecting and Recording the Signal

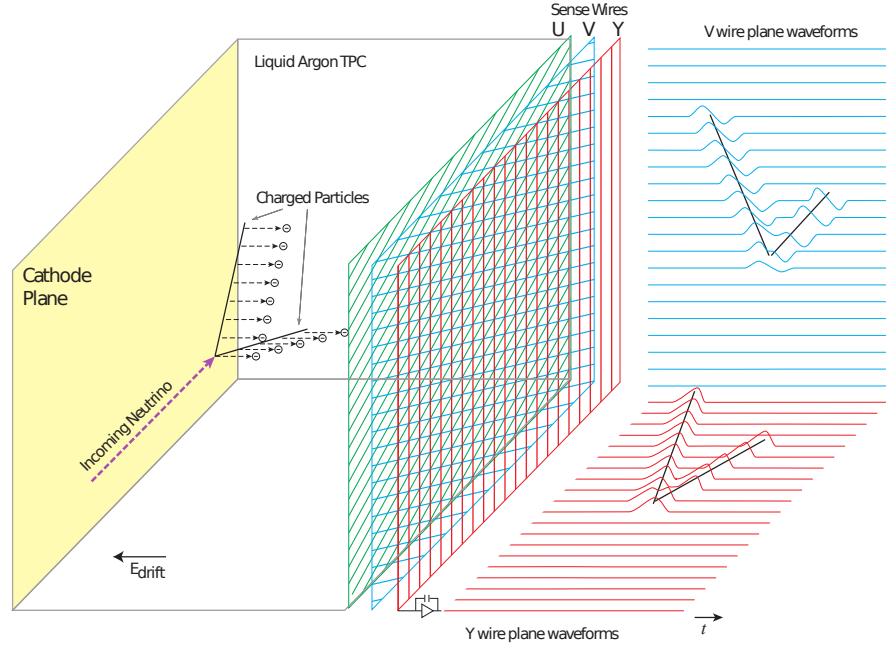
The ionisation electrons and scintillation photons produced by a relativistic particle crossing the argon are collected and recorded by two different systems, responsible for charge and light collection. In order to transport the ionisation electrons to the detector instrumentation, a uniform electric field of 273 V/cm is applied which causes the electrons to drift towards a series of sense wires. A set of photomultiplier tubes is used for photon collection.

Charge drift is behind the basic operating principle of a time projection chamber (TPC). This section will describe in further detail how this is accomplished, outlining in the process MicroBooNE's technical design.

### 3.2.1 The MicroBooNE Time Projection Chamber

MicroBooNE's TPC is a cuboid 10.36 m long in the beam-direction, 2.32 m in the vertical direction, and 2.56 m in the horizontal drift direction, housed inside a cylindrical cryostat filled with 170 tons of liquid argon. The TPC comprises a field cage, in which a high electric field is maintained by a cathode plane, made of a sheet of stainless steel and kept at an electric potential of -70 kV, and an anode plane, sitting on the opposite side of the field cage. When ionisation electrons are produced inside the TPC, the constant electric field is responsible for the drifting of the ionisation charge which moves towards the anode. The anode plane houses the sense wires which are distributed across three different wire planes, with 3 mm spacing between each wire on a plane and 3 mm spacing between each plane.

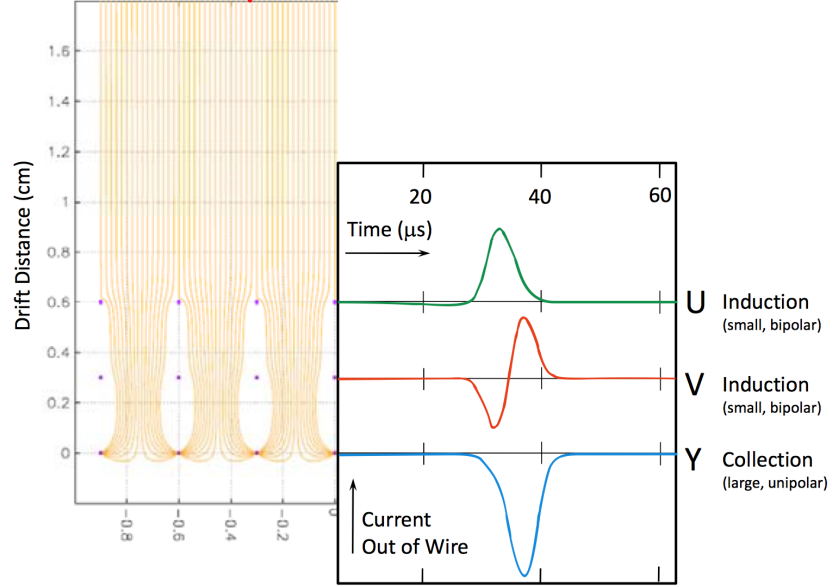
The sets of wires in the three wire planes are oriented at  $60^\circ$  with respect to each other (with respect to the vertical and listed in order towards the anode:  $U$  plane with 2400 wires at  $-60^\circ$ ,  $V$  plane with 2400 wires at  $60^\circ$  and  $Y$  plane with 3456 wires at  $0^\circ$ ). Figure 3.3 illustrates the basic operating principle of a time projection chamber and the location of the components of the TPC with respect to each other.



**Figure 3.3:** Diagram illustrating the basic operating principle of a time projection chamber (from Ref. [35]). Charged particles traversing the detector produce a trail of ionisation electrons which then drift towards the sense wires on the anode plane due to the induced electric field. The signals induced on the sense wires by the drifting charge are then collected for digitisation.

A bias voltage is applied across the three wire planes which causes the electrons to drift towards the furthest plane ( $Y$ ). While they cross the first two planes ( $U$  and  $V$ ), moving between the set of wires, they induce a current. The current is negative when they approach and turns positive when they move away. The end result of their transition is an induced bipolar waveform. The last wire plane ( $Y$ ) is positively biased and collects the entire charge of the electron, producing a unipolar negative

waveform. Because of the difference in waveforms, the  $U$  and  $V$  planes are referred to as *induction* planes, while the  $Y$  plane is referred to as *collection* plane. Figure 3.4 shows the different signals induced on the three wire planes.

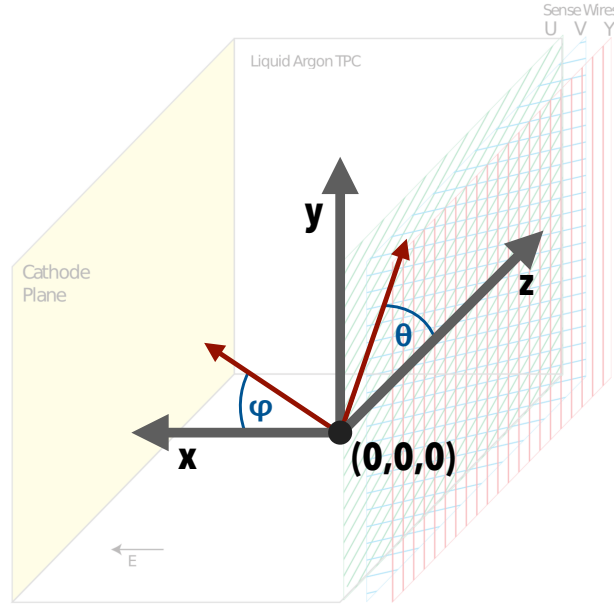


**Figure 3.4:** Diagram illustrating the different signals induced on the three wire planes by the ionising electrons (right) and the equipotential lines of the electric field (left) (from Ref. [39]). In the left figure, the  $x$  axis represents the beam-line direction whereas the  $y$  axis represents the distance to the anode plane (electrons drift *downwards* following the equipotential lines). The purple dots in the left figure show the location of the wires on the 2d-projection of the TPC. The shapes of the bipolar and unipolar waveform collected by the wires are shown in the right figure.

Because of the different orientation of the wires, the set of pulses recorded by each plane is a unique 2-d representation of the ionisation trail (see figure 3.3). Provided a  $t_0$  of the interaction is known (which can be obtained from light information, see section 3.1.3), the time  $t$  of the pulse can be used to define the coordinate along the drift direction axis. The three representations from the three planes can then be combined to obtain a 3-d stereoscopic reconstruction of the event [35].

### Internal Coordinate System

The internal MicroBooNE coordinate system is shown in figure 3.5 which illustrates the direction of the three axes:  $x$  along the drift direction, increasing with distance from the wire plane (the anode is located at  $x = 0$ ),  $y$  is the vertical direction and  $z$  is the *forward* direction, collinear with the neutrino beam direction.  $\theta$  is the polar angle, defined in the  $yz$  plane ( $\theta = 0$  is collinear to the beam direction,  $\theta = \pi/2$  is perpendicular to it).  $\phi$  is the azimuthal angle, defined in the  $xy$  plane ( $\phi = 0$  is parallel to the drift direction,  $\phi = \pi/2$  is parallel to the vertical direction).



**Figure 3.5:** MicroBooNE internal coordinate system.

#### 3.2.2 Cryogenics and Argon Purity

Argon at ambient temperature is gaseous. In order to keep it in a liquid state it has to be cooled to a temperature at 87 K. For this reason, a cryogenic vessel surrounds the TPC. The cryostat is coupled to a cryogenic system which keeps the temperature constant and ensures that the temperature gradient is minimised and that a uniform

electron drift velocity can be maintained. A pump system constantly recycles and filters the argon to keep the argon purity above the minimum operational threshold.

LArTPCs rely heavily on the ability of the ionisation electrons to reach the wire plane in a straight trajectory at constant velocity. The presence of contaminants capable of capturing or attracting an electron (such as electronegative molecules, like  $O_2$ ) can negatively affect the detector energy and position resolution. In addition, light quenching from nitrogen contamination can attenuate the scintillation photons, with consequences on trigger efficiency and timing [35]. This is further discussed in section 3.3.2.

#### 3.2.3 TPC Readout and Signal Processing

The collection (induction) of charge is readout as analog signals. The signal has to travel for  $\approx 20$  m before reaching the readout electronics. An important component of the readout is the set of *Application-Specific Integrated Circuits* (ASICs) which operates in the cold, inside the cryostat, immediately adjacent to the TPC wires.

##### Cold Electronics

The ASICs operate as low-noise pre-amplifiers, ensuring that the signal has been shaped and amplified sufficiently in order to survive the travel to the readout without being covered by increasing noise. Performing these operations in the cold allows MicroBooNE to achieve a high signal-to-electronics-noise ratio. Each ASIC has 16 channels, connected to a set of adjacent 16 wires. It is configured to operate with the optimal settings of 14 mV/fC gain and 2  $\mu$ s peaking time. Cold cables transport the pre-amplified analog signal to dedicated feedthroughs mounted on the cryostat [35].

##### Warm Electronics

Faraday cages mounted on the external, warm side of the feedthrough provide shielding for the intermediate amplifiers located inside. The intermediate amplifiers, located on the warm side, provide a  $\approx 12$  dB gain in order to make the analog signal suitable

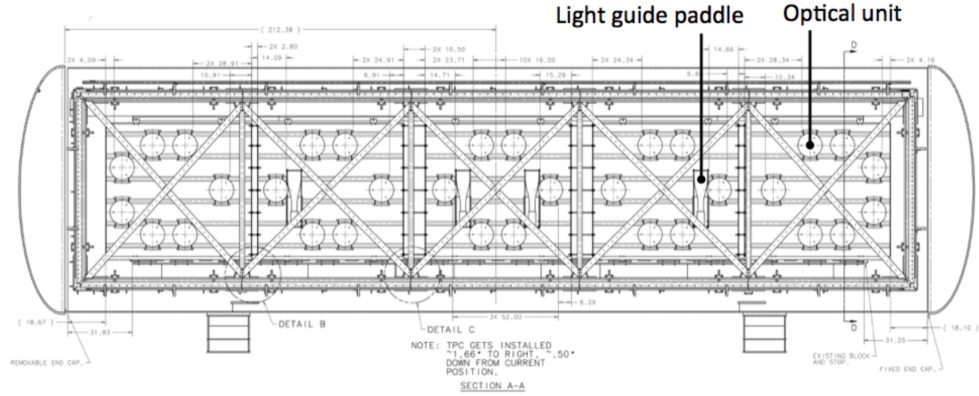
for transmission along a  $\approx 20$  m long cable to the readout electronics.

Once transported to the readout electronics system, the induction and collection signals are ready for being digitised, processed and readout. 130 custom-designed Analog-to-Digital Converter Front-End readout Modules (ADC/FEMs) are distributed over nine TPC readout crates. The readout crates are synchronised to a common 16 MHz clock and digitise the signal continuously. Each channel baseline can be configured independently, with low setting (450 ADC) for the collection channel or in the middle of the dynamic range (2055 ADC) for induction channel, in order to ensure that both induction and collection signals can take full advantage of the ADC analog input range. The dynamic range is set so that a MIP produced near the cathode (at the lower end) and a highly-ionising stopping proton near the anode (at the higher end) can be successfully observed with no saturation. Digitised signal is then passed to a Front End Module (FEM) which contains a Field-Programmable Gate Array (FPGA) chip performing downsampling from 16 to 2 MHz (resulting in 500 ns wide time-ticks). Each FPGA chip converts the waveform of 64 channels into time ordered waveforms of three 1.6 ms-wide frames, each counting 3,200 time-ticks. The frames are arranged so that one frame precedes the trigger time and the other two immediately follows it. The TPC output data is then sent to the Data Acquisition system (DAQ), described in section 3.2.6, where it is merged with the light collection system data [35].

#### 3.2.4 Light Collection System

While the information carried by the ionisation charge is processed by the TPC system, the task of detecting the scintillation photons is performed by the light collection system. For this purpose, MicroBooNE employs an array of 32 8-inch PhotoMultiplier Tubes (PMTs) and four light paddles which are located behind the wire plane, in the cold. The paddles propagate the captured light to the end of the bar where a silicon photomultiplier (SiPM) detects the photons. These paddles were included in the light collection system for R&D purposes but the measured signal is not used in this analysis. Figure 3.6 shows the location of the PMTs and paddles on a side view of the MicroBooNE detector [35].

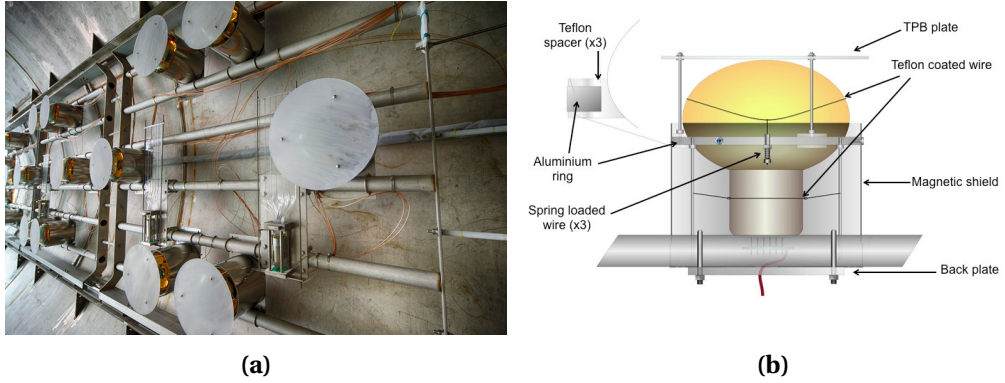
### 3.2. COLLECTING AND RECORDING THE SIGNAL



**Figure 3.6:** MicroBooNE light collection system, consisting of a primary system of 32 PMTs and a secondary system of four lightguide paddles [40].

### Detection

Each of the 32 optical units consists of a cryogenic Hamamatsu 5912-02MOD PMT, positioned behind an acrylic plate coated with an organic fluor called tetraphenylbutadiene (TPB). Figure 3.7 shows a diagram of the components of the optical unit and their location inside the cryostat.

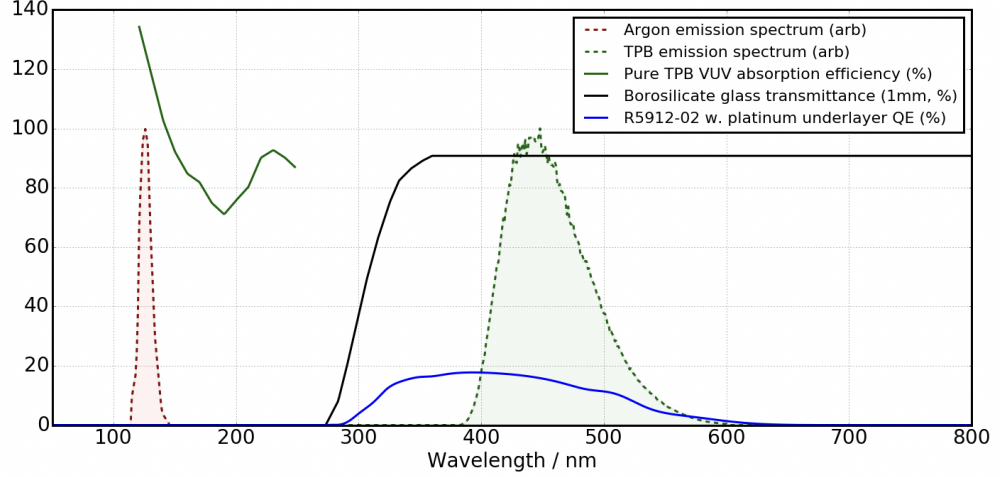


**Figure 3.7:** Close-up of a MicroBooNE PMT (left) and diagram showing its components (right).

The TPB coating is extremely important for maintaining a high photon detection efficiency. In section 3.1.3 we have underlined how the emission of scintillation light in the ultra-violet prevents the argon from reabsorbing its own light, enhancing the efficiency at varying travelling distances. A major drawback is however the fact that



photomultiplier tubes efficiency peaks in the visible spectrum ( $\approx 400$  nm)) and quickly drops to zero below few hundreds nanometers. [Figure 3.8](#) illustrates the challenges involved in the detection of 128 nm scintillation light.



**Figure 3.8:** Comparing spectra for the emission of scintillation light in argon, re-emission by TPB, and efficiencies for the instrumentation used. Spectra are normalised to arbitrary units. The TPB absorbs the ultra-violet light from argon scintillation (red dashed spectrum) with high efficiency (solid green) and re-emits visible light (green dashed spectrum) which is transmitted with high efficiency (solid black) by the borosilicate (glass) window of the PMT. In solid blue is then shown the quantum efficiency of the platinum underlayer of the PMT (from Ref. [41]).

The TPB-coated plate mounted on top of the PMT takes care of the issue by absorbing the 128 nm scintillation photons with high efficiency and re-emitting light in the optical window (blue wavelength at  $\approx 425$  nm) where the glass transmittance is maximal and PMT quantum efficiency is highest ( $\approx 20\%$ ) [35].

## Digitisation and Readout

The PMT electronics are similar to those used for the readout of TPC information, with similar process. Among some of the differences are the different digitisation clock, set at 64 MHz (with 15.625 ns time-ticks) and the use of a unipolar response function with a 60 ns rise time for the pulse shaping. The 32 PMT channels do not use a front-end ASIC but send the signal directly to a shaper located in the warm. A splitter circuit



splits the signal in high-gain (HG) and low-gain (LG) in order to increase the PMT dynamic range [35]. Unlike TPC data, which is read continuously, PMT data is collected in two different modes:

- **23.4  $\mu$ s-wide unbiased beam-gate readout**, which surrounds the neutrino beam spill. The unbiased readout records activity in coincidence with a neutrino spill and it is triggered by a signal sent by the Fermilab Accelerator Division facility.
- **0.6  $\mu$ s-wide triggered *cosmic discriminator* readout**, which reads all the data outside of the unbiased beam-gate readout and applies a “cosmic discriminator” threshold condition, set at 9.5 photoelectrons (PE), which has the purpose of recording light activity produced by CR muons. The threshold is applied independently to each PMT.

#### 3.2.5 Cosmic Ray Tagger

Aside from the TPC and PMT instrumentation, a third additional sub-system detects particle activity in the surrounding of the detectors. Although its data is not used in this analysis, it is described in this section since its use in future analyses may allow for a higher signal-to-background ratio.

The MicroBooNE detector sits in a pit  $\approx 6$  meters below the ground, with no overburden shielding, and is thus exposed to a high rate of CR muons, crossing the detector at 5 kHz. Due to the large readout window ( $\approx 4.8$  ms) required to collect the slow-drifting ionization charge, an average of 24 crossing CR muons are expected in each event, leading to a large expected number of background events mimicking the tracks (and showers) observed as the result of neutrino interaction. An additional sub-system, designed to identify and reject CR muons and known as Cosmic Ray Tagger (CRT), has thus been integrated to the set of MicroBooNE detectors.

The CRT consists of a series of scintillating panels installed around the cryostat, with each panel comprised of interleaved layers of plastic scintillating strips. When CR muons traversing the CRT modules interact with the plastic material they produce scintillation photons which are collected by SiPMs. The signal, which is digitised and read out by a customised front-end board, can then be used to reconstruct the

trajectory of the crossing muon by interpolating the entry and exit points. In addition, CRT uses timing information to match CRT tracks with TPC and PMT signals, in order to tag the tracks as background cosmogenic muons which can then be filtered out of the event.

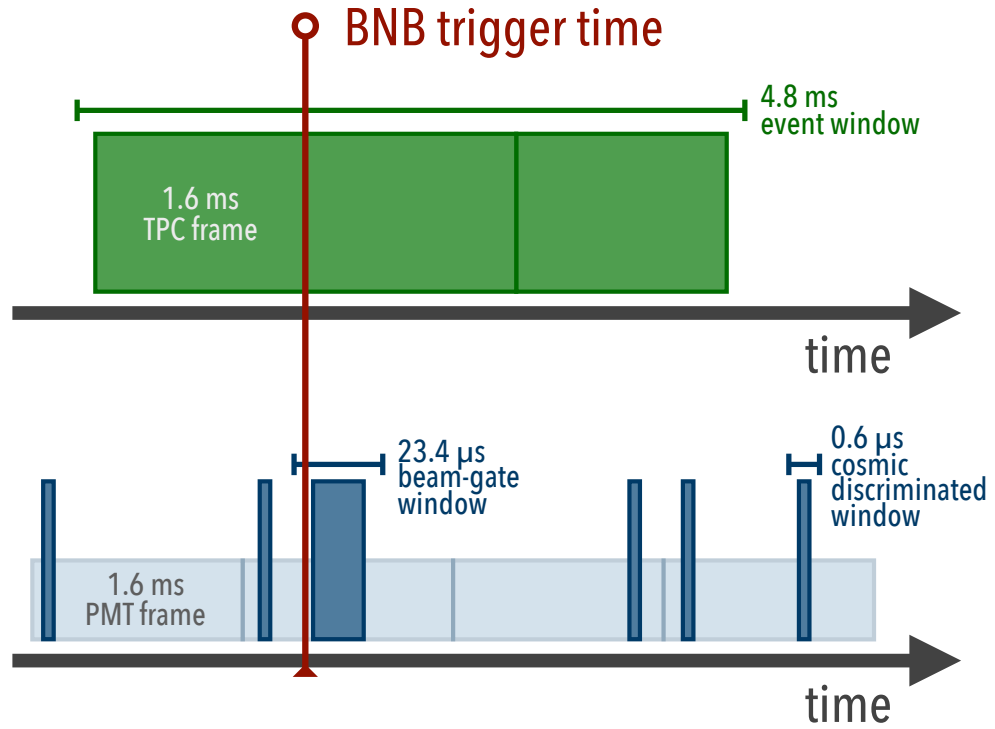
The CRT sub-system was installed and tested in October 2017 and the data was reconstructed in early 2018 so that this analysis does not use CRT information [42].

#### 3.2.6 Data Acquisition

From detection inside the cryostat to digitisation in the readout electronics system, the two TPC and PMT streams are kept separate and record detector activity independently. The TPC system records activity passively and continuously while the PMT system has an active role in processing and issuing triggers.

An externally issued trigger provides the  $t_0$  around which the two data streams are aligned (one of the advantages of accelerator-based neutrino experiments is that the arrival time and time profile of the neutrino pulse is well-defined). The external trigger can be either issued by the beam division, in correspondence to an incoming neutrino spill, or fired by a “pulser” at regular intervals outside of the neutrino spill window in order to provide a sample for characterising the background (details of the trigger system are discussed in section 4). Three 1.6 ms-wide TPC frames are formed which are aligned to the trigger time in a way such that the first 1.6 ms-wide frame precedes the trigger time and the other two immediately follows it. From the two PMT data modes described in section 3.2.4, the unbiased beam-gate readout and the cosmic discriminator readout, windows are selected from the four 1.6 ms-wide PMT frames surrounding the trigger time. A diagram of the activity recorded by the two data streams is shown in dark coloured blocks in figure 3.9.

One of the important MicroBooNE’s sub-systems is the DAQ system. The DAQ system’s purpose is to process the data received from the readout electronics and writing it to local disk before transferring it to long term storage. It also monitors the data flow and configures and control the readout electronics during data-taking periods [35]. It accomplishes the processing and writing tasks by moving the data received by the back-end electronics into an internal buffer where it is checked for



**Figure 3.9:** Diagram of the activity recorded by the TPC (green) and PMT (blue) data streams. TPC activity is processed continuously and once a trigger is issued, three 1.6 ms-wide TPC frames are formed surrounding the trigger, one frame precedes the trigger while two follow it, forming a 4.8 ms-wide event window. The PMT data streams process only PMT activity in 0.6  $\mu$ s-wide windows from the individual PMTs that pass the *cosmic discriminator* threshold condition. In addition, an unbiased 23.4  $\mu$ s-wide window following the trigger is processed with no threshold requirement. The unbiased beam-gate window and the cosmic-discriminated windows contained in the four PMT frames surrounding the trigger time are then recorded (in the diagram, only dark colour blocks are recorded, while the light ones are not). Block widths are not to scale.

consistency and assembled into an *event* [35].

### Data Rates

At this point, it is important to consider the rate at which the DAQ needs to process data to understand MicroBooNE's requirement for high-level triggers. With 9,600 samples

(500 ns-wide time-ticks at 2 MHz after FPGA downsampling, covering the 4.8 ms-wide TPC window), 8,256 channels (combining all the wires from the three planes) and 16 bits per words, a total of 150 MB are required per event. Huffman encoding scales that down by a factor of  $\approx 5$ , bringing the total volume down to 30 MB/event. At an average beam trigger-rate of 5 Hz that would imply a data volume of 13 TB/day.

The expected rate of neutrino beam spills resulting in a neutrino interaction is  $\approx 1/600$ . As a consequence, the vast majority of the 13 TB/day data would contain exclusively CR muons, with a small fraction of that data actually containing a neutrino interaction. A much more efficient and inexpensive way of processing the data is thus achieved by applying a high-level software trigger that can reduce the data rate by filtering and saving to disk only events which are likely to contain a neutrino interaction. The software triggers employ information recorded by the PMT unbiased 23.4  $\mu$ s beam-gate window in order to make a decision on an event-by-event basis. The different trigger types will be described in [section 4](#).

While in principle rates up to 20 Hz can be handled by MicroBooNE's DAQ, the real bottleneck is the disk writing speed which is limited to  $\approx 300$  MB/s, bringing the maximum allowed DAQ rate to  $\approx 16$  Hz. Since the data rate from the beamline triggers amounts to 5 Hz, this leaves space for  $\approx 11$  Hz that can be used to process off-beam data (events occurring outside of the beam spill which contain exclusively cosmogenic activity).

Events that pass the consistency checks and software trigger requirements are placed into an offline queue and then copied to another server via the internal DAQ network. The raw information is saved as a binary file. A separate step, called *swizzling*, converts the binary file into the *art-ROOT* file format using the LArSoft software framework (both software tools are described in [section 5](#)) [[35](#)].

## 3.3 Physics and Detector Effects

Up until now we have neglected the physics involved in the transport of the signal particles (ionisation electrons and scintillation photons) from the site of production, along the trail of the ionising charged particle, to the detector instrumentation on the

anode plane, thus assuming that the signal particles travel unperturbed towards the sense wires and PMTs. We take now a step back to take a further look at the physics of drifting and light collection and the physics effects arising from the detector operation. This is an extremely important step, since understanding all of the detector effects is necessary to correctly model the sources of systematic uncertainty and to provide a calibration scale to convert the measured signal charge on a wire ( $dQ/dx$ ) into the original energy loss due to ionisation ( $dE/dx$ ), thus enabling calorimetry. In order to properly estimate these factors it is necessary to consider all the circumstances that can hinder or perturb the transport of ionisation electrons and scintillation photons, and the efficiencies associated with the TPC and PMT instrumentation.

#### 3.3.1 Recombination

Electron-ion recombination is one of the first effects to consider. It can occur immediately in the few nanoseconds following the production of ionisation charge. Some of the ionisation electrons produced by ionising charged particles might not be able to completely liberate themselves from the parent argon ion they originated from and recombine with it to form neutral argon atoms again. If not taken into account, this effect can result in an underestimation of the particle energy loss due to ionisation.

A popular model for the description of this effect is the Box Model [43]. The inverse Box Model equation has the form:

$$\frac{dE}{dx} = \frac{1}{\beta} \left[ \exp \left( \beta W_{\text{ion}} \cdot \frac{dQ}{dx} \right) - \alpha \right], \quad (3.4)$$

where  $W_{\text{ion}}$  is the energy required to ionise an argon atom (equal to 23.6 eV), and  $\alpha$  and  $\beta$  are tunable parameters (conventionally, in the standard Box Model,  $\alpha$  is set to 1). Recombination effects are highly dependent on the intensity of the electric field applied (a higher electric field can reduce the effect of recombination, by providing a stronger “pull” towards the anode plane) and that dependency folds into the  $\beta$  parameter. Studies of electron recombination using highly ionising particle in the ArgoNeuT liquid argon TPC [44] have demonstrated that a “Modified Box Model” with parameters  $\alpha = 0.93$  and  $\beta = 0.30$  MeV/cm better describes the observed effect in LArTPCs. For MIPs, the amount of ionisation electrons undergoing recombination is

≈40%.

### 3.3.2 Attenuation and Argon Purity

Once the ionisation charge is drifting, few effects can perturb its motion and in certain cases arrest it altogether. For the latter case, we can consider attenuation. Attenuation of the signal can be caused by the capture, in transit, of the electrons (or photons) by highly electronegative contaminants such as oxygen and water which can reduce the intensity of the signal recorded at the wires. We will treat the two signal particles, electrons and photons, separately.

#### Ionisation Electron Attenuation

Much like the recombination effect, neglecting attenuation can result in underestimating the particle energy and impacts track reconstruction efficiency. Unlike recombination, attenuation is drift-distance dependent (for larger travel distances, electrons have a higher probability of being captured).

Argon purity and attenuation effects are thus linked, so that an expression for argon purity can be modelled by measuring the average time  $\tau_e$  that an electron can travel for before being captured by impurities. Taking oxygen as an example, which is the most significant contaminant, and assuming that the free electron concentration ( $[e^-]$ ) is smaller than the contaminant concentration ( $[O_2]$ ), we write the following expression for the decrease of electrons over time:

$$\frac{d[e^-]}{dt} = -k_e[O_2][e^-] \quad \rightarrow \quad [e^-](t) = [e^-](0) \cdot \exp\left(-\frac{t}{\tau_e}\right), \quad (3.5)$$

where  $k_e$  is the rate constant and where we can define an electron lifetime  $\tau_e$  as follows:

$$\frac{1}{\tau_e} = k_e[O_2]. \quad (3.6)$$

Electron lifetime (and consequently contaminant concentration and argon purity) can be determined by measuring the change in measured charge versus distance from

the anode plane for tracks that have a constant, known  $dE/dx$ . Through-going CR muon (MIPs) tracks are particularly suited for this task [45]. For example, by measuring the ratio between the charge originated at the anode ( $Q_A$ ) and the charge originated at the cathode ( $Q_C$ ), and reaching the anode after a time  $t$ , one obtains:

$$\frac{Q_A}{Q_C} = \exp\left(-\frac{t}{\tau_e}\right). \quad (3.7)$$

A high electron lifetime and, consequently, a very low impurity concentration, is an extremely important operational requirement for a LArTPC, with its importance increasing with larger volumes and longer drift distances. Figure 3.10 shows the impact of different values of electron lifetime (argon purity) on the attenuation effect. MicroBooNE's maximum drift distance is 2.56 meters and during stable periods of purity, the lowest recorded lifetime is 18 ms. Figure 3.11 shows the variation in  $Q_A/Q_C$  as a function of time. As can be seen from the figure, an average attenuation factor of  $\approx 5\%$  is observed for stable purity periods [46].

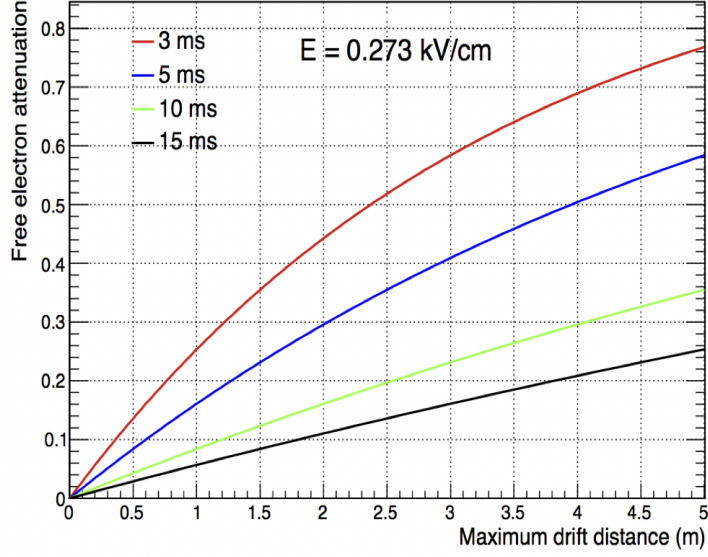
#### Scintillation Photon Attenuation

Attenuation effects due to contaminants in argon can occur in scintillation photon detection as well. The effects can be caused both by decreased light production (independent of drift distance) and by direct contaminant absorption (with a drift distance dependence).

In the first case, *quenching*<sup>1</sup> processes can reduce the amount of emitted photons, due to two-body collisions between “impurity molecules” (such as  $N_2$  and  $O_2$ ) and  $Ar_2^*$  excimer states which would otherwise decay radiatively with emission of scintillation photons. In section 3.1.3 we have discussed the mechanisms for scintillation light production, outlining the role of the two decay time constants  $\tau_S$  and  $\tau_T$  and corresponding amplitudes  $A_S$  and  $A_T$ , for the two excitation states.

While in pure argon the two relative amplitudes  $A_S$  and  $A_T$  are constrained to unity, the presence of contaminants can decrease the lifetime  $\tau_S$  and  $\tau_T$  of each component which is reflected in equation (3.3) by the substitution of the amplitudes  $A_S$  and  $A_T$

<sup>1</sup>Decrease in fluorescence intensity of a substance due to non-radiative processes.



**Figure 3.10:** Attenuation effect (y axis) as a function of distance from the anode (x axis), determined for an electric field of 273 V/cm (MicroBooNE’s operational field). Different coloured curves correspond to different electron lifetimes  $\tau_e$  and different purities. The red line corresponds to the lowest purity scenario for MicroBooNE design goal electron drift-lifetime (3 ms) while the black line corresponds to the highest purity and is close to the lowest purity observed in MicroBooNE during stable purity periods (18 ms). MicroBooNE has a maximum drift distance of 2.56 meters. In 100% pure liquid argon, electrons drift forever with infinite lifetime and no attenuation at any value of drift distance.

with the effective (quenched) amplitudes  $A'_S$  and  $A'_T$ .

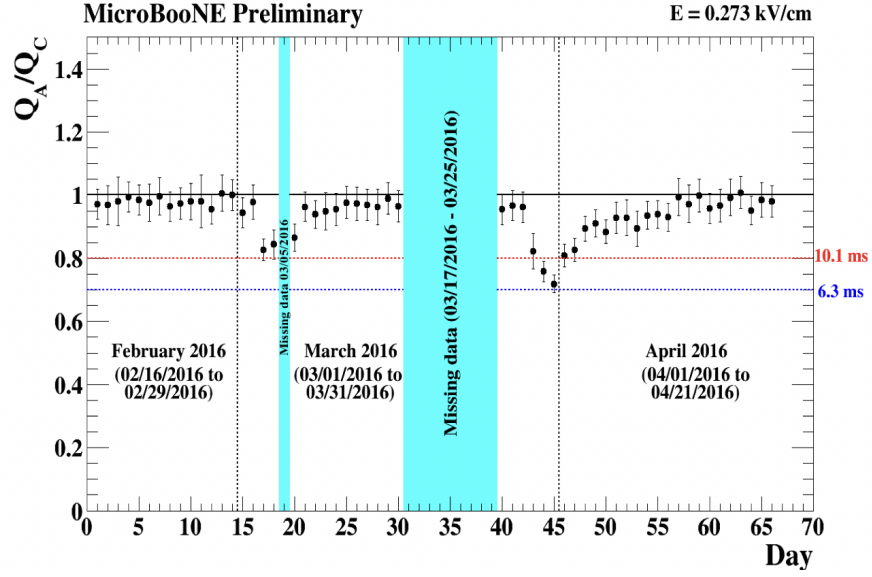
An overall *quenching factor*  $Q_F$  can be defined:

$$Q_F = A'_S + A'_T, \quad (3.8)$$

which represents the survival probability of the scintillation photons ( $Q_F = 1$  in pure argon).

Measurements of nitrogen and oxygen contamination effects in liquid argon have been performed by the WArP R&D programme [38] which has characterised the behaviour of  $Q_F$  with respect to increasing quantities of contaminants. Appreciable effects on the slow component have been found to appear above  $\mathcal{O}(1 \text{ ppm})$  nitrogen

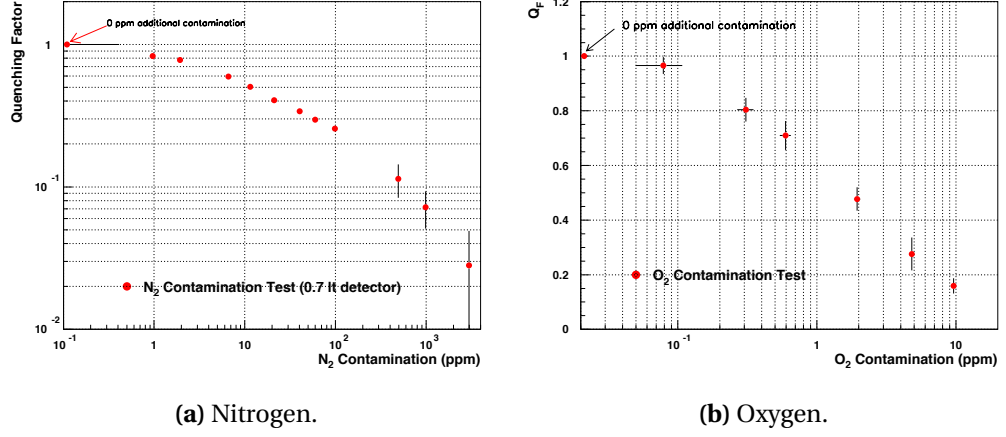




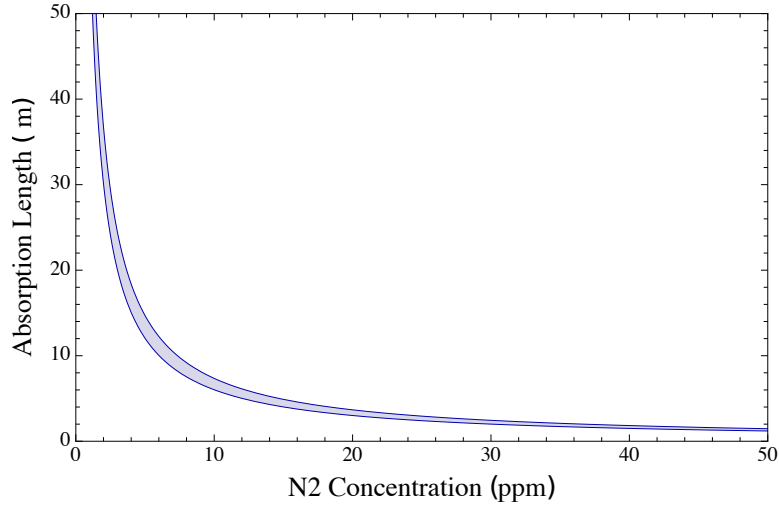
**Figure 3.11:** Variations in  $Q_A/Q_C$  (ionisation electron survival probability due to contaminant capture) as a function of time, for Run 1 period data. The horizontal red and blue line correspond to different value of electron lifetime, with data during stable periods observing a minimum lifetime of 18 ms. Vertical cyan bands corresponding to periods where data for argon purity studies was not available. The data shown contains space charge corrections (which are discussed in [section 3.3.4](#)).

concentration and  $\mathcal{O}(0.1 \text{ ppm})$  oxygen concentration.

Aside from quenching, which prevents light production, attenuation can be caused by absorption, which refers to the loss of the scintillation photon while in transit due to interaction with contaminant molecules (predominantly nitrogen). Unlike quenching, absorption affects the slow and fast components in the same manner, since the photon is captured after production. Measurements of absorption performed at Fermilab [47], shown in [figure 3.13](#), have found that the absorbing effect is relatively weak compared to other processes, with a loose requirement of 5 ppm (a much more stringent purity requirement of  $<1 \text{ ppm}$  is required to prevent quenching effects) nitrogen contamination guaranteeing more than 10 meters of absorption length. For reference, MicroBooNE’s maximum photon travel distance is  $\mathcal{O}(10)$  meters.



**Figure 3.12:** Quenching factor (scintillation photon survival probability) as a function of increasing levels of nitrogen (left) and oxygen (right) contaminations in liquid argon, measured by the WArP programme [38]. Appreciable effects are noticeable above  $\mathcal{O}(1$  ppm) nitrogen concentration and  $\mathcal{O}(0.1$  ppm) oxygen concentration.



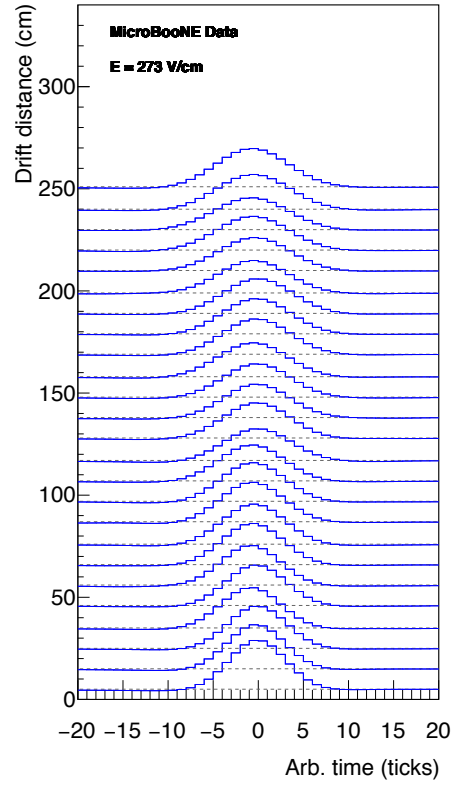
**Figure 3.13:** Measurement of absorption length as a function of increasing quantities of nitrogen contaminants by Fermilab [47]. The shaded region corresponds to the  $1\sigma$  bound for the measurement.

#### 3.3.3 Diffusion

A different type of motion perturbation that the ionisation electrons can undergo is diffusion. Diffusion governs the changes in electron trajectory due to various effects, such as inelastic collision, that cause the shape of a cloud of electrons moving col-

linearly in an electric field to grow in volume. Generally speaking, diffusion causes a smearing of the initial cloud of electrons, with its effects increasing with drift distance and with a non-isotropic behaviour (due to the vectorial action of the electric field). The immediate effects of such smearing are a worsening of both spatial and temporal resolutions.

Diffusion (and its effect on resolution) can be generally parameterised in terms of two components: transverse diffusion ( $D_T$ , perpendicular to the drift and electric field direction) and longitudinal diffusion ( $D_L$ , parallel to the drift and field direction).



**Figure 3.14:** Average waveform as a function of drift distance. All waveforms have been shifted on the  $x$  axis (1 tick  $\simeq 0.5 \mu\text{s}$ ) to be aligned to 0 and they have been shifted on the  $y$  axis to aid visibility. Lower waveforms have been obtained closer to the anode plane while higher waveforms are originated from electrons farther away from the anode plane and travelling for larger drift distances. It is possible to notice how larger drift distances cause the effects of diffusion to be more pronounced and the pulses to become more smeared.

For  $D_L$  we can imagine that while the electron cloud moves with a certain defined

drift velocity, the individual electrons do not. Some electrons will arrive earlier and some later, so that the Gaussian profile resulting from the response function on the wire to the charge deposition will be broadened. If we denote the original profile (for no diffusion) as  $\sigma_0$  and the broader profile (due to longitudinal diffusion effects) as  $\sigma$ , we can write the following expression:

$$\sigma^2 \simeq \frac{2D_L}{\nu_d^3} x + \sigma_0^2, \quad (3.9)$$

where  $x$  is the drift distance and  $\nu_d$  is the drift velocity (assumed to be 114 cm/ms in MicroBooNE). The effects of the extra drift direction dependent term are visible on [figure 3.14](#) where waveforms resulting from the collection of electrons originated at increasing drift coordinates are superimposed.

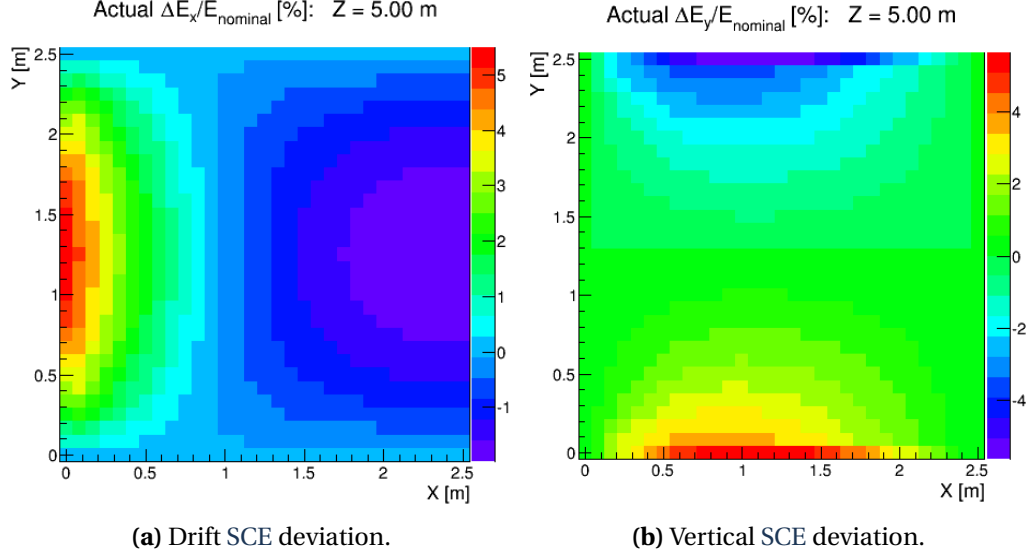
The transverse component is much more difficult to identify and measure. It causes a broadening of the cross section of the electron cloud with respect to the anode plane. Essentially, some of the electrons that would normally be measured on a wire  $N$  migrate on the neighbouring wires  $N + i$  and  $N - i$  with  $i$  being proportional to the magnitude of the  $D_T$  effect on the electron cloud. A particularly challenging task is to disentangle this effect from the generation of induced charge on a wire caused by a neighbouring wire.

#### 3.3.4 Space Charge

One of the last effects that can modify the trajectory of the ionisation electrons is the space charge effect (SCE). The assumption that the electrons drift towards the anode plane in a straight line and at a constant velocity is valid only if the electric field is perfectly uniform inside the TPC. Different field effects can cause distortions in the electric field and one of the most impactful is the presence of slowly moving positive ions which alter the intensity and direction of the electric field.

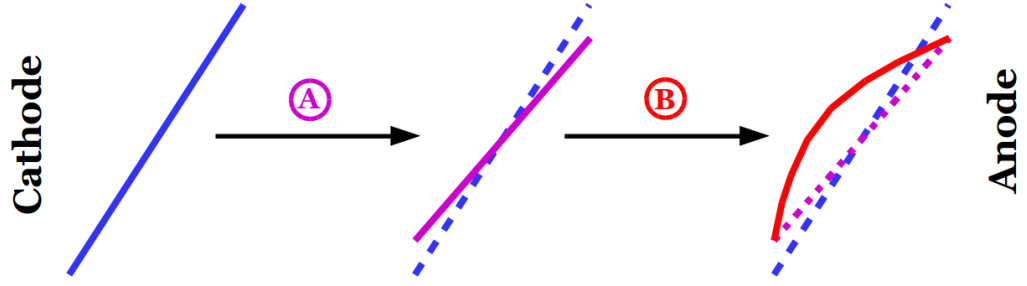
As previously stated, MicroBooNE is a surface detector with no overburden shielding, exposed to a high 5 Hz rate of crossing CR muons. The amount of positive argon ions generated by the CR tracks drifts relatively slowly, causing a build-up to arise predominantly near the cathode where the  $\text{Ar}^+$  ions drift. The build up is characterised

by a steady flow of ions which causes a distortion of the electric field. A simulation of the distortion caused by the electric field is shown in figure 3.15.



**Figure 3.15:** Variation of the electric field components due to space charge. Both figures show a view of the detector from the front panel ( $x$ /drift direction on the  $x$  axis and  $y$ /vertical direction on the  $y$  axis), for a cross section in the centre of the detector ( $z = 5$  meters). The left figure shows the variation of the  $X$  component of the electric field (the one parallel do the drift direction), the right figure shows the variation of the  $Y$  component of the electric field (perpendicular to the drift direction, parallel to the vertical). The areas with the most distortion are near the anode and on the top and bottom of the TPC. The electric field components are normalised to 500 V/cm. Figure is from Ref. [48].

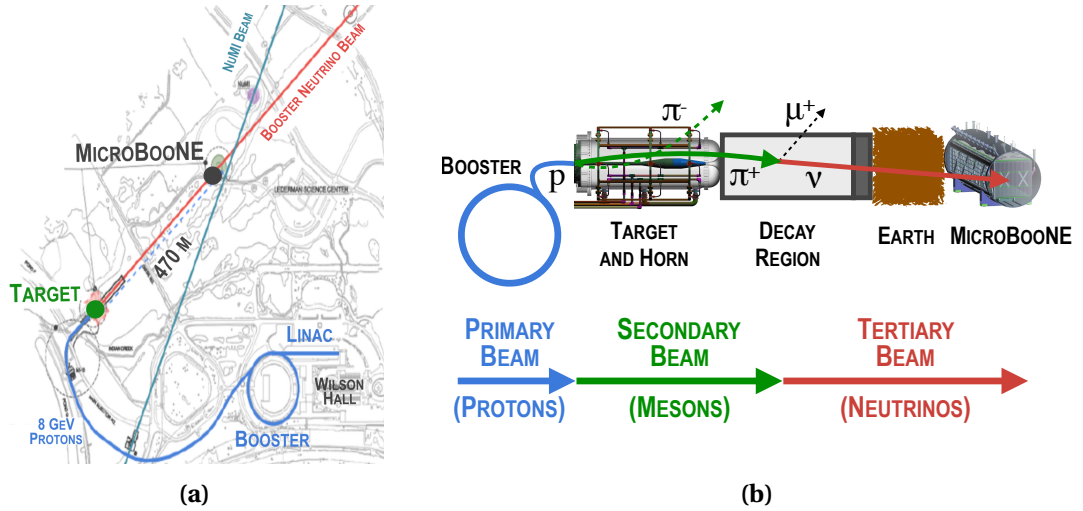
The effects of SCE are particularly important in the reconstruction of tracks where they can cause the tracks to appear “bent” and “rotated”. A diagram illustrating such effects is shown in figure 3.16. A method for calibrating the effects of SCE using data consists of measuring the effects of distortions on localised charge depositions at a known energy (e.g., MicroBooNE can use a pulsed UV laser to ionise liquid argon via multi-photon absorption). The measurements obtained can be used to generate a map of the electric field distortion due to SCE.



**Figure 3.16:** Diagram of the effects of space charge deviations on the reconstruction of tracks. Two distinct features are responsible for the deformation of tracks. A squeezing of the sides of the tracks, resembling a rotation (A) and a bowing of the track towards the cathode, most pronounced near the centre of the TPC (B). Figure is from Ref. [48].

### 3.4 The Booster Neutrino Beamline

The MicroBooNE detector collects data from neutrino interactions thanks to the constant exposition to a flux of neutrinos with energies  $\mathcal{O}(1 \text{ GeV})$  produced by the Booster Neutrino Beamline (BNB) at Fermilab. The Booster synchrotron accelerates protons to  $\approx 8 \text{ GeV}$  which subsequently impact on a beryllium target, producing secondary mesons which are focused by a magnetic horn and whose decay gives rise to a neutrino beam.



**Figure 3.17:** Map showing the respective locations of the Booster, target hall (containing the magnetic horn) and MicroBooNE (a). Diagram showing the primary, secondary and tertiary beams that compose the BNB (b).

### 3.4.1 Proton Beam

The production of the protons that are accelerated by the Booster synchrotron starts at the Fermilab Pre-Accelerator which contains two  $H^2$  beam sources. Hydrogen gas is ionised by means of a high-voltage arc. A cone-shaped electrode subsequently extracts  $H^-$  ions from the plasma, accelerating them to 35 keV [49]. The beam is subsequently *chopped* using an Einzel lens which periodically interrupts the beam; the purpose of this operation is to transform the continuous stream into a pulsed 100  $\mu s$ -wide structure at 15 Hz that can match the requirements of the radio frequency quadrupole cavities used for further acceleration to 750 keV.

The purpose of the LINAC linear accelerator is to accelerate the ions from 750 keV to 400 MeV via two series of RF cavities, so that they can be injected in the Fermilab Booster (see [figure 3.17a](#)). The 400 MeV beam is steered towards the Booster in a 2.2  $\mu s$ -wide time interval which corresponds to the revolution period of the protons inside the Booster (at 400 MeV).

