MANCHESTER 1824
The University of Manchester

Search for millicharged particles using the MicroBooNE detector

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

2024

Luciano G. Arellano
Department of Physics and Astronomy
School of Natural Sciences

Contents

Abstract

We search for millicharged particles using a single-electron scatter signature in the MicroBooNE detector. Using 1.07×10^{21} protons-on-target of data from the Neutrinos at the Main Injector (NuMI) beam, we observe no significant excess compared to the expected backgrounds. We set limits on the production of millicharged particles from the decay of light neutral mesons, covering a mass range of 15 to 400 MeV, and electric charges between 5.17×10^{-4} and 1.88×10^{-2} times the elementary charge. These results improve upon previous constraints from the ArgoNeuT liquid argon-time projection chamber (LArTPC), though they do not represent the most stringent limits to date. This is the first search for millicharged particles using a single-scatter signature in a LArTPC. The analysis targets electron recoils with energies as low as 20 MeV, where the Wire-Cell reconstruction framework yields only a 1 % selection efficiency in its default configuration. We also present the expected sensitivity for a complementary search using a multiple *blip* scatter signature, aligned with the NuMI beam target, with energies as low as 0.5 MeV, similar to the search performed by ArgoNeuT. This blip-based channel shows lower sensitivity than the single-scatter search for electric charge ranges between 3.75×10^{-3} and 1.13×10^{-2}. Only at the highest mass point of 400 MeV does the blip-based search surpass the sensitivity of the Wire-Cell based single-scatter channel.

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning

Copyright statement

Acknowledgements

I would like to thank, first of all, my supervisor Justin Evans, who has guided me throughout my PhD. Despite being one of the busiest people I have ever known, he always — without exceptions — made time to meet with me, answer all my questions, and provide guidance on both my analysis and career, all the while being jolly and polite, constantly offering words of encouragement.

I am deeply grateful to Pawel Guzowski, without whom this analysis would have been impossible. He contributed directly as the developer of the millicharged particle event generator and the first versions of the analysis, and provided guidance to me and others as a postdoc at Manchester and convener of the MicroBooNE APE group. He could have done this work better than me in one month — my only goal is to become 10 % of the physicist that he is.

I would also like to thank Dr. Yu-Dai Tsai, who originally proposed the idea of using higher-energy single-scatters for this search.

I am grateful to my co-supervisor, Stefan Söldner-Rembold, for his mentorship and for instilling a goal-oriented approach to science — and also for his (sometimes misunderstood) sense of humour. Together, Justin and Stefan made for an outstanding supervisory team, and fostered a great environment within the neutrino group. I would also like to thank Elena Gramellini, who helped me greatly with my work on SBND, and with navigating the challenges of being a PhD student. Thanks as well to Nicola McConkey for her help with the QA of DUNE boards (which, though not included in this thesis, was very valuable), her suggestions on the SBND TPC commissioning group meetings, and her positive words of encouragement. Thanks to Anyssa Navrer-Agasson for her support as APE group convener in MicroBooNE and for her constant willingness to help — not just me with my analysis, but everyone else's too.

To David Marsden and Luis Mora Lepin, thank you for your friendship, the unforgettable moments, and so many incredible scenes. I am sure we also talked about physics at some point, but honestly, who could remember. Thanks to the Manchester neutrino group, and the particle physics group in general. I am a little sad knowing that I will probably never find a better working environment than the one I had at Manchester, but I am also glad that I got to experience it.

I would like to thank the brilliant people of the MicroBooNE collaboration. I am happy to have been even a small part of this experiment, so full of inspiring people. In particular, I would like to thank the APE conveners, past and present — Pawel Guzowski, Ivan Lepetic, Daisy Kalra, and Anyssa Navrer-Agasson — all of whom supported me in my search for particles that had very little chance of existing. I would also like to thank Maya Wospakrik and Ryan Dorrill, who guided me through my electron lifetime task, and Will Foreman, who developed the MicroBooNE *blip* reconstruction tools and taught

me how to use them. Thanks also to blip pioneers, Ivan Lepetic and Avinay Bhat.

From SBND, I would like to thank Michelle Stancari. Her thorough and complete understanding of the detector and all of its components was truly awe-inspiring, and her guidance was invaluable. I am also grateful to the TPC commissioning conveners, Lauren Yates and Nicola McConkey, for their help. Thanks also to my friends from Fermilab (and one memorable trip to the UK), Norman Martinez and Diego Andrade (also a fellow blip lover), and my flatmates Harry Scott and Robert Darby.

I would like to acknowledge the support of Agencia Nacional de Investigación y Desarrollo (ANID) which provided funding for my PhD through the Becas Chile program. I also acknowledge support from the Universities Research Association (URA), which made my one-year stay at Fermilab possible.

Finally, and above all, I want to thank my family. Everything that I am and that I have, I owe to them (by which I mean only the positives). In particular, my niece Flavia is in my thoughts; I am looking forward to finally meeting her in Chile in the next few months, now that this work is done.

Mi padre siendo tan pobre dejó
una herencia fastuosa.

(Inti-Illimani, "Samba Lando")

Part I.

Introduction

1. Introduction

The Standard Model (SM) has been remarkably successful in describing the fundamental building blocks of the universe — the elementary particles and their interactions — yet it is known to be incomplete. It lacks a mechanism by which neutrinos can acquire mass, fails to explain the matter-antimatter asymmetry of the universe, and does not account for dark matter.

Neutrinos, in particular, are known to exhibit beyond the Standard Model (BSM) properties, and their study may provide key insights into new physics. In the SM, neutrinos appear only as left-handed particles (and antineutrinos as right-handed), which is consistent with the SM assumption that they are massless. However, the observation of neutrino oscillations implies that neutrinos must have a mass, albeit very small. The Higgs mechanism, which gives mass to other fermions and bosons, requires both left- and right-handed components, but right-handed neutrinos are absent in the SM.

Experimentally, neutrinos are difficult to detect. They interact only via the weak force — and gravity, which is far too feeble to be relevant in particle physics experiments. The "intensity frontier" strategy addresses this challenge by using massive detectors with sophisticated reconstruction capabilities, placed in close proximity to intense neutrino sources. These same features, however, also enable intensity frontier neutrino experiments to search for other BSM phenomena.

The MicroBooNE detector is a liquid-argon time projection chamber (LArTPC) experiment located at Fermilab. During its operation between 2015 and 2020, it collected a vast dataset of neutrino–argon interactions from two different neutrino beams. The LArTPC technology provides excellent spatial resolution and calorimetric reconstruction capabilities, enabling the use of advanced reconstruction methods and the exploration of complex topologies. This has previously allowed for a wide range of new physics searches, resulting in numerous published results in exotic models such as heavy neutral leptons, Higgs portal scalars, and light dark matter.

This thesis presents a search for millicharged particles using the MicroBooNE detector. Millicharged particles are hypothetical new particles that violate the electric charge quantisation empirically observed in the SM. They can be produced in the decays of neutral mesons, which are abundantly generated in the same accelerator collisions that produce neutrino beams. Millicharged particles would then travel largely unimpeded towards MicroBooNE, where they could interact feebly with electrons in the detector's active volume, via the electromagnetic interaction.

In chapter 2, we introduce millicharged particles, outlining their theoretical motivation, describing the two different topologies used in our search, and providing a phenomenological overview of their production in the NuMI neutrino beam and subsequent interaction with the MicroBooNE detector.

Chapter 3 provides a description of the MicroBooNE experiment, including an overview of LArTPC technology, a summary of the neutrino beams at Fermilab, and a brief discussion of SBND, a current-generation LArTPC experiment, which is now operational.

Chapter 4 takes a deeper look into electron lifetime, a critical parameter related to argon purity in LArTPC operations. This chapter presents work the author performed to validate and calibrate the electron lifetime in Runs 4 and 5 in MicroBooNE, respectively, as well as a pilot study conducted during SBND commissioning, which the author performed. As a more self-contained chapter, it is somewhat distinct from the central theme of the thesis — the search for millicharged particles in MicroBooNE.

Chapter 5 describes the simulation and reconstruction framework used in the search for millicharged particles, including the treatment of backgrounds and the data samples employed. Chapter 6 details the event selection criteria applied to both simulation and data, as well as the Boosted Decision Trees (BDTs) trained to classify signal and background events.

Systematic uncertainties are discussed in chapter 7, and the final results of the analysis are presented in chapter 8. A summary of the findings and their implications is given in chapter 9.

The author was responsible for running the millicharged particle Monte Carlo event generator, originally developed by Pawel Guzowski, within the *LArSoft* framework. This included validating the implementation of the underlying physics equations, generating the signal simulation samples used in this work, and processing them through the *LArSoft* reconstruction chain using existing workflows developed by the MicroBooNE collaboration. The author then used the outputs to train two BDTs, developing training code based on earlier versions written by Pawel Guzowski. Neutrino-interaction and beam-off background samples used in BDT training were produced by the MicroBooNE collaboration. The author performed the assessment of all relevant sources of systematic uncertainty, which required generating additional signal simulation samples, as well as the use of existing background simulations produced by the MicroBooNE collaboration. The author performed limit setting using the outputs of the BDTs, with original code developed to interface with the *pyhf* framework.

The author also led the validation of the electron lifetime calibration in Run 4, adapting existing macros for data processing and plotting, to reproduce previous results. For Run 5, the author performed the calibration, adapting and modifying the existing macros for their application to new samples, developing new analysis tools as needed, thus generating new results. In the SBND pilot study, the author developed code to match TPC and CRT tracks using the existing commissioning *ntuple* format and implemented tools to extract the electron lifetime. Additionally, the author generated Monte Carlo samples of muon tracks with electron lifetimes different from 10 ms.

2. Millicharged particles

In this chapter, we present the concept of a millicharged particle (mCP), a new Beyond the Standard Model particle whose electric charge is not an integer multiple of the elementary charge e. In section 2.1 we present the theoretical motivation behind them, a mechanism via which they could arise, and the EDGES anomaly which sparked a resurgence of this idea in recent years. In section 2.2 we outline the basic search strategy presented in this thesis, in which we will search for mCPs in the MicroBooNE experiment using a discrete number of electrons scatters as a signature. We then describe in some detail the phenomenology of mCP production and interaction with our detector in sections 2.3 and 2.4 respectively. We end the chapter with a discussion of the current experimental limits in section 2.5.

2.1. Motivation

2.1.1. Charge quantisation

Shortly after his discovery of the electron using cathode rays in 1897 [1], J.J. Thomson measured the charge of gas ionised by ultraviolet light in 1899 [2] and concluded that the magnitude of its negative charge was about 6×10^{-10} esu, and was "equal to the positive charge of the hydrogen atom", meaning the proton. Later, between 1910 and 1913 [3, 4] a series of experiments by Millikan used oil droplets under the effects of an electric field and gravity to establish that all electric charges are integer multiples of a fundamental unit, the elementary charge e. He measured its value as $(4.774 \pm 0.009) \times 10^{-10}$ esu, only around a 0.6 % difference from the current best measurements. Current experimental limits on the deviation from unity of the proton-electron charge ratio are $\epsilon_{\mathrm{p-e}} < 1 \times 10^{-21}$ [5], and the idea of charge quantisation has been solidified; since 2019, the elementary charge is defined in the SI as the exact value $1.602\,176\,634 \times 10^{-19}$ C [6].

In the Standard Model (SM) however, electric charge quantisation is only an empirically observed symmetry. As a contrast, electric charge conservation is a fundamental requirement for the formulation of a locally invariant $U(1)$ gauge theory of quantum electrodynamics (QED). Gauge symmetries are at the heart of the SM, and for example in QED they are used to reach a covariant formulation that is consistent with the massless photon observed in nature [7]. There is no such fundamental requirement or theoretical motivation for charge quantisation in the SM, where observable elementary particles have a charge of either $\pm 1e$ (charged leptons and W bosons), or zero (photons, gluons, Z boson, the Higgs boson, or neutrinos). Quarks have a fractional electric charge of $+2/3\,e$ or $-1/3\,e$, however colour confinement forbids them from being directly observed; they only manifest as hadrons whose charge only takes integer values, such as

the Δ^{++} baryon, which has a charge of $+2\,e$, to give a non $\pm 1\,e$ example. Furthermore, the electric charge of quarks is still quantised, only at a lower value which can be made integer by rescaling all electric charges by a factor of 3.

In Grand Unified Theories, the reason that the proton has the same charge as the electron can be explained due to a yet unobserved decay of the proton, for example $p \to e^+ + \pi^0$ [8]; still this does not necessarily exclude the existence of additional elementary particles with an arbitrary charge that is a non-integer multiple of e. In 1931, Dirac postulated the existence of magnetic monopoles as a new particle with a magnetic charge [9], which would lead to a full symmetrisation of Maxwell's equations. In that same publication, Dirac showed that the existence of a single magnetic monopole in the Universe leads to electric charge quantisation. However, magnetic monopoles have not yet been observed experimentally [10], and both charge quantisation violation and Dirac magnetic monopoles persist as two Beyond the Standard Model scenarios worth exploring, yet fundamentally opposing ideas, the existence of one negating the other.

2.1.2. A minimal model for millicharged particles

One example of a minimal extension to the SM which results in a particle with a non-integer multiple of the elementary charge is described in Ref. [11]. In this model, there are two symmetries, $U_1(1)$ and $U_2(1)$, in which fermions f_1, f_2, f_{12} carry different charges 1, 2, or both (there could also be fermions f_0 which do not have either charge). Depending on the charge they carry, fermions couple to the vector fields A_1^μ, A_2^μ, which could then mix, producing off-diagonal terms. The photon can be associated to an effective field $A_1^{'\mu}$ which couples to both f_1 and f_2 under different ratios depending on the "bare" couplings and the masses of f_{12} mediators. Meanwhile, a new "paraphoton" $A_2^{'\mu}$ does not need to couple to the SM fermions f_1, therefore the proton can have the same charge as the electron, and only new fermions f_2 acquire an electric charge which is "shifted" multiplicatively by a factor ϵ.

These new fermions are usually referred to as minicharged or millicharged particles. We will use the latter, and often use the abbreviation mCP. The electric charge of these particles Q_χ is related to the elementary charge e by

$$\epsilon = \frac{Q_\chi}{e}. \tag{2.1}$$

As implied by the name millicharge, it is assumed that the charge Q_χ is much smaller than e ($\epsilon < 0.001$), which would result in a reduced electric coupling, leading to feeble interactions which have escaped detection. The adimensional coupling ϵ is a free parameter of the theory, as well as the non-zero mass of the mCP, m_χ.

2.1.3. The EDGES anomaly and beam-dump searches

In 2018, the Experiment to Detect the Global Epoch of Reionization Signature (EDGES) published the results of its search for the alteration of the 21 cm hyperfine line of hydrogen due to the formation of stars in the early universe [12]. They found an absorption profile

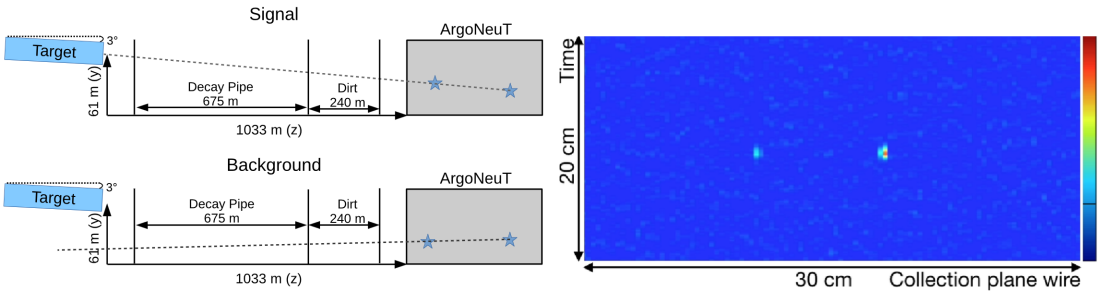


Figure 2.1.: Left: Outline of ArgoNeuT's mCP search strategy. Signal is characterised by multiple scatters aligned with the beam target, whereas background is not aligned. ArgoNeuT observed one signal-like event (shown in the right figure), which is consistent with background expectations. From Ref. [17].

consistent with the expected spectral distortion on the cosmic microwave background, but its amplitude was twice the expected value, indicating that the temperature of the early universe could have been half of its expected value. It was quickly suggested that cooling could be explained by interactions of the gas with Dark Matter, and that a 1 % component of Dark Matter being millicharged would be enough to explain this anomaly [13].

Beam dump experiments emerged as candidates for searching mCPs produced from the decays of mesons [14]. As mCPs carry a small electric charge, they could electrically interact in a detector resulting in a small amount of energy deposited per unit length. Ref. [14] proposes a new scintillator detector located in either Fermilab or CERN to search for them, however it was quickly realised that existing neutrino experiments, such as those located in Fermilab, could search for mCPs [14, 15], as the same low-energy detection capabilities and high spatial resolution that allow them to detect the feeble neutrino interactions would also enable them to search for mCPs.

The ArgoNeuT liquid argon time projection chamber (LArTPC) demonstrated pioneering MeV-scale reconstruction capabilities with the technology [16], and shortly after published world leading limits on the production of millicharged particles [17]. ArgoNeuT's search strategy is outlined in Ref. [18] and shown in figure 2.1. Millicharged particles would travel in a straight line from the beam target, where they are produced, towards the detector, where they scatter off electrons in the argon producing only a few discrete energy depositions above the detection threshold. The line joining any two of these small energy depositions, which are usually referred to as *blips* [19], would have an arbitrary orientation if due to background, but would point back towards the target if due to mCPs.

2.2. Search channels in MicroBooNE

In this thesis, we will present a search for millicharged particles using the MicroBooNE LArTPC neutrino detector, which sits downstream of the same NuMI neutrino beam

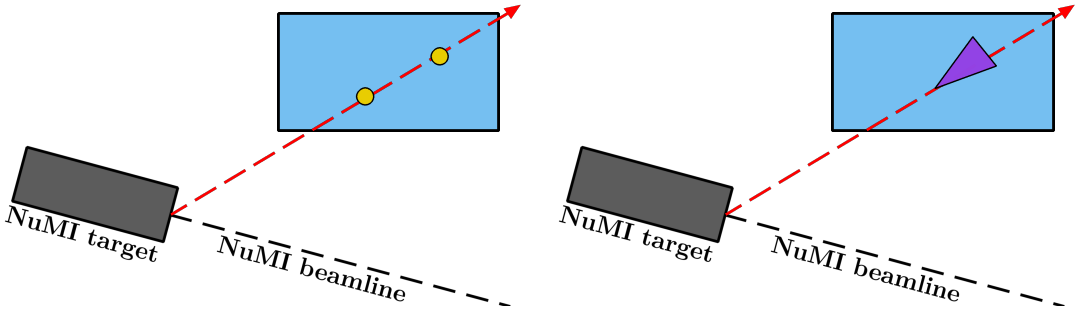


Figure 2.2.: Diagram of the two distinct topologies we will search for in this thesis (not to scale). The light blue square represents the MicroBooNE detector, and the dashed red arrow the passage of a millicharged particle originating from the NuMI beam. Left: multiple low-energy *blips* (yellow circles) aligned with the beam target. Right: a single high-energy electron scatter will form a low-energy shower-like deposition (purple cone).

as ArgoNeuT, albeit on a highly off-axis angle. Like ArgoNeuT, we will also look for low-energy electron scatters that can be geometrically traced back to the beam target. For this, we explore two different topologies, which are sketched in figure 2.2.

Like ArgoNeuT, we will search for multiple $\mathcal{O}(1)$ MeV scale *blips* aligned with the beam target. This channel targets the lowest possible energy depositions that a LArTPC is able to detect. Blips by themselves are point-like and have no clear directionality, hence a multi-blip search is necessary. Additionally, we will search for single electron scatters which are energetic enough to produce a deposition with a clear directionality. These $\mathcal{O}(10)$ MeV scale scatters can be reconstructed as a low energy track or shower-like topology. This will push the low energy range that can be reconstructed with these tools, which are targeted at reconstructing the $\mathcal{O}(100)$ MeV products of neutrino interactions, however a single interaction is more likely than multiple.

2.3. Millicharged particle production

Being electrically charged, mCPs could be produced by any photon-mediated processes. We will not make any further assumptions other than their feeble electric charge, and being massive, so we will look for mCP pair creation (from virtual photons). We can relate the decay branching ratios of known processes which create pairs of leptons to that of mCP creation, as illustrated in figure 2.3. The only other consideration we must make is that of the phase-space of the decay.

In the Standard Model, the coupling strength of the electromagnetic field to particles, $\sqrt{\alpha}$, is related to the fundamental charge as

$$\alpha = \frac{e^2}{4\pi\varepsilon_0\hbar c}, \tag{2.2}$$

where in this equation, e is the elementary charge, ε_0 the vacuum permittivity (not to

be confused with the millicharge ϵ), $\hbar$ the reduced Planck constant, and c the speed of light. By considering a differently charged particle $Q_\chi = \epsilon e$, α will be scaled by a factor of ϵ^2 which can be factored out, as illustrated in the vertex of the coupling in figure 2.3.

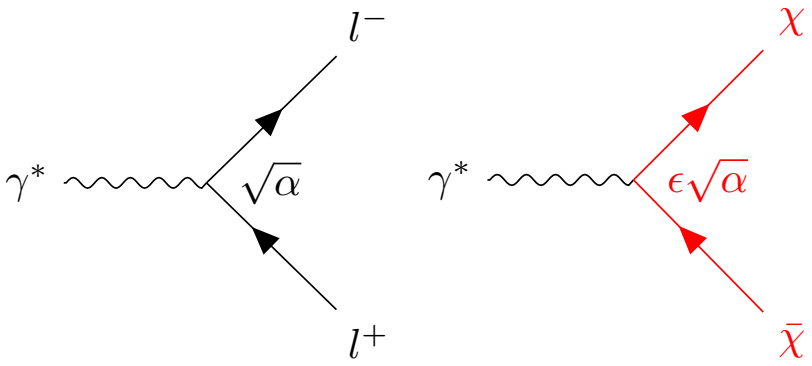


Figure 2.3.: Under our minimal assumptions, millicharged particle pairs ($\chi\bar{\chi}$, in red) can be created from any process where a virtual photon produces a lepton pair (l^+l^-). In most cases, the processes we consider will be e^+e^- pairs. The coupling constant will be scaled by a factor of ϵ.

At the lowest energies (0–400 MeV), we can obtain virtual photons from the decay of neutral mesons such as π^0 and η. For higher energies their production would be dominated by the decay of bound states such as J/ψ or Υ particles, the Drell-Yan process, and bremsstrahlung. In this thesis, we will only search for mCPs from the the decay of light neutral mesons, as we have a well-validated prediction for their flux from proton-on-target collisions in the NuMI neutrino beam. In the following subsections we will study some of the possibilities for two- and three-body decay of light neutral mesons shown in table 2.1.

Table 2.1.: Light neutral mesons which we will consider for mCP production

Meson	Mass (MeV)
π^0	134.98
η	547.86
η'	957.78
ρ	775.26

2.3.1. Production from two-body decay of vector mesons

Neutral mesons of spin 1 (vector mesons) can decay directly to an e^+e^- pair, and these can be related to mCP pair creation by [20]

$$\mathrm{Br}\left(V \to \chi\bar{\chi}\right) = \epsilon^2 \cdot \mathrm{Br}\left(V \to e^+e^-\right)\left(1 + 2\frac{m_\chi^2}{M_V^2}\right)\sqrt{1 - 4\frac{m_\chi^2}{M_V^2}}. \tag{2.3}$$

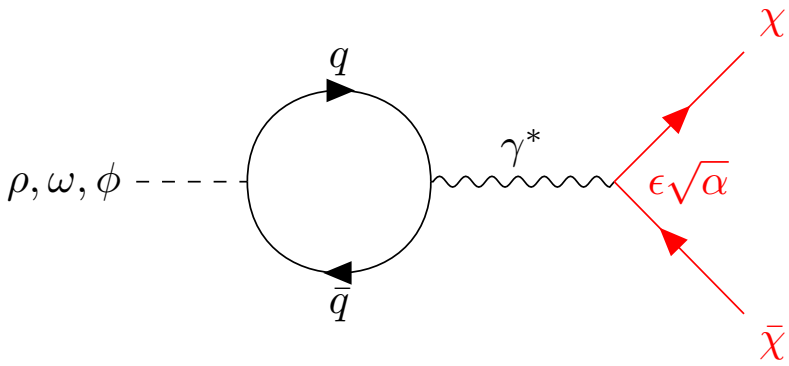


Figure 2.4.: Two-body decay of spin 1 light neutral mesons. At tree level they can decay via a quark loop into an e^+e^- pair, which we can replace with a pair of millicharged particles and reweight the branching ratio. The replacement works for every higher order diagram as well.

Here, V is the vector meson, M_V is its mass, and m_χ is the mass of the mCP. In our range of energies, V could be a ρ, ω or ϕ meson, however we do not have a well-validated flux simulation for ω or ϕ mesons, so we only consider decays from ρ mesons. This will limit our sensitivity to mCPs above 200 MeV, but not in a way that will change the position of our results relative to the that of other experiments with higher sensitivity. The branching ratio of ρ mesons to e^+e^- is shown in table 2.2, which is represented by the $\mathrm{Br}\,(V \to e^+e^-)$ term in equation (2.3). This branching ratio is modified by a "phase-space suppression" factor (term in parentheses and square root) given by the ratios between the masses of the meson and the mCP. The ϵ^2 factor comes from the mCP pair creation vertex and is a modification to the usual α coupling, which is contained within the branching ratio. The tree level Feynman diagram of this decay is shown in figure 2.4.

Table 2.2.: Known branching ratios to which the two- and three-body production of millicharged particles can be related.

Two-body decays	Branching ratio
$\rho \to e^+e^-$	4.72×10^{-5}
Three-body decays	**Branching ratio**
$\pi^0 \to \gamma\gamma$	0.988
$\eta \to \gamma\gamma$	0.394
$\eta' \to \gamma\gamma$	0.0221

2.3.2. Production from three-body decay

For three-body decay, in the energy ranges we consider a (virtual) photon can be produced from the decay of pseudoscalar neutral mesons (π^0, η, η'). We will relate the

branching ratios leading to mCP pair creation to the branching ratios shown in table 2.2. In the three-body case, the phase-space suppression factor is more involved as we must integrate over the possible invariant masses q^2 of the virtual photon which will give rise to the mCP pair. We have [20]

$$\begin{aligned}\mathrm{Br}\left(P \to \gamma\chi\bar{\chi}\right) =& \epsilon^2 \cdot \mathrm{Br}\left(P \to \gamma\gamma\right) \frac{2\alpha}{3\pi} \int_{4m_\chi^2}^{M_P^2} \left[\frac{dq^2}{q^2}\left(1 + 2\frac{m_\chi^2}{q^2}\right)\right. \\ & \left.\left(1 - 4\frac{m_\chi^2}{q^2}\right)^{\frac{1}{2}} \left(1 - \frac{q^2}{M_P^2}\right)^3 \left|F_{P\gamma}\left(q^2\right)\right|^2\right],\end{aligned} \tag{2.4}$$

where P is the pseudoscalar meson which can be either π^0, η, or η', with $\mathrm{Br}(P \to \gamma\gamma)$ as shown in table 2.2. The kinematically accessible range gives us the integration limits: the lower limit is (the square of) two times the mass of the millicharge (for pair generation) $(2m_\chi)^2$, while the upper integration limit is the mass of the decaying meson P, M_P, squared. A factor of 2 appears outside the integral due to the freedom to choose any of the two photons as the virtual one. Finally, this equation also contains form factor terms, which are given by [20]

$$\left|F_{\pi^0\gamma}(q^2)\right|^2 = \left(1 + a_\pi \frac{q^2}{m_\pi^2}\right)^2, \tag{2.5}$$

$$\left|F_{\eta\gamma}(q^2)\right|^2 = \left(1 - \frac{q^2}{\Lambda_\eta^2}\right)^{-2}, \tag{2.6}$$

$$\left|F_{\eta'\gamma}(q^2)\right|^2 = \frac{\Lambda_{\eta'}^2\left(\Lambda_{\eta'}^2 + \gamma_{\eta'}^2\right)}{\left(\Lambda_{\eta'}^2 - q^2\right)^2 + \Lambda_{\eta'}^2\gamma_{\eta'}^2}, \tag{2.7}$$

where $a_\pi = 0.11$, $\Lambda_\eta = 0.72\,\mathrm{GeV}$ and $\gamma_{\eta'} = 0.13\,\mathrm{GeV}$. The tree-level Feynman diagrams of these 3-body decay processes can be seen in figure 2.5, where we also show a possible production mechanism from the decay of the ω and ϕ vector mesons which is described in Ref. [20]. Although not considered in our search, this mechanism has been implemented in a bespoke event generator which will be described in chapter 5, so it can be used in future searches.

2.4. Detector interaction

After being produced, mCPs must travel from the beam target to the detector through matter. The interactions of mCP with matter will cause energy loss (via rare hard collisions) and a change in their direction (via soft scattering off nuclei). It has been estimated [18] that the energy loss in 500 m of matter would only be a few MeV, and their deflection after each successive hit only around 10^{-3} rad, both negligible when we consider mCPs in the GeV scale. The centre of the MicroBooNE detector is approximately 679 m from the NuMI beam target, so the estimations for ArgoNeuT of Ref. [18] still apply.

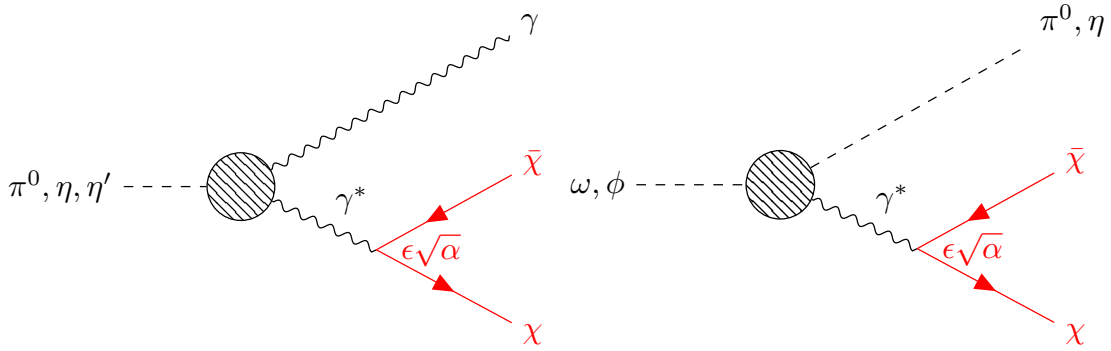


Figure 2.5.: Three-body decays of pseudoscalar (left) and vector (right) mesons producing millicharged particle pairs. These processes exist in the SM when replacing the mCP pair (shown here in red) with a pair of massive leptons, given kinematic constraints.

Once mCPs enter the detector volume, they may interact with its liquid argon active mass via elastic soft scattering with the electrons in the argon. At tree level, the interaction is described by the Feynman diagram in figure 2.6. In the laboratory frame for an mCP incident along the z axis, the 4-momenta shown in figure 2.6 are as follows:

$$p_e = (m_e, 0, 0, 0), \tag{2.8}$$

$$p_\chi = (E_\chi, 0, 0, \sqrt{E_\chi^2 - m_\chi^2}), \tag{2.9}$$

$$p_{e'} = (m_e + E_r, \vec{p}_{e'}), \tag{2.10}$$

$$p_{\chi'} = (E_\chi - E_r, \vec{p}_{\chi'}). \tag{2.11}$$

Here E_χ is the energy of the incident mCP, m_e is the mass of the electron and E_r is the recoil energy transferred from the mCP to the electron. We are considering a free electron at rest in the lab frame, which holds as we will consider momentum exchanges $E_r \geq 100$ keV, which is well above the atomic binding scale. The electron will be scattered with an angle φ with respect to the incoming mCP, obtained from momentum conservation as

$$\vec{p}_{\chi'} = \vec{p}_\chi - \vec{p}_{e'}, \tag{2.12}$$

$$p_{\chi'}^2 = p_\chi^2 + p_{e'}^2 - 2p_{e'}p_\chi \cos(\varphi), \tag{2.13}$$

where, $p_\chi = \sqrt{E_\chi^2 - m_\chi^2}$, $p_{\chi'} = \sqrt{(E_\chi - E_r)^2 - m_\chi^2}$ and $p_{e'} = \sqrt{E_r(E_r + 2m_e)}$ are the momentum magnitudes obtained from the energy-momentum relation, from which we find

$$\cos(\varphi) = \frac{E_r(E_\chi + m_e)}{\sqrt{E_r(E_r + 2m_e)(E_\chi^2 - m_\chi^2)}}. \tag{2.14}$$

As we will see, in the momentum transfer scale we are in, $E_r \ll E_\chi$, the mCP will not change its angle significantly, as is shown in Ref. [18]. Furthermore, the kinematics

in general will be the same whether we consider χ or $\bar{\chi}$; in addition we have no way of distinguishing between them, so in what follows we will refer to both particle and antiparticle generically as mCP.

The differential cross-section with respect to the recoil energy E_r can be obtained from the diagram in fig. 2.6, and is given by [18]

$$\frac{d\sigma}{dE_r} = \pi\alpha^2\epsilon^2 \frac{2E_\chi^2 m_e + E_r^2 m_e - E_r\left(m_\chi^2 + m_e(2E_\chi + m_e)\right)}{E_r^2\left(E_\chi^2 - m_\chi^2\right)m_e^2}. \tag{2.15}$$

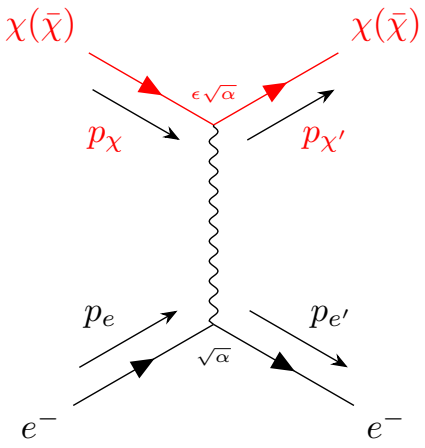


Figure 2.6.: Tree level diagram of the interaction between an (anti)mCP, shown in red, and one of the electrons bound to the argon in the detector. Due to the differences in energy scale, the electron can be considered to be initially at rest. The momenta in the laboratory frame are shown.

The differential cross-section in equation (2.15) diverges when E_r approaches zero, so we must integrate between a low-threshold cut-off $E_r^{\min}$ and the maximal kinematically allowed recoil energy $E_r^{\max}$. In this case, the total cross-section is [18]

$$\sigma\left(E_r^{\min}, E_r^{\max}\right) = \pi\alpha^2\epsilon^2 \left\{ \frac{m_e(E_r^{\max} - E_r^{\min})(2E_\chi^2 + E_r^{\max}E_r^{\min})}{E_r^{\max}E_r^{\min}(E_\chi^2 - m_\chi^2)m_e^2} - \frac{\left(m_\chi^2 + m_e(2E_\chi + m_e)\right)\log\left(\frac{E_r^{\max}}{E_r^{\min}}\right)}{(E_\chi^2 - m_\chi^2)m_e^2} \right\}. \tag{2.16}$$

The maximal kinematically allowed recoil energy is given by [18]

$$E_r^{\max} = \frac{\left(E_\chi^2 - m_\chi^2\right)m_e}{m_\chi^2 + 2E_\chi m_e + m_e^2}, \tag{2.17}$$

and therefore for given values of ϵ, m_e and E_χ, the total cross-section in equation (2.16)

depends only on $E_r^{\min}$ and it scales as

$$\sigma\left(E_r^{\min}\right) \propto \frac{a}{E_r^{\min}} - bE_r^{\min} + c\log\left(\frac{E_r^{\max}}{E_r^{\min}}\right), \tag{2.18}$$

with a, b and c constants. For values of $E_r^{\min}$ near zero this tends to infinity and the cross-section will be dominated by soft collisions. The low-energy detection threshold will determine $E_r^{\min}$, and thus the sensitivity to mCPs. In future sections, when discussing the simulation of interactions, we will choose a specific low-energy recoil cutoff based on MicroBooNE's low-energy detection efficiency and available computation time.