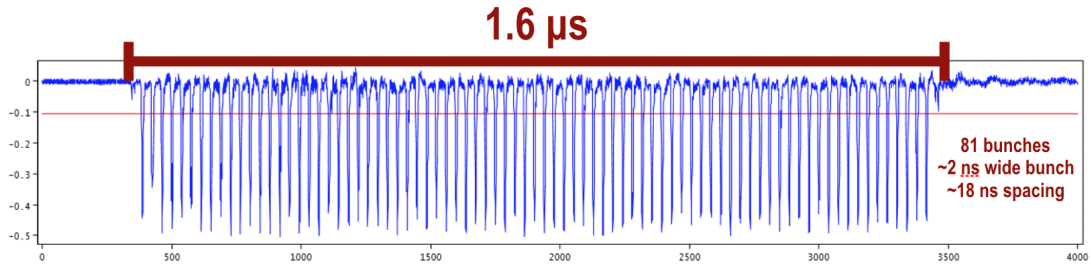
After they arrive at the Booster, the  $H^-$  ions are filtered through a stripping foil which removes the electrons from the ions, producing a proton beam. The proton beam can at this point go through a *paraphasing* process which captures the continuous beam into RF buckets. This step is needed because at injection, the proton beam is *unbunched* and it has a continuous structure so that cannot be accelerated. During the paraphasing process, the RF cavities accelerate some protons and decelerate others in order to align them in 84 *bunches*. The net acceleration is zero during this step.

The set of aligned 84 bunches is referred to as *batch*. After alignment, the resonant frequency of the cavities can increase (37.8 to 52.8 MHz) in order to accelerate the protons. This increases the energy of the beam which goes from 400 MeV at injection to 8 GeV at extraction. As the energy increases so does the revolution period which goes from 2.2  $\mu s$  to 1.6  $\mu s$ .

Kicker magnets provide a fast acting field that can deflect the beam for extraction. Although small, the ramp-up time of the magnet is non-zero. Extracting protons during ramp-up time would cause the extracted beam to be “sprayed” instead of being collimated. For this reason, a segment of the batch where three contiguous bunches should be is left empty, bringing down the total number of delivered bunches from 84

to 81. The empty segment allows the kicker magnet enough time to ramp up to full field intensity and send the beam off to the target.

The BNB delivers  $1.6\ \mu\text{s}$ -wide proton batches with a 15 Hz frequency on to a beryllium target. The 81 bunches that comprise the batch have a  $\approx 2\ \text{ns}$ -wide Gaussian profile with a  $\approx 18\ \text{ns}$  spacing between each bunch. Figure 3.18 shows the timing structure of a BNB batch. An important term used to describe a proton beam is *spill*. Spill refers to the timing structure of the extraction beam. In general, multiple batches can be part of a spill. Since the BNB is extracted directly from the Booster (i.e., does not use an additional synchrotron to accelerate the protons further) the timing structure of the spill matches the batch timing structure ( $\text{spill} = \text{batch} = 1.6\ \mu\text{s}$ ) and the two terms can be used interchangeably [50].



**Figure 3.18:** BNB batch/spill timing structure showing the 81 2 ns-wide bunches. From Ref. [51].

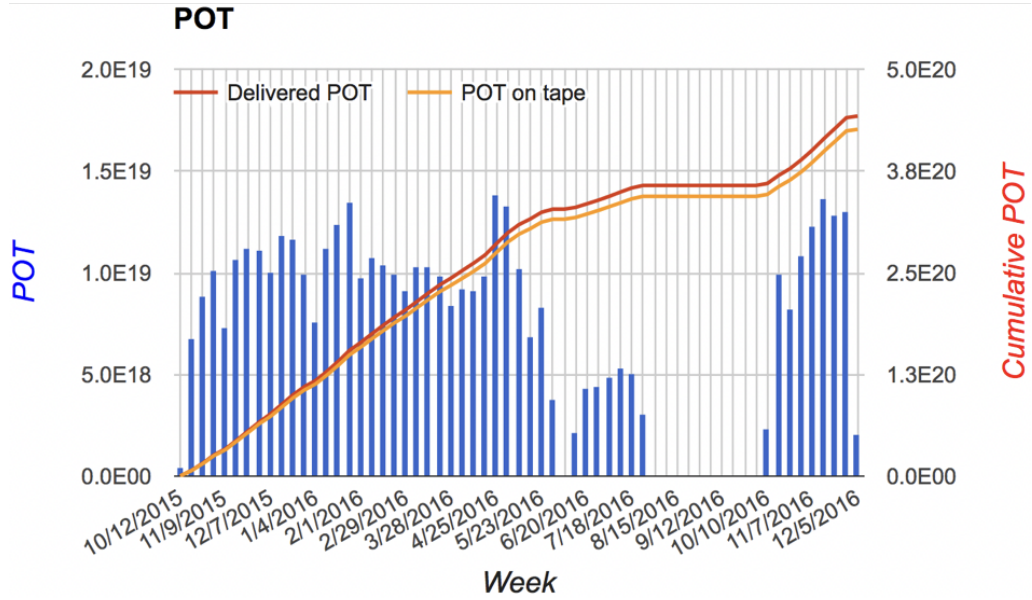
## Beam Toroids

The purpose of the beam toroids is to measure the amount of beam current that passes through its aperture which allows the measurement of the amount of beam sent to a particular area and to monitor the conditions of the beamline. Since the proton beam is equivalent to an electric current, the magnetic field surrounding it can impinge on the toroid and induce an electric current which is directly proportional to the beam intensity.

The use of toroids is important for neutrino analysis. The amount of protons delivered to the neutrino generating target (protons on target, POT) can be used as a unit of measure of data volume and as a tool for normalising different datasets containing neutrino interaction events to the same denominator. Figure 3.19 shows



the amount of recorded and delivered POT at MicroBooNE over the first 5 months of operation [50].

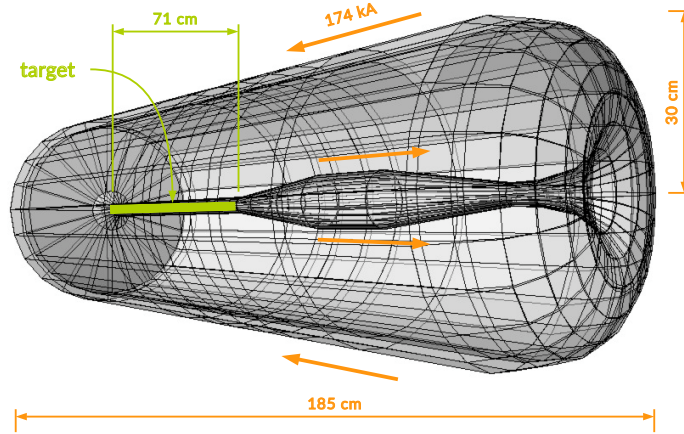


**Figure 3.19:** Recorded POT per week (blue) and cumulatively delivered and recorded POT (red and yellow) during the first 5 months of runtime. From Ref. [52].

### 3.4.2 Beam Target and Focusing Horn

The BNB uses a solid target upon which the 8 GeV protons impact, in order to create a secondary beam of short lived mesons (see figure 3.17b). The target is composed of seven beryllium cylinders, with a total length of 71.1 cm, corresponding to 1.7 proton interaction lengths. Upon impact, protons interact with it to produce a diffuse flux of pions and kaons with a wide angular distribution. In order to produce a collimated neutrino beam the kaons and pions need to be focused, which is performed by a horn electromagnet.

The focusing horn is a 185 cm long, aluminium alloy toroidal electromagnet which surrounds the beryllium target. Figure 3.20 shows the location of the target with respect to the horn. When pulsed, a positive current at 174 kA circulates in the toroids, generating a magnetic field reaching 1.5 T.

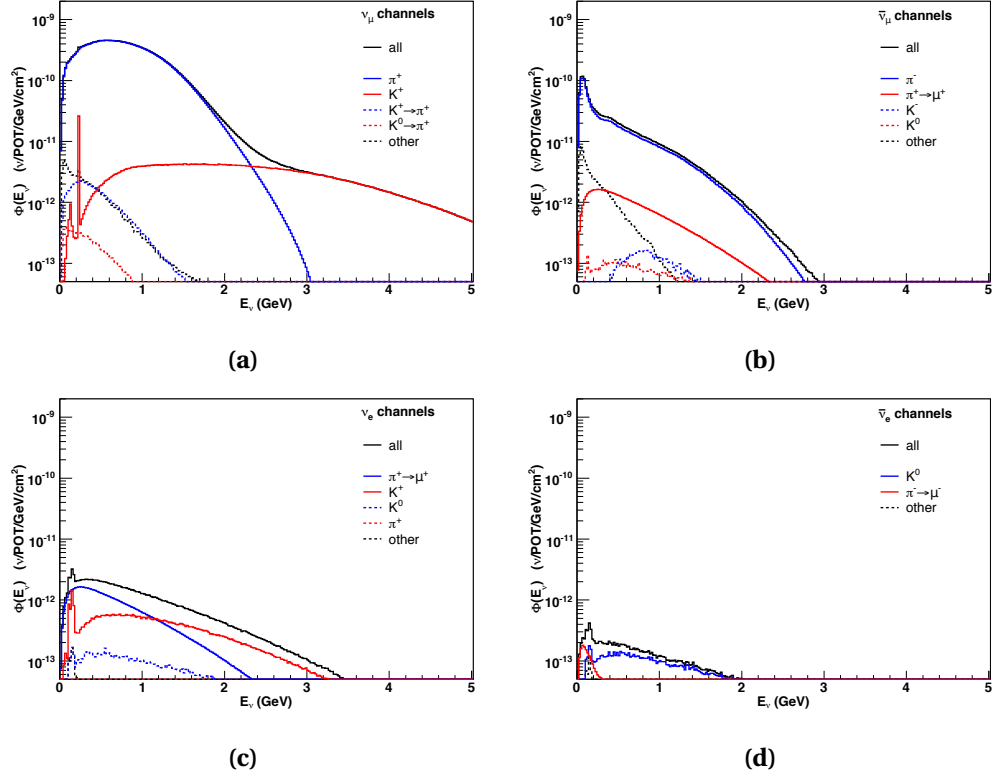


**Figure 3.20:** Model of the focusing horn. In green is shown the location of the beryllium target, in orange the direction of the electric field. From Ref. [53].

The pions and kaons trajectories, as they move through the collimator, are bent by the action of the magnetic field according to their charge. For a positive current, positive mesons are focused along the beamline, while negative mesons are diverted from it, resulting in a neutrino flux (as opposed to anti-neutrino). The polarity of the horn can be changed, in order to focus negatively charged mesons and produce an anti-neutrino beam. While the instantaneous delivered beam frequency is 15 Hz the target and horn need cool-down periods to avoid stress to the equipment, which brings the total average beam frequency to  $\approx 5$  Hz.

A collimator lets the focused mesons move to a decay chamber while a 2.14 m thick block of concrete absorbs all the particles which are not part of the secondary meson beam. The decay region which immediately follows is 50 m long cylindrical volume containing air. Pions and kaons decay in flight in this volume to mostly muons and neutrinos, producing the tertiary neutrino beam. A beam stop placed at the end of the volume absorbs all particles but neutrinos which can proceed through 470 m of dirt towards the detector [53].

Underground network fibers connect the horn hall to the MicroBooNE readout, along the 470 meters route separating target from detector. This allows the Fermilab



**Figure 3.21:** Components of the MiniBooNE flux with horn set in neutrino mode. The black line is the total predicted flux for  $\nu_\mu$  (a),  $\bar{\nu}_\mu$  (b),  $\nu_e$  (c), and  $\bar{\nu}_e$  (d). The coloured lines indicate the sub-component of the flux originated from a specific meson decay while the dashed black line groups all the lower contributions. Figures are from Ref. [53].

Accelerator Division to send trigger signals to the MicroBooNE DAQ, notifying that a neutrino beam is heading towards the detector.

### 3.4.3 Beam Composition

The neutrino beam resulting from the meson decay is largely composed of  $\nu_\mu$  with small contaminations of  $\bar{\nu}_\mu$  and  $\nu_e$ . Figure 3.21 shows the flux of neutrinos for different flavours at the MiniBooNE detector. The MiniBooNE detector is a predecessor of MicroBooNE located on the same beamline but at a closer distance, so that the flux shapes are very similar up to a  $1/r^2$  scaling factor.

Figure 3.21 highlights the importance of the helicity suppression mechanism (dis-

cussed in [section 2.2](#)) on the SM neutrino flux. The muonic ( $a$  and  $b$ ) and electronic ( $c$  and  $d$ ) components of the SM neutrino flux differ by several orders of magnitude. This effect is of particular relevance to the study of HNLs. As the neutrino mass increases the effect becomes less prominent. For HNLs, a helicity *anti-suppression* mechanism occurs with respect to the active neutrinos where the individual components of the BNB flux receive a boost proportional to the HNL mass.

# Triggers

The MicroBooNE detector is located on the surface with no overburden shielding and is thus exposed to a high rate of [CR](#) muons, crossing the detector at 5 kHz. An average of 24 crossing muons are expected in each event, due to the large readout window ( $\approx 4.8$  ms) which is necessary for the ionisation charge resulting from particle interaction to cross the width of the detector. A set of triggers is used to record detector activity from the smallest possible timing window surrounding an event and for filtering events that do not contain a signal, thus reducing background rate and disk writing.

A special class of triggers, called [HNL](#) triggers, have been developed, tested and implemented by the author in March 2017, and have been continuously operating since then. The [HNL](#) triggers are a special class of triggers tuned to exploit the different time of flight between [SM](#) neutrinos and [HNLs](#) in order to collect a data sample with no [SM](#) neutrino interaction background. This is the first application of such a technique in [HNL](#) searches and the data collected by these triggers constitutes the data analysed in this thesis.

In [section 4.1](#) the different classes of triggers used at MicroBooNE are outlined. [Section 4.2](#) describes the setup and the performance of the additional [HNL](#) trigger. [Section 4.3](#) lists the datastreams resulting from different combinations of triggers which will define the datasets used for the [HNL](#) analysis described in [section 6](#).

## 4.1 MicroBooNE Triggers

In accelerator neutrino experiments, the time  $t_0$  of the arrival of the neutrino spill is communicated in advance by the accelerator. While not all the neutrino spills lead to an interaction, the timing range in which the interaction might occur is well-defined. A trigger can thus be set up in order to require a condition in exact coincidence with the duration of a neutrino spill. Two classes of triggers are used in the MicroBooNE experiments:

- **Hardware triggers**, which decide whether to send the readout to the DAQ. This class of triggers depends exclusively on an externally issued signal. The BNB hardware trigger, for example, is issued by the accelerator whenever a neutrino spill is approaching the detector.
- **Software triggers**, which are applied to events passing the hardware triggers and consist of algorithms performing logic operations on the PMT output (e.g., a minimum amount of light has been produced and detected by  $> n$  PMTs) in order to decide whether an event should be saved or discarded. The purpose of software triggers is to discard all triggered events that are known not to contain events of interest.

### 4.1.1 Hardware Triggers

The hardware triggers are divided into beam, cosmic and supernova triggers. The beam triggers are issued whenever a neutrino spill approaches the detector. The signals issuing the BNB and NuMI hardware triggers are sent by the accelerator when the respective neutrino beams are sent to the detector.

In section 3.2.6 we have introduced the maximum disk writing speed which allows for a  $\approx 16$  Hz data rate. The BNB beam produces an average 5 Hz rate. The remaining 11 Hz disk write rate can thus be used to record CR muon background data which can be used for background subtraction. An external (EXT) trigger is used which is issued by an external pulser, issuing triggers at regular intervals: EXT-BNB and EXT-NuMI. The cosmic external triggers are designed to emulate the readout states of the two

beam triggers (BNB and NuMI) but to be issued exclusively *not* in coincidence with a neutrino spill (i.e., when no neutrino beam is arriving at the detector). These events thus contain exclusively cosmogenic activity with no neutrino interaction and can be used to characterise the background of the data collected by the beam triggers. By counting the number of issued hardware triggers of the beam and the EXT datasets, the normalised background from the EXT trigger can be subtracted from the beam data to obtain a signal distribution.

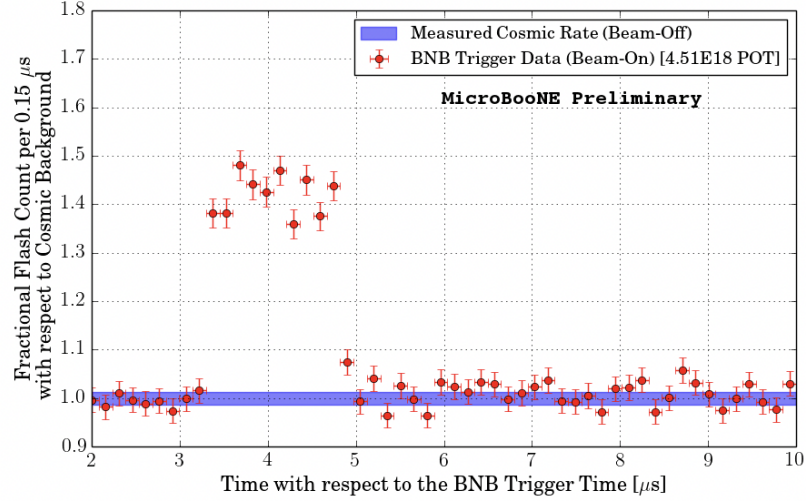
A separate class of hardware trigger is the supernova trigger which operates by using a constant circular buffer which stores approximately two full days of data. The Supernova Early Warning System (SNEWS) [54] is a network of neutrino telescopes constantly monitoring the skies for signs of a supernova. If at least two detectors report a coincidental observation of a supernova, a warning is issued to SNEWS. Upon receiving a SNEWS trigger, the data contained in the buffer are saved to disk, while in the case no trigger is issued the data get overwritten daily [55].

### 4.1.2 Software Triggers

Due to the low neutrino cross sections, only  $\approx 0.16\%$  of the neutrino spills traversing the detector contain a neutrino interaction. While the beam hardware triggers are issued for each neutrino spill, recording all spills would be both inefficient and expensive. In MicroBooNE a second category of triggers is employed, called software triggers.

The software triggers are configurable generic algorithms which run online and analyse each event after readout in order to decide whether that event should be saved to disk or discarded. They operate using information from the PMT system waveforms, performing logic operations on the amount of light observed in the detector and its distribution among the 32 PMTs. Figure 4.1 shows the distribution of flash times with respect to the trigger time. An excess in coincidence with the beam spill can be observed between  $3\ \mu\text{s}$  and  $5\ \mu\text{s}$ .

The BNB *software* trigger (not to be confused with the BNB hardware trigger), for example, requires a minimum of 6.5 photoelectrons (PEs) recorded by the PMT system in a timing window surrounding the neutrino spill in order to filter all the BNB events that contain exclusively cosmogenics (with a lower light activity). The BNB software



**Figure 4.1:** Distribution of flash times with respect to trigger time for BNB triggered events. The distribution is normalised to the expected average cosmogenic rate. The blue band centred at 1 shows the baseline for the cosmogenic background as measured with the *off-beam* CR muons data, with the width showing the measured uncertainty. The excess observed between 3 and 5  $\mu\text{s}$  corresponds to the flashes created by the interaction of neutrino produced by the BNB in coincidence with a spill.

trigger can be applied to data produced by the BNB *hardware* trigger, to filter events that do not contain a neutrino interaction, and by the EXT hardware trigger, to emulate the effects of the BNB software trigger on the BNB hardware trigger, thus making the two datasets directly comparable.

An additional sets of software triggers is the HNL triggers. The HNL triggers look specifically for delayed HNLs that reach the detector at the end of the neutrino spill. The PE threshold for HNL triggers is also higher (10.5 PE). This class of software triggers will be discussed in more detail in [section 4.2](#).

An additional configurable trigger parameter is the *prescaling factor*. The prescaling factor allows the user to configure the fraction of data written to disk without altering the other parameters of the trigger, thus allowing triggers with extremely high rate to be reduced to manageable levels. The prescale fraction is expressed as a number between 0 and 1 (e.g., for a prescaling factor of 0.7, 70% of data is saved to disk while the rest is discarded). This factor needs to be taken into account when normalising datasets to the same POT.



## 4.2 HNL Trigger

The  $\mathcal{O}(100)$  MeV mass of the HNLs and the  $\approx 500$  m distance travelled from source to detector have a non-negligible effect on the HNL time of arrival at MicroBooNE. SM neutrinos are produced at the BNB in a spill with a  $1.6 \mu\text{s}$  width (see figure 3.18). The time of arrival of the neutrino as a function of energy can be written as:

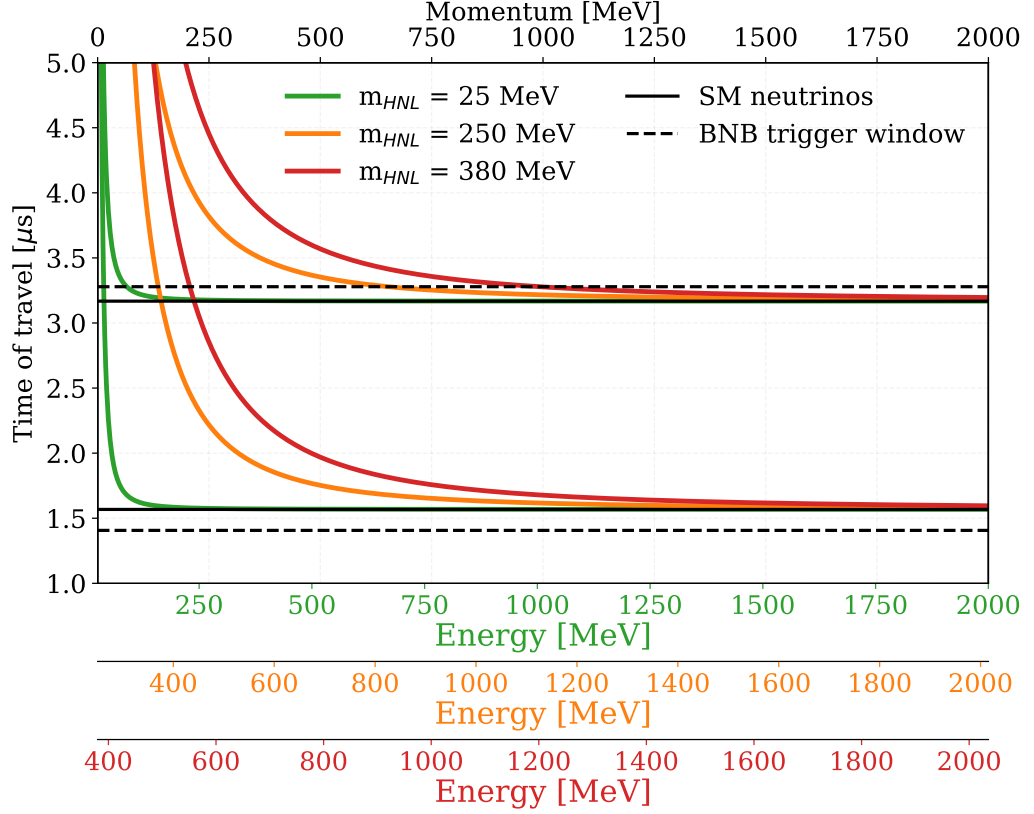
$$t(E) = \frac{470 \text{ m}}{c \sqrt{1 - \left(\frac{mc^2}{E}\right)^2}} + t_0, \quad (4.1)$$

where the numerator is the distance between the target and the centre of the detector and the denominator is the relativistic speed of the neutrino. For relativistic SM neutrinos, which have masses  $10^6$  smaller than their  $\mathcal{O}(\text{GeV})$  BNB energy, equation (4.1) reduces to  $470 \text{ m}/c$ , independent of the neutrino energy. Assuming the neutrinos are distributed uniformly in the neutrino spill timing range ( $1.6 \mu\text{s}$ ), the first neutrino in the spill would take  $1.54 \mu\text{s}$  from the beginning of the spill to reach the MicroBooNE detector while the last neutrino in the spill would arrive after  $3.14 \mu\text{s}$ . This defines a timing window in which the SM neutrino interaction is expected to occur. The BNB trigger operates in a slightly larger timing window ( $[1.4, 3.3] \mu\text{s}$ ), taking into account possible resolution effects.

HNLs are not ultra-relativistic particles. For the mass values probed in this work, they exhibit a non-negligible delay below 2 GeV with respect to the SM neutrinos that causes a fraction of the particles to reach the detector *after* the BNB detector has stopped recording.

Figure 4.2 shows the time of travel for SM neutrinos and HNLs. It highlights how the HNL time of arrival depends strongly on their initial momentum. For lower initial momenta, their speed becomes sub-luminal and the time required to travel from BNB to MicroBooNE increases exponentially. Events corresponding to an HNL decay with a time of arrival larger than  $3.3 \mu\text{s}$  are not recorded by the BNB software trigger. It becomes thus necessary to setup an additional trigger to record these events corresponding to *delayed* HNL arrival. Besides offering an increased event rate, an additional HNL software trigger would provide access to a dataset containing *late* HNL decays without any background originating from SM neutrino interaction. The most relevant

#### 4.2. HNL TRIGGER

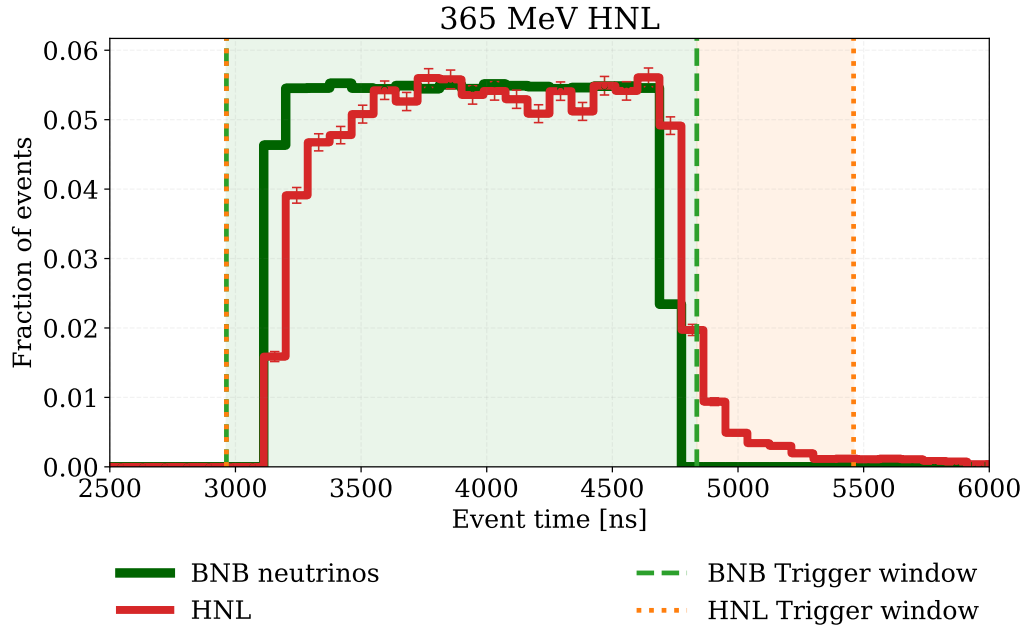


**Figure 4.2:** Time of travel for SM neutrinos and HNL of different mass, travelling from BNB to MicroBooNE detector. Black solid lines indicate time of arrivals for SM neutrinos, produced in a beam spill  $1.6 \mu\text{s}$  wide. Lines represent time of arrival for the first and last neutrino in the spill and neutrinos are expected to arrive at any time between the two solid lines. Their time of arrival is independent of their initial momentum due to their negligible mass. The black dashed lines indicate the start and end time of the BNB trigger, used to record SM neutrino activity. The coloured solid lines (and corresponding bands) indicate the time of arrival for HNL of different mass. The figure shows how the time of arrival depends on the HNL mass and initial momentum and how the BNB trigger increasingly loses efficiency for HNL with higher mass and lower momentum.

source of background for this dataset are only CR muons which are easier to reject, thus boosting the sensitivity for HNL searches.

For simplicity, we will use the word trigger to indicate *software* triggers, unless otherwise specified. The HNL trigger has been developed to perform the same logic as the BNB trigger on the PMT readout but with a higher threshold which is set at

10.5 PEs (instead of 6.5). The HNL trigger starts at the same time as the BNB trigger, 140 ns before the neutrino spill arrival, however it extends for an additional 624 ns after the end of the BNB trigger ( $1/3$  of the BNB spill duration). The exact fraction of HNL decays that fall in the extended timing window depends on the mass of the HNL and can be obtained by convolving the timing information from figure 4.2 with the decay energy spectrum of HNLs (which will be described in section 5.2). Figure 4.3 shows the distribution of the time of arrival for a Monte Carlo simulation of SM neutrino and HNLs with a mass of 365 MeV.



**Figure 4.3:** Timing distribution of the time of arrival of SM neutrinos (dark green solid) and decay of HNLs (for a mass of 365 MeV, red solid) at the MicroBooNE detector. The exact fraction of HNL events falling inside the HNL trigger depends on the HNL mass. The  $x$  axis shows time in nano-seconds from the start of the BNB hardware trigger. The histograms are area normalised and the dashed vertical lines correspond to the start and end time of the BNB (light green) and HNL (orange) trigger windows.

The decision on whether the event is saved by the HNL trigger is made based on the waveforms of the high-gain channel of each PMT. A low threshold of 10 ADC (0.5 PE) is applied individually to each PMT waveform to minimise noise. The sum of all the recorded PE from all the PMTs is then summed and this value is compared to

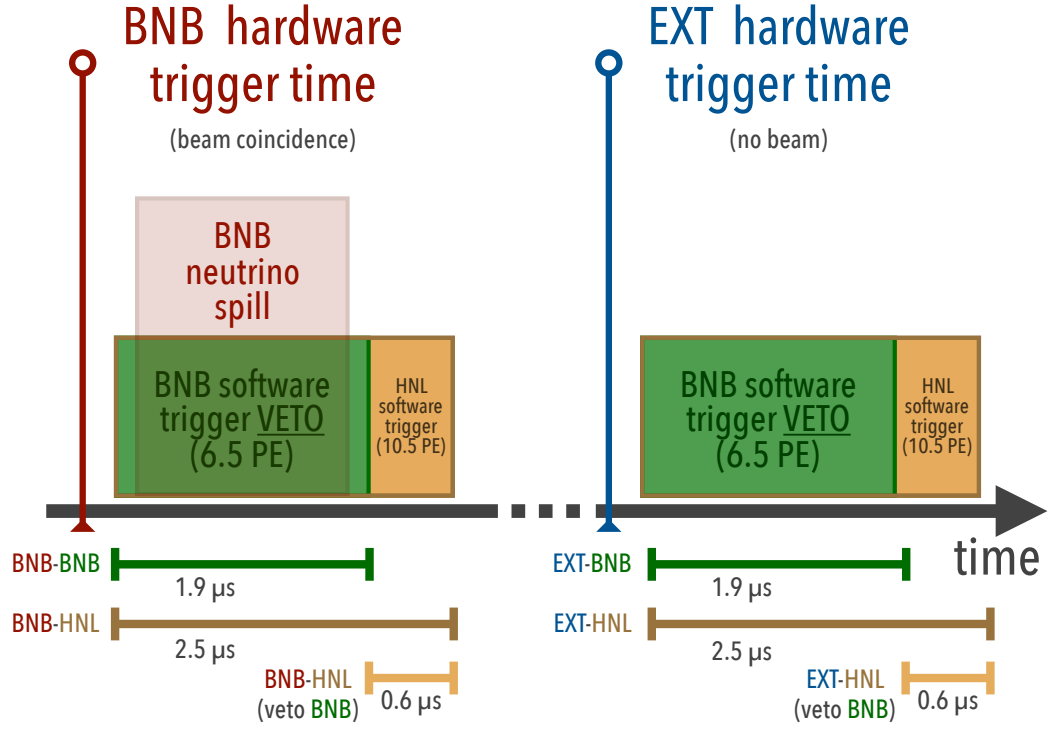
the total threshold. Table 4.1 shows the configuration of the HNL and BNB software triggers.

Software Triggers		
Variable	HNL Trigger	BNB Trigger
Timing window start	2.9 $\mu$ s	2.9 $\mu$ s
Timing window end	5.4 $\mu$ s	4.8 $\mu$ s
Timing window size	2.5 $\mu$ s	1.9 $\mu$ s
Single PMT threshold	0.5 PE	0.5 PE
Total PE threshold	10.5 PE	6.5 PE

**Table 4.1:** Trigger configurations for the HNL and BNB software trigger.

A special dataset is created from the events that pass the HNL trigger but not the BNB trigger, thereby applying a veto on any light activity above a PMT threshold of 6.5 PE before the start of the *extended* timing region, selecting events passing the trigger in the late timing region and vetoing light activity in the BNB timing region (figure 4.3). This rejects the background from SM neutrino interactions ensuring that only activity from the extended HNL trigger window is recorded. In the rest of this thesis we will refer exclusively to the data collected in the extended timing window on which the BNB software trigger veto has been applied.

An additional application of the HNL software trigger is to couple it to the EXT hardware trigger. The HNL software trigger is applied in a timing region *not* in coincidence with a beam spill (off-beam). This provides a data sample that contains the same background as the HNL software trigger without the signal and can be used to characterise the background. In order for the two samples to contain equivalent background (the HNL software trigger run with the BNB hardware trigger, in coincidence with beam spill and the HNL software trigger run with the EXT hardware trigger), both sample definitions include a BNB software trigger *veto* on the timing region preceding them. Figure 4.4 illustrates the relationship between the different hardware and software trigger types.



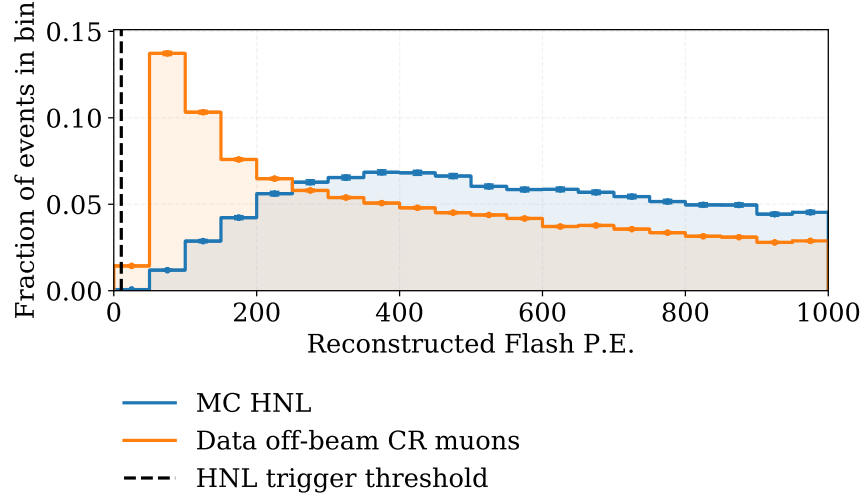
**Figure 4.4:** Diagram of the hardware-software trigger combinations.

### 4.2.1 HNL Trigger Performance in Monte Carlo

The HNL trigger has been developed, tested and implemented by the author in March 2017, shortly before the MicroBooNE detector shutdown taking place in summer, during which the beamline is not on. It is part of the standard array of triggers operating during standard runs and it is operating and continuously recording data as of this date.

The efficiency of the HNL trigger has been determined with MC simulations of the HNL signal. For all mass values tested, the PMT threshold is located away from the bulk of the light production distribution, so that the HNL trigger efficiency is  $\approx 100\%$ . Figure 4.5 shows the location of the HNL trigger threshold with respect to the PE distributions from reconstructed flashes for HNL and CR muons.

Although the efficiency for PMT light collection is  $\approx 100\%$ , effects due to the limited time resolution of the PMTs (15.6 ns) can lower the efficiency. These effects are



**Figure 4.5:** PE production distribution for reconstructed flashes from a MC HNL sample (mass 325 MeV, shown in blue) and an off-beam CR muons data sample (in orange). The figure shows that the HNL software trigger threshold (black dashed, at 10.5 PE) is located away from the bulk of the distribution for HNLs which provides 100% efficiency for HNL. The off-beam CR distribution drops in the first bin due to the limited efficiency of flash reconstruction with CR muons. Reconstructed flash efficiency for HNL is 100%.

relatively small, if they can be accommodated by adding a padding at the beginning and end of the trigger timing window. This is not possible with the HNL trigger for two reasons:

- The HNL time of arrival distribution is given by an exponential function. As a consequence, unlike SM BNB neutrinos, the timing distribution does not have a well-defined end time. This effect can be made arbitrarily small by extending the duration of the trigger window. Due to the exponential nature of the timing distribution, the number of HNLs failing the trigger conditions at the end of the tail becomes negligible.
- The action of the BNB trigger veto causes the HNLs arriving before the start time of the extended HNL trigger timing window to be rejected. Smearing on the time of arrival caused by the PMT timing resolution can cause HNLs around the start time to be rejected even if their time of arrival is expected to be within the

HNL trigger timing window. The exponential nature of the timing distribution, which starts shortly before the start of the HNL trigger window, makes the effect relevant for all mass values. This is shown in figure 4.6.

Taking into account all the above mentioned effect, the total trigger efficiency is estimated to be  $\approx 80 - 85\%$ , with the efficiency decreasing at larger masses due to the increasingly steepness of the exponential at larger masses. Figure 4.6 shows the HNL trigger pass fraction as a function of HNL time of arrival and the trigger efficiency as a function of HNL mass.

### 4.2.2 HNL Trigger Performance in Data

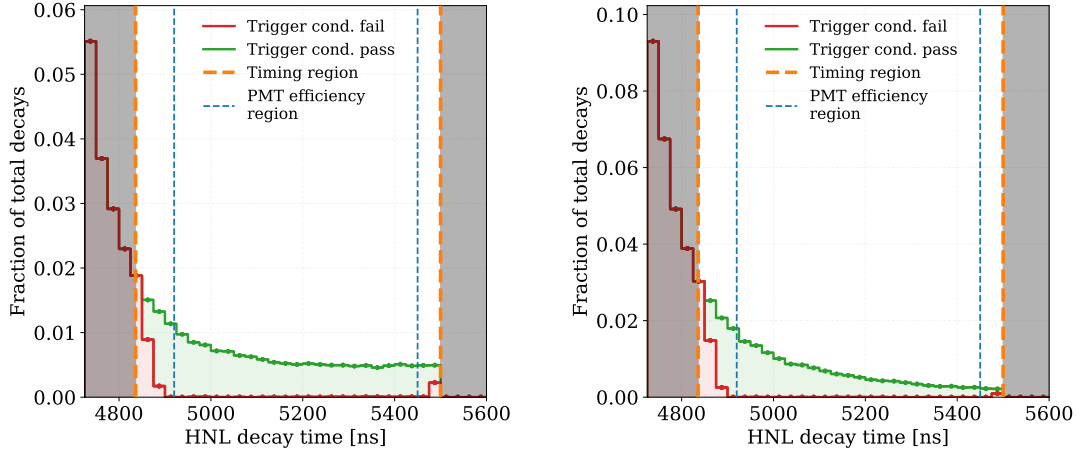
During the Run 3 of the MicroBooNE detector, for a period of seven months between October 2017 and June 2018, the HNL trigger has collected 217,000 events from  $5.6 \times 10^7$  issued triggers in BNB mode (on-beam) and 91,000 events from  $2.3 \times 10^7$  issued triggers in EXT mode (off-beam). Table 4.2 summarises the status of the HNL trigger during Run 3.

HNL Trigger Data	
Start period	March 2017
Number of issued triggers	$\approx 5.6 \times 10^7$
Equivalent POT	$2.18 \times 10^{20}$
Pass fraction	0.4%
Number of events collected	$\approx 217,000$

**Table 4.2:** Statistics regarding the *on-beam* HNL trigger (BNB-HNL). The data refers to the output of the “HNL veto BNB” trigger up to the date the data for this analysis was collected (June 2018). As of the date of writing this document, the HNL is running as part of the standard set of MicroBooNE triggers.

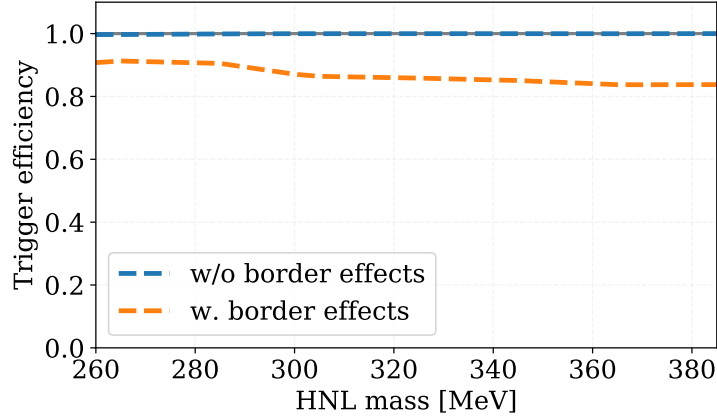
In September 2017, one of the 32 PMT was lost due to a faulty connection. The effects of the lost PMT on the trigger efficiency have been tested in MC, but the effects have been found to be negligible. A decline in light yield (measured as average PE/cm from CR muons) has been observed since the end of 2016, with a sharp decline in November and December 2016 and May and July 2017. Since September 2017 the light

## 4.2. HNL TRIGGER



(a) Efficiency for HNL with mass 285 MeV.

(b) Efficiency for HNL with mass 345 MeV.



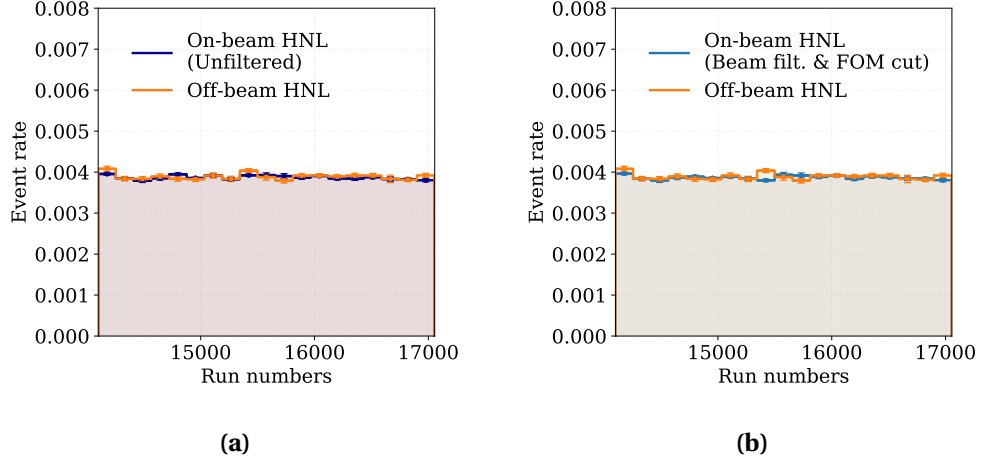
(c) Efficiency for HNL of different mass values.

**Figure 4.6:** HNL (veto BNB) trigger efficiency for HNL with a mass of 285 MeV (a), mass of 345 MeV (b) and for different mass values (c). The region between the two blue dashed lines in (a) and (b) indicates the timing region in which the trigger efficiency, based exclusively on the efficiency on light collection and without taking into account border effects, is estimated. Any event with a time-of-flight larger than 4,836 ns (indicated by the orange vertical dashed line) is used to determine the trigger efficiency taking into account border effects. Histograms in (a) and (b) are stacked, where the red histogram shows the events failing the triggering condition and green the event fulfilling it. Figure (c) shows the two types of determined efficiency for different mass values.

yield has stabilised, leaving the HNL data unaffected by the decline. Tests in the data collected by the HNL software trigger confirm that the on-beam and off-beam HNL



triggers have been running in stable conditions at the expected rate. Figure 4.7 shows the rate of events collected by the on-beam and off-beam triggers as a function of time.



**Figure 4.7:** Fraction of events passing the HNL trigger. Both the off-beam (coupled to the EXT hardware trigger, in orange) and the on-beam (coupled to the BNB hardware trigger, in blue) have run in stable conditions throughout the data taking period. (a) shows the event rates for all the events collected by the triggers, (b) shows the event rates after a selection on the beam quality (affecting only on-beam) has been applied, which shows the event rate unaffected.

## 4.3 Data Streams

In MicroBooNE we indicate with data streams the events processed by a certain combination of hardware and software triggers. Since the same name can be used to indicate both a hardware trigger, software trigger and data stream, we outline below the name used for the data streams analysed in this work. We indicate in parenthesis the hardware and software triggers combination that produces it (e.g., hardware-software):

- **On-beam BNB** (BNB-BNB), data stream using the BNB hardware trigger, fired in coincidence with the BNB neutrino spill, and the BNB software trigger, requiring 6.5 PE in the timing window surrounding the BNB neutrino spill in order to

increase the chance of an interaction occurring in the event. Despite the requirements of the hardware and software trigger the data streams can also contain events with no neutrino interaction, triggered by accidental cosmogenic activity generating enough light to satisfy the software trigger conditions. A “neutrino interaction sample” can be obtained by subtracting a purely cosmogenic (*off-beam BNB*) sample.

- **Off-beam BNB** (EXT-BNB), data stream using the EXT hardware trigger, fired at regular intervals *not* in coincidence with a beam spill and thus containing only cosmogenic activity, and the BNB software trigger which selects cosmogenic activity with a total 6.5 PE thus selecting only the CR muons that would be contained in the *on-beam BNB* data stream. The data collected by this data stream can be used to subtract background from data from the *on-beam BNB* data stream.
- **On-beam HNL** (BNB-HNL), data stream using the BNB hardware trigger, fired in coincidence with the BNB neutrino spill, and the HNL software trigger, requiring 10.5 PE in the timing window following the BNB neutrino spill in order to detect possible activity caused by a delayed HNL. Like the BNB data stream, this data stream can also contain events with no HNL decay, but triggered by accidental cosmogenic activity generating enough light to satisfy the software trigger conditions. This data stream automatically vetoes activity from the *on-beam BNB* data stream (i.e., if there is enough light to trigger the BNB software trigger in the timing window preceding the HNL software trigger, the event is rejected).
- **Off-beam HNL** (EXT-HNL), data stream using the EXT hardware trigger, fired at regular intervals *not* in coincidence with a beam spill and thus containing only cosmogenic activity, and the HNL software trigger which selects cosmogenic activity with a total 10.5 PE *after* the end of the neutrino spill thus selecting only the CR muons that would be contained in the *on-beam HNL* data stream. The data collected by this data stream can be used to characterise the background from the *on-beam HNL* data stream. This data stream automatically vetoes activity from the *off-beam HNL* data stream (i.e., if there is enough light to trigger the *off-beam BNB* software trigger in the timing window preceding the HNL software trigger, the event is rejected).

# Simulation and Data

This analysis involves the comparison of quantities measured by the detector (often *reconstructed* from low-level observables) and the same quantities generated by a model that needs to be tested, which are simulated and generated via Monte Carlo (MC) methods. In MicroBooNE, as well as other LArTPC programmes, a unifying generic analysis software called LArSoft takes care of MC generation, physics and detector simulation, as well as reconstruction of high-level variables from the detector measurements.

A custom LArSoft module called HNLGen, used for the generation of HNL events has been developed in LArSoft by the author. HNLGen takes care of the simulation of HNL decays and their products and it interfaces with the other modules of the LArSoft suite for further simulation.

LArSoft is also used for the analysis of both data and MC. Prior to reconstruction and analysis, the application of a data quality filter and data calibration is necessary. Data calibration is also a pre-requisite for the use of cosmic data in the generation of cosmic overlay MC samples. The author has performed the data calibration for the Run 3 data of MicroBooNE. This analysis is the first to make use of Run 3 data.

This section will introduce some of the LArSoft concept related to the HNL analysis in section 5.1. In sections 5.2 to 5.4 the MC generation of the HNL flux and decay events will be discussed, as well as other samples used for comparison. Section 5.5 describes the procedure used for calibrating the data. Section 5.6 describes how reconstruction is performed, both in MC and data, and in section 5.7 are described the different samples obtained from the trigger datastreams and used in this analysis.

## 5.1 Software Framework

The detector-agnostic software package **LArSoft** [56] is designed to provide a framework for the simulation, reconstruction and analysis of **LArTPCs** detector data. It is built on top of the Fermilab-supported Analysis Reconstruction Tool (*art*) [57], which allows access to event records and to build collections of events. **LArSoft** can be interfaced to external software packages such as GENIE [58], CORSIKA [59], GEANT4 [60, 61] and Pandora [62, 63] for, respectively, neutrino interaction, cosmic-ray shower, and particle propagation simulation, and for pattern recognition reconstruction. The foundation of **LArSoft** are native *art*-ROOT files, which are container files that can hold all the data objects representing a specific event such as MC “truth” particle trajectories (for simulation), arrays of wire and **PMT** waveforms and high-level reconstructed quantities such as particle identification (PID), direction and momentum. The file format is based on the CERN-ROOT [64] software package, thus allowing the use of ROOT’s native software tools. While **LArSoft** is detector-agnostic (the base **LArSoft** toolkit is shared across all **LArTPCs** experiments), specific subpackages are developed which are specifically tuned for the geometry and specification of the detector considered. In the case of MicroBooNE, the subpackage is called *uboonecode*.

All the MC simulated data and reconstructed data in this analysis have been processed with a family of product versions known as v8 or MCC8. Table 5.1 shows the products and respective versions used in this analysis. The external products are common across all MCC8 samples and their use is listed as follows:

- **GENIE** is used for simulating SM neutrino interaction with the MicroBooNE detector. It is not used for the MC HNL samples (since they contain only HNL decays). It is used however to generate a MC simulation of the BNB SM neutrino flux and BNB dirt (interaction occurring outside the TPC) samples, used for trigger and selection workflow validation, and background rejection estimation.
- **CORSIKA**, used for the simulation of air showers producing the CR muons crossing the detector. It is not used for the MC HNL samples, which take advantage of a novel background production method called *cosmic overlay* (described in more detail in section 5.2.3). It is used for the MC generation of cosmic-ray muons present in the MC BNB and dirt samples.

### 5.1. SOFTWARE FRAMEWORK

Application	Product	Version
Common to all samples	genie	v2_12_2
	corsika	v7.4003
	geant4	v4_10_1_p03d
	pandora	v03_00_00
MC HNL Simulation and HNL Data Reconstruction (1st part)	larsoft	v06_26_01_17
	uboonecode	v06_26_01_27
HNL Data Reconstruction (2nd part)	larsoft	v06_26_01_17
	uboonecode	v06_26_01_31
MC BNB Simulation + Cosmics + Dirt	larsoft	v06_26_01_11
	uboonecode	v06_26_01_14
On-beam and off-beam BNB Data Reconstruction	larsoft	v06_26_01_10
	uboonecode	v06_26_01_11

**Table 5.1:** Software package versions used in this analysis, which are part of MCC8. The application column lists the specific samples concerning a combination of products. The data samples are described in more details in [section 5.2](#) and [section 5.5](#). All external packages have a fixed version for the MCC8 samples (although not all samples necessarily use all the products). The HNL data sample has been split in two parts and reconstructed with two different uboonecode versions due to different availability of the time dependent data calibrations at the time of data processing. The versions differ only for changes to the MC generators and the availability of detector databases but do not affect data reconstruction performance.

- **GEANT4**, (also shortened *G4*), used for the simulation of the passage of particles through matter. It is responsible for generating the ionisation electrons and scintillation photons produced by the passage of secondaries and it is used for all MC-generated samples.
- **Pandora** pattern recognition software, responsible for interpreting the detector reconstructed quantities, such as discretised energy deposits on the wires, in terms of high-level information, such as **PID**, trajectory length and direction, etc. Pandora is used across all MC and data samples and will be described in more detail in [section 5.6](#).

The set of **LArSoft** products is designed to generate a simulation of the detector

response for a MC-generated sample that is an ideal representation of the output of the data collection and readout. MC-generated and data samples are then processed by a set of algorithms for reconstruction (discussed in [section 5.6](#)) and analysis ([section 6](#)). [Figure 5.1](#) illustrates the LArSoft simulation and reconstruction workflow and the relationships between the different products.

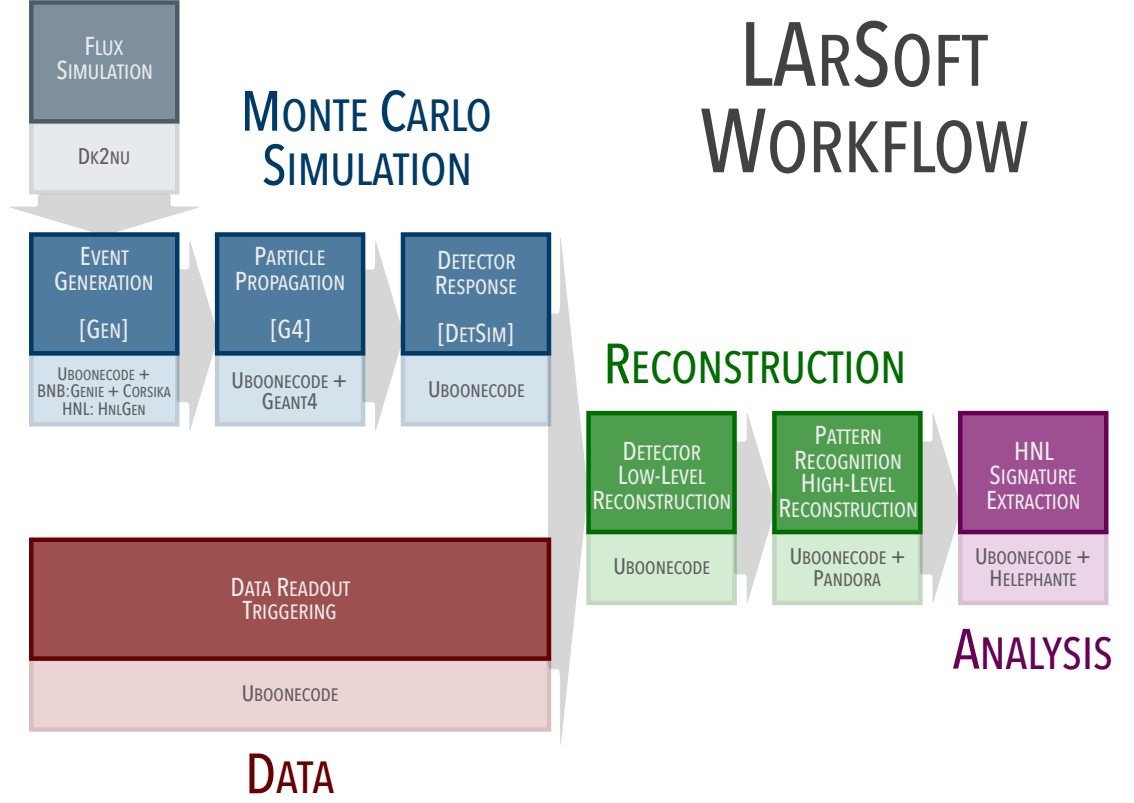
## 5.2 Generation

The simulation of the MicroBooNE detector consists of the three steps shown in [figure 5.1](#): generation of the secondary particles resulting from interaction or decay, propagation of secondaries through the detector and detector drift and signal collection simulation. We describe the three steps for what concerns the generation of HNLs, since this is the sample of primary interest, leaving the details of the generation of BNB SM neutrino and dirt interaction for [section 5.2.5](#) and [section 5.2.6](#).