2.4.1. Relativistic limit for the cross-section

In the ArgoNeuT search, a relativistic approximation was used. The differential cross-section equation (2.15) can be rearranged slightly as

$$\frac{d\sigma}{dE_r} = \pi\alpha^2\epsilon^2 \frac{2E_\chi^2 m_e - E_r\left[-E_r m_e + m_\chi^2 + 2E_\chi m_e + m_e^2\right]}{E_r^2(E_\chi^2 - m_\chi^2)m_e^2}. \tag{2.19}$$

In the relativistic approximation we have $E_\chi \gg m_\chi$, and in the mass ranges we are looking for $m_\chi \gg m_e$. Since low-thresholds dominate, we will also have $E_\chi \gg E_r$ and so we can disregard the terms in parentheses in the numerator as they will all be small when compared to the first term $2E_\chi^2 m_e$. For the denominator, the term in parentheses approximates to E_χ^2, so we will be left with

$$\frac{d\sigma}{dE_r} \simeq \frac{2\pi\alpha^2\epsilon^2}{m_e E_r^2}. \tag{2.20}$$

This is the same as equation (3.2) in Ref. [18], and equation (1) in Ref. [17]. However, we will show in section 5.4 that the energy spectrum of the scattered electrons generated using this approximation looks very different to that of our full simulation of the kinematics, which we will use instead. One of the advantages of this approximation is that it does not depend on the mCP mass m_χ, but we have found that, depending on the mass and the scattered electron reconstruction efficiency, this could result in either an under or overestimation of the sensitivity, which a full simulation of the kinematics would avoid.

2.4.2. Mean free path and probability of detector hit

The scattering cross section can be converted into a mean free path using [21]

$$\lambda(E_r^{\min}) = \frac{m_A}{ZN_A\rho\sigma(E_r^{\min})}, \tag{2.21}$$

where $Z = 18$, $m_A = 39.948$ g/mol, $\rho = 1.3954$ g/cm^3 are the atomic number, atomic mass and liquid density of argon and N_A is Avogadro's number.

The probability of an mCP interacting with argon will depend upon the relation between the mean free path λ and the length L which the mCP crosses in the active mass of the detector. If the ratio L/λ is small, the probability of an interaction above the detector threshold will be very small. We are also interested in the cases of more than one interaction with different argon atoms as the mCP traverses the active mass of the detector. We can assume that these interactions are independent of each other in each event, and thus in the limit of many events we should obtain a Poisson distribution for the number of n-hit events as

$$p_{n\text{-hits}}\left(n, L/\lambda\right) = \frac{1}{n!}\left(\frac{L}{\lambda(E_r^{\min})}\right)^n e^{-L/\lambda} \qquad \text{for } L/\lambda \ll 1. \tag{2.22}$$

A generalisation for the $L > \lambda$ case (due for example to larger charge ratio ϵ) can be found in Ref. [18]. It should be remembered that the detector interaction cross section depends on factors such as the incident mCP energy, so this probability must be calculated on an event-by-event basis.

2.5. Current experimental limits

The current experimental limits on the existence of mCPs are shown in figure 2.7. This is the range of ϵ and m_χ accessible to beam-dump experiments. Cosmological limits, which we will not explore, exist outside this range.

At the lowest mCP mass values up to 35 MeV limits are set by a reinterpretation [22] of LSND electron-neutrino scattering data [23]. The 1998 experiment carried out in SLAC [24] is an early example search, and the only purpose-built mCP experiment still on this list, with world-leading limits between 60 and 85 MeV. Save for the SLAC-dominated region, the SENSEI experiment [25], a dark-matter experiment that uses the Skipper-CCD technology, has leading limits between 35 and 250 MeV. SENSEI uses data collected in 2020 from the same Fermilab NuMI beam that ArgoNeuT used and we will use. Since then, SENSEI has moved to SNOLAB [26]. Above 250 MeV the exclusion is dominated by a reinterpretation of data from the Big European Bubble Chamber (BEBC) [27].

Two new purpose-built millicharged particle experiments at CERN are expected to place stringent limits in the near future. The milliQan experiment released partial results in 2020 using a prototype "demonstrator" with a size of around 1 % that of the full detector [28] and is currently taking data. The FORMOSA collaboration [29] similarly installed a demonstrator in early 2024 in the same tunnel that houses FASER.

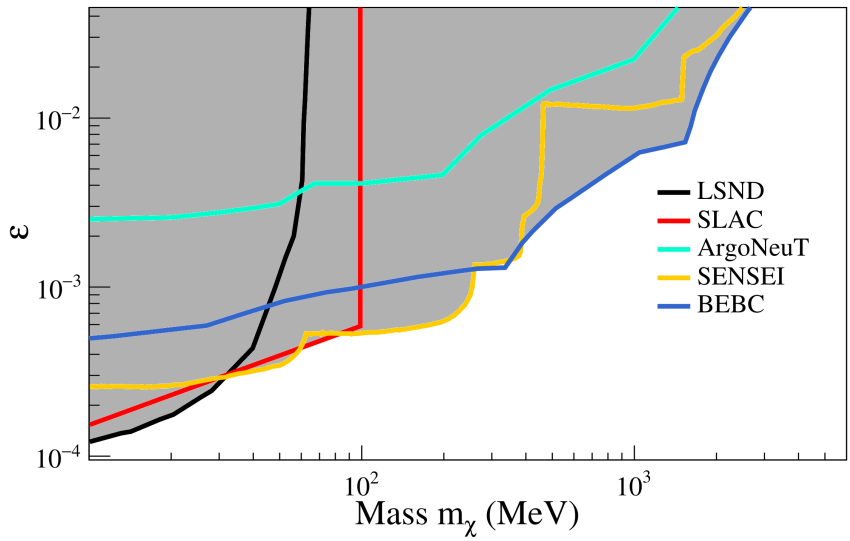


Figure 2.7.: Current experimental limits on the production of millicharged particles using accelerator experiments. The greyed-out area corresponds to excluded regions in the mass m_χ (horizontal axis) and electromagnetic coupling ϵ (vertical axis) 2D parameter space.

3. The MicroBooNE experiment at Fermilab

In this chapter, we will give an overview of the MicroBooNE liquid argon time projection chamber (LArTPC) detector. First, we will give an introduction to LArTPC detector technology in a general way in section 3.1. Then we will discuss the specifics of the MicroBooNE detector in section 3.2, going through its main subsystems and the data it collected during its operation time. We will then zoom out and discuss the neutrino beams at the Fermi National Accelerator Laboratory (Fermilab) in section 3.3. We then end this chapter with a brief description of the SBND experiment, a next generation LArTPC detector which has recently started operations.

3.1. LArTPC technology

A liquid-argon time projection chamber (LArTPC) for neutrino detection was first proposed by Carlo Rubbia in 1977 [30] as a technology that can combine the high spatial resolution of bubble chambers with the triggering and fast readout capabilities of electronic detectors. Early examples include prototypes for the ICARUS project [31].

A large mass of liquid argon doubles as both neutrino target and detector medium. Charged particles produced from neutrino-argon interactions ionise and excite the nearby argon atoms, freeing electrons and producing scintillation photons. An electric field is applied which drifts the ionisation electron towards readout wires located in the anode. An image can be formed by combining the 2D wire plane information with the drift time. A diagram of this readout principle is shown in figure 3.1. The timing resolution can be vastly improved with the addition of a light collection system to detect the fast-travelling scintillation light, usually placed behind the wire planes, which can also be used for triggering the detector.

3.1.1. Liquid argon, ionisation, and scintillation

Pure liquid argon (LAr) has many characteristics that make it desirable for this type of detector, which were already outlined in its original proposal [30]. As argon is a noble gas, it does not attach electrons, enabling long drifting distances and hence a large TPC can be built. At the same time, its high density of $1.40\,\mathrm{g\,cm^{-3}}$ make each nucleus a "big target", increasing the probability of neutrino interactions compared to hydrogen-based bubble chambers, or water-based neutrino detectors, so detectors can be made compact. Its nuclear interaction length of 85.7 cm makes it a good active medium for hadronic calorimetry, while its short radiation length of 14.0 cm is excellent for electromagnetic

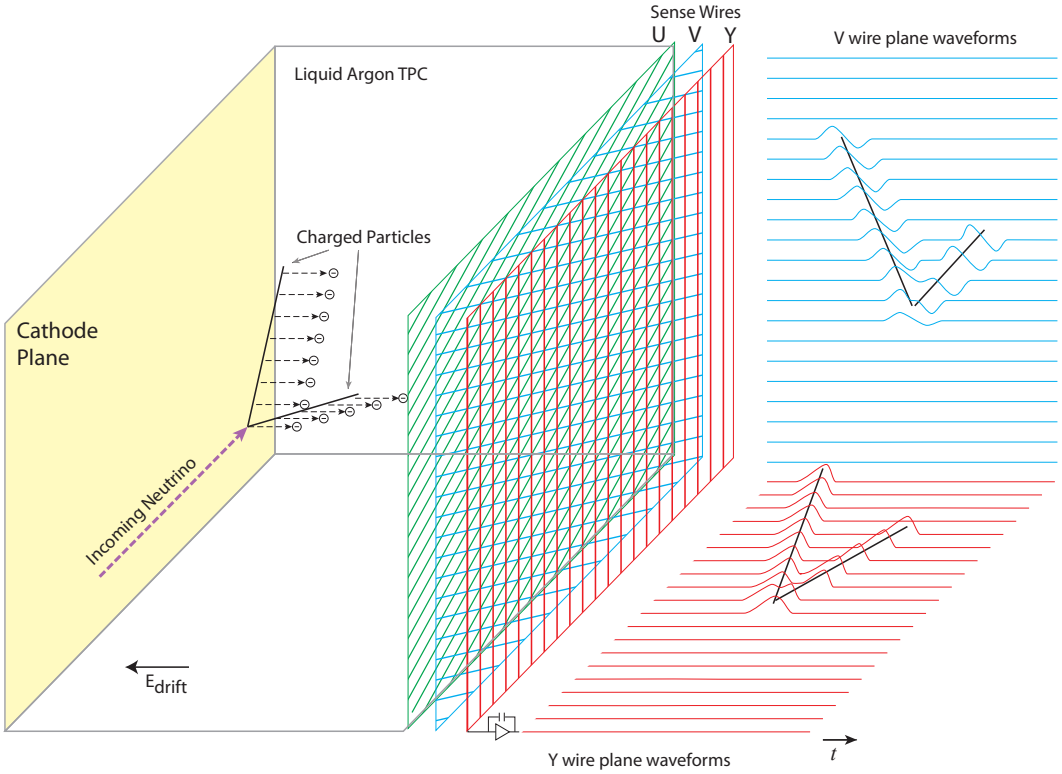


Figure 3.1.: Diagram of the readout principles of the MicroBooNE LArTPC. An incoming (invisible) neutrino enters the detector, producing ionisation electrons which are made to drift towards three wire planes, labelled U, V and Y. The signals from the three 2D planes is read out, with time forming a third dimension. Diagram from Ref. [32].

calorimetry. Finally, LAr is easy and cheap to obtain compared to other noble gases such as xenon, as argon makes up around 1 % of the Earth's atmosphere.

Argon is ionised at a relatively low 26.3 eV. Some of the ionised electrons will recombine with the ion unless an electric field separates them quickly [33], so the large electron mobility of LAr is another advantage. If the LAr is highly purified, the ionisation electrons can drift distances of the order of meters using electric fields of the order of $1\,\mathrm{kV\,cm^{-1}}$ [34].

LAr is also a very good scintillator, producing about 24 000 photons per MeV at $500\,\mathrm{V\,cm^{-1}}$ drift field [32]. Scintillation light originates either from the prompt decay of single excimer states with a lifetime of around 6 ns, or the slower decay of triplet excimer states with a lifetime of 1.6 µs. Both components consist of photons with an ultraviolet wavelength of 128 nm, which is not absorbed by LAr, making it a transparent medium to its own scintillation.

3.1.2. Charge deposition topologies in a LArTPC

The passage of charged particles through liquid argon can be separated into two main categories based on the distinct topologies they generate. Minimum ionising particles (MIPs) will produce a *track* topology as they steadily lose energy through Coulomb interactions with the electrons in the argon while travelling in a mostly straight line. As they come to a stop, their mean energy loss per unit length reaches a maximum, a phenomenon known as a Bragg peak. Typical track-producing particles originating from neutrino interactions are protons, charged mesons, or muons, the latter of which produce very long tracks. The profile of their energy loss as they stop can be used to discriminate the particle type, as shown in figure 3.2. In this figure, we see two distinct populations around the expected values for protons and muons, given their distinct profile of energy loss per unit length. The Bragg peak is the sharp rise on low end of the residual range, which is the distance measured back from the track stopping point.

The second main topology is known as a *shower*. Electrons, due to their low mass, will not behave like MIPs, and their energy loss will instead be dominated by bremsstrahlung. As they travel through argon, the highly energetic photons they radiate will produce new electron-positron pairs, which can then in turn radiate new high-energy photons, and so on. This cascading reaction widens in the travel direction of the initiating particle, and ends when no new pairs can be created. Showers can also be initiated by high-energy photons, in which case the first few centimetres of the shower will be an e^+e^- pair rather than a single electron, which results in double the ionisation energy deposited. The start of the shower can then be used to distinguish between electron and photon-initiated showers, as shown in figure 3.3.

Photons, and neutral particles in general, do not deposit energy via ionisation, so by themselves they are invisible to the detector and can only be detected by the secondary charged particles that they produce. Therefore, an interaction that produces both a photon and a charged particle will show a distinctive gap between the interaction vertex and the start of the shower, signalling the travel distance of the photon before the first electron-positron pair was created. The powerful electron/photon separation capabilities

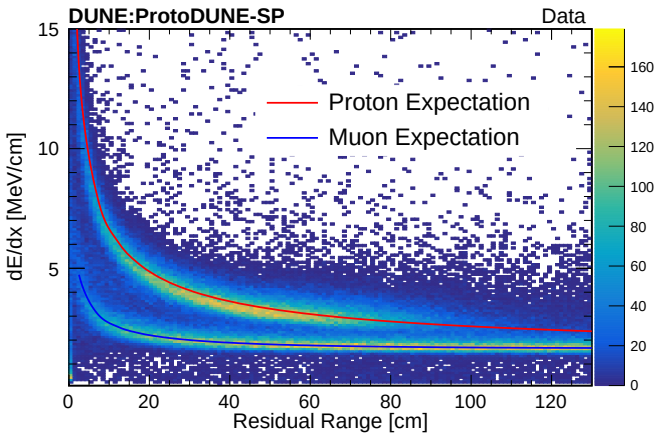
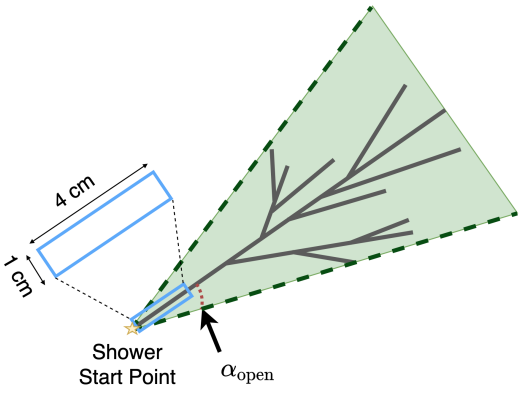
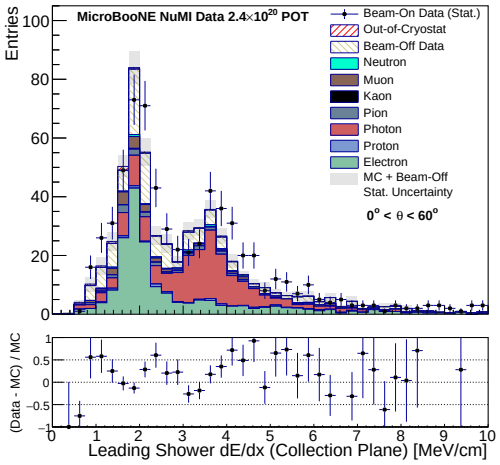


Figure 3.2.: Data from the ProtoDUNE-SP LArTPC, showing separation between protons and muons in their energy loss profile. The red and blue lines are the expected values for each particle type. From Ref. [35].



(a) Diagram of an electromagnetic shower shows a widening characterised by the opening angle α_{open}. The first 4 cm are used for electron/photon separation.



(b) Data (markers) and simulation (filled histograms) comparison for the leading reconstructed shower in the MicroBooNE experiment, using data from the NuMI beam. The main contributors are electrons (green histogram) and photons (red histogram).

Figure 3.3.: The energy deposited by the initial few cm of a shower (diagram in (a)) can be used to distinguish between electrons and photons, as shown in (b). Both figures from Ref. [36].

of a LArTPC give it an advantage over water Cherenkov detectors.

In recent years, there has been an increasing interest in the study of MeV-scale physics using LArTPCs [37]. At this energy scale, charged particles will only travel a very short distance before coming to a stop, producing a point-like isolated deposition, regardless of its type.

3.1.3. Detector effects

The drift of ionisation electrons is affected by many different transport effects. First, the electrons may reattach to the ionised argon atoms unless quickly separated by the action of an electric field, an effect known as recombination.

The electric field inside the TPC is distorted by the remnant ionised argon atoms, an effect known as the space charge effect (SCE). As they are positively charged, the argon ions drift towards the cathode, in the opposite direction compared to the ionisation electrons. They drift much more slowly due to their heavier mass. This problem affects surface LArTPCs as the high incidence of ionising cosmic ray muons leads to an accumulation of Ar^+. Even in ideal conditions, the drifting electrons diffuse longitudinally and transversally as they travel, affecting reconstruction in all three dimensions.

Finally, impurities in the liquid argon can capture the electrons as they drift, effectively quenching the charge which will arrive to the readout wires. This effect, which affects especially interactions produced far from the readout wires in the anode, is known as electron lifetime, and was one of the main concerns in the initial proposal for LArTPCs.

We will discuss these effects more in detail in chapter 4, when discussing electron lifetime calibration in the MicroBooNE LArTPC experiment.

3.2. The MicroBooNE experiment

MicroBooNE is a LArTPC neutrino detector located in the Fermilab campus in Batavia, Illinois, USA [32]. It uses 85 tonnes of liquid argon as active mass in a TPC volume of $2.54\,\mathrm{m} \times 2.33\,\mathrm{m} \times 10.37\,\mathrm{m}$. It collected data from neutrino interactions between 2015 and 2020, obtaining the largest dataset of neutrino interactions on liquid argon to date.

Situated at a short-baseline distance of 468 m from the Booster Neutrino Beam (BNB) target, its main physics objective is the study of the low-energy excess (LEE) of electron-like events observed by the MiniBooNE experiment [38] (and previously by LSND [39]), which was situated a similar distance downstream in the same beamline.

In 2021, the MicroBooNE collaboration released its first set of results, using half of its full dataset to search for evidence of the LEE in multiple final state topologies using different reconstruction methods [40]. It searched in charged-current ν_e interactions without pions using the Pandora reconstruction framework [41], quasi-elastic ν_e interactions using deep learning reconstruction [42], single photons from the radiative decay of neutral-current Δ baryons using Pandora [43], and an inclusive search of charged-current ν_e interactions using the Wire-Cell framework [44]. Very recently, MicroBooNE published the first of its second set of results of a search for the LEE in charge ν_e interactions without pions [45], in the first search to use the full collected MicroBooNE

BNB dataset. No significant excess has been found, and the MiniBooNE LEE remains unexplained.

In addition to this, MicroBooNE has a very extensive cross-section measurement programme [46–60]. More importantly for this thesis, MicroBooNE has a rich programme of searches for Beyond the Standard Model (BSM) physics. In 2020, MicroBooNE published the first search for heavy neutral leptons (HNLs) in a LArTPC [61], looking for HNL decays into a pion and a muon, using data from the BNB. It has since expanded upon this result using data from the Neutrinos at the Main Injector (NuMI) beam [62], searched for Higgs portal scalars[62, 63], decays of HNLs to e^+e^- and π^0 final states [64], a search for a dark matter portal in the form of "dark tridents" [65], and demonstrated the use of a LArTPC to search for neutron-antineutron oscillations [66].

3.2.1. Time projection chamber (TPC)

The MicroBooNE TPC has three main components, the cathode, field cage, and anode wire planes. The origin of the MicroBooNE coordinate system is situated in the anode, at the middle of its height, on the upstream face (defined as the face which the BNB beam first encounters). The x axis is parallel to the horizon, from anode to cathode, the y coordinate increases with height, and the z coordinate follows the direction of the BNB beam from up- to downstream. The cathode and anode are opposing faces of the rectangular cuboid shape of the detector, with the field cage enclosing the volume. Photographs of the outside and inside the TPC are shown in figure 3.4.

The cathode is an assembly of 9 stainless steel sheets where a negative voltage is applied, making it an equipotential surface. The flatness of the cathode is established from a survey with more than 10 000 points using a laser tracker, from which a "best fit plane" is defined. The maximum deviations from this best fit plane in either direction are 6.6 mm and −6.5 mm [32]. The cathode surface is parallel to the anode within 0.0413° [32]. The field cage defines the *active volume* of the detector, and its main job is to create a region with a uniform electric field. Its structure consists of 64 sets of steel tubes, shaped as rectangular loops, covering the perimeter and running parallel to the cathode and anode planes. These loops are held in place by a rib structure made out of G-10, a composite material composed of fibreglass laminate bonded by epoxy resin, commonly used for structural support. These field cage loops are connected via a resistor chain which divides the voltage in equal steps, to enforce the electric field uniformity. The nominal electric field is of $273\,\mathrm{V\,cm^{-1}}$, which results in an electron drift velocity of $0.11\,\mathrm{cm\,\mu s^{-1}}$.

An anode frame holds the three planes of wires, also called anode planes, labelled U, V and Y. Coming from the cathode, the first two planes that ionisation electrons encounter are called induction planes, U and V, which are orientated at ±60° from the vertical line. Bias voltages of −200 V, 0 V, and 440 V are applied to the three wire planes to ensure that ionisation electrons continue to travel past the induction planes and are only collected in the third and final Y plane (also referred to as the collection plane, with vertically oriented wires), a condition refred to as *wire transparency*. The 2400 wires in each induction plane and 3456 wires in the collection plane are separated by

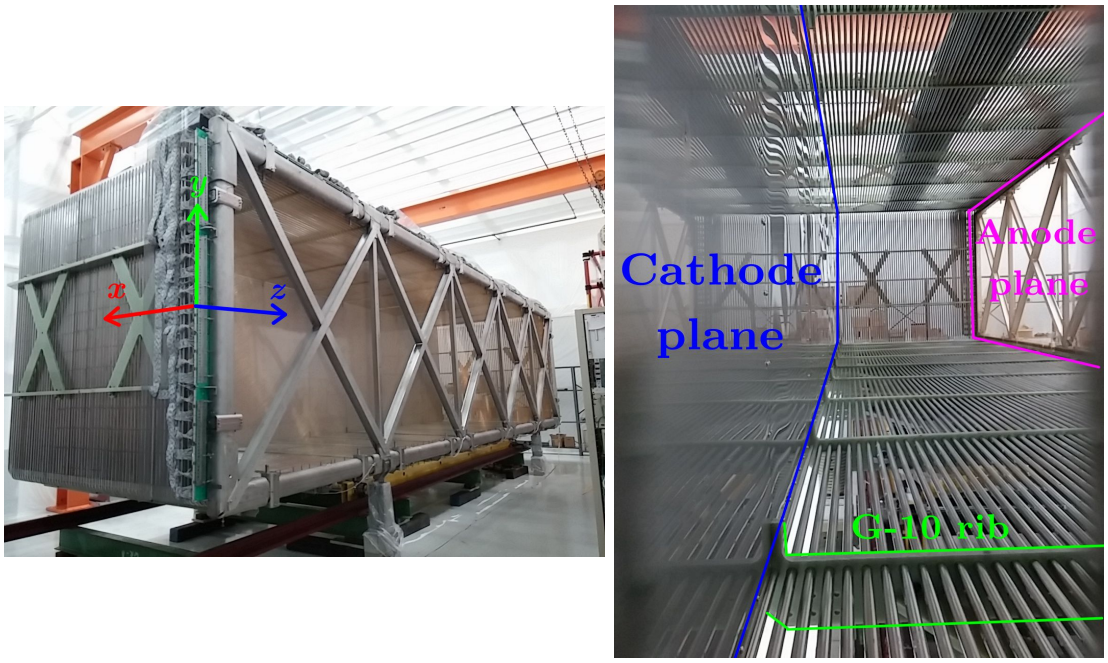


Figure 3.4.: Left: outside view of the TPC before it was placed inside the cryostat, showing the upstream face of the detector and the anode plane. The MicroBooNE coordinate system has been drawn. Right: inside view from the upstream face. The tubular field cage loops which make up the field cage are held together by ribs made out of G-10. Figures adapted from [32].

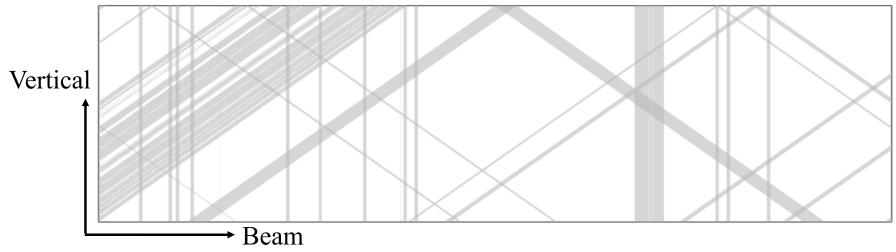


Figure 3.5.: Non-functioning regions in the three anode planes superimposed. Some regions have overlap of dead wires from multiple planes. From Ref. [68].

3 mm between each other, and the anode planes themselves are also separated by 3 mm. The signal is amplified and shaped by application-specific integrated circuits (ASICs) submerged in the liquid argon, then the signal is digitised by analog-to-digital converters (ADCs) with a sampling rate of 2 MHz that determines the TPC drift time resolution of 0.5 µs clock "ticks".

Early during commissioning, a number of readout channels were deemed non-functioning due to various issues with the ASICs, shorted wires, or noisy channels [67]. The distribution of the unusable wires in the yz plane is shown in figure 3.5. These "dead wires" cover an area of approximately 30 % of the yz plane, and their impact on reconstruction and detector systematics will be discussed as they arise.

3.2.2. Light-collection system

MicroBooNE has an array of 32 8-inch photomultiplier tubes (PMTs) for light detection. These are distributed in the yz plane behind the anode plane in 5 groups arranged in a circular shape, as shown in the photograph in figure 3.6. A wavelength-shifting plate coated with tetraphenyl-butadiene (TPB) is placed in front of them in order to shift the vacuum ultraviolet wavelength 128 nm scintillation photons into the visible-light region which the PMTs are most sensitive to.

Scintillation photons traverse the detector in nanoseconds, which is orders of magnitude faster than the slow drifting of ionisation electrons, which can take milliseconds. The light-collection system enables a precise timing calibration of the event by which a t_0 can be defined when a beam interaction takes place, for which its readout waveforms are digitised with a higher sampling rate of 64 MHz, independently of the TPC readout. During reconstruction, light activity can be matched to charge readout for the precise timing of every interaction inside the TPC. This light-to-charge matching facilitates the removal of cosmic activity [68] and enables nanosecond-time event resolution [69]. Additionally, light can be used for triggering the detector, which will be discussed in the next subsection.

Figure 3.6.: Photograph of the PMT array mounted on the inner cryostat wall before the TPC was installed. A group of six PMTs is arranged in a circular array. The flat wavelength-shifting TPB plates are placed in front of them. From Ref. [32].

3.2.3. Trigger and readout

Neutrino interactions are only expected to be in narrow µs-sized *beam spill* time windows, which are signalled by the upstream neutrino beams which will be discussed in section 3.3. To capture all the relevant charge information in neutrino interactions, MicroBooNE records 4.8 ms (equivalent to 9600 clock ticks) of TPC activity around an unbiased 23.4 µs *beamgate window*; this TPC time frame is longer than the maximum drift time of electrons of 2.3 ms. However, due to the low neutrino cross-section, most beam spills will not actually contain a neutrino interaction. MicroBooNE uses hardware and software triggers for deciding whether to store the readout information or not.

The Fermilab accelerators broadcast a continuous on/off signal every 100 ns. When a beam spill is about to occur, the signal is switched to on, which determines the hardware beam-on trigger. The hardware trigger signal is sent to the PMTs, which are normally zero-supressed, to start reading the unbiased 23.4 µs beamgate window and look for a rapid increase in their waveforms. The software trigger activates if the PMT waveforms show activity greater than 9.5 photo-electrons (PE), and the event is then recorded.

In addition to the beam-on hardware trigger, MicroBooNE uses a function generator signal as hardware trigger when the beam is off. This beam-off trigger, also known as EXT, is used for recording samples of beam-off events which either pass the software trigger due to cosmic activity, or are unbiased towards it (EXT-unbiased), as the study of events without neutrino interactions is useful for calibration and reducing systematic uncertainties.

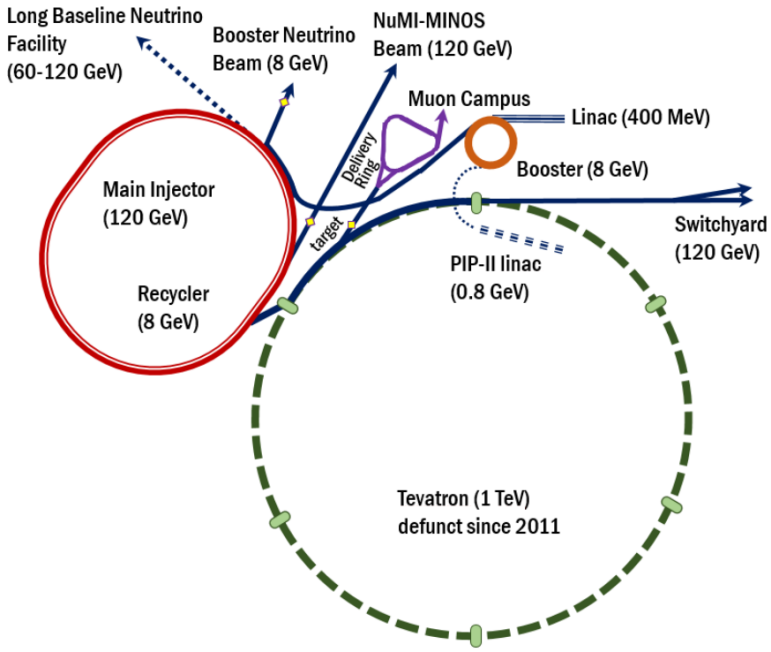


Figure 3.7.: Diagram of the Fermilab accelerator complex. Figure from Ref. [71].

3.2.4. Cosmic ray tagger

MicroBooNE is a surface detector with no overburden shielding, which exposes it to a high cosmic ray muon rate. This results in tens of cosmic muons being recorded in each event, which can be a background to muon-neutrino interactions. A cosmic-ray tagger (CRT) system was installed and became operational during 2017 [70]. The CRT consists of four planes covering the top, bottom and side faces of the cryostat. Each plane is composed of two layers of independent modules, each of which is composed of a series of scintillating strips arranged in parallel, with photosensors on their ends. When a muon crosses a CRT plane, the photosensor of the strip it crossed will tag its position along the axis in which the strips are arranged. The second layer of modules is set up with its strips perpendicular to the first, and combining both layers gives an x, y position of the muon when crossing the plane.

3.3. Neutrino beams at Fermilab

MicroBooNE sits downstream from two different neutrino beams, the Booster Neutrino Beam (BNB) and the Neutrinos at the Main Injector (NuMI) beam. Both produce neutrinos in fundamentally the same way, from the decay of pions produced in the collisions of protons on a target. The protons are accelerated in the Fermilab accelerator complex [71], a diagram of which is shown in figure 3.7.

For our purposes, the main differences between both beams will be the beam intensity,

the energy of the protons, the material of the target, and the orientation of the beam with respect to MicroBooNE, all of which determine the neutrino flux seen by the detector. We will first describe the Fermilab proton accelerators, then describe the NuMI beam, and then only briefly describe the BNB, as it is the former which is used in the main analysis presented in this thesis. We will end this section with some comments on the data collected by MicroBooNE.

3.3.1. The Fermilab accelerator complex

Starting from a pulsed source of ionised hydrogen, protons are accelerated to 400 MeV in a linear accelerator that feeds into the 474 m circumference Booster rapid cycling synchrotron. The Booster accelerates the protons to 8 GeV with typical intensities of 5.0×10^{12} protons per pulse at 85–90 % efficiency [72]. The Booster can deliver up to 84 "bunches" (as the pulses are usually referred to) per cycle. The full output of one accelerator cycle is referred to as its *beam-spill*, which we mentioned in section 3.2.3 when discussing the MicroBooNE readout windows. The BNB spill packs its bunches in a *batch* in a time window of 1.6 µs.

If not delivered to the BNB, the booster spills are passed on to the Recycler, a 3.3 km storage ring made from permanent magnets. The Recycler can store up to 6 full BNB spills, but through a technique called *slip-stacking* [71, 72] it can stack up to 12 batches. Six new batches are injected to the Recycler while the six already present slightly decelerate, after which they can be made to "slip together" just at the point at which they are delivered to the Main Injector synchrotron. The Main Injector shares a tunnel with the Recycler, and is capable of accelerating the protons up to 150 GeV, which was used in the past to inject protons into the Tevatron [72]; the Main Injector is now operated at 120 GeV to obtain optimal neutrino rates in the NuMI beam [73]. The 6+6 1.6 ms stacked batches result in a NuMI beam spill window of 9.6 ms. A diagram summarising the concepts of bunch, batch, and spills is shown in figure 3.8.

3.3.2. The NuMI beam

The NuMI beam [73] is produced from the collision of 120 GeV protons delivered by the Main Accelerator into a graphite target. This high-energy collision produces abundant charged pions and kaons, which are focused using a series of magnetic horns to increase their on-axis flux. The charged mesons then pass down a decay pipe with a length of 675 m, which ensures all of them will have decayed by the end. Pions decay predominantly to a muon and a muon neutrino, which, after crossing a hadron monitor, travel through 5 m of absorber material which stops most hadrons. The muons and neutrinos then travel through 240 m of rock, which stops muons but cannot stop the beam of neutrinos, which now travel toward the on-axis detectors downstream. This process is illustrated in the diagram in figure 3.9, and is a common template shared by neutrino beams originating from accelerators.

The NuMI beam has provided an on-axis neutrino flux to the MINOS experiment [75], MINERνA [76], and the ArgoNeuT LArTPC [77], among others. The focusing horns

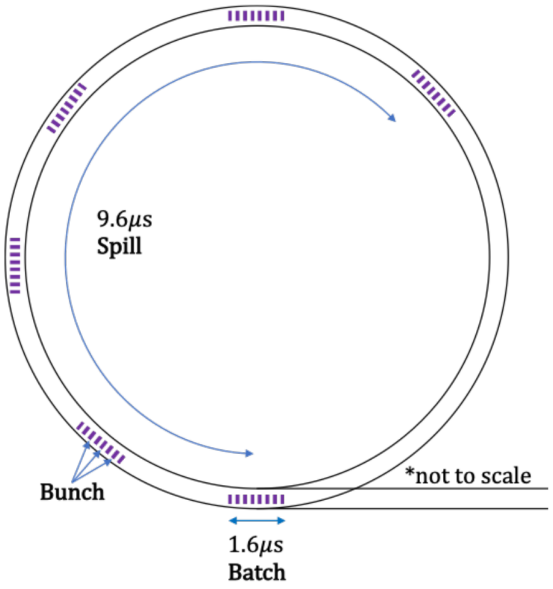


Figure 3.8.: The NuMI beam structure. Proton pulses form bunches, and the Booster beam delivers a batch of "width" 1.6 µs each cycle to the BNB (one BNB spill) or the NuMI beam. In NuMI these are stacked and grouped in a longer 9.6 µs NuMI spill. Figure from Ref. [74].

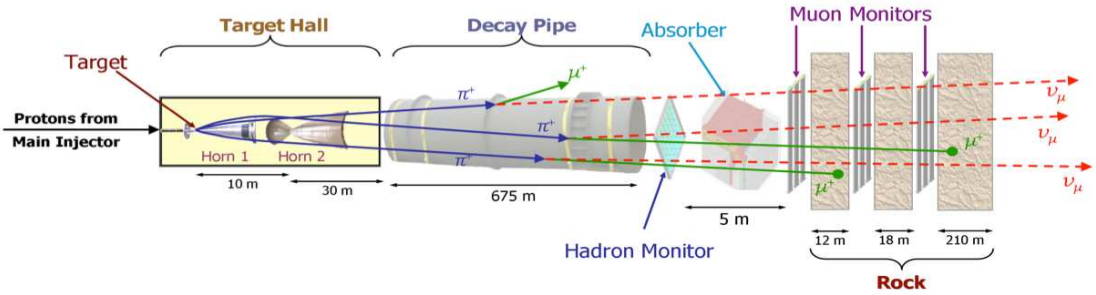


Figure 3.9.: The NuMI neutrino beam from the collision of protons on target. From Ref. [73].

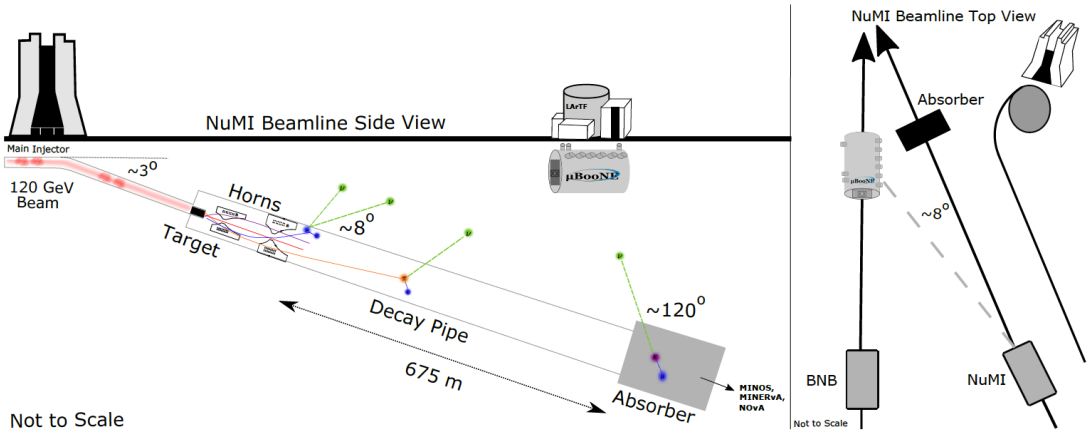


Figure 3.10.: The position of the MicroBooNE detector relative to the NuMI neutrino beam target. Left shows a side view, where the downward angle of the NuMI beamline can be appreciated. Right shows a top view, where the on-axis BNB is also shown. From Ref. [49].

work by using a magnetic field to deflect hadrons, effectively acting as hadronic lenses increasing the hadron flux in a flexible energy range. Their operating current can be reversed, which has the effect of focusing either positively or negatively charged mesons. In forward horn current (FHC) mode, positive mesons are focused, which decay to a positive muon and a muon neutrino. In reverse horn current (RHC) mode, negative mesons result in a muon and antineutrinos. Thus, the horn current determines whether the resulting beam is predominantly neutrinos or antineutrinos.

For MicroBooNE, NuMI provides a highly off-axis neutrino beam, with a wide range of incoming angles ranging from 8° near the target up to 120° for neutrinos produced from the decay of kaons at rest in the absorber. A diagram of the position and orientation of MicroBooNE relative to the NuMI beam is shown in figure 3.10. The NuMI neutrino flux at MicroBooNE has a ν_e component of 4 % above 250 MeV, which is 10 times larger than in the highly pure ν_μ BNB [49]. The 120 GeV proton-on-target energy of the NuMI beam is equivalent to a centre of mass energy of 15 GeV, which can produce heavy neutral mesons such as the J/ψ or Υ. It also results in a higher yield of mesons compared to the lower-energy BNB.

3.3.3. The Booster Neutrino Beam (BNB)

The BNB produces neutrinos from the collisions of 8 GeV protons on a beryllium target. It provides an on-axis neutrino flux to the SBND, MicroBooNE and ICARUS experiments, the three of which comprise the Short-Baseline Neutrino programme at Fermilab [78], and in the past it provided a neutrino flux for the MiniBooNE experiment.

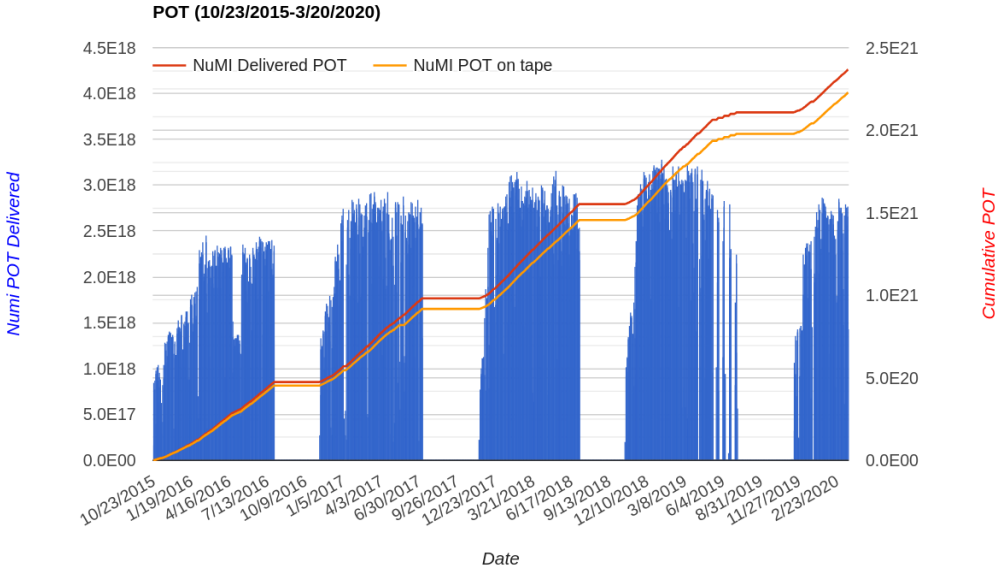


Figure 3.11.: NuMI collected POT. Each of the five distinct columns represents a MicroBooNE Run. Data acquisition downtime results in not all POT delivered by NuMI being recorded to tape storage for offline analysis. This figure was generated using MicroBooNE POT accounting tools.

3.3.4. MicroBooNE beam data

MicroBooNE keeps a count of the protons which are delivered to either NuMI or BNB by tracking the rate of their respective beam spills during data taking. This count is referred to as *protons on target* (POT), and is used to normalise data from each beam independently. The beam spill rate depends on the beam conditions at the time of data taking, and the POT is recorded multiple times during a single data taking session, even when events do not trigger the detector. A data taking session is called a run (with lowercase r), and may last up to seven hours. Depending on beam and detector conditions, data-quality selections can be performed on a per-run basis, and the runs passing these selections are referred to as *good runs*.

MicroBooNE collected neutrino beam data between 2015 and 2020, which is separated in five Runs (uppercase R) of roughly one year each. The POT recorded from NuMI is shown in figure 3.11. The data recorded on tape storage for offline analysis is 2.23×10^{21} POT before good run selections. The main analysis presented in this thesis will use 1.07×10^{21} POT of data from Runs 1, 2, and 3, as the remaining data from Runs 4 and 5 is in the late stages of processing. We plan to use the full dataset of Runs 1 through 5 in a publication of these results, which is in preparation.

3.4. The SBND experiment

The Short-Baseline Near Detector (SBND) [78][1] is a 112 t active mass 4 m × 4 m × 5 m LArTPC located on-axis with the BNB, at a very short baseline of 110 m fro the proton target. Its size and close proximity to the BNB target will expose it to a tremendous flux of on-axis neutrinos, and it is expected that the size of its dataset of neutrino-argon interactions will surpass that of MicroBooNE. SBND was filled with liquid argon during March 2023, and by December 2024 it has completed its transition from commissioning to operations.

Figure 3.12 shows the main components of the SBND detector. SBND has a cathode plane running along the centre of the detector and two anode plane assemblies (APAs) on each side, which means it is effectively two TPCs. This keeps the drift distance at a size of 2 m, comparable to MicroBooNE. SBND is currently operating at its nominal bias voltage, producing a drift electric field of 500 V cm^{-1}. The four APAs went through extensive testing during their construction, with an emphasis on quality assurance of the tension of the wires [79]. Its photon detection system (PDS) is considerably more powerful than MicroBooNE's [80], featuring a total of 120 photomultiplier tubes in addition to a novel X-ARAPUCA [81] light trap to increase photon collection using silicon photomultiplier arrays. Its cosmic ray tagger system (CRT) has full 4π surface coverage and has been fully integrated in its design since the beginning, which will help with the removal of cosmic activity.

[1] In the proposal cited here, what will become SBND is referred to by the name LAr1-ND.

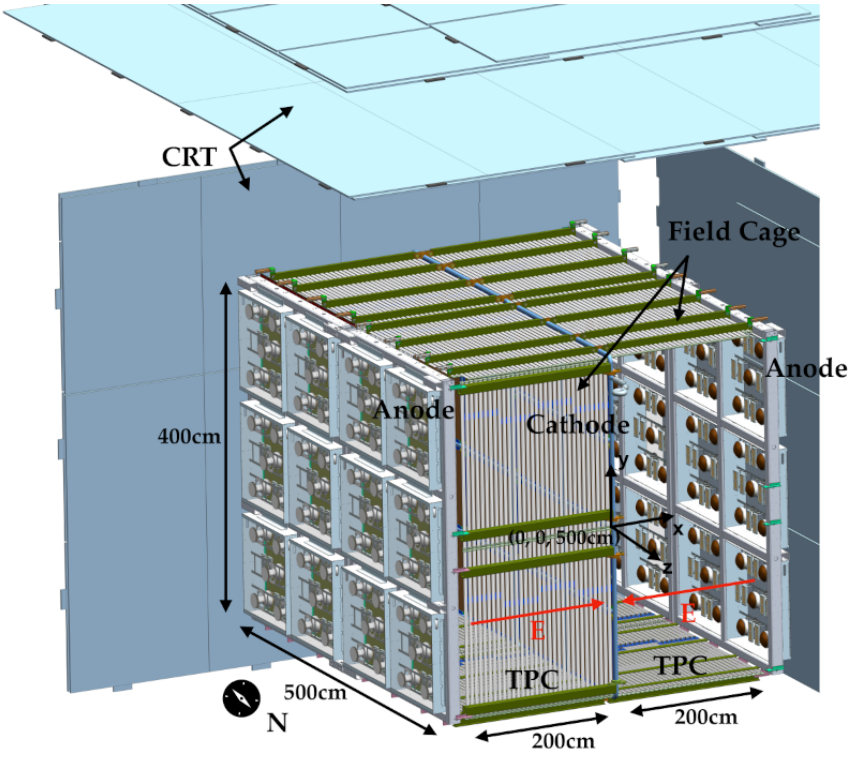


Figure 3.12.: Computer rendering of the SBND detector, showing its main subsystems. Figure from Ref. [80].

4. Electron Lifetime

In this chapter, we will explain in some detail the concept of electron lifetime and how it is measured and corrected for in LArTPCs. While not directly related to the main topic of a search for millicharged particles in MicroBooNE, it is an important aspect of LArTPC calibration directly related to the argon purity. Furthermore, during the course of my PhD I directly participated in the studies related to electron lifetime described in this chapter, performing a validation of Run 4 and a calibration of Run 5 in MicroBooNE using both existing analysis tools and developing some of my own. I also did a pilot study for a measurement of the electron lifetime using early commissioning data in SBND. This is an opportunity to document this effort while having a closer look into one of the many electron transport effects in liquid argon. As a result of this, this chapter is more self-contained than the rest of the thesis. All the work presented in this chapter is original unless explicitly stated otherwise.

In section 4.1 we define the concept of electron lifetime in the context of LArTPC neutrino experiments, and outline some general aspects of its measurement. We also mention additional electron transport effects to provide context and to give a fuller picture of what electron lifetime calibration entails. In section 4.2 we discuss how the electron lifetime calibration for runs 4 and 5 of MicroBooNE was performed. Finally, in section 4.3 we show a pilot study for an online/semi-online measurement of electron lifetime in SBND, in the context of its commissioning.

4.1. Electron Lifetime in LArTPCs

As the ionisation electrons drift in the liquid argon towards the readout wires, electronegative impurities in the detector, such as molecules containing oxygen [82–84], may capture them. This reduces the amount of charge that reaches the wire planes, especially that of interactions produced far from the anode planes, whose electrons have to drift further through the argon and have more opportunities to be captured. This effectively attenuates the measured signals, which will result ultimately in non-uniformity in the reconstruction of particle energy. This quenching effect, sometimes called electron attenuation, is therefore an indicator of low argon purity. We can quantify this effect by defining the electron lifetime τ as the exponential decay

$$n = n_0 \exp\left(-\frac{t_d}{\tau}\right), \tag{4.1}$$

where an initial number of ionisation electrons n_0 is reduced to n after a certain drift time t_d, thus reducing the amount of charge reaching the wires. In an ideal scenario, as

electrons drift in the x direction, the drift time will be correlated to the drift distance x_d via the nominal drift velocity v_d as $x_d = v_d t_d$, where v_d depends on the applied electric field. Therefore equation (4.1) could be written as

$$n = n_0 \exp\left(-\frac{x_d}{v_d \tau}\right). \tag{4.2}$$

In reality, the drift velocity is not constant or uniform because the local electric field is not. Electrons also do not drift constantly or unidimensionally, due to a variety of transport effects. The space charge effect (SCE) [85] is caused by the counterpart to the ionisation electrons, the remnant ionisation argon atoms (Ar^+), which drift much more slowly and towards the cathode. The accumulation of Ar^+ in the active volume leads to distortions in the local electric field, which causes the position of ionisation electrons to be misreconstructed. Surface detectors, such as MicroBooNE and SBND, are particularly susceptible to this effect due to the high incidence of ionising cosmic muons.

Variations in the electric field also play a role in the recombination effect [33]. If the ionisation electrons are not spatially separated quickly from the ions, they might recombine with them, resulting in loss of signal reaching the wires. Under higher electric fields, separation between these occurs more quickly, which decreases the chance of recombination. Recombination in LArTPCs is usually corrected at the calorimetry stage, when the deposited energy per unit length dE/dx[1] is reconstructed from the measured deposited charge per unit length dQ/dx using a modified box model [86, 87]

$$\frac{dE}{dx} = \frac{\exp\left(\frac{\frac{dQ}{dx}}{C_{\text{cal}}} \frac{\beta' W_{\text{ion}}}{\rho \varepsilon}\right) - \alpha}{\frac{\beta'}{\rho \varepsilon}}, \tag{4.3}$$

where C_{cal} is a calibration constant, in which MicroBooNE normalises 1 ADC to 200 electrons [87], W_{ion} is the work function of argon — i.e. the energy required to ionise an argon atom — with a value of 23.6×10^{-6} MeV/electron, ρ is the liquid argon density of $1.38\,\text{g}\,\text{cm}^{-1}$, and ε is the MicroBooNE drift electric field of $273\,\text{kV}\,\text{cm}^{-1}$. Finally, α and β' are the parameters of the modified box model, which are adjusted to experimental data collected by the ArgoNeuT experiment [86], in which they measure $\alpha = 0.93$ and $\beta' = 0.212\,(\text{kV/cm})(\text{g/cm}^2)/\text{MeV}$. The values of these parameters can be modified to account for systematic uncertainties, as will be described in section 7.3. The recombination correction stage is performed *after* the electron lifetime corrections are applied, and thus we will not elaborate further on it in this thesis.

Drifting electrons are also subject to diffusion [88]. Rather than individual particles, ionisation electrons in an electric field can be better thought of as a drifting cloud, whose shape can vary non-isotropically and can be modelled separately in longitudinal and transverse components (with respect to the drift direction). This can result in different drift times or transverse velocity components, which affect reconstruction.

[1] Please note here dx refers to a small length along the trajectory of an ionising particle crossing the detector, which is not necessarily in the x drift direction.

A full calibration of the detector must take all of these effects into account, and the order in which they are applied matters, as each step relies on assumptions from underlying models. For example, the electron lifetime calibration is performed after uniformity in the drift direction has been established, and thus a stable drift velocity can be assumed (within some uncertainty). Therefore, the assumption that $v_d = x_d/t_d$ is justified, and we can use equations (4.1)and (4.2) interchangeably.

As both MicroBooNE and SBND are surface detectors, the abundant cosmic muons which cross them can be used for electron lifetime calibration. These relativistic muons are minimum ionizing particles (MIPs), and as they cross the full (or most of the) length of the detector, their energy loss per unit length can be considered constant and to follow a Landau distribution [89, 90]. For example, anode-cathode crossing (AC) tracks, which have one of their ends near the anode and the other near the cathode, are very useful for calibrating effects across the drift direction such as the SCE and electron attenuation [87].

Typically, some selection based on the incidence angle is applied to the crossing muon tracks. For example, tracks that are parallel to the readout wires — i.e., those with angles of $\pm 60°$ or $90°$ with respect to the horizontal in the anode plane — can result in misreconstructed signals. This is due to a fundamental limitation of using only three wire planes to reconstruct full 3D information; we rely on the trajectories of particles generating signals across multiple wires. Similarly, tracks that are parallel to the drift direction will produce a continuous stream of electrons arriving at the readout wires, which makes resolving the individual *hits* required to reconstruct a useful signal difficult. Signal reconstruction in MicroBooNE will be discussed in section 5.6, and the specific selections applied to the MicroBooNE and SBND studies will be described in sections 4.2.1 and 4.3.3 respectively.

The most common method to perform a lifetime calibration in a LArTPC will be presented using the case of MicroBooNE in the following section, but in basic terms a slicing of the detector is made across its drift direction x, and hits from AC tracks are sorted into the different bins across this slicing, and then compared to measure the attenuation across the drift direction.

In the case of SBND, which has two TPCs with anode plane assemblies on each end and a cathode in the middle, a single muon can cross the detector anode-to-anode and could be used in the calibration of both TPCs simultaneously, however we will still refer to this and similar situations as anode-to-cathode, and treat the tracks reconstructed in each TPC as separate entities.

4.2. Electron lifetime in MicroBooNE

MicroBooNE has a drift distance of 2.54 m between the cathode and anode, and a nominal drift velocity of $1.098\,\mathrm{mm\,\mu s^{-1}}$, therefore a nominal full drift time of 2.31 ms. MicroBooNE considers an electron lifetime above 37 ms to be essentially infinite, and if the measured values are above this, they are recorded as 1 s. During Runs 1 to 3, MicroBooNE's electron lifetime was high for the entire period, and therefore no electron

lifetime corrections were applied to data, opting instead to not include the few days of low lifetime in its set of good runs. However, early during Run 4, a significant drop in electron lifetime was observed, which is believed to have been the result of a delivery of "bad argon" to the detector. This event affected especially the data taken in mid-to-late September 2017, but high electron lifetime could not be recovered during the entirety of Run 4. Therefore, lifetime correction is considered critical for this run in particular, and applied to all of its data and simulation samples, as well as that of Run 5. In the following subsections, we will present a complete picture of the lifetime correction in MicroBooNE for Run 4, and then briefly mention the results for Run 5, which used the same methods.

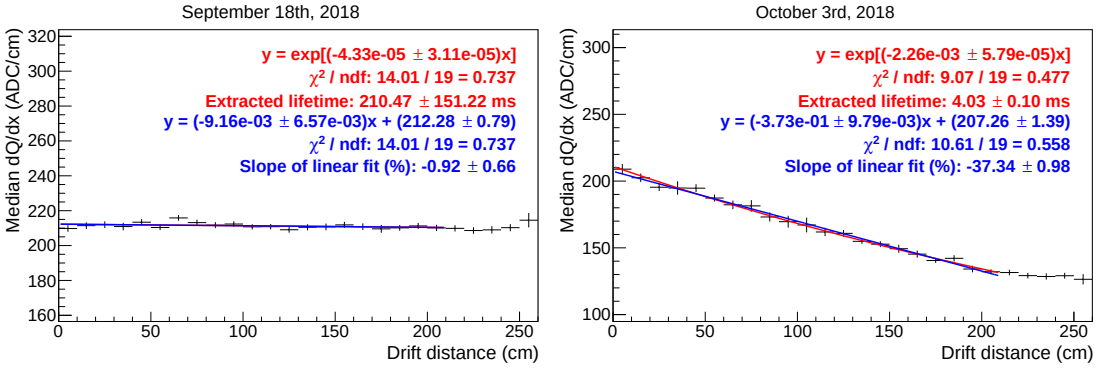
In MicroBooNE, corrections to the various transport effects are applied sequentially. At the lifetime correction stage, the SCE has been corrected for, and a *YZ* uniformity map correcting for misconfigured or dead TPC channels and transverse diffusion has been applied, therefore it should be assumed that all figures in this section contain these corrections. Longitudinal diffusion and recombination are simulated, but not corrected for at this stage. We will only present some discussion on the *YZ* uniformity map in subsection 4.2.4; this will be done in order to keep the discussion focused on electron lifetime, and because I had direct involvement in the *YZ* correction for Run 5, so it will be presented only in the context of this Run. A full description of the calibration of MicroBooNE in Runs 1 through 3 can be found in Ref. [87].

4.2.1. Lifetime extraction in Run 4

For lifetime extraction, MicroBooNE uses anode-cathode crossing (AC) cosmic muon tracks, which deposit energy in the full length of the drift direction. The AC tracks used for calibration have to fulfil some selection criteria. First, their projection length along the detector x coordinate should be greater than 250 cm, to ensure a good coverage of the entire drift distance while allowing some room for error due to inefficiencies in track reconstruction. The horizontal track angle θ_{XZ} should not be too parallel ($< 15°$) to the drift direction, which is enforced with the selections $\theta_{XZ} < 75°$ and $105° < \theta_{XZ}$. Similarly, to avoid the tracks being too parallel to the collection plane wires, the vertical angle θ_{YZ} must be outside the range $[80°, 100°]$.

The electron lifetime is extracted offline on a per-day basis, and stored in a `postgreSQL` database as lifetime in milliseconds, and its error, for each calendar date; the sample of cosmic muons used is obtained from *good runs* starting between 00:00 and 23:59 in Fermilab local time — either Central Standard Time (CST) or Central Daylight Time (CDT), depending on the date — on any given day. Using a full day's worth of data ensures sufficient statistics for the calculation of each lifetime value on every day of the Run.

The detector is divided into 26 bins of approximately 10 cm in size along the drift direction, which was found to provide sufficient granularity for the measurement while maintaining adequate statistics in most cases. The median value of the deposited charge per unit length dQ/dx in each of these bins by the AC tracks is used as a measure of the energy lost due to a low electron lifetime, i.e. the median value will be biased



(a) (b)

Figure 4.1.: Median dQ/dx per x-position bin in the collection plane for two different dates. The left figure (a) shows data collected on September 18th 2018. The right figure, (b), shows data from October 3, 2018. Both figures show exponential and linear fits to the data (in red and blue respectively). In the left figure, both fits are completely overlapping.

towards lower values for bins further from the anodes. The median values of the dQ/dx distributions are then compiled into new histograms, two example dates of which are shown in figure 4.1. The two days shown in figure 4.1 are representative of a day with normal purity (September 18th, 2018), where we observe a flat distribution of the median dQ/dx throughout the drift distance, and a day of low purity (October 3rd, 2018), where we see a reduction in the median dQ/dx for greater drift distance, which is indicative of low electron lifetime conditions.

Figure 4.1 has two different fits applied to the data. First, an exponential fit is used to extract a value for the electron lifetime using equation (4.2). We see, however, that the lifetime extracted for the normal purity day (figure 4.1(a)) of (210.47 ± 151.22) ms has a high uncertainty. This is because the distribution of data is flat rather than exponential, which results in a poor exponential fit. This is in contrast to the case of low purity days, such as the one shown in Fig. 4.1(b), where the extracted lifetime of (4.03 ± 0.10) ms has a much smaller uncertainty. When studying the electron lifetime of Run 4, which includes both days of normal and low purity, the MicroBooNE collaboration considered that the large error bars in days of normal purity were not helpful when visualising the overall argon purity, especially when considering that lifetimes above 37 ms are considered infinite given the size of the detector. For this reason, a linear fit is also applied to the data, in which case, the slope of the fit can be interpreted as a measure of the argon purity. Under normal purity the distribution of dQ/dx is uniform across bins, resulting in a slope close to zero, whereas in low purity days, there is a marked difference in dQ/dx depending on the drift distance, resulting in a negative slope. The slope is multiplied by a factor of 100, and interpreted as a "percentual" deviation from a zero slope. In this ad-hoc model — which has so far only been used internally in the MicroBooNE collaboration — it was found empirically through analysis of the data, that

the benchmark lifetime of 37 ms closely corresponds to a linear fit slope of 0.05, or a 5 % deviation from a flat distribution.

The extracted lifetimes and linear fit slopes for the entire Run 4 are shown in figure 4.2, in the top and bottom figure respectively. In this figure, every data point represents one day of detector operation. We can appreciate a short period of 8 days of "normal purity" at the start of the run, followed by low purity throughout the full run, with dips in the bottom figure representing periods of very low purity due to deliveries of bad argon, issues with the filter and high-voltage (HV) system, the cryostat, or general purity issues. Starting near the end of January 2019, the electron lifetime oscillated around 21.6 ms on good days, which was found to correspond to an 8.5 % deviation from a flat slope of the linear fit. Under normal purity conditions, the data points in the bottom figure in Fig. 4.2 would be centred between the red dashed lines corresponding to a lifetime above 37 ms, or a 5 % slope deviation from a flat distribution. Only the first 8 data points are within this range.

4.2.2. Lifetime correction and validation

To obtain a well-calibrated detector, we perform a correction to the deposited charge given the knowledge of the daily lifetime that was extracted. This is done by applying an x-dependent correction to every hit based on its x position and the extracted daily lifetime, with the aim of recovering a uniform distribution of charge across the drift distance. The lifetime-corrected charge per unit length $(dQ/dx)_{\tau\ \mathrm{corr}}$ is set as

$$\left(\frac{dQ}{dx}\right)_{\tau\ \mathrm{corr}} = \exp\left(\frac{x_d}{\tau v_d}\right)\left(\frac{dQ}{dx}\right)_{YZ_{\mathrm{corr}}}, \tag{4.4}$$

where $(dQ/dx)_{YZ_{\mathrm{corr}}}$ is the (lifetime uncorrected) reconstructed value. The YZ_{corr} subscript indicates that a YZ-plane correction, which we will discuss in subsection 4.2.4, has already been performed. Inside the exponential term, τ is the electron lifetime value in the database for the day we are correcting for, x_d is the position of the hit along the drift distance, and v_d is the drift velocity. Since τ is the lifetime value extracted by fitting the analogous negative exponential $\exp(-x/\tau v_d)$ it follows that this correction aims to cancel out the decay in dQ/dx due to electron lifetime.

After the lifetime corrections have been applied to data, a sample was run through the same analysis to perform a validation. If the corrections were applied correctly, a new lifetime extraction upon the lifetime-corrected subsample would show high lifetime values. The sample used for validation is a subsample of Run 4, from September 14th to October 29th, 2018. This range of dates covers a wide range of purity conditions and extracted electron lifetimes. The reasoning is that if the lifetime correction is applied correctly to this representative period, then we can be confident that corrections are applied correctly throughout Run 4.

The selection of muon tracks used for extracting a lifetime value is the validation sample is the same as described in section 4.2.1. The extracted values of the electron lifetime for the validation sample are shown in figure 4.3. In the top figure, we observe that all extracted lifetime values are above the 37 ms threshold. There are many points

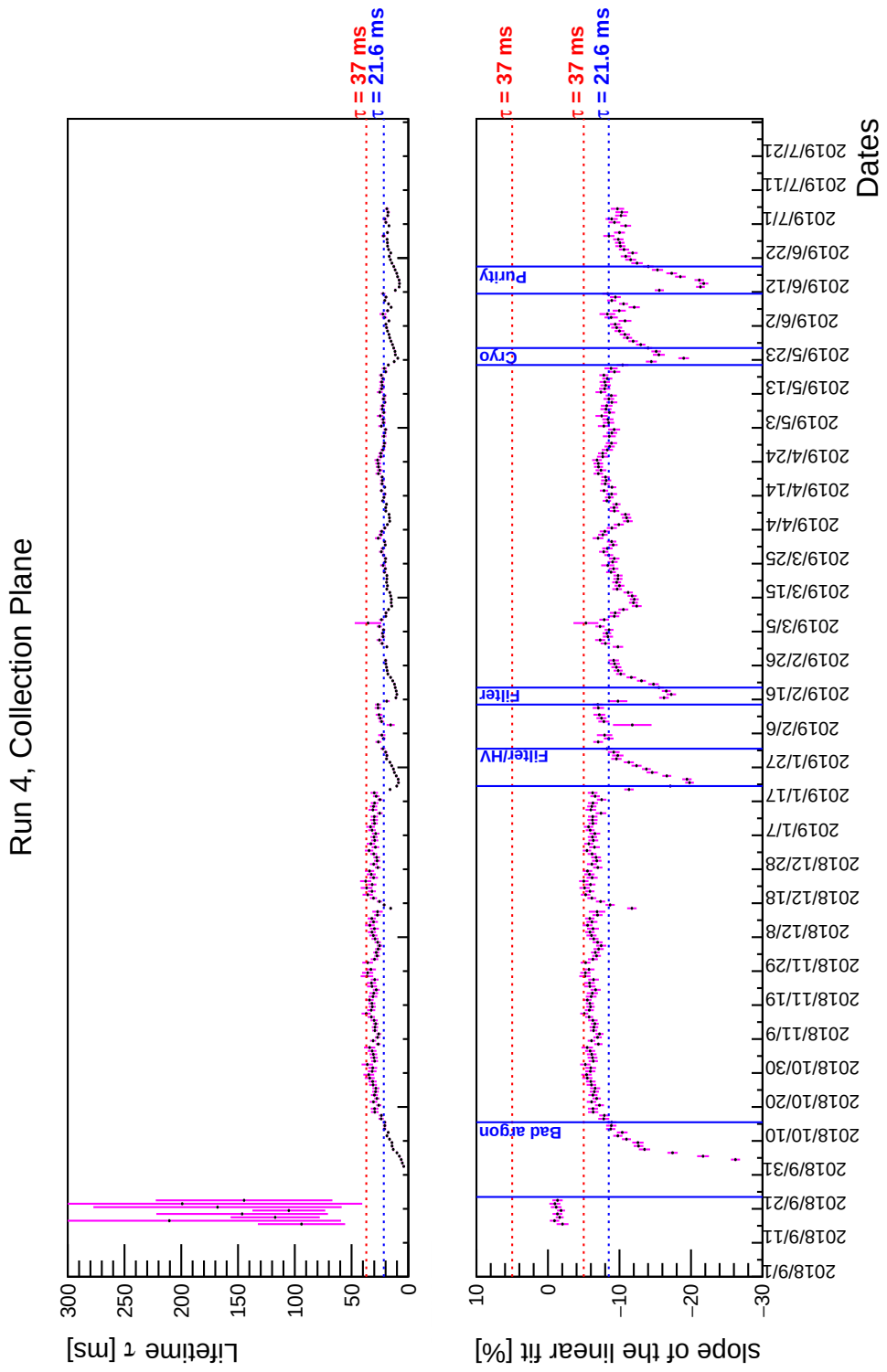


Figure 4.2.: Extracted electron lifetime for every day of Run 4. The top figure shows the extracted lifetime per day, while the bottom figure shows the slope of the linear fit (times 100). Periods where the purity dips considerably are tagged using blue vertical bars and text in the bottom figure.

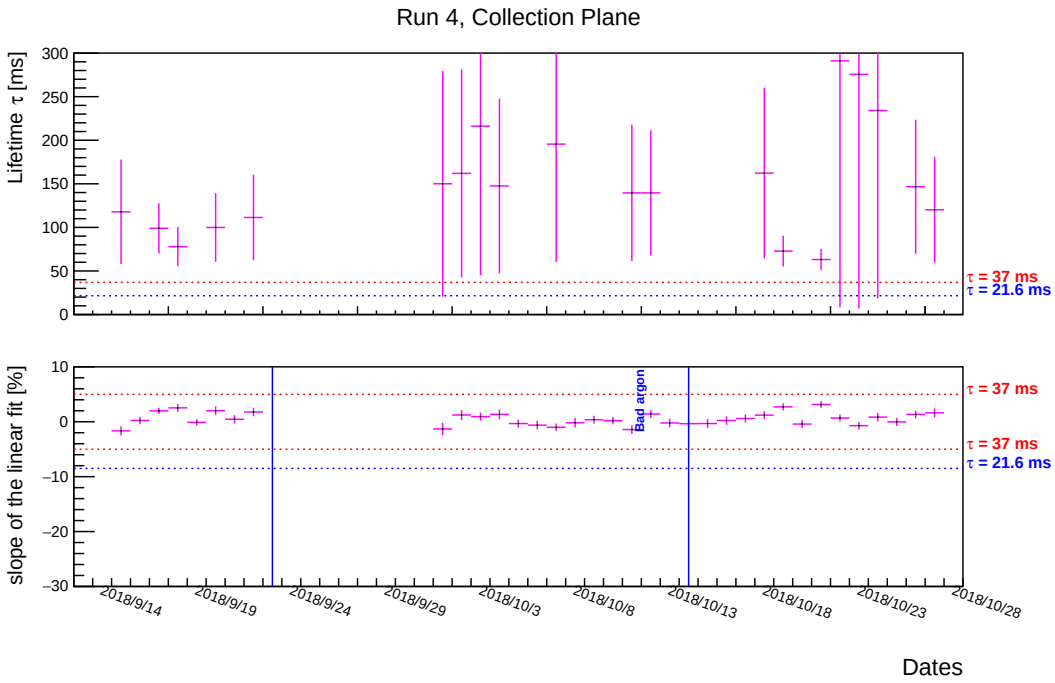


Figure 4.3.: Extracted electron lifetime (top) and slope of the linear fit (bottom) for the electron-lifetime corrected sample. The period of "bad argon" delivery with the lowest lifetime before corrections is shown in the bottom figure between vertical blue lines. The horizontal red and blue dashed lines in the top and bottom figures represent lifetimes of 37 and 21.6 ms. Some data points in the top figures are missing due to being outside the vertical range.

which are outside the vertical range of the figure, which is already very high as a value of 37 ms is considered infinite given the size of the detector. The slope of the linear fits shown in the bottom figure of 4.3 all lie within the 5% zones (which is equivalent to 37 ms, shown as red dashed lines) compared to a flat distribution. The bottom figure also shows the blue dashed line which represents the pre-correction lifetime of 21.6 ms around which most days in Run 4 oscillated. The fact that most data points are well above this value indicates that the electron lifetime correction has been applied correctly, by making the deposited charge uniform throughout the drift direction on days of low purity, while not negatively impacting days of normal purity.

4.2.3. Closure test on Monte Carlo

So far we have discussed the electron lifetime corrections to data, but in order for our Monte Carlo (MC) simulation samples to model the data correctly, they too must take into account the daily conditions of the detector, including the electron lifetime. In order to add this effect to our simulation, a quenching of the deposited charge per unit

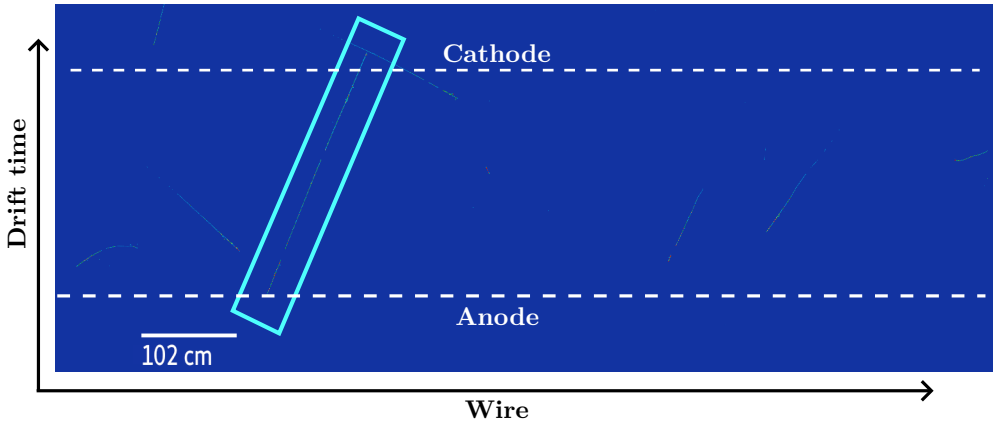


Figure 4.4.: Event display of a simulated anode-cathode crossing (AC) track (inside teal box) used for the closure test in the detector collection plane. As the readout window is longer than the drift time, activity beyond the anode and cathode is recorded. The other tracks visible in the image are from beam-off data overlaid alongside the Monte Carlo track.

length $(dQ/dx)_{\text{sim}}$ on simulations is applied based on the daily electron lifetime and hit position. MicroBooNE applies this for the first time in Run 4; previous runs used simulation with an "infinite lifetime", as only data from days with high lifetime were used.

To check that this quenching is correctly applied to our simulation, a closure test was performed, in which a value for the electron lifetime was extracted using a sample of simulated AC tracks overlaid alongside beam-off data. If the electron lifetime extracted from simulated tracks only matches the values in the database, which was obtained using data, that would mean that the lifetime smearing is being correctly applied to the simulation. An event display of an example simulated AC track is shown in figure 4.4.

We perform a selection of the simulated tracks only by looping over all the tracks in the event, and then looping over their component hits, and counting the fraction of them which are associated with the truth information. If the track has 90 % of its hits associated with the truth, then the track is classified as simulation, and used in the extraction. We can refer to this fraction as *track purity*, and its distribution for tracks is shown in figure 4.5. As shown in this figure, the vast majority of tracks in the sample have a track purity either close to 0.0 or close to 1.0, with the population between 0.2 and 0.8 being orders of magnitude smaller.