### 5.2.1 HNL Generator

The Monte Carlo generator LArSoft module used for the generation of HNLs events is called HNLGen. It is based on the InFlight [65] HNL decay simulation software and has been modified for use in LArSoft by the author. HNLGen samples an energy spectrum distribution of the HNL decay (discussed in [section 5.2.2](#)) and generates a HNL object with an assigned energy and momentum. Unlike a SM neutrino interaction simulation, where a flux of SM neutrinos *traversing* the detector is provided and the simulation of the interaction is performed to produce an event rate scaled by interaction cross-section, for HNLs the flux provided is of HNLs *decaying* in the detector (not traversing it). In this case, the information regarding the decay is already convolved in the flux and the determination of the related scaling is performed previously, during the generation of the HNL spectrum. This procedure simplifies the interface with LArSoft.

Once a value of energy has been sampled from the HNL decay energy distribution, a momentum vector is assigned whose direction is chosen to be collinear with the beam propagation ( $z$ ) axis. In this case, the HNL momentum vector in the lab frame is



**Figure 5.1:** Summary of LArSoft MC simulation, data readout, reconstruction, and analysis workflow. Event generation takes care of creating data products for the “truth” level secondary particle produced by SM neutrino interaction or HNL decay. The propagation of the secondary particles through the detector, computation of the energy loss and relative production of *signal* particles (ionisation electrons and scintillation photons) is performed at the G4 stage. The *detsim* stage performs the propagation of the signal particles to the detector readout instrumentation, their collection, and the simulation of the LArTPC physics effects affecting such measurements (outlined in section 3.3). The output of the MC simulation stages is designed to be an ideal simulation of the equivalent output produced the actual data readout and triggering stages. Both streams (MC and data) can thus be processed for low-level reconstruction, which includes signal deconvolution and hit-finding, and high-level reconstruction, which performs pattern recognition. The output of the reconstruction stage is then passed through an analyser module tuned to identify HNL candidates which are subsequently analysed off-line.

uniquely determined for a sampled value of energy. The *delay* in the time of arrival is also uniquely determined for a determined energy value. A time  $t_{\text{decay}}$  of the decay is assigned to the HNL:

$$t_{\text{decay}} = t_{\text{travel}} + t_{\text{spill}} + \frac{d}{c} \left( \frac{1}{\beta} - 1 \right), \quad (5.1)$$

where  $t_{\text{travel}}$  is a fixed offset with respect to the BNB hardware trigger, which includes also the time of travel of SM neutrinos from target to detector,  $t_{\text{spill}}$  is a value randomly sampled from a uniform distribution  $[0, 1.6] \mu\text{s}$ , representing the fact that the neutrino might have originated from any bunch in the  $1.6 \mu\text{s}$ -wide BNB spill (see figure 3.18). The third term, which contains the target-detector distance  $d$ , the speed of light  $c$  and the relativistic Lorentz factor  $\gamma$  of the neutrino, is the delay experienced by the HNL with respect to the SM neutrino. The equation is written in this form in order to be consistent with the BNB SM neutrino generation in LArSoft, where only the first two terms are considered. In addition, the coordinates  $x$ ,  $y$ , and  $z$  of the decay are sampled from a uniform distribution, sampling a volume larger than the active LArTPC in order to take into account inefficiencies due to not completely contained events.

Having determined a four-vector of the HNL (energy and momentum) and a four-vector for its decay (spatial and time coordinates), we generate the two decay particles ( $\mu$  and  $\pi$ ) in the rest frame of the HNL. Since this is a two-body decay, the energy and momentum of the two particles is uniquely determined for a specific HNL mass. The HNL considered here is a Majorana neutrino. We simulate thus an isotropic decay with randomly sampled lepton direction and with equally split  $\mu^- \pi^+$  and  $\mu^+ \pi^-$  decays. Each individual decay is generally not isotropic. For a given charge of the lepton, the preferential direction in which it is emitted is HNL spin-dependent. However, changing the lepton charge flips the emitted angular distribution with respect to the HNL spin. As a consequence, measuring the combined decays of both channels returns an isotropic distribution in which the two angular dependencies complement each other. Since MicroBooNE is insensitive to the charge responsible for the energy deposit, it cannot distinguish between the two decay channels and it observes the sum of the two decay channels. For the simulation we neglect the polarisation effects and simulate the two decay channels with an isotropic distribution.

The  $\mu$  and  $\pi$  back-to-back momentum vectors are then Lorentz boosted into the



lab frame. The space-time and momentum-energy four-momenta are then added to the list of “truth” particles of the event (“truth” is a label used to indicate MC-generated objects and related quantities, as opposed to *reconstructed* quantities) which will then be processed by the G4 and detector simulation steps.

### 5.2.2 HNL Flux and Decay Spectrum

An important element of the HNL MC simulation is the generation of a flux and related decay energy spectrum to provide to the generator. To first approximation the HNL flux can be obtained by scaling the individual components of the SM neutrino flux (shown in figure 3.21) by the mass-mixing matrix element relevant to the lepton emitted in the decay (figure 2.1b) and by the kinematic enhancement factor (equation (2.4)) that takes care of helicity un-suppression and kinematic boundaries at mass threshold. In doing so however one neglects kinematic effects related to the Lorentz boost of the HNL that are different from SM neutrino production.

#### Flux Simulation

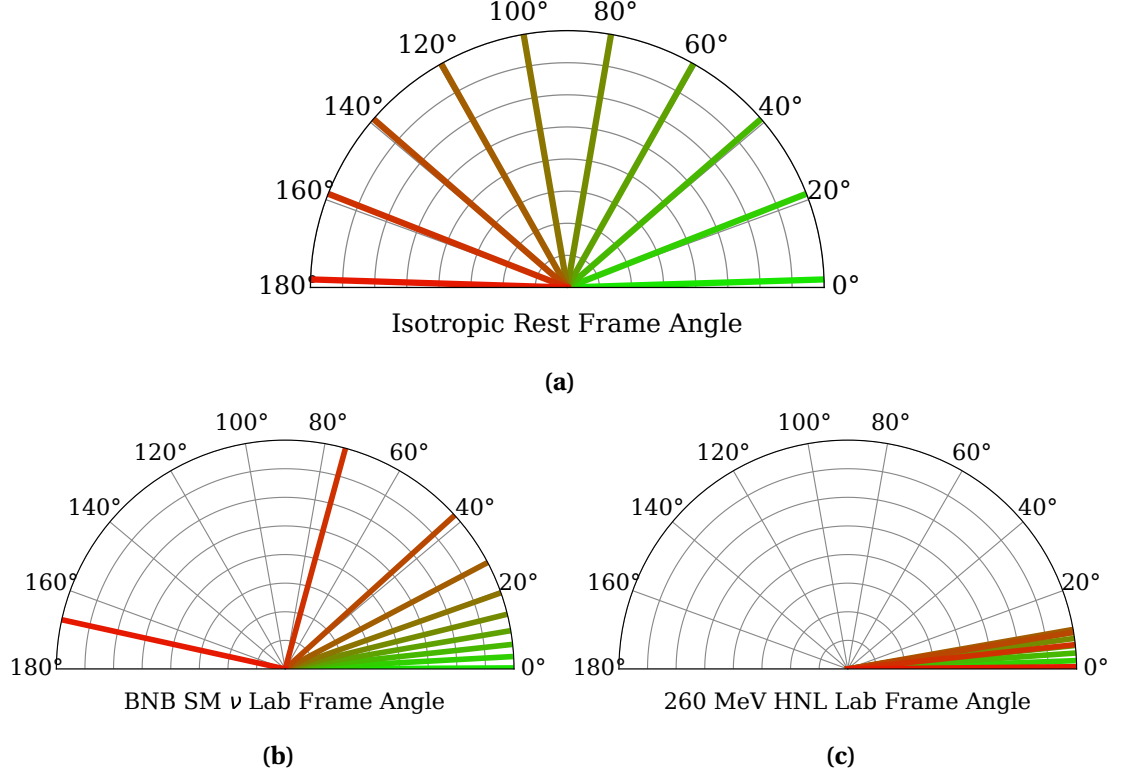
In the parent meson rest frame, SM neutrinos are replaced with HNLs to obtain an equivalent HNL isotropic distribution. At this stage, aside from the global scaling factors introduced by the helicity effects, the simulation is indifferent to the mass of the neutrino emitted. To obtain the SM neutrino flux the meson’s daughter particles are boosted in the lab frame and their trajectory propagated forward to verify intersection with the detector. Of all the simulated SM neutrinos only the ones intersecting the detector are included in the flux and their energy added to the energy spectrum.

For a parent meson of a given momentum directed towards the MicroBooNE detector, the angle  $\theta_{\text{meson}}$  between the parent meson direction and the produced neutrino in the meson’s rest frame can assume any value between 0 and  $2\pi$ , where for 0 the neutrino travels in the same direction as the parent meson and for  $2\pi$  in the opposite direction. By assuming a fixed parent meson momentum we can see that each value  $\theta_{\text{meson}}$  has a unique value of  $\theta_{\text{lab}}$  associated to it, where  $\theta_{\text{lab}}$  is the same angle between parent meson and daughter neutrino but boosted in the lab frame. The effect of a Lorentz boost produces a monotonic function, for lower angles in the rest frame,

lower angles result in the lab frame, due to having assumed the neutrino to be massless. Since we assumed that the meson is directed towards MicroBooNE, the intersection of the neutrino with the detector is then bound on the condition that  $\theta_{\text{lab}} < \theta_{\text{lab}}^{\text{crit}}$ , where  $\theta_{\text{lab}}^{\text{crit}}$  is the angle between the beamline and the line connecting the target and the detector edges. This translates into an equivalent condition  $\theta_{\text{meson}} < \theta_{\text{meson}}^{\text{crit}}$ , where only neutrinos produced sufficiently *forward* in the meson's rest frame can intersect the detector. **SM** neutrinos produced *backwards* in the parent meson's rest frame (i.e., away from the detector) are always travelling backwards independently of the parent original momentum since the **SM** neutrino mass can be assumed to negligible and there is no frame of reference that can reverse the momentum direction (in the massless assumption).

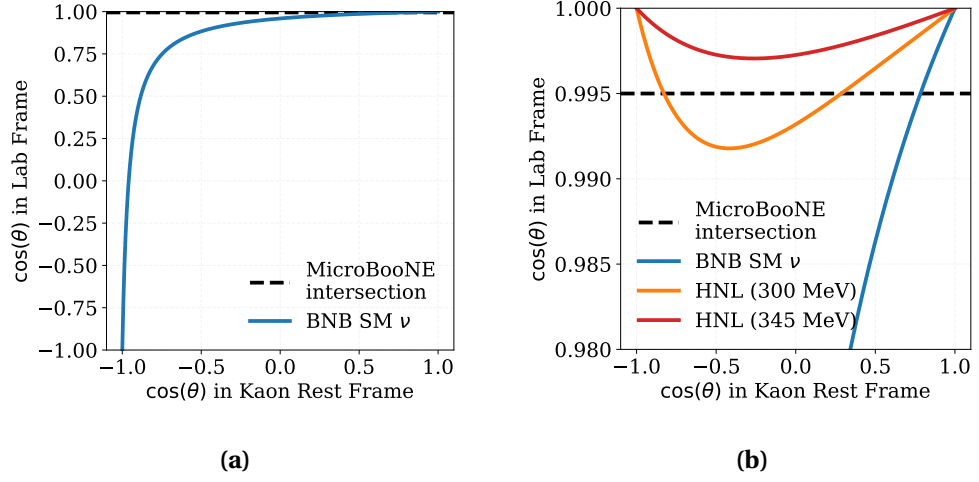
Due to the **HNL** mass however, the effect of a Lorentz boost into the lab frame produces a different relationship between  $\theta_{\text{meson}}$  and  $\theta_{\text{lab}}$ . **HNLs** have less transverse momentum available than **SM** neutrinos and the **HNLs** are more collimated, travelling preferentially in the parent meson direction. This means that if we assume the same  $\theta_{\text{meson}}$  for a **HNL** and a **SM** neutrino emitted in the same decay (i.e., they are emitted in the same direction),  $\theta_{\text{lab}}^{\text{HNL}} < \theta_{\text{lab}}^{\nu}$  which conversely implies  $\theta_{\text{meson}}^{\text{crit, HNL}} < \theta_{\text{meson}}^{\text{crit, } \nu}$ . This is equivalent to assuming that the critical angle in the parent meson's rest frame required for detector intersection is larger for **HNLs** than for **SM** neutrinos and that a larger fraction of **HNLs** are included into the flux, thus increasing the number of **HNLs** received, with respect to **SM** neutrinos, and softening the spectrum at the same time. Furthermore, the  $\theta_{\text{meson}}$  distribution for **HNLs** accepts two conditions for intersection due to the admission of **HNLs** that are emitted backwards. For a sufficiently low **HNL** momentum directed backwards, the Lorentz boost can flip the direction of the **HNL** directing it towards the detector. Figure 5.2 shows the effect of a Lorentz boost on the angular distribution in the lab frame. Figure 5.3 shows the requirement for interception with the MicroBooNE detector. The second requirement, for originally backwards going **HNLs**, correspond to low momentum particles and their inclusion causes an increase of the low energy peak of the flux spectrum, further softening the energy spectrum. The effects become more prominent at larger **HNL** masses and near the kinematic threshold where less transverse momentum is available and an increasingly fraction of softer **HNLs** reaches the MicroBooNE detector. For certain **HNL** mass values, the mass can be sufficiently high that all **HNLs** produced by a parent meson, regardless of their direction in the parent frame, intersect the MicroBooNE detector, provided

the parent meson is travelling in a direction sufficiently parallel to the beam direction. **Figure 5.4** shows the fraction of **HNL** and **SM** neutrinos emitted from a parent kaon that has a sufficiently low  $\theta_{\text{lab}}$  to intersect the detector.



**Figure 5.2:** Lorentz transformation of the  $\theta_{\text{meson}}$  angle between parent meson and daughter into  $\theta_{\text{lab}}$ . (a) shows equispaced directions in the rest frame of the parent meson, (b) and (c) their corresponding Lorentz transformed for the **SM BNB** neutrino and **HNL** case, respectively. The figures are produced for a **HNL** mass of 260 MeV and a parent kaon with a momentum of 1.68 GeV. The figures are different for different **HNL** masses and parent momentum considered, although higher masses of **HNL** increase the collimation effect [66].

The simulation is performed with `dk2nu`, using the same meson flux simulation as the **BNB SM** neutrinos and reweighting the mass of the neutrino to match the **HNL** mass studied. Energy distributions are produced for 10 different mass points. Since the mass-mixing matrix elements affect only the absolute normalisation of the flux there is no need for producing fluxes for different matrix element values. The same spectrum shape can be scaled to match different mass-mixing matrix element values. The work on the flux simulation and the production of the flux files has been performed by

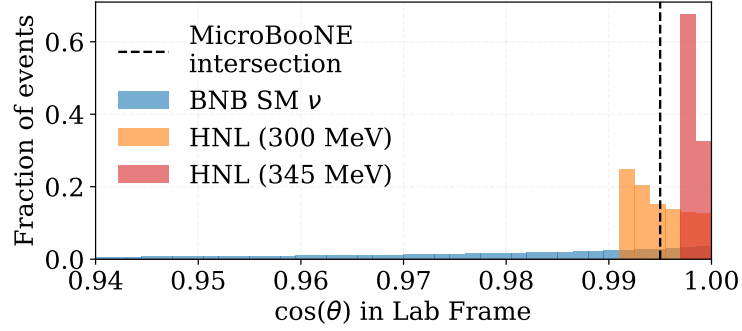


**Figure 5.3:** HNL and SM neutrino intersection angle for a kaon parent with a fixed momentum of 1.68 GeV. Figure (a) shows the relationship between the  $\cos \theta_{\text{meson}}$  ( $x$  axis) angle between the emitted neutrino and the parent meson, in the parent meson's rest frame and the equivalent boosted  $\cos \theta_{\text{lab}}$  in the lab frame ( $y$  axis). The relationship is shown for a randomly sampled meson directed towards the detector. It is thus for display only, as choosing a different meson would show a different relationship but analogous results. The intersection with the detector is guaranteed for a minimum  $\cos \theta_{\text{meson}}^{\text{crit}}$  value (shown by the horizontal dashed line). Neutrinos emitted with a value lower than this do not enter the detector and are not part of the simulated flux. Figure (b) shows a comparison for the same angles and parent meson between SM neutrino (blue) and HNL (orange and red). The  $y$  axis has been zoomed on to show better the HNL sample features. The larger mass causes the HNL to be more forward boosted. The critical value  $\cos \theta_{\text{meson}}^{\text{crit}}$  is smaller thus allowing a larger number of HNLs to intersect the detector. Furthermore a second condition for intersection with negative  $\cos \theta_{\text{meson}}$  (emitted backwards, away from the detector) is available, where the HNL has sufficiently low transverse momentum that the momentum projection on the beamline is flipped towards the detector, causing the HNL to reach MicroBooNE and be included in the flux. For a sufficiently massive HNL, any direction result in the HNL reaching the MicroBooNE detector [66].

Owen Goodwin. Sample distributions are shown in figure 5.5.

## Decay Energy Spectrum

The HNLGen LArSoft module takes as input a binned distribution of the number of decays occurring in the detector as a function of HNL energy. This is produced from



**Figure 5.4:** Probability distribution of lab frame angles for **SM** neutrinos and **HNLs**. On the  $x$  axis is the  $\cos \theta_{\text{lab}}$  angle between parent meson and emitted neutrino in the lab frame. On the  $y$  axis is the probability of a **SM** neutrino (blue) and **HNL** (orange) emitted isotropically by a kaon decay with a fixed momentum. The figure is thus for display only, since changing the parent meson momentum and direction would yield different results but with analogous conclusions. It can be considered as the marginal distribution of [figure 5.3](#) on the  $y$  axis. Regardless of the original direction of the **HNL** in the parent meson rest frame, due to Lorentz boost the  $\cos \theta_{\text{lab}}$  is always larger than 0.99 (for this particular case in study) causing a much larger fraction of **HNL** than **SM** neutrinos to reach the detector (the minimum angle necessary to reach the detector is shown by the vertical dashed line). Sufficiently massive **HNLs** always reach the detector [\[66\]](#).

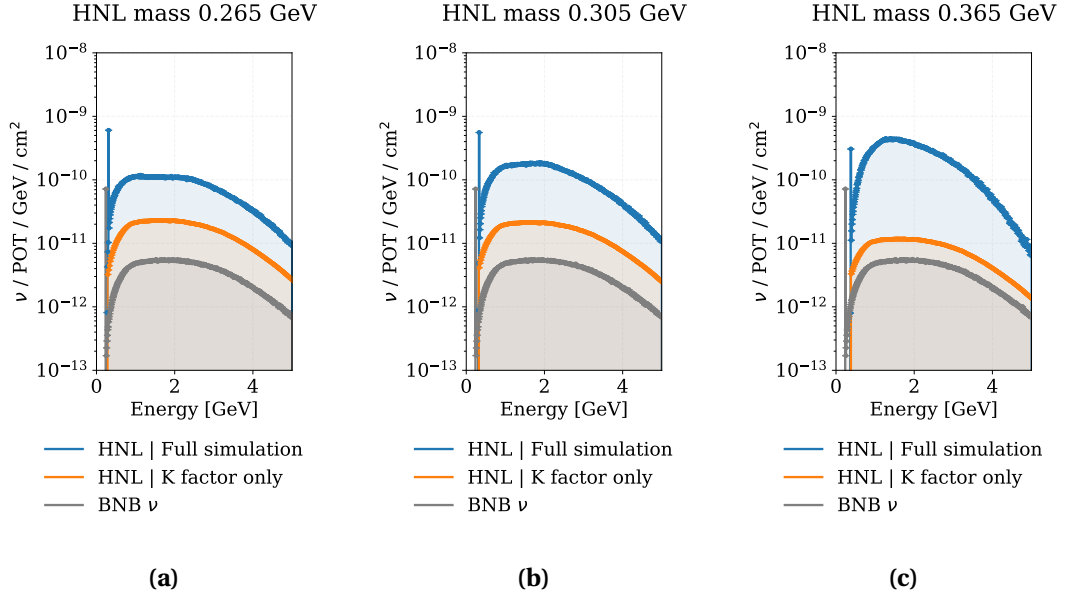
the **HNL** fluxes aforementioned by multiplying the **HNL** flux by the decay probability:

$$N_{\text{decays}}(E) \propto \phi_N(E) \times P_{\text{decay}}(E), \quad (5.2)$$

where we can write the decay probability  $P_{\text{decay}}(E)$  as the product of the probability of the **HNL** surviving to the detector, the probability of the **HNL** of decaying between the start and the end of the detector and the probability of the **HNL** to decay to the specific decay channel we are probing ( $\mu\pi$ ):

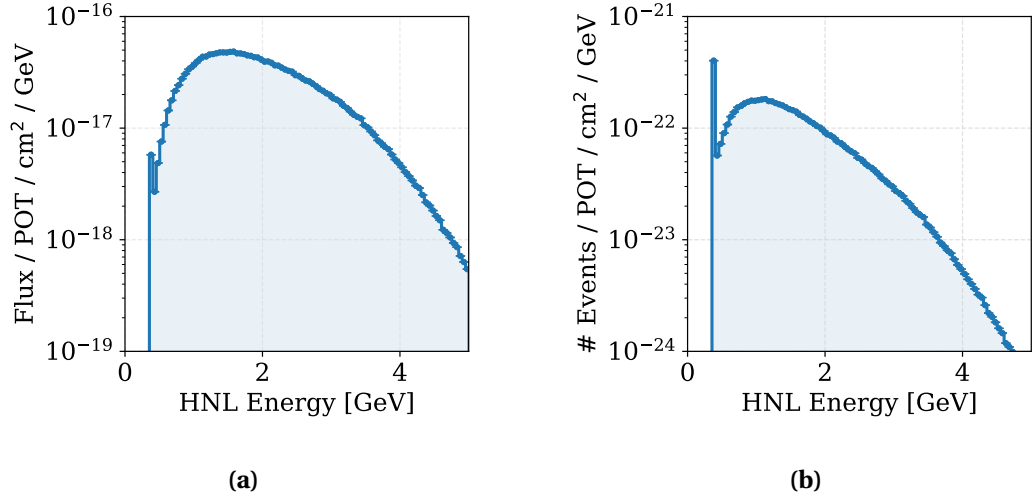
$$P_{\text{decay}}(E) = e^{-\frac{\Gamma_T d}{\gamma \beta}} \left( 1 - e^{-\frac{\Gamma_{\mu\pi} L}{\gamma \beta}} \right) \frac{\Gamma_{\mu\pi}}{\Gamma_T}, \quad (5.3)$$

where  $\Gamma_T$  indicates the total decay width,  $\Gamma_{\mu\pi}$  is the decay width of the  $\mu\pi$  decay channel ([equation \(2.7\)](#)).  $d$  is the distance to the MicroBooNE detector (463 m) and



**Figure 5.5:** Comparison of HNL and SM neutrino fluxes at the MicroBooNE detector for three different HNL mass values: (a) 265 MeV, (b) 305 MeV and (c) 365 MeV. The gray histogram shows the BNB SM neutrino flux for comparison. In orange is shown the HNL flux calculated using only the kinematic factor from [equation \(2.4\)](#) and neglecting the effect of the Lorentz boost. In blue is the HNL flux which contains both the kinematic factor and the full simulation of the HNL kinematic at production. The figures show how the kinematic effects become increasingly relevant at larger HNL mass values. The spike in the first bin is due to the decay at rest of kaons producing low-energy HNLs. The HNL flux are shown for a Majorana HNL, where the neutrino and anti-neutrino production channel are summed together and the mass-mixing matrix element  $|U_{\mu 4}|^2$  is set to 1, in order to make the comparison with the SM neutrino flux more readable (the electronic and tauonic coupling are assumed to be zero).

$l$  is the length of the detector travelled by the HNL (along the  $z$  axis, here 10.45 m, using the TPC definition). The energy spectrum of the HNL  $\mu\pi$  decays is shown in [figure 5.6](#), which shows the effects on the energy spectrum of the decay probability. Due to the decay probability dependence on  $1/\gamma\beta$ , the relevance of the lower momentum bins increases exponentially. This is drastically different from the case of neutrino interactions, where the dependence of the event occurring (cross-section) is directly proportional to the neutrino energy, thus favouring higher energy bins.



**Figure 5.6:** Comparison between the energy spectrum of the HNL traversing the MicroBooNE detector (a) and the HNL decays to  $\mu\pi$  occurring in the same volume (b). The latter energy spectrum is obtained by multiplying the former by the decay probability, shown in Equation 5.3. The spike in the first bin is due to the decay at rest of kaons producing low-energy HNLs. Both histograms are obtained for a 370 MeV HNL with a  $|U_{\mu 4}|^2=10^{-7}$ . Figure (b) shows the increasing importance of lower momentum HNLs due to decay probability increasing for smaller values of  $\gamma\beta$ .

### 5.2.3 HNL Cosmic Overlay

The HNL samples produced for this study are one of the first MC samples produced by the MicroBooNE collaboration that take advantage of a new first-of-a-kind MC generation technique: *cosmic overlay* (CO).

In standard MC samples, the CORSIKA Cosmic-Ray generator package is used to simulate the interaction of CR muons with the detector [67]. CORSIKA simulates the showers produced by the interaction of CR primaries with the Earth’s atmosphere and propagates the secondaries to the MicroBooNE detector, including the CR muons that cross the detector at 5 Hz. The MC generated particles produced by CORSIKA are then added to the list of MC particles generated by the signal generator (GENIE for SM neutrino interaction or HNLGen for HNLs) and the full list is then processed by particle propagator services and detector simulation modules.

An alternative method of including CR muons from real measured data has been

developed by the MicroBooNE collaboration in 2018. This is possible thanks to the accumulation of  $\approx 10^{21}$  POT of CR background events over the years of MicroBooNE operation. CO uses as initial input a real event obtained from a *off-beam* unbiased stream. That is, data collected by the EXT hardware trigger (*not* in coincidence with beam spills) on which no software trigger has been applied. This provides an unbiased sample of CR muons that matches the CR muons expected in a real signal event. Once the waveforms of the TPC and PMT systems are simulated for the HNL, they are *superimposed* on top of the waveforms obtained from the real data.

### Cosmic Overlay Workflow

The workflow for CO starts from a file containing *off-beam* unbiased events. All the MC CO datasets generated for this analysis use as input *off-beam* unbiased data from Run 3 (dataset defined as `prod_extunbiased_swizzle_crt_inclusive_v6`). The following conditions are imposed on the dataset, prior to use for overlay:

- *Run number*  $\geq 13697$ . This condition ensures that the *off-beam* unbiased dataset was collected during the same period of time as the data collected by the HNL triggers.
- *Runs contained in good runs list* (defname: `goodRuns_Run3_part1of2`). This condition ensures that all the runs have passed the data quality requirements (described in [section 5.5.2](#)), which filters runs in which the detector was operating in standard conditions.
- *Event count*  $> 40$ . Each of the files from the *off-beam* unbiased dataset contains a varying number of events. Files with a very small number of events affect the efficiency of data processing, since a number of operations (e.g. readout, file declaration) performed on the file use a similar amount of resources and time independent of the number of events contained in the file. This condition increases the efficiency of the workflow.

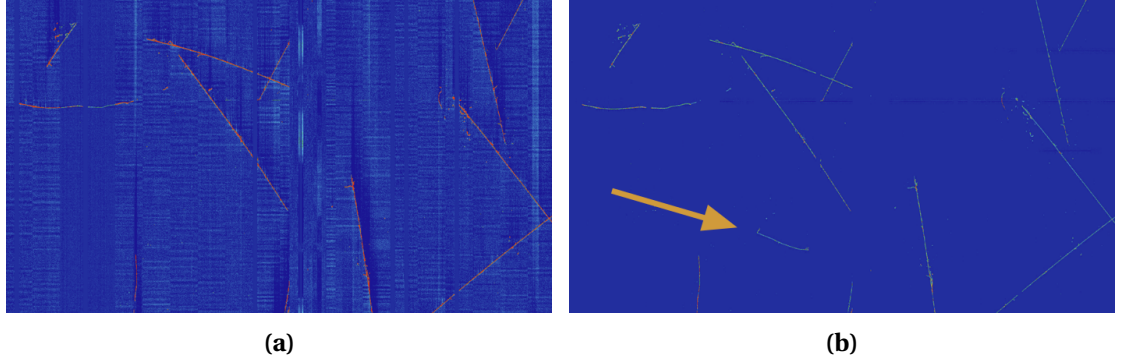
Each event from this dataset is taken as an input by the MC HNL generator. MC HNL “truth” particles are generated as per standard, and the relevant data products added



to the event in question. The propagation (G4) and detector simulation of the HNL daughter particles is performed in the same way as the standard MC generation with the exception of the detector noise which is *not* simulated. Since the waveforms from data, containing real data noise, are added on the MC generated waveforms, the MC CO event *inherits* the noise from the underlying data event. In addition, channels (i.e., wires) which are not operating optimally in the *off-beam* unbiased data are *masked* in both data and MC detector simulation. For the TPC wires, a database keeps track of the time-dependent status of each channel. Wires from two different planes can come into contact or the ASICs can saturate. The MC HNL waveform is suppressed for channels which are labelled as non-functioning by the database in the period of time in which the *off-beam* unbiased event was taken.

An operation of *pre-calibration* is also performed on the MC HNL data after detector simulation. The procedure of data calibration (outlined in [section 5.5.2](#)) produces a time-dependent 3-d map of  $x$ ,  $y$ ,  $z$  dependent *correction factors* which compensate for several of the detector effects described in [section 3.3](#). The correction factors can be inverted and applied to the MC generated signal in order to *degrade* the signal and simulate the detector effects (e.g. the deterioration of the signal as a function of increasing distance from the wire plane due to diffusion or attenuation). This avoids the use of specific models for the simulation of detector effects and the need for systematics representing possible mismodelling. The database of the gains used by each PMT is also queried and the MC PMT waveform are scaled by the data gain and re-mapped in order to match the data indexes, since MC and data use different PMT mapping. In the *overlaying* stage, the MC and data data products are merged. The MC generated waveform from each PMT and TPC is added to the corresponding one from data.

The resulting event consists of a MC HNL added to a real data event, which inherits all the properties of real data. Aside from containing real data CR tracks, thus avoiding the need to use MC CORSIKA CR and estimate model uncertainties, the dataset produced with CO contains real noise and real detector effects. [Figure 5.7](#) shows a data event display before and after the overlaying of MC HNL and signal deconvolution.



**Figure 5.7:** Display of an off-beam event (CR muons only) showing the raw waveform (a). The MC HNL signal is then overlaid on top of the data event and the signal on the wire is deconvolved producing the event display in (b). The overlaid HNL is indicated by the yellow arrow.

#### 5.2.4 HNL Sample Normalisation

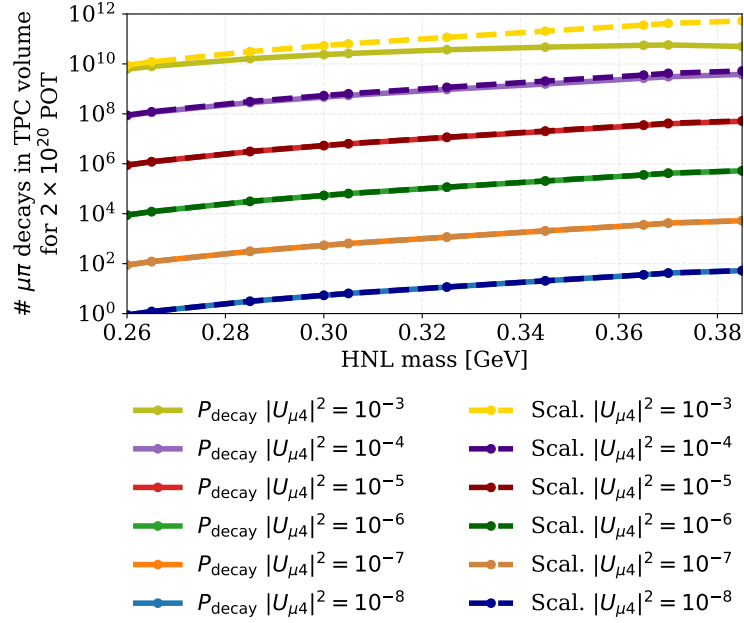
Samples from data and from MC simulation are often collected (or generated) for a different total number of events. In order to allow comparison between the samples it is necessary to provide a normalisation factor to renormalise the distributions of observables. For the generated MC HNL sample, we determine a number of decays occurring in the detector for  $2 \times 10^{20}$  POT and  $|U_{\mu 4}|^2 = 10^{-7}$  as a reference for each HNL mass value and scale accordingly for the POT and  $|U_{\mu 4}|^2$  needed.

We determine the number of decays as follows:

$$N_{\text{decays}} = \int_{m_N}^{\infty} dE_N \phi_N(E_N) P_{\text{decay}}(E_N) \text{POT } A, \quad (5.4)$$

where  $m_N$  is the HNL mass and  $E_N$  its energy.  $\phi_N(E_N)$  is the neutrino flux (shown in figure 5.5 for different masses),  $P_{\text{decay}}(E_N)$  is the neutrino decay probability (defined in equation (5.3)),  $A$  the cross-sectional area of the detector ( $2.60 \times 2.56 \text{ m}^2$ ) and POT is the number of protons on target. Figure 5.8 shows the calculated number of HNL decays for different values of the mass-mixing matrix element.

The LArSoft event generator is tuned to perform a selection on the HNLs that reach the detector in a timing window [4725 ns, 5500 ns] surrounding the HNL software trigger. This fraction is also included in the normalisation factor. Figure 5.9 shows

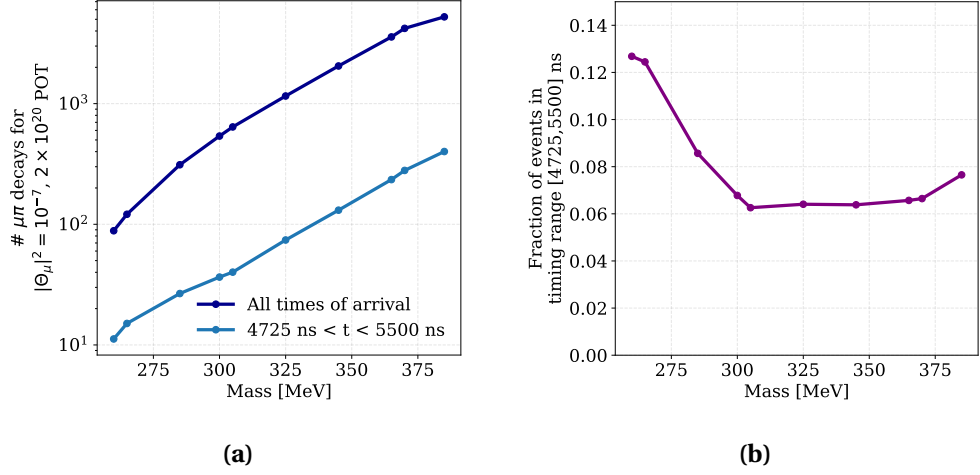


**Figure 5.8:** Number of HNL decays to  $\mu\pi$  for a Majorana HNL. The number of decays is considered in the TPC volume and for an exposure of  $2 \times 10^{20}$  POTs. In solid lines are shown the number of events calculated with the decay probability, using equation (5.4), while in dashed lines the number of events is calculated with the approximation of equation (2.8), which neglects the number of decays occurring before the detector and scales the number of decays by a  $|U_{\mu 4}|^2$  factor. The figure shows the number of decays scales with  $|U_{\mu 4}|^2$  up to values of  $|U_{\mu 4}|^2 = 10^{-4}$ . For higher mass-mixing matrix element values the detector is located at distances in the same order of magnitude as the decay length, and a non-negligible number of HNLs decays before reaching the detector.

the fraction of events falling inside the timing window surrounding the HNL software trigger.

### 5.2.5 Other Monte Carlo Samples: BNB Neutrinos

MC simulated events of SM BNB interactions are used as a control sample in comparison with BNB data to estimate the effects of event selection between data and MC. MC BNB simulation uses as input the simulated neutrino flux from the BNB. Simulated neutrinos are handled by the GENIE event generator which simulates the



**Figure 5.9:** Total number of HNL decays to  $\mu\pi$  pairs for  $|U_{\mu 4}|^2 = 10^{-7}$  and  $2 \times 10^{20}$  POTs occurring in the detector (dark blue) and number of decays occurring the timing window surrounding the HNL software trigger (a). In figure (b) is shown the fraction of decays from the total number that falls inside the timing window surrounding the HNL software trigger.

interaction of the neutrino with a target atom or nucleus. The GENIE version used in this analysis is v2\_12\_2. The interaction models used by this version are shown in table 5.2.

Model element	Implementation
Nuclear model	Bodek-Ritchie
Quasi-elastic	Llewellyn-Smith
Meson exchange currents	Empirical
Resonant	Rein-Seghal
Coherent	Rein-Seghal
Final-state interactions	hA

**Table 5.2:** Model implementation in GENIE v2\_12\_2.

For the overall nuclear model, the Bodek-Ritchie Relativistic Fermi Gas (RFG) is used, which combines a low energy tail from the standard non-interacting Fermi gas model with a phenomenological high-momentum tail observed in electron scattering measurements [68]. GENIE combines different models for each of the interaction modes. Quasi-elastic scattering events are simulated with the Llewellyn-Smith formalism, while the resonant and coherent pion production channels are simulated with the Rein-Seghal models. The hA model is used for final-state particle interac-

tion. The meson exchange current simulation is motivated experimentally, with the quasi-elastic cross section tuned such that the event rate matches those observed by MiniBooNE [58].

The MC BNB samples used in this work do not contain CO CR but use CORSIKA for the simulation of CR background. CORSIKA generates particles in the atmosphere and propagates the particle showers, tracking the particles which intersect with the MicroBooNE detector [59]. The list of MC “truth” particles resulting from BNB interaction and the CR CORSIKA particles are then passed to G4 for particle propagation in liquid argon.

#### 5.2.6 Other Monte Carlo Samples: Dirt

The *dirt* sample is produced in a manner analogous to the BNB sample. While MC BNB simulates neutrino interaction in the liquid argon, the dirt generation takes care of producing a sample of activity in the detector resulting from the interaction of neutrino with the material surrounding the TPC. The sample is called dirt, since the vast majority of particles generated outside the TPC originate from the dirt surrounding the hall in which the MicroBooNE detector is contained (the hall is below ground level). Interactions with the metal cryostat are also possible but less likely.

## 5.3 Particle Propagation

The output of GENIE, HNLGen and CORSIKA at particle simulation is a list of secondary particles resulting from either the interaction or decay of a neutrino and from the CR secondaries crossing the detector. The particle propagation step relies on GEANT4 to fully simulate the chain of daughter particles resulting from the interaction with the liquid argon (e.g., an electron produced by neutrino interaction can be tracked by GEANT4 in order to simulate the series of electrons and photons constituting the electromagnetic shower resulting from the electron bremsstrahlung).

Each of the daughter particles is tracked, and the energy deposit at each step recorded. The final output of the particle propagation stage is a list of particles which

includes their timing information, full trajectories and energy deposit along those trajectory. The GEANT4 module relies on a Geometry Description Markup Language (GDML) model of the MicroBooNE detector and its environment, including the internal composition of the TPC, cryostat, detector hall, and dirt surrounding it. At this stage, free electrons produced by the charged particles are instantaneously transported to the readout wires. The full simulation of the physics and detector effects affecting the drift are then taken care by the successive step, detector simulation.

Intermediate data product called *MCTrack* and *MCShower* are also produced at this stage, which are used in estimating the efficiency of the reconstruction algorithm. *MCTrack* and *MCShower* are still *truth* level particles, but they are convolved with electronic and detector effects, constituting the data objects that would be visible to MicroBooNE if the reconstruction were perfect. An example of *MCTrack* is the trajectory of a crossing CR muons, going through the detector. While the *truth* level particle data product would contain information regarding the muon trajectory throughout the whole environment, the corresponding *MCTrack* object contains only the portion of the trajectory going through the TPC for which ionisation electrons are produced and can drift to the wires. The comparison of *MCTracks* to reconstructed tracks provides a tool for estimating reconstruction efficiencies.

## 5.4 Detector Effects

The physics and detector effects listed in [section 3.3](#) affect the propagation of the ionisation electrons and scintillation photons responsible for the readout waveforms. These effects need to be fully simulated in standard MC generation to produce a simulation that agrees with data. The detector simulation includes the distortion of the tracks due to SCE, recombination, attenuation and diffusion.

The values for the parameters describing the physical effects are typically based on models or external measurements (the models and related uncertainties are described in more detail in the systematics discussion, in [section 8](#)). Few detector effects related to the electronics are not simulated at this stage. For example, the effects due to misconfiguration of the ASIC channels are added at the signal processing stage, by deconvolving the signal with an incorrect response and re-convolving with the

standard one.

One of the effects which is missing from the iteration of the MC software version used in this analysis (MCC8), is the detector simulation of the induced charge on the wires neighbouring the channel collecting the electron, known as Dynamically Induced Charge (DIC). The most direct consequence of the lack of this effect in simulation is a poor agreement in data and MC for the charge measured by the two induction planes. MCC8 analyses focus on calorimetry information obtained from the collection plane only, which shows better agreement in data and MC. An additional limitation is also introduced in the angular phase space. Tracks perpendicular to the wire plane, and thus being represented in reconstruction by a single or few channels are reconstructed with more difficulty. While the effects of DIC on calorimetry do not affect this analysis, which focuses on topological features, the effect on reconstruction efficiency are taken into account in the systematics treatment (discussed in [section 8](#)). Implementation of DIC effects is one of the collaboration priorities and will be included in the next iteration of the simulation software (MCC9).

One of the differences between standard MC detector simulation and CO detector simulation is that in CO the noise is not simulated. Noiseless waveforms with signals only from particle are generated, and the resulting waveforms are then added to the waveforms from the underlying real data event. In both cases, the main output of the detector simulation stage is a series of waveforms for each TPC and PMT channel of the MicroBooNE detector, in an analogous way to the output of the real datastreams.

## 5.5 Data

The data collected by the triggers, introduced in [section 4](#), is constituted of ROOT files containing the raw waveforms from the TPC and PMT systems, data from the triggering system, and meta-data pertaining each event (e.g., run number, POT number, etc.). Datasets in MicroBooNE are divided according to the run in which they were taken, and the combination of hardware and software triggers producing them.

### 5.5.1 Data Quality

Ensuring that the quality of the individual runs collected by the MicroBooNE detector for physics analysis is up to standard is an essential step of any analysis chain. The beam and data quality requirements ensure that no anomalous behaviour is exhibited by the MicroBooNE detector, and if that is the case, that such anomalous runs are excluded from a physics analysis dataset. This is performed by applying a series of filters that query a database of beam and detector operational status.

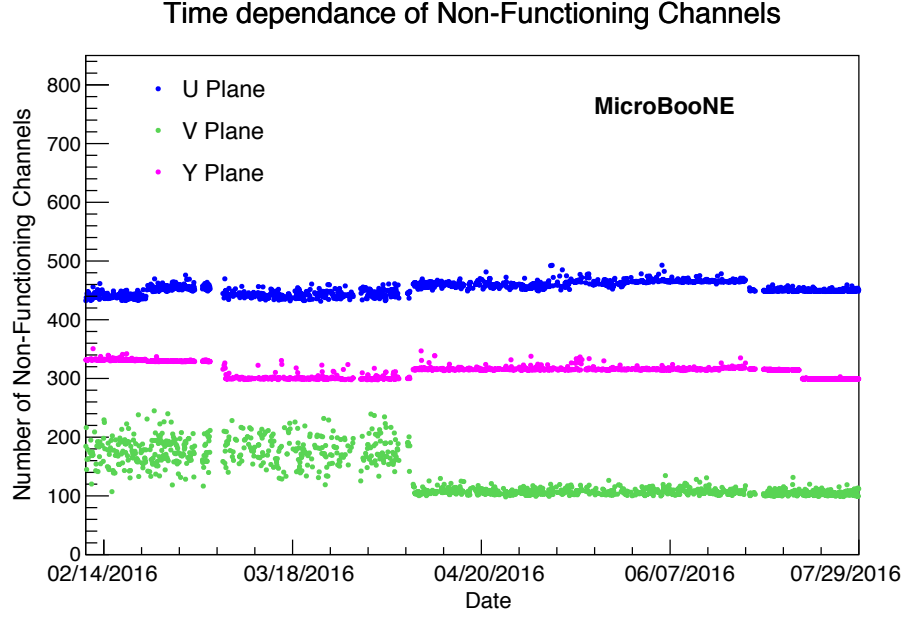
#### Channel Status

A pre-requisite for data quality monitoring is the determination of a set of (run-dependent) status labels and corrections to be assigned to each readout channel before signal processing. That is because the variables used for data quality measurements use reconstructed quantities and it is thus necessary to ensure that the operational status of each channel is taken into account before performing signal processing and event reconstruction.

Diagnostics of MicroBooNE data have found the detector to have  $\approx 10\%$  of non-functional channels, with various causes of failure attributed to their non-operational status. A comprehensive list of possible sources of channel malfunction is contained in Ref. [69]. These causes need to be fully accounted for and this operation can be divided into the determination of a database of *bad channels* (labelled as such mostly due to dead ASICs, excessive noise and/or chirping), and the determination of a list of *misconfigured channels*, found by analysing average hit amplitudes. Figure 5.10 illustrates the average number of non-functioning channels as a function of time over five month of data from Run 1.

Misconfiguration of MicroBooNE channels is caused by a default setting of the ASICs that the channels feed into. The 8256 MicroBooNE TPC readout channels feed into 516 analog front-end ASICs, with 16 channels per ASICs. Among the different settings available, a gain and peaking time of 14 mV/fC and 2  $\mu$ s were determined to be optimal. However, raw data analysis in the time and frequency domain determined 224 sub-optimal channels (belonging to 14 ASICs) generating electric noise consistent with

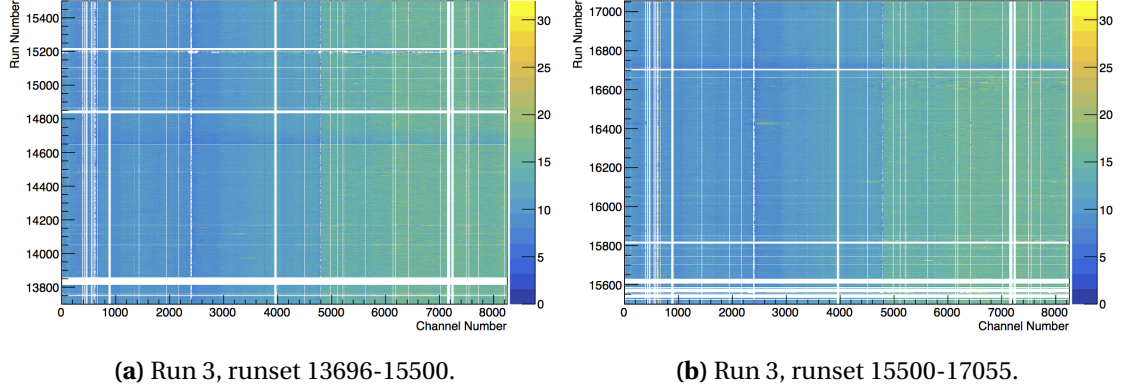




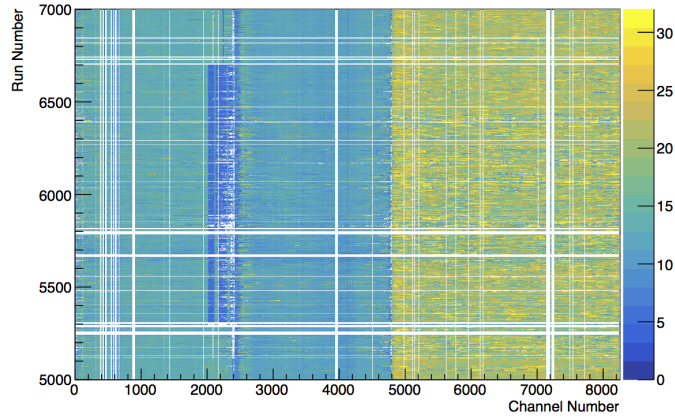
**Figure 5.10:** Average number of non-functioning channels over  $\approx 10$  events per run as a function of time, from Run 1 data. The discontinuities in the U and V planes are due to two ASICs (32 channels) being temporarily recovered. The difference in behaviour between the V plane and the other two is due to the effect of ASIC saturation affecting in particular the V plane. After the replacement of a faulty cryogenic pump in April 2016 the occurrence of ASIC saturation becomes negligible. Figure is from Ref. [69].

the factory settings of the ASICs (4.7 mV/fC and 1  $\mu$ s). The cause of this misconfiguration is believed to be due to damage to the configuration signal lines from electrostatic discharge during the installation of the detector. Misconfigured channels are analysed using the average amplitude of the Gaussian curve fitted to the energy deposition candidates on the channel waveform, each defining a *hit*. This was performed for each channel in a run and then averaged over half of an ASICs (8 channels) expecting the average amplitude to be relatively constant across all channels in a plane. The reason for averaging over 8 channels instead of 16 (number of channels in an ASICs) is due to the fact that a single ASICs can be connected to two sets of 8 channel from different planes, however sets of 8 channels in an ASICs are from spatially adjacent channels. Sudden deviations from baseline, shared across 8 contiguous channels, can be attributed to misconfigured ASICs. The determination of misconfigured channels for the portion of Run 3 used for the HNL analysis has been performed in two different phases (each time a new channel database has been available for subsequent ranges

of Run 3). Figure 5.11 shows the average hit amplitudes as a function of run number for Run 3. For comparison, the same data from Run 1, where misconfigured ASICs were present, is shown in figure 5.12.



**Figure 5.11:** Average hit amplitude across the different runs of Run 3. No misconfigured channels are present.



**Figure 5.12:** Average hit amplitude across the different runs of Run 1. Misconfigured channels in the 2000-2400 block are visible during the run period 5300-6900.

Information regarding misconfigured channels is uploaded to a database, which is queried by the reconstruction algorithms so that they can correct for the misconfigured status of a certain channel. No new misconfigured channels have been found in the range of interest for this analysis (13696-17055).

### Beam Quality

Quality criteria are applied separately for beam and detector. Beam quality criteria do not affect data collected by the *off-beam* EXT hardware trigger and they require the beam to operate in stable conditions. Beam selections ensure that variables such as horn current, toroids current and figure of merit (FOM) value are within standard limits. FOM is an indicator of the fraction of beam that goes on target. It is determined by beam profile monitors that track the beam alignment. A beam database stores beam condition data and the list of runs that pass the beam quality criteria. During reconstruction, the database is queried by LArSoft which filters out events collected during not optimal beam conditions. The database is also queried during normalisation and POT counting. POT numbers are available for data before and after beam quality requirements [70].

### Data Quality

Data quality uses a combination of pre-selection criteria based on variables from the slow monitor system, which keeps track of the conditions of the detector, and selection criteria, which are based on reconstructed quantities. The pre-selection criteria are static (variables used and allowed range are not calculated relatively to a median of the dataset) and query the slow monitor database, ensuring that values for variables attributed to the different MicroBooNE sub-systems (e.g. TPC Drift Voltage, PMT High Voltage, DAQ monitoring variables, etc.) are within allowed range. Furthermore, *selection criteria* are applied, which are *dynamic* and are calculated relatively to a dataset median for events that have passed optical filtering. Variables used include number and coordinates of flashes, number of tracks and number of vertices; each of which has to be within  $3\sigma$  of the median of the whole dataset. An average of 10% POT data is removed from the *on-beam* BNB and *off-beam* EXT datasets after the application of beam quality criteria [71].

### 5.5.2 Data Calibration

The charge per unit length released by a charged particle crossing the detector, defined as  $dE/dx$  and measured in electrons/cm, drifts towards the wire plane where it produces a signal on the sense wire, which gets filtered, deconvolved and processed into a hit with an associated  $dQ/dx$ , measured in ADC/cm. The  $dx$  indicates the length of the track as seen by a wire on a wire plane and not as the element of infinitesimal distance in the *drift* ( $X$ ) direction. As such, for the same segment length along a track, the  $dx$  element is dependent on the angle of the track with respect to the wire pitch. The precise relationship between the energy released ( $dE/dx$ ) and the energy deposit measured ( $dQ/dx$ ) needs to be established in order to perform calorimetric measurements (i.e., associate an energy scale to particle energy measurements). We thus write:

$$dQ/dx = C \, dE/dx, \quad (5.5)$$

where  $C$  is a calorimetric constant, converting the ADC value measured by the waveform into number of electrons. Several detector effects (described in [section 3.3](#)), which are both space and time dependent, can cause a not uniform response across different wires and different periods of time.  $C$  is thus generally a function of detector coordinates and time. As an example, attenuation, which is caused by electron capture by argon impurities is strongly dependent on both  $x$  and  $t$ . For larger distances drift distances the capture probability increases and fewer electrons reach the wire plane for the same energy deposit, causing  $C$  to decrease at larger  $x$ . The purity level of the detector is also fluctuating over time, causing a time dependence. We thus write:

$$C = C_{\text{abs}} \cdot C_{\text{rel}}(x, y, z, t), \quad (5.6)$$

where  $C_{\text{rel}}(x, y, z, t)$  is a space and time dependent sets of constant, which makes the detector response uniform to an arbitrary ADC count (210 ADC/cm for the collection plane), and  $C_{\text{abs}}$  is an absolute constant that converts 210 ADC/cm to electron count.