Once the AC tracks from simulations are selected, the lifetime extraction proceeds in the same manner as described in section 4.2.1: we store the hits from AC tracks in 26 histograms according to their position in the x coordinate, then compile the median values in a new histogram, and perform an exponential fit to extract a value for the lifetime. The result of the lifetime extraction is shown alongside of the values used in the database in figure 4.6(a). In this figure, we see a good agreement between MC

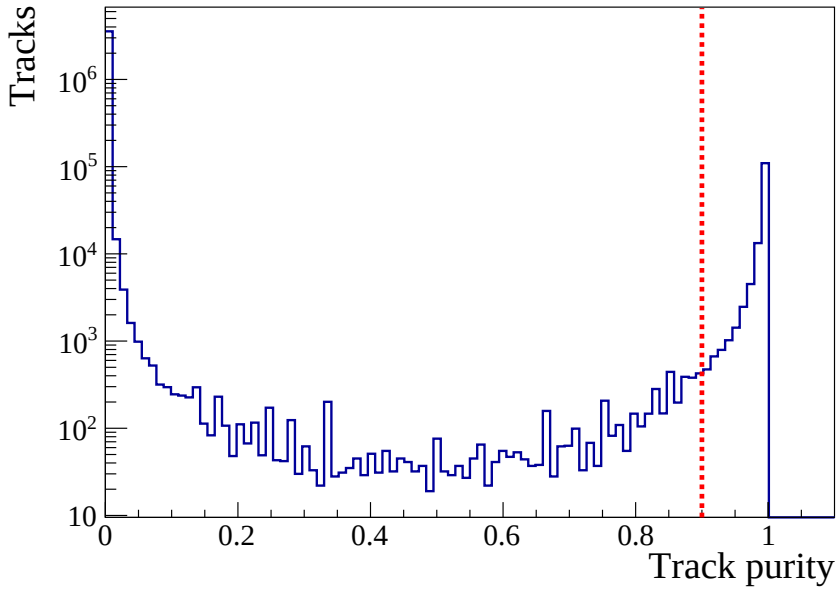
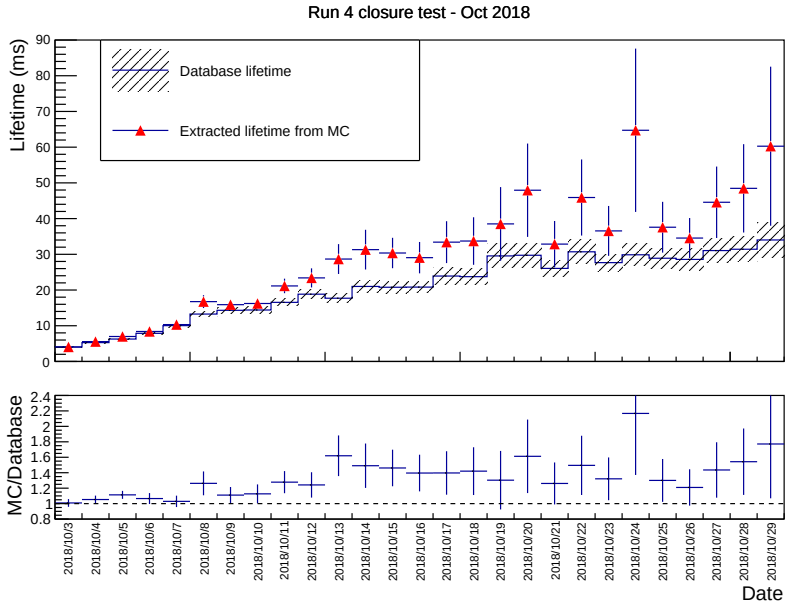


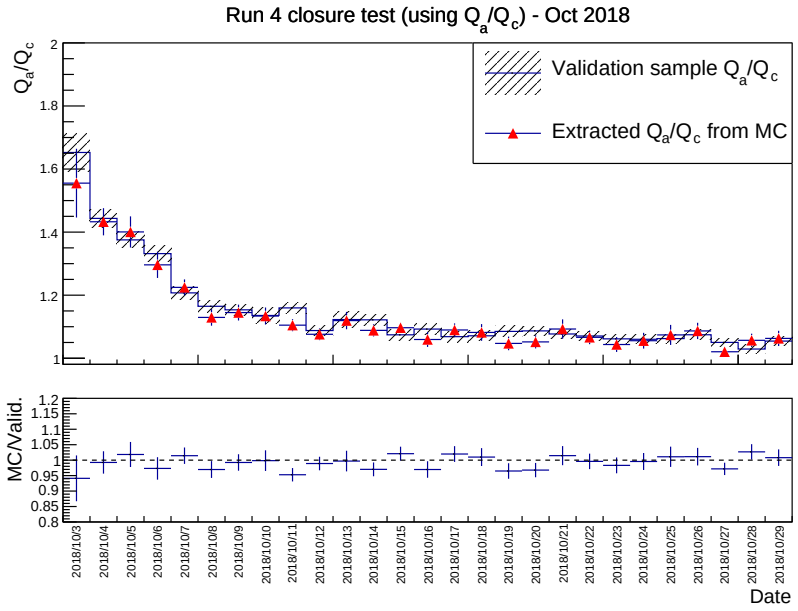
Figure 4.5.: Distribution of tracks in the closure test sample according to the fraction of their hits associated to truth (track purity). The red dotted line shows the value of 0.9 above which we classify a track as truth in our closure study.

and the database for days of lowest purity, but a poor agreement for days of higher purity. A possible reason for this could be the previously mentioned fact that the higher the extracted lifetime, the exponential fits have a higher uncertainty. In this case, the effect is compounded by the fact that this high-uncertainty lifetime was applied to simulations. This would not explain why there is a bias towards higher lifetimes in this second extraction, however.

Another commonly used metric is the ratio of charge present at the anode versus the charge present at the cathode, or Q_A/Q_C. While this value does not contain the full information of the charge distribution across the full drift distance, and is only the ratio of the median charge in the first and last bins of the x-slicing, it is still useful as a simple and commonly used indicator of the quenching of the charge as it traverses through the detector. A comparison of the Q_A/Q_C between MC AC tracks and beam-off data is shown in figure 4.6(b). The beam-off data is taken from the same events, simply inverting the track selection to those below 0.9 track purity, i.e. those below the dotted red line in figure 4.5. Figure 4.6(b) shows a much better agreement between data and simulation, and thus shows that the electron lifetime smearing has been applied to the simulation correctly on days of high lifetime. The uncertainties shown in figure 4.6(b) do not come from the lifetime extraction, rather they come from the uncertaintines from the first and last bins (from the median charge in the anode and cathode) of the drift distance histograms.



(a)



(b)

Figure 4.6.: Two different validations of the application of electron lifetime smearing to MC samples. (a): Electron lifetime as extracted from MC (red triangles) compared to values in the database (solid line). Uncertainties from the exponential fit in both cases. (b): Ratio of the charge at the anode (Q_a) and the charge at the cathode (Q_c) for both MC (red triangle markers), and the beam-off data validation sample (solid line). Uncertainties are from statistics in the anode and cathode bins only.

4.2.4. YZ uniformity correction for Run 5

MicroBooNE performs a calibration in the YZ plane, which has not been discussed thus far, for clarity of presentation. The YZ uniformity correction aims to correct for misconfigured TPC channels, dead channels, and transverse diffusion in the YZ plane, which is perpendicular to the drift direction x in which the electron lifetime correction is applied. We will return to equation (4.4) and take a step back to focus on the right-hand side YZ corrected term

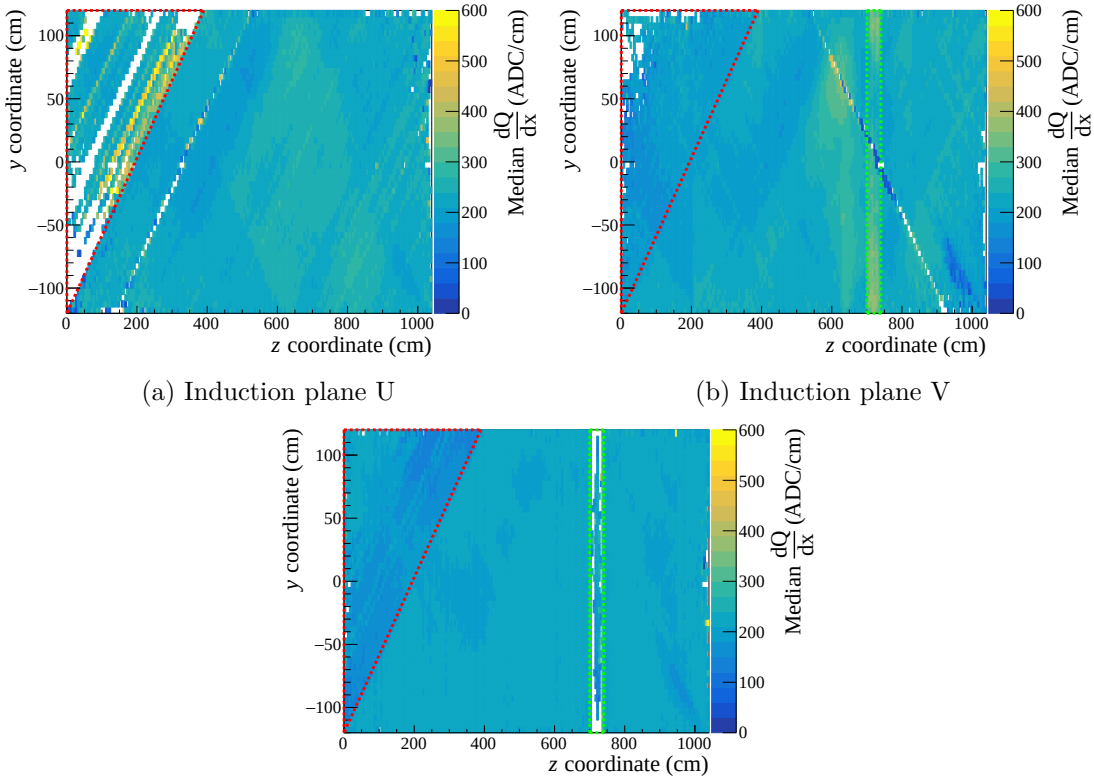
$$\left(\frac{dQ}{dx}\right)^{i,j}_{YZ_{\text{corr}}} = YZ^{i,j}_{\text{correction}} \left(\frac{dQ}{dx}\right)^{i,j}_{\text{SCE}}. \tag{4.5}$$

Here, a correction term $YZ_{\text{correction}}$ is applied to the charge per unit length, after SCE corrections. All terms in this equation contain two superindices, i and j, as they correspond to bins in a two-dimensional histogram. The detector is segmented in 5×5 cm^2 bins in the YZ plane, and the same selection of crossing muon tracks which was used for the lifetime correction is now used to fill a 2D histogram with the medians of the charge deposited per unit length, integrated over the x axis; this corresponds to the $(dQ/dx)^{i,j}_{\text{SCE}}$ term in equation (4.5) and we refer to these values as the *local medians*. We show these histograms for each of the three wire planes in figure 4.7. As a side note, it would be ideal to fully voxelise the detector in three dimensions to perform this correction more finely, but limited statistics prevent us from doing this [87], hence why we integrate over the x axis.

The YZ maps shown in figure 4.7 have empty regions with no data, which can be seen in white. These originate predominantly from shorted wires, such as in the upper left corner in the induction plane U, shown within a dashed red boundary in figure 4.7(a). It is believed that a large number of wires in this plane are in contact with one wire of plane V [67]. As a result of this, some wires in the U plane do not retain their bias, affecting the surrounding electric field, and in some cases collecting some of the charge before it reaches the other wire planes, which can be seen as an irregular region of high dQ/dx (yellow bins) within the same red-dashed boundary, and darkened regions of diminished charge collection in the same region in the other planes. Likewise, shorted wires in the collection plane can be seen as a vertical empty line around 700 cm in figure 4.7(c), shown as a boundary delimited by a dashed green line. In this case, the induction plane V collects the charge, resulting in a high dQ/dx region seen in figure 4.7(b).

Other features displayed in figure 4.7 are caused by misconfigured TPC channels and dead wires, as described in Ref. [67], all of which are well understood by the collaboration. Most importantly as we will see, they are consistent between Runs, meaning that since the initial event — early during commissioning — which caused wires to become shorted, the readout channels of the detector have remained stable. The main goal of this subsection is to establish whether this continues to be the case in Run 5, considering the issues related to argon purity observed during Run 4.

To obtain the correction term $YZ^{i,j}_{\text{correction}}$, a global median value of the dQ/dx histograms is obtained across the YZ maps in figure 4.7. For each plane, the local median dQ/dx values comprising the two-dimensional bins of the YZ maps are histogrammed,



(a) Induction plane U

(b) Induction plane V

(c) Collection plane

Figure 4.7.: Map of the median charge per unit length integrated in the x axis, as measured by the three planes of wires during Run 5, before YZ uniformity correction. Large sets of shorted wires in plane U (red dashed boundary in upper left corner) and the collection plane (green dashed boundary between $700 \leq z \leq 740$ cm) result in regions without data, and affect the electric field in surrounding regions, which can be seen in the other planes delimited by the same boundaries. Other minor features are due to misconfigured TPC channels and other dead wires.

with the result shown in figure 4.8. The histograms in figure 4.8 thus correspond to the values of the z axis in their counterparts in figure 4.7, so for example the tail observed between values of dQ/dx of 150 and 200 ADC/cm are due to the region of shorted wires in the upper left corner in figure 4.7(c).

The global median in each plane is the median of the histograms of local medians, shown as red dashed lines in figure 4.8. These medians are rounded to the nearest integer, which is sufficient for our purposes as we are prioritising stability during Runs and do not wish to convey a false sense of precision. We see in figure 4.8 that the global medians are not strongly biased due to the features observed in the histograms, although the maximum range has been set as 400 ADC/cm, which excludes the highest values seen as yellow bins in figure 4.7, as these are considered anomalous.

The values for the global medians of all three planes in Run 5 are shown in the last row of table 4.1, alongside the values of all Runs. In this table we see that after a dip in the values during Run 4, the global median dQ/dx has returned to values more consistent with previous Runs, which were not affected by purity issues. Finally, the correction term in bin i, j, $YZ^{i,j}_{\text{correction}}$ in equation (4.5) is defined as the global median divided by the local median in bin i, j.

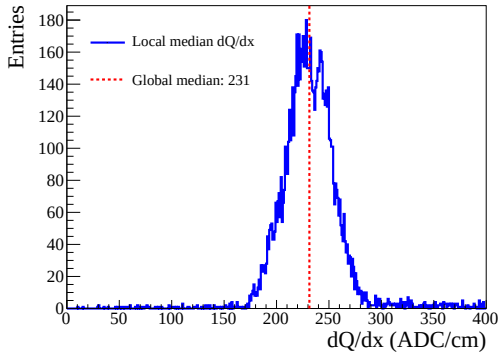
Table 4.1.: Global Median dQ/dx (in ADC/cm) for all runs. Values are stored to the nearest integer. We note a dip in the values of Run 4 compared to the rest of the runs, due to low purity.

	Plane 0	Plane 1	Plane 2
Run 1	235	223	210
Run 2	231	225	213
Run 3	234	222	209
Run 4	219	213	201
Run 5	231	222	215

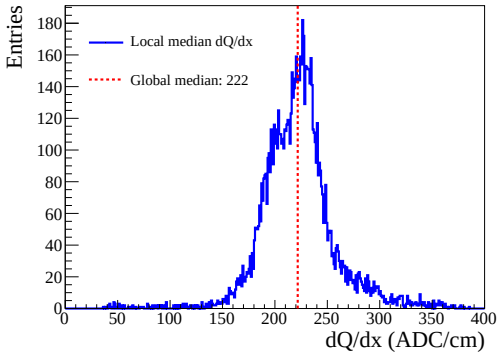
By following this procedure, the $YZ_{\text{correction}}$ two-dimensional histograms are stored and applied to all samples in MicroBooNE. We can make a further validation by comparing these $YZ_{\text{correction}}$ histograms among different runs by dividing their corrections bin-by-bin among any given two runs. We compare each plane of Run 5 to Run 2 in figure 4.9. The correction maps between runs are similar in every bin, for which these distributions have a narrow peak at 1. We subtract 1.0 from the mean value of these distributions to define a bias, which we show in table 4.2 for all three planes, comparing the Run 5 map to those of Runs 2 and 3. A bias under 2% in all cases makes this compatible with previous runs, and thus in practice we continue reusing the data map from previous runs. This also shows that the features observed in the histograms in figures 4.7 and 4.8 are consistent between these Runs.

4.2.5. Run 5 calibration

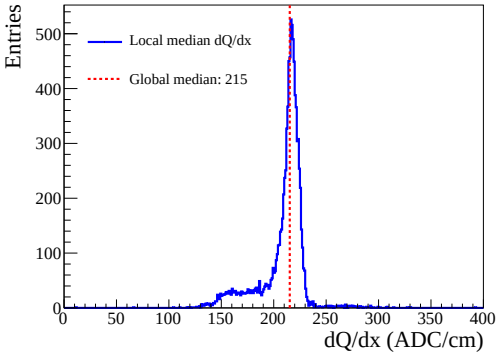
We extract the lifetime of Run 5 following the same procedure outlined in section 4.2.1. We find that the purity has returned to normal levels on average, with only a dip near



(a) Induction plane U



(b) Induction plane V

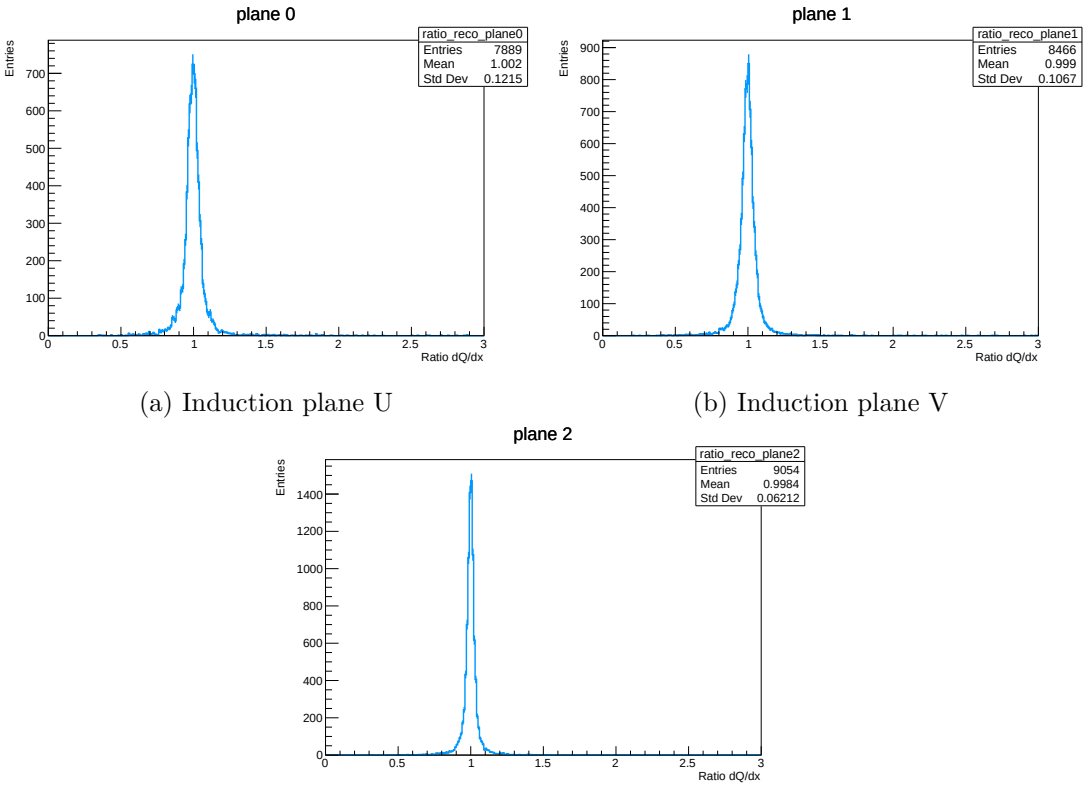


(c) Collection plane

Figure 4.8.: Local dQ/dx for the integrated 2D YZ maps for Run 5, before correction. The global median rounded to the nearest integer extracted from these histograms is shown as a red dashed line.

Table 4.2.: Bias extracted from the ratio of correction factors between two different runs. For example, the biases for Run 5/Run 2 are extracted from figure 4.9.

	Plane 0	Plane 1	Plane 2
Run 5/Run 2	+0.2%	−0.1%	−0.2%
Run 5/Run 3	−0.5%	−0.9%	+1.9%



(a) Induction plane U

(b) Induction plane V

(c) Collection plane

Figure 4.9.: Bin-to-bin ratio of the correction factors for runs 5 and 2, $YZ^{\text{Run 5}}_{\text{correction}} / YZ^{\text{Run 2}}_{\text{correction}}$, for all three wire planes. Small differences between these correction maps make the values differ from 1.0. A bias is extracted from the mean value (shown in the upper right corner).

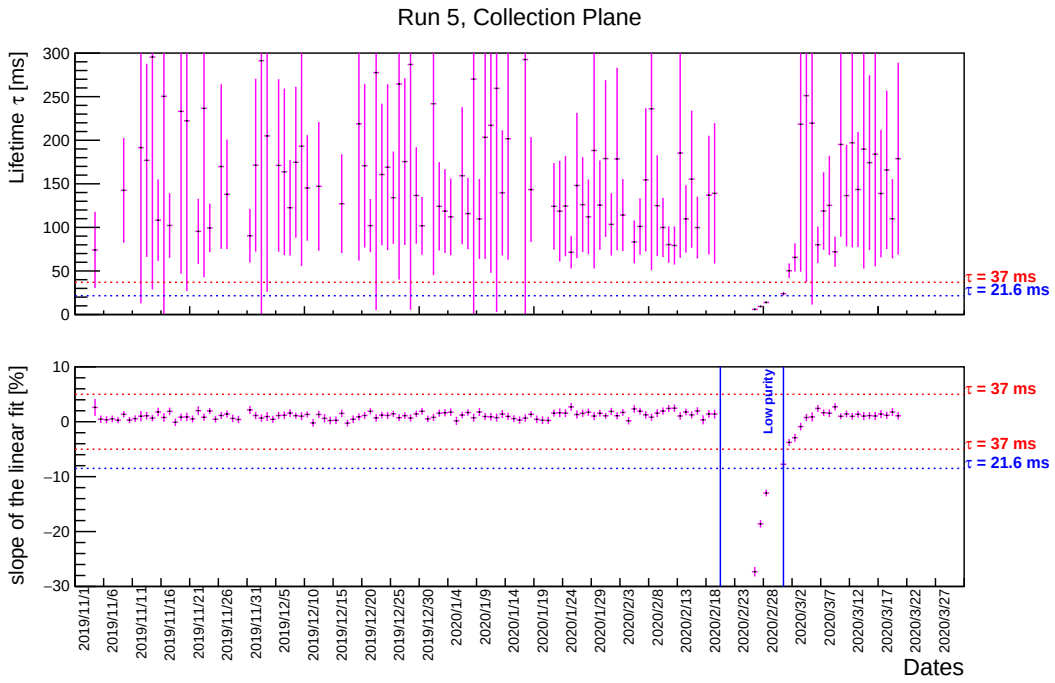


Figure 4.10.: Extracted values for electron lifetime during Run 5. Only a short period of low lifetime occurred near the end of February 2020, where the values of lifetime dip below the 37 ms horizontal red dashed line. The 21.6 ms horizontal blue dashed line which was the average lifetime during Run 4 is also shown for comparison. The period of low purity in Run 5 is shown delimited between vertical blue lines in the bottom figure.

the end of February and beginning of March, as shown in figure 4.10 in the bottom panel between vertical blue lines. The extracted values of the lifetime alongside their errors were then delivered to the database to be used by the collaboration.

4.3. Pilot study of an SBND commissioning measurement

In this section, we will describe a strategy we explored for performing a measurement of the electron lifetime in SBND during its commissioning phase. This strategy uses fast reconstruction products, and combines information from the time-projection chamber (TPC) and the cosmic-ray tagger (CRT) subsystems, which could also be used in studies other than electron lifetime measurement. The work described in this section is based solely on Monte Carlo simulations, and documents studies I carried out during a visit to Fermilab. While this strategy was ultimately not adopted by the collaboration, it served as a useful exercise in working with simplified reconstruction tools and data structures.

For commissioning, SBND prepared a file format containing a reduced set of variables containing the essential reconstruction information, which we will refer to as the *com-*

missioning trees. The aim was to use these computationally light files to perform quick measurements with reasonable accuracy for quick checks, without the need to process the early commissioning data through the full analysis chain. We will describe only the subset of variables which were used in our study of electron lifetime as we need them.

In a LArTPC the drift direction is correlated to both the x spatial direction and time, which presents some limitations. For example, the TPC subsystem by itself cannot differentiate between a hit occurring at x_1, t_1 and a second hit x_2, t_2 occurring at a later time $t_2 > t_1$ further downstream, $x_2 < x_1$, if their time difference coincides with the time it takes for the charge produced at the x_1 to reach x_2, that is $x_1 + v_d(t_2 - t_1) = x_2$. It is for this reason that different subsystems are used to provide timing, mainly the photon detection system (PDS). In our previous example, the scintillation light of hits 1 and 2 would travel orders of magnitude faster than the drift electrons, and unless the time difference $t_2 > t_1$ is in the nanosecond range, the time difference in arrival of photons measured by the PDS would be enough to resolve the hits in time. The PDS is essential for timing reconstruction in the case of neutrino interactions, which originate on the inside of the TPC. In the case of muons crossing the detector however, the CRT subsystem can also provide very accurate and easy to reconstruct spatial and timing information, and could be used in place of the PDS for this task.

There is one case in which a TPC track's timing can be known using only information from the TPC, and that is if the track crosses one of the anodes. In this case the drift time at the crossing of the anode end is zero, and a t_0 can be defined which can then be compared to the start of the readout window. Using anode crossing muons is in fact what is typically done in a LArTPC for calibration, as we have established for MicroBooNE, but there are some advantages to breaking free from this restriction. To make a lifetime measurement, a large number of geometrically appropriate anode-cathode crossing muons must be collected, each of which will only make a few hits in each respective x-bin. On the other hand, a single front-to-back crossing muon with a small θ_{XZ} angle will provide $\mathcal{O}(1000)$ hits in one of the x-bins. An online or semi-online electron lifetime measuring scheme could be devised in which the measurement is continuously updated, as incoming muons deposit hits in appropriate x bins, and stale ones are phased out, while making use of a wider variety of muon track geometries. This is the strategy we will explore in this section.

4.3.1. Commissioning trees

Instead of the usual hit reconstruction, which involves fitting a Gaussian function to a deconvolved waveform, the commissioning trees record what we will call a *fast hit*, which is obtained by integrating the area under the curve of the ADC of a raw waveform. A diagram of how a fast hit is constructed is shown in figure 4.11. A threshold of 10 ADC is set, the algorithm loops over the time ticks of the waveforms and a hit width is defined as a region above threshold, while the peak time of the region is stored. A tail 3 times the length of the width is added on each side. If the resulting tail is fewer than 100 ticks, this time length is used instead. The charge is obtained by summing the ADC values for each time tick between the end of the tails. The extension of the tails over

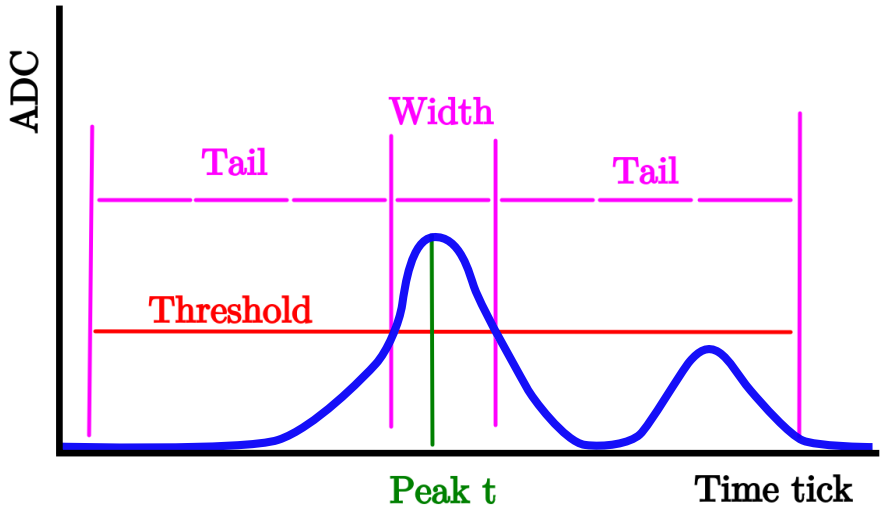


Figure 4.11.: Diagram of how a fast hit is reconstructed. A raw signal waveform (blue) in the collection wireplane is scanned for regions above a set ADC threshold (shown in red), and a fast hit is formed including long tails. The peak position is shown in green. In this specific example, the signal displays two clear peaks, in order to highlight one of the limitations of fast hit reconstruction.

a large window of time is done to capture the full charge even on highly diffused hits. Figure 4.11 also highlights one way in which this fast reconstruction may fail, by using a hardcoded tail length of triple the width above threshold, all activity within that wide region will be counted as part of the same hits, when in reality it may be due to multiple hits.

Fast hits give a reasonably good description of the charge deposited by a particle which crosses multiple wires in a straight line, such as a crossing muon, but will have some trouble differentiating two objects which are close to each other, such as the case of overlapping muon tracks or delta rays knocked out by the muon tracks (in which cases points with overlapping hits could be reconstructed as a single hit with double the charge). Similarly to regular hits, they are unsuited for reconstructing tracks which are parallel to the wires, or travelling directly towards them.

The main objective of fast hits is a fast reconstruction of muon tracks, which are widely used in calibration. The commissioning trees contain a *fast muon track* product, or simply *TPC track*. TPC Tracks are constructed using a Hough transform to find straight lines from consecutive fast hits in adjacent wires. Normally track reconstruction demands hit-coincidence in all three wire planes, but in the commissioning trees only a match between the collection plane (plane 2) and either of the induction planes (0 and 1) is considered sufficient, and the hits of only one induction plane are stored. An example of reconstructed TPC tracks and their constituent hits from a simulation of muons using `CORSIKA` [91] is shown in figure 4.12. We will do a more detailed explanation of the

simulation samples used when discussing the electron lifetime extraction in subsection 4.3.3.

The commissioning trees also contain information about the CRT subsystem. In particular they contain information about CRT hits and CRT tracks. The SBND CRT subsystem consists of 7 planes surrounding the 6 faces of the cryostat for full 4π coverage, plus an extra plane 2 m over the top plane to provide additional spatial resolution from the abundant comsic muons raining from above. Each plane is composed of an arrangement of modules, each of which is in turn composed of flat and thin $1\,\text{cm} \times 10\,\text{cm} \times 400\,\text{cm}$ strips of scintillating material arranged in rows. The strips have embedded in them two optical fibres running through their length, on opposing corners. See figure 4.13 for a diagram of the strips and module arrangement.

When a charged particle (for example a muon) crosses one of the strips, the scintillation light produced will be guided through the fibres towards silicon photomultipliers (SiPMs) attached on their ends. Using the relative time of arrival at each of the four SiPMs, the position along the length of the strip can be determined. CRT planes are formed of an arrangement of CRT modules covering a full face of the detector, and as a two-layer perpendicular stack such that strips run in perpendicular orientations. In this way, a time coincident hit in strips from two perpendicular modules can provide a *CRT hit*, which contains two-dimensional position information on the CRT plane, as well as timing. The placement position of the CRT plane itself provides the third coordinate, resulting in a CRT hit containing the full 3D position of a hit plus timing information.

Finally, the time of flight of relativistic muons crossing the distance between two CRT planes is known, therefore a *CRT track* is defined when two CRT hits happen within a sensible timeframe on different CRT planes. These are also stored in the commissioning trees as two endpoints in 3D coordinates from which a vector can be easily formed, plus the time at which the track happened as an average of the CRT hit time.

A summary of the information stored in the commissioning trees which has been discussed in this section and will be used in our study is shown in table 4.3.

Table 4.3.: Reconstruction products present in the SBND commissioning trees which will be used in our study.

Reconstruction product	Subsystem	Information
Fast hit	TPC	Fast estimate of deposited charge. Accurate y, z coordinates.
TPC track	TPC	Fast hits forming a straight line. θ_{XZ}, θ_{YZ}.
CRT hit	CRT	Accurate x, y, z, t information. from two perpendicular strips.
CRT track	CRT	Accurate endpoints x_1, y_1, z_1 and x_2, y_2, z_2 from two CRT hits.

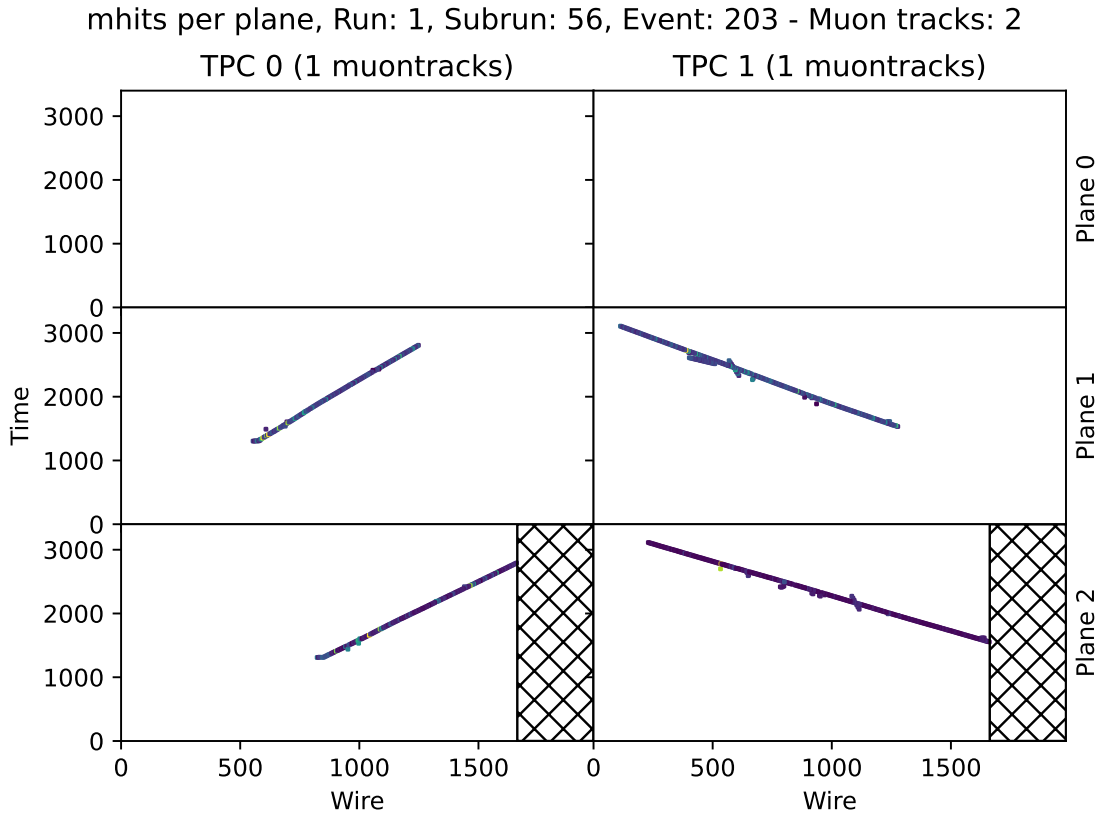


Figure 4.12.: Event display of reconstructed fast hits in the commissioning trees. The two SBND TPCs are shown in the left and right columns, and the three different planes in each TPC are in shown as rows. The collection plane in the bottom row has fewer wires than the induction planes, which is why it is drawn with a hatched area. The event shows two muon tracks, one in each TPC, both appearing in planes 1 and 2. The run, subrun and event information in CORSIKA simulations is used for bookkeeping, but does not relate to data-taking runs.

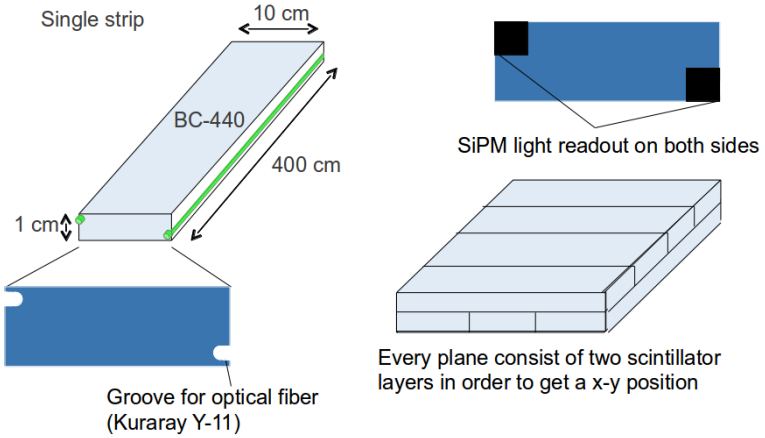


Figure 4.13.: The SBND CRT modules are made of an array of strips of scintillating material, with optical fibres running through them to guide the light towards silicon photomultipliers (SiPMs) placed at each end. Stacking the modules perpendicular to each other gives two-dimensional information on the position of a CRT hit. Figure from Ref. [92].

4.3.2. TPC-CRT matching for front-to-back crossing tracks

TPC tracks are stored with an arbitrary x position for their endpoints, and this needs to be corrected using information from matching CRT tracks. It was found that the x position stored for their endpoints was incorrect as a result of an error in the code, and was always recorded as being on one of the anodes for front-to-back crossing tracks. This results in every front-to-back crossing track having a θ_{XZ} value of zero if calculated with their endpoints. However, the $y-z$ information is stored correctly, and it was checked that the angular $(\theta_{XZ}^{\mathrm{TPC}}, \theta_{YZ}^{\mathrm{TPC}})$ information is recorded independently before the error in the endpoints occurred and was also verified as reliable.

For CRT tracks, a vector can be formed using its endpoints, from which angular information $(\theta_{XZ}^{\mathrm{CRT}}, \theta_{YZ}^{\mathrm{CRT}})$ can be obtained. To match TPC and CRT tracks, we compare their respective θ_{XZ} and θ_{YZ} angles; if both angles are within 2° of each other, we classify them as a potential match. To avoid incorrect matches from tracks which have the same angular orientation but are in a different position in the y–z plane, we evaluate y near the centre of the detector at $z = 200\,\mathrm{cm}$ for both tracks and see that they match within 2 cm. If both of these conditions are satisfied, we establish a match and save both tracks.

We then use the x information of the CRT-track to calculate the x correction needed for the TPC-track, as illustrated in the bottom panel in figure 4.14. We consider the middle z position of the TPC-track in the detector, z_{TPC}, which for front-back crossing tracks will be in the centre of the detector, then using the CRT-track endpoints in the x–z plane we find the x position of the CRT track, x_{CRT}, at the same z coordinate, which we define as z_{center} and for which $z_{\mathrm{TPC}} = z_{\mathrm{CRT}} = z_{\mathrm{center}}$. We assume that the TPC track should also pass through this point and use this information to correct the endpoints of

the TPC-track. The upstream and downstream corrected endpoints, (z_1, x_1) and (z_2, x_2) respectively, will be located at

$$x_1 = x_{\text{center}} - (z_{\text{center}} - z_1) \tan(\theta_{XZ}), \tag{4.6}$$

$$x_2 = x_{\text{center}} + (z_2 - z_{\text{center}}) \tan(\theta_{XZ}), \tag{4.7}$$

where the angle θ_{XZ} with respect to the beamline is shown in the bottom panel in figure 4.14. With this, we now store the endpoints as a new *corrected* TPC-track, alongside its matching CRT-track.

In figure 4.14 the uncorrected track in yellow is drawn using the endpoint information, thus the incorrect information in the x position can be seen as the track appears to be flat on the anode, whereas its XZ angle variable is recorded as non-zero and shown on the figure as $\theta_{XZ}^{\text{TPC}} = -4.20°$.

4.3.3. Electron lifetime extraction in SBND

The simulation samples contain muons generated using `CORSIKA`. A filter is applied to keep only events which contain at least one front-to-back crossing muon, but additional muons could be present in the event to simulate cosmic activity. The detector simulation includes the space-charge effect as well as diffusion, and samples with different values of the electron lifetime are generated, of 1, 3, 5, 7, 10 and 15 ms. The simulation followed the commissioning workflow, and the information stored corresponds to the fast products described in the previous subsections.

For our measurement we used around 200 tracks across all bins. We divided the x-coordinate of SBND into 30 equally sized bins, with 15 bins of 13.33 cm in each TPC, while filling all bins with enough tracks. As we will see, we were able to extract a good value of the dQ/dx distribution in bins containing as few as 3 tracks. Front-to-back crossing tracks with a narrow angle $\theta_{XZ} < 3.0°$ were selected, such that we can assign the full track to a single x-bin, that in which its x coordinate lies at the midpoint in the z coordinate. This results in some spillover of tracks to adjacent bins near the TPC edges, but this is considered acceptable for this proof-of-principle measurement. Using these single-x-bin tracks comes with the benefit of reducing any drift-induced biases upon their constituent hits. An example of the binning is shown in figure 4.15.

The charge distribution from the hits belonging to these tracks is then plotted and fitted. In MicroBooNE, the median value of the dQ/dx distribution was used to extract the lifetime as it is a simpler measurement which will yield good results using the high statistics of an offline measurement. As we are aiming towards an online measurement using more limited statistics, we will rely more heavily on the fact that the energy deposition of minimum ionizing particles should follow a Landau distribution. Furthermore, the readout error of the wires can be approximated as a Gaussian, and therefore we perform a fit of the convolution of a Landau function and a Gaussian centred around zero to model the measured deposited charge per unit length. Additionally, a second Gaussian fit is applied, centred around twice the dQ/dx of the most probable value, to account for doubly counted hits, although this second fit is not used in the final measurement

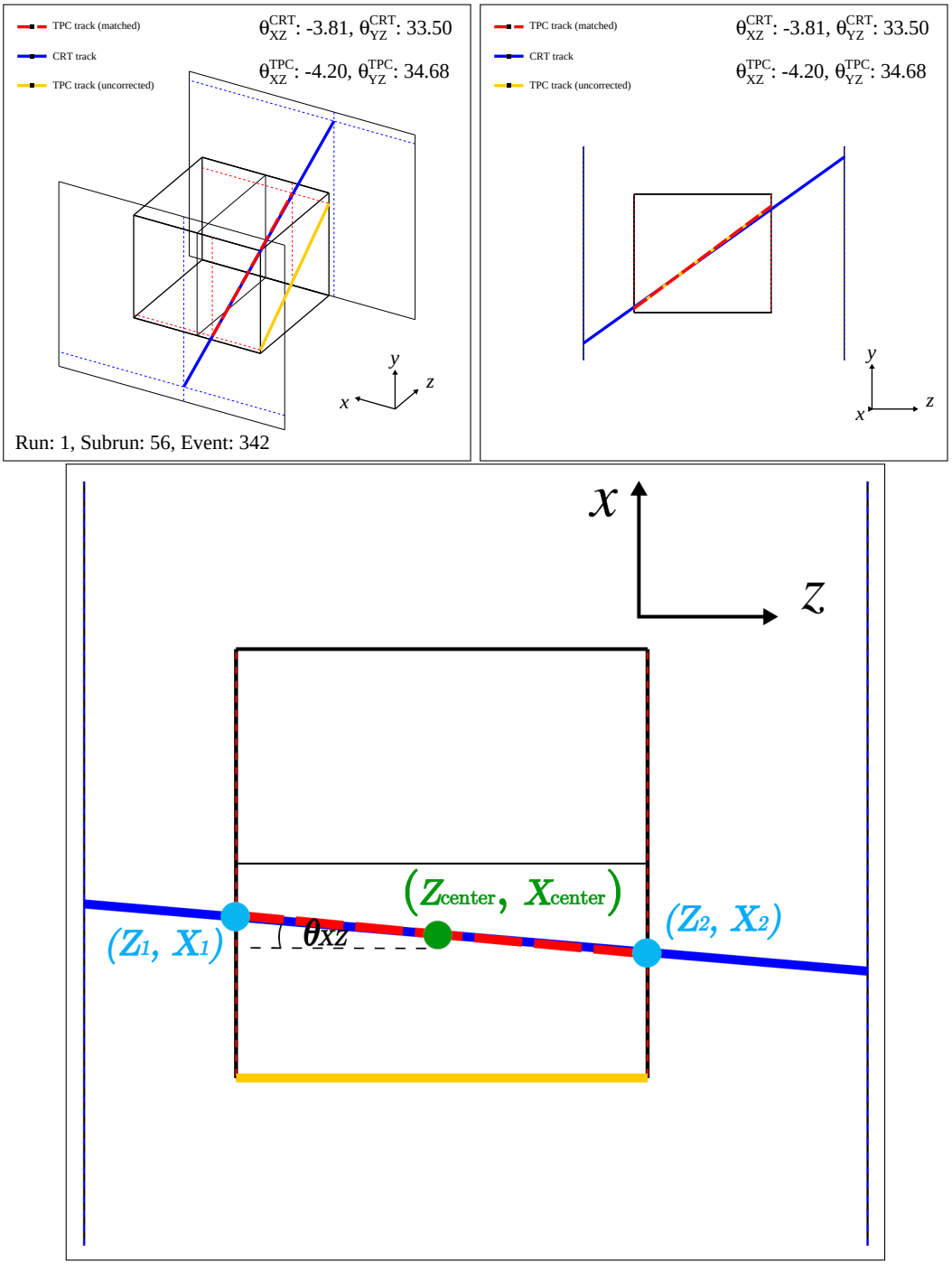


Figure 4.14.: TPC-CRT track matching and correction. The CRT track (blue) is matched to the corrected TPC track (red dashed line). The yellow line represents the uncorrected TPC track, whose angular information does not match its trajectory. Toy geometries of the SBND detector show the front and back CRT panels and the TPCs from three different perspectives. The bottom figure shows the trigonometry used to determine the points of entrance and exit to the TPC (z_1, x_1) and (z_2, x_2) from an anchored centre point $(z_{\text{center}}, x_{\text{center}})$ and the known angle θ_{XZ}.

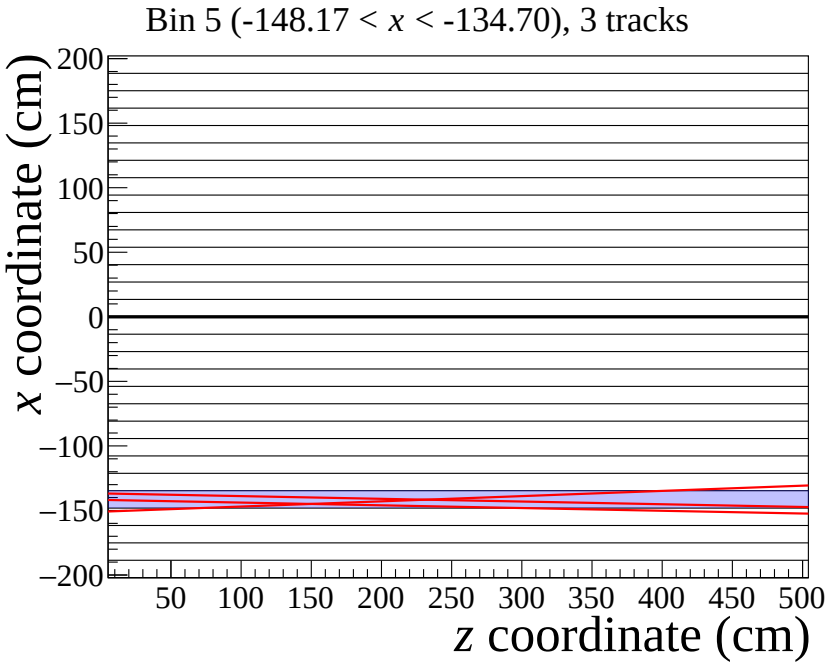


Figure 4.15.: Slicing of the SBND TPCs in the x coordinate into 15 bins each, for 30 bins total, as a diagram of the detector from a top view. The red lines show the trajectories of the 3 muon tracks crossing bin 5, which is highlighted as a purple shaded area.

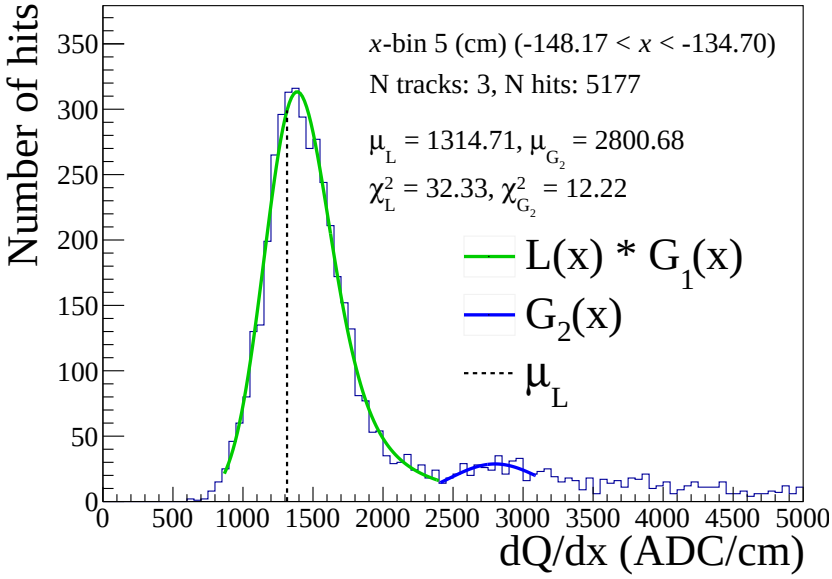
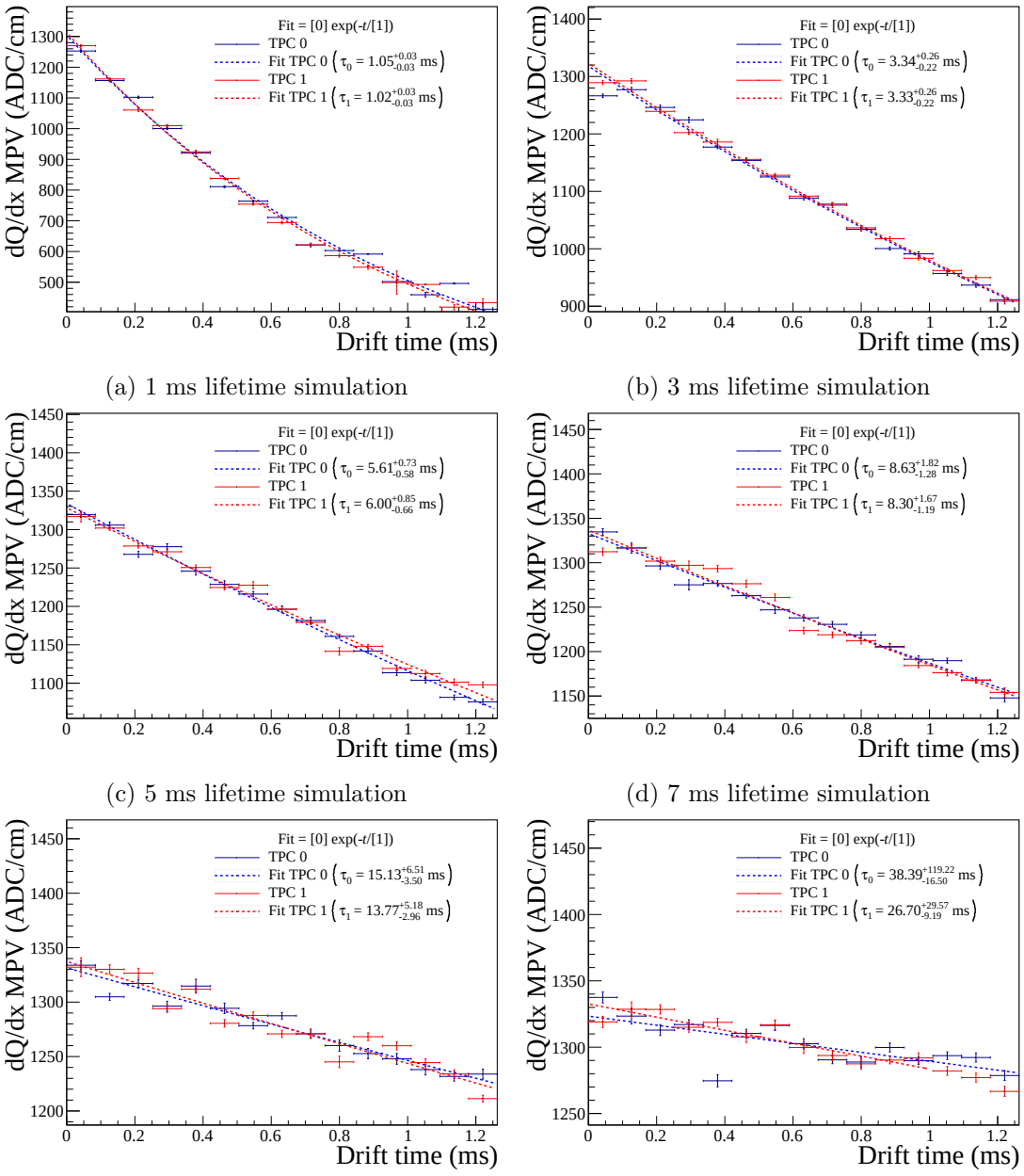


Figure 4.16.: Distribution of the charge per unit length for x-bin 5, which contains only 3 tracks. The fit of a Landau convolved with a Gaussian is shown in green. From this fit, we extract the most probable value μ_L. A second Gaussian fit (blue) centred at double the value of μ_L is also shown.

but could be used to subtract background. An example x-bin alongside its fits is shown in figure 4.16. In this figure, only 3 tracks were used to perform the fits, however given the geometry of front-to-back crossing hits, these 3 tracks provide over 5000 hits, from which good fits can be applied. The example shown in figure 4.16 is the bin with the lowest number of tracks in this sample, thus bins with more tracks have better statistics.

By performing these fits and extracting the most probable value μ_L of the convolved Landau and Gaussian, we follow the same procedure as in MicroBooNE and compile them onto histograms according to their drift time. We make the assumption of constant drift velocity v_d using the nominal SBND value of $160.0\,\mathrm{cm\,ms^{-1}}$, and use $v_d = x_d/t_d$ to make a conversion from our drift distance (x_d) slice histograms into drift time (t_d) sliced ones, taking into consideration that the drift time is zero at the cathode position $x = 0\,\mathrm{cm}$. Therefore, our 30 bin slicing of the detector along the drift coordinate becomes a 15 bin slicing in drift time. We obtain two lifetime values, one for each TPC, which are shown for all simulation samples in figure 4.17. Due to the strong non-linearity of the lifetime parameter, we use asymmetric uncertainties in its fit, estimated using the *MINOS* algorithm, as implemented in *ROOT* [93].

For every different lifetime simulation between 1 and 15 ms we see a good agreement between the lifetime values extracted in each TPC. We also observe a good agreement for low values of simulated lifetime compared to the extracted ($< 5\,\mathrm{ms}$). Figure 4.18 shows the extracted lifetime values with respect to the true simulated values in figure 4.18. In this figure we see a bias becoming evident, as the extracted lifetimes tend to overesti-



(a) 1 ms lifetime simulation

(b) 3 ms lifetime simulation

(c) 5 ms lifetime simulation

(d) 7 ms lifetime simulation

(e) 10 ms lifetime simulation

(f) 15 ms lifetime simulation

Figure 4.17.: Most probable values of dQ/dx from Landau-Gaussian fits per drift time for different simulated values of electron lifetime. Different fits (dashed lines) are applied to each of the two TPCs, shown in blue and red respectively. A value of electron lifetime is extracted from these fits, and shown in the legend.

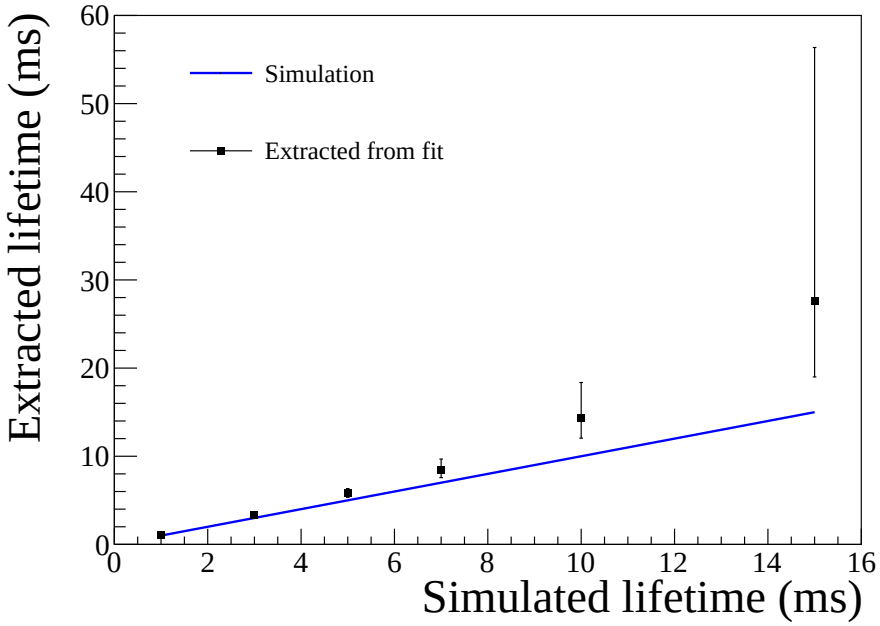


Figure 4.18.: The lifetimes extracted from simulation (markers) obtained by averaging the lifetimes obtained for each TPC. Blue line shows what 1:1 agreement between simulated and extracted would look like. For simulated lifetimes above 5 ms the extraction is biased towards overestimating the lifetime value.

mate the true simulated values. Such deviations are expected, as we are not applying corrections for the space charge effect or diffusion — both of which are present in the simulation — and these sources of systematic uncertainty are not reflected in the error bars.

4.4. Electron lifetime summary

In this chapter, we have presented a detailed study of the measurement and correction of electron lifetime effects in LArTPCs, which are closely tied to argon purity. For the MicroBooNE experiment, we outlined the full procedure for performing electron lifetime corrections based on the measured energy deposition per unit length of cosmic muons traversing the detector. We also examined the factors that led to a reduction in electron lifetime during MicroBooNE's Run 4 data-taking period, from September 2018 to July 2019. An electron lifetime of 37 ms is sufficient to avoid degrading data quality in MicroBooNE, given its maximum electron drift time of 2.3 ms — this condition was met in most runs during Runs 1–3. However, during most of Run 4, the electron lifetime fluctuated around 21.6 ms.

The work presented in section 4.2 is not part of any previous publication. Ref. [87] presents results using data from Runs 1–3, where electron lifetime is mentioned, but

the substantial efforts undertaken by the collaboration to make Run 4 data usable are not addressed. More recently, MicroBooNE published its first results incorporating data from all five of its Runs [45], an achievement made possible by the corrections detailed in this chapter. Notably, the lifetime correction applied in MicroBooNE represents the largest-scale electron lifetime correction implemented in any LArTPC experiment to date. Although the millicharged particle search presented in the rest of this thesis uses only data from Runs 1–3, the full dataset will be used in an upcoming publication, which is currently in preparation.

For SBND, we have shown in section 4.3 a pilot study designed to measure the electron lifetime quickly, in the context of the detector's commissioning. This study uses reconstruction tools with simplified pattern recognition algorithms, intended to enable a rapid measurement that could, for example, inform operations in the case of a "bad argon delivery", such as occurred in MicroBooNE. We have shown that we can match tracks detected in the CRT and TPC subsystems, bypassing the need to involve the PDS. This is very valuable as CRT reconstruction is computationally fast and can be used to trigger the detector and store events to a separate stream for analysis. We have also shown that using front-to-back crossing tracks, as few as 3 tracks in a bin can result in sufficient hit statistics for measuring the charge-per-unit-length. Using fewer tracks would enable this to be an online or semi-online measurement, as opposed to an offline measurement as described in MicroBooNE.

The SBND pilot study presented serves as a proof of concept, but many of its parameters would need to be optimised for full implementation, for example, the geometry of the crossing tracks. If diagonally crossing muon tracks, which deposit fewer hits per bin, were used instead of front-to-back tracks, fewer tracks could still provide a sufficient electron lifetime estimate. The binning was also not optimised and was chosen based on our simulation sample. A full study would need to consider the rate of crossing muons detected during normal operations and select appropriate track geometries and binnings that allow for a reasonably good measurement of the electron lifetime.

SBND is now beyond its commissioning phase and is fully operational, collecting a large amount of neutrino interaction data. Its electron lifetime is periodically measured offline, using calibration tools based on robust and sophisticated reconstruction frameworks. These measurements also take into account transport effects, such as diffusion, that were neglected in this chapter, but do not yet incorporate the SCE. The measured lifetime is generally high, with only occasional short dips following disruptions in the cryostat.

Part II.

Search for millicharged particles in the MicroBooNE detector

5. Simulation and reconstruction

In this chapter, we will describe our simulation of millicharged particles (mCP) and the scatter electrons they produce as they cross the detector. We will then give a general overview of particle reconstruction in MicroBooNE before discussing the reconstruction of our scattered electron signal in particular. We will also briefly discuss the neutrino background simulation.

In MicroBooNE, both simulation and reconstruction are integrated in the `LArSoft` framework [94, 95]. Starting with a bespoke mCP event generator, which takes as input the NuMI beam flux, we simulate the interactions of the mCP with electrons in the liquid argon volume. We then follow the standard MicroBooNE reconstruction workflow for Monte Carlo simulations, including the overlaying of beam-off cosmic background data.

The structure of this chapter closely follows the same order in which simulation and reconstruction chain steps are applied. In section 5.1 we describe the simulation of the neutral meson flux from the NuMI beam towards MicroBooNE. In section 5.2 we discuss how the protons-on-target (POT) normalisation discussed in section 3.3.4 is realised in our Monte Carlo signal simulation, which we record as an *unweighted* set of events. Then, in section 5.3 we describe the generation of the mCPs from the decay of the neutral meson flux, as well as some general characteristics of their kinematics. In section 5.4 we describe our simulation of electron scatters.

We will search for mCPs in two topologically distinct signal channels, one with multiple electron scatters of order $\mathcal{O}(1\,\text{MeV})$, and one with a single-electron of $\mathcal{O}(10\,\text{MeV})$. These two topologies will present only a few differences in kinematics at the simulation stage, but the tools which will be used in their reconstruction will be vastly different.

We describe the general reconstruction methods in MicroBooNE in section 5.6, in particular the *BlipReco* [19] tools we use for the low-energy multiple-scatter search, and the *Wire-Cell* framework [96, 97] we use for the high-energy single scatter. We then show the application of these reconstruction tools to our signal simulation in sections 5.7 and 5.8. We will then briefly describe the main backgrounds in the search in section 5.9. We end this chapter with a description of all the data and Monte Carlo simulation samples that are used in the searches in section 5.10.

5.1. Neutral meson flux

An extensive sample of Monte Carlo (MC) simulation generated using the `g4numi` [98] package is used to model the NuMI beam induced fluxes in MicroBooNE. The `g4numi` package is based on `Geant4` [99], and aims to simulate full hadronic production in the

NuMI beamline. It has been used by multiple Fermilab experiments downstream of the NuMI beam, such as MINERνA and NOνA. Neutrino analyses typically use the charged meson flux as input to the `GENIE` event generator [100], however we are interested in the neutral meson flux, as discussed in chapter 2.3.

The output flux files store information of the produced mesons, such as their type, and kinematic information at the time of production. For generating mCPs we are limited by the available flux files, which do not store information about ω or ϕ mesons, so these channels are not available to us. For the same reason, we also do not consider the decays of heavy neutral mesons such as J/ψ or Υ. This will limit our sensitivity to mCPs above 200 MeV, but not in a way that will change the position of our results relative to the that of other experiments with higher sensitivity.

A weight is applied to each event in order to constrain the proton-on-carbon interactions in the target based on data from the NA49 and MIPP experiments, as described in Ref. [98]. This "Flux weight" w_{Flux} is stored independently of POT normalisation, and we will detail how we account for it and other weight factors in what follows.

5.2. Monte Carlo POT normalisation

MicroBooNE normalises its data and simulation to the protons-on-target (POT) delivered by the neutrino beam, in this case NuMI. Each `g4numi` flux file has 500,000 entries representing physics events, and in a naive scenario we would assume this corresponds to the same number of POT. However, event generators often use weight factors to take into account the probabilities of a given interaction occurring, and the real number of POT is obtained by summing the weights across all events. Our goal, however, will be to generate unweighted events and use POT normalisation to account for all probabilities. This will make it easier to compare our generated mCP signal to other MicroBooNE samples, such as data or neutrino simulations.

In our signal, we have three weight factors to consider. First, the flux weight we have already mentioned, which corrects the number of simulated mesons. Second, a branching ratio weight, which accounts for the probability of a meson decaying into a mCP. Finally, a scatter weight, which is the probability that a mCP crossing the detector will generate n electron scatters above a given minimum recoil energy E_r^{min}. The latter two weights will be discussed in future subsections.

The event generator will loop through the events in the meson flux and "attempt" to generate a millicharged particle which crosses the MicroBooNE detector. This attempt may fail if there is no suitable meson in this event, for example if the mCP fails to cross the detector, or due to kinematic constraints in the decay or the scatters. Once a suitable event is found in the flux, the millicharged particle and electron scatters are generated. The kinematics of the meson, mCP, and electron scatters will determine a total event weight w_{total} for the event given by

$$w^e_{\text{total}} = w^e_{\text{Flux}} w^e_{\text{BR}} w^e_{n\text{-scatters}}, \tag{5.1}$$

where w_{Flux}, w_{BR}, and $w_{n\text{-scatters}}$ are the flux, branching ratio, and scatter weights

respectively, and the superscript e reminds us that this is a per-event weight which depends on the kinematics of the meson, mCP and scatter electrons.

In order to obtain unweighted events, we use rejection sampling — a technique by which a target distribution can be reproduced by probabilistically accepting or rejecting weighted events. For this, we need a value which is larger than any given w^e_{total}, which we call the maximum weight $w_{\max}$ (the value of which is to be determined). For every mCP event we generate, we also generate a random number u uniformly between 0 and $w_{\max}$. We then compare the event weight w^e_{total} to u, and if $u \geq w^e_{\text{total}}$, we reject the event e.

As we loop through the entries in the flux file, we keep only the events which have been not been rejected, either due to kinematic constraints or our sampling, and the POT counting will be given by

$$\text{MC POT} = \frac{\text{Entries looped through}}{w_{\max}}. \tag{5.2}$$

All the weight information has been encoded in this normalisation. Events which have a high total weight have a proportionally high probability of not being rejected. This means fewer entries would need to be looped through to accept them, keeping the MC POT low. The opposite is true for low w^e_{total} events.

The only remaining task is to find a value for $w_{\max}$. This is done by generating a large MC simulation sample which has the same configuration parameters but is not subject to rejection sampling. In this case, we do store the weight information, and we record the value of $w_{\max}$ as the highest value of w^e_{total} in all the events, plus an additional 10 % for safety. Choosing a value of $w_{\max}$ that is too high will result in more events failing rejection sampling, which means an inefficient simulation. Choosing a value that is too low will result in incorrect POT accounting, which is why the event generator has been configured to return an error if it encounters an event where $w^e_{\text{total}} \geq w_{\max}$.

The configurable parameters of our event generator are the mCP coupling factor ϵ, the mCP mass m_χ, the minimum electron recoil energy $E_r^{\min}$ and the minimum number of scatters to generate n_{scatters}. Generating a new sample only to obtain a value for $w_{\max}$ may seem like an expensive extra step, but this task is computationally very light as it does not require the entire detector simulation, only the event generation step. This step also has to be run only once, as the distribution of the total weight per event is completely determined by the chosen parameters, therefore the same value of $w_{\max}(\epsilon, m_\chi, E_r^{\min}, n_{\text{scatters}})$ can be reused when generating a new MC sample with the same parameters. Examples of MC samples used to obtain values for the maximum weight are shown in figure 5.1.

The MC POT shown in equation (5.2) is a number which can be used as a scaling correction. Our mCP sample S will be normalised to a certain number of beam-on data POT as

$$\hat{S} = S \times \frac{\text{Data POT}}{\text{MC POT}}, \tag{5.3}$$

where we use the hat notation $\hat{S}$ to represent the data-normalised sample. So for example, if we want to search for millicharged particles in a beam-on data sample equivalent

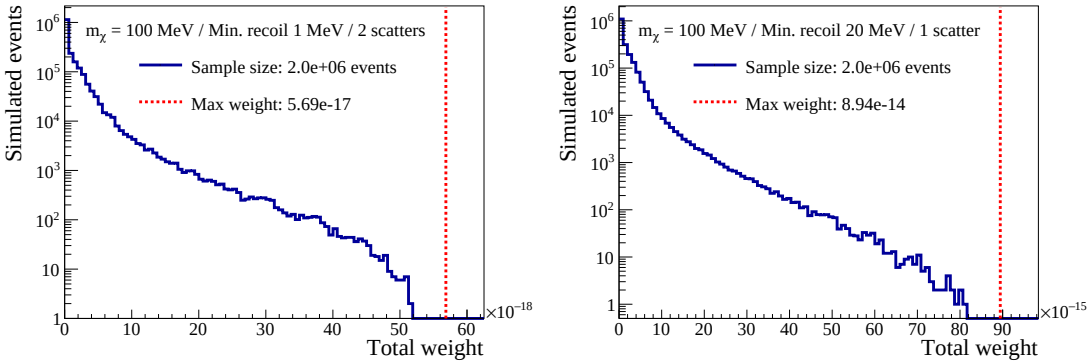


Figure 5.1.: MC samples used to obtain the maximum weight in two different event generator configurations. Left shows a two-scatter sample with a low $E_r^{\min}$ of 1 MeV, while right shows single-scatter events with a higher $E_r^{\min}$ of 20 MeV. Both samples contain 2 million events, and are of a mCP mass m_χ of 100 MeV and coupling ϵ of 0.001. The maximum weight value selected is shown as a dashed red line, which is a factor 1.1 above the highest weight of any event in the sample.

to 1.0×10^{21} POT, and our simulation sample has a POT of 2.0×10^{26} POT, we scale our MC sample by a factor of $1.0 \times 10^{21} / 2.0 \times 10^{26} = 5 \times 10^{-6}$. An equivalent POT accounting strategy is used in the simulation of neutrinos, although the method for obtaining unweighted events may be different, and we will not discuss them in this thesis.

Finally, we note that the branching ratio weight w_{BR} will always contain a factor of ϵ^2, as it originates from the branching ratio equations (2.3) and (2.4). The scatter weight $w_{n\text{-scatters}}$ will have a factor of ϵ^2 for each electron scatter generated, as it involves the sampling of the cross-section equation (2.16). Therefore, the normalised sample will be proportional to a factor $\epsilon^{2+2n_{\text{scatters}}}$ which is contained in $w_{\max}$ in equation (5.2),

$$\hat{S} \propto \epsilon^{2+2n_{\text{scatters}}}. \tag{5.4}$$

Factoring ϵ out of the sample in this manner allows us to use a single simulation of millicharges with a parameter ϵ_0 which can then be easily rescaled. For any two given normalised samples $\hat{S}$ and $\hat{S}_0$ which differ only in ϵ [1],

$$\frac{\hat{S}(\epsilon, m_\chi, E_r^{\min}, n_{\text{scatters}})}{\epsilon^{2+2n_{\text{scatters}}}} = \frac{\hat{S}_0(\epsilon_0, m_\chi, E_r^{\min}, n_{\text{scatters}})}{\epsilon_0^{2+2n_{\text{scatters}}}} \tag{5.5}$$

from which we obtain

$$\epsilon = \epsilon_0 \left(\frac{\hat{S}}{\hat{S}_0} \right)^{\frac{1}{2+2n_{\text{scatters}}}}. \tag{5.6}$$

[1] This assumes that the samples have sufficiently high statistics. Equalities among independent MC simulation samples can be considered an abuse of notation.

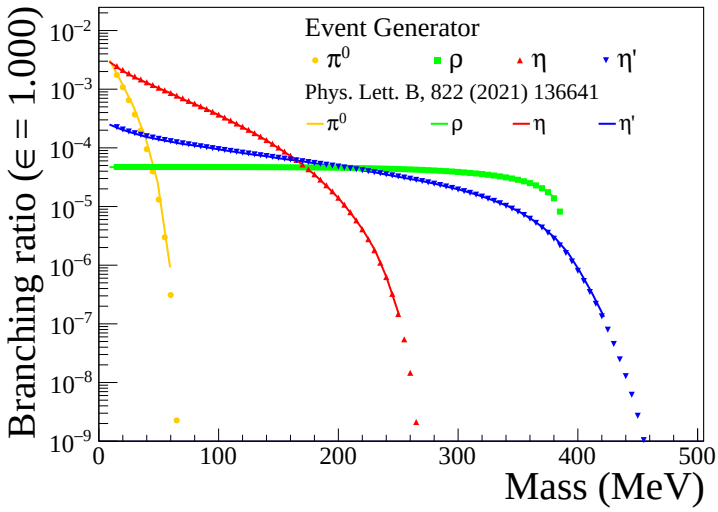


Figure 5.2.: Branching ratios used for the decay of neutral mesons into mCPs. Lines show a digitised version of the data points in Ref. [20], while markers show our implementation in the event generator. Agreement shows that our implementation of the equations in Ref. [20] is correct.