The purpose of data calibration is the determination of the constants  $C_{\text{abs}}$  and  $C_{\text{rel}}(x, y, z, t)$  which are stored in a database and queried by [LArSoft](#) during reconstruction. This allows for the correction of the waveform in data at reconstruction, making calorimetry measurements possible. Data calibration for Run 3 has been performed

by the author, following the same procedure performed for Run 1 and outlined in Ref. [72].

$C_{\text{abs}}$  has been determined by the MicroBooNE collaboration using Run 1 data and a dataset of stopping muons from neutrino interactions.  $C_{\text{abs}}$  needs to be determined only once, since temporal variations are taken into account by  $C_{\text{rel}}$ .  $C_{\text{rel}}$  has been determined for Run 3, in the course of this analysis, using through-going CR muons. CR muons are MIP particles in the range 4 GeV - 5 GeV with a relatively constant energy deposition at  $\approx 1.7$  MeV/cm. An ideal dataset with an infinite number of CR would produce a sample where by summing all the energy deposition from all the tracks, a constant median  $dE/dx$  would be spread across the whole detector. In reality statistics are limited so that the detector needs to be voxelised and the size of the voxels must be such that the median  $dE/dx$  of each voxel is still constant. While the *local* median  $dE/dx$  is assumed to be identical for each voxel, the median  $dQ/dx$  is not, due to the physics and detector effects. Measuring the ratio between each local median  $dQ/dx$  and the *global* median  $dQ/dx$  provides the relative corrections  $C_{\text{rel}}$ .

In reality, due to limited statistics of through-going CR passing the selection criteria a full voxelisation of the detector is not possible. Corrections are thus assumed to factorise and they are generated independently in the  $yz$  plane (perpendicular to the drift direction), in the  $x$  axis (parallel to drift direction) and in time [72].

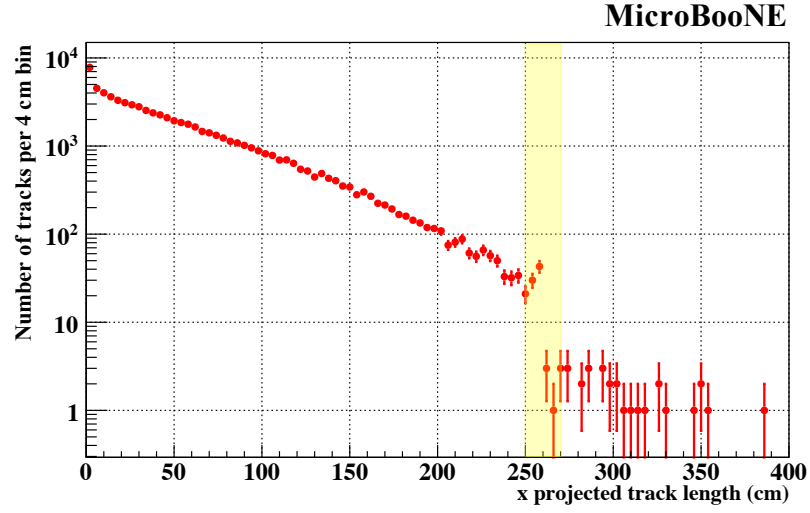
### Through-Going Cosmic-Ray Muons Correction Sample

For the calibration study, through-going CR muons are selected, which cross both the anode and the cathode. Anode-cathode CR muons span the whole  $x$  coordinate, thus covering all values of drift distance. Furthermore, the proximity to the anode plane allows the  $t_0$  to be determined with precision, since the  $t_0$  is assumed to coincide with the minimum drift coordinate.

The following criteria are required (coordinate definitions are shown in figure 3.5):

- *Projected  $x$  length in range* [250 cm, 270 cm]. This is a requirement for selecting anode-cathode crossing tracks. The projected length of a sample of CR muons is shown in figure 5.13.

- $\theta_{xz}$  *outside of range*  $[75^\circ, 105^\circ]$ . The requirement excludes tracks that are perpendicular to the wire plane, and due to the small number of hits representing them (and DIC effects) are poorly reconstructed.
- $\theta_{yz}$  *outside of range*  $[80^\circ, 100^\circ]$ . The requirement removes track which are nearly parallel to the collection plane wires, which are also poorly reconstructed.



**Figure 5.13:** Projected length on the  $x$  axis of a sample of CR muons from Run 1. The coloured yellow band shows the  $x$  projected length for anode-cathode crossing CR. The sharp increase at the beginning of the yellow band is due to the geometric constraint of the detector which has a maximum drift distance of  $\approx 250$  cm. Tracks with a projected  $x$  length of 250 cm can be produced by any CR muon which is crossing both the anode and the cathode. The boundaries of the yellow band are used as criteria to select the CR muons used for the calibration of the detector. Only  $\approx 0.13\%$  of CR muons are anode-cathode crossing, which makes factorisation necessary in order to have sufficient statistics in each voxel. Figure is from Ref. [72].

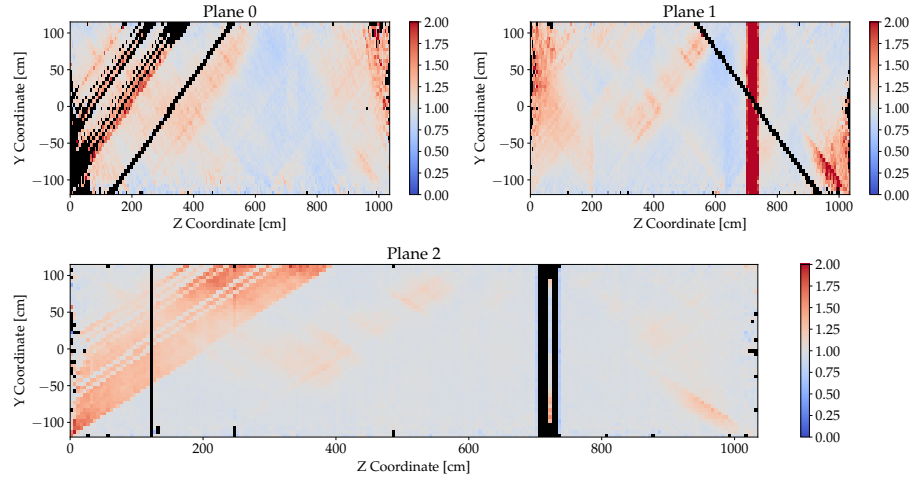
## YZ Corrections

YZ plane corrections deal with space charge effects, misconfigured or cross-connected anode plane channels and transverse diffusion. In order to determine correction values, the YZ plane is segmented into 5 cm by 5 cm cells and the median  $dQ/dx$  value for each cell (*local median*) is calculated. A  $y$  and  $z$  dependent correction factor

is derived by dividing the *global median* (the  $dQ/dx$  median across the entire  $YZ$  plane) by the *local median* of each cell:

$$C(y, z) = \frac{(dQ/dx)_{\text{global}}}{(dQ/dx)_{\text{local}}} \quad (5.7)$$

The  $YZ$  corrections are shown in [figure 5.14](#).



**Figure 5.14:**  $YZ$  correction factors with data from the run period 13696-15500.

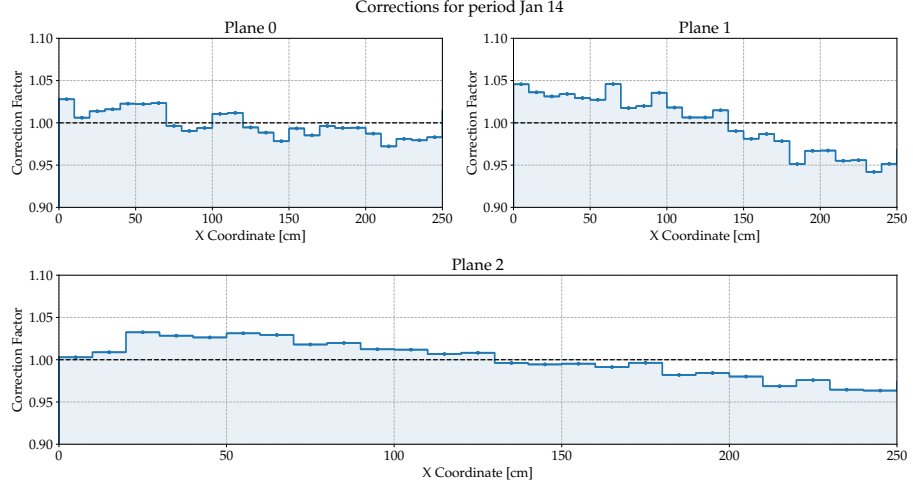
## **X Corrections**

$X$  corrections deal mainly with effects such as space charge, electron attenuation and longitudinal diffusion. Corrections are calculated for each day, and due to the requirement for high statistics, corrections are derived only for days that have a minimum of 40 crossing tracks leaving hits in the collection plane after all angular cuts. The previously determined  $YZ$  corrections are applied to all reconstructed  $dQ/dx$  values. The local  $dQ/dx$  median for each 5 cm segment along the  $x$  axis is then calculated and divided by the global median calculated across the whole  $x$  range. A correction factor for each 5 cm-long segment is then calculated.

$$(dQ/dX)_{\text{corrected}} = C(y, z) \times (dQ/dX)_{\text{reconstructed}}, \quad (5.8)$$

$$C(x) = \frac{(dQ/dx)_{\text{corrected, global}}}{(dQ/dx)_{\text{corrected, local}}}, \quad (5.9)$$

The  $X$  corrections are shown in figure 5.15.



**Figure 5.15:**  $X$  correction factors with data from the run period 13696-15500 (January 14). Different planes are affected by detector effects with different magnitude. The rise in the first 10 cm is caused prevalently by electron attenuation, while the successive fall is mostly a consequence of SCE. These effects are most evident in *plane 2* (Y plane).

## T Corrections

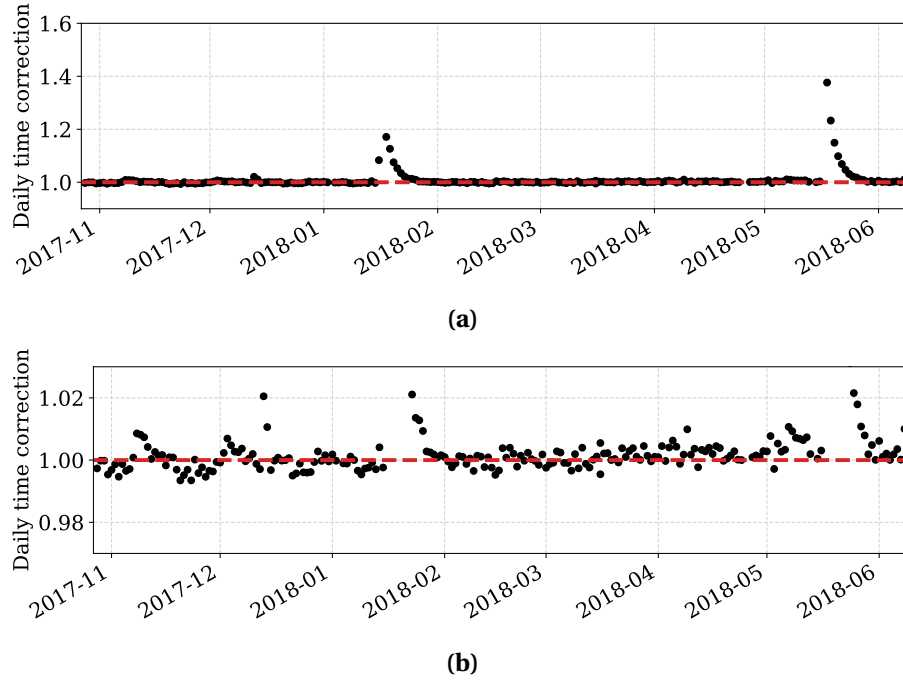
$T$  corrections deal with any temporal variation in the detector response. By ensuring that the detector response is uniform over time, it is possible to use the same absolute corrections across different runs and time periods. Time corrections are calculated on a daily basis, by measuring the *corrected*  $dQ/dx$  response (using both  $YZ$  and  $X$  corrections) and taking the ratio to a reference value (210 ADC/cm for the collection plane).

$$(dQ/dx)_{\text{corrected}} = C(y, z) \times C(x) \times (dQ/dx)_{\text{reconstructed}}, \quad (5.10)$$

$$C(t) = \frac{(dQ/dx)_{\text{reference}}}{(dQ/dx)_{\text{corrected, local}}}, \quad (5.11)$$

The  $T$  corrections are shown in figure 5.16.





**Figure 5.16:**  $T$  correction factors in the collection plane for data from the run period 13696-17055. After  $YZ$  and  $X$  corrections, variations are of the order of  $< 1\%$ . The two peaks in January and May in (a) correspond to periods of time in which the argon pumps were malfunctioning and the purity dropped drastically. The slow tail over the successive days corresponds to the progressive purification of the detector after the replacement of the pump. In (b) is shown the same figure zoomed on the  $y$  axis. The  $\mathcal{O}(0.01)$  increase in the median has also been observed during Run 1. The causes for the upward trend are attributed to a build-up of SCE during run time.

## 5.6 Object Reconstruction and Identification

LArSoft reconstruction modules have the task of interpreting the data from the raw waveforms of either data or simulated MC into physics quantities that can be studied by the analysers. The reconstruction of the data from the TPC and PMT waveforms is performed separately.

For TPC data, the reconstruction is performed in the following steps:

- *Noise filtering*, which removes known sources of noise from the waveforms, due mostly to the ASICs, voltage regulators and high voltage power supply.

- *Deconvolution*, used to deconvolve the detector response from the measured waveform and obtain a signal waveform, where peaks can be directly correlated to electron clusters originated from energy depositions.
- *Hit finding*, performed by fitting Gaussians to the peaks of deconvolved signal waveforms to generate *hits*, corresponding to energy depositions.
- *Clustering*, which groups hits into 2-d objects, which are 2-d projections of the higher-level tracks and showers in the detector.
- *3D object reconstruction*, performing further grouping on the clusters to reconstruct particles and perform PID.

Each of these steps builds higher-order reconstructed quantities. The final output of the reconstruction modules is a series of reconstructed particles with PID and associated direction and momenta.

### 5.6.1 Noise Filtering

The signal detected on the TPC wires, shaped and amplified by the ASICs in the cold, has to travel from inside the cryostat to the DAQ racks. During collection, shaping, amplification, and transport, different source of electronic noise can build up. Noise characterisation and filtering methods have been developed by the MicroBooNE collaboration (see Ref. [69]), in which different sources of inherent noise, generated by the ASICs, have been found, such as: “white series” noise due to thermal fluctuations, low-frequency “pink” noise due to charge trapping and de-trapping in the input transistor and “white parallel” noise due to the transistor bias current and resistors providing wire bias voltage.

Excess noise, in addition to the inherent ASIC noise has also been observed. The two most important sources are:

- Low-frequency noise from the voltage regulator which provides a stable voltage for the cold ASICs is highly correlated across channels on the same service board that contain the same regulator. This coherent noise is mitigated via correction waveforms constructed on a per sample basis.

- Harmonic noise is induced on the anode wire plane by fluctuations in the cathode potential, due to the ripple frequency of the high voltage power supply. The two highest peaks are near 36 kHz and 108 kHz and the effects of the high voltage power supply noise is strongest in the first induction plane, which is positioned in front of the cathode, and gets progressively weaker for the other two planes, which are shielded from it. Noise is removed by masking discrete bins in the frequency domain.
- *Burst* noise at high-frequencies (900 kHz) is observed, which has no clear position dependence and occurs intermittently. The source of the noise has not been conclusively identified but is suspected to originate from the PMT high voltage power supply. The noise is not directly mitigated since it is naturally attenuated to acceptable levels by the anti-alias ASIC filter.

Figure 5.17 illustrates the principal sources of excess noise and the effects on a raw waveform after applying a noise filter. The full effects on an event display of the offline noise filter are shown in figure 5.18. More details on the characterisation of the noise and the application of the filters can be found in Ref. [69].

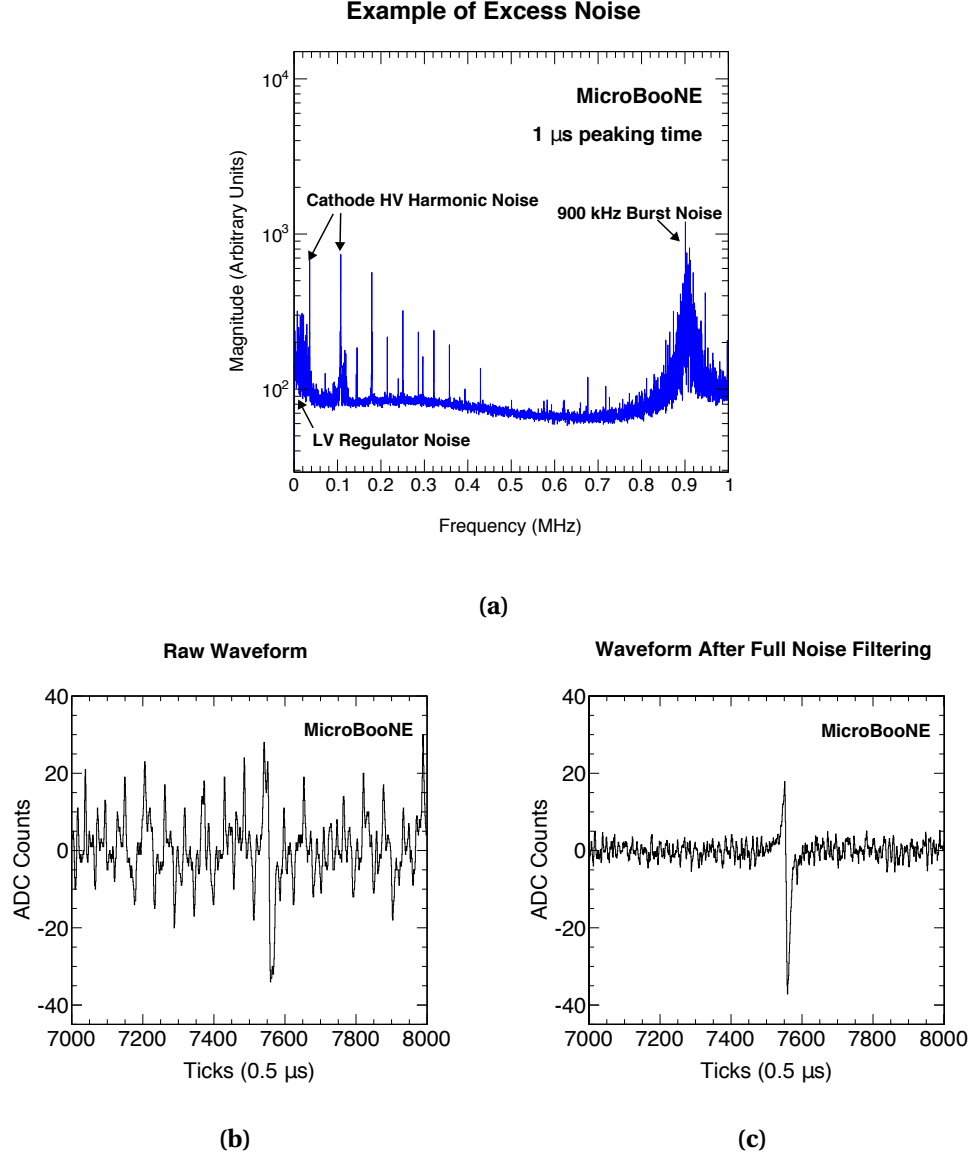
### 5.6.2 Deconvolution

The signal read by each of the TPC channels is a combination of the charge induced by the electrons, noise and detector effects. The aim of deconvolution is to obtain a signal waveform from a measured waveform, removing the detector response. In order to apply deconvolution, we model the measured signal  $M(t')$  as a convolution of the original signal  $S(t)$  and the detector response function  $R(t, t')$ :

$$M(t') = \int_{-\infty}^{\infty} R(t - t') \cdot S(t) dt, \quad (5.12)$$

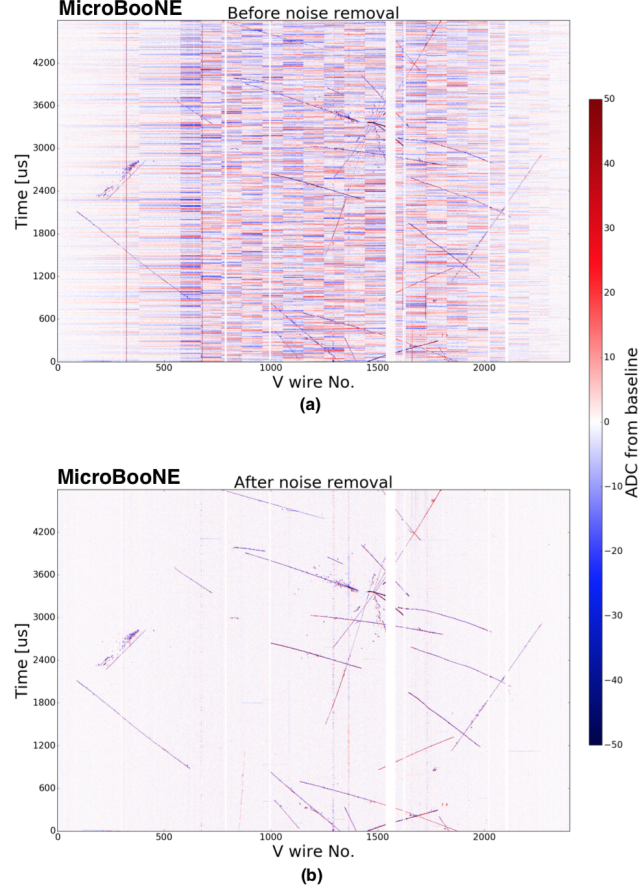
where  $t$  is the signal time,  $t'$  is the time the signal is measured, and we have assumed the detector response to be time-invariant ( $R(t, t') = R(t' - t)$ ). Equation (5.12) can be solved for the original signal by performing a Fourier transformation:

$$S(\omega) = \frac{M(\omega)}{R(\omega)}, \quad (5.13)$$



**Figure 5.17:** Average noise magnitude in the frequency domain for the factory  $1\ \mu\text{s}$  peaking time  $4.7\ \text{mV/fC}$  gain ASIC settings (a). These do not correspond to the MicroBooNE default settings but facilitate display of the noise (the noise is at the same frequencies for default settings). In (b) and (c) is shown the same U plane raw waveform from data before and after the full noise filtering chain. Figures are from Ref. [69].

where  $\omega$  is the frequency. Once  $M(\omega)$  is measured and  $R(\omega)$  correctly modelled,  $S(t)$  can be obtained by applying the inverse Fourier transform to the ratio of  $M(\omega)$  and  $R(\omega)$ . Two complications, arising from the measured signal  $M(t')$  containing ad-



**Figure 5.18:** 2-d event display of the V plane showing the raw signal before (a) and after (b) offline noise filtering. The figure shows how a clean event signature can be recovered after subtracting the identified sources of noise. Figure is from Ref. [69].

ditional contributions from electronic noise sources and the response function  $R$  decreasing at high frequencies, cause the noise spectrum to be artificially amplified by equation (5.13). An additional term for a low-pass filter function  $F(\omega)$  is thus introduced, which attenuates the high frequency noise:

$$S(\omega) = \frac{M(\omega)}{R(\omega)} \cdot F(\omega). \quad (5.14)$$

While this one dimensional deconvolution technique is fast and computationally light, it assumes that the measured signal on a wire depends exclusively on the charge deposition on that wire, while neglecting the induced signal on the other wires and the

impact on the response function  $R$ . For a more realistic treatment of this effect, known as Dynamically Induced Charge (DIC), the measured signal  $M_i$  for a wire  $i$  would have to be expanded to include the responses for the charge drifting in the neighbouring wire regions:

$$M_i(t') = \int_{-\infty}^{\infty} \left[ \dots + R_1(t-t') \cdot S_{i-1}(t) + R_0(t-t') \cdot S_i(t) + R_1(t-t') \cdot S_{i+1}(t) + \dots \right] dt, \quad (5.15)$$

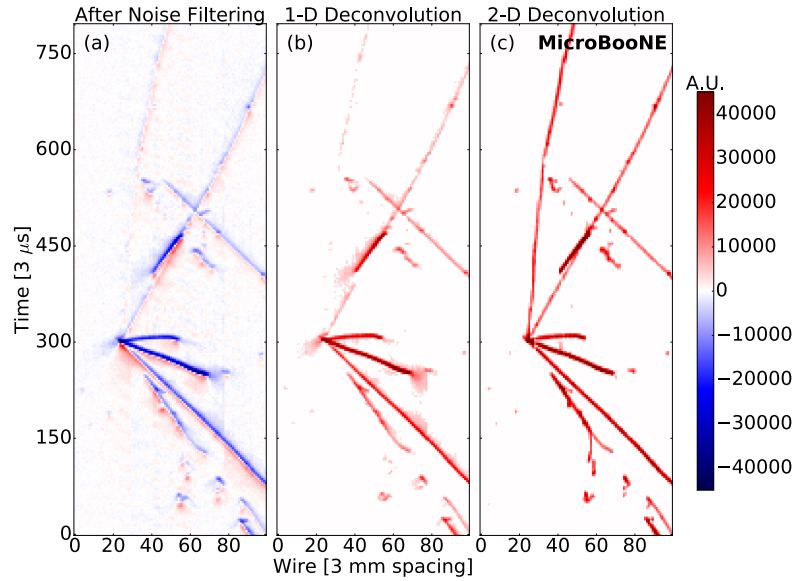
where  $M_i$  is the signal measured on the wire  $i$ .  $S_{i-1}$ ,  $S_i$ ,  $S_{i+1}$  represent the signal inside the boundaries of wire  $i$  and its adjacent neighbouring wires.  $R_0$  is the average response for ionisation charge passing in the area surrounding the wire  $i$ , while  $R_1$  is the average response for ionisation charge drifting in the area surrounding the neighbouring wires  $i-1$  and  $i+1$ . Terms up to  $R_n$  can be included for increasing description accuracy, albeit with diminishing impact.

2-d deconvolution provides a more robust representation of the electronic noise and provides better data-MC agreement. It is particularly efficient in recovering tracks and showers directed towards the wire plane, which cannot be properly reconstructed with traditional 1-d deconvolution. Figure 5.19 illustrates the effect of applying 1-d and 2-d deconvolution to the waveforms of a MicroBooNE data event. Although 2-d deconvolution will be implemented in the next iteration of LArSoft it is not part of the standard LArSoft suite (MCC8) used for this analysis. More details on the signal processing and related algorithms can be found in Ref. [73, 74].

### 5.6.3 Hit Finding

Peaks in the deconvolved waveform correspond to electron clouds drifting past the wires, or collected by them. A hit-finder algorithm performs fits of Gaussian functions to the peaks of the deconvolved waveform. For each Gaussian, the mean of the Gaussian provides the time the signal was measured while the integral of the Gaussian (pulse-area) provides the amount of charge deposited which is proportional to the number of electrons and the amount of energy deposited by the original particle responsible for the signal.

The two quantities are used to construct *hit* objects. Using the  $t_0$  of the interaction,



**Figure 5.19:** Effects on the waveforms of the 1-d and 2-d deconvolution. On the  $x$  axis are individual wires, for a given wire ( $x$  coordinate), the  $y$  axis indicates the time of the waveform measured by the wire, the  $z$  axis in red and blue is directly proportional to intensity of the current measured by the wire and the signal collected (in arbitrary units, calculated from average baseline subtracted ADC scaled by 250 per  $3 \mu\text{s}$ ). The figure shows the importance of 2-d deconvolution for tracks perpendicular to the wire plane (in this figure, parallel to the  $y$  axis), which in the 1-d deconvolution case are not visible. Figure is from Ref. [73].

and taking advantage of the constant drift speed, the time of the pulse can be used to provide, in the drift direction, an  $x$  coordinate for the hit.

#### 5.6.4 Clustering and 3D Object Reconstruction

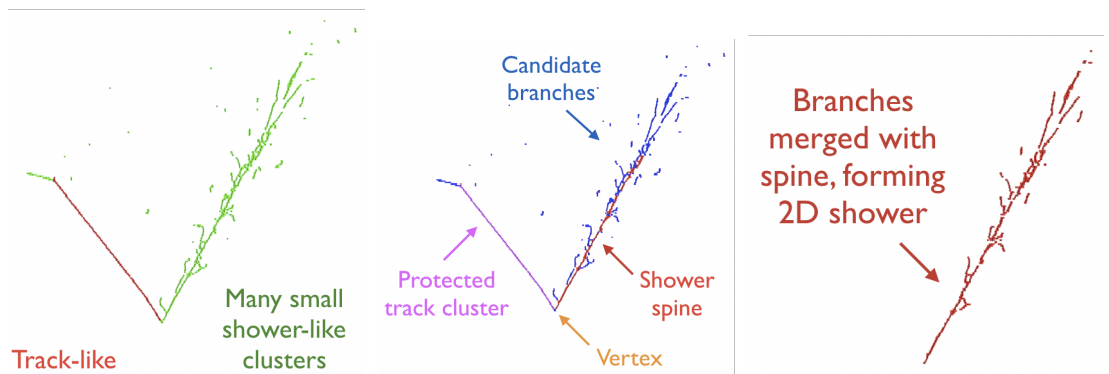
Various techniques are used at MicroBooNE to combine the information from the hits into 3-d reconstructed particles. This analysis make use of the algorithms provided by the Pandora pattern recognition software [62], which can be interfaced with LArSoft reconstruction modules. Pandora is based on algorithms that create a series of reconstructed objects of increasing complexity and satisfying harsher topological association requirements.

The ensemble of hits from a wire plane, an example of which is shown in figure 5.20,



**Figure 5.20:** Hits from the induction plane grouped into clusters (indicated with a random colouring scheme) by the Pandora clustering algorithms. Figure is from Ref. [63].

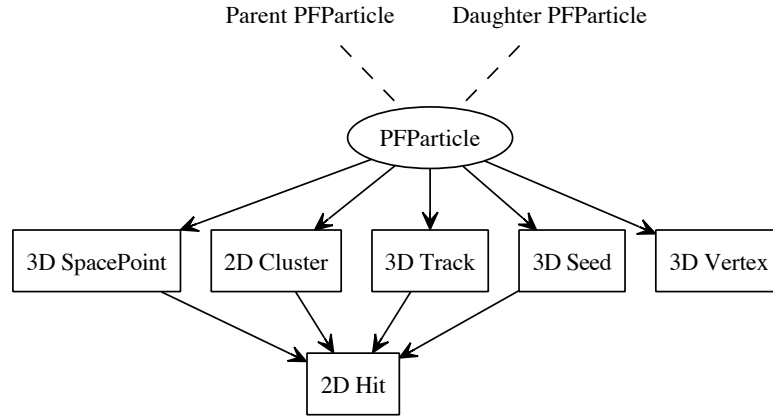
can be thought of as a 2-d projected *snapshot* of the interaction occurring in the detector in a 4.8 ms wide timing window. Clustering algorithms group together hits in a plane that are in geometrical proximity and create geometrical shapes consistent with expectations of components of 2-d projections of tracks and showers. The initial clustering aims at providing high purity (hits exclusively from one particle), at the cost of completeness (not all hits from a particle are included in the cluster), by constructing continuous and unambiguous lines of hits (see figure 5.20). Successive refinement algorithm can then split or merge clusters.



**Figure 5.21:** Refinement of clustering process and formation of a 2-d shower. Figure is from Ref. [63].



Combinations of 2-d clusters are then used to create plausible candidate *3-d vertex* positions and where three consistent track-like clusters can be matched across the three planes, *3-d tracks*. 2-d showers are built merging together consistent shower-like clusters. The process is shown in [figure 5.21](#).



**Figure 5.22:** Relationship and dependencies of the Pandora objects. From the 2-d hits, Pandora constructs different 2-d and 3-d objects which are combined to create PFPParticle candidates. Figure is from Ref. [63].

*3-d spacepoints* (3-d hits) can also be created by combining 2-d hit information for clusters that have been matched together across different planes. The final output of Pandora is a collection of 3-d reconstructed particles (termed *PFParticles*), each of which correspond to a distinct track or shower in the event. PFParticles have associated reconstructed 3-d spacepoints, 3-d position and direction (*3-d seeds*) and a *3-d vertex* for their origin. In addition, topological association between PFParticles are used to identify parent/daughter relationships and construct a particle hierarchy. The relationship between the Pandora data objects are shown in [figure 5.22](#).

## Cosmic and Neutrino Passes

An important feature of the application of Pandora reconstruction algorithms in LArSoft is its two-pass approach:

- The **PandoraCosmic** pass is configured to run algorithms optimised for the reconstruction of CR muons and their delta rays daughters (high energy electrons

*branching* out of a CR track). The reconstruction is extremely track-oriented and it aims at identifying unambiguous CR muons that are through-going. The reconstructed start point of the CR track is assumed to be the highest  $y$  coordinate. 2-d hits which have been associated to *PandoraCosmic* tracks are *removed* from the list of 2-d hits available to the successive algorithms.

- The **PandoraNu** pass performs standard reconstruction *after* the *PandoraCosmic* pass and on the remaining 2-d hits that have not been removed by *PandoraCosmic* after having been matched to CR tracks. The pass is configured to identify neutrino interaction vertex, with particular emphasis on the purity and completeness of the *secondaries* (particles emerging directly from the 3-d vertex).

More information on the Pandora software development kit can be found in Ref. [62] and on the application of Pandora to pattern recognition in LArTPCs in Ref. [63].

### 5.6.5 Optical Reconstruction

The optical reconstruction uses as input the raw waveform from the PMT system and produces reconstructed *flashes*, representing optical activity detected by multiple PMTs in time coincidence and caused by either neutrino or CR muon interaction. Reconstruction is performed by matching the high and low gain channels (described in section 3.2.4), correcting saturated waveforms from the high-channels with the information of the low-gain channels. The waveform baseline is then calculated and a peak-searching algorithm is used to identify pulses above threshold. Once coincident pulses are determined across different PMTs, an integration window of 8  $\mu\text{s}$  is applied to collect late light.

Mapping the location of the PMTs responsible for the pulses of the flash and calculating the mean position weighted by the PE observed by each PMT allows to associate  $y$  and  $z$  coordinates on the anode plane to the reconstructed flash. Each flash is thus characterised by the total PE measured by the associated PMTs, the  $y$  and  $z$  coordinates and the time of the flash in  $\mu\text{s}$  relative to the hardware trigger.

## 5.7 Datasets

In this section are outlined the datasets used in this work and their role in the analysis. The data from the [HNL](#) software trigger used here has been collected over Micro-BooNE's Run 3, which lasted from October 2017 to July 2018. The portion of time between June and the end of Run 3 has been excluded, since an updated channel status database was not available for that month at the time of this analysis. A total of  $1.97 \times 10^{20}$  POTs have been collected for the [HNL](#) sample, after beam and run quality filters. All datasets are normalised relatively to the amount of POTs collected by the *on-beam HNL* data sample. [Figure 5.23](#) shows a diagram of the datasets and their use.

### 5.7.1 Test Samples

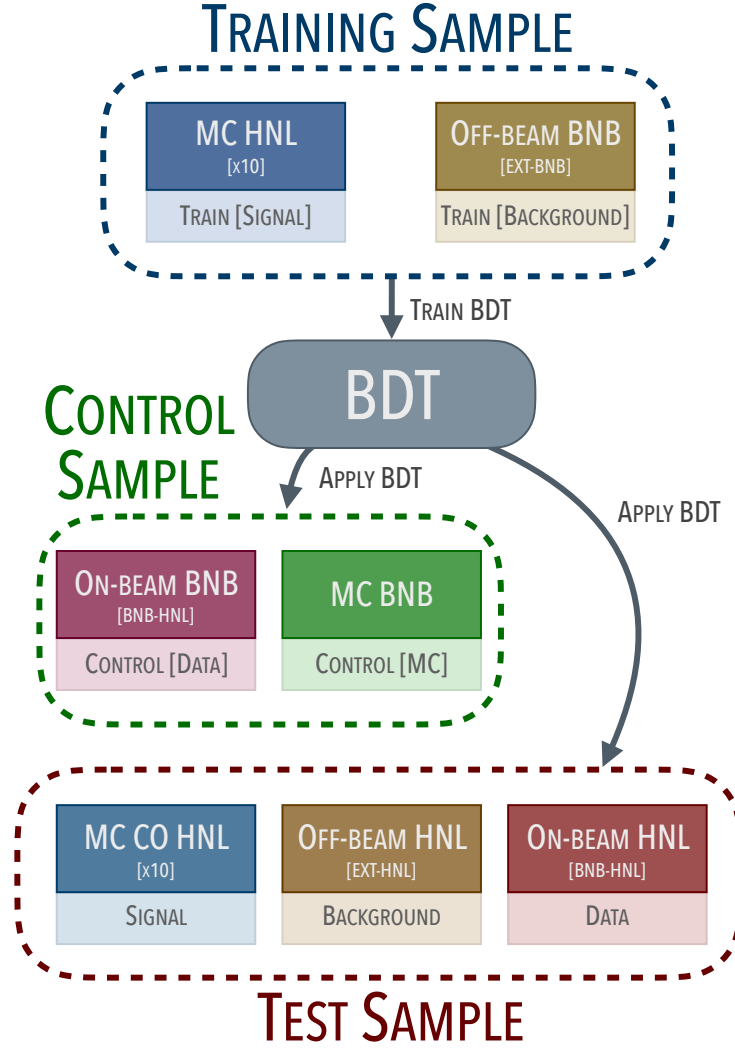
The three main datasets used for testing the [HNL](#) hypothesis are termed *test samples*. They comprise the data sample, and the background and signal sample. If signal exists, the data sample would be described by the sum of the background and signal sample. They are defined as following:

#### On-Beam Data Sample: BNB-HNL

The on-beam data sample is obtained by the [BNB](#) hardware trigger, in coincidence with a beam spill, and the [HNL](#) software trigger, vetoing the timing region preceding it. It is a data sample assumed to contain CR muons and (if existing with sufficiently high mass-mixing matrix element values) [HNLs](#). [Table 5.3](#) shows more information regarding the dataset. The values shown in the table are used as a baseline for normalising the other datasets. A beam-quality and data-quality filters have been applied to the dataset. POT count and number of hardware triggers are determined after the filter.

#### Off-Beam Background Sample: EXT-HNL

The off-beam data sample is obtained by the [EXT](#) hardware trigger, *not* in coincidence with a beam spill, and the [HNL](#) software trigger, vetoing the timing region preceding



**Figure 5.23:** Datasets used in this analysis and their role in the analysis workflow. Training samples are used for the training of a BDT which is then subsequently applied to the remaining datasets. The control samples ensure that the workflow is not sensitive to MC-data differences, while the test samples are used to test the HNL hypothesis.

it. Although no BNB interaction is present in the timing window preceding the HNL window, the purpose of the veto is to have the distribution of CR muons match the same distribution from the data *on-beam* HNL (BNB-HNL) dataset, up to timing and veto effects. It is assumed to be a dataset containing purely CR muons. If no signal is present, *off-beam* HNL and *on-beam* HNL would correspond to the same dataset, up to statistical differences. The difference in the number of events available in this

<i>On-beam HNL</i> Sample Statistics	
Time range covered	October 2017 - June 2018
Event count	199,226
POTs	$1.97 \times 10^{20}$
Hardware triggers	47,459,339
Norm. factor	1

**Table 5.3:** *On-beam HNL* (BNB-HNL) dataset used for this analysis, after beam quality selections.

<i>Off-beam HNL</i> Sample Statistics	
Time range covered	October 2017 - June 2018
Event count	90,869
Hardware triggers	23,332,437
Norm. factor	2.03

**Table 5.4:** *Off-beam HNL* (EXT-HNL) dataset used for this analysis.

dataset, with respect to *on-beam HNL*, is due to a pre-scaling factor of 0.2 and different number of hardware trigger issued. Table 5.4 shows more information regarding the dataset. A data-quality filter has been applied to the dataset. POT count and number of hardware triggers are determined after the filter.

### Simulated Signal Sample: MC CO HNL

The simulated signal sample, MC CO HNL, is a simulation of the HNL decay signal expected in MicroBooNE. It uses the CO technique, outlined in section 5.2.3, to overlay data CR muons and data noise on top of the simulated MC events. Samples are produced for 10 different mass hypotheses ([260, 265, 285, 300, 305, 325, 345, 365, 370, 385] MeV). Generation of samples for different mass-mixing matrix elements is not necessary since this parameter does not affect production or decay kinematics and overall scaling of the sample is sufficient. The *off-beam* unbiased data used for the overlay is from the period October 2017 - March 2018. While this does not cover the whole time range of the data and background datasets, studies performed during the calibration of the detector during Run 3 have confirmed that no fundamental operational differences are present in the detector between the period October 2017 - March 2018 and the remaining period March 2018 - June 2018 (e.g., no new misconfigured

MC CO HNL Sample Statistics	
“Time range” covered	October 2017 - March 2018
Event count	$\approx 120,000$
Norm. factor (for $ U_{\mu 4} ^2 = 10^{-7}$ )	$\mathcal{O}(10^{-4})$

**Table 5.5:** MC CO HNL datasets used for this analysis. The time range shown in the table reflects only the period of time from which the CR muons sample (*off-beam* unbiased) used for the overlay is taken from, since the MC simulated HNLs do not have any time period dependency. The event count is shown only approximately, since each of the 10 datasets for each mass value has a different event count. The same occurs for the normalisation factor, since samples are normalised to the *on-beam* HNL POT count, by using their event count as denominator and different number of events are expected for each mass point (see [section 5.2.4](#)).

ASICs or shorted channels have appeared), so that the two halves of the dataset can be considered equivalent up to the expected amount of statistical fluctuations (see [figure 5.16](#)). [Table 5.5](#) shows more information regarding the datasets. A data-quality filter has been applied to the *off-beam* unbiased data used for the CO procedure.

### 5.7.2 Control Samples

While the efficiency for MC simulated signal events can be calculated, this does not necessarily implies that the same efficiency would be observed in data (e.g., intrinsic MC-data effects not taken into account might be responsible for a lower efficiency for data).

Due to the fact that *data signal* samples (real HNL decays) are not available for comparison, we use a dataset of BNB interaction with event topology similar to the HNL decay and compare it to its respective simulation in order to quantify possible differences in the analysis workflow performance between data and MC.

#### Control Data Sample: Data *On-beam* BNB

Data collected by the BNB hardware trigger and the BNB software trigger (*on-beam* BNB, BNB-BNB), operating during the neutrino beam spill and collecting neutrino

<i>On-beam BNB</i> Sample Statistics	
Time range covered	February 2016 - June 2016
Event count	192,043
POTs	$4.9 \times 10^{19}$
Hardware triggers	10,948,876
Norm. factor	4.34

**Table 5.6:** Data *on-beam BNB* dataset used for this analysis.

<i>MC BNB</i> Sample Statistics	
Event count	74,600
POTs simulated	$7.54 \times 10^{19}$
Norm. factor	2.61

**Table 5.7:** MC BNB dataset used for this analysis.

interaction events. At the time of the analysis, reconstructed *BNB* data for Run 3 was not available, and Run 1 data was used. While operational differences exists between Run 1 and Run 3, the dataset is used only for internal comparison with the associated MC sample, which has been tuned for Run 1. Run 1 *BNB* data has been extensively studied by the MicroBooNE collaboration and it has been used for several analyses. Table 5.6 shows more information regarding the dataset. A beam-quality and data-quality filters have been applied to the dataset. POT count and number of hardware triggers are determined after the filter. In all the figures and efficiencies shown in the remaining part of this thesis, the *off-beam BNB* data has been subtracted from the *on-beam BNB* dataset.

### Control Monte Carlo Sample: MC BNB

The MC generated *BNB* sample comprises GENIE-generated neutrino interactions and CORSIKA-generated CR muons. An additional sample is created where the neutrino interaction is simulated outside of the cryostat (*dirt* interaction) and the two samples are added together after normalisation. Table 5.7 and table 5.8 show more information regarding the datasets.

---

<i>MC BNB Dirt Sample Statistics</i>	
Event count	1,020,044
POTs simulated	$2.7 \times 10^{21}$
Norm. factor	0.07

---

**Table 5.8:** MC BNB dirt dataset used for this analysis.

### 5.7.3 Training Samples

The final steps of the analysis chain involve the use of a Boosted Decision Tree (BDT) for the discrimination of the signal and background in the data sample. The BDT needs to be trained on samples containing pure signal features and pure background features. Ideally one would use subsets of the same signal and background sample used for testing (MC CO HNL and *off-beam* HNL), however two factors need to be taken into account:

- The HNL candidates found in the MC CO HNL sample and providing a *signal* sample contain actually a mixture of correctly identified HNLs and HNL candidates corresponding to CR muons, constituting the background. While the true coordinates and trajectories of the tracks from the “truth” HNL vertex can be used and *reco-truth* matching methods are available in LArSoft (to verify whether a reconstructed objects does indeed match its simulated counterpart), a complication arises from the fact that the CR muons are provided from data, and their true coordinates and trajectories are not known. This worsens the efficiency of the *reco-truth* matching algorithms, which for CORSIKA MC generated CR muons can count on the “truth” information to reject false matchings. A consequence of this is that after the *reco-truth* matching the “pure” HNL sample from the MC CO HNL would still contain accidental CR muons.

Although computationally more expensive, a better method is the generation of a MC HNL sample for each mass point. The MC HNL samples contain exclusively HNL decays, generated in the same fashion as the MC CO HNL samples, but without any data or CORSIKA-generated CR muons. A drawback of this method is that the HNLs in the MC sample differ slightly from the MC CO HNLs. The noise on the waveforms is simulated instead of provided by data, and data-driven corrections are not applied. Furthermore, *hybrid* candidates not provided by the



linear combination of HNL and CR candidates (e.g., due to a CR muon *crossing* a HNL and preventing reconstruction or misplacing the vertex) are not taken into account. These effects however are considered to be negligible and do not constitute a source of uncertainty, since the samples are used exclusively for the *training* of the BDT. Once the BDT has been trained, its efficiency is uniquely determined by the *test* samples (i.e., MC CO HNL and *off-beam* HNL), and the use of any arbitrary sample for its training can be likened to the placing of an arbitrary selection on a sample variable (e.g., the efficiency of the selection is independent of the particular reasoning for that selection).

- The statistics available for the background *off-beam* HNL sample are limited (see table 5.4). The advantage of using a data sample for background, which does not differ from the background in the data sample due to mismodelled MC, is however mitigated by the fact that additional statistics cannot be produced (as in the MC case). Splitting the sample in a subset for training and testing would result in a poorly trained BDT and a low-statistics background sample. For this reason, we train the BDT with a sample of *off-beam* BNB CR muons, for which 10 times more events are available. The *off-beam* BNB sample differ from the *off-beam* HNL for few reasons: two different software triggers with two different PMT thresholds are applied (BNB: 130 ADC and HNL: 210 ADC). The BNB software trigger does not performs the same veto in the region preceding it (it does apply a form of optical filtering, but with different settings) and the data collected is from Run 1, which was operated differently from Run 3. The same reasoning proposed for the use of the MC HNL sample holds though, where the variables used for the training of the BDT, based mostly on topological structures, are not sensitive to these differences. After quality selections, the distributions of the variables used for the training of the BDT are compatible, up to statistical differences, and the BDT shows better performance if trained with larger statistics from the *off-beam* HNL when tested on the *test* samples.

### Training Signal Sample: MC HNL

The simulated signal sample, MC HNL, is a simulation of the HNL decay signal expected in MicroBooNE, with no data or CORSIKA-generated CR muons and simulated

---

<i>MC HNL</i> Sample Statistics	
Event count	$\approx 150,000$
Norm. factor (for $ U_{\mu 4} ^2 = 10^{-7}$ )	$\mathcal{O}(10^{-4})$

---

**Table 5.9:** MC HNL datasets used for this analysis.

<i>Off-beam BNB</i> Sample Statistics	
Time range covered	February 2016 - October 2016
Event count	983,718
Hardware triggers	77,633,890
Norm. factor	0.6

---

**Table 5.10:** *Off-beam BNB* (EXT-BNB) dataset used for this analysis.

noise. Samples are produced for 10 different mass hypotheses ([260, 265, 285, 300, 305, 325, 345, 365, 370, 385] MeV) and the production uses the same identical parameters used for the MC CO HNL samples. Table 5.9 shows more information regarding the datasets.

### Training Background Sample: Data *off-beam BNB*

Data collected by the EXT hardware trigger, *not* in coincidence with a neutrino spill, and the BNB software trigger during Run 1. The *off-beam BNB* sample is expected to contain exclusively CR muons, much like the *off-beam HNL* sample albeit with a different distribution due to different PMT thresholds. Table 5.10 shows more information regarding the dataset. A data-quality filter has been applied to the dataset. Number of hardware triggers is determined after the filter.

### Data Sample Summary

A summary of the data samples used in this analysis, their content and their function is shown in table 5.11.

Data Samples Summary		
Usage	Sample Name	Sample Content
<i>Training samples</i>	<b>MC HNL</b>	HNL <i>simulation</i> (no overlay)
	<b>Off-beam BNB</b>	CR muon <i>data</i> (from BNB trigger)
<i>Control samples</i>	<b>On-beam BNB</b>	Neutrino interaction and CR muon <i>data</i>
	<b>MC BNB</b>	Neutrino interaction and CR muon <i>simulation</i>
<i>Testing samples</i>	<b>MC CO HNL</b>	HNL <i>simulation</i> (with overlay)
	<b>Off-beam HNL</b>	CR muon <i>data</i> (from HNL trigger)
	<b>On-beam HNL</b>	CR muon <i>data</i> with possible HNL <i>data</i> (from HNL trigger)

**Table 5.11:** Summary of the samples used in this analysis. *Training samples* are used to train the BDT for signal selection. *Control samples* are used to verify that the BDT is capable of selecting signal-like events in data and that the BDT (and the entire selection workflow) is not sensitive to data-MC differences. *Testing samples* are the *signal*, *background* and *data* (*signal* plus *background*) samples and are used to test the HNL hypothesis (whether signal events are contained in the *data* sample).

# Event Selection

The aim of this analysis is to identify and reconstruct [HNL](#) candidates. [Section 6.1](#) describes how signal candidates are identified in an event, and how high-level reconstructed variables are constructed and assigned to the candidates. In [section 6.2](#) are described the pre-selection requirements applied to the candidates. A second series of pre-selection requirements based on kinematic quantities are described in [section 6.3](#). The efficiencies of both types of pre-selection are given in [section 6.4](#).

## 6.1 Signal Definition

The decay of [HNLs](#) to  $\mu\pi$  pairs is characterised by two secondary tracks sharing a common vertex, with no additional activity around that vertex. Candidates are identified using a custom built [LArSoft](#) analyser module, [HELEPHANTE](#) (HELEPHANT Explorer), written by the author.

The aim of the module is to construct a database of minimally-defined [HNL](#) candidate objects contained in a dataset, which can be subsequently analysed offline. The original [art](#)-ROOT files in which the events are stored contain all the information available for a specific event, from the raw [TPC](#) and [PMT](#) waveforms to the full list of *pandoraCosmic* and *pandoraNu* reconstructed PFParticles and associated objects, which makes the datasets unsuitable for fast analysis and interactive workflow (e.g., for the test sample, *on-beam HNL*, the size of the dataset is of the order of  $\approx 4$  TB).

The output of [HELEPHANTE](#) is in a file format dependent only on CERN-ROOT

libraries that can be analysed offline (with no LArSoft dependency) and contains variables related to the HNL candidates and meta-data regarding the event. Two different ROOT TTrees are available for event-wise information (as in the case of MC simulation, the *truth* information of the generated event is unique for each event) and candidate-wise information, where each entry corresponds to a single candidate and individual event numbers can occur multiple times (e.g., an event may have three HNL candidates with three invariant mass values contained in the same event). Run, subrun, event and neutrino ID numbers (unique number defining each candidate in an event) can be used to correlate one to the other. The use of HELEPHANTE significantly reduces the size of the dataset and makes it available for offline analysis (e.g., for the *on-beam* HNL sample, the sample size reduces from  $\approx 4$  TB to  $\approx 100$  MB, in the minimal configuration). Several options are available in HELEPHANTE which can increase or reduce the number of data products available to the analyser according to the needs of a particular analysis. E.g., storing hits coordinates in the HELEPHANTE output can be enabled to visually inspect the HNL candidates although this can significantly increase the size of the file due to the  $\mathcal{O}(100)$  hits associated with each candidate track.

### 6.1.1 Candidate Identification

Candidate identification is performed on a sample of PFParticles that are produced by the *PandoraNu* algorithms and that have not been filtered out by the *PandoraCosmic* selection. The *PandoraCosmic* pass removes tracks that can be unambiguously identified as CR muons. A series of requirements is then imposed on the *PandoraNu* PFParticles, which aims at selecting HNL candidates.

- **C1**, *PFParticle must be a primary*. This identifies reconstructed neutrino interaction vertices identified by Pandora. Although a neutrino flavour tag is associated to the neutrino Pandora PFParticle, this information is not used since Pandora assigns flavours based on the expected topology of a neutrino interaction.
- **C2**, *The primary PFParticle must have two associated track-like PFParticle daughters*. This selection aims at finding the muon and pion associated with the HNL.
- **C3**, *No additional daughter particles must be associated to the primary PFParticle*.

This selection ensures that with the exception of the two track-like daughters no other activity is observed near the vertex.

- **C4**, *The PFPparticle must be contained in a pre-selection volume, defined in table 6.1.* The selection allows for a direct conversion to and from *wire-tick* coordinates (hit coordinates in each 2-d plane view, see section 3) and  $x$ ,  $y$  and  $z$  coordinates.