This means that instead of sweeping the 2-dimensional parameter space (m_χ, ϵ) with different MC simulation samples, we only have to simulate different mCP masses m_χ, which drastically reduces computation time.

5.3. Millicharged particle simulation

We produce mCP pairs $\chi\bar{\chi}$ from the decay of neutral mesons as described by the branching ratios described in equations (2.3) and (2.4), depending on the parent meson, which determines the weight w_{BR}. The branching ratios from our simulation can be seen in figure 5.2, where we compare the values in our event generator to those of Ref. [20].

The light neutral mesons which we consider decay promptly via an electromagnetic process (as opposed to charged mesons or heavy neutral mesons, which decay via a weak process) with a lifetime $\tau < 1 \times 10^{-17}\,\mathrm{s}$, thus the travel distance can be assumed to be zero, and we set the initial position of the mCP pair to the position of the neutral meson. Most of the neutral mesons from the flux are generated in the target itself, which means that most of the mCPs will be as well. This means that there will be little variation in their incidence angles θ and ϕ on the MicroBooNE detector. A diagram of the MicroBooNE coordinate system along with these angles and an incident mCP can be seen in figure 5.3.

The $\chi\bar{\chi}$ pair is generated in its rest frame as a back-to-back pair. We use polar coordinates (θ, ϕ) and take random numbers from a flat distribution for $\cos(\theta)$ between $(-1, 1)$

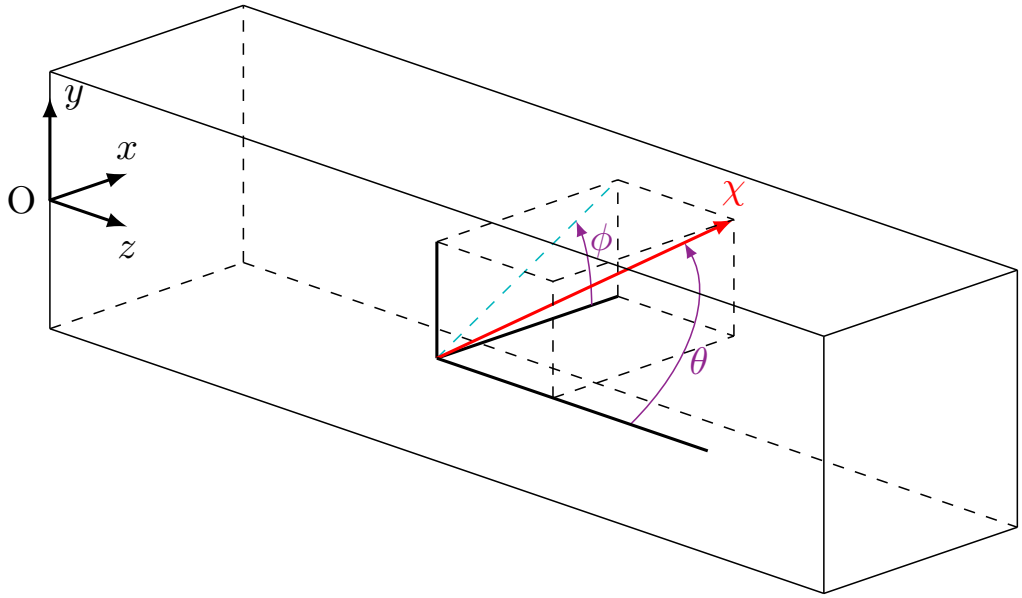


Figure 5.3.: MicroBooNE uses a Cartesian coordinate system, whose origin is shown in the figure as O, with the z coordinate on-axis with the BNB and x increasing against the drift direction. Most millicharged particles (χ, in red) coming from the NuMI target would enter the MicroBooNE detector through the anode face ($-x$) and exit on the cathode side ($+x$). We define a spherical coordinate system from the Cartesian coordinates as usual (in purple), with θ the angle of the particle with respect to the z axis, and the ϕ azimuth angle in the x–y plane, starting from $+x$. The detector volume is to scale, but the typical mCP direction has been altered to better display the angles.

and for ϕ between $(0, 2\pi)$ and give the pair an orientation (in Cartesian coordinates) as

$$\vec{p_1} = -\vec{p_2} = p^{\mathrm{CM}} \left(\cos(\phi)\sin(\theta), \sin(\phi)\sin(\theta), \cos(\theta)\right), \tag{5.7}$$

where p^{CM} is the momentum magnitude, which is obtained from the energy and mass of each mCP as $p^{\mathrm{CM}} = \sqrt{(E_\chi^{\mathrm{CM}})^2 - m_\chi^2}$. Here we have been using the superscript CM to emphasize that we are in the centre of mass frame of the mCP pair. The energy E_χ^{CM} is generated differently for the cases of two- and three-body decay. For two body decay, we simply take the energy as half the mass of the parent neutral meson M as $E_\chi^{\mathrm{CM}} = M/2$.

For three-body decay, we proceed as two consecutive two-body decays. First, we split the available phase space between the virtual photon which will give rise to the mCP pair, and a third particle Y (a real photon) which we will not be interested in. The invariant mass of the virtual photon q^2 is distributed according to the phase space suppression factor, which is the term inside the integral in equation (2.4), and is generated as a random value between $4m_\chi^2$ and M_P^2, the limits of the integrals where M_P is of the parent pseudo-scalar neutral meson. Ref. [20] also includes the case of decays of vector mesons, which in our ranges of energy would be ω and ϕ mesons. Although we do not include those decays in the search, their decays have been implemented in the event generator in case we obtain access to a reliable flux in the future. In these decays, the outgoing third particle Y can have a mass M_Y, as shown in figure 2.5. As its mass and that of the virtual photon are different in general, the momentum magnitude is obtained using the more general K\"{a}ll\'{e}n function λ [101],

$$|\vec{p_Y}| = |\vec{q}| = \frac{1}{2M}\sqrt{\lambda(M^2, M_Y^2, q^2)}, \tag{5.8}$$

where M is the mass of the parent meson (either vector or pseudoscalar). The energy then follows from the energy-momentum equation, and the orientation of these particles is sampled randomly as in (5.7). Finally, both particles are boosted back from the decaying meson center-of-mass frame into the laboratory frame. For pseudoscalar meson decays, the "third particle" Y is a (real) photon with zero mass, therefore it can be considered a special case.

We then use the 4-momentum of the generated virtual photon to produce a pair of mCPs in the rest frame of the virtual photon. The available energy is given by the rest mass $\sqrt{q^2}$, which, due to the interval in which we sampled, will always be bigger than the $2m_\chi$ we need. This is a simple two-body decay in which each mCP gets half of the energy, the momentum magnitude is obtained from the energy-momentum relation, and the particle orientation is uniformly generated according to equation (5.7). Finally the mCPs are boosted back into the laboratory frame.

Initially, we generated $2 \times 6 \times 95$ MC samples corresponding to different configurations of the horn current (FHC and RHC), minimum recoil thresholds (0.1, 0.5, 1.0, 5, 10 and 20 MeV), and millicharged particle masses between 5 and 475 MeV every 5 MeV. However, carrying out this many samples through the entire detector simulation chain is computationally expensive, hence we will reduce the number of samples by using only FHC, selecting two minimum recoil thresholds, and 12 mass points for a total of $1 \times 2 \times 12$ MC samples.

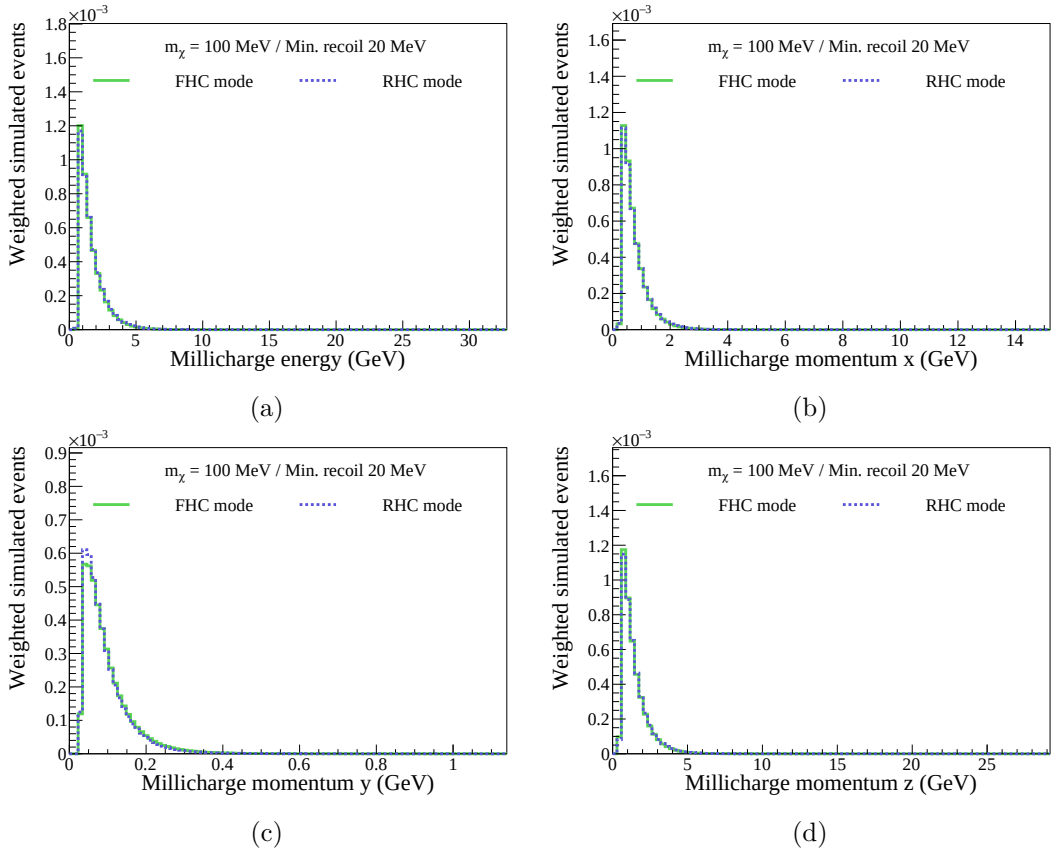
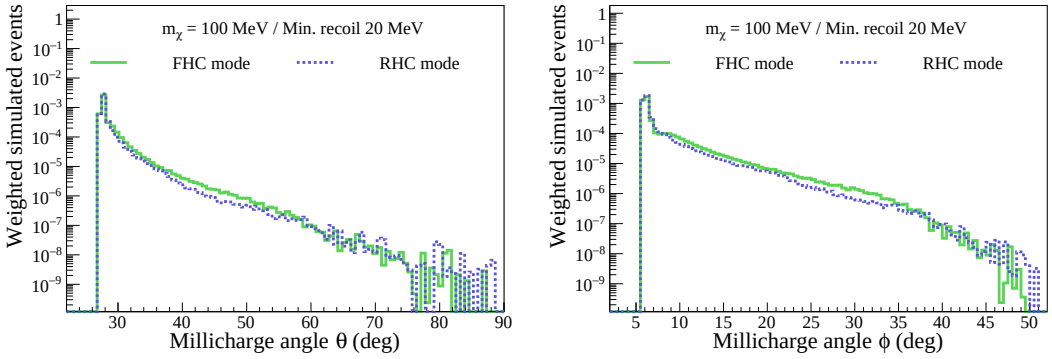


Figure 5.4.: Comparison between FHC and RHC modes, showing (a) Energy and (b-d) momentum distributions for mCPs of mass m_χ of 100 MeV and 1 detector interaction.

5.3.1. Horn current comparison

The energy and momentum distribution for a sample of mCPs of mass 100 MeV with a minimum recoil energy of 1 MeV is shown in figure 5.4. We see no significant differences between the forward and reverse horn currents (FHC and RHC). This is expected, as the different currents are designed to focus the on-axis flux of charged pions. Our decays are from neutral mesons, which are not affected by the horn current. The deflection of mCP by the magnetic horns would, in principle, be suppressed by a factor ϵ. However, we do not simulate this effect, as the mCP flux originates from the decay of neutral mesons, which happens predominantly at the target — which is located upstream from the focusing horns — or in the NuMI absorber, well downstream from them. In both cases, the horns are not expected to cause significant deflection.

More important than the energies and momenta in Cartesian coordinates, we are interested in the angles of incidence of the mCP, which will be close to the angles of the



(a) (b)

Figure 5.5.: Comparison between FHC and RHC modes for a mCP signal of mass 100 MeV showing the angles (a) θ and (b) ϕ in the MicroBooNE coordinate system.

vector pointing from the NuMI target towards MicroBooNE, given by

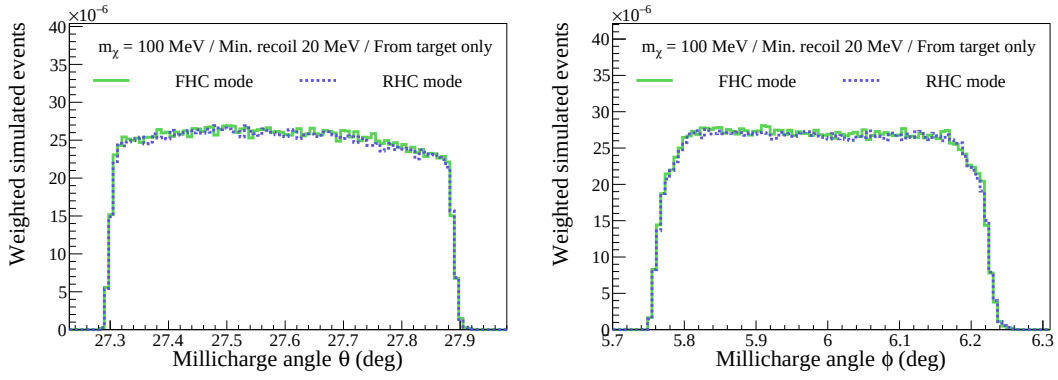
$$\theta_{\text{NuMI}} = 27.46°, \qquad \phi_{\text{NuMI}} = 6.20°. \tag{5.9}$$

In figure 5.5 the angular orientation of our mCP simulation in MicroBooNE coordinates is shown. As expected, the angle values are highly peaked around the nominal values. The off-peak angle values are due to the decay of mesons along the decay pipe. By applying a selection on the meson decay position within a box of dimensions 250 cm × 16 cm × 500 cm centred on the target, we obtain the distribution shown in figure 5.6. We note that this box is significantly larger than the target itself, and the selection is simply intended to isolate decays far from the target region. In figure 5.6 we observe that the horizontal axis scale has drastically changed, with no events outside these narrow bands less than one degree wide.

Our multiple low-energy scatter channel heavily depends on this angular information; therefore, we will perform a selection on the angle variables, as will be discussed in section 6.4.2. We do not apply such a selection in the single-scatter channel. However, since the majority of the mCP flux originates near the target, we can neglect contributions from the absorber. From this point on, we use simulations with forward horn current only.

5.3.2. Millicharge mass comparison

Figure 5.7 shows a comparison of the millicharge energy for different values of the mCP mass m_χ, for minimum recoil energies of 1 and 20 MeV. The differences between the mass points are only due to kinematics, which affect the number of events via three different mechanisms. The decay of mesons into mCP pairs becomes inaccessible once the sum of the mass of the mCP pairs $2m_\chi$ goes above the mass of the parent meson. The phase-space suppression factor also depends on the ratio of the mass of the mCP



(a) (b)

Figure 5.6.: Comparison between FHC and RHC modes for a mCP signal of mass 100 MeV showing the angles (a) θ and (b) ϕ in the MicroBooNE coordinate system after a selection based performed based on the decay position of the meson originating the mCP.

compared to its parent meson. Both of these effects can be appreciated on the vertical axis of figure 5.7, as a gradual decrease in the overall number of events for increasing m_χ.

The third mechanism is the fact that the minimum recoil energy limits the available phase space when generating a viable mCP which can produce scatters. This is very noticeable in the case where the minimum recoil energy is set to $E_r^{\text{min}} = 20\,\text{MeV}$, appearing as sharp cutoffs at low mCP energies (as can be seen on the horizontal axis in figure 5.7 for increasing values of m_χ. Let us remember that there is a maximal kinematically allowed recoil energy, given by equation (2.17),

$$E_r^{\text{max}} = \frac{\left(E_\chi^2 - m_\chi^2\right) m_e}{m_\chi^2 + 2E_\chi m_e + m_e^2}.$$

According to this equation, for a mCP mass of $m_\chi = 400\,\text{MeV}$, the maximum recoil energy E_r^{max} is above 20 MeV (and thus above E_r^{min}) only when the mCP energy E_χ is above 2.55 GeV, which is the cutoff point we see in figure 5.7. This severely limits our sensitivity to higher values of m_χ when using a high threshold for the recoil energy, as we would need very high energy mesons to cross this cutoff point. This effect is much less noticeable in the left figure in 5.7 which uses a minimum recoil of 1 MeV which we will use in the low energy scatters channel.

5.4. Electron scatter simulation

We now simulate the recoil electrons, which are the visible particles that we will reconstruct. We will look for recoil electrons in two different channels. In one channel, we look for single electron scatters with a relatively high energy of the order of 10 MeV. In

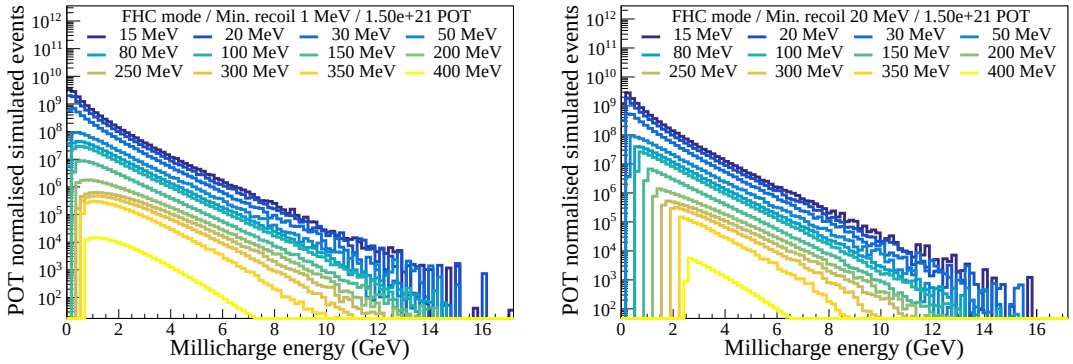


Figure 5.7.: mCP energy for different mCP mass simulations in FHC mode, for minimum recoil energies of (left) 1 MeV and (right) 20 MeV. All signals are normalised to a POT of 1.5×10^{21}.

a second channel, we look for two lower-energy scatters (order of 1 MeV) aligned with the detector target. To aid this, we can select the number of scatters, and their energy threshold in our event generator.

The minimum value for the generated electron energy $E_r^{\min}$ is a very important aspect. As discussed in chapter 2, equation (2.18) shows that the cross-section of the detector interaction scales as $1/E_r^{\min}$, which means the recoil electrons favour low energies. However, there is no clear-cut value for the minimum recoil energy $E_r^{\min}$ from a physics standpoint. The reconstruction methods which we will use in our two channels are very different, but generally there is a decreasing efficiency for lower recoil energies. Therefore, if we set $E_r^{\min}$ too low, we will be simulating recoils which have a very low reconstruction efficiency, and will have to generate more simulations to obtain a Monte Carlo sample with sufficient statistics; if we set the threshold too high, we will not correctly model the interaction of mCPs in our detector, and as we have seen, we limit the phase space for high mCP masses. So our general plan will be to set it as low as possible within computation considerations, and taking into consideration the reconstruction efficiency, for which we will discuss it in the corresponding reconstruction section of each channel.

Recoil electrons are generated in a uniform distribution across the length L which the mCP crosses in the detector volume. The detector time of the electrons is calculated using information of the time of flight of the mCP. For the recoil energy for the electron, we sample a value of E_r according to the distribution of the differential cross-section (2.15). The sampling is done between the values of $E_r^{\min}$ (which we set as parameter of the simulation) and a maximum recoil energy $E_r^{\max}$ given by kinematics (see eq. (2.17)). We obtain the distribution by using the rejection sampling method on the differential cross-section distribution.

For the low-energy two-scatter channel, a histogram of the resulting energies for different $E_r^{\min}$ low-threshold values of 100, 500 and 1000 keV can be seen in figure 5.8, for a mCP mass of 100 MeV. As the differential cross section depends on the mCP energy E_χ,

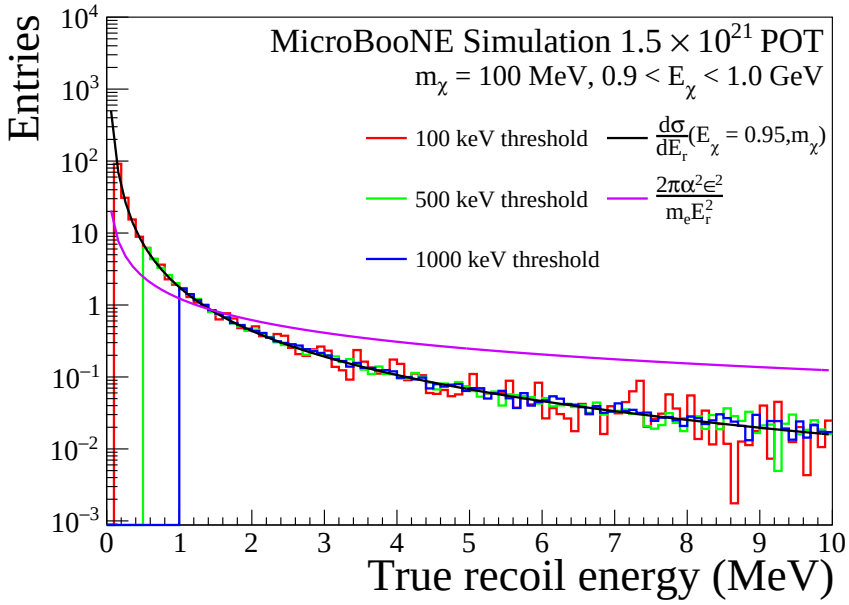


Figure 5.8.: Recoil energy of the scattered electrons for a selection of mCP energy for different low-threshold configurations, alongside the full differential cross section function (black) and its relativistic approximation (violet). The cut-off on the left side of the histograms shows the effect of threshold selection.

figure 5.8 shows only a subset of the generated sample for which $0.9 < E_\chi < 1.0\,\text{GeV}$. We also show the relativistic approximation from eq. (2.20), and see that it is not a good fit, even if at $E_\chi = 10m_\chi$ we should be well into the relativistic regime. For this reason, we will not use the relativistic approximation in the following, as our full simulation of the kinematics proves to be a better fit.

The resulting energy of the electron is $m_e + E_r$, and its momentum magnitude $p_e = \sqrt{(m_e + E_r)^2 - m_e^2}$. We should mention that unlike in the previous particle generation which we have described, this time we are already in the lab frame and there is no need for boosting. Finally, the orientation of the electrons is generated with an angle φ with respect to the mCP, as was shown in equation (2.14),

$$\cos(\varphi) = \frac{E_r(E_\chi + m_e)}{\sqrt{E_r(E_r + 2m_e)(E_\chi^2 - m_\chi^2)}}.$$

We can see the angular dependence on the minimum recoil energy in figure 5.9 where the generated scatter angle is shown for two different values of E_r^{min} of 1 and 20 MeV. We see that the harder momentum transfer of 20 MeV forces the electrons to be more forward.

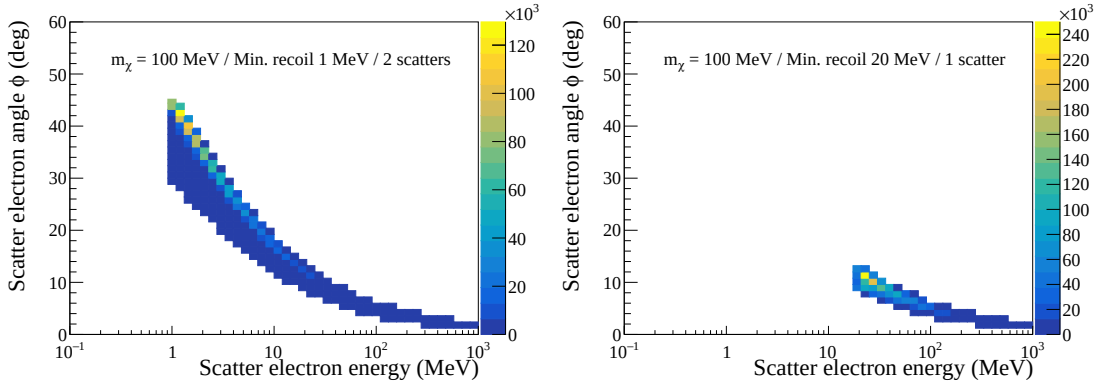


Figure 5.9.: Angle of the generated scatter electrons with respect to the mCP direction for minimum recoil energies of (left) 1 MeV and (right) 20 MeV.

5.4.1. Probability of multiple hits in the detector

As we have mentioned, the number of generated electron scatters is a configurable parameter of the event generator, which we have set to either one or two scatters. However, we still would like to consider the cases of 3, 4, etc. scatters, even if they have a low probability of being realised. The probability of generating n scatters corresponds to the scatter weight w_{scatter}, and more scatters results in a lower weight.

Let us remember that the probability of n-hits in the detector is given by a Poisson distribution we have discussed in equation (2.22),

$$p_{n\text{-hits}}(n, L/\lambda) = \frac{(L/\lambda)^n}{n!} e^{-L/\lambda}. \tag{5.10}$$

This probability is now weighted against the probability of generating $n+1$ hits using rejection sampling over the cumulative distributions for a given L/λ,

$$p_{(n+1)\text{-hits}}(n+1, L/\lambda) = \frac{(L/\lambda)^{n+1}}{(n+1)!} e^{-L/\lambda}. \tag{5.11}$$

In both these equations, L is the length of the detector which is crossed by the mCP, and λ the mean free path, which, as shown in equation (2.21), depends on the cross-section, which in turn depends on the mCP energy and mass. Therefore, the probability of generating n scatters is calculated event by event given the geometry and kinematics of the mCP, with a weighted random chance of increasing the number of scatters from n to $n+1$. The process by which the number of scatters is increased is iterated, for a chance to increase from $n+1$ to $n+2$ scatters, and so on. In practice, the probability of such cases is very low.

5.5. Data driven cosmic backgrounds

As a surface detector, events in MicroBooNE typically contain an important component of cosmic muons. Thus, a realistic simulation of signal must also include this activity.

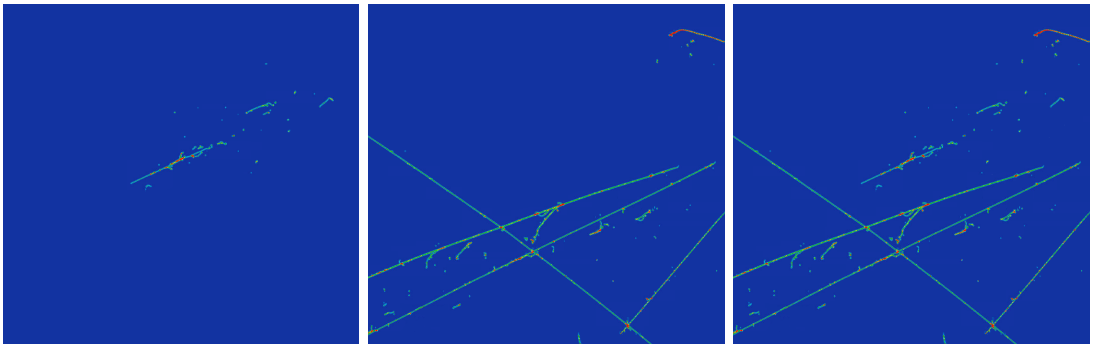


Figure 5.10.: Overlay of MC simulation and beam-off data. On the left, we show the MC simulation of a 30 MeV electron scatter. The middle figure shows beam-off data due to cosmic muons. On the right, we have overlaid the MC signal with the beam-off data at the waveform level. All images show the collection plane, with wires on the horizontal axis and time on the vertical axis. In this toy example used for illustration, the images have been manually edited.

Rather than using Monte Carlo simulations which may introduce additional systematic uncertainties, MicroBooNE uses events with an unbiased trigger — i.e. not subject to 9.5 photo-electron threshold described in section 3.2.3 — which are outside of the beam spill window, and thus contain only cosmic activity and electronic noise. These beam-off events are then mixed with the MC signal before the detector simulation stage, as exemplified in figure 5.10. In MicroBooNE this process of mixing MC simulation and beam-off data as well as the resulting sample are referred to as "overlay".

After the overlay stage, subsequent stages apply the same reconstruction methods for MC simulations and data. In what follows, when talking about the generated MC signal specifically, we will refer to it by the commonly used term "truth". During the entire analysis chain, LArSoft keeps an internal record of which reconstruction products are matched (even in part) to truth information, so in the example shown in figure 5.10 after reconstruction the shower would ideally be "truth matched" while the cosmic muon tracks would not. This could be impacted by partial truth matches; for example, a hit could contain depositions from both truth and beam-off data if these are in close proximity. Such a hit could end up as part of a track or shower, where the rest of the constituent hits are either from truth or beam-off data. For this reason, the fraction of truth-deposited energy to reconstructed energy is often used to classify a product as truth, and it is ultimately up to the analysers to decide what fraction of truth energy is sufficient for this. We briefly discussed this as part of the closure test on Monte Carlo study of electron lifetime calibration, in section 4.2.3.

5.6. Reconstruction in MicroBooNE

We have presented the implementation of the theoretical model for mCPs in a newly developed event generator written by Dr. Pawel Guzowski for the LArSoft framework. The "end product" of our simulation will be either one or multiple isolated electron scatters generated along the trajectory of the mCP inside the detector, which can then be reconstructed.

MicroBooNE, as most other LArTPCs, is designed to be sensitive to neutrino interactions in the energy range of $\mathcal{O}(100\,\mathrm{MeV})$, which result in tracks and showers that are reconstructed by high-level pattern recognition frameworks. In this section we will discuss reconstruction in MicroBooNE in general, starting from the low-level reconstruction of ionisation electrons and scintillation light, and finishing with a description of the high-level Wire-Cell framework, which reconstructs entire neutrino interactions. We will also discuss MeV-scale reconstruction, which is a relatively recent development in LArTPC reconstruction, in which small point-like isolated depositions known as *blips* provide information on an energy scale which is outside of the scope of track and shower reconstruction algorithms. We will then apply these reconstruction tools to our electron scatters in their respective energy scales.

5.6.1. Raw waveforms, noise filtering and 2D deconvolution

As ionisation charge travels near the induction plane wires towards the collection plane, a bipolar signal is induced, as they travel towards, and then away from them. The charge is then collected in the collection plane wires, where a unipolar signal is registered. These signals, known as *raw waveforms*, go through a number of stages to filter the noise due to the TPC and PMT electronics [67].

The raw waveforms are effectively a convolution of the distribution of the ionisation electrons, the electric field response to their propagation, and the electronics response [102]. For calorimetry, we make a conversion between charge and energy, and as charge is dictated by the number of ionisation electrons, we need to extract the number of electrons by applying a deconvolution to the raw waveforms.

Instead of a 1-dimensional deconvolution on the signal of a single wire per plane, MicroBooNE performs a 2-dimensional deconvolution, which also considers neighbouring wires. A comparison of the resulting waveforms after noise filtering, and applying 1D and 2D deconvolution, is shown in figure 5.11.

5.6.2. Hits and clusters

The basic building blocks for practically all high-level reconstruction of TPC activity are *hits*. Hits represent the passage of discretised ionisation electrons in close proximity to the readout wires. They are formed by fitting Gaussian functions to the deconvolved waveforms, with each hit corresponding to one fitted Gaussian. An example from the ArgoNeuT LArTPC can be seen in figure 5.12.

Hits exist on a wire, and have a time tick position and a width σ, also measured in

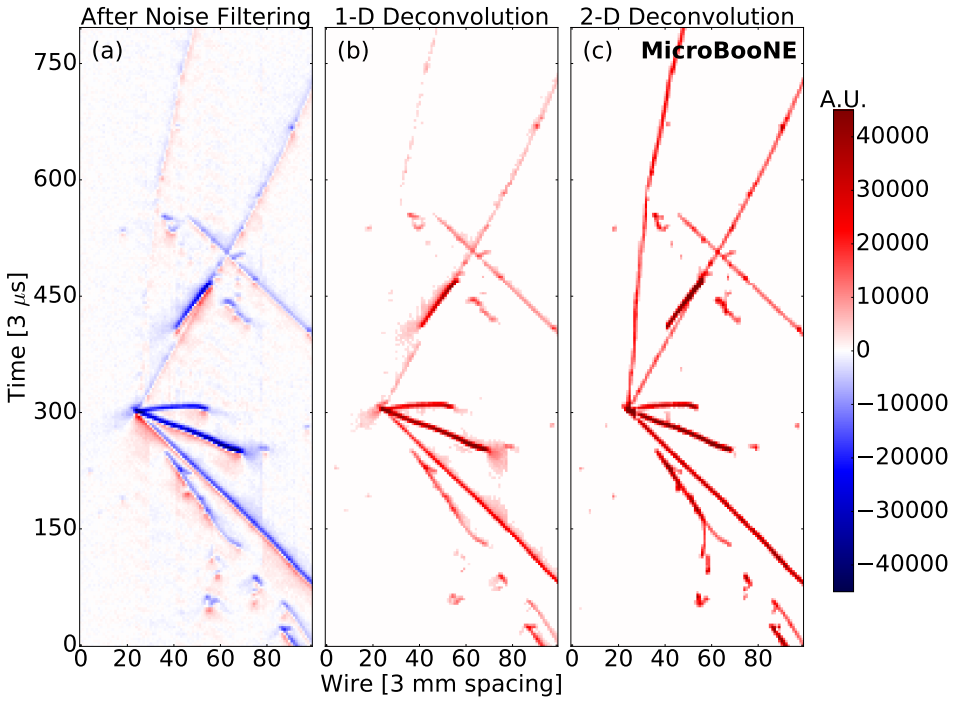


Figure 5.11.: A neutrino candidate event is shown (a) after noise filtering, (b) after 1D deconvolution, (c) after 2D deconvolution. The horizontal axis represents the wire number on the "U" induction plane, the vertical axis shows the time steps of 3 µs. The colour scale shows amplitude on an arbitrary unit scale. The non-deconvolved waveform in (a) is still a bipolar signal which alternates between negative and positive values. Figure from Ref. [102].

time ticks, given by the parameters of the fitted Gaussian. The number of electrons is obtained from a calibration factor applied to the measured ADC in the waveform. Then the area under the curve corresponds to the charge Q, which is proportional to the number of electrons. The number of electrons will ultimately be converted to energy after calibration.

The waveforms are not completely flat even in the absence of ionisation electrons, due to irreducible electronics noise. For this reason, hits must have a lower reconstruction threshold. In MicroBooNE, this threshold is determined based on several criteria: the signal must be above the root mean squared noise by 2σ the induction planes and above 3σ in the collection plane; the peak amplitude must have a minimum of 1000 electrons; and the mean amplitude of the hit must be of at least 500 electrons.

As can be seen in figure 5.12 it is common for hits to be clustered together. Clusters are reconstructed independently in each of the three wireplanes, although clusters in one plane without a matching counterpart in one of the others are considered to be due to electronic noise and discarded.

5.6.3. Light reconstruction

The scintillation light produced by interactions in the detector travels much faster than the drifting of ionisation electrons. For this reason, the reconstructed information of light signals from the photomultiplier tubes (PMTs) is crucial for precise timing of interactions.

As mentioned in section 3.2.2, the light readout from the 32 PMTs records light-waveforms independently of the TPC system. The time and number of photoelectrons (PE) in an interaction is reconstructed offline from the deconvolved PMT waveforms. A *flash* is constructed when a group of PMTs records signals within a short window of time, as shown in figure 5.13. Flashes reveal information of the position of the interaction, given the relative PEs recorded in each PMT, whose relative position is known.

5.6.4. The Wire-Cell reconstruction framework

Wire-Cell [96, 97] is an event reconstruction framework developed for MicroBooNE based on 3D tomographic imaging. In a nutshell, Wire-Cell uses the deconvolved charge depositions which arrive at the readout planes in a narrow 2 µs time window to define a 2D time-slice "image", which is in effect a tomographic cross-section of the detector. By concatenating the time-slices over the full readout time window, a full 3D reconstruction of the detector activity is generated.

The most fundamental units in Wire-Cell reconstruction are called *cells*, which are determined by the 2D wire overlap geometry in the detector in each time-slice. A diagram of this geometry is shown in figure 5.14, in which an individual cell is shown as a black triangle. In the scenario depicted, signals are found in 8 U wires, 5 V wires, and 6 Y wires. There are 55 cells contained in the region of overlap of these wires, which is a larger number of "unknowns" than the "known" information from the 19 wires. This

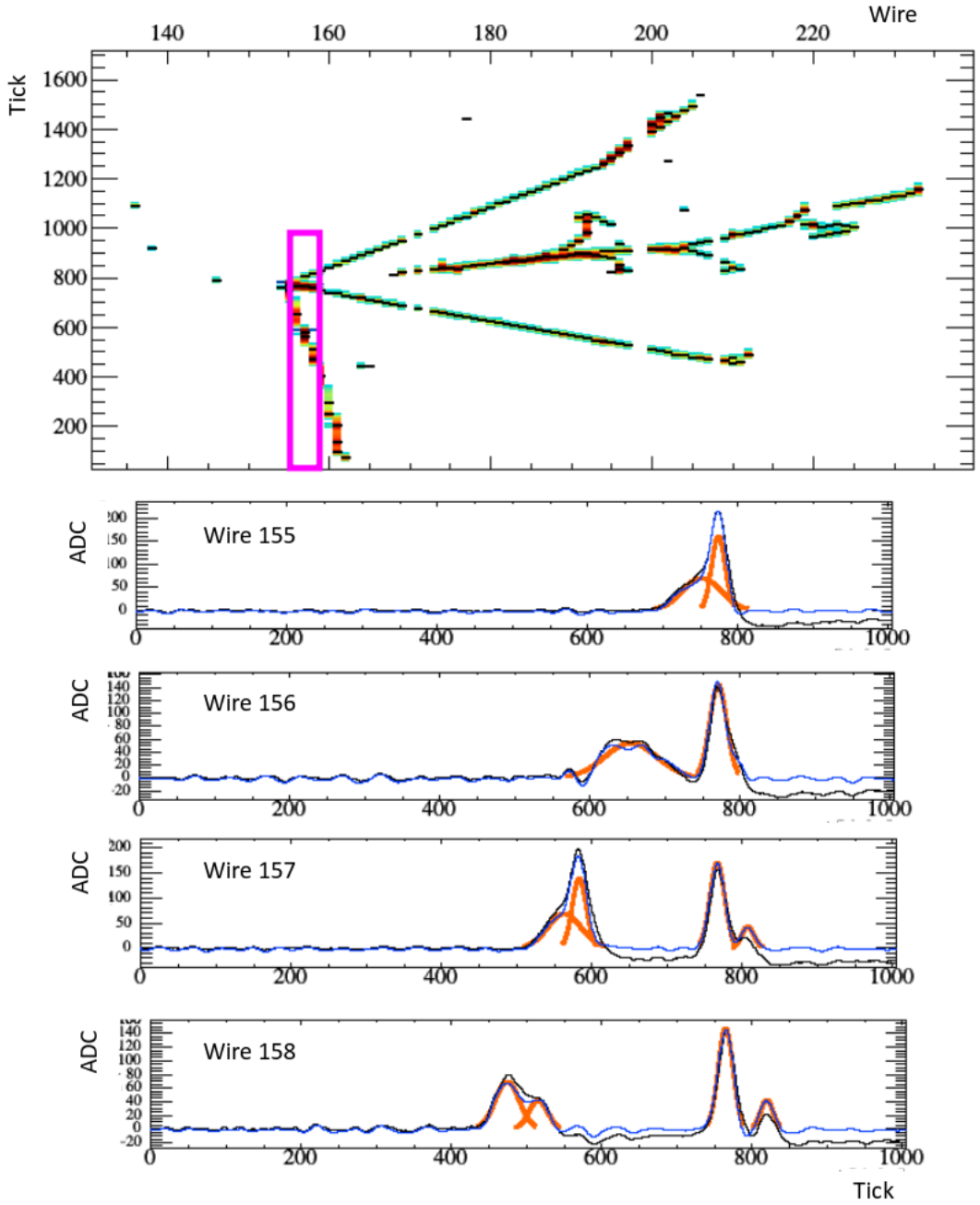


Figure 5.12.: Event display from the ArgoNeuT detector of a neutrino interaction showcasing the reconstruction of hits. In the topmost figure, a neutrino interaction is shown in which the horizontal axis is the consecutive wires of the LArTPC, and the vertical axis is the readout time. The four figures at the bottom show the waveforms ("raw" in black, and deconvolved in blue) from four different wires (vertical slices enclosed by a pink contour in the topmost figure). Gaussian fits to the waveforms are shown in orange, each of which is considered a hit. Figure from Ref. [103].

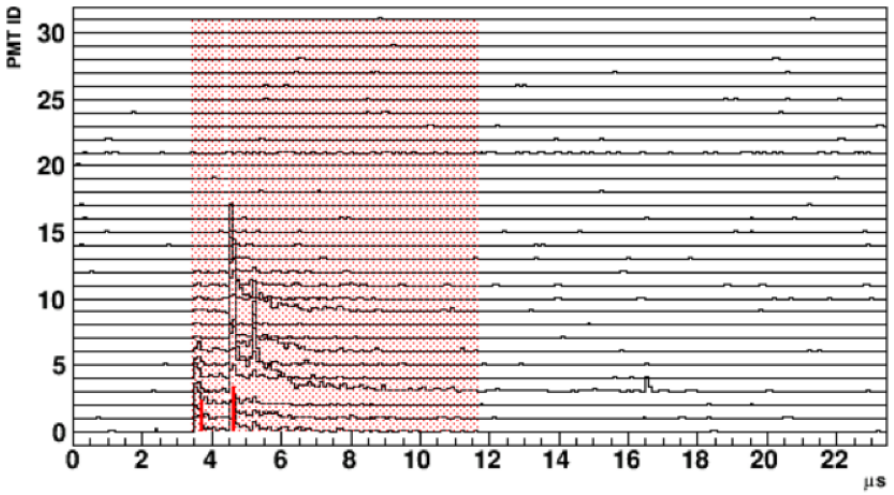


Figure 5.13.: Light flash reconstruction. The deconvolved readout from the 32 PMTs is shown as black lines. In this example, two flashes, indicated by vertical red lines, initiate subsequent flash windows (red shaded area). Figure from Ref. [97].

represents a fundamental ambiguity introduced by having only three readout planes, as opposed to a fully pixelated readout.

To mitigate this readout ambiguity, the entire region of overlapping wires is treated as a unit and its cells merged in a *tiling* procedure. The merged cells are referred to as a *blob*, shown as a blue contour in figure 5.14, and the continuous set of wires in each plane is referred as a *wire bundle* (hence a blob is also defined by the overlap of three wire bundles). This tiling reduces the number of variables fed into a following solving step, in which a system of linear equations is constructed to solve the charge, which has been integrated for the entire time-slice, into the respective blobs in which physical activity has occurred.

The ambiguity in readout introduces a number of fake blobs, referred to as *ghosts*, in which no true physical charge is present. The fact that there are many regions in the detector with non-functional wires in which the tiling has to be done with only two wire bundles significantly increases the number of ghosts in each event. In an ideal scenario, the system of linear equations would assign its true charge to each "real" blob, and zero charge to ghosts. However, the system of equations has no unique solution, and is treated as an optimisation problem instead. This charge solving makes use of additional constraints such as the sparsity of the activity in the LArTPC, and an approximation technique known as *compressed lensing* [104] to reduce computation time.

A further de-ghosting step is performed, now considering the concatenated time-slice information in 3D space. The remaining ghosts are usually present in areas where one wire planes is non-functional. The effects of this step are shown in figure 5.15. The entire step process described (tiling, charge solving, and de-ghosting) is iterative, with the number of iterations based on empirical evaluation and described in [97].

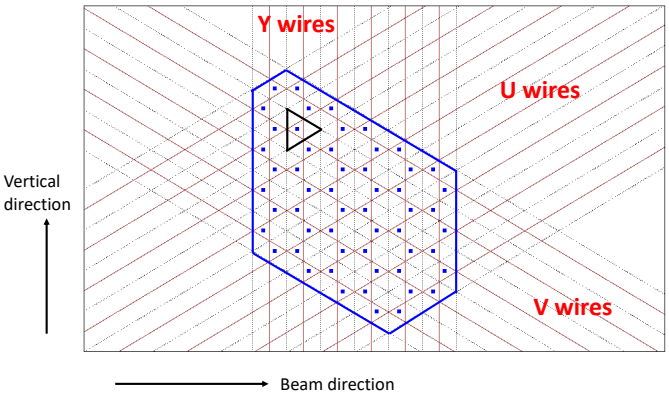


Figure 5.14.: Diagram of a time-slice in Wire-Cell. The solid lines represent wires, with dotted lines on each side representing the region covered by each wire. The triangular region of overlapping coverage by the three wires defines a cell, one of which is shown in black, and which is the smallest unit of coverage given the geometry of the wires. In this example, there are 55 possible cells represented by blue squares, and the entire region of overlap in which activity has been detected is known as a *blob*, and shown as the solid blue line. Figure from Ref. [97].

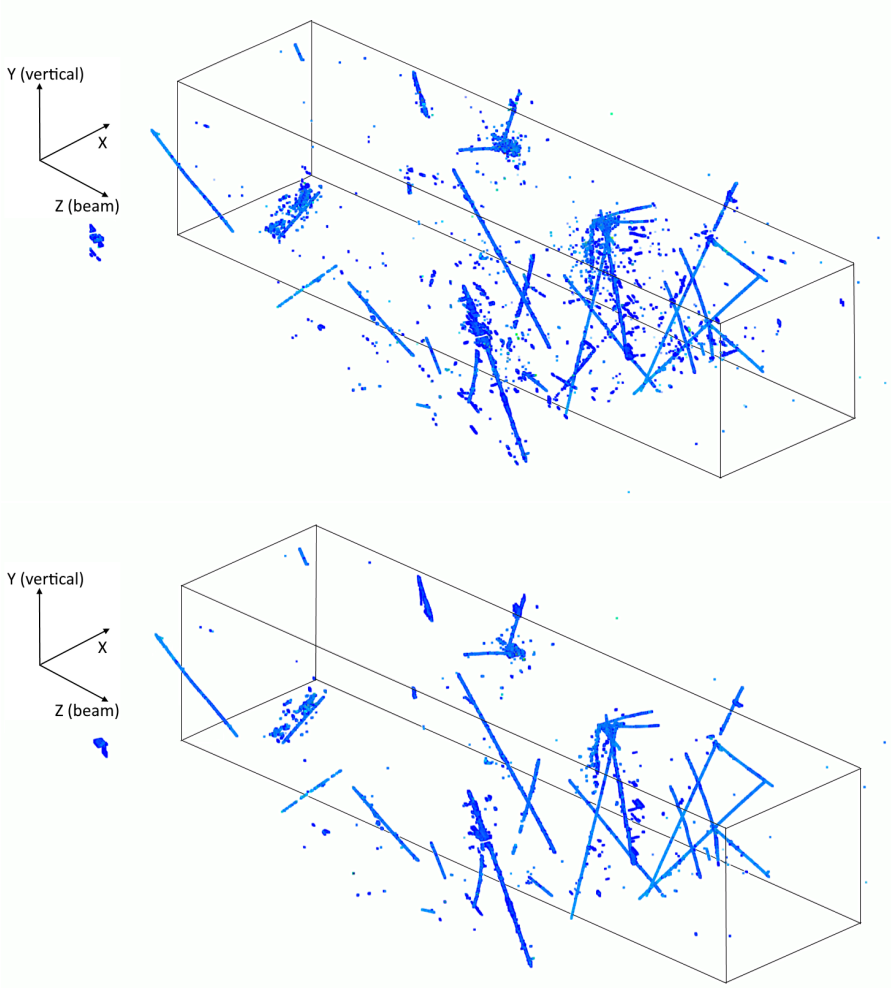


Figure 5.15.: Comparison of one MicroBooNE data event before (top) and after (bottom) deghosting has been performed. The colour scale represents the relative amount of charge in each blob. Most of the activity present is due to cosmic rays, for which further selection is needed to remove. Figures from Ref. [97].

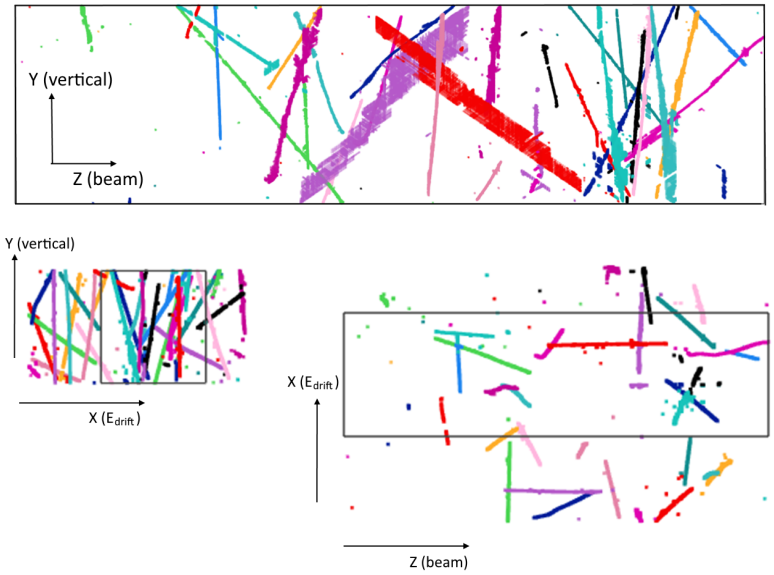


Figure 5.16.: MicroBooNE data event after Wire-Cell 3D clustering. Each reconstructed 3D cluster is shown in a different colour, with colours maintained across the three views of the detector shown. The black rectangles in each view represent the edges of the TPC. Figures from Ref. [97].

The result of reconstruction at this stage is an event-topology-independent 3D image of the ionisation electrons, which is then used as the basis for 3D clustering (not to be confused with the hit clusters discussed in the previous section). The goal of this 3D clustering is creating a solid base that would allow a later "high-level" stage to separate the activity in each event according to its physics origin, for example cosmic muons, or individual particle states in beam-related activity such as protons, pions, electrons, etc. An example of the final 3D clusters reconstructed in a MicroBooNE data event are shown in figure 5.16.

Finally, a many-to-many charge to light matching is performed between the 3D clusters and the light flashes described in section 5.6.3. The x-position of the 3D cluster at this stage not time-calibrated, and only given by its wire readout time. Its position is changed along this drift/time axis, and a prediction of the signals it would generate on the PMT is done using a photon library. The prediction is then compared to the signal recorded in the flash being compared to. A diagram of this matching is shown in figure 5.17. Typically, there are more flashes (40-50 per event) than 3D clusters (20-30), mainly because light from interactions outside the TPC but inside the cryostat is also reconstructed. Despite this, there still may be 3D clusters left without an associated flash, due to inefficiencies in light reconstruction. It should be emphasised that the isolated light-matched neutrino interaction cluster, shown in figure 5.17, is only one of many matched clusters at this stage. The removal of clusters due to cosmic activity will be discussed in section 6.2.

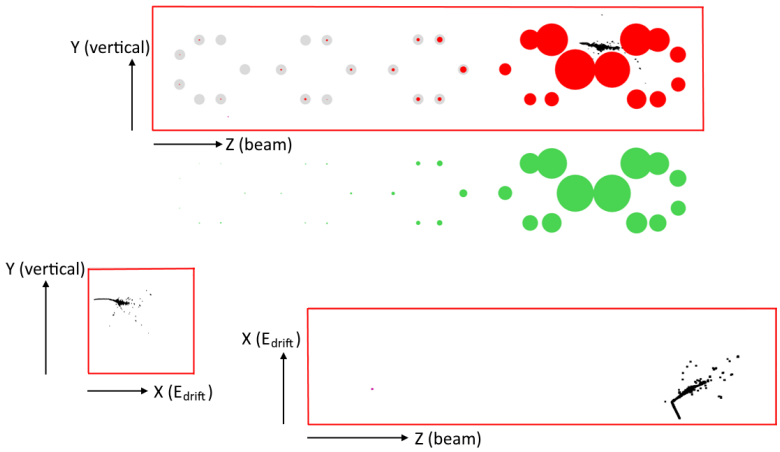


Figure 5.17.: Light matching for a neutrino interaction 3D cluster. Three views of the TPC are shown as red rectangles. The top figure also shows the PMT arrangement, with red circles indicating the reconstructed PEs in each PMT, a bigger circle representing more PEs. The green circles represent the predicted PEs, given the 3D cluster being tested. We see very good agreement between the size and position of the red and green circles, indicating a match. Figure from Ref. [97].

5.6.5. MeV scale reconstruction

In recent years, there has been an increasing interest in the study of MeV-scale physics using LArTPCs [37]. Particles at this energy scale will only travel a short distance (in the millimetre scale) in the liquid argon before coming to a stop. This means they will not be able to form long tracks or showers, and instead manifest as short depositions spanning only a few wires, which are known as *blips*. An example of blips alongside tracks and showers can be seen in figure 5.18.

The study of MeV scale physics in LArTPCs was pioneered by ArgoNeuT [16], after which it was quickly adopted by MicroBooNE [105–109]. The reconstruction and analysis tools for this new topology have now reached a high level of maturity in the form of a new LArSoft module named `BlipReco`, dedicated to blip reconstruction and which is described in Ref. [109]. MicroBooNE has recently demonstrated the sophisticated energy calibration and particle identification capabilities of BlipReco in Ref. [19]. There are many possible applications of blip reconstruction in LArTPCs [110], such as the tagging of neutrons (which are by themselves invisible in a LArTPC) via the de-excitation photons they produce [111], or the constraint of backgrounds in supernova burst detection [112, 113]. They have potential application in BSM searches [114], searches for millicharged particles in particular, as shown by ArgoNeuT [17], which acted as the main motivation behind our search.

Blips are constructed by matching hits from different wireplanes based on wire and time proximity. Hits are required in the collection plane (as it is used for calorimetry)

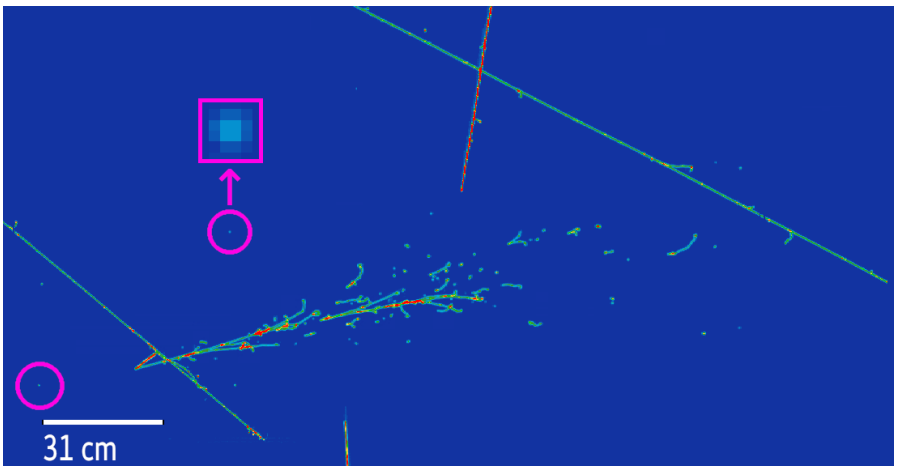


Figure 5.18.: Simulation of a NuMI ν_e interaction in MicroBooNE highlighting two blips (magenta circles), one of which has been zoomed in (magenta square). While many more point-like depositions can be seen near the overlaid cosmic tracks, or in the electron shower from the neutrino interaction, the blip reconstruction discussed in this chapter is targeted towards *isolated* depositions, such as the ones highlighted.

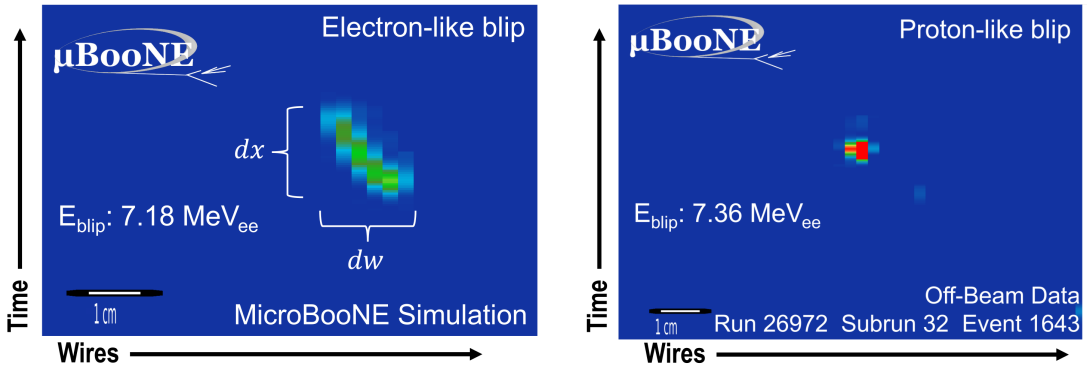


Figure 5.19.: Event displays of (left) electron-like and (right) proton like blips. The electron-like blip spans 6 wires, which are denoted by dw in the figure, and its extent along the drift coordinate is shown as dx. From Ref. [19].

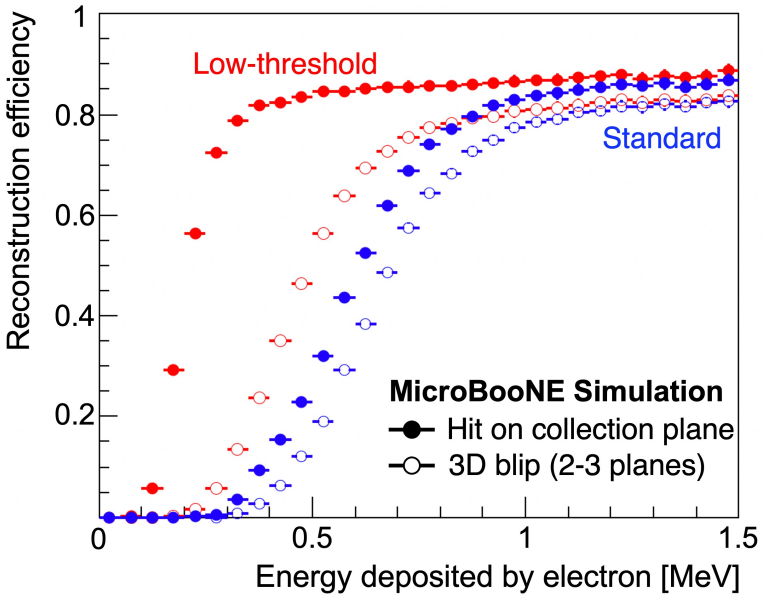


Figure 5.20.: Hit and blip reconstruction efficiency for different hit-forming threshold configurations. The standard thresholds have been used in most MicroBooNE publications. Ref. [109], from which this figure is taken, is an example of an analysis that uses a low-threshold hit reconstruction data sample.

and at least one of the induction planes. The resolution and spatial extent of blips in the YZ plane is determined by the 3 mm wire pitch, while in the drift direction X its length and position are determined from the drift velocity and the time information of the first and last hits comprising it. This is shown in figure 5.19, and the same information will determine the 3D position of blips. The shortest distance to a track is recorded, as blips due to cosmic activity are abundant, and this facilitates the removal of blips which are within a certain distance of a track. Energy calibration for tracks and showers in MicroBooNE is typically obtained from the charge per unit length, dQ/dx. However, the small spatial extent dx of blips makes this approach impractical. Instead, an "electron-equivalent" energy is estimated using simplified assumptions, as described in Ref. [109], which reports energy biases between 10 % and 20 % in the 0.1 MeV to 1 MeV range.

MicroBooNE's main goal is the reconstruction of neutrinos of $\mathcal{O}(100\,\text{MeV})$, and the hit-forming thresholds described in subsection 5.6.2 have been set with the aim of facilitating the identification of track and shower topologies. Blip reconstruction would benefit from less stringent hit-forming thresholds, but as hits are formed very early in the reconstruction stages, obtaining a sample with lowered thresholds implies the reprocessing of vast datasets. A sample with low thresholds was collected and analysed for the studies in Refs. [108, 109], and the impact of the threshold on hit and blip reconstruction efficiency is shown in figure 5.20.

In our blip search channel, we will use data from Runs 1, 2 and 3 from NuMI with

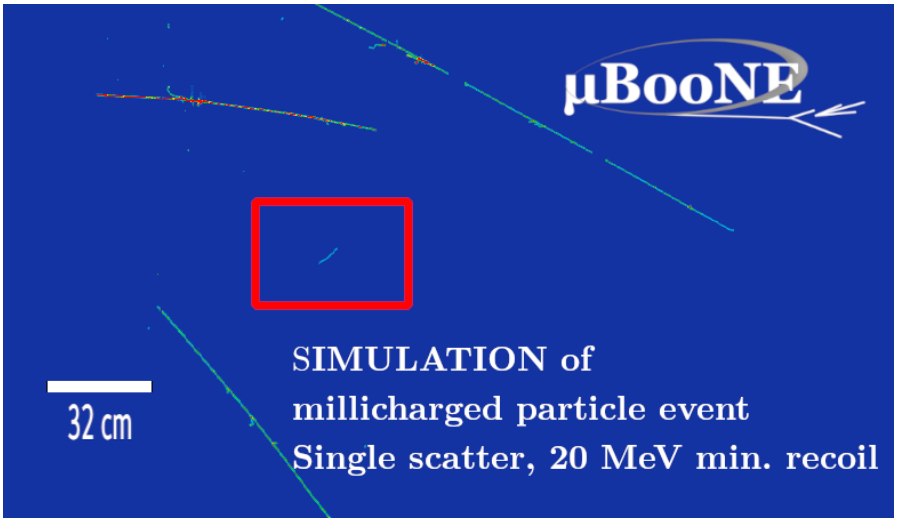


Figure 5.21.: Event display of a single-scatter mCP interaction (inside red box) alongside overlaid cosmic from beam-off data. The view is from the collection plane, with wires on the horizontal axis and time in the vertical axis.

the "normal" thresholds described in section 5.6.2. The sample of data processed with lowered thresholds is from a period with no beam activity, and thus could not be used in our search for mCPs. Using low hit forming thresholds significantly increases the number of blips due to cosmic activity, which are a background in our search. We performed preliminary studies which did not show a significant increase in sensitivity to mCPs when using lowered thresholds.

5.7. Wire-Cell single-scatter reconstruction

Wire-Cell is designed to reconstruct neutrino interactions at the $\mathcal{O}(100\,\mathrm{MeV})$ scale [97, 115]. Given the similarities between our signal and neutrino interactions containing electrons in the final state, such as ν_e charge-current interactions and especially to a neutral-current elastic electron scattering, we will apply more or less the standard Wire-Cell reconstruction and event selection, and effectively Wire-Cell will treat our signal as a neutrino event. In this section, we will describe Wire-Cell neutrino event reconstruction. The main difference between our signal and neutrino interactions is the generally lower energy of our scatters, which will cause Wire-Cell to fail to reconstruct a neutrino interaction, which will affect our signal efficiency. An event display of a simulation of a mCP interaction with cosmic beam-off data overlaid alongside it is shown in figure 5.21.

As MicroBooNE is located on the surface, one of Wire-Cell's primary concerns is the removal of the large number of cosmic muons present in each event while retaining the relatively sparse ν_μ and ν_e interactions. In Wire-Cell this is referred to as the *generic neutrino selection*, and can achieve a rejection of 99.98% of the cosmic-ray backgrounds, resulting in a contamination of only 14.9% of events (9.7% when above 200 MeV of

visible energy) while retaining 87.6% efficiency for inclusive ν_e CC interactions [68]. An example of the application of this selection to one event is shown in figure 5.22. Minor adjustments to this selection will be noted as relevant and summarised in subsection 6.2, when discussing event selection.

5.7.1. Generic neutrino selection

The first step in the selection is to remove all the activity outside the beam spill time window. This is accomplished by matching charge depositions in the TPC to flashes reconstructed by the light-collection system, which have precise timing information; then removing TPC activity which is outside the beam window. This matching is described in detail in Ref. [97], and a summary has been given in section 5.6.4.

The remaining clusters are in-time with the beam-spill, and could be produced by the interactions of particles originating from the beam, light-mismatched events (LMM), or cosmic tracks which are randomly coincident with this time window. These cosmic tracks can be further subdivided into through-going muons (TGM), which cross the entirety of the TPC, or stopped muons (STM), which enter through one face of the detector but stop within the active volume. In what follows, we will present the main concepts of the generic neutrino selection, whereas a more technical description can be found in Ref. [68].

A cluster is identified as a TGM by finding its highest and lowest points in the three coordinates x, y and z, and checking that both extreme points lie at the boundaries of the TPC. The boundary considered must take into account the space charge effect (SCE), which distorts the electric field and thus affects the drifting of electrons to the wires. The SCE creates an "effective detector boundary", which manifests as a reduction from the active volume as defined by the wireplanes and cathode plane, increasing with the distance from the anode plane. This effective boundary is shown in figure 5.23, alongside an example of a TGM. An extra 3 cm cut is applied on all faces of this volume, which defines the *fiducial volume* of the detector. A cluster is defined as *fully contained* (FC) if all its extreme points lay within the fiducial volume. We will require our events to be FC, which is not something required in the generic neutrino selection.

Stopped muons (STMs) can be distinguished from beam activity thanks to the fact that they will exhibit a Bragg peak (increased dQ/dx) as they stop; this is a telltale sign of a track that is at the end its trajectory (as opposed to starting it, as it might do if its origins were due to a neutrino interaction). To identify a Bragg peak, the 3D imaging information from Wire-Cell is projected onto the 2D space of the wire planes, and a series of Kolmogorov-Smirnov [116, 117] (KS) tests against the Bragg-peak/flat-MIP hypotheses. An example STM that was identified in MicroBooNE data using this procedure is shown in figure 5.24.

While a first pass of light-to-TPC charge matching has already been performed to remove the out-of-beam-time activity, the reduced background now allows for a second, more careful pass to remove light-mismatched (LMM) events. The majority of LMM events arise from low-energy activity, which are typically small, point-like clusters with low light intensity. The generic neutrino selection records whether an event is classified

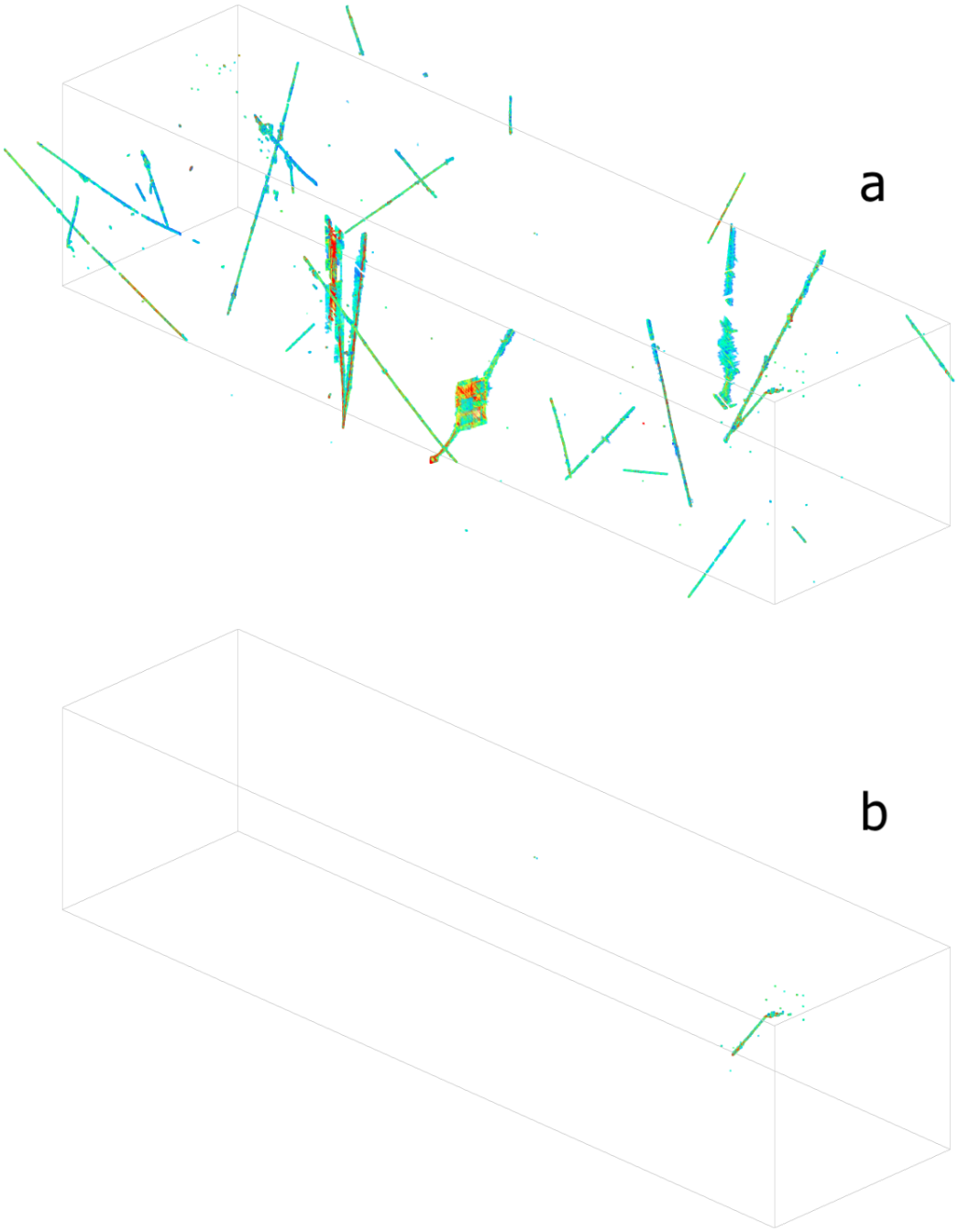


Figure 5.22.: Example event display showing the reconstructed TPC activity (a) before and (b) after generic neutrino selection is applied. Most of the activity in the top figure has been categorised as cosmic interactions, and a single neutrino-interaction candidate remains in the bottom figure. Source: Ref. [115].

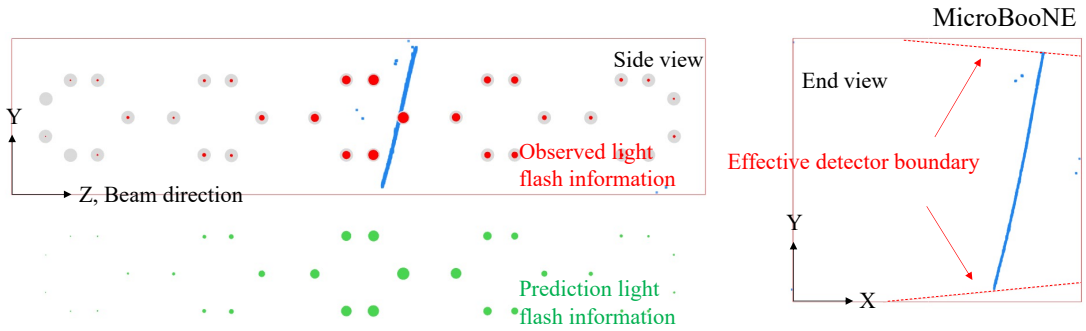


Figure 5.23.: Example of a through-going muon (TGM) in MicroBooNE data. The left panel shows a side view of the TPC, where the thick blue line within the detector corresponds to the TGM. The red circles drawn inside the detector represent the flash information observed using the light-detection system, while the green circles outside indicate the predicted flash information, based on the charge depositions. The right panel shows the TGM's endpoints crossing both sides of the effective detector boundary (dashed red lines) in the x–y plane. Source: Ref. [68].