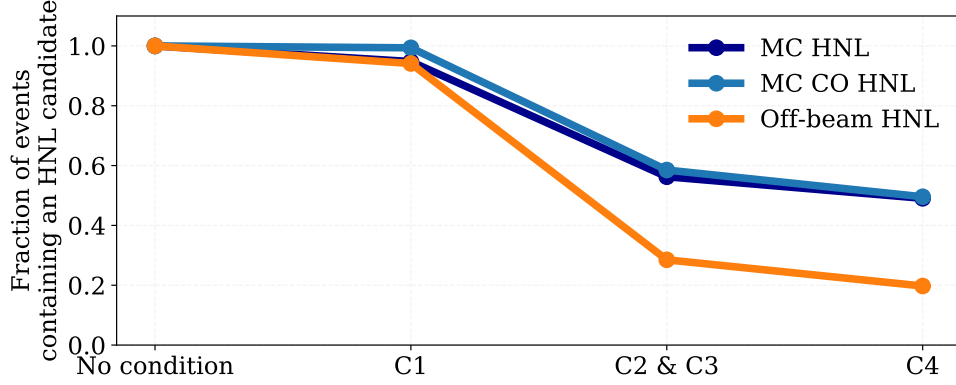
<i>Pre-selection volume selection values</i>		
<b>Coordinate</b>	<b>Minimum</b>	<b>Maximum</b>
$x$	10 cm	246.35 cm
$y$	-105.5 cm	107.5 cm
$z$	10.1 cm	1026.9 cm

**Table 6.1:** Boundaries of the volume in which HNL candidates are built.

For PFPparticle that have fulfilled these requirements, *DecayVertex* objects are constructed by HELEPHANTE. The *DecayVertex* object contains pointers to the data objects representing the two reconstructed 3-d tracks and the 3-d vertices and 2-d hits associated to them. From those it constructs HNL-related information, such as opening angle between the two tracks, HNL invariant mass and direction, etc. The efficiency for the number of events that pass all these requirements and contain at least one successfully reconstructed candidate is shown in figure 6.1 for the MC generated signal sample and the *off-beam* HNL background sample.

Figure 6.1 shows that the presence of additional CR muons in the MC CO HNL sample (see section 5.7) increases the fraction of events surviving the selection, because despite each event containing only one MC “truth” HNL, multiple HNL candidates can be constructed in each event from the CR muons. Figure 6.2 shows the number of HNL candidates found in each event of MC and data.

Despite the fact that  $> 90\%$  of HNL events contain a fully reconstructed neutrino candidate, requiring exactly two tracks decreases the efficiency to  $\approx 50\%$ , since  $> 40\%$  of the reconstructed neutrino vertices in the HNL sample have  $\geq 1$  showers. Figure 6.3 shows the daughter particle multiplicities for each of the reconstructed neutrinos in the MC HNL and *off-beam* HNL samples with CR muons only. The majority of reconstructed neutrino objects in the MC HNL sample contains two tracks as direct daughters ( $\approx 45\%$ ), while the second most likely topology is a track-shower object.

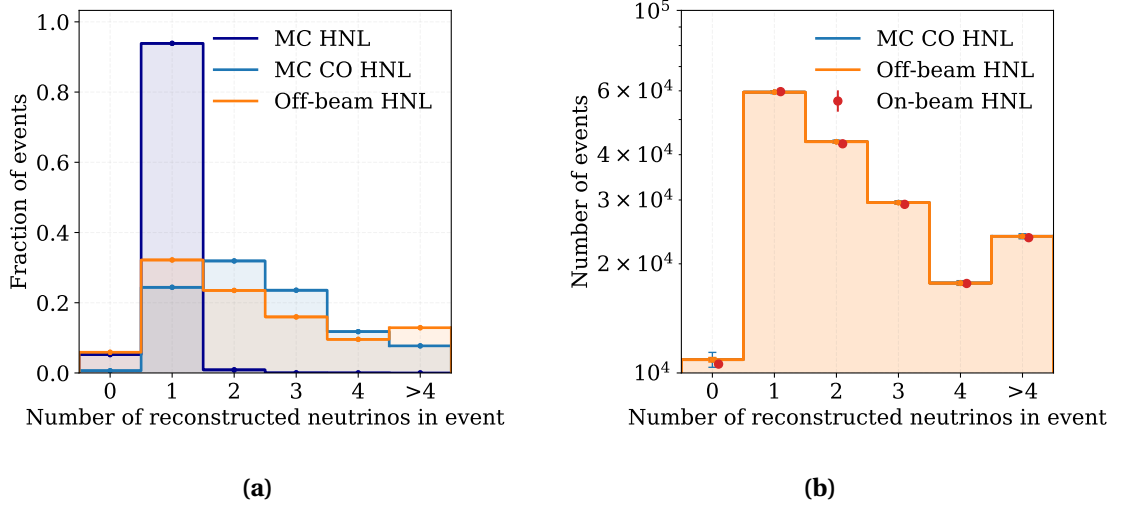


**Figure 6.1:** Fraction of events containing an HNL candidate for the simulated signal-only sample (MC HNL in light blue), simulated signal with background sample (MC CO HNL in dark blue) and the background *off-beam* HNL sample from the EXT-HNL trigger (orange). A sample HNL with a 370 MeV mass is used and  $|U_{\mu 4}|^2 = 10^{-7}$ . The effects of the candidate identification requirements also increase the signal-to-background ratio by a factor of 2.5.

This can be caused, for example, by a low-energy pion produced in the decay which decays into a muon subsequently decaying to an electron ( $\pi \rightarrow \mu \rightarrow e$ ), with both tracks having a length smaller than the TPC resolution, due to their very low energy. An example MC event of this is shown in figures 6.4 and 6.5.

Additional HNL events might be recovered by allowing the DecayVertex object to use track-shower objects as well for the determination of HNL candidates. Albeit this would correspond to a  $\approx 40\%$  increase in the amount of collected events, the impact on the sensitivity would not scale linearly since reconstructed kinematic variables from a track-shower object (such as invariant mass, opening angle and directionality) have a worse resolution than the ones obtained from a track-track object, which could degrade the efficiency of the BDT.

The normalised absolute number of HNL events remaining after candidate identification, for a fixed  $|U_{\mu 4}|^2$ , is obtained by scaling the total number of MC HNL events generated to the number of expected events (see figure 5.8) and multiplying by the timing and trigger efficiency and the two-track selection efficiency. Figure 6.6 shows the effects on the efficiency of requiring the events to pass the trigger and to contain at least an HNL candidate (two-track object within containment volume).

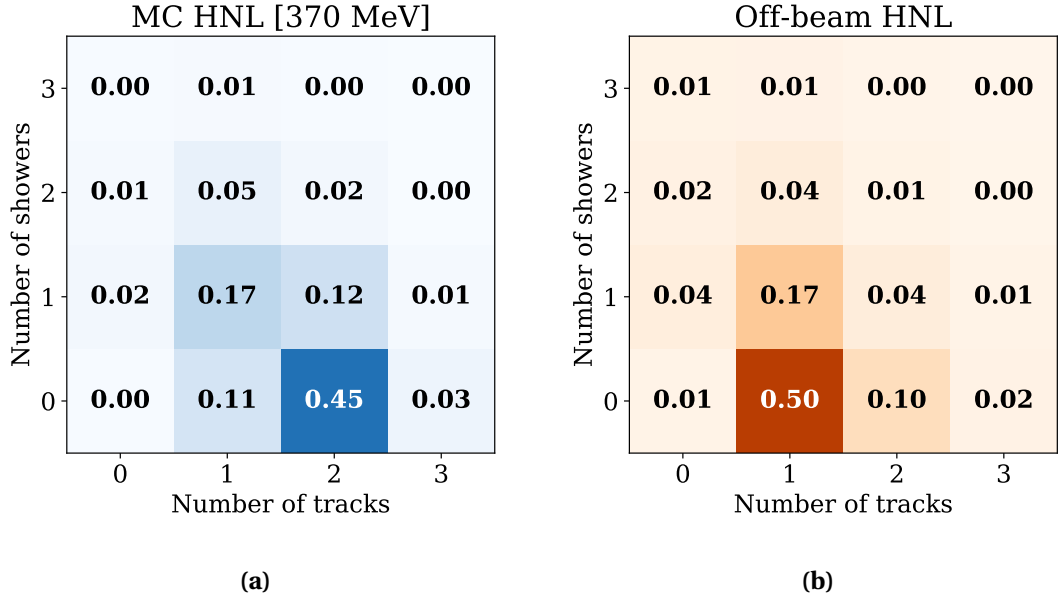


**Figure 6.2:** Distribution for the number of reconstructed neutrinos (before a selection on track multiplicity is applied) per event. Figure (a) shows the fraction of events associated with a neutrino multiplicity for MC HNL samples and the background *off-beam HNL* data. The distribution shows how in the MC HNL sample the vast majority of events contain a successfully reconstructed neutrino object and that a small fraction ( $< 10\%$ ) does not contain a reconstructed neutrino object. In the case of data CR samples (or MC CO) events may contain multiple HNL candidates, which increases the average event multiplicity. Figure (b) shows the distribution of the number of events in each category for data (*on-beam HNL* in red), background (*off-beam HNL* in orange) and simulated signal (MC CO HNL). The signal is simulated with a mass of 370 MeV and  $|U_{\mu 4}|^2 = 10^{-7}$ , and is stacked on top of the background distribution, although the low signal-to-background ratio before any selection does not allow to see the signal in this figure.

The efficiencies are calculated at the *event level*. While an event might pass the selection criteria, multiple HNL candidates can be found within it, and none of them is guaranteed to be the correct one, e.g. all HNL candidates in the event might be from CR muons, despite an HNL being present in the event.

Figure 6.6a shows the timing and trigger efficiency, for the *generated* events that pass the HNL trigger requirement, while vetoing the BNB software trigger for both the pure MC HNL sample and the MC CO HNL sample. This efficiency does not represent a trigger efficiency, which is instead discussed in section 4.2, since MC events are generated in a larger timing window surrounding the trigger window to better estimate resolution effects. This efficiency thus convolves the intrinsic efficiency of the trigger

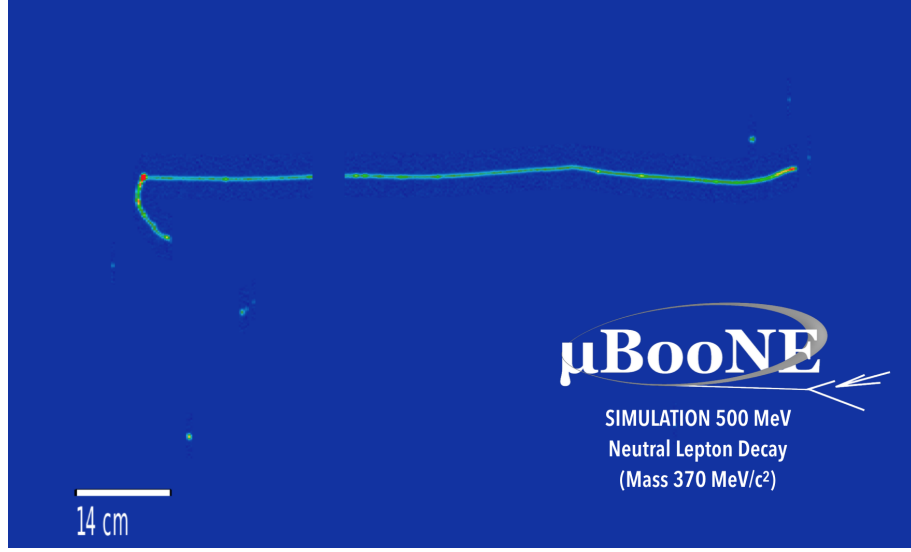




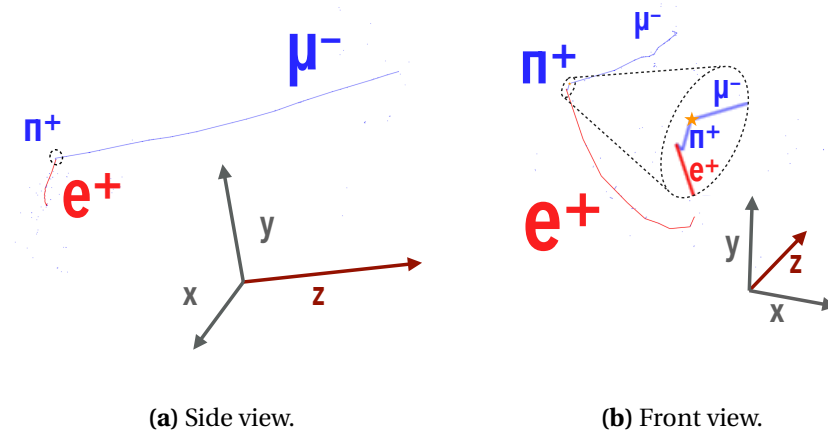
**Figure 6.3:** Number of track and shower daughters associated to each reconstructed neutrino for two different sample types, *MC CO HNL* (a) and *off-beam HNL* (b). The *MC HNL* candidates are more likely to have a two-track object reconstructed ( $\approx 45\%$ ), with a track-shower topology being the second most likely. For the *off-beam HNL CR* muon sample the most likely topology is a single track object ( $\approx 50\%$ ) while the fraction of two-track objects amounts to only  $\approx 10\%$ .

and the amount of *HNLs* generated outside of the trigger window and it is dependent on the timing range used for the generation of the *MC* sample. Its use is thus exclusively for normalisation purposes. The figure also shows how the use of *CO* allows for the estimation of the trigger efficiency in *MC*. *CR* muons are present across the entire width of the timing window and might fulfil the *BNB* software trigger condition, thus preventing an event from passing the selection. The effect is independent of the *HNL* mass, being exclusively due to *CR* muons and it lowers the passing fraction by few percents.

Figure 6.6b shows the efficiency for the requirement that at least one reconstructed two-track object contained in the volume defined in table 6.1 be present in the event. The seemingly higher efficiency of the *MC CO HNL* sample is an artefact of calculating the efficiency at the event level. The presence of *CR* muons in the event increase the number of *HNL* candidates in an event thus increasing the likelihood that at least one candidate is present, even though it might not necessarily correspond to a real



**Figure 6.4:** Simulated MC HNL decay (370 MeV mass with an energy of 500 MeV) to a  $\mu\pi$  pair resulting into a reconstructed track-shower object. The shower comes from an electron produced by a quick  $\pi \rightarrow \mu \rightarrow e$  decay, occurring at the vertex within a region below MicroBooNE spatial resolution.



**Figure 6.5:** MC “truth” information for the event shown in figure 6.4. Figure (a) shows a lateral view, from behind the anode plane. The muon is moving in the beam direction, along the  $z$  axis. Figure (b) shows the same event from a frontal view (looking down the beamline) with a zoom around the decay vertex (orange star). A pion quickly decays in a chain, producing an electron which is then used by the reconstruction to create the track-shower object. Comparing with figure 6.4 shows that the decay occurs below MicroBooNE spatial resolution.

HNL decay. The efficiency in [figure 6.6b](#) is calculated after requiring the event to have passed the trigger requirements so that the denominator is the number of events passing the trigger. [Figure 6.6c](#) shows the combined effects of both selection, which is approximately constant across all mass values probed, sitting at  $\approx 15\%$ .

Sample	In Timing Window	Trigger and Timing Selection	Efficiency
MC HNL	276	100	$0.362 \pm 0.001$
MC CO HNL	276	98	$0.355 \pm 0.001$

**Table 6.2:** Expected number of events for a HNL with a mass of 370 MeV,  $|U_{\mu 4}|^2 = 10^{-7}$  and  $1.97 \times 10^{20}$  POTs. The second column correspond to the expected number of events in the timing window [4,725 ns, 5,500 ns] surrounding the HNL software trigger, while the third column is the number of events surviving after the HNL software trigger with a veto on the BNB software trigger has been applied. The last column shows the efficiencies for this selection.

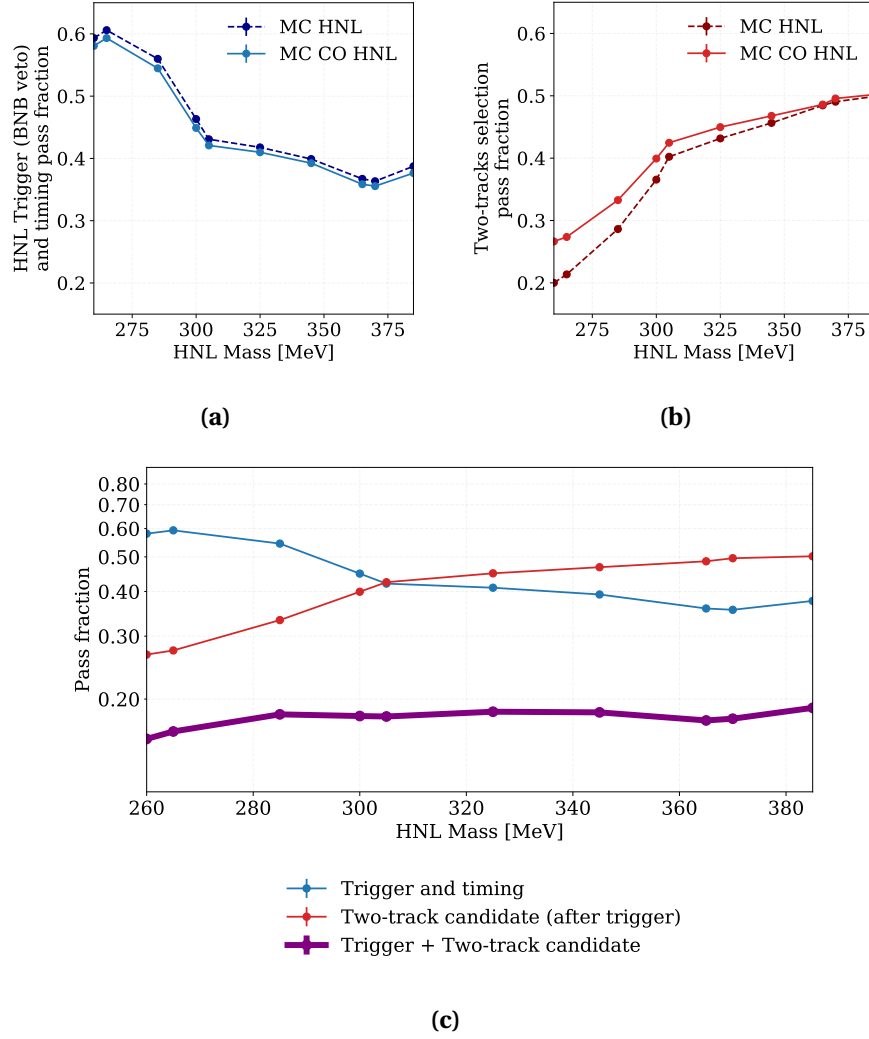
Sample	Before Selection	Two-Tracks Selection	Efficiency
MC HNL	100	49	$0.490 \pm 0.001$
MC CO HNL	98	49	$0.500 \pm 0.001$
<i>On-beam HNL</i>	183,591	36,803	$0.200 \pm 0.001$
<i>Off-beam HNL</i>	184,828	36,488	$0.197 \pm 0.001$
<i>On-beam BNB</i>	269,878	81,247	$0.301 \pm 0.001$
MC BNB	226,476	70,321	$0.310 \pm 0.001$

**Table 6.3:** Expected number of events for a HNL with a mass of 370 MeV,  $|U_{\mu 4}|^2 = 10^{-7}$  and  $1.97 \times 10^{20}$  POTs. The second column correspond to the expected number of events before candidates are identified, while the third column is the number of events surviving the requirement that at least one two-track HNL candidate object is present in the event. The last column shows the efficiencies for this selection.

### 6.1.2 High-level Variables

Variables such as the 3-d opening angle can be calculated by HELEPHANTE and assigned to each HNL candidate in an unequivocal way (e.g., the opening angle is the angle between the two tracks in the plane determined by their associated vectors). For other variables, such as momenta and mass, certain assumptions or models must be used.

## 6.1. SIGNAL DEFINITION



**Figure 6.6:** Mass-dependent efficiencies at the event level for the timing and trigger selection (a) and contained two-track object selection (b) for the pure MC HNL sample and the MC CO HNL sample. The two-track selection efficiency in (b) is calculated *after* the timing and trigger selection has been applied. The denominator is the number of events passing the timing and trigger selection. Figure (c) shows in purple the total efficiency (applying both triggers and contained two-track selection) for the MC CO HNL sample only. Overlapping in red and blue are the two individual efficiencies contributing to the total. Statistical uncertainties are negligible due to the high number of MC statistics involved.

The momentum and energy of each track can be determined in MicroBooNE via three methods:

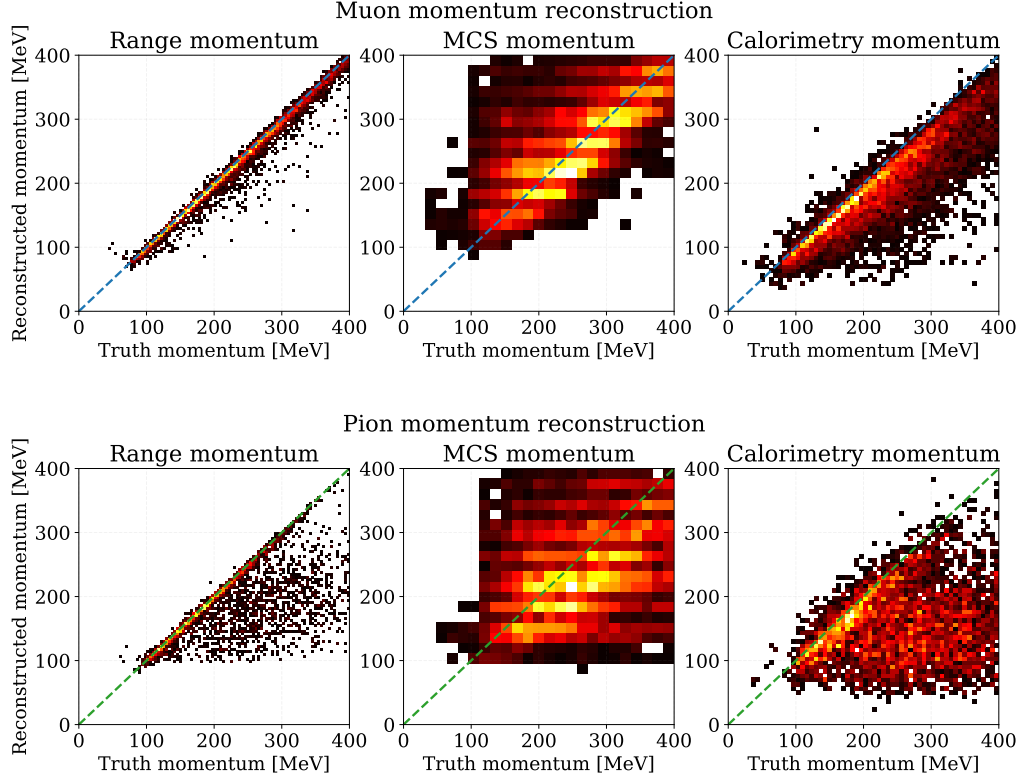
- **Range**, which uses the particle's track length and its mass to determine its energy. By integrating over energy loss (see Bethe-Bloch, [equation \(3.1\)](#)):

$$R(T) = \int_0^T \left[ \frac{dE}{dx} \right]^{-1} dE, \quad (6.1)$$

where  $R$  is the track length and  $T$  its total energy. [Equation \(6.1\)](#) shows that, provided the particle comes at rest and does not re-interact, the length can be used to determine its original starting energy. Range can be an effective method for measuring the particle momentum, although the efficiency is dependent on whether the particle is fully contained and comes at rest.

- **Multiple Coulomb Scattering (MCS)**, which uses the many small-angle scatters of the charged particle, mostly due to Coulomb scattering off the nuclei, as it traverses the medium. The mean angle can be used to determine the momentum of the particle. One of the main advantages of [MCS](#) is that its efficiency is not dependent on containment or the particle coming at rest.
- **Calorimetry**, which integrates the  $dE/dx$  depositions from the single hits to obtain the total energy. Its efficiency depends on the particle being fully contained and coming at rest. The calorimetry method provides the kinetic energy  $T$ , so that the momentum  $p$  must be determined via  $p = \sqrt{(T + m)^2 - m^2}$ , thus requiring the mass to be known.

HNLGen can be configured to simulate [HNL](#) decays in *single particle mode*. This produces the same standard decays produced in the regular samples, but masks one of the two particles, so that it is not tracked by [G4](#) or the detector simulation. This allows to study single particle reconstruction efficiency, while maintaining the kinematics and energy distributions of [HNL](#) decays. Using *single particle mode* and a sample [HNL](#) with mass 370 MeV, the momentum reconstruction efficiency for the three methods are determined. [Figure 6.7](#) compares the reconstructed momentum with the generated momentum for both muons and pions. The figure shows range method provides the best efficiency for momentum reconstruction.



**Figure 6.7:** Generated ( $x$  axis) particle momentum versus reconstructed ( $y$  axis) particle momentum using the range method (left column), MCS (centre column) and calorimetry (right column) for muons (top row) and pions (bottom row). For the range and calorimetry methods the correct mass has been used for the momentum estimation (assuming perfect particle identification). The horizontal features in the MCS momentum are due to the limited size of the steps of the fitting algorithm. To obtain the pion range momentum, the input range value and output momentum was rescaled.

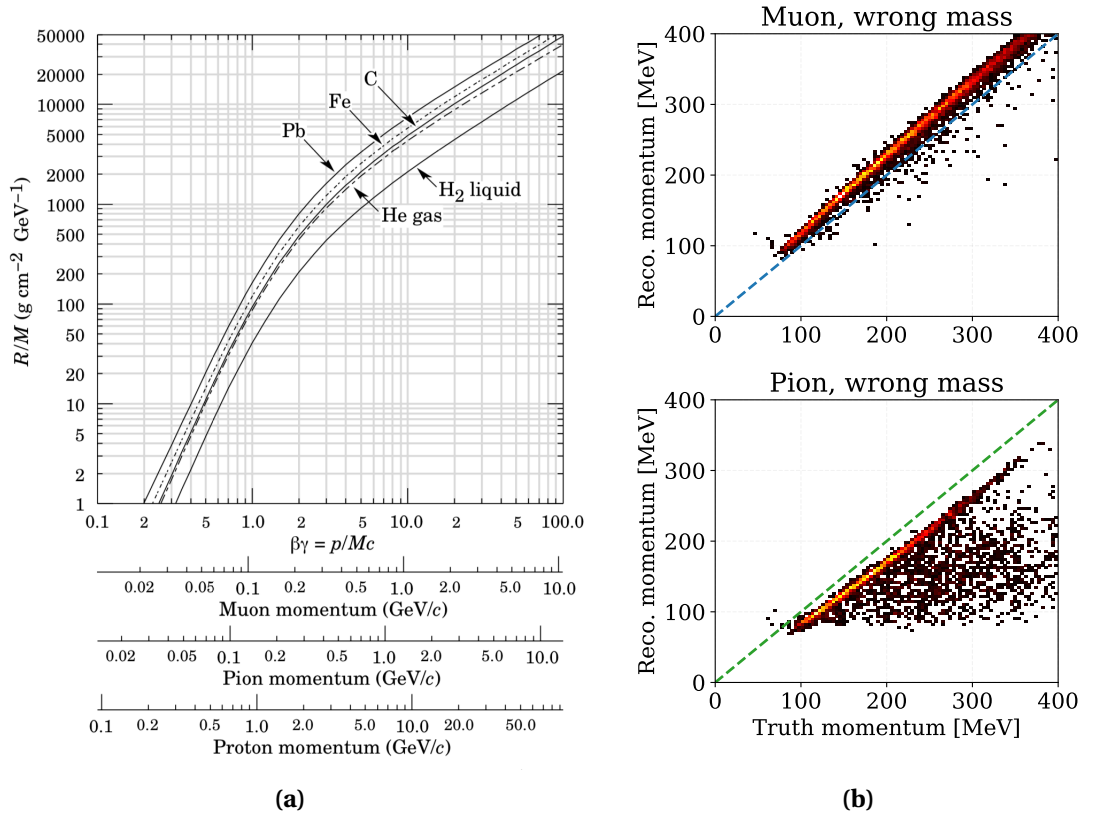
### Determining Momentum by Range

Although the range method has a better resolution than the MCS method, it requires an assumption to be made about the mass of the particle. Since pions and muons are difficult to discriminate in liquid argon, the range algorithms in LArSoft provide the option for determining the momentum of the tracks for muons and protons only (no option for pion is provided). This is done by interpolating the range obtained for the track to range-momentum tables from Ref. [75] for muons and from the PSTAR programme [76] for protons. In order to use the same algorithm which transforms

$R/M$  to  $\beta\gamma = p/M$  (see figure 6.8a), we scale the input range by  $m_\mu/m_\pi$  and the output momentum by  $m_\pi/m_\mu$ , as shown in equation (6.2).

$$f\left(\frac{R}{m_\mu}\right) = \beta\gamma m_\mu = p \quad \rightarrow \quad f\left(\frac{R}{m_\mu} \cdot \frac{m_\mu}{m_\pi}\right) = \beta\gamma m_\mu \cdot \frac{m_\pi}{m_\mu} = p \cdot \frac{m_\pi}{m_\mu}, \quad (6.2)$$

where  $f$  is the interpolating function, which takes as input a range value and provides a momentum value. If the default algorithm is used for pions, the reconstructed momentum would be underestimated (see figure 6.8b).



**Figure 6.8:** Function used by the interpolating function in LArSoft to determine the range momentum (a). It is expressed in terms of  $R/M$  and  $\beta\gamma = p/M$ , which require rescaling for pions. The figure is taken from Ref. [22]. Using the wrong assumption causes the algorithm to overestimate the momentum for muons and to underestimate the momentum for pions (b).

## Mass Hypothesis

From the reconstructed momentum vectors and energies of the two particles the total momentum and total energy is derived. Since the decay channel studied in this analysis is a two-body decay, these values are assigned to the **HNL**. From the **HNL** momentum vector, the angles ( $\theta$  and  $\phi$ ) and invariant mass of the **HNL** is also derived. All these quantities are dependent on the assumption made about the two particles. Although it is assumed that the two tracks in a **HNL** candidate are a muon and a pion, there is no reliable method to perform particle identification. We calculate all variables which are dependent on the particle identification assumption in both hypothesis, which we order based on track length:

- **Hypothesis 1:** Longest track is a muon, shortest track is a pion.
- **Hypothesis 2:** Longest track is a pion, shortest track is a muon.

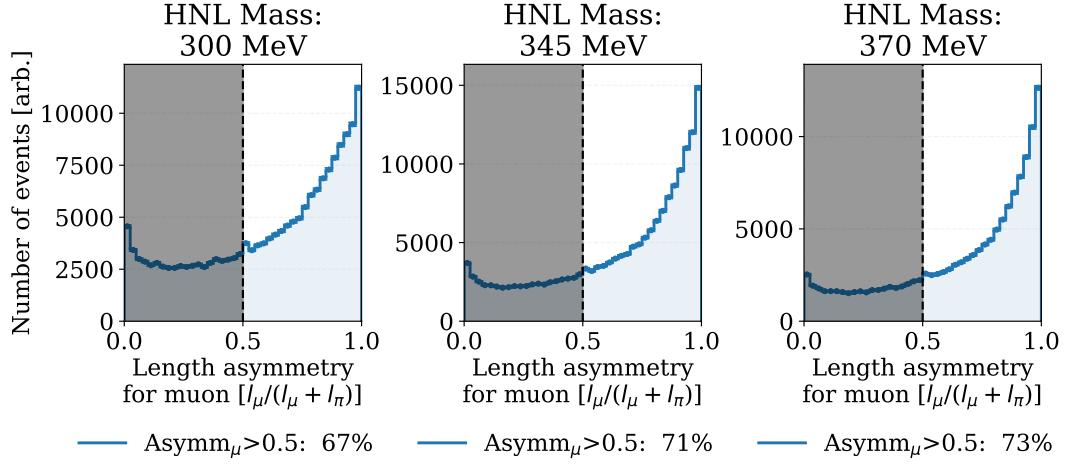
*Hypothesis 1* is more likely to be correct for all mass values, since the muon is more likely to travel across the detector without re-interacting. This is illustrated in [figure 6.9](#), which shows the ratio between the length of the muon and the sum of the lengths of the two tracks for different **HNL** mass values. We use thus *Hypothesis 1* as the choice for the determination of all the kinematic variables assigned to the **HNL** candidate. *Hypothesis 1* is used for all the figures and results shown in the rest of the thesis, unless otherwise stated.

Due to the similar mass values of the muon and the pion, the choice of hypothesis has a small impact on the determination of the kinematic quantities. [Figure 6.10](#) shows the reconstructed invariant mass resolution in the two hypotheses for different **HNL** mass values.

## 6.2 Pre-selections

Once **HNL** candidates have been selected, a series of pre-selection requirements are imposed which have the two-fold purpose of improving the quality of the candidates



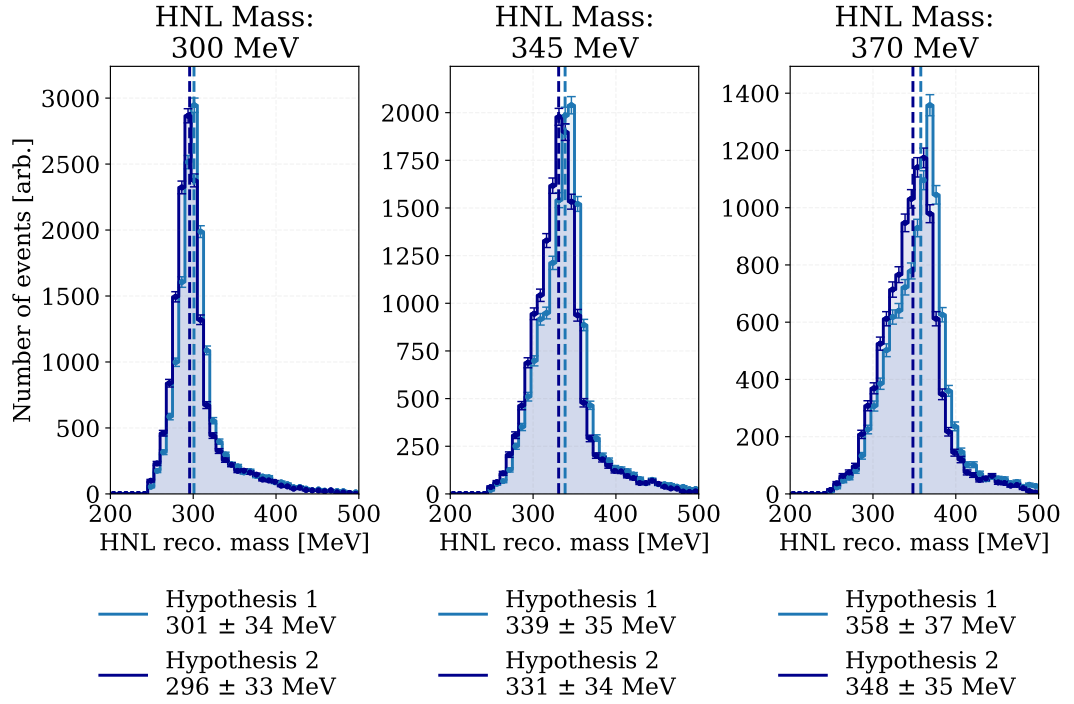


**Figure 6.9:** Track length asymmetry between the muon and the pion resulting from a HNL decay, calculated as the muon length divided by the sum of the lengths of the two tracks. A value of the variable higher than 0.5 indicates that the muon track is longer than the pion track. The figures are produced before any pre-selection (only a two-vertex object is required). The default assumption for track reconstruction assumes that the longest track is a muon. The figure shows that the assumption is correct in  $\approx 70\%$  of cases.

(by removing candidate objects which are likely to be misreconstructed) and of reducing the cosmic background (the pre-selection affects largely the cosmic background sample). At this stage, multiple candidates per event are considered and *ensemble* distributions of all the candidates found in a sample are analysed. Table 6.4 shows a summary of the pre-selection requirements applied to the samples.

### 6.2.1 Fiducial Volume

The *fiducial volume* (FV) is an inner region of the detector, fully contained in the TPC, that is defined to characterise HNL candidates that are less likely to be misreconstructed due to edge effects or due to dead wires. Reconstructed HNL candidates near the boundaries of the TPC are more likely to have their tracks only partially reconstructed. Analogously, vertices situated in a dead region of the detector might be mis-reconstructed. In addition, the placing of geometrical selections at the edges of the detector has the added benefit of removing a large fraction of background CR



**Figure 6.10:** Reconstructed **HNL** invariant mass using the momenta of the two daughter particles as derived by the range method. The invariant mass is shown for both hypothesis cases and is calculated after all the pre-selection requirements are applied (which are discussed further in this section). The label reports the median and standard deviation of the (non-Gaussian) distributions.

muons, thus achieving the dual purpose of improving the quality of the signal and increasing the signal-to-background ratio.

Figure 6.11 shows a schematic of the  $x$ ,  $y$  and  $z$  boundary regions used to define the fiducial volume. The larger 35 cm buffer region at the top and bottom of the detector is used to remove a larger fraction of **CR** muons since through-going muons have their vertices preferably placed where the **CR** muons enters or exits the detector. A larger 85 cm buffer region is also used in the downstream section of the TPC to improve the quality of the selected **HNL** events, by allowing forward-going tracks to have enough space to be correctly reconstructed. The 100 cm gap between 675 cm and 775 cm is due to a section of the detector in which a set of shorted wire on the collection plane is responsible for a dead region (see figure 5.14).

## 6.2. PRE-SELECTIONS

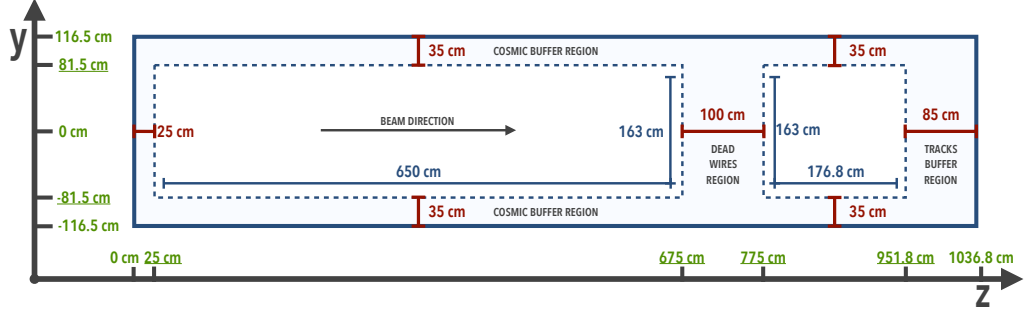
Pre-selection requirements		
Name	Variable Used	Requirement
Fiducial volume	HNL vertex $x$ coordinate	$> 12 \text{ cm and } < 244.35 \text{ cm}$
	HNL vertex $y$ coordinate	$> -80.5 \text{ cm and } < 80.5 \text{ cm}$
	HNL vertex $z$ coordinate	$(> 25 \text{ cm and } < 675 \text{ cm}) \text{ or } (> 775 \text{ cm and } < 951.8 \text{ cm})$
Vertex-track distance	Distance between vertex and farthest track start point	$< 5 \text{ cm}$
Minimum number of hits	Number of hits of smallest track	$> 30 \text{ hits on collection plane}$
Flash PE	PE of largest flash in event	$> 0 \text{ PE}$
Vertex-flash distance	2-d distance between HNL and largest flash	$< 150 \text{ cm}$
Track containment	$x$ coordinate of end point farthest from centre	$> 12 \text{ cm and } < 240 \text{ cm}$
	$y$ coordinate of end point farthest from centre	$> -98 \text{ cm and } < 98 \text{ cm}$
	$z$ coordinate of end point farthest from centre	$> 15 \text{ cm and } < 1010 \text{ cm}$
Opening angle	3-d angle between tracks	$< 2.8 \text{ radians } (160^\circ)$
Invariant mass	Range-calculated HNL candidate invariant mass	$< 0.5 \text{ GeV}$

**Table 6.4:** Selection flow showing the pre-selection requirements imposed on all the samples used for analysis.

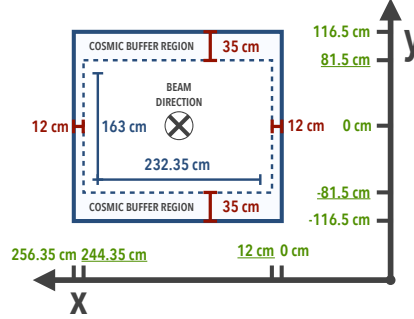
Figure 6.12 shows the effect of the FV requirements on the distributions of re-constructed vertex location distributions for the HNL and *off-beam BNB CR* muons sample and table 6.5 shows the efficiency for this selection.

### 6.2.2 Vertex-Track Distance

The Pandora reconstruction algorithm can occasionally misreconstruct events. A flag of the occurrence of misreconstruction is the start of a Pandora secondary track being situated at a large distance from the neutrino vertex that originated it. We thus select only HNL candidates where both tracks start at a distance smaller than 5 cm from the neutrino vertex location. The selection requirement reduces the efficiency of signal by 8% and for background by 25%. Figure 6.13 shows how the variable used to place



(a) Side view.



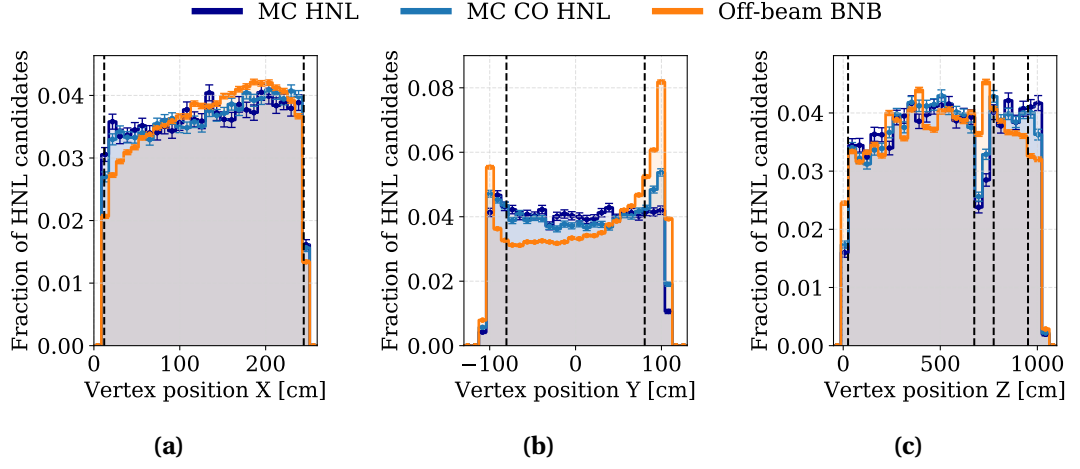
(b) Front view.

**Figure 6.11:** Fiducial volume used for this analysis. Solid blue line represents the instrumented TPC. The white region surrounded by the dashed blue lines is the fiducial volume. Events occurring in the light blue regions are thus rejected.

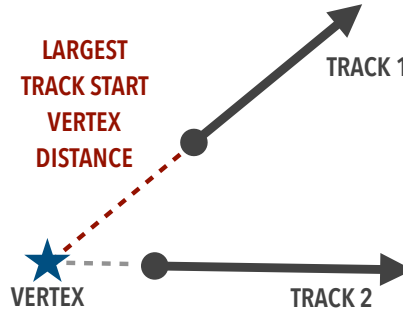
the selection is determined and [figure 6.14](#) shows the effects on the distribution for different sample types. [Table 6.6](#) shows the effect of the selection on the efficiency and the number of events resulting after the selection.

### 6.2.3 Minimum Number of Hits

The Pandora reconstruction algorithm uses clusters of hits for constructing track objects. A minimum number of hits (30) on the collection plane is required for both tracks in order to increase the quality of the candidates. [Figure 6.15](#) shows the effect on the distribution for different sample types. The number of hits is correlated to track



**Figure 6.12:** Reconstructed position of the HNL candidate vertex in  $x$  (a),  $y$  (b) and  $z$  (c) coordinates. The rise in the  $x$  direction is due mostly to SCE. The selection applied on the  $y$  coordinate has a larger effect on the CR sample, due to many candidates resulting from through-going CR muons starting at the top or the bottom of the detector. In the  $z$  dimension, the dip in the region between 675 cm and 775 cm is caused by a set of unresponsive wires, and the candidates whose coordinates are contained between these two  $z$  values are rejected. Histograms are not stacked and they are area normalised to illustrate the fraction of candidates passing the selection for each sample. The HNL samples are shown for a HNL mass of 370 MeV.



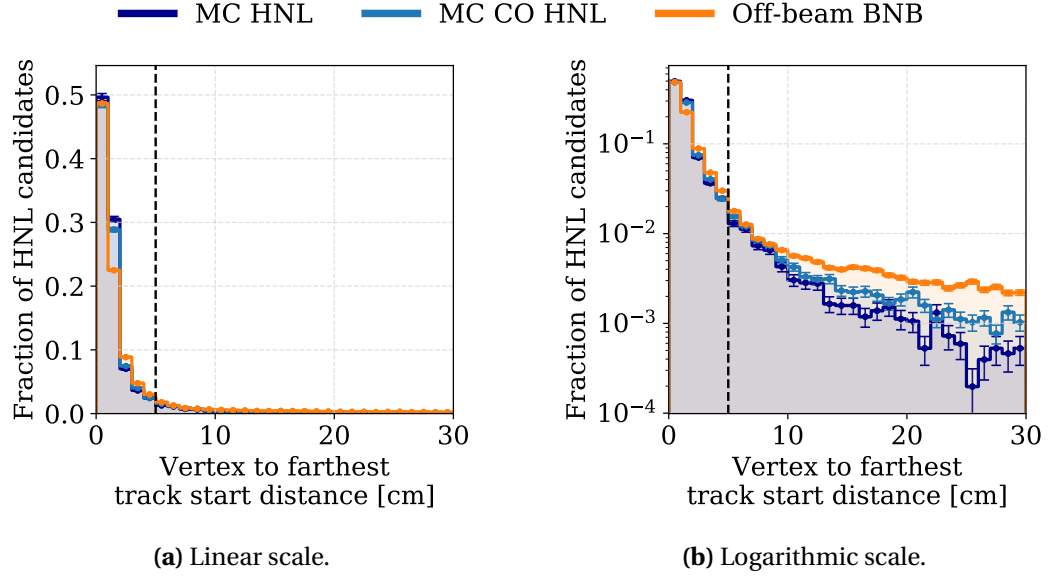
**Figure 6.13:** Illustration of the variable used for the determination of the vertex-track distance selection.

length (and directionality), since it corresponds to a selection on the length projected on the  $z$  axis (beamline direction). The selection has a minimal impact on both signal and background, although the impact increases for smaller HNL masses, which have shorter tracks. Table 6.7 reports the effect of the selection on the efficiency and the number of events resulting after the selection.

*Fiducial volume selection efficiency*

Sample	Efficiency (this selection)	Cumulative efficiency	Number of candidates
MC HNL	0.613 ( $\pm 0.004$ )	0.613 ( $\pm 0.002$ )	30
MC CO HNL	0.584 ( $\pm 0.003$ )	0.584 ( $\pm 0.002$ )	32
Off-beam BNB	0.516 ( $\pm 0.002$ )	0.516 ( $\pm 0.002$ )	68,955
MC BNB	0.534 ( $\pm 0.004$ )	0.534 ( $\pm 0.002$ )	42,414
On-beam BNB	0.531 ( $\pm 0.002$ )	0.531 ( $\pm 0.002$ )	43,078
On-beam HNL	0.520 ( $\pm 0.003$ )	0.520 ( $\pm 0.002$ )	21,811
Off-beam HNL	0.519 ( $\pm 0.005$ )	0.519 ( $\pm 0.002$ )	21,501

**Table 6.5:** *Fiducial volume selection efficiencies for different sample types.* The *efficiency* column shows the fraction of HNL candidates surviving this specific selection, from the previous selection. *Number of candidates* shows the number of candidates that are left in each data sample after applying this selection and after normalising to  $1.97 \times 10^{20}$  POTs. The HNL samples are shown for a HNL mass of 370 MeV and  $|U_{\mu 4}|^2 = 10^{-7}$ . Uncertainties shown are statistical only.

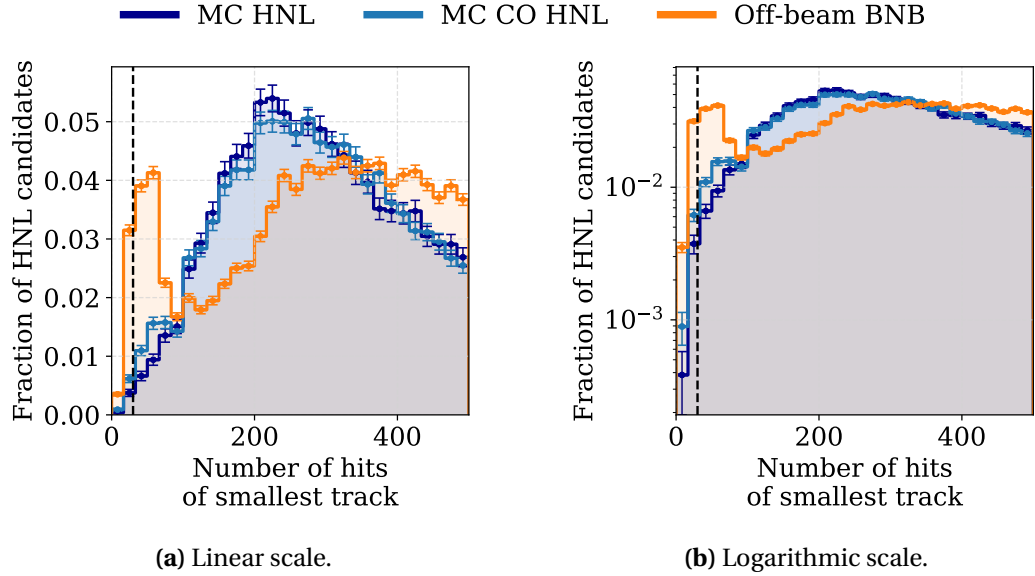


**Figure 6.14:** Distance between the reconstructed vertex and the start point of the track which is located farthest away from the vertex (as shown in figure 6.13). The black dashed line correspond to the selection placed on the samples (candidates to the left of the line are kept). Format used for this figure is identical to the format used for figure 6.12.

Vertex-track distance selection efficiency

Sample	Efficiency (this selection)	Cumulative efficiency	Number of candidates
MC HNL	0.925 ( $\pm 0.002$ )	0.567 ( $\pm 0.002$ )	28
MC CO HNL	0.857 ( $\pm 0.002$ )	0.501 ( $\pm 0.002$ )	27
Off-beam BNB	0.751 ( $\pm 0.002$ )	0.387 ( $\pm 0.002$ )	51,772
MC BNB	0.766 ( $\pm 0.004$ )	0.409 ( $\pm 0.002$ )	32,471
On-beam BNB	0.746 ( $\pm 0.002$ )	0.396 ( $\pm 0.002$ )	32,120
On-beam HNL	0.749 ( $\pm 0.003$ )	0.390 ( $\pm 0.002$ )	16,339
Off-beam HNL	0.750 ( $\pm 0.005$ )	0.389 ( $\pm 0.002$ )	16,126

**Table 6.6:** Vertex-Track distance selection efficiencies for different sample types. Format for this table is identical to the format used for [table 6.5](#).



**Figure 6.15:** Number of hits of the track with the smallest number of hits. The black dashed line correspond to the selection placed on the sample (candidates to the right of the line are kept). Format used for this figure is identical to the format used for [figure 6.12](#).

#### 6.2.4 Flash PE

Albeit the HNL software trigger performs already a selection based on light intensity, using information from the raw waveforms of the PMT system, additional discrimination between signal and background can be achieved by looking at the reconstructed optical flash objects, constructed by LArSoft and described in [section 5.6.5](#).

## 6.2. PRE-SELECTIONS

<i>Minimum number of hits selection efficiency</i>			
<b>Sample</b>	<b>Efficiency (this selection)</b>	<b>Cumulative efficiency</b>	<b>Number of candidates</b>
MC HNL	0.9975 ( $\pm 0.0004$ )	0.565 ( $\pm 0.002$ )	28
MC CO HNL	0.9961 ( $\pm 0.0004$ )	0.499 ( $\pm 0.002$ )	27
Off-beam BNB	0.9871 ( $\pm 0.0004$ )	0.382 ( $\pm 0.002$ )	51,105
MC BNB	0.9925 ( $\pm 0.0008$ )	0.406 ( $\pm 0.002$ )	32,228
On-beam BNB	0.9944 ( $\pm 0.0004$ )	0.394 ( $\pm 0.002$ )	31,939
On-beam HNL	0.9870 ( $\pm 0.0009$ )	0.385 ( $\pm 0.002$ )	16,126
Off-beam HNL	0.989 ( $\pm 0.001$ )	0.385 ( $\pm 0.002$ )	15,942

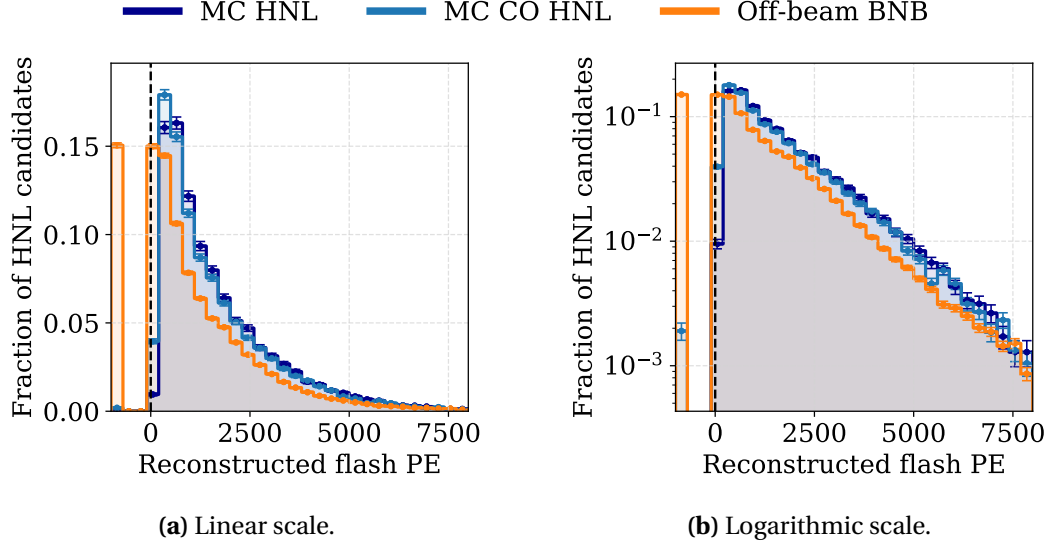
**Table 6.7:** *Minimum number of hits* efficiencies for different sample types. Format for this table is identical to the format used for [table 6.5](#).

Reconstructed optical flashes correlate light produced by adjacent PMTs. The PMT PE intensity is correlated with the amount of photons released in the area surrounding the vertex and can be used to discriminate signal from background CR muons. In addition, the *lack* of a reconstructed flash can also be used for discrimination. Albeit CR muons may produce sufficient light in the detector to fulfil the trigger condition, the spatial correlation of the light produced is not sufficient to create a reconstructed optical flashes. To each HNL candidate we associate the reconstructed flash with the largest PE in the event. This means that if multiple HNL candidates are present in the event, one of which may be the generated HNL and the others are from CR muons, *all* candidates get assigned the same PE value. As a consequence, this selection requirement operates at the *event-level*, and not at the *candidate-level*, like the other selection requirements in this section (e.g. HNL candidates from CR muons are not removed as long as they are in the same event as a successfully reconstructed HNL).

We perform a selection on HNL candidates with an assigned PE larger than 0. The selection has a particular high efficiency for signal, close to 100%, since events have already been selected by the software trigger for sufficient light production, and in all cases the spatial correlation is sufficient to create reconstructed optical flashes. The fact that both the signal-only and the signal with CO sample have the same efficiency is due to the fact that the presence of a HNL is sufficient to have an event pass the selection (and all the other candidates contained within). For background the rejection efficiency is  $\approx 15\%$ . [Figure 6.16](#) shows the effect on the distribution for different sample types. [Table 6.8](#) reports the effect of the selection on the efficiency and the number of



events resulting after the selection.



**Figure 6.16:** Number of PE for the largest reconstructed flash in the event in which each candidate is contained. HNL candidates from the same event have the same identical PE score. HNL candidates from an event in which no optical flash has been reconstructed have a -999 PE value assigned. The selection is made for  $PE > 0$  (candidates to the right of the vertical dashed line are kept). Format used for this figure is identical to the format used for figure 6.12.

<i>Flash PE selection efficiency</i>			
Sample	Efficiency (this selection)	Cumulative efficiency	Number of candidates
MC HNL	1	0.565 ( $\pm 0.002$ )	28
MC CO HNL	0.9981 ( $\pm 0.0003$ )	0.498 ( $\pm 0.002$ )	27
Off-beam BNB	0.850 ( $\pm 0.001$ )	0.325 ( $\pm 0.002$ )	43,457
MC BNB	0.822 ( $\pm 0.004$ )	0.334 ( $\pm 0.002$ )	26,484
On-beam BNB	0.962 ( $\pm 0.001$ )	0.379 ( $\pm 0.002$ )	30,713
On-beam HNL	0.646 ( $\pm 0.005$ )	0.249 ( $\pm 0.002$ )	10,421
Off-beam HNL	0.652 ( $\pm 0.007$ )	0.251 ( $\pm 0.002$ )	10,400

**Table 6.8:** Flash distance efficiencies for different sample types. Format for this table is identical to the format used for table 6.5.

### 6.2.5 Vertex-Flash Distance

In addition to PE information, the spatial coordinates of the optical flash can be used to be put into correlation with the projection on the wire plane of the HNL candidate vertex. We perform a simple form of flash-matching used by other MCC8 analyses where we assume that the largest flash produced in an event will be largely correlated with the signal (here, HNL) than the cosmic background. We use as metric the 2-d distance between the  $y$  and  $z$  coordinates of the largest optical flash in the event and the  $y$  and  $z$  coordinates of the HNL candidate.

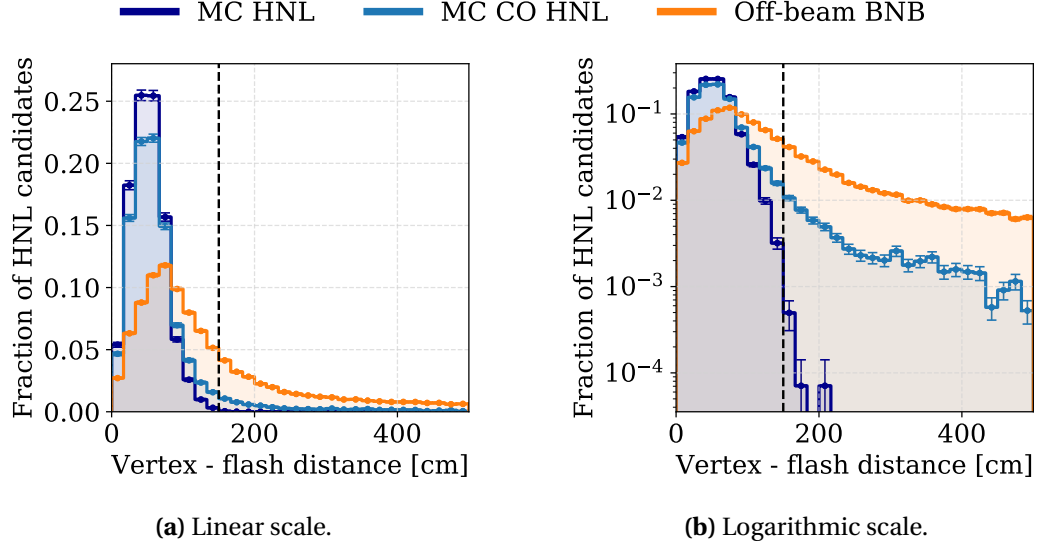
Figure 6.17 shows the effect on the distribution for different sample types. The signal sample has a peak at  $\approx 50$  cm with a narrow  $\approx 30$  cm width, while the cosmics sample has a peak at  $\approx 75$  cm with a broader width and an additional tail caused by uncorrelated flash - (CR) HNL candidate pairs. We remove all the candidates located farther than 150 cm from the largest flash, keeping the signal sample unaltered ( $\approx 100\%$  efficiency) while removing  $\approx 30\%$  of the background cosmics. Table 6.9 reports the effect of the selection on the efficiency and the number of events resulting after the selection.

<i>Flash distance selection efficiency</i>			
<b>Sample</b>	<b>Efficiency (this selection)</b>	<b>Cumulative efficiency</b>	<b>Number of candidates</b>
MC HNL	0.9994 ( $\pm 0.0002$ )	0.565 ( $\pm 0.002$ )	28
MC CO HNL	0.928 ( $\pm 0.002$ )	0.462 ( $\pm 0.002$ )	25
Off-beam BNB	0.649 ( $\pm 0.002$ )	0.211 ( $\pm 0.002$ )	28,190
MC BNB	0.754 ( $\pm 0.005$ )	0.252 ( $\pm 0.002$ )	19,962
On-beam BNB	0.752 ( $\pm 0.003$ )	0.285 ( $\pm 0.002$ )	23,089
On-beam HNL	0.722 ( $\pm 0.005$ )	0.180 ( $\pm 0.002$ )	7,527
Off-beam HNL	0.720 ( $\pm 0.007$ )	0.181 ( $\pm 0.002$ )	7,487

**Table 6.9:** *Flash distance* efficiencies for different sample types. Format for this table is identical to the format used for table 6.5.

### 6.2.6 Track Containment

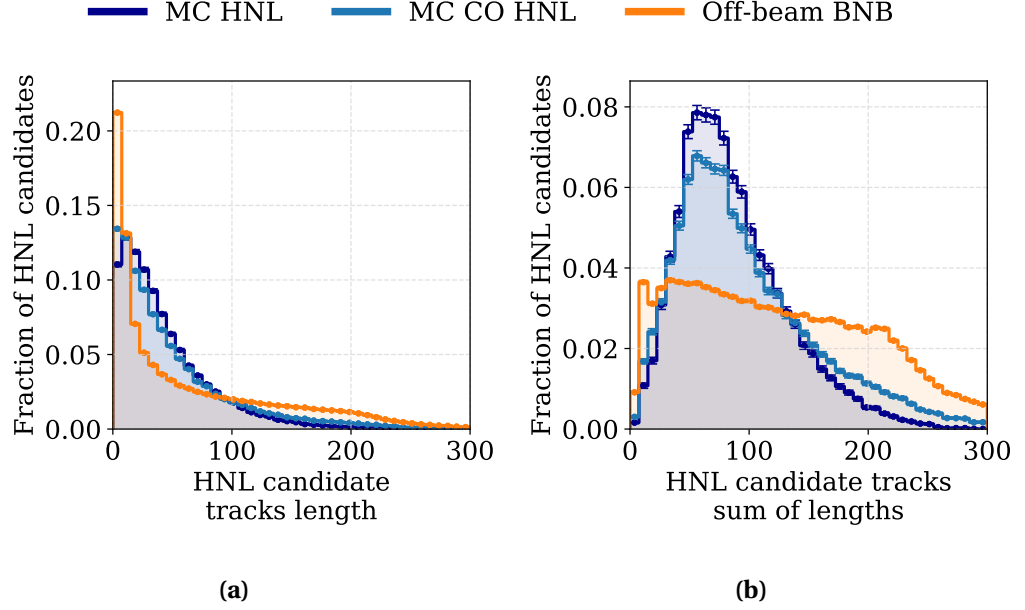
Due to their low energy, caused by the bias of the HNL software trigger, selecting late, low momentum, HNLs, the two tracks associated with a signal candidate have



**Figure 6.17:** 2-d distance between the largest flash and the projection on the  $yz$  plane of the HNL candidate vertex. The black dashed line correspond to the selection placed on the sample (candidates to the left of the line are kept). Format used for this figure is identical to the format used for figure 6.12.

short lengths, averaging around  $\approx 25$  cm (see figure 6.18a). The cosmics background has instead both a larger short-track component in the few cm region, caused by the reconstruction algorithm breaking a track in two near the start/end point, or by  $\delta$  rays, and a larger tail, caused by the remaining part of the CR muon track. This can be seen in figure 6.18b where the sum of the lengths of the two tracks has a much larger components for the cosmic sample, spanning the 250 cm range (corresponding to the width/height of the detector).

Because of these large lengths and their preferential up-down directionality, CR muons often cross the detector with a resulting end track point situated near the top or bottom of the detector. HNL signal candidates, because of their short track length and the effect of the fiducial volume selections which creates a buffer region for the tracks, will have their end point located within the volume with no preference for the edge locations. This is also affected by the larger 85 cm buffer region at the end of the  $z$  axis (see figure 6.11), enforced by the FV selection. HNL tracks travel preferentially in the beam direction and vetoing the last 85 cm provides a section of the detector where the tracks resulting from vertices near the end of the fiducial volume can develop.



**Figure 6.18:** Length of any HNL candidate tracks (a). For each HNL candidate, each of the two secondary track length is measured and added to the histogram (two entries per HNL candidate). In (b) the length of the two-tracks is summed together (one entry per HNL candidate). Histograms are not stacked and they are area normalised. The HNL samples are shown for a HNL mass of 370 MeV. The distributions *do not* have any of the previously introduced selection applied to them.

In order to filter exiting tracks we determine for each HNL candidate the track whose end point is located farthest away from the centre of the detector. This is done separately for each dimension so that there is no correlation to an individual track (e.g. the farthest value from the centre of the detector in the  $x$  axis might be taken from the end point of the first track, while the farthest value from the centre in the  $y$  axis might be taken from the end point of the second track). The ensemble of the three end points thus defined can be used to determine the *smallest* possible box that contains both tracks. We then place a selection for each axis and require both the  $x$ ,  $y$  and  $z$  selections to be satisfied for each candidate. This is equivalent to requiring that the box thus defined is no closer than a certain value to the edges of the instrumented TPC. Figure 6.19 shows the effect on the distribution for different sample types. The dip structure at the centre of each axis is caused by the method used for the selection, which is choosing one of the two tracks which ends farthest away from the centre. To populate the centre region, both tracks would have to end very close to the centre

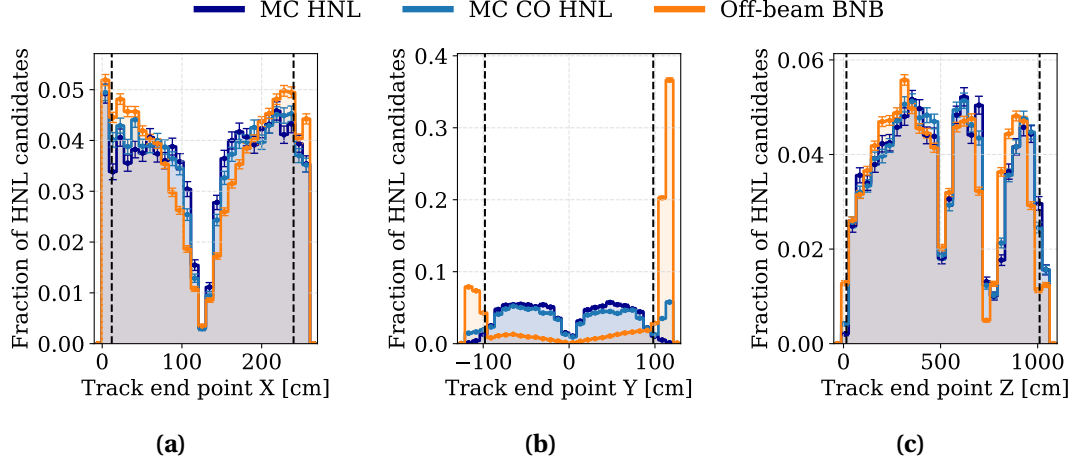
which is unlikely.

The selection improves the quality of the candidates by ensuring that both tracks are fully contained and, as it can be seen from [figure 6.19b](#), it improves the purity by removing  $\approx 85\%$  of the cosmic background, while removing only  $\approx 20\%$  of the signal. The large difference between the small amount of down-ward cosmics exiting the detector from the bottom panel ( $< -100$  cm) and the large amount of up-ward cosmics exiting from the top panel ( $> 100$  cm) is due to the *PandoraCosmic* pass (see [section 5.6](#)). The vast majority of cosmic background should be directed downward, however, the *PandoraCosmic* algorithm removes all the hits belonging to successfully reconstructed cosmics before passing them over to the *PandoraNu* reconstruction algorithm. Since *PandoraCosmics* favours downward going tracks, the remaining tracks surviving the pass will preferentially be directed upward.

In the  $z$  axis, the second dip at  $\approx 750$  cm is caused by a set of unresponsive wires. The lack of an increase at the end of the  $z$  axis for the signal HNL sample, which would be expected for forward-going tracks, exiting the detector through its downstream panel, is due to the effect of the fiducial volume selection which creates a buffer region for the HNL tracks. [Table 6.10](#) reports the effect of the selection on the efficiency and the number of events resulting after the selection.

<i>Track containment selection efficiency</i>			
<b>Sample</b>	<b>Efficiency (this selection)</b>	<b>Cumulative efficiency</b>	<b>Number of candidates</b>
MC HNL	0.811 ( $\pm 0.004$ )	0.458 ( $\pm 0.002$ )	23
MC CO HNL	0.708 ( $\pm 0.004$ )	0.327 ( $\pm 0.002$ )	18
<i>Off-beam</i> BNB	0.158 ( $\pm 0.004$ )	0.0332 ( $\pm 0.0008$ )	4,441
MC BNB	0.302 ( $\pm 0.010$ )	0.076 ( $\pm 0.001$ )	6,021
<i>On-beam</i> BNB	0.275 ( $\pm 0.006$ )	0.078 ( $\pm 0.001$ )	6,344
<i>On-beam</i> HNL	0.15 ( $\pm 0.01$ )	0.0272 ( $\pm 0.0008$ )	1,138
<i>Off-beam</i> HNL	0.15 ( $\pm 0.02$ )	0.0265 ( $\pm 0.0007$ )	1,096

**Table 6.10:** *Track containment* efficiencies for different sample types. Format for this table is identical to the format used for [table 6.5](#).



**Figure 6.19:**  $x$ ,  $y$  and  $z$  coordinates of the end point of the track whose 1-d projection on each axis ends farthest away from the centre of the detector, for each HNL candidates, for each axis. Coordinates for the three plots are not necessarily correlated to the same track. For each dimension, the value farthest away from the centre of the axis is selected. The black dashed lines correspond to the selections placed on the sample (candidates contained between the two dashed lines, *for each dimension*, are kept). Format used for this figure is identical to the format used for figure 6.12.

## 6.3 Kinematic Selections

A further series of selections is applied based on the candidate kinematics. While kinematic selections can be particularly effective in improving the signal-to-background ratio, they lack generalisability to all HNL masses values that can be probed by MicroBooNE. Different HNL mass values lead to different kinematics of the decay so that the optimal value of a selection can vary across different mass values (e.g. an optimal selection on momentum at 370 MeV might cut into signal for a 300 MeV HNL).

We reserve the use of kinematic variables for the training of a BDT. Since different BDTs are trained for each mass point, they can better exploit the correlation between kinematic features, such as between opening angle and forward momentum. The kinematic features correlations are mass-dependent and a mass-dependent tuning is needed in order to take advantage of the features for maximising the signal/background discrimination power. We make an exception for only two selections on kinematic variables: *opening angle* and *invariant mass*. The selections are placed in

regions where the impact on the signal can confidently considered to be minimal for all mass values while removing large fractions of background, thus removing regions of the phase-space where separation is trivial and allowing the BDT to focus the training on more challenging regions of phase-space. Although using the forward direction for selection is effective at high HNL masses (due to its high peak at small angles), its use is not particularly suited for lower HNL masses, where the much shorter length of the tracks reduce the resolution, and it is thus not used for the kinematic selections.

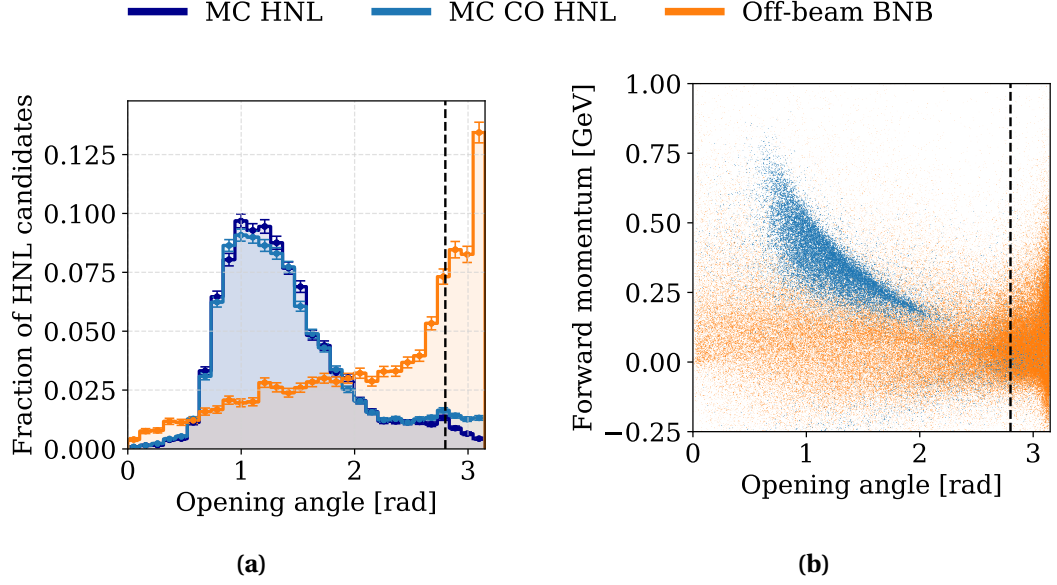
#### Opening Angle

In the rest frame of the HNL the two decay particles are emitted back-to-back, and the opening angle ( $\eta$ ) observed by the detector is described by the Lorentz boost in the  $z$  direction of a flat angle. This is responsible for the distribution of the opening angle between the two tracks to be peaked at a value, function of the mass of the HNL and its energy spectrum, as seen in figure 6.20a. For HNL decays, the opening angle  $\eta$  is strongly correlated with forward momentum, making both variables powerful in separating HNLs from CR muons, as shown in figure 6.20b. The CR sample does not see any particular correlation and the vast majority of the HNL candidates occur for  $\eta \approx \pi$ . This is due to the large contribution to the two-track topology of CR muons from *broken tracks*.

We define as *broken tracks* single muon tracks which the Pandora reconstruction algorithm has split in two, placing the neutrino vertex along the trajectory of the track and having the two reconstructed tracks emitted back-to-back. A large fraction of HNL candidates in the cosmic sample are produced by this occurrence and are associated with a large opening angle ( $\eta \approx \pi$ ). We place a relatively high selection on opening angle ( $\eta < 2.8$ ) for tracks which are almost back-to-back. The selection removes a large amount of background cosmics (by  $\approx 30\%$ , with a negligible  $\approx 2\%$  impact on HNL, at 370 MeV), shown in table 6.11.

#### Invariant Mass

We construct the invariant mass of the HNL candidate by using the *range* method (see section 6.1.2) to determine the momentum of both tracks and combine this



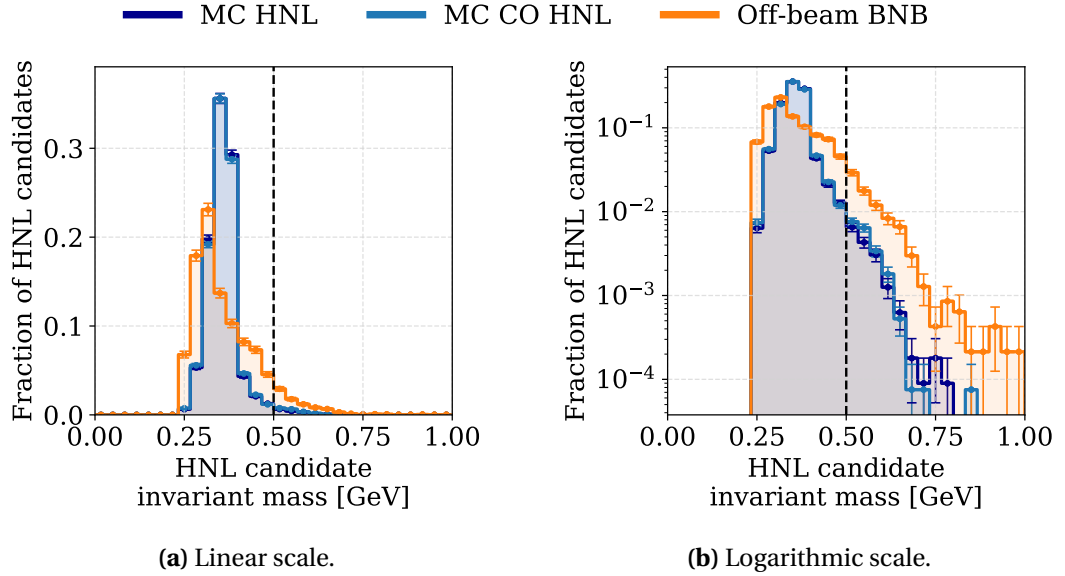
**Figure 6.20:** (a) 3-d opening angle ( $\eta$ ) between the two tracks associated with the HNL candidate. The peak in the signal sample is dependent on the mass and the energy spectrum of the HNL, while for the cosmics background sample, the vast majority of the candidates is located at  $\eta \approx \pi$ , corresponding to a flat angle (*broken tracks*). All the previous selections have been applied to the histograms (shown are the resulting histograms from the cumulative application of the selections). Histograms are not stacked and area normalised to illustrate the fraction of candidates passing the selection for each sample. (b) shows the correlation between opening angle and forward momentum, which is dependent on the mass of the HNL. For illustration purposes, in order to maintain enough statistics to make the features visible, (b) shows the distributions *before selections are applied*. The HNL samples are shown for a HNL mass of 370 MeV.

information to derive their energy and subsequently the invariant mass of the parent particle. The selection is applied for HNL candidates with an invariant mass value from *Hypothesis 1* (longest track muon, shortest track pion) smaller than 500 MeV, which is considerably larger than the highest value of HNL mass that can be probed by this analysis (388 MeV). The selection has a minimal effect on the signal ( $< 2\%$ ) while removing  $\approx 10\%$  of the background. Figure 6.21 shows the effects of the selection on distributions of signal and background. In table 6.12 are reported the efficiencies for this selection.



Opening angle selection efficiency			
Sample	Efficiency (this selection)	Cumulative efficiency	Number of candidates
MC HNL	0.976 ( $\pm 0.001$ )	0.447 ( $\pm 0.002$ )	22
MC CO HNL	0.954 ( $\pm 0.002$ )	0.312 ( $\pm 0.002$ )	17
Off-beam BNB	0.670 ( $\pm 0.007$ )	0.0223 ( $\pm 0.0007$ )	2,976
MC BNB	0.865 ( $\pm 0.008$ )	0.066 ( $\pm 0.001$ )	5,207
On-beam BNB	0.779 ( $\pm 0.006$ )	0.061 ( $\pm 0.001$ )	4,942
On-beam HNL	0.64 ( $\pm 0.02$ )	0.0173 ( $\pm 0.0006$ )	724
Off-beam HNL	0.65 ( $\pm 0.03$ )	0.0172 ( $\pm 0.0006$ )	712

**Table 6.11:** Opening angle efficiencies for different sample types. Format for this table is identical to the format used for [table 6.5](#).



**Figure 6.21:** Invariant mass of the HNL candidates, reconstructed with the momentum of the two tracks as determined with the *range* method. *Hypothesis 1* invariant mass is shown here (longest track is a muon, shortest track is a pion). The selection is represented by the dashed vertical line, events to the left of it are kept by the selection. Format used for this figure is identical to the format used for [figure 6.12](#).

## 6.4 Pre-selection Efficiency Summary

The purpose of the pre-selections is to improve the quality of the HNL candidates which are provided to the BDT in order to increase its performance, by filtering the

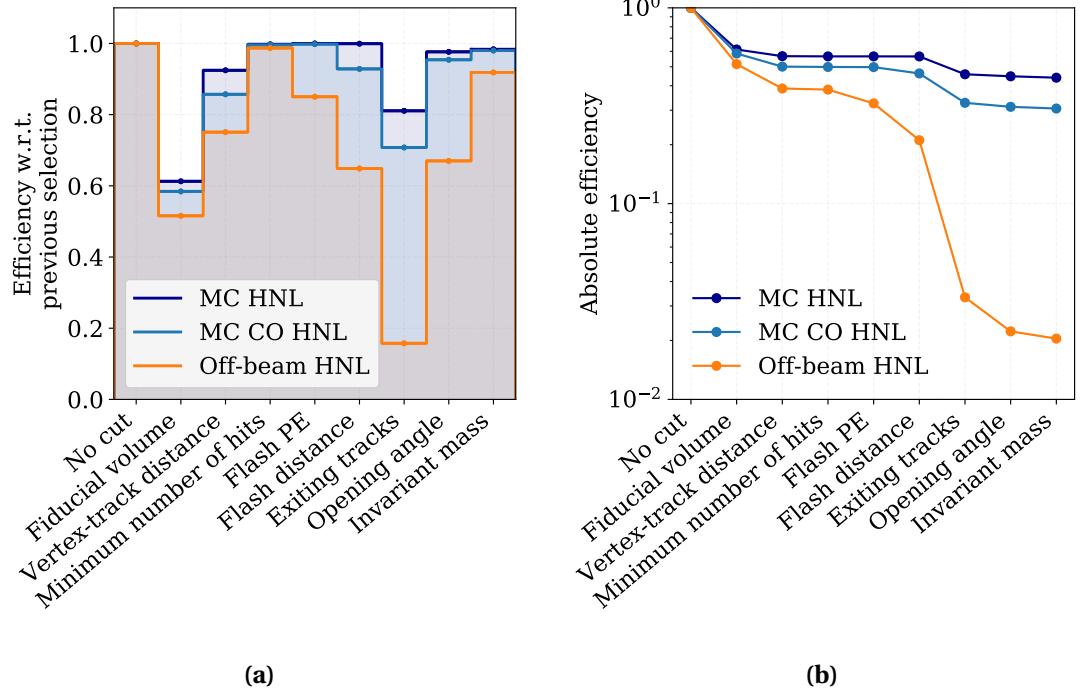
#### 6.4. PRE-SELECTION EFFICIENCY SUMMARY

<i>Invariant mass selection efficiency</i>			
<b>Sample</b>	<b>Efficiency (this selection)</b>	<b>Cumulative efficiency</b>	<b>Number of candidates</b>
MC HNL	0.984 ( $\pm 0.001$ )	0.440 ( $\pm 0.002$ )	22
MC CO HNL	0.980 ( $\pm 0.001$ )	0.306 ( $\pm 0.002$ )	17
<i>Off-beam</i> BNB	0.919 ( $\pm 0.004$ )	0.0204 ( $\pm 0.0007$ )	2,733
MC BNB	0.820 ( $\pm 0.010$ )	0.054 ( $\pm 0.001$ )	4,267
<i>On-beam</i> BNB	0.804 ( $\pm 0.006$ )	0.049 ( $\pm 0.001$ )	3,972
<i>On-beam</i> HNL	0.92 ( $\pm 0.01$ )	0.0160 ( $\pm 0.0006$ )	669
<i>Off-beam</i> HNL	0.92 ( $\pm 0.02$ )	0.0158 ( $\pm 0.0006$ )	653

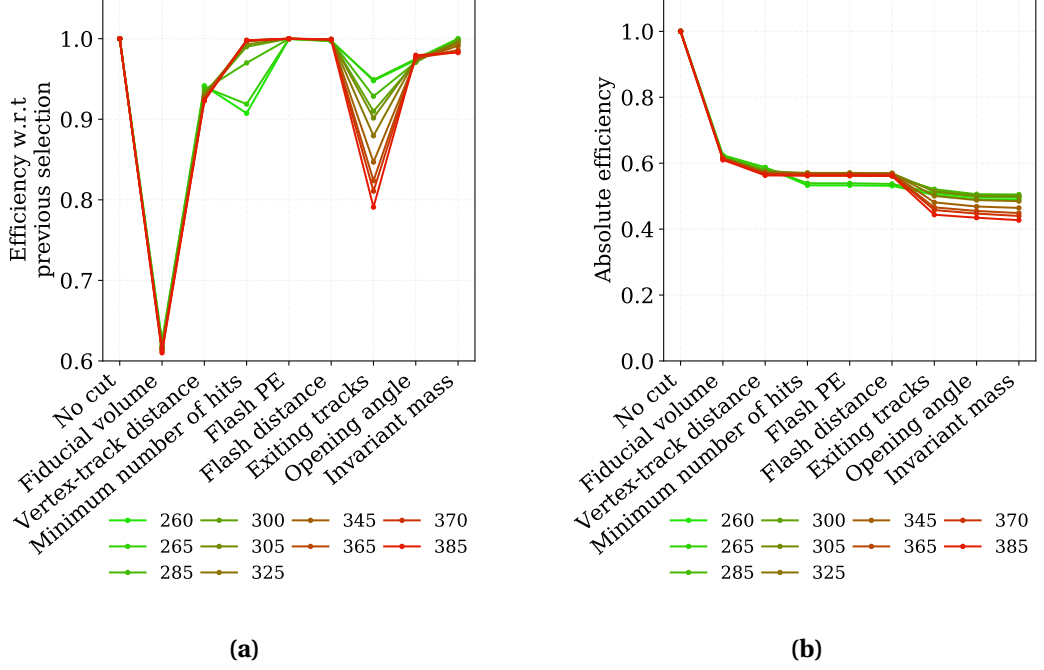
**Table 6.12:** *Invariant mass* efficiencies for different sample types. Format for this table is identical to the format used for [table 6.5](#).

training sample for candidates that have a well-defined signature. An associated benefit to the application of pre-selection is also an increase of the signal-to-background ratio since all selections have an higher efficiency for signal than background. [Figure 6.22a](#) shows the effect of the pre-selections on MC HNL, MC CO HNL and data *off-beam* HNL samples. The total effect of the pre-selection on the efficiency is a reduction of the number of signal events by a factor of  $\approx 2$  (0.44) while the cosmics background sample is reduced by a factor  $\approx 50$  (0.02).

Up until now, the pre-selection efficiencies have been shown for a single mass hypothesis. [Figure 6.23](#) shows the efficiencies of each pre-selection requirements for all the HNL mass values generated in this analysis. [Figure 6.23](#) illustrates how although the pre-selection requirements affect the HNLs differently according to their mass, the combined effects of all the pre-selection requirement is very similar, and at the end of the selection chain all the mass hypotheses have a similar efficiency located at  $\approx 50\%$ . [Tables 6.13](#) to [6.15](#) summarises the impact of the pre-selections on sample normalisation.



**Figure 6.22:** Effects of the pre-selections on different sample types, at the *candidate* level. (a) shows the relative efficiency of each selection, the fraction of events surviving each selection *after* all the previous selections have been applied. Consequently it shows the effects of an individual selection on the sample and the possible correlations with other selections (i.e. for each bin, the numerator is the number of candidates surviving the selection, while the denominator is the number of candidates surviving all the previous selections). (b) shows the cumulative effect of the selections on the *total* number of candidates (the denominator is, for all bins, the original number of candidates before the pre-selection).



**Figure 6.23:** Effects of the pre-selections requirements on different samples of MC HNLs (no CO), for increasing value of hypothesis mass (green is used for the lowest mass value, 260 MeV and red for the highest mass value, 385 MeV). (a) shows the cumulative effects of the pre-selection requirements on the samples. Although the different samples branch out at the first requirement, the combined effect of all requirements returns a similar final efficiency for all mass hypotheses at  $\approx 50\%$ . (b) shows the effect of each individual selection requirement after all the previous selections have been applied. It shows in particular how the *minimum number of hits* requirements is more taxing for the lower mass values (at lower masses, the two daughter particles have less available energy in the rest frame and their tracks are shorter, thus producing fewer hits along their trajectories). On the other side, the containment selection for *exiting tracks* is more demanding for higher masses, since their tracks are more likely to be longer and have a higher probability of reaching the edges of the detector despite the FV requirement.

#### 6.4. PRE-SELECTION EFFICIENCY SUMMARY

Number of candidates for MC HNL and CR muons		
Selection	Sample Type	
	MC HNL	Off-beam BNB
No cut	1 [49]	1 [133,675]
Fiducial volume	0.613 ( $\pm 0.004$ ) [30]	0.516 ( $\pm 0.002$ ) [68,955]
Vertex-track distance	0.567 ( $\pm 0.004$ ) [28]	0.387 ( $\pm 0.002$ ) [51,772]
Minimum number of hits	0.565 ( $\pm 0.004$ ) [28]	0.382 ( $\pm 0.002$ ) [51,105]
Flash PE	0.565 ( $\pm 0.004$ ) [28]	0.325 ( $\pm 0.002$ ) [43,457]
Flash distance	0.565 ( $\pm 0.004$ ) [28]	0.211 ( $\pm 0.002$ ) [28,190]
Track containment	0.458 ( $\pm 0.005$ ) [23]	0.033 ( $\pm 0.002$ ) [4,441]
Opening angle	0.447 ( $\pm 0.005$ ) [22]	0.022 ( $\pm 0.002$ ) [2,976]
Invariant mass	0.440 ( $\pm 0.005$ ) [22]	0.020 ( $\pm 0.002$ ) [2,733]

**Table 6.13:** Selection flow comparing the number of candidates surviving all the selections for MC HNL without CO and the off-beam BNB CR muon samples. These samples are used to train the BDT. The HNL values are generated for a HNL mass of 370 MeV and  $|U_{\mu 4}|^2 = 10^{-7}$ . Uncertainties are statistical only. Square brackets indicate the number of events remaining after the application of each selection and after renormalisation to  $1.97 \times 10^{20}$  POT.

Number of candidates for HNL data		
Selection	Sample Type	
	On-beam HNL	Off-beam HNL
No cut	1 [41,914]	1 [41,426]
Fiducial volume	0.520 ( $\pm 0.003$ ) [21,811]	0.519 ( $\pm 0.005$ ) [21,501]
Vertex-track distance	0.390 ( $\pm 0.004$ ) [16,339]	0.389 ( $\pm 0.005$ ) [16,126]
Minimum number of hits	0.385 ( $\pm 0.004$ ) [16,126]	0.385 ( $\pm 0.005$ ) [15,942]
Flash PE	0.249 ( $\pm 0.004$ ) [10,421]	0.251 ( $\pm 0.006$ ) [10,400]
Flash distance	0.180 ( $\pm 0.004$ ) [7,527]	0.181 ( $\pm 0.006$ ) [7,487]
Track containment	0.027 ( $\pm 0.005$ ) [1,138]	0.026 ( $\pm 0.007$ ) [1,096]
Opening angle	0.017 ( $\pm 0.005$ ) [724]	0.017 ( $\pm 0.007$ ) [712]
Invariant mass	0.016 ( $\pm 0.005$ ) [669]	0.016 ( $\pm 0.007$ ) [653]

**Table 6.14:** Selection flow comparing the number of events surviving all the selections for MC CO HNL and the BNB-HNL (data) and EXT-HNL (background) samples. These samples are then provided to the BDT for further discrimination. The HNL values are generated for a HNL mass of 370 MeV and  $|U_{\mu 4}|^2 = 10^{-7}$ . Uncertainties are statistical only. Square brackets indicate the number of events remaining after the application of each selection and after renormalisation to  $1.97 \times 10^{20}$  POT.

Number of candidates for BNB data		
Selection	Sample Type	
	MC BNB	<i>On-beam BNB</i>
No cut	1 [79,365]	1 [81,112]
Fiducial volume	0.534 ( $\pm 0.004$ ) [42,414]	0.531 ( $\pm 0.002$ ) [43,078]
Vertex-track distance	0.409 ( $\pm 0.004$ ) [32,471]	0.396 ( $\pm 0.003$ ) [32,120]
Minimum number of hits	0.406 ( $\pm 0.004$ ) [32,228]	0.394 ( $\pm 0.003$ ) [31,939]
Flash PE	0.334 ( $\pm 0.005$ ) [26,484]	0.379 ( $\pm 0.003$ ) [30,713]
Flash distance	0.252 ( $\pm 0.005$ ) [19,962]	0.285 ( $\pm 0.003$ ) [23,089]
Track containment	0.076 ( $\pm 0.006$ ) [6,021]	0.078 ( $\pm 0.003$ ) [6,344]
Opening angle	0.066 ( $\pm 0.006$ ) [5,207]	0.061 ( $\pm 0.003$ ) [4,942]
Invariant mass	0.054 ( $\pm 0.006$ ) [4,267]	0.049 ( $\pm 0.003$ ) [3,972]

**Table 6.15:** Selection flow comparing the number of candidates surviving all the selections for the control sample of MC generated BNB interactions (with simulated CORSIKA CR muons and dirt interactions) and data BNB interaction events (from which the *off-beam BNB* CR muons sample is subtracted). Uncertainties are statistical only. Square brackets indicate the number of events remaining after the application of each selection and after renormalisation to  $1.97 \times 10^{20}$  POT.

# Boosted Decision Trees

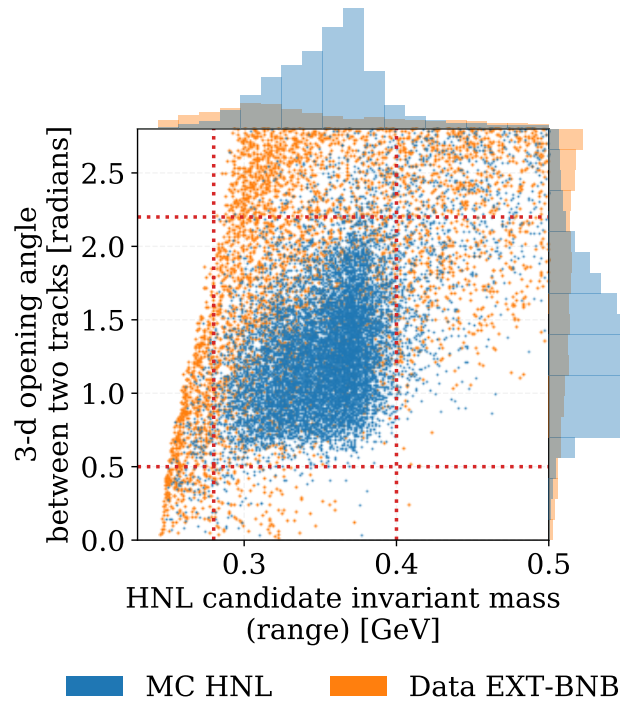
The application of *rectangular* selections has limited performance. A single selection may not offer an acceptable signal acceptance or background rejection level despite the fact that in higher-dimensionality variable space the two datasets are clearly separable. [Figure 7.1](#) shows an example of this, using correlated variables from the signal and background samples used for the BDT training. As an example, a simple rectangular selection on invariant mass and opening angle (shown in [figure 7.1](#)), which keeps a signal efficiency of 85% can only reduce background to 40% and provides an improvement of signal-to-background ratio of 2. For comparison, BDT performance, which is further discussed in [section 7.2](#), can increase the signal-to-background ratio by a factor of  $\approx 10$ , depending on HNL mass.

## 7.1 BDT Definition

An extension of the rectangular selection approach into a multivariate technique is the application of *decision trees* [\[77\]](#). Decision trees are recursive applications of rectangular selections which continue to analyse candidates that fail a particular selection. The basic operation of a (two-classes) decision tree algorithm proceeds as follows:

- Sort all events according to each variable.
- For each variable, find the splitting value that gives the best separation between the two classes.

- Select the variable that provides the best separation. Split the sample in two nodes, one for events that satisfy the criterion, one for the ones that fail it.
- For each node, repeats the process from the first step, until a stopping condition is reached.



**Figure 7.1:** HNL candidate invariant mass is shown on the  $x$  axis, calculated using the *range* momentum of the two tracks in *Hypothesis 1* (longest track muon, shortest track pion). 3-d opening angle between the two tracks is shown on the  $y$  axis. The two samples shown, pure MC HNL in blue and CR muons from EXT-BNB (*off-beam BNB*) in orange are used to train the BDT. The figures shows how the correlation between the two variables allows for high discrimination between the two samples, although the application of rectangular selections (shown in dotted red lines) cannot provide both optimal efficiency and rejection.

The terminal nodes of the tree are called *leaves*. A *signal leaf* contains mostly signal events, while a *background leaf* contains mostly background events. The stopping condition, which turns a final node into a leaf, depends on multiple conditions:



- A minimum number of samples can be required at each node, for statistical significance.
- The events have been perfectly classified (e.g., the purity of the leaf is either 1 or 0).
- The purity cannot be further improved by any choice of splitting value.
- The tree depth (number of iterative nodes) has reached the maximum allowed value.

Each leaf is associated with a score which represents the probability of an event reaching that particular leaf to be signal or background. More information on the application of decision trees in high energy physics can be found in Ref. [77]. Recent applications of decision trees make also use of boosting algorithms, producing *Boosted Decision Trees* (BDT). The basic operation of a boosting algorithm (which can be applied to any classifier) can be described as following:

- Train a classifier  $T_1$  on a sample of  $N$  events.
- Train a classifier  $T_2$  on a sample of  $N$  events, half of which were misclassified by  $T_1$ .
- Build  $T_3$  on events where  $T_1$  and  $T_2$  disagree.

It is thus an ensemble technique, where rather than producing an extremely good discriminant, a series of weak classifiers is produced. These weak classifiers generally perform poorly on their own, as they are characterised by high bias, so that they have low discrimination power (bias is caused by *under-fitting* and is an indicator of poor performance) but they are also characterised by low variance, so that they generalise well to new data that is not part of the training set (variance is represented by the sensitivity to small fluctuations in the training set, it is caused by *over-fitting*). These weak classifier are generally characterised by a very low maximum tree depth.

The goal of boosting is to combine such weak classifiers, in order to reduce bias while keeping the variance low, into a better classifier, with better performance [77]. For boosted trees, the final score of an event is given by the weighted sum of the scores

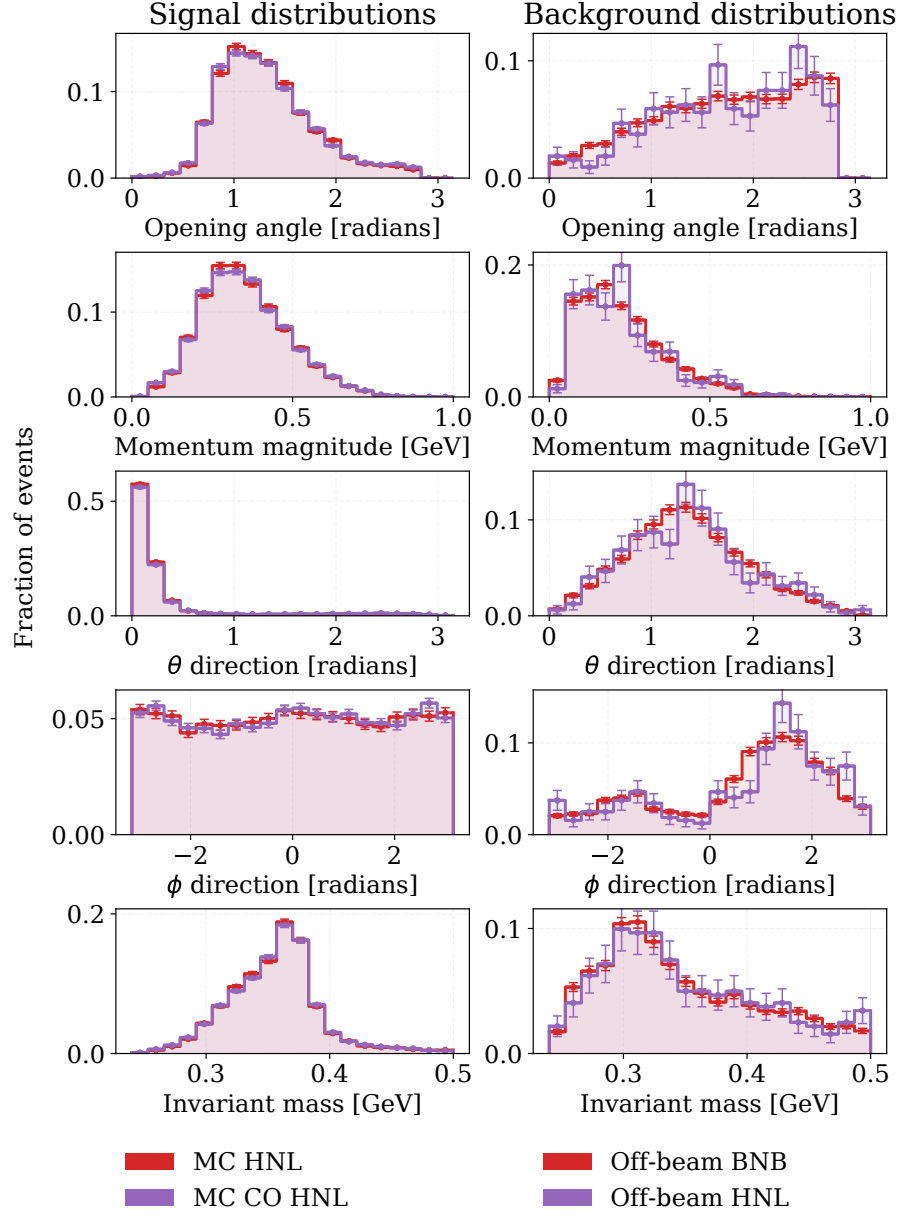
of the individual trees. *Gradient boosting* is the particular application of boosting where boosting is treated as an optimisation algorithm on a suitable cost function, and is thus similar to a gradient descent algorithm used to find the minimum of a function.

## 7.2 Application to HNL Samples

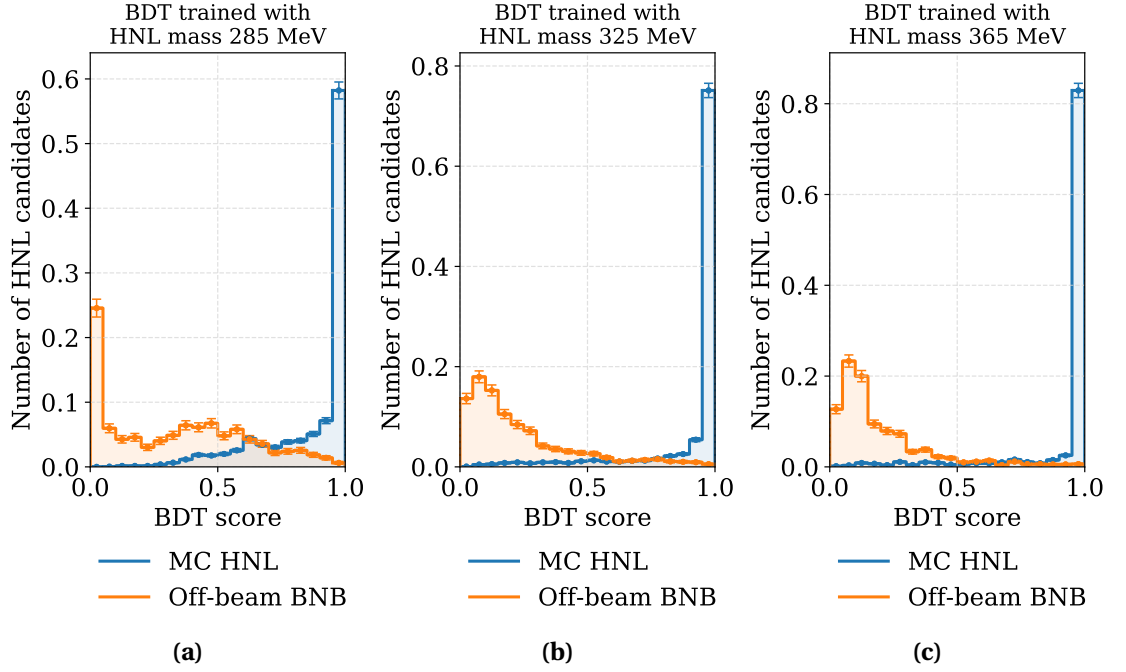
In this analysis, we use *XGBoost* [78] to train the BDT on the MC HNL and *off-beam* BNB samples (see section 5.7). As explained in section 5.7, while the train and test samples differ in definition, after quality selections the distributions used are sufficiently similar, due to the quality selection removing the events which are differently characterised in the two samples (e.g., the flash PE selection removing the difference due to different software triggers), and the increase in statistics results in a better efficiency. Figure 7.2 shows the variables used for the BDT training after the pre-selections for both the training samples and the testing samples. The variables used to train the BDT are the following:

- 3-d opening angle between the two tracks ( $\eta$ ).
- Reconstructed HNL total momentum magnitude.
- Reconstructed HNL  $\theta$  angle.
- Reconstructed HNL  $\phi$  angle.
- Reconstructed HNL invariant mass.

All quantities, with the exception of the opening angle which is purely geometrical, are derived using the momenta of the two tracks as derived by the range method, using *Hypothesis 1* (longest track muon, shortest track pion). The XGBoost algorithm uses 100 estimators with a maximum tree depth of 3. We train a different BDT for each mass point using  $\approx 5000$  MC HNL events for each mass point (with no CO) for the signal and  $\approx 3000$  off-beam BNB CR muons events. We reserve  $\approx 2000$  signal events and  $\approx 1300$  background events to estimate the performance of the trees, which is shown in figure 7.3.



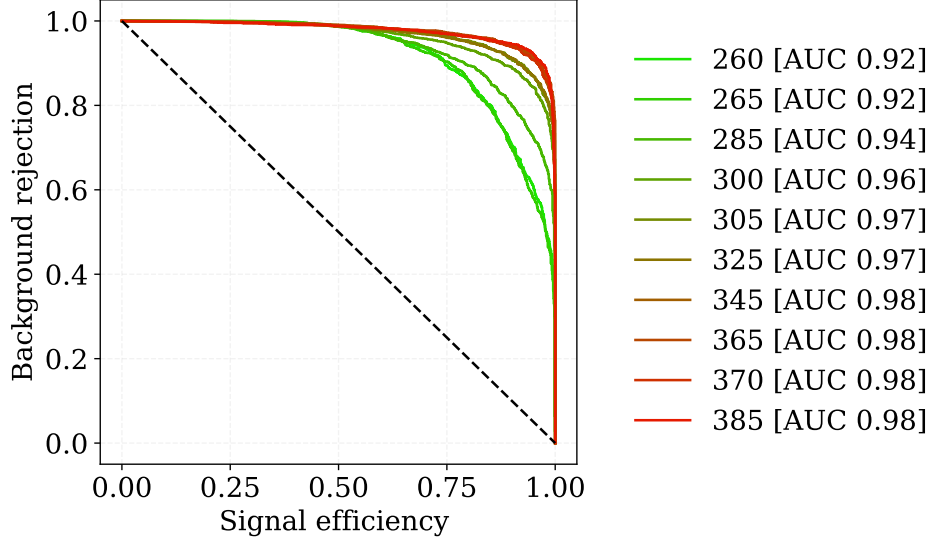
**Figure 7.2:** Variables used to train the BDT for both the training samples (in red) and the testing samples (in purple). The left column shows the distributions for the signal samples, while the right column shows the distributions for the background sample. The y axis is fraction of candidates from the full dataset in each bin, since the distributions are area normalised to emphasise the differences in shape. The figure shows that aside from statistical fluctuations the samples from training and testing have the same distributions.



**Figure 7.3:** BDT score (probability for events to be signal) from a MC HNL sample (285 MeV left, 325 MeV centre, 365 MeV right) and an off-beam BNB CR cosmics sample. The BDTs perform better for larger mass values, as the kinematic features of the decay become more distinct with respect to CR muons.

A common figure of merit for estimating the performance of BDT is the Receiving Operator Characteristic curve (ROC), which we show here as the true positive rate (the number of correctly identified signal events) against the true negative rate (the number of correctly identified background events), for different thresholds on the BDT score shown in figure 7.3. The quality of a BDT attributed to a ROC curve can be quantified by calculating the Area Under Curve (AUC) integral. This corresponds to the probability that a randomly chosen signal event would get a higher BDT score than a randomly chosen background event. The farthest away from the diagonal, the better the performance of the BDT, as this would allow more stringent selections on the BDT score that can maximise efficiency (number of selected signal events over total number of signal events) and purity (number of selected signal events over total number of selected events). The ROC curves and AUC values are shown in figure 7.4. The figure shows how, for increasing values of the mass of the HNL, the performance of the BDTs increases. This is due to the decay of more massive HNLs producing more distinctive kinematic features, as the energy of the particles increases and consequently their

length.



**Figure 7.4:** ROC curves for the BDTs trained for each mass point (from green to red, in order of increasing HNL mass values). The AUC score shows that for increasing mass values the performance of the BDT increases.

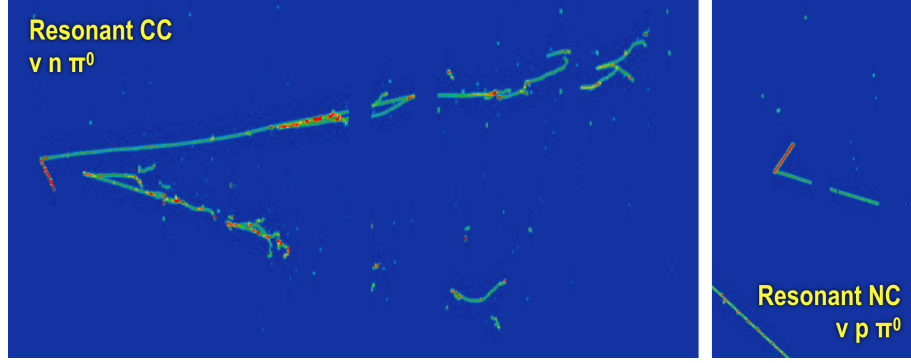
The BDTs show very good performances with average AUC values around 0.95 for all masses. With the data samples available for this analysis, the limiting factor is the amount of events available in the background sample to model the background distribution in the signal-like region. The BDTs are successively applied to the test samples (MC CO HNL, on-beam HNL and off-beam HNL) in order to explore the existence of an excess in the on-beam HNL sample over the off-beam HNL sample in the signal-like region of the BDT score distributions. These distributions are provided to a confidence limit evaluator program (COLLIE, discussed in section 9) in order to extract a limit in the HNL parameter space.

# Control Sample and Systematic Uncertainties

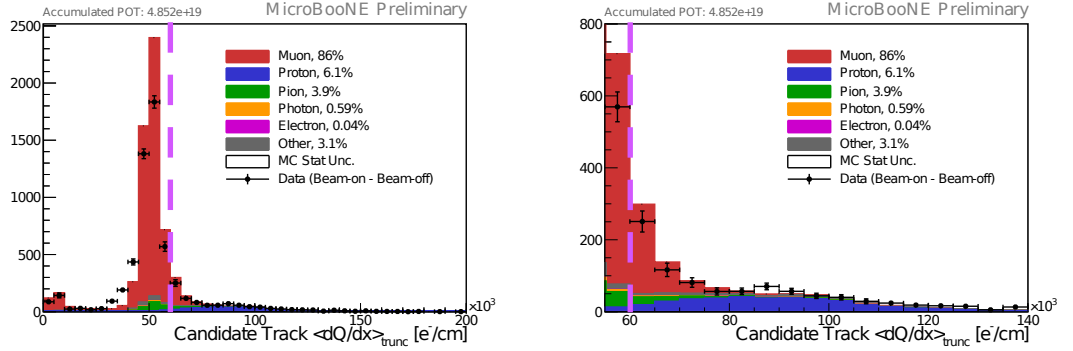
In this section the methods used for the determination of systematic uncertainties associated with the analysis are discussed. The final [BDT](#) score distribution for the expected signal ([HNL](#) decays) sample provided to the limit setting algorithm depends on a series of assumptions regarding the [HNL](#) production mechanism, its propagation to the detector and the response of the detector to the signals produced in the [TPC](#). By estimating the range of uncertainties associated with each underlying models, the uncertainties can be propagated to the final [BDT](#) distribution and provide a total systematic uncertainty associated with the limit bound. In [section 8.1](#) we show the results from the application of the [BDTs](#) to the control sample for validation purposes. The approaches used for the propagation of the uncertainties are different across different uncertainty source, we discuss the two main methods in [section 8.2](#). [Sections 8.3](#) to [8.5](#) describe the source of systematic uncertainties analysed in this work. A summary of the systematic uncertainty study is presented in [section 8.6](#).

## 8.1 Control Sample

The *on-beam* [BNB](#) data and [MC BNB](#) samples provide a mean for demonstrating, in data, that the selection and [BDT](#) chains are able to select events with [HNL](#)-like topologies. The two [MC](#)-generated samples, [MC BNB](#) and [MC HNL](#), differ under few aspects. Aside from the obvious different content of the events (neutrino interaction

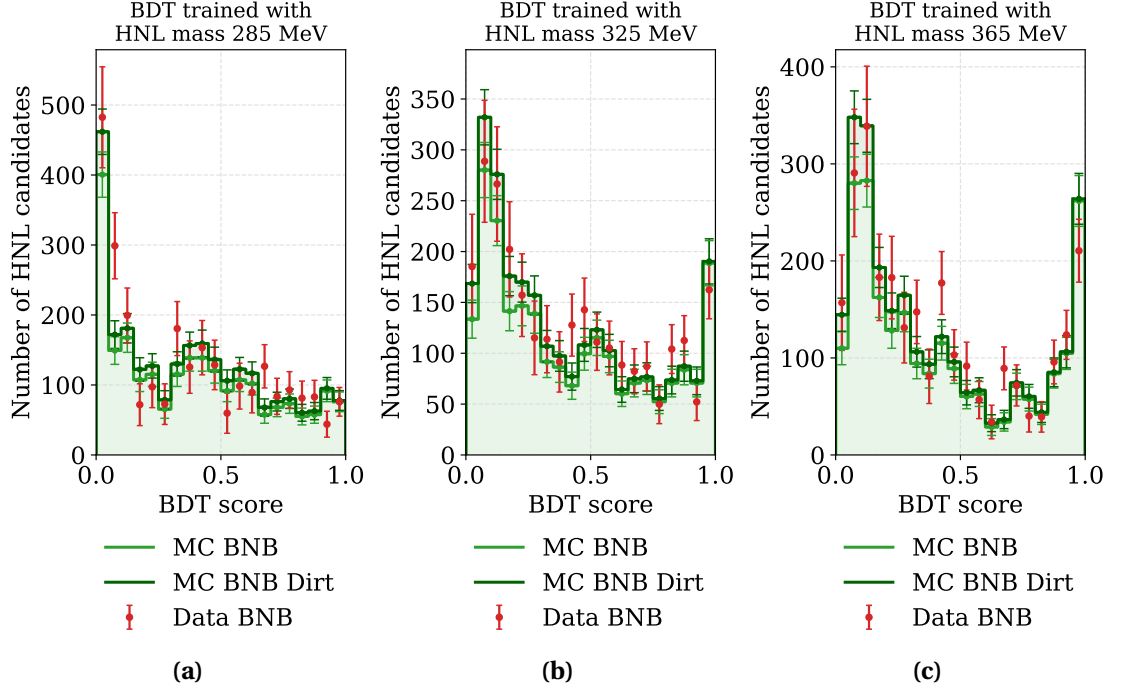


**Figure 8.1:** Event displays from a MC BNB sample where the final products of the  $\nu$  interaction produce two-tracks which are reconstructed as a HNL candidate with high HNL-like BDT score ( $\text{BDT} < 0.1$ ). The left figure shows a resonant CC interaction where the  $\pi^0$  (the invisible particle decaying to two electron shows, few cm away from the vertex) has not been reconstructed and a proton (short red track) instead has been associated with the muon (long track), creating a two-track object. The right figure shows a resonant NC interaction, where the proton (short red track) forms a two-track object with the muon (long track).



**Figure 8.2:** Median  $dQ/dx$  distributions for tracks contained in the BNB sample. The figure on the right is an enlargement of the figure on the left, to illustrate the region prevalently populated by protons (blue). A selection on  $dQ/dx < 60 e^-$  (purple dashed line) is applied for the BNB sample in order to demonstrate a better MC/data agreement when using exclusively muons (red). Figures are from Ref. [79].

and HNL decays), the BNB sample is produced with simulated CORSIKA CR muons and simulated underlying noise, it is produced with previous iterations of the MC generation and reconstruction algorithms (see table 5.1) and it is compared to data from a previous run period (Run 1, whereas Run 3 is used for the HNL studies). Regardless



**Figure 8.3:** BDT score (probability for events to be signal) for a simulated BNB  $\nu$  interaction sample and an on-beam data sample (with CR muons contribution subtracted using an off-beam sample) for 285 MeV (left), 325 MeV (centre), 365 MeV (right).

of these differences, it is able to provide events with distinctive HNL-like features. Figure 8.1 shows two event displays from a MC BNB sample, where the final products of the  $\nu$  interaction produce two-tracks which are reconstructed as a HNL candidate with high HNL-like BDT score.

A common feature of many HNL candidates from the BNB sample is the presence of a proton in the final state, which is recognisable by the higher energy loss. Protons from the HNL candidates, which are abundant in the BNB control sample, are not present in neither the MC HNL signal nor the data background sample used in this analysis, and as a consequence, any inefficiency related to their simulation or reconstruction does not apply to the HNL search. Furthermore the models for the production of protons as a result of neutrino-nucleus scattering are not well constrained due to the limited availability of measurements [80]. Particularly, proton production models in MicroBooNE show up to 20% variation depending on the GENIE model used [81]. We apply a selection on median  $dQ/dx < 60 e^-/\text{cm}$  (see figure 8.2) to remove protons. The  $dQ/dx$  is applied exclusively on the control sample, and not on the analysis samples



(HNL and CR muons), since no protons are expected in neither of the analysis samples. Once protons are removed from the BNB sample, the data and MC samples are found in good agreement. The results of the control sample thus validate the performance of the BDT with respect to data, and provide positive evidence that if HNL events were present in data, the BDT would be able to correctly identify and separate them.

## 8.2 Uncertainty Determination

Two different methods are available in MicroBooNE for the propagation of the uncertainties. “Multi-sim” methods creates pseudo-universes with random variations of the base parameters, whereas the less computationally-intensive “uni-sim” variations use single-variation simulations to estimate the shift from a central value with no variation.

### 8.2.1 Uni-sim Approach

The set of models used for the generation of MC simulated events may be described by a series of parameters  $i$  with associated values  $p$ , where each  $p^i$  can describe a single value, such as a diffusion coefficient, or a more complex instance, such as the particular equation used to model a physics effect. The standard set of parameters used for MC simulation is denoted with  $p_{CV}$ , where CV indicates the central value of reference.

The uni-sim approach to uncertainty is designed to fully characterise a source of uncertainty by varying a single parameter  $p^i$  during the MC generation stage, where  $p^i$  is responsible for a particular source of uncertainty, and compare the generated MC sample with the baseline MC sample, defined here as central value (CV) sample, which is generated with the nominal values  $p_{CV}$  for all the parameters.

In the case in which a parameter  $i$  has uncertainties associated to it, the variation samples can be produced by performing the simulation in the  $p_{CV}^i + 1\sigma^i$  and  $p_{CV}^i - 1\sigma^i$  cases, while keeping all the other parameters to their default CV, and comparing the produced samples with the CV samples. For models which do not have an associated

$\sigma$  uncertainty interval, such as in the case of SCE, the model may be switched off entirely. In this case, the CV sample is generated modelling the physics effect, while the variation sample does not include the physics effect at the simulation stage. The uni-sim “on/off” procedure allows one to determine the magnitude of a physics effect where the associated underlying uncertainties are not known, although analysing each effect independently neglects correlation and might lead to an overestimation of the uncertainties. The uni-sim approach is particularly suited for uncertainties intervening at computationally intensive stages of the Monte Carlo generation procedure which cannot be factorised (e.g., detector effects).

In order to minimise the impact of statistical variations, the same underlying MC truth events are used for all the variations. E.g., detector effects can be estimated by varying the parameters and running the detector simulation on the same collection of events, with the same MC “truth” particles, so that while the response of the detector might vary as a result of a parameter changed from its CV, the underlying event is exactly the same (e.g., a muon travelling for 50 cm before producing to a Michel electron). This also allows for the comparison of the effect of the variation on an event-by-event basis for validation.

### 8.2.2 Multi-sim Approach

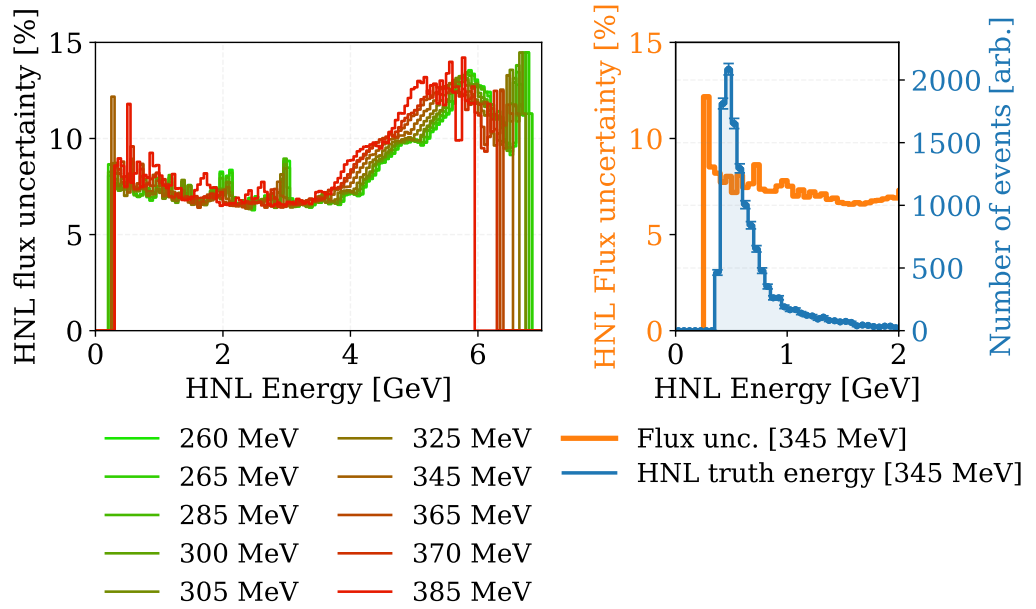
The multi-sim approach is an extension to the uni-sim approach. Instead of generating a single sample where a parameter  $i$  is varied by  $\pm 1\sigma$  with respect to its CV,  $N$  samples are generated, where for each sample a random value of  $p^i$  is sampled from the Gaussian distribution with centre  $p_{CV}^i$  and width  $\sigma^i$ . Multi-sim methods provide a more realistic depiction of the uncertainty associated with a certain effect. Furthermore, if multiple parameters are varied (sampled) simultaneously, correlations between these parameters can be taken into account.

While the multi-sim method represents a more comprehensive approach to uncertainty estimation, the generation of  $N$  samples makes it computationally more intensive and unsuitable for later stages of MC simulation. In MicroBooNE it is used for the estimation of flux uncertainties, which occur at the early stages of the MC generation procedure and can be factorised with respect to the remaining stages of

simulation. I.e., the output of the flux simulation is a flux spectrum which can be provided to the simulation software.

The multi-sim method returns a series of weights which can be used to weight individual events of the down-stream distributions (e.g., the flux uncertainties return energy-dependent weights).

### 8.3 Flux and POT Count Systematics



**Figure 8.4:** Systematic uncertainties (left) on the HNL flux. From green to red are shown uncertainties for HNL mass of increasing value. The flux systematic is weakly depending on the HNL mass, with the largest differences at high energies ( $> 4$  GeV). The figure on the right shows the uncertainty for a HNL with a mass of 345 MeV overlaid on top of the “truth” energy spectrum of a MC HNL sample, after all the selections have been applied. The figure shows how only the first few bins of the flux uncertainties are relevant to this analysis.

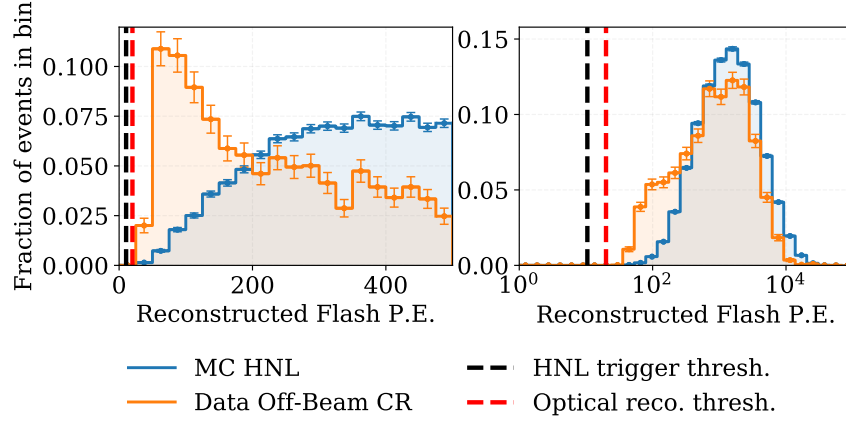
Uncertainties on the HNL flux can be calculated with the same methods used for the BNB flux. This is done with the multi-sim method, by generating 1000 simulations of the HNL flux where all the parameters responsible for the flux simulation are

simultaneously varied. The parameters are mainly divided in “hadron production systematics”, in which parameters are varied by smearing the production cross-section for a particular meson ( $\pi$  production is neglected since only  $K$  mesons contribute to the production channel studied in this analysis) and “beamline systematics”, related to the beamline instrumentation and in which the parameters reflect the uncertainties on the current circulating in the focusing horn, the depth by which the current penetrates in the horn conductor and the pion and nucleon cross sections on aluminium and beryllium [82, 83].

The flux systematic uncertainties are calculated by generating 1000 multi-sim variation universes, which are used to produce 1000 histograms of the HNL flux. For each energy bin of the flux, the distribution of all the values from the 1000 universes is calculated and the mean value of the bin from the central value flux simulation is used as central value. The spread with respect to the mean results Gaussian. The  $1\sigma$  value associated with each bin is used as the uncertainty for that bin. The uncertainty on each bin can then be used to weigh each event of the MC HNL sample. Figure 8.4 shows the flux systematic uncertainties determined with this method. For all masses the uncertainty on the flux is  $\approx 8\%$ . A fixed 2% uncertainty is assigned to the POT counting performed with the beam toroid [83].

## 8.4 Trigger Systematics

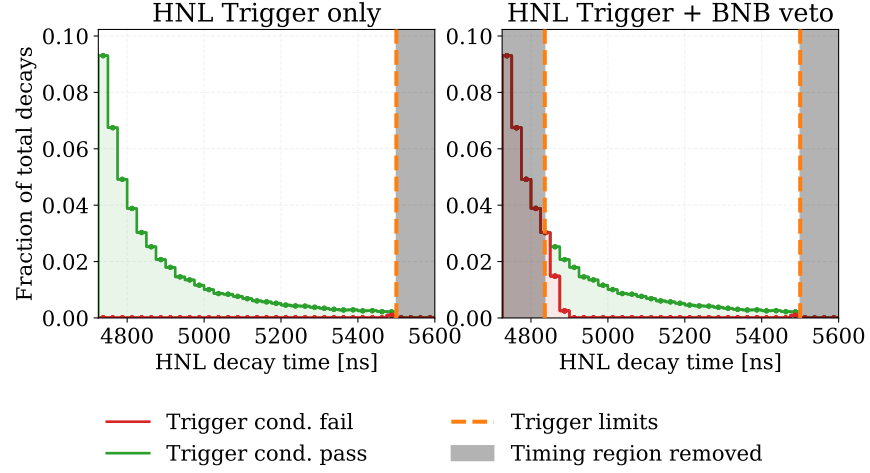
The systematic uncertainties on the efficiency of the trigger can be divided in uncertainties related to the light collection and timing resolution. Studies on the HNL trigger light collection efficiency have been discussed in section 4.2, where an efficiency close to 100% has been observed for all masses. Uncertainties in the light collection are superseded by the optical requirements of the pre-selection (see section 6.2.4), which imposes the requirement that the candidate must have a reconstructed flash assigned to it. Since reconstructed flashes are constructed only if more than 20 PE have been observed in an event, this is equivalent to applying a 20 PE threshold on the event, which supersedes the 10.5 PE threshold imposed by the HNL trigger. Both threshold values are placed at the very tail of the HNL light production distribution, which is shown in figure 8.5, such that variations around that value would have a negligible



**Figure 8.5:** PE of the reconstructed flash assigned to the HNL candidate for MC HNL sample (with a mass of 345 MeV, in blue) and data off-beam CR sample (collected by the BNB software trigger, in orange). The two distributions are shown before pre-selection, (a) shows the distributions near the PE threshold for the HNL software trigger (10.5 PE) and the threshold for a reconstructed optical flash to be formed (20 PE). (b) shows the same distributions on a  $x$ -log scale. The two figures show that both thresholds are placed at the very tail of the MC HNL distributions and that variations around that value are not bound to affect the efficiency.

impact on the efficiency for candidates surviving the selection.

The major source of trigger systematic uncertainty is related to the timing resolution of the PMTs and the effects of vetoing on the BNB timing window. The efficiency across the whole HNL software trigger is  $\approx 100\%$ , however due to the BNB software trigger veto, only the last portion (1/5) is used. Due to the limited resolution of the PMT instrumentation ( $\approx 15.6$  ns), events which are produced in the allowed timing regions may be accidentally vetoed. This effect is shown in figure 8.7, where the effects of the BNB veto cause inefficiencies at the edges of the timing range which defines the “HNL veto BNB” timing window. Efficiencies due to the limited timing resolution are estimated by shifting the whole distributions by  $\pm 15.6$  ns, the timing resolution of the PMTs. While this assumes that *all* the events would be shifted in a correlated manner, it adequately covers the uncertainty range assumed for this source of systematics. The  $\pm 1\sigma$  trigger efficiency boundaries are shown in figure 8.7. The uncertainty thus derived is estimated to be  $\approx 5$ -10%.

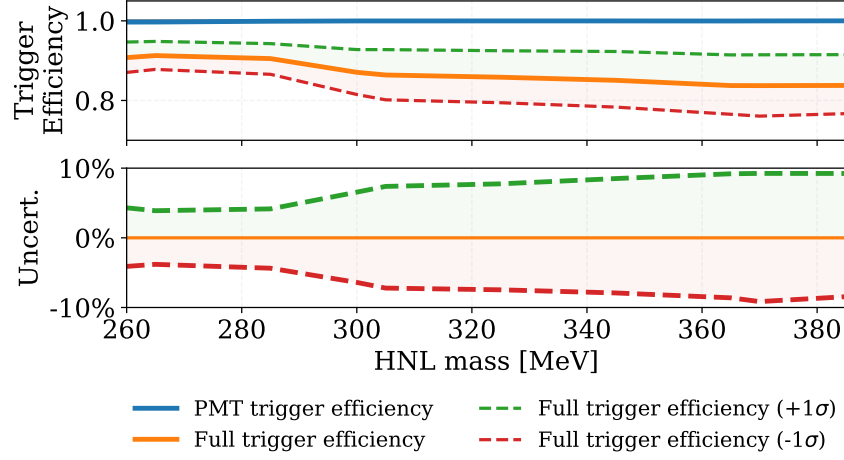


**Figure 8.6:** Pass rate for a HNL sample (mass 345 MeV). In (a) is shown the pass rate with the requirement that only the HNL software trigger conditions are satisfied. The figure in (b) shows the same distribution, while requiring both the HNL software trigger conditions to be satisfied and the BNB software trigger to be vetoed. As can be seen from the plot, a fraction of vetoed events “spill” into the timing range which defines the “HNL veto BNB” timing window, lowering the total efficiency.

## 8.5 Detector Systematics

The MC simulation of the detector response includes a large number of physics effects, described in more details in [section 3.3](#), that can affect both the ionisation electrons and scintillation photons during propagation (e.g., SCE, diffusion) and the signal measured by the detector instrumentation (e.g., DIC, noise). For fully MC generated samples, each of the parameters governing the models representing the physical effects need to be varied in order to estimate the effect of the model/parameter uncertainty on the HNL sample.

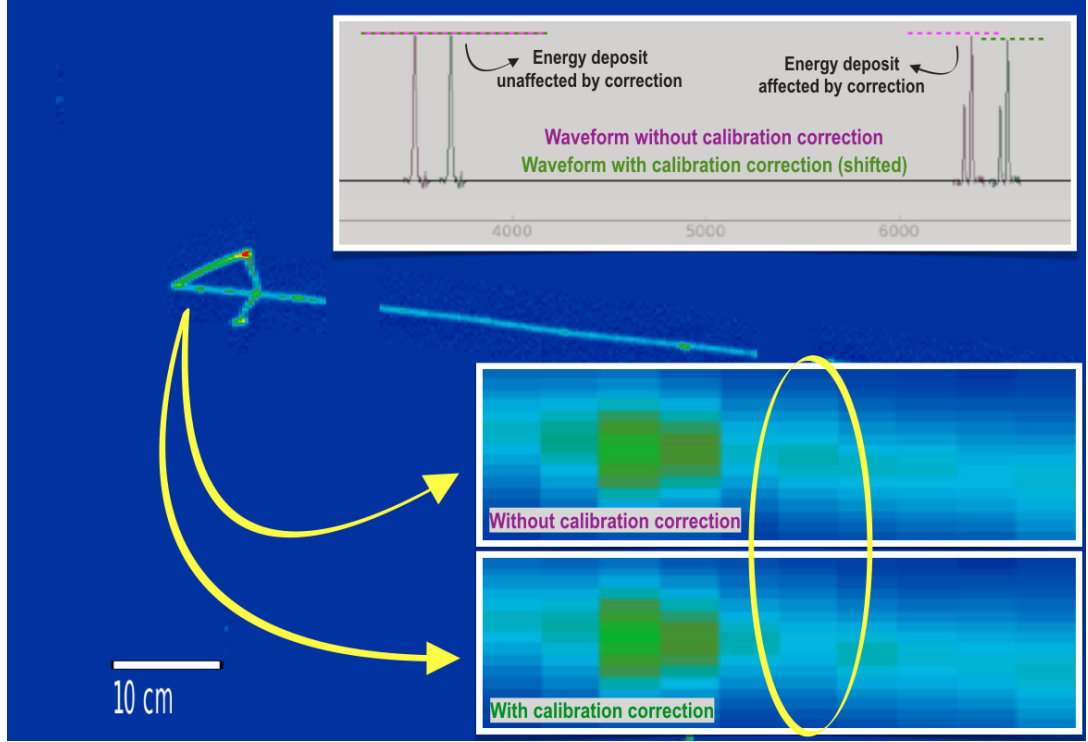
One of the advantages of the MC CO procedure (described in [section 5.2.3](#)) is that the MC event is overlaid on top of the raw readout of the off-beam (CR-only) data event, thus inheriting the original data event noise. This method avoids the use of noise models for noise simulation. Furthermore, no dead channels need to be simulated, since the MC CO procedure does not overlay the MC signal on inactive channels. In addition, each data event is associated with a database entry describing the  $x$ ,



**Figure 8.7:** Trigger efficiency (top). In blue is shown the trigger efficiency and in orange the total efficiency, taking into account the edge effects from the BNB veto. The green and red lines are the  $\pm 1\sigma$  variations in total trigger efficiency obtained by shifting the timing distribution by  $\pm 15.6 ns$  (PMT timing resolution). The bottom figure shows the deviation of the  $\pm 1\sigma$  variations with respect to the total trigger efficiency, which provide the trigger systematic uncertainties.

$y$ ,  $z$  and  $t$ -dependent correction constant (discussed in [section 5.5.2](#)) determined with the calibration procedures which encapsulate the result of many detector effects (e.g., SCE, misconfigured and cross-connected channels, longitudinal and transverse diffusion, attenuation) on the charge collected by the wires in a correlated fashion.

In order to have the MC event match the underlying data event, the simulated detector effects are removed from the simulated signal and the MC simulated collected charge is *degraded* by the inverse of the calibration correction. The calibration correction takes care of normalising the charge deposition in data at coordinates  $x$ ,  $y$ ,  $z$  and  $t$  in such a way that the detector response is uniform across all the TPC. Conversely, the inverse of the calibration correction  $C(x, y, z, t)$  reflects the result of all the physics effects on the energy deposition at coordinates  $x$ ,  $y$ ,  $z$  and  $t$  in an *average* fashion ( $y$  and  $z$  corrections are determined over periods encompassing several months,  $x$  and  $t$  are determined daily). By *degrading* the MC simulated event with the calibration corrections, the *averaged* physical effects, determined from data on the day the underlying data event was taken, are applied on the MC simulated energy



**Figure 8.8:** Simulated MC HNL decay. The yellow arrows point to two zoomed version of the first few centimetres of the muon track. In the zoomed picture on top (purple writing) the standard detector simulation has been applied, while in the zoomed picture on the bottom (green writing) the signal has been degraded by the calibration corrections. The yellow circle highlights a hit in which the two procedure differs in the amount of collected charge obtained from the same energy deposition. On top is shown a waveform corresponding to the same wire. The same waveform is shown in purple for no calibration correction and in green (shifted on the  $x$  axis to facilitate comparison) for the calibration correction sample. The waveform illustrates the drift direction dependency of the calibration correction. While the first hit is unaffected by the calibration correction, the second hit is down-scaled by the effects of the calibration corrections.

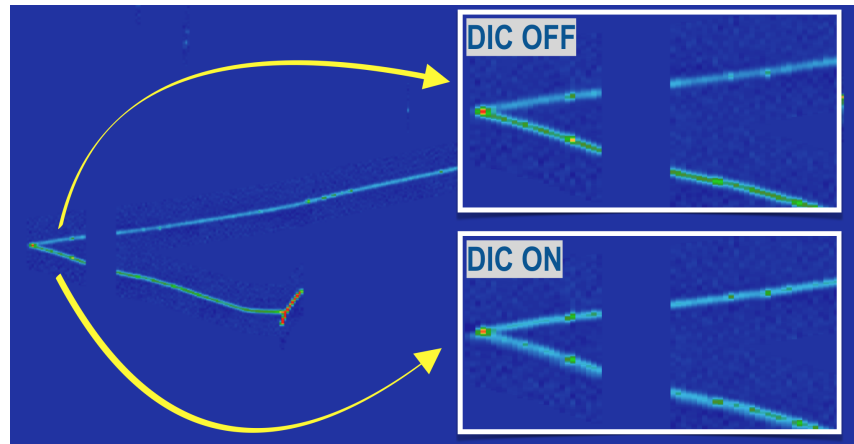
depositions. E.g., a MC track simulated near the cathode will have a correspondingly large drift distance.

The calibration corrections for large drift distances ( $x$ ) compensate for larger diffusion and attenuation, so that using the inverse of the correction ensures that the MC generated signal is degraded accordingly. The application of the calibration corrections on the simulated MC event represents an improvement over the simulated detector effects, since they reflect the time dependency of the effects and they are





**Figure 8.9:** Simulated MC HNL decay. In green, yellow and red are shown the muon and pion trajectories simulated with the standard G4 settings, which include the SCE effect, producing the characteristic bending features for long tracks. In monochromatic blue is shown the same event, where the muon and pion tracks have been simulated with G4, but switching off the SCE effects. The figure shows how the SCE can be considered negligible for very short track, but can drastically affect longer tracks.



**Figure 8.10:** Simulated MC HNL decay. The two zoomed images on the right of the figure show the same vertex with DIC effect disabled (top) and enabled (bottom). The figure shows how including the DIC effects leads to a worsening of the resolution, which can affect the reconstruction algorithms.

built to match the underlying data event.

### 8.5.1 Calibration

A method for estimating the uncertainty on the application of the calibration corrections with the uni-sim approach is to generate a sample in which the calibration corrections are not applied, but the standard simulated detector effects are used instead. By comparing this sample to the CV sample with corrections applied an uncertainty on the calibration method (which encapsulate the uncertainties related to the detector effects) can be estimated.

### 8.5.2 Space Charge Effect

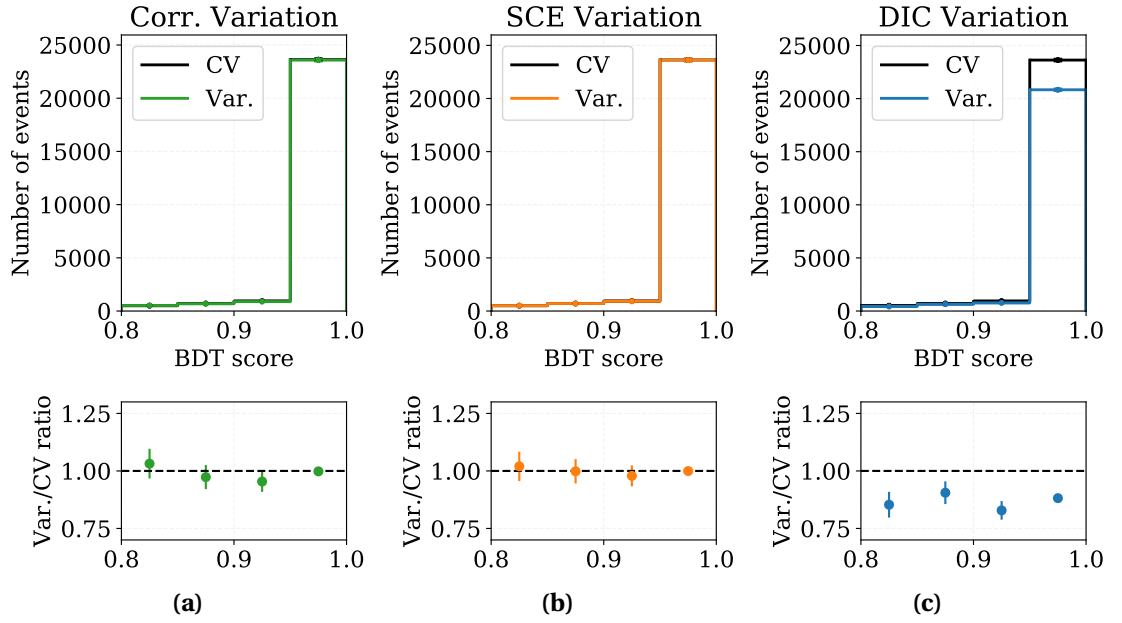
The calibration uncertainty captures the effect due to variations in the electric fields caused by slowly-drifting ions produced by the CR muons (SCE, described in [section 3.3.4](#)). However it does so in an *averaged* way. The individual energy deposits are modified by the action of the corrections, but the track trajectories are unaffected. To estimate the impact of the uncertainties on the SCE model on the reconstruction efficiency, a separate source of uncertainty needs to be taken into account. This is performed with the uni-sim approach by switching the SCE model at the G4 stage, thus producing a sample which neglects the effects of the variations of the electric fields due to slow-drifting ions on the MC simulated tracks. [Figure 8.9](#) shows two superimposed displays of the same “truth” event, simulated with the SCE model switched on and off.

### 8.5.3 Dynamically Induced Charge

The term Dynamically Induced Charge (DIC, described also in [section 5.6.2](#)) refers to the charge induced on the wires beyond the nearest wire. MCC8 does not use DIC modelling for standard simulation, so that the charge drifting past the sense wires is assigned to the nearest channels. DIC effects have the greatest impact on the induction wire planes, in particular for tracks which are orthogonal to the wire plane (see [figure 5.19](#)). In order to estimate the limitations of the current MCC8 MC modelling, we produce variation samples which simulate the induction effects up to the  $N - 10$  and  $N + 10$  wires, where  $N$  is the wire nearest to the drifting charge. [Figure 8.10](#) shows the same simulated MC HNL event with DIC effect enabled and disabled.

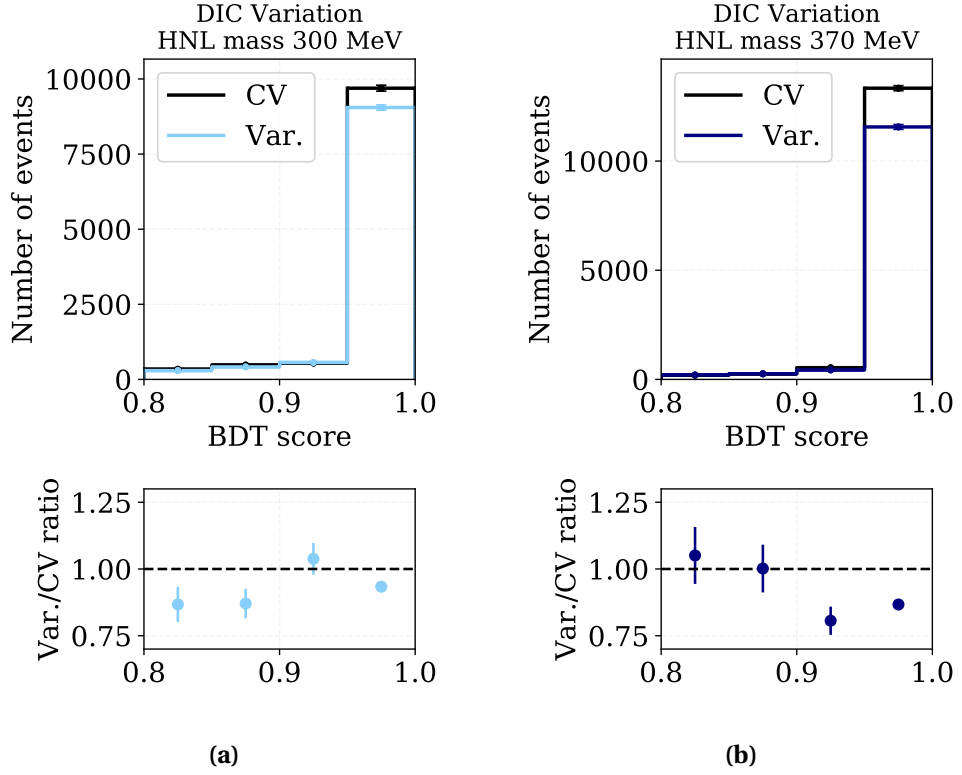
### 8.5.4 Detector Systematics Summary

Due to the computationally intense nature of the detector uncertainty determinations, which are performed with uni-sim methods and require MC samples to be generated for each variations and to be propagated through the whole simulation workflow, the systematics related to the calibration corrections and SCE effects are determined for a single mass point (345 MeV). The DIC systematic uncertainties are determined for two additional mass points (300 MeV and 370 MeV) due to their larger contribution to the total uncertainty. 220,000 events are generated for the CV sample and 220,000 events are generated for each variation sample.



**Figure 8.11:** BDT score distribution for a simulated MC CO HNL sample with a mass of 345 MeV. The CV sample has been produced with the default values for all the detector simulation parameters (shown in black, in the first two plots is underneath the coloured histograms), while the variation samples are indicated with different colours (green for calibration uncertainties, orange for SCE and blue for DIC).

All samples share the same underlying MC-generated events, in order to avoid effects due to statistical fluctuations. The samples were produced in a tighter timing window surrounding the HNL trigger (4830 ns to 5500 ns) than the standard MC HNL samples, to improve the MC efficiency. As a consequence the efficiency for the events that survive the timing selection is higher than the timing selection efficiency for

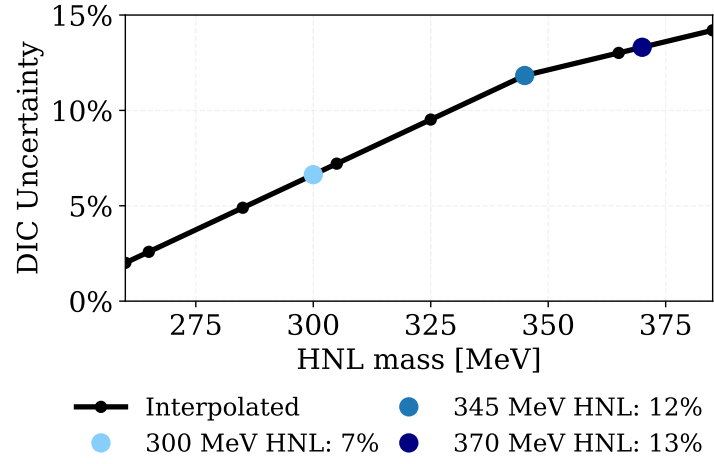


**Figure 8.12:** BDT score distribution for simulated **MC CO HNL** sample with masses of 300 MeV (a) and 370 MeV (b), respectively.

the standard samples (shown in [figure 6.6](#)). This higher efficiency is an artefact of generating fewer **MC** events in timing regions that are successively removed by the selection requirements.

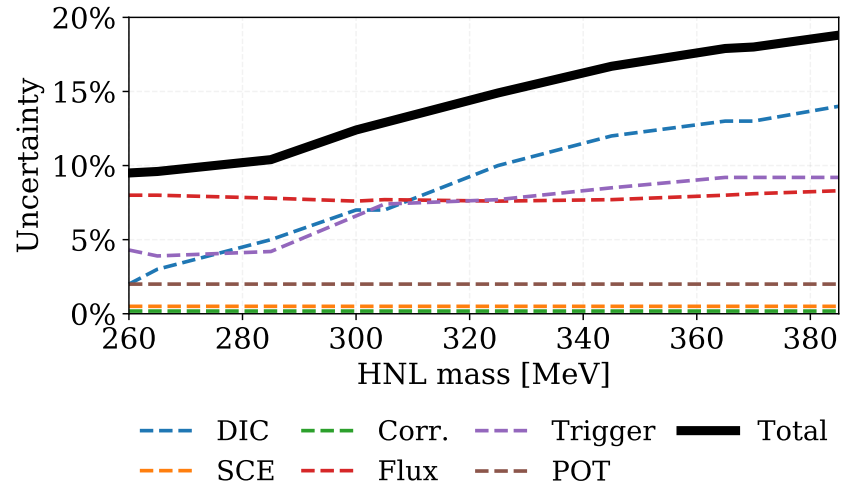
By comparing the variation samples to the **CV** samples, the uncertainties can be determined. The **DIC** uncertainty dominates over the other two, with an uncertainty of 12%, while the calibration and **SCE** uncertainties are on the order of 0.3% - 0.4%. The effect of the detector effects on the **BDT** distribution for **MC HNL** signal is shown in [figure 8.11](#). [Table 8.1](#) summarises the uncertainties as determined from the histograms in [figure 8.11](#).

Due to the larger contribution of the **DIC** effect to the total detector systematic uncertainty, two additional samples with **HNL** masses of 300 MeV and 370 MeV are produced to determine the dependency of the **DIC** uncertainty with respect to the **HNL** mass, which are shown in [figures 8.12a](#) and [8.12b](#). Based on the three samples,



(a)

**Figure 8.13:** Mass dependency of the DIC effect obtained by interpolating and extrapolating the systematic uncertainty values obtained for the three samples simulated.



**Figure 8.14:** Systematic uncertainties as a function of mass. The figure shows how the main three sources of uncertainties are provided by the trigger, flux and DIC systematics.

DIC uncertainties for all the other mass points are obtained by interpolating and extrapolating to the known values.

Detector Variation Samples			
Variation	Events	Events (BDT > 0.95)	Fractional change [%] (BDT > 0.95)
CV sample [ $m_N = 345$ MeV]	30,038	23,627	-
Calibration variation	29,894	23,584	-0.18%
SCE variation	30,089	23,615	-0.05%
DIC variation	26,472	20,832	-12%
CV sample [ $m_N = 300$ MeV]	13,736	9,693	-
DIC variation	12,602	9,050	-7%
CV sample [ $m_N = 370$ MeV]	16,834	13,339	-
DIC variation	14,575	11,564	-13%

**Table 8.1:** Comparison of the number of events in each variation sample. The second column shows the number of events remaining in each sample after the application of all the pre-selections and BDT requirements (only one candidate per event). In the third column a further requirement for HNL-like signature is required (BDT > 0.95, the same requirement imposed before limit extraction). The fourth column uses the event number of the third column to extract fractional uncertainties. The table shows how the DIC uncertainty dominates over the other two.

Systematic Uncertainty Summary	
Systematic Source	Relative Uncertainty
Flux	8%
POT	2%
Trigger	3% - 8%
Calibration	0.4%
SCE	0.3%
DIC	7% - 12%
Total	10% - 19%

**Table 8.2:** Systematic uncertainties for the HNL sample with high BDT score (BDT > 0.95). In the rows with two numbers are shown the uncertainty values for the lowest mass (260 MeV) and the highest mass (385 MeV) studied.

The mass dependency of the DIC effect is shown in figure 8.13a. The associated systematic increases at higher masses due to the DIC effect affecting predominantly longer tracks. As the mass of the HNL increases, so does the energy available in the HNL rest frame and the momentum of the daughter particles, which results in longer tracks. The effect does not increase linearly since the limited volume of the TPC and the addition of the containment selection requirement reduces the differences in the

track length distributions at increasingly larger masses.

## 8.6 Systematics Summary

The systematic uncertainties are estimated with different methods. Among them, the largest sources of systematic uncertainty are determined to be **DIC** ( $\approx 12\%$ ) and the flux and trigger uncertainties ( $\approx 8\%$  and  $\approx 7\%$ , respectively). A key feature of the **HNL** samples which drastically reduces the impact of detector systematics (compared to other MicroBooNE analyses) is the use of the **MC CO** technique, which allows for the production of samples which have no simulated noise or background, thus entirely removing the impact of the uncertainties associated with the underlying **MC** models. [Figure 8.14](#) shows the systematic uncertainties as a function of **HNL** mass. In [table 8.2](#) are shown the values for the uncertainty systematics studied in this section.

# Results

The data obtained by the application of the BDTs can be used to perform statistical test on the hypothesis that the measured data can be described by a sum of the background dataset and a signal dataset. This section describes the procedure used to determine the limits on the mass-mixing matrix element and gives the results.

## 9.1 Limit Extraction Procedure

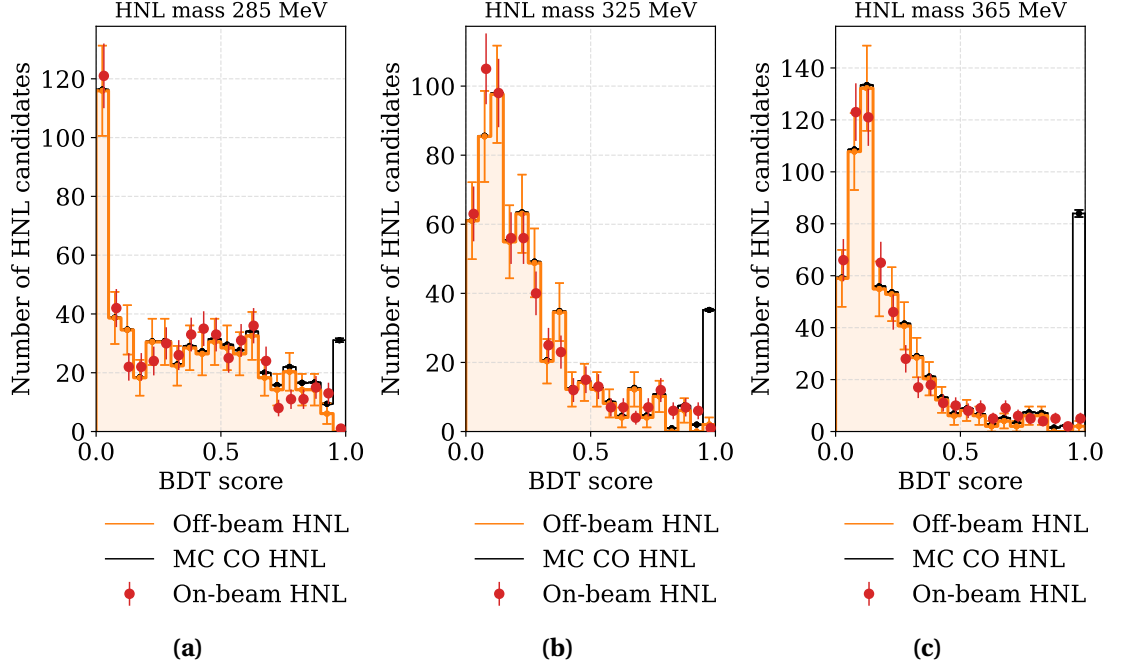
To determine the limits on the  $|U_{\mu 4}|^2$  mass-mixing matrix element, the trained BDTs are applied to the *off-beam HNL*, *on-beam HNL* and MC CO HNL (with generated mass value matching the mass value the BDT was trained on) simulated samples. The resulting BDT distributions are shown in [figure 9.1](#) for three mass points. The data and background distributions are found in good agreement. As shown in [figure 9.1](#) the HNL signal is contained predominantly in the last bin (0.95 - 1.00 BDT score). We therefore define two regions:

- Signal-depleted **control region** (BDT score in interval [0.5 - 0.95]), which has higher statistics in both the on-beam and off-beam sample and is used to test the relative normalisation of the samples.
- Signal-rich **test region** (BDT score in interval [0.95 - 1.00]), which is used to test the HNL hypothesis.

The efficiencies for this selection are shown in [table 9.1](#). At this stage, if multiple



candidates are still found in an event, the candidates with the highest BDT score is kept, while the others are rejected. Tables 9.2 and 9.3 show the number of events contained in the two regions. The limiting factor is the statistics available in the data background sample, which is smaller than the signal sample. The event number shown for the signal is a reference point for a fixed mass-mixing element  $|U_{\mu 4}|^2 = 10^{-7}$ .



**Figure 9.1:** BDT score distribution for the data *on-beam HNL* (red) and background *off-beam HNL* (orange) data samples. The signal *MC CO HNL* sample with the same mass used to train the BDT (black) is added on top of the background histogram. The *MC CO HNL* sample is normalised by the scaling factor determined during the limit setting procedure, which gives the minimum amount of expected HNL candidates that can be excluded at the 90% CL, and by a further fixed factor of 10 to make it visible in the figure.

In order to determine whether a signal is contained in the data sample, a discriminant to perform a statistical analysis is needed. The discriminant operates by testing two different hypotheses:

- **Null hypothesis** (or background hypothesis). This assumes that the observed data is provided exclusively by the background distribution. The hypothesis can be described by:

$$N_{\text{events}}^{\text{NULL}} = R_B \times \varepsilon_B \times n_{\text{ht}}, \quad (9.1)$$

<i>Test region</i>	
BDT selection efficiency	
HNL Mass [MeV]	Efficiency
260	0.40 ( $\pm 0.01$ )
265	0.43 ( $\pm 0.02$ )
285	0.59 ( $\pm 0.02$ )
300	0.71 ( $\pm 0.01$ )
305	0.74 ( $\pm 0.01$ )
325	0.76 ( $\pm 0.01$ )
345	0.79 ( $\pm 0.02$ )
365	0.81 ( $\pm 0.02$ )
370	0.80 ( $\pm 0.01$ )
385	0.80 ( $\pm 0.01$ )

**Table 9.1:** Signal efficiency after a selection on BDT score  $> 0.95$ . Uncertainties are statistical only.

<b>Control region ( <math>0.5 \leq \text{BDT} &lt; 0.95</math> )</b>			
HNL Mass [MeV]	Signal	Background	Data
	MC CO HNL (at $ U_{\mu 4} ^2 = 10^{-7}$ )	Off-beam HNL	On-beam HNL
260	0.2	169 ( $\pm 19$ )	170
265	0.3	185 ( $\pm 19$ )	205
285	0.4	175 ( $\pm 19$ )	174
300	0.5	126 ( $\pm 16$ )	121
305	0.4	61 ( $\pm 11$ )	80
325	0.8	57 ( $\pm 11$ )	69
345	1	59 ( $\pm 11$ )	75
365	1	35 ( $\pm 8$ )	53
370	2	37 ( $\pm 9$ )	47
385	2	20 ( $\pm 6$ )	28

**Table 9.2:** Number of events in the control region, defined for BDT score in the range  $[0.5 - 0.95]$ , for the data, background and signal sample. The off-beam sample has statistical uncertainties only, which are determined as  $\sqrt{N}$  of the number  $N$  of events before re-scaling.

where  $R_b$  is the background rate (provided by CR muons),  $\varepsilon_b$  is the overall efficiency for triggering and selecting from the background sample an HNL candidate after pre-selection and with high BDT signal-like score, and  $n_{\text{ht}}$  is the number of hardware trigger issued (or analogously, the number of events recorded prior

Test region (BDT $\geq 0.95$ )			
HNL Mass [MeV]	Signal	Background	Data
	MC CO HNL (at $ U_{\mu 4} ^2 = 10^{-7}$ )	Off-beam HNL	On-beam HNL
260	0.1	$< 3.7$	1
265	0.2	$2 (\pm 2)$	3
285	0.8	$< 3.7$	1
300	1	$2 (\pm 2)$	4
305	2	$2 (\pm 2)$	0
325	3	$2 (\pm 2)$	1
345	6	$2 (\pm 2)$	4
365	10	$2 (\pm 2)$	5
370	12	$2 (\pm 2)$	4
385	18	$< 3.7$	4

**Table 9.3:** Number of events in the test region, defined for BDT score in the range  $[0.95 - 1]$ , for the data, background and signal sample. The off-beam sample has statistical uncertainties only, which are determined as  $\sqrt{N}$  of the number  $N$  of events before re-scaling. For the background sample,  $< 3.7$  indicates the re-normalised 68% CL upper limit for a Poisson distribution with mean 0 for BDTs that do not contain background events in the test region [84].

to any trigger requirement), which is used an overall normalisation factor.

- **Test hypothesis** (or signal plus background hypothesis). This assumes that the observed data is provided by the sum of the signal and background distributions. The hypothesis can be described by:

$$\begin{aligned}
 N_{\text{events}}^{\text{TEST}}(m_N, |U_{\mu 4}|^2) &= [R_B \times \varepsilon_B \times n_{\text{ht}}] + [R_S(m_N, |U_{\mu 4}|^2) \times \varepsilon_S \times \text{POT}] \\
 &\simeq [R_B \times n_{\text{ht}}] + [|U_{\mu 4}|^4 \times R_S(m_N) \times \varepsilon_S \times \text{POT}],
 \end{aligned} \tag{9.2}$$

where  $N_{\text{events}}^{\text{TEST}}(m_N, |U_{\mu 4}|^2)$  directly depends on the mass and mass-mixing matrix element of the HNL studied.  $R_S(m_N, |U_{\mu 4}|^2)$  is the rate of HNL decays occurring in the MicroBooNE detector, which depends on the HNL mass and mass-mixing matrix element,  $\varepsilon_S$  is the overall efficiency for triggering and selecting a HNL candidate from the HNL sample after pre-selection and with high BDT signal-like score, and POT is the number of protons on target, which works as a normalisation factor. For the range of mass-mixing matrix element stud-

ied in this analysis, the decay rate depends on  $|U_{\mu 4}|^4$  (a factor of  $|U_{\mu 4}|^2$  acts at production and another factor of  $|U_{\mu 4}|^2$  is involved at decay) so that we can write  $R_S(m_N, |U_{\mu 4}|^2) \simeq |U_{\mu 4}|^4 \times R_S(m_N)$ . This approximation stops being valid for mass-mixing matrix element values larger than  $10^{-3}$  (see [figure 5.8](#)), due to the number of decays occurring in-flight before reaching the detector becoming non-negligible, causing a depletion of the flux observed at the detector. The values probed here are around  $10^{-8}$ - $10^{-6}$ , making the assumption valid at all masses.

The test hypothesis reduces to the null hypothesis in the case in which the mass-mixing element reduces to zero. A statistical test can be performed to evaluate whether the observed dataset  $N_{\text{events}}^{\text{observed}}$  is better described by the null hypothesis or by the test hypothesis. Since the value of  $|U_{\mu 4}|^2$  is not known a priori it can assume any value for testing. Two outcomes are possible, after hypothesis testing:

- The signal dataset  $N_{\text{events}}^{\text{observed}}$  observes an excess with respect to  $N_{\text{events}}^{\text{NULL}}$ . The magnitude of the excess can be used to determine the  $|U_{\mu 4}|^4$  mass-mixing element from the  $N_{\text{events}}^{\text{TEST}}(m_N, |U_{\mu 4}|^2)$  hypothesis and the relevant confidence level that can be placed on the hypothesis.
- The signal dataset  $N_{\text{events}}^{\text{observed}}$  is consistent with the  $N_{\text{events}}^{\text{NULL}}$  hypothesis. This excludes all  $N_{\text{events}}^{\text{TEST}}(m_N, |U_{\mu 4}|^2)$  with a value of  $|U_{\mu 4}|^2$  so that  $N_{\text{events}}^{\text{TEST}}$  is significantly different from  $N_{\text{events}}^{\text{observed}}$ . Below a determined value of the signal intensity  $|U_{\mu 4}|_{\text{limit}}^2$ , the contribution of the signal to the  $N_{\text{events}}^{\text{TEST}}$  is so small that  $N_{\text{events}}^{\text{TEST}}$  and  $N_{\text{events}}^{\text{NULL}}$  are statistically indistinguishable and  $N_{\text{events}}^{\text{observed}}$  is consistent with both. In that case the experiment is not able to exclude the possibility of HNL with mass-mixing matrix element below  $|U_{\mu 4}|_{\text{limit}}^2$  and  $|U_{\mu 4}|_{\text{limit}}^2$  is used as a limit on all the values that can be excluded. The value depends on the [CL](#) required.

A statistical test can be performed by producing *ensemble* experiments (also known as pseudo-experiments), via [MC](#) methods. This is performed by constructing *probability distribution functions* (PDF) for  $N_{\text{events}}^{\text{NULL}}$  and  $N_{\text{events}}^{\text{TEST}}$ . We assume that the datasets are governed by a Poisson distribution with mean  $N_{\text{events}}^{\text{NULL}}$  in the null hypothesis and  $N_{\text{events}}^{\text{TEST}}$  in the test hypothesis and that the observed dataset is randomly sampled from one of the distributions. Pseudo-experiments can be generated for the two hypotheses

and the distributions resulting from the pseudo-experiment ensembles will describe the PDFs for the two hypotheses. The effect of the uncertainties of the nuisance parameters, such as  $\varepsilon$  and POT, is to have each of the pseudo-experiments sampled from a Poisson distribution where the mean of the distribution is also randomly varied, with the variation described by the uncertainty on the nuisance parameters (sampled from a Gaussian function in this study). From the PDFs a likelihood can be evaluated for a dataset to be described by one of the two hypotheses. Assuming Poisson PDFs:

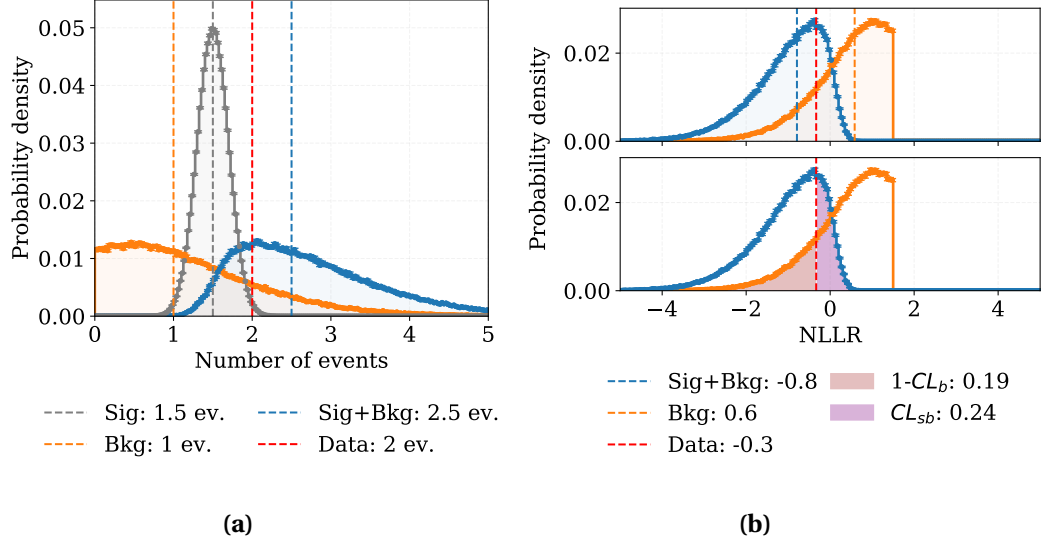
$$\begin{aligned} L^{\text{NULL}}(b, d) &= \frac{e^{-b} b^d}{d!}, \\ L^{\text{TEST}}(s, b, d) &= \frac{e^{-(b+s)} (b+s)^d}{d!}, \end{aligned} \quad (9.3)$$

where  $s$ ,  $b$ , and  $d$  are the expected signal and background, and observed data, number of events in a determined bin. For multiple bins, the product of the likelihoods gives the total likelihood.

A commonly used test statistics is the negative log-likelihood ratio, which is expressed as:

$$\begin{aligned} \text{NLLR}(s, b, d) &= \log\left(\frac{L^{\text{NULL}}(b, d)}{L^{\text{TEST}}(s, b, d)}\right) = \log(L^{\text{NULL}}(b, d)) - \log(L^{\text{TEST}}(s, b, d)), \\ &= s - d \log\left(1 + \frac{s}{b}\right). \end{aligned} \quad (9.4)$$

Figure 9.2 illustrates how PDFs can be built from expected and observed values and what NLLR distributions may results from such distributions. A high NLLR value indicates a dataset is more likely to have resulted from the background distribution, whereas a low NLLR value indicates the dataset is more likely to have resulted from the signal plus background distribution. An overlap is generally expected between the two distributions, for NLLR values where no hypothesis can be preferred over the other. The amount of overlapping is also an indicator of the sensitivity of the experiments. I.e., if the two distributions are not overlapping a conclusive statement can be made on whether a dataset is more background or signal plus background like, whereas for completely overlapping distributions no statistically significant statement can be made.



**Figure 9.2:** Toy PDF distributions obtained from the pseudo-experiments generated with statistics similar to the ones obtained in this study (a). Using Poisson distribution, pseudo-experiments are generated around the background expected number of events (orange) for the null hypothesis. The test hypothesis is obtained summing the number of signal and background events. The signal mean value is obtained by using high-statistics MC and scaling down to the expected normalisation. As a consequence the pseudo-experiments distribution is dominated by systematics which have a Gaussian profile, resulting in the gray distribution. The signal plus background distribution (blue) is thus obtained from a convolution of the background (orange) and signal (gray) distributions.  $NLLR$  distribution values are obtained by calculating the  $NLLR$  for each pseudo-experiment, producing the distributions in figure (b). The cut-off in the background  $NLLR$  distribution is due to the Poisson dominated PDF used for the pseudo-experiments. The  $CL_{sb}$  and  $1-CL_b$  can be obtained by integrating the  $NLLR$  distribution up to and from the  $NLLR_d$  value.

A statistical test used to quantify the degree of agreement to one hypothesis or the other makes use of “Confidence Levels” (CL). A CL ( $CL_{sb}$ ) for the test hypothesis is constructed as the probability of the test hypothesis to be more background like than the observed data, whereas  $CL_b$  is constructed as the probability of the null hypothesis to be more background like than data. A more robust test statistics is the use of  $CL_s$  defined as:

$$CL_s = \frac{CL_{sb}}{CL_b}, \quad (9.5)$$

which protects against poor background modelling, that can produce false exclusions

[85]. For exclusion,  $\text{CL}_s$  is calculated as a function of the parameters until  $1 - \text{CL}_s < \alpha$ , where  $\alpha$  is a fraction, set for this analysis at  $\alpha = 0.1$ , so that the exclusion is performed at the 90% CL.

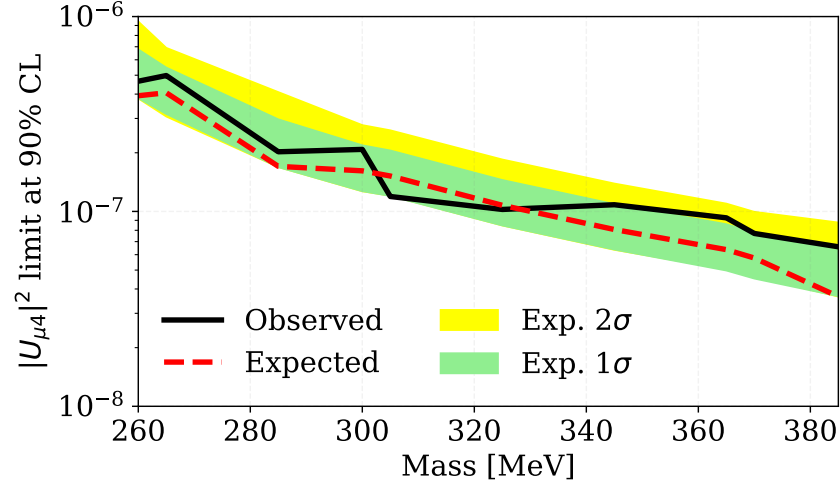
## 9.2 Mass-mixing Matrix Element Limits

The event count from [table 9.2](#) and [table 9.3](#) are provided as a two-binned histogram to COLLIE, a CLs method software toolkit for confidence level limit evaluation [85–87]. The aim of COLLIE is to generate the pseudo-experiments and to use a Poisson log-likelihood ratio test statistic to determine confidence intervals. [Table 9.4](#) shows the 90% CL exclusion limits that can be placed on the  $|U_{\mu 4}|^2$  mass-mixing matrix element from the event count shown in [tables 9.2](#) and [9.3](#). The same limits are also illustrated in [figure 9.3](#). [Figure 9.4](#) shows the same limits with respect to the other experiments, discussed in [section 2.4](#).

MicroBooNE limits on $ U_{\mu 4} ^2 (\times 10^7)$					
HNL Mass [MeV]	Observed	Median Expected	Exp. $1\sigma$	Exp. $2\sigma$	1- $\text{CL}_b$
260	4.65	3.92	3.81-6.78	3.81-9.34	0.65
265	4.98	4.06	3.16-5.49	3.16-6.91	0.72
285	2.03	1.70	1.69-2.97	1.69-4.09	0.64
300	2.08	1.62	1.27-2.19	1.27-2.77	0.78
305	1.19	1.52	1.20-2.05	1.20-2.61	0.28
325	1.02	1.08	0.84-0.14	0.84-0.18	0.45
345	1.08	0.80	0.63-0.10	0.63-0.13	0.83
365	0.92	0.63	0.50-0.86	0.50-1.10	0.90
370	0.77	0.57	0.45-0.77	0.45-0.99	0.82
385	0.65	0.36	0.36-0.63	0.36-0.87	0.86

**Table 9.4:** 90% CL limits for a Majorana HNL decaying via the  $\mu\pi$  channel in MicroBooNE, obtained from the measured events shown in [tables 9.2](#) and [9.3](#), from data collected by the delayed HNL trigger throughout Run 3, corresponding to  $1.97 \times 10^{20}$  POT. The observed, median expected,  $1\sigma$  and  $2\sigma$  columns have been multiplied by a factor of  $10^7$ .

The HNL mass dependent systematic uncertainties described in [section 8](#) (calibration, SCE, DIC, flux, POT, and trigger) are assigned to the signal sample only (MC CO HNL). Limits are extracted by convolving the Gaussian distribution describing

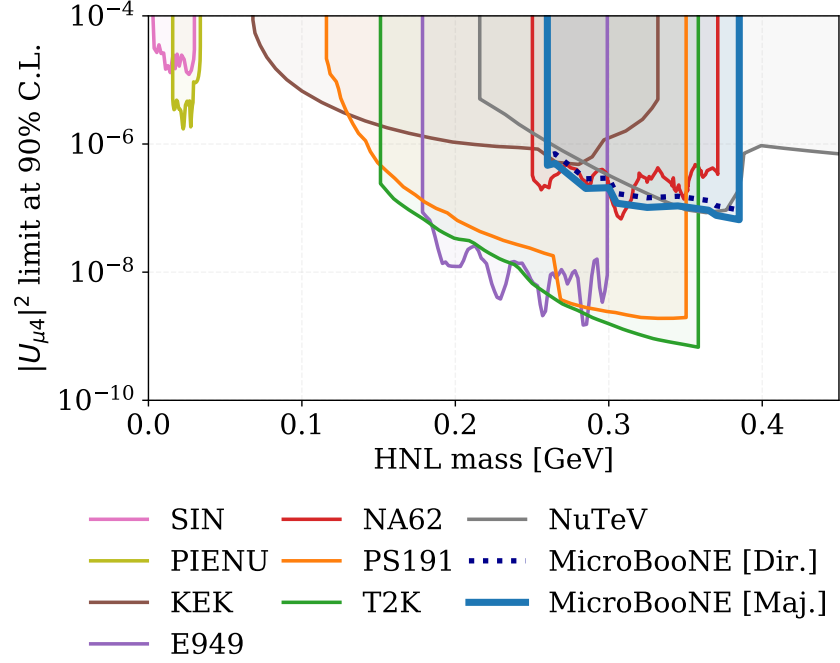


**Figure 9.3:** 90% CL limits for a Majorana HNL decaying via the  $\mu\pi$  channel in MicroBooNE, obtained from the measured events shown in tables 9.2 and 9.3, from data collected by the delayed HNL trigger throughout Run 3, corresponding to  $1.97 \times 10^{20}$  POT. The lack of a visible lower  $2\sigma$  bound is due to the 68% and 95% lower interval bounds assuming the same value.

the systematic uncertainties with the Poisson distribution used for the PDF of the pseudo-experiments generated in the signal plus background hypothesis. The systematic uncertainties on the signal are found to have a negligible impact on the limit obtained, since the limits is largely dominated by the amount of statistics available in the background sample, which is smaller than the data sample. Systematic uncertainties for the background only hypothesis are derived from the  $\sqrt{N}$  statistical uncertainties on the background sample (*off-beam HNL*) used for the generation of the pseudo-experiments, before re-scaling to the signal sample normalisation (i.e., after re-scaling the sample by a factor 2.034, the statistical uncertainties are re-scaled by the same factor, to obtain  $2.034\sqrt{N}$ ).

No significant excess consistent with a Majorana HNL has been observed at MicroBooNE, and in the lack of a signal, limits at the 90% CL can be placed in the mass region 260-385 MeV for  $|U_{\mu 4}|^2$  in the range  $[7 \times 10^{-8}, 5 \times 10^{-7}]$ . The results are comparable with the limit set by NA62 [8] and NuTeV [34]. PS191 [31, 32] and T2K [33] place more stringent limits in the mass range [260, 360] MeV by two orders of magnitude, although the location of the detectors, which are both positioned with a  $\approx 2^\circ$  off-axis angle prevents them from fully extending the same sensitivity to the kinematically





**Figure 9.4:** 90% CL limits for HNL in the  $m_N - |U_{\mu 4}|^2$  parameter space. The blue solid line indicates the limit set by the MicroBooNE experiment for from the analysis shown in this work, which used  $1.97 \times 10^{20}$  POT data collected by the delayed HNL trigger during the Run 3 of MicroBooNE. The dotted blue line show the same result scaled by a factor of 2 for Dirac HNLs, for which the  $\mu\pi$  decay channel has its rate halved with respect to Majorana.

available mass range. The detectors of the SBN and DUNE programmes are located on-axis, which allows them to extend the sensitivity to HNLs to 385 MeV HNL mass. A complementary extension to this analysis can be performed by analysing the *on-beam BNB* dataset, which would provide similar sensitivities but with a higher background. For Dirac neutrinos, the rate expected for the same mass-mixing matrix element is  $1/2$ , so that in figure 9.4 limits on Dirac HNL are shown by applying a reducing factor of  $1/\sqrt{2}$ . The angular distribution for Dirac HNLs is expected to differ from the Majorana HNL angular distribution by less than 30%, which does not significantly affect the limit [66].

# Conclusions

In this thesis, HNLs are studied using data corresponding to  $1.97 \times 10^{20}$  POT, collected by the MicroBooNE detector during its Run 3, in the period of time between October 2017 and June 2018. The search focuses on the decay of Majorana HNLs produced by kaon decay and decaying to  $\mu\pi$  pairs, which is the dominant decay channel above HNL masses of 245 MeV and offers direct sensitivity to the  $|U_{\mu 4}|^2$  mass-mixing matrix element in the mass range 245-388 MeV.

This search takes advantage of the slow time-of-flight of the HNLs, which due to their mass, arrive at the detector with a delay with respect to the SM active neutrinos produced by the same production channel. HNLs in this timing window are calculated to constitute 5%-10% of the total HNL flux, depending on the mass. A special HNL trigger is set up to record delayed events in a timing region with no SM interaction background, and where the dominant source of background is determined to be from CR muons. The HNL trigger is the first of its kind ever used for HNL searches, and its use demonstrates the validity of a new method for HNL searches that can be extended to other beam-dump like experiments. The HNL trigger has an efficiency of  $\approx 80\%$ -84%. Sources of inefficiencies are due to the limited time resolution of the PMTs and the effect of the veto preceding the HNL trigger. The HNL trigger is currently part of the standard set of MicroBooNE triggers and operates in stable conditions with a fixed event rate of 0.4%.

The use of a background dataset provided by real data in an off-beam configuration of the HNL trigger, analogous to the on-beam configuration, provides a data and background samples which are not affected by systematic uncertainties related to

---

MC-based studies. The use of a novel MC technique (CO) which combines real data and simulated data for the generation of background CR muons and noise in MC-generated sample also results in lower systematic uncertainties usually associated with the MC generation process.

The analysis studies HNL decays at ten benchmark mass points in the [260-385] MeV mass range. For each mass point, MC signal samples are generated and filtered through a series of pre-selection requirements before being processed by a BDT. HNL candidates are constructed from reconstructed neutrino vertices with two-track objects associated to them. Efficiencies for successfully reconstructed HNL candidates are 20%-50% depending on HNL mass. A series of pre-selections are applied to the HNL candidates, with the aim of improving the quality of the resulting sample. Signal efficiencies after pre-selection vary across mass values, with typical efficiencies at  $\approx 4\%$  and background rejection at  $\approx 0.001\%$ . A series of BDTs is trained for each simulated mass point to further discriminate signal from background. The BDTs show good performance with AUC metric  $\approx 0.95$ .

Sources of systematic uncertainties associated with the signal HNL are studied, using uni-sim and multi-sim approaches. The detector related systematics, SCE and calibration, are subdominant with systematics below the 1% level across all mass values. The flux, trigger and DIC effect systematics have the largest impact on the total systematic uncertainty, which is at the 10%-19% level.

A control sample from BNB  $\nu$ -interaction data is used to test the sensitivity of the pre-selection and BDTs to data-MC differences, which shows good agreement between data and MC and validates the selection procedure. Results are divided in two bins and the CLs method is used to extract limits on the HNL mass-mixing matrix elements. In the absence of a signal, limits are placed for  $|U_{\mu 4}|^2$  in the range  $[7 \times 10^{-8}, 5 \times 10^{-7}]$ . The impact of the systematic signal uncertainties on the limit is found to be negligible, since the limit is dominated by the Poisson distribution governing the low-level background. MicroBooNE's result are currently only statistically limited. The limits set by MicroBooNE are comparable with the results obtained by NA62 and NuTeV.

This search is the first of its kind ever performed in a LArTPC and it paves the way for similar searches to be performed in the upcoming experiments part of the SBN and

---

DUNE programmes. Since this search is currently statistically limited, and it uses only a subset of the data available to MicroBooNE, the inclusion of additional detectors, POTs, and extended timing ranges that include the *on-beam BNB* dataset can boost the sensitivity by few order of magnitudes. Searches in the timing region coincident with the SM BNB neutrino interaction window can increase the signal rate by a factor of  $\approx 20$ . Although additional background caused by neutrino interaction may mimic the signal, the excellent spatial resolution of liquid argon TPC can be used to remove neutrino interaction background by applying more stringent selections on the quality of the vertex. Additional signal events may also be recovered by investigating the track-shower objects that do not pass the track-track requirements and their possible inclusion in the signal sample. Timing resolution is an important factor affecting efficiency, improvements in the timing resolution of the optical instrumentation can provide more precise triggering and the possible reduction of the buffer window separating the veto timing window from the signal window, thus allowing the inclusion of more events with a smaller delay in their time-of-flight. The combination of data from multiple detectors positioned at different distances from the beamline target, such as the detectors of the SBN and DUNE programmes, can be employed to increase the sensitivity by allowing the use of timing windows probing HNLs with different delays in their time-of-flight and looking at the ratio in the expected number of events, which would be different for different HNL masses. In addition, the detectors of the SBN and DUNE programmes are uniquely positioned on-axis and can take full advantage of the boosting effects of the HNLs, especially at higher masses, which offer larger fluxes and sensitivity to mass values that cannot be probed by off-axis experiments.

# Glossary

**art** *Analysis Reconstruction Tools*. Fermilab-supported analysis tools on top of which LArSoft is built. It allows access to *event* structures and the construction of event collections. See [section 5.1](#). 48, 80, 128

HELEPHANTE *HELEPHANT Explorer*. LArSoft analyser module, used to find preliminary HNL (HELEPHANT) candidates. See [section 5.3](#). 13, 128–130, 135

**ADC** *Analog-to-Digital Converter units*. Units of output for the PMT and TPC waveforms after digitisation. After calibration, these can be converted into physical units (e.g., for PMTs 1 ADC  $\approx$  0.05 PE). 42, 71

**ASIC** *Application-Specific Integrated Circuits*. Set of circuits placed inside the cryostat which pre-amplify and shape the signal from the wires, before sending it outside the cryostat vessel. See [section 3.2.3](#). 41, 44, 93, 98, 100–102, 109–112, 122

**AUC** *Area Under Curve*. Area under the ROC curve of a classifier. It expresses the probability that a randomly chosen signal event would get a higher BDT score than a randomly chosen background event. 168, 169, 199

**BDT** *Boosted Decision Tree*. Predictive model based on a tree-like model of decisions. See [section 7](#). 120, 124, 125, 127, 131, 154, 155, 157, 161, 163–173, 183, 184, 186, 188–191, 199, 201

**BNB** *Booster Neutrino Beamline*. MicroBooNE’s primary source of neutrinos, using 8 GeV protons impinging on a target located  $\approx$  470 m from the detector. The acronym is also used to indicate data collected by the BNB trigger (i.e., data containing neutrino interactions from BNB neutrinos). See [section 3.4](#). 32, 33, 58, 60, 61, 64, 66–72, 74–78, 80–82, 84, 87, 90, 95, 97, 103, 119–127, 132, 133, 135, 143, 146–150, 153, 157, 158, 161, 162, 164, 166, 168, 170–173, 175, 177–179, 197, 199, 200

**CL** *Confidence Level*. Also known as coverage. Level of confidence that the true value of a measured parameters lie in the confidence interval associated with the

- confidence level. With the term CLs method is indicated a statistical method for performing interval estimation on model parameters. 189, 191, 192, 194–197, 199
- CO** *Cosmic Overlay*. Monte Carlo sample generation technique which uses the waveforms from real data containing exclusively cosmic-ray crossing muons on which the Monte Carlo samples are *overlaid*. Cosmic Overlay samples do not need to simulate cosmic-ray muon interaction or noise, since these are provided by the input data samples. See [section 5.2.3](#). 91–93, 97, 99, 121, 122, 124–127, 130–133, 135, 136, 146–150, 153, 157, 158, 160, 161, 166, 169, 178, 183, 184, 187–191, 195, 199
- COLLIE** *CO*nfidence *L*evel *L*imit *E*valuator. Software toolkit used to perform hypothesis testing based on the CLs method, by producing ensembles of pseudo-experiments. 169, 195
- CR** *Cosmic-Ray*. High energy radiation originating from outside the Solar System which impacts on the Earth atmosphere and produce showers of secondary particles which reach the surface. Since MicroBooNE does not have a significant overburden cosmic-ray muons continuously cross the detector at 5Hz, producing a source of background. See [section 5.2.3](#). 34, 45, 48, 51, 56, 65, 66, 68, 70, 73–75, 78, 80, 91–94, 97, 98, 105, 106, 117–127, 129, 130, 132, 133, 141–143, 145, 148, 150, 151, 155, 161, 162, 164, 166, 168, 171–173, 177, 178, 182, 190, 198, 199
- CRT** *Cosmic Ray Tagger*. MicroBooNE’s subsystem consisting of scintillating panels surrounding the cryostat, used for tagging cosmic tracks crossing the detector. See [section 3.2.5](#). 45, 46
- CV** *Central Value*. The nominal value assigned to the parameters of the models used for the Monte Carlo generation of simulated events. Uni-sim uncertainty estimations are performed by producing samples where each parameter is changed from its central value to the central value  $\pm 1\sigma$ . In the case of models with no assigned  $\sigma$ , the CV may represent the instance of having the models switched on or off (such as in the case of DIC or SCE). “CV samples” refer to Monte Carlo samples generated with the nominal CV. See [section 8.2](#). 173, 174, 182–184, 186
- DAQ** *Data AcQuisition*. MicroBooNE’s subsystem responsible for processing the data received by the readout electronics, assembling events and saving them to disk. See [section 3.2.6](#). 42, 46–48, 63, 66, 103, 110
- DIC** *Dynamically Induced Charge*. Detector effect caused by the charge induced on the neighbouring wires of a wire collecting an electron. Not modelled in MCC8. See [section 5.3](#). 99, 106, 114, 178, 181–187, 195, 199

- DUNE** *Deep Underground Neutrino Experiment*. Long-baseline (1,300 km) neutrino programme, using a combination of near and far LArTPCs to measure various neutrino properties.. 27, 32, 197, 200
- EXT** *EXTernal trigger*. Hardware trigger recording data *not* in coincidence with a neutrino spill. As a consequence, data collected by the EXT trigger contains exclusively cosmogenic activity which is used to characterise the background of data collected by the other triggers. EXT is used in this work to describe both the hardware trigger and the dataset obtained from the EXT hardware trigger. See section 4.1. 66–68, 72, 75, 77, 78, 92, 103, 119, 121, 126, 131, 161, 164
- FV** *Fiducial Volume*. A volume contained within the instrumented TPC volume. Vertices located in the FV are better reconstructed, due to the length of the tracks associated with them being more likely to be fully contained. See section 6.2.1. 141, 143, 151, 160
- G4** *GEANT4*. Toolkit for the simulation of the passage of particles through matter. With G4 step is indicated the MC procedure of simulating the signal particles (ionisation electrons and scintillation photons) produced in the detector resulting from the energy loss of the secondary particles resulting from interaction or decay. See section 5.3. 81, 83, 85, 93, 97, 137, 181, 182
- GDML** *Geometry Description Markup Language*. Application-independent geometry description format developed at CERN based on XML . Used to describe the geometry of the MicroBooNE detector and surrounding environment. See section 5.3. 98
- HNL** *Heavy Neutral Lepton*. Right-handed neutral lepton with a mass higher than the three active ( $\nu_e, \nu_\mu, \nu_\tau$ ) neutrinos and which doesn't interact weakly but only via mass-mixing with the active neutrinos. Also denoted with  $N$  and the term HELEPHANT (HEavy LEpton, Produced by HAdrons, Neutral and Tardyonic). 9, 11–13, 16–31, 35, 64, 65, 68–96, 119–122, 124–137, 140–143, 145–158, 160, 161, 163, 164, 166, 168–173, 175–178, 180–184, 186–192, 195–200
- LArSoft** *Liquid Argon Software*. Software package toolkit, used to perform simulation, analysis and reconstruction with LArTPCs. See section 5.1. 12, 48, 79–84, 88, 94, 103, 104, 109, 114, 115, 117, 124, 128, 129, 138, 139, 147
- LArTPC** *Liquid Argon Time Projection Chamber*. Neutrino detector employing a time projection chamber immersed in liquid argon. See section 3.2. 27, 32, 33, 41, 49, 51, 79, 80, 83, 84, 118, 199

- MC** *Monte Carlo* methods. Broad class of algorithms relying on random sampling to obtain numerical results. Used to generate simulations of detector datasets. 73–75, 79–83, 85, 91–95, 97–99, 109, 114, 120–127, 129–136, 146–150, 153, 157, 158, 160–162, 164, 166, 168–184, 187–192, 194, 195, 199
- MCS** *Multiple Coulomb Scattering*. Small-angle deflection of the trajectory of a charged particle traversing a medium due to Coulomb scattering from nuclei. The mean angle can be used to determine the momentum of the particle. 137, 138
- MIP** *Minimum Ionising Particle*. Particle whose mean energy loss rate through matter is close to a minimum. Cosmic-ray muons are an example of MIP. See [section 3.1.1](#). 34, 36, 42, 49, 51, 105
- NLLR** *Negative Log-Likelihood Ratio*. Test statistics used to test two competing statistical models based on the ratio of their likelihoods. Used in the context of CLs methods. 193, 194
- NuMI** *Neutrinos at the Main Injector*. Neutrino beam produced with 120 GeV protons from the Main Injector and thus with higher energies than BNB. The NuMI beam is used by the MINOS, MINERvA and NOvA experiments. MicroBooNE is off-axis with respect to NuMI but still receives a measurable flux which is used for several studies. See [section 3.4.1](#). 66, 67
- PDF** *Probability Distribution Function*. Function describing the probability that a random variable would be drawn within a particular range of values. E.g., for counting experiments, the probability distribution function of measuring  $n$  events can be described by a Poisson function. 36, 192–194, 196
- PE** *PhotoElectron*. Electron produced in photomultiplier tubes via photoelectric effect by successfully detected photons. PE is used as a unit of measure of photomultiplier tubes output. 45, 67, 68, 71–75, 77, 78, 118, 143, 148–150, 166, 176, 177
- PID** *Particle IDentification*. Process of using detector information to infer the type of particle. The most common PID is the separation of a particle candidate into a track ( $\mu$ ,  $\pi$ ,  $p$ ) or shower ( $e$ ,  $\gamma$ ) type based on topology.  $dE/dx$  allows for the further separation of  $\mu$  from  $p$  and  $e$  from  $\gamma$ . 80, 81, 110
- PMT** *PhotoMultiplier Tube*. Light detector. MicroBooNE employs 32 of them, placed behind the wire plane. See [section 3.2.4](#). 42–49, 66, 67, 70–75, 80, 92, 93, 99, 103, 109, 111, 118, 125, 126, 128, 147, 148, 177, 179, 198
- POT** *Protons On Target*. The number of protons delivered on the beryllium target, which produce the neutrino beam. Since POT is directly proportional to the number of neutrinos delivered to the detector, it is used as a normalisation



- factor for data samples and as a performance indicator for data volume. See [section 3.4.1](#). 60, 61, 68, 75, 92, 94–96, 99, 103, 119, 121–124, 135, 146, 161, 162, 176, 186, 191, 193, 195–198, 200
- RFG** *Relativistic Fermi Gas*. Model for an ensemble of large number of non-interacting fermions at relativistic speed, used for the nuclear model in neutrino interactions. The Bodek-Ritchie RFG is used in this work. See [section 5.2.5](#). 96
- ROC** *Receiver Operating Characteristic curve*. Graphical curve which illustrates the discriminating power of a binary classifier system. The farthest away from the diagonal, the better the performance of the classifier. 168, 169
- SBN** *Short Baseline Neutrino*. Short-baseline ( $\approx 600$  m) neutrino programme using an array of three LArTPCs (SBND, MicroBooNE, ICARUS) and the Booster Neutrino Beam to perform neutrino oscillation measurements.. 27, 32, 197, 199, 200
- SCE** *Space Charge Effect*. The distortion of the electric field caused by slowly drifting ions produced by high rate crossing cosmic-ray muons. See [section 3.3.4](#). 56, 57, 98, 108, 109, 145, 174, 178, 179, 181–184, 186, 195, 199
- SiPM** *Silicon PhotoMultiplier*. Solid-state single-photon detectors. SiPMs are used in the four light paddles inside the cryostat and in the Cosmic Ray Tagger. See [section 3.2.4](#) and [section 3.2.5](#). 42, 45
- SM** *Standard Model*. Theory describing elementary particles and the electromagnetic, weak, and strong interactions between them. 11, 14–16, 19–21, 26, 30, 64, 65, 69–72, 74, 80, 82–91, 95, 198, 200
- TPB** *TetraPhenyl Butadiene*. Organic fluor used as wavelength shifter. A special panel coated in TPB is placed in front of each PMT unit to convert ultraviolet light emitted by argon in the visible spectrum for higher photon detection efficiency. See [section 3.2.4](#). 43, 44
- TPC** *Time Projection Chamber*. Particle detector using electric fields to perform three-dimensional reconstruction of particle trajectory. See [section 3.2](#). 37–42, 44–49, 56–58, 80, 90, 92, 93, 95, 97–100, 103, 109–111, 128, 131, 141, 144, 152, 170, 179, 186, 200

## References

- [1] Evans, J. et al. *Uncertainties in Atmospheric Muon-Neutrino Fluxes Arising from Cosmic-Ray Primaries*. In: *Phys. Rev. D* 95.2 (2017), p. 023012. arXiv: [1612.03219 \[astro-ph.HE\]](#).
- [2] Fukuda, Y. et al. *Evidence for oscillation of atmospheric neutrinos*. In: *Phys. Rev. Lett.* 81 (1998), pp. 1562–1567. arXiv: [hep-ex/9807003 \[hep-ex\]](#).
- [3] Aharmim, B. et al. *Electron energy spectra, fluxes, and day-night asymmetries of B-8 solar neutrinos from measurements with NaCl dissolved in the heavy-water detector at the Sudbury Neutrino Observatory*. In: *Phys. Rev. C* 72 (2005), p. 055502. arXiv: [nucl-ex/0502021 \[nucl-ex\]](#).
- [4] Gorbunov, D. and Shaposhnikov, M. *How to find neutral leptons of the  $\nu$ MSM?* In: *JHEP* 10 (2007). [Erratum: *JHEP* 11,101(2013)], p. 015. arXiv: [0705.1729 \[hep-ph\]](#).
- [5] Giunti, C. and Kim, C. W. *Fundamentals of Neutrino Physics and Astrophysics*. 2007. ISBN: 9780198508717.
- [6] Ballett, P., Pascoli, S. and Ross-Lonergan, M. *MeV-scale sterile neutrino decays at the Fermilab Short-Baseline Neutrino program*. In: *JHEP* 04 (2017), p. 102. arXiv: [1610.08512 \[hep-ph\]](#).
- [7] Abazajian, K. N. et al. *Light Sterile Neutrinos: A White Paper*. In: (2012). arXiv: [1204.5379 \[hep-ph\]](#).
- [8] Cortina Gil, E. et al. *Search for heavy neutral lepton production in  $K^+$  decays*. In: *Phys. Lett. B* 778 (2018), pp. 137–145. arXiv: [1712.00297 \[hep-ex\]](#).
- [9] Alekhin, S. et al. *A facility to Search for Hidden Particles at the CERN SPS: the SHiP physics case*. In: *Rept. Prog. Phys.* 79.12 (2016), p. 124201. arXiv: [1504.04855 \[hep-ph\]](#).
- [10] Wolfram, S. *Weak Decays*. In: *Nukleonika* 26 (1981), pp. 273–310.
- [11] Shrock, R. E. *New Tests For, and Bounds On, Neutrino Masses and Lepton Mixing*. In: *Phys. Lett.* 96B (1980), pp. 159–164.

- 
- [12] Shrock, R. E. *General Theory of Weak Leptonic and Semileptonic Decays. 1. Leptonic Pseudoscalar Meson Decays, with Associated Tests For, and Bounds on, Neutrino Masses and Lepton Mixing.* In: *Phys. Rev. D* 24 (1981), p. 1232.
  - [13] Aguilar-Arevalo, A. et al. *Evidence for neutrino oscillations from the observation of anti-neutrino(electron) appearance in a anti-neutrino(muon) beam.* In: *Phys. Rev. D* 64 (2001), p. 112007. arXiv: [hep-ex/0104049](#) [[hep-ex](#)].
  - [14] Aguilar-Arevalo, A. A. et al. *A Combined  $\nu_\mu \rightarrow \nu_e$  and  $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$  Oscillation Analysis of the MiniBooNE Excesses.* In: 2012. arXiv: [1207.4809](#) [[hep-ex](#)]. URL: <http://lss.fnal.gov/archive/2012/pub/fermilab-pub-12-394-ad-ppd.pdf>.
  - [15] Aguilar-Arevalo, A. A. et al. *Improved Search for  $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$  Oscillations in the MiniBooNE Experiment.* In: *Phys. Rev. Lett.* 110 (2013), p. 161801. arXiv: [1303.2588](#) [[hep-ex](#)].
  - [16] Aguilar-Arevalo, A. A. et al. *Significant Excess of ElectronLike Events in the Mini-BooNE Short-Baseline Neutrino Experiment.* In: *Phys. Rev. Lett.* 121.22 (2018), p. 221801. arXiv: [1805.12028](#) [[hep-ex](#)].
  - [17] Akhmedov, E. K. and Smirnov, A. *Paradoxes of neutrino oscillations.* In: *Phys. Atom. Nucl.* 72 (2009), pp. 1363–1381. arXiv: [0905.1903](#) [[hep-ph](#)].
  - [18] Atre, A. et al. *The Search for Heavy Majorana Neutrinos.* In: *JHEP* 05 (2009), p. 030. arXiv: [0901.3589](#) [[hep-ph](#)].
  - [19] Ballett, P., Boschi, T. and Pascoli, S. *Heavy Neutral Leptons from low-scale seesaws at the DUNE Near Detector.* In: (2019). arXiv: [1905.00284](#) [[hep-ph](#)].
  - [20] Asaka, T., Eijima, S. and Watanabe, A. *Heavy neutrino search in accelerator-based experiments.* In: *JHEP* 03 (2013), p. 125. arXiv: [1212.1062](#) [[hep-ph](#)].
  - [21] Schael, S. et al. *Precision electroweak measurements on the Z resonance.* In: *Phys. Rept.* 427 (2006), pp. 257–454. arXiv: [hep-ex/0509008](#) [[hep-ex](#)].
  - [22] Tanabashi, M. et al. *Review of Particle Physics.* In: *Phys. Rev. D* 98.3 (2018), p. 030001.
  - [23] Drewes, M. and Garbrecht, B. *Combining experimental and cosmological constraints on heavy neutrinos.* In: *Nucl. Phys. B* 921 (2017), pp. 250–315. arXiv: [1502.00477](#) [[hep-ph](#)].
  - [24] Abela, R. et al. *Search for an Admixture of Heavy Neutrino in Pion Decay.* In: *Phys. Lett.* 105B (1981). [Erratum: *Phys. Lett.* 106B, 513 (1981)], pp. 263–266.
  - [25] Minehart, R. C. et al. *A Search for Admixture of Massive Neutrinos in the Decay  $\pi \rightarrow \mu \nu$ .* In: *Phys. Rev. Lett.* 52 (1984), pp. 804–807.
  - [26] Daum, M. et al. *Search for Admixtures of Massive Neutrinos in the Decay  $\pi^+ \rightarrow \mu^+ \text{Neutrino}$ .* In: *Phys. Rev. D* 36 (1987), p. 2624.

- 
- [27] Aguilar-Arevalo, A. et al. *Search for Heavy Neutrinos in  $\pi \rightarrow \mu \nu$  Decay*. In: (2019). arXiv: [1904.03269 \[hep-ex\]](#).
- [28] Asano, Y. et al. *Search for a Heavy Neutrino Emitted in  $K^+ \rightarrow \mu^+ \nu$  Neutrino Decay*. In: *Phys. Lett.* 104B (1981), pp. 84–88.
- [29] Hayano, R. S. et al. *Heavy Neutrino Search Using  $K(\mu 2)$  Decay*. In: *Phys. Rev. Lett.* 49 (1982), p. 1305.
- [30] Artamonov, A. V. et al. *Search for heavy neutrinos in  $K^+ \rightarrow \mu^+ \nu_H$  decays*. In: *Phys. Rev.* D91.5 (2015). [Erratum: *Phys. Rev.* D91, no. 5, 059903 (2015)], p. 052001. arXiv: [1411.3963 \[hep-ex\]](#).
- [31] Bernardi, G. et al. *Search for Neutrino Decay*. In: *Phys. Lett.* 166B (1986), pp. 479–483.
- [32] Bernardi, G. et al. *Further Limits on Heavy Neutrino Couplings*. In: *Phys. Lett.* B203 (1988), pp. 332–334.
- [33] Abe, K. et al. *Search for heavy neutrinos with the T2K near detector ND280*. In: (2019). arXiv: [1902.07598 \[hep-ex\]](#).
- [34] Vaitaitis, A. et al. *Search for neutral heavy leptons in a high-energy neutrino beam*. In: *Phys. Rev. Lett.* 83 (1999), pp. 4943–4946. arXiv: [hep-ex/9908011 \[hep-ex\]](#).
- [35] Acciarri, R. et al. *Design and Construction of the MicroBooNE Detector*. In: *JINST* 12.02 (2017), P02017. arXiv: [1612.05824 \[physics.ins-det\]](#).
- [36] Antonello, M. et al. *A Proposal for a Three Detector Short-Baseline Neutrino Oscillation Program in the Fermilab Booster Neutrino Beam*. In: (2015). arXiv: [1503.01520 \[physics.ins-det\]](#).
- [37] Acciarri, R. et al. *Long-Baseline Neutrino Facility (LBNF) and Deep Underground Neutrino Experiment (DUNE)*. In: (2016). arXiv: [1601.05471 \[physics.ins-det\]](#).
- [38] Acciarri, R. et al. *Effects of Nitrogen and Oxygen contamination in liquid Argon*. In: *Nucl. Phys. Proc. Suppl.* 197 (2009). [*Nucl. Instrum. Meth.* 607, 169 (2009)], pp. 70–73.
- [39] Joshi, J. and Qian, X. *Signal Processing in the MicroBooNE LArTPC*. In: *Proceedings, Meeting of the APS Division of Particles and Fields (DPF 2015): Ann Arbor, Michigan, USA, 4-8 Aug 2015*. 2015. arXiv: [1511.00317 \[physics.ins-det\]](#).
- [40] Katori, T. *The MicroBooNE light collection system*. In: *JINST* 8 (2013), p. C10011. arXiv: [1307.5256 \[physics.ins-det\]](#).
- [41] Jones, B. J. P. “Sterile Neutrinos in Cold Climates”. PhD thesis. MIT, 2015. URL: <http://lss.fnal.gov/archive/thesis/2000/fermilab-thesis-2015-17.pdf>.

- 
- [42] Adams, C. et al. *Design and construction of the MicroBooNE Cosmic Ray Tagger system*. In: *JINST* 14.04 (2019), P04004. arXiv: [1901.02862 \[physics.ins-det\]](#).
- [43] Thomas, J. and Imel, D. A. *Recombination of electron-ion pairs in liquid argon and liquid xenon*. In: *Phys. Rev.* A36 (1987), pp. 614–616.
- [44] Acciarri, R. et al. *A Study of Electron Recombination Using Highly Ionizing Particles in the ArgoNeuT Liquid Argon TPC*. In: *JINST* 8 (2013), P08005. arXiv: [1306.1712 \[physics.ins-det\]](#).
- [45] Acciarri, R. et al. *Oxygen contamination in liquid Argon: Combined effects on ionization electron charge and scintillation light*. In: *JINST* 5 (2010), P05003. arXiv: [0804.1222 \[nucl-ex\]](#).
- [46] The MicroBooNE Collaboration. *A Measurement of the Attenuation of Drifting Electrons in the MicroBooNE LArTPC*. MicroBooNE Public Note: MICROBOONE-NOTE-1026-PUB. URL: <https://microboone.fnal.gov/wp-content/uploads/MICROBOONE-NOTE-1026-PUB.pdf>.
- [47] Jones, B. J. P. et al. *A Measurement of the Absorption of Liquid Argon Scintillation Light by Dissolved Nitrogen at the Part-Per-Million Level*. In: *JINST* 8 (2013). [Erratum: *JINST* 8, E09001 (2013)], P07011. arXiv: [1306.4605 \[physics.ins-det\]](#).
- [48] Mooney, M. *The MicroBooNE Experiment and the Impact of Space Charge Effects*. In: *Proceedings, Meeting of the APS Division of Particles and Fields (DPF 2015): Ann Arbor, Michigan, USA, 4-8 Aug 2015*. 2015. arXiv: [1511.01563 \[physics.ins-det\]](#).
- [49] Barak, D., Harrison, B. and Watts, A. *LINAC Rookie Book: RIL*. Fermilab Accelerator Division, Operations Department, October 2015. URL: [http://operations.fnal.gov/rookie\\_books/RIL.pdf](http://operations.fnal.gov/rookie_books/RIL.pdf).
- [50] Barak, D., Harrison, B. and Watts, A. *Concepts Rookie Book*. Fermilab Accelerator Division, Operations Department, December 2013. URL: [http://operations.fnal.gov/rookie\\_books/concepts.pdf](http://operations.fnal.gov/rookie_books/concepts.pdf).
- [51] Pavlovic, Z. *Booster Neutrino Beamline Upgrades*. Neutrino Short Baseline Meeting, September 2014. URL: [https://indico.cern.ch/event/335564/contributions/1728708/attachments/655673/901518/sbn\\_bnb.pdf](https://indico.cern.ch/event/335564/contributions/1728708/attachments/655673/901518/sbn_bnb.pdf).
- [52] Tsai, Y.-T. *MicroBooNE and its Cross Section Measurement*. In: *Proceedings, Prospects in Neutrino Physics (NuPhys2016): London, UK, December 12-14, 2016*. 2017. arXiv: [1705.07800 \[hep-ex\]](#). URL: <http://lss.fnal.gov/archive/2017/conf/fermilab-conf-17-127-nd.pdf>.
- [53] Aguilar-Arevalo, A. A. et al. *The Neutrino Flux prediction at MiniBooNE*. In: *Phys. Rev.* D79 (2009), p. 072002. arXiv: [0806.1449 \[hep-ex\]](#).
- [54] Vigorito, C. *SNEWS: The supernova early warning system*. In: *J. Phys. Conf. Ser.* 309 (2011), p. 012026.

- 
- [55] An, R. et al. *Software Trigger Documentation*. MicroBooNE Internal Documentation: DocDB 5205. URL: <https://microboone-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=5205>.
- [56] Church, E. D. *LArSoft: A Software Package for Liquid Argon Time Projection Drift Chambers*. In: (2013). arXiv: [1311.6774](https://arxiv.org/abs/1311.6774) [physics.ins-det].
- [57] Green, C. et al. *The Art Framework*. In: *J. Phys. Conf. Ser.* 396 (2012), p. 022020.
- [58] Andreopoulos, C. et al. *The GENIE Neutrino Monte Carlo Generator: Physics and User Manual*. In: (2015). arXiv: [1510.05494](https://arxiv.org/abs/1510.05494) [hep-ph].
- [59] Heck, D. et al. *CORSIKA: A Monte Carlo code to simulate extensive air showers*. In: (1998).
- [60] Agostinelli, S. et al. *GEANT4: A Simulation toolkit*. In: *Nucl. Instrum. Meth.* A506 (2003), pp. 250–303.
- [61] Allison, J. et al. *Recent developments in Geant4*. In: *Nucl. Instrum. Meth.* A835 (2016), pp. 186–225.
- [62] Marshall, J. S. and Thomson, M. A. *The Pandora Software Development Kit for Pattern Recognition*. In: *Eur. Phys. J.* C75.9 (2015), p. 439. arXiv: [1506.05348](https://arxiv.org/abs/1506.05348) [physics.data-an].
- [63] Acciarri, R. et al. *The Pandora multi-algorithm approach to automated pattern recognition of cosmic-ray muon and neutrino events in the MicroBooNE detector*. In: *Eur. Phys. J.* C78.1 (2018), p. 82. arXiv: [1708.03135](https://arxiv.org/abs/1708.03135) [hep-ex].
- [64] Brun, R. and Rademakers, F. *ROOT: An object oriented data analysis framework*. In: *Nucl. Instrum. Meth.* A389 (1997), pp. 81–86.
- [65] Ross-Lonergan, M. and Ballett, P. *InFlight Generator*. Private Communications.
- [66] Goodwin, O. *Private communication*. 2019.
- [67] The MicroBooNE Collaboration. *Cosmic Shielding Studies at MicroBooNE*. MicroBooNE Public Note: MICROBOONE-NOTE-1005-PUB. URL: <https://microboone.fnal.gov/wp-content/uploads/MICROBOONE-NOTE-1005-PUB.pdf>.
- [68] Bodek, A. and Ritchie, J. L. *Fermi Motion Effects in Deep Inelastic Lepton Scattering from Nuclear Targets*. In: *Phys. Rev.* D23 (1981), p. 1070.
- [69] Acciarri, R. et al. *Noise Characterization and Filtering in the MicroBooNE Liquid Argon TPC*. In: *JINST* 12.08 (2017), P08003. arXiv: [1705.07341](https://arxiv.org/abs/1705.07341) [physics.ins-det].
- [70] Pavlovic, Z. *POT Counting and Beam Quality Cuts*. MicroBooNE Internal Documentation: DocDB 11078. URL: <https://microboone-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=11078>.



- 
- [71] Guzowski, P. *Good Runs Guide*. MicroBooNE Internal Documentation: DocDB 13933. URL: <https://microboone-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=13933>.
- [72] Adams, C. et al. *Calibration of the charge and energy response of the MicroBooNE liquid argon time projection chamber using muons and protons*. In: (2019). arXiv: [1907.11736](https://arxiv.org/abs/1907.11736) [physics.ins-det].
- [73] Adams, C. et al. *Ionization electron signal processing in single phase LArTPCs. Part I. Algorithm Description and quantitative evaluation with MicroBooNE simulation*. In: JINST 13.07 (2018), P07006. arXiv: [1802.08709](https://arxiv.org/abs/1802.08709) [physics.ins-det].
- [74] Adams, C. et al. *Ionization electron signal processing in single phase LArTPCs. Part II. Data/simulation comparison and performance in MicroBooNE*. In: JINST 13.07 (2018), P07007. arXiv: [1804.02583](https://arxiv.org/abs/1804.02583) [physics.ins-det].
- [75] Groom, D. E., Mokhov, N. V. and Striganov, S. I. *Muon stopping power and range tables 10-MeV to 100-TeV*. In: *Atom. Data Nucl. Data Tabl.* 78 (2001), pp. 183–356.
- [76] Seltzer, S. *Stopping-Powers and Range Tables for Electrons, Protons, and Helium Ions, NIST Standard Reference Database 124*. 1993. URL: <http://www.nist.gov/pml/data/star/index.cfm>.
- [77] Coadou, Y. *Boosted Decision Trees and Applications*. In: *EPJ Web Conf.* 55 (2013), p. 02004.
- [78] Chen, T. and Guestrin, C. *XGBoost: A Scalable Tree Boosting System*. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. San Francisco, California, USA: ACM, 2016, pp. 785–794. ISBN: 978-1-4503-4232-2. URL: <http://doi.acm.org/10.1145/2939672.2939785>.
- [79] The MicroBooNE Collaboration. *First Muon-Neutrino Charged-Current Inclusive Differential Cross Section Measurement for MicroBooNE Run 1 Data*. MicroBooNE Public Note: MICROBOONE-NOTE-1045-PUB. URL: <https://microboone.fnal.gov/wp-content/uploads/MICROBOONE-NOTE-1045-PUB.pdf>.
- [80] Alvarez-Ruso, L. et al. *NuSTEC White Paper: Status and challenges of neutrino–nucleus scattering*. In: *Prog. Part. Nucl. Phys.* 100 (2018), pp. 1–68. arXiv: [1706.03621](https://arxiv.org/abs/1706.03621) [hep-ph].
- [81] The MicroBooNE Collaboration. *Selection of  $\nu_\mu$  charged–current induced interactions with  $N>0$  protons and performance of events with  $N=2$  protons in the final state in the MicroBooNE detector from the BNB*. MicroBooNE Public Note: MICROBOONE-NOTE-1056-PUB. URL: <https://microboone.fnal.gov/wp-content/uploads/MICROBOONE-NOTE-1056-PUB.pdf>.

## REFERENCES

---

- [82] Zennamo, J. and Pavlović, Ž. *MicroBooNE Flux and Flux Uncertainties*. MicroBooNE Internal Documentation: DocDB 8622. URL: <https://microboone-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=8622>.
- [83] The MicroBooNE Collaboration. *Booster Neutrino Flux Prediction at MicroBooNE*. MicroBooNE Public Note: MICROBOONE-NOTE-1031-PUB. URL: <https://microboone.fnal.gov/wp-content/uploads/MICROBOONE-NOTE-1031-PUB.pdf>.
- [84] Garwood, F. *Fiducial limits for the Poisson distribution*. In: *Biometrika* 28.3/4 (1936), pp. 437–442.
- [85] Fisher, W. *Collie: A Confidence Level Limit Evaluator*. D0-NOTE-5595. 2009.
- [86] Fisher, W. *Systematics and limit calculations*. FERMILAB-TM-2386-E, D0-NOTE-5309. 2006.
- [87] Fishchler, M. *Optimizing the Performance and Structure of the D0 Collie Confidence Limit Evaluator*. FERMILAB-TM-2468-CD. 2010.