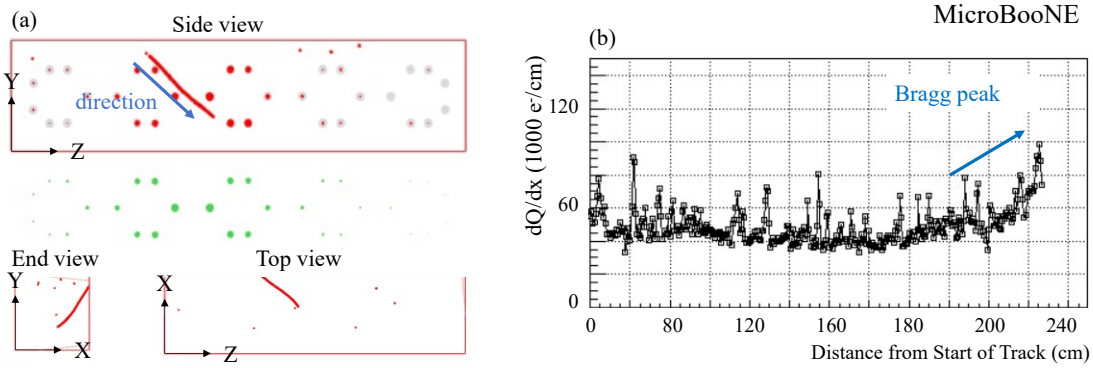


Figure 5.24.: Example of a stopped muon (STM) in MicroBooNE data. Fig. (a) shows the side, end, and top views of the MicroBooNE TPC, where the thick red line contained within is the STM. The red circles drawn inside the detector represent the flash information observed using the light-detection system, while the green circles outside indicate the predicted flash information, based on the charge depositions. Fig. (b) shows the charge deposited per unit length as a function of track position. Source: Ref. [68].

as *low-energy*, and separately tracks the minimum cluster length. In our selection, we lower the minimum cluster length threshold from 15 cm to 0 cm, but we retain the veto on low-energy events, despite the significant impact on our signal efficiency. This is necessary because an event may have a cluster length of 0 cm but still not be classified as low-energy. By doing this, we aim to remove low-energy point-like clusters, which are better targeted by our dedicated tools in the *blip* channel, where we have more effective and well-validated methods for background rejection.

Another potential cause of LMM events is that light produced near the cathode plane has a lower reconstruction efficiency, as it is the furthest from the light-detection system. To identify them, a KS test is performed between the predicted and observed light patterns. If the result of the KS test is extremely low, the cluster is immediately discarded; if it shows only a small disagreement, the cluster is checked against other flashes [68]. Let us remember that TPC clusters obtain their timing information, and therefore their x position, only when matched to a flash; hence, as the matching of different flashes are tested, the x coordinate of the clusters as a whole are tested. Now that the effective detector boundary has been calculated in a previous step, these clusters can be tested against the TGM or STM hypotheses. In the absence of any flashes matching the cluster, the cluster is moved in the x direction to find whether it matches the edges and be classified as a TGM. In the absence of a match, it is assumed that the flash was lost.

5.7.2. Wire-Cell 3D pattern recognition

After applying the generic neutrino selection, the cosmic-ray backgrounds have been greatly reduced and a more sophisticated 3D pattern-recognition stage can be performed. The main goal of this final selection is a full reconstruction of the "particle flow" of a neutrino interaction, meaning the identification of all final state particles from the 3D clusters, including their kinematics and parentage relation. The methods summarised here are described in full in Ref. [115].

The first step is to reconstruct the neutrino vertex. A vertex is a point at which many different clusters converge, and vertices can be categorised as either the neutrino (also called primary) vertex — produced by the initial neutrino interaction — or secondary vertices, which arise from subsequent interactions of particles generated in the initial event. As the neutrino is the originator of the entire interaction, all the energy, momentum, and the type of secondary particle generated will depend upon it, thus finding its vertex is fundamental for particle flow reconstruction. The neutrino vertex is found using two different methods: a "traditional" series of pattern-finding algorithms, and a deep-learning (DL) based method.

Traditional vertex reconstruction relies on knowledge of neutrino interactions. For example, the primary vertex tends to be more upstream and have more tracks connected to it. Second, the fact that the initial particles in the interaction (the neutrino and argon nucleus) are invisible, means that all particles will seem to "travel out of" the primary vertex, whereas in a secondary vertex some may travel towards it. To determine the direction a particle is travelling, the identification of clusters as either tracks or showers is made at this point. The travel direction in tracks can then be determined by a Bragg

peak, while showers will widen in their travel direction.

The DL method uses a regressional segmentation network which takes as input only the voxelised position and charge of the clusters and outputs a single "confidence value" whose value becomes larger closer to the primary vertex [115]. The DL method shows improvements over the traditional method, however the final vertex selection is made using a hybrid method. If the vertex found using the traditional method lies within 2 cm of the DL-based one, the DL vertex is used; otherwise, the traditional candidate is retained. This approach accommodates the existing pattern recognition workflow, which still relies on many non-DL methods. However, it is expected that Wire-Cell will continue to incorporate more DL-based techniques in future iterations.

Once the vertex is found, the particle flow tree is reconstructed. Starting from the primary vertex, the electromagnetic (EM) showers are clustered from what are initially multiple smaller 3D clusters. Second, neutral pions are reconstructed. Around 99 % of the time π^0s decay to a pair of photons, and the reconstruction of their invariant mass can be used to validate the EM energy reconstruction itself. They have also been used to constrain the background in single EM shower searches, such as in ν_e CC interactions [44] and single photon events [43]. Neutral current interactions could produce a single π^0 and no additional visible products in the event, in which case the primary vertex is likely to have been erroneously connected, therefore a dedicated algorithm exists to test for this case. During this step, Wire-Cell applies a variety of specialised "taggers", designed to classify clusters according to their origin, such as cosmic rays, neutral pions, and other common background or signal sources.

Track reconstruction (for muons, charged pions and protons) has been done during the traditional neutrino vertex finding stage, which completes all possible final state particles commonly expected in MicroBooNE. With the particle flow tree from the neutrino vertex completed, the neutrino energy is obtained from the sum of the kinetic energy of all particles in the particle flow tree, adding the masses of each muon, pion, or electron present, and a binding energy of 8.6 MeV per proton identified. Finally, Wire-Cell uses the information from its many taggers and the reconstructed kinematics of the interaction as input to a number of boosted decision trees (BDTs) which assign the interaction a probability of being a ν_e or ν_μ interaction. BDTs will be described in section 6.1, as we will use this tool ourselves to classify events as originating from a mCP interaction.

5.7.3. mCP signal event reconstruction

In each event, Wire-Cell will first try to find a neutrino event candidate by looking for light-charge matched clusters in time with the beam time window. In a very small fraction (< 0.1 %) of our signal events, our generated electron scatters will fail to deposit any energy in the detector. It is also possible that Wire-Cell is not able to find a neutrino candidate in our signal events, for example due to light flash misreconstruction. Finally, it is possible that Wire-Cell does find a neutrino candidate, but which is due to the cosmic activity overlaid alongside our generated electron scatter, in which case it should be considered a "bad match". We show the fraction of events which pass this first series of tests in figure 5.25. Here we see that around 30 % of our events will generate a flash

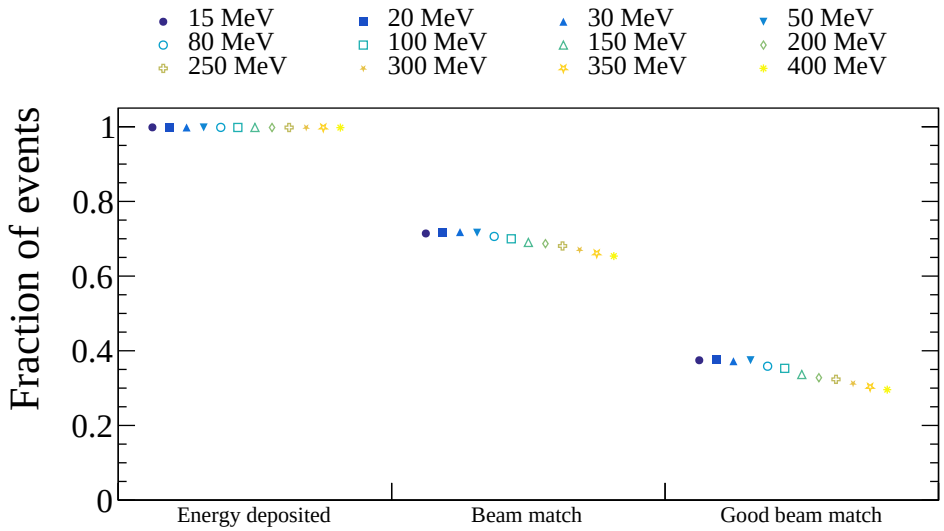


Figure 5.25.: Fraction of events in which the scatter electrons deposit energy, Wire-Cell is able to find a flash-charge match in time with the beam, in which those matches are due to our signal instead of cosmic activity. The fractions for different masses are shown with different markers and colours, and they are offset within their bins for better visualisation.

and charge signal that Wire-Cell is able to match to beam time, i.e. a "good beam match".

The reconstruction and selection efficiency of single electron scatters from a mCP signal of 15 MeV is shown in figure 5.26. We will discuss the effect of the selections applied in section 6.2, but for now we use this to justify our choice of a simulation threshold of 20 MeV, as the efficiency at this point is (1.49 ± 0.09) %, especially considering that the efficiency further decreases for higher values of m_χ.

In events in which a neutrino interaction has been successfully reconstructed, the reconstructed energy and vertex position is similar to that of the true electron scatter values, as shown in figure 5.27. The reconstructed energy shows a cut off below 20 MeV, as this is the lowest energy we simulate in this channel. It also shows some bias towards higher energy, but ultimately this is not important as the energy will not be an input for our multivariate search, we only care that a neutrino interaction has been reconstructed with some accuracy, and figure 5.27 shows that for a large population of events, it has.

The comparison between true and reconstructed vertex positions in the three coordinates is also shown in figure 5.27, where we observe that, in the vast majority of events, the vertex position is well reconstructed, as indicated by the dense population of events

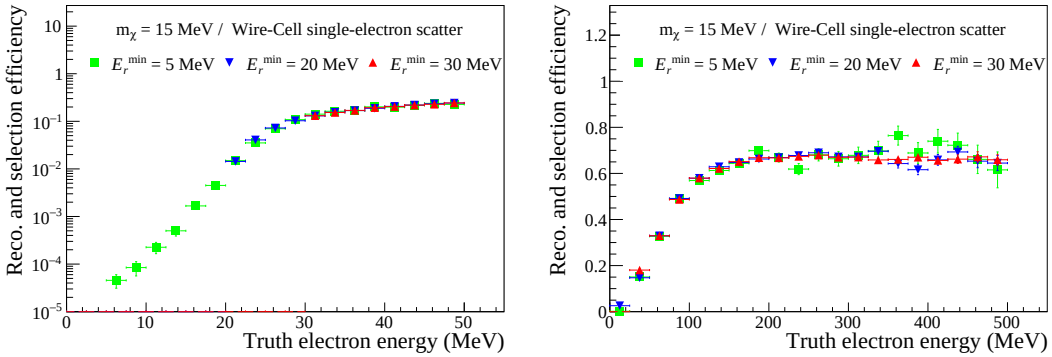


Figure 5.26.: Reconstruction efficiency for different values of truth electron scatter energy, for different minimum recoil energy, while maintaining all other simulation parameters. Event selection has also been applied. Left shows the detail at low energy ranges, using a logarithmic scale in the vertical axis. Right shows the efficiency in the high energy range, where fluctuations at the highest energies are due to the limited simulation samples used. The error bars are from statistics only.

along the diagonal. In all three coordinates we observe tails in the 2D histograms: in x and z these tails appear above the diagonal (i.e. reconstructed values are biased high), while in y, the tails are more symmetric around the diagonal. This can be explained by cases where a secondary vertex is misidentified as the primary one. The asymmetry between the x/z and y coordinates can be understood by considering the NuMI beam's incidence angle, with $\theta_{\text{NuMI}} = 27.46°$ and $\phi_{\text{NuMI}} = 6.20°$, corresponding to a unit vector of approximately $(0.46, 0.05, 0.89)$ in Cartesian coordinates. As the scatter electron tends to be forward-going along this direction, reconstruction biases are more pronounced in x and z, whereas the smaller y component results in a more symmetric distribution of vertex differences in that coordinate.

We also observe a population of events where the vertex has been reconstructed with a significantly higher x value (over 50 cm above the true value), as well as populations at the extreme y values (i.e., near the top and bottom of the detector). Both features can be explained by artifacts introduced by the effective detector boundary, discussed in section 5.7.1, and illustrated in the right panel of figure 5.23. This effective detector boundary is implemented in Wire-Cell to account for electric field distortions caused by the space charge effect. Finally, we observe a horizontal band of events in the z around 700 cm, and a smaller one around 400 cm. These correspond to regions non-functioning wires in the collection plane. As described in section , such areas lead to the reconstruction of a large number of fake blobs (or *ghosts*) in Wire-Cell.

Wire-Cell reconstructs the shower momentum in the (x, y, z) coordinates, from which we construct the (θ, ϕ) angles and compare them to those of the orientation from NuMI $\theta_{\text{NuMI}} = 27.46°$ and $\phi_{\text{NuMI}} = 6.20°$. This is shown in figure 5.28. One limitation is that, in the configuration used, Wire-Cell will not reconstruct showers with a kinetic energy

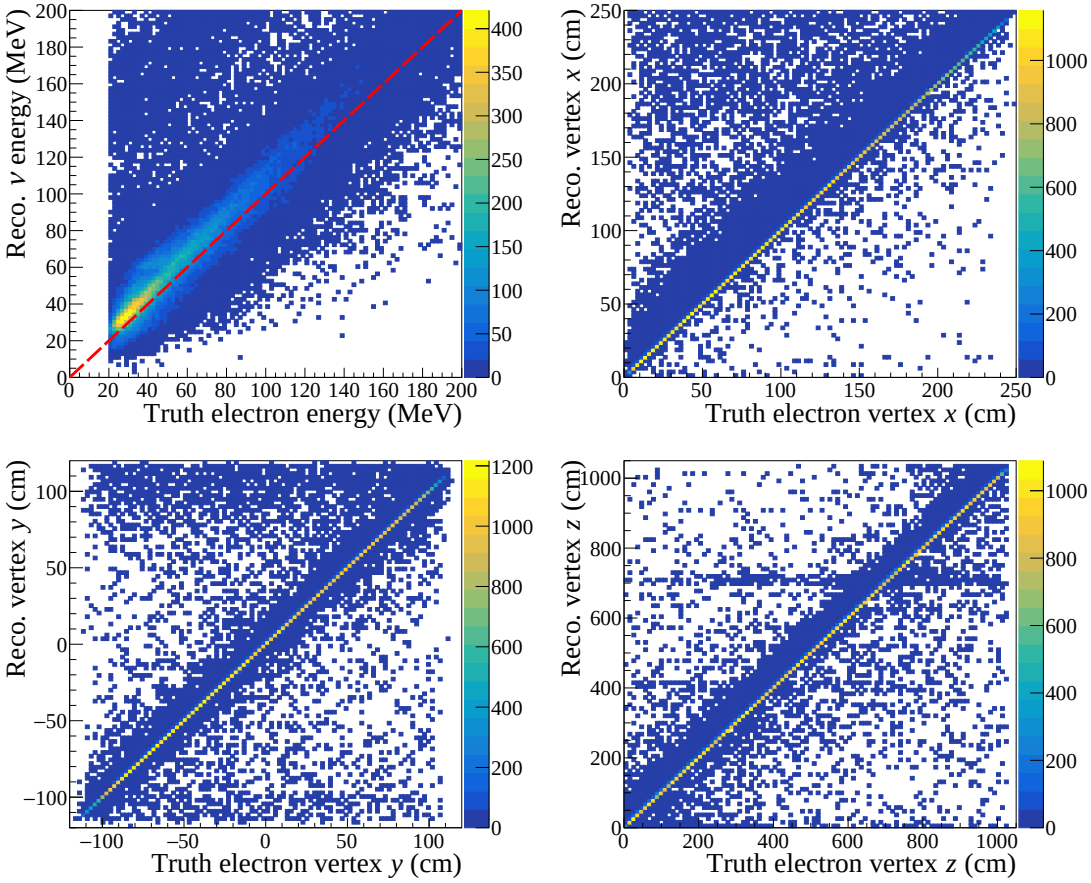


Figure 5.27.: Reconstructed neutrino energy and vertex compared to the true scatter electron energy, after mCP event selection.

below 70 MeV, as seen in the bottom panel in figure 5.28. Nevertheless, the associated charge depositions are still used to reconstruct a neutrino interaction, and the events are not discarded.

We observe that the reconstructed shower angles for a large population of events is well reconstructed, and centred around the expected values, with some smearing due to the non-zero angle of the electron scatter with respect to the mCP. The top panels in figure 5.28 display a zoomed-in view, focusing on the angular region around these main populations around the diagonals. However, the full distributions extend beyond this range, and both θ and ϕ exhibit long tails spanning the full angular range, from $0°$ to $180°$ in θ and $-180°$ to $180°$ in ϕ. These extended tails, not shown in figure 5.28, are likely due to events in which cosmic activity was misreconstructed as the primary neutrino interaction.

In the top panels, we can observe two populations for the truth shower angle. This can again be explained by the small vertical (i.e. y-axis) component of the mCPs incidence direction. Since electron scatters are generated at angle with respect to the trajectory of the mCP, the relatively large components in the x and z coordinates smooth out any bimodal structure. In contrasts, the small y means that deviations are more impactful, allowing a clearer separation into two populations. This explains why the effect is more noticeable in the ϕ angle, which depends on the ratio $\phi = \arctan(y/x)$ and is therefore more sensitive to variations in y. Meanwhile, the polar angle $\theta = \arccos(z/\sqrt{x^2+y^2+z^2})$ is dominated by the large z component and less affected by changes in y, leading to less visible structures.

5.8. Blip multiple-scatters reconstruction

The reconstruction efficiency of blips has been studied in Ref. [109] and is shown in figure 5.20. We observed the same results in our simulation, and chose to use a minimum recoil energy E_r^{min} of 0.5 MeV, which has an efficiency of 17.5 %. The signature we target with this channel is of multiple blips aligned with the detector target, as blips by themselves have no directional information, so as we need to reconstruct at least two blips, this lowers our event efficiency to 3 %. Figure 5.20 only shows the efficiency in the lower blip reconstruction range, but as the scatter electrons become too energetic the depositions start to become more track-like in nature, lowering the efficiency. We show the high-end energy range of blip reconstruction efficiency in figure 5.29. If we combine this information with the reconstruction efficiency obtained for Wire-Cell, we identify a "blind spot" in reconstruction efficiency, shown in figure 5.30, although it should be emphasized that in the case of Wire-Cell this efficiency already includes the effects of event selection, and Wire-Cell parameters were not optimised for low-energy searches.

The blip position distribution of signal and background blips along the x coordinate and the YZ plane is shown in figure 5.31, where we have used a signal sample with a mCP mass m_χ of 100 MeV. As previous studies have shown [19, 106, 107], there is an increased number of blips in the positions of the G-10 ribs used to support the field cage tubes. These *hot-spots* of activity can be seen in the YZ plane in the background distribution

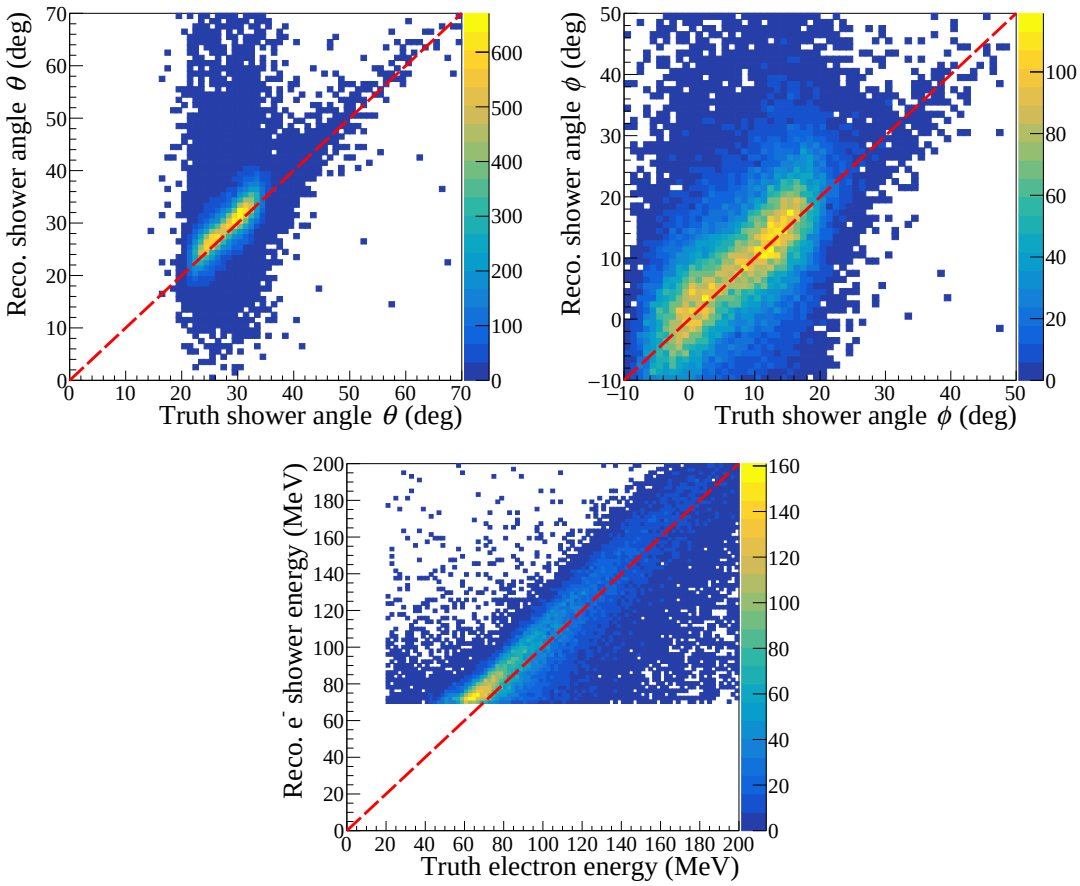


Figure 5.28.: Comparison between the true (simulated) and reconstructed shower angle orientations (top panels), and between between true shower energy and the reconstructed neutrino energy (bottom panel), for mCP signal sample.

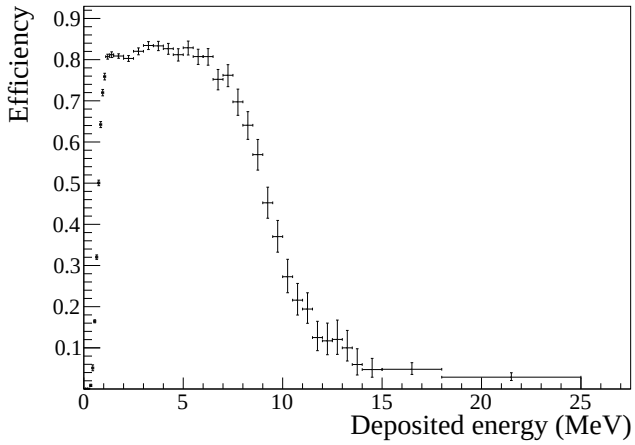


Figure 5.29.: Upper energy range of blip reconstruction efficiency, compared to figure 5.20. We see a dropoff for blips depositing over 5 MeV, as they start to resemble short tracks.

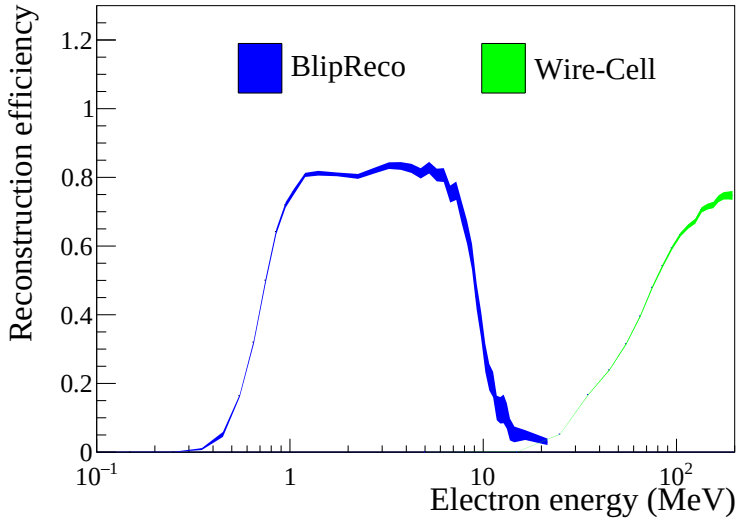


Figure 5.30.: Reconstruction efficiency of single electron scatters using either BlipReco or Wire-Cell (blue and green, respectively). The width of the band represents the uncertainties due to statistics. Wire-Cell reconstruction efficiency reaches a plateau above the energy range shown.

near the top and bottom of the vertical y axis. These hot-spots are a prime candidate for removal due to their decreased signal-to-background ratio. We also see that `BlipReco` saves the entire readout window, which extends beyond the main drift window. We only simulate mCPs in time with the beam spill, therefore we can remove blips outside this range. We will come back to this when discussing event selection in section 6.4.1, but for now we will continue discussing blip reconstruction without applying any selection to them.

Each event in MicroBooNE has on average around 85 background blips, as shown in figure 5.32. In this figure, we also show the number of reconstructed signal blips compared to the number of truth blips. The simulation parameter we used calls for at least two electron scatters, with a small probability of more than two scatters occurring. However, we see a large range of "truth blips" in the right panel of figure 5.32. The primary electron scatters are not in a one-to-one correspondence with the truth blips. On the low primary electron energy range, their depositions may have a low efficiency, resulting in fewer than two truth blips per event. In the high scatter range, their depositions may be more resemblant of tracks or showers, in which case none or multiple blips may be reconstructed. We observed a small fraction of events in which there are hundreds of truth associated blips, well outside the range shown in figure 5.32. It is this truth-association information which must be used in obtaining an efficiency, as was shown in figure 5.29. This effect can also be seen in the reconstructed energy, as shown in figure 5.33. The figure shows values of truth energy below the minimum recoil energy of 0.5 MeV used, as the primary scattered electron may produce additional low-energy depositions.

The main advantage of using LArTPCs to study blips is the millimetre-scale spatial resolution allowed by the wire spacing, which we can use to look for blips aligned with the NuMI beam target. A blip by itself has no orientation, which is why we need multiple blips. As will be explained in detail in section 6.4, we will consider blip-pairs to make the final selections, where the pair orientation is given by the angles of the line connecting the blips. We use the angles (θ, ϕ) in MicroBooNE coordinates, the distributions of which are shown in figure 5.34 for truth, reconstructed signal, and background blips. By signal pairs we mean that both blips in the pair are associated to the primary simulated electron recoil, whereas for background pairs neither of them are. There is an extra "mixed" case in which one of the blips is from signal and one from background, which will be explored when making selections on the pairs. For now, we note that the truth signal is very highly focused on a single bin in angle space around the angles of the vector pointing from the NuMI target towards MicroBooNE, $\theta_{\text{NuMI}} = 27.46°$ and $\phi_{\text{NuMI}} = 6.20°$. Those pairs whose orientation is not in this bin or an adjacent one are distributed uniformly. These correspond to secondary low-efficiency blips, and we can see in figure 5.34(b) that they are not reconstructed. The background blip-pair distribution in figure 5.34(c) shows features due to the geometry of the detector. A small ϕ angle indicates that both blips are on the same horizontal plane rather than vertical, while a small θ angle indicates small deviation from the Z axis, both of which are explained from the detector being much longer in the Z direction.

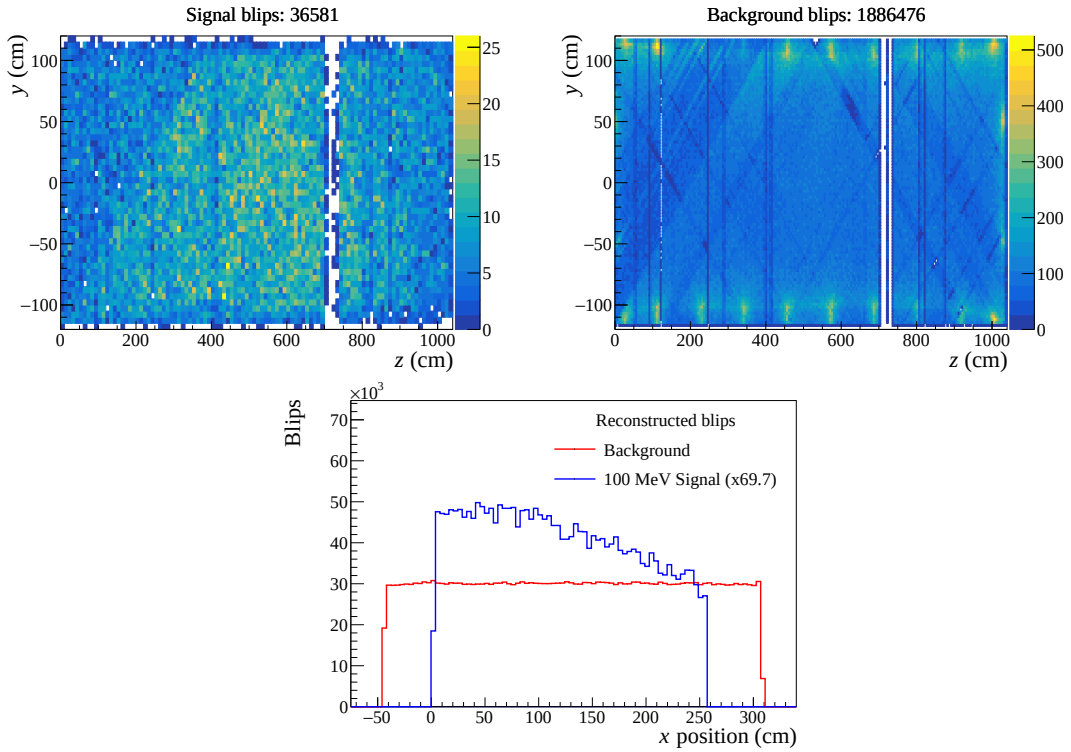


Figure 5.31.: Distribution of signal (top left) and background (top right) blips in the YZ plane. On the YZ plane the dead wire regions are visible through decreased reconstruction. The background distribution shows "hot spots" of activity as high as 500 blips per bin around the positions of G-10 struts. The bottom figure shows the distribution of signal and background blips along the drift direction. We see that blips are reconstructed through the entire readout window but signal is only present in time with the beam, hence on an effectively shorter x distribution. Signal blips decrease further from the anode (higher x) due to detector effects.

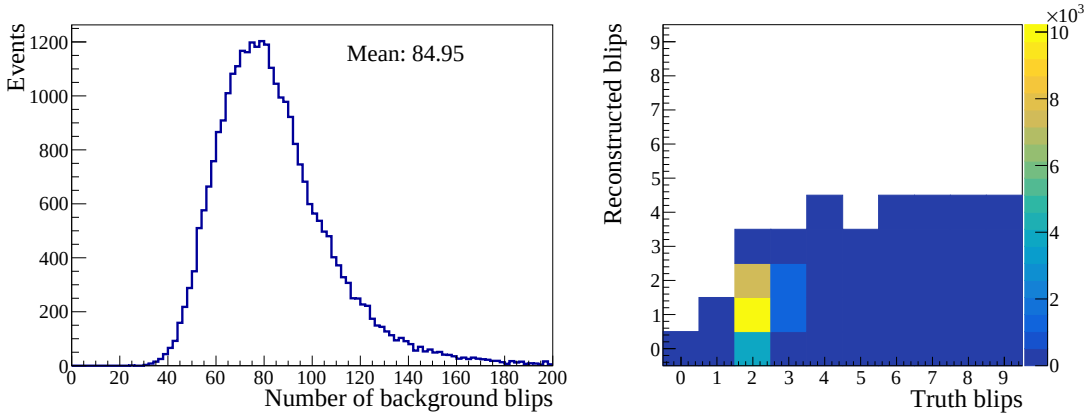


Figure 5.32.: Number of blips per event for an overlay sample. Left shows the number of background blips, which in this sample averages 84.95 per event. Right shows the number of reconstructed signal blips per truth blip.

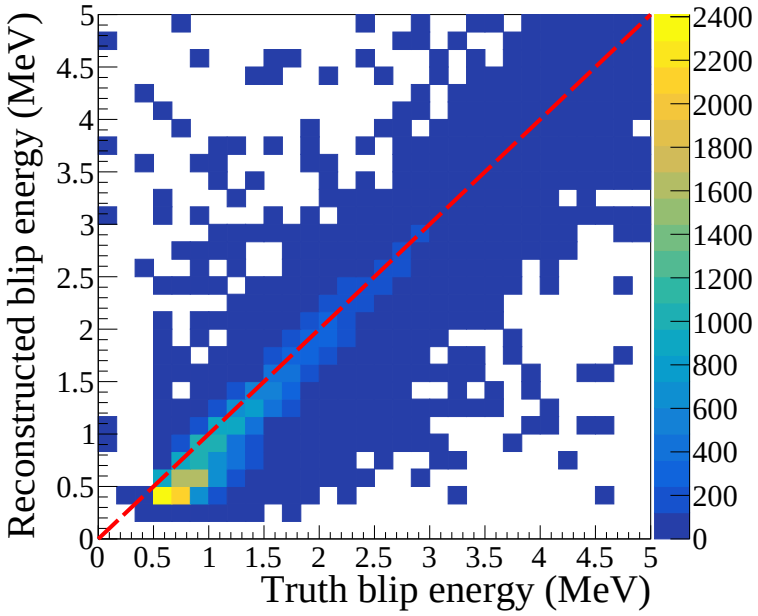
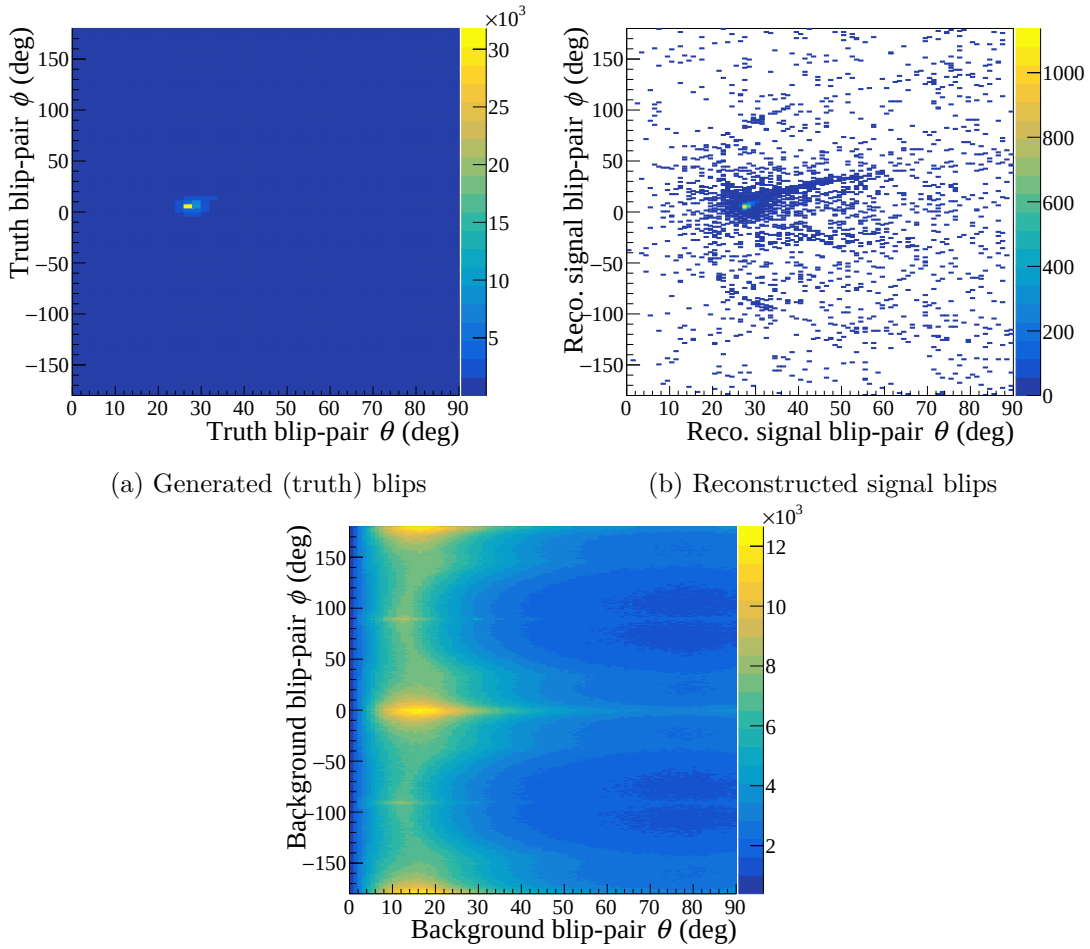


Figure 5.33.: Reconstructed energy for signal blips. The red dashed line indicates an ideal one-to-one energy reconstruction.



(a) Generated (truth) blips

(b) Reconstructed signal blips

(c) Background blips

Figure 5.34.: Distribution of the angles of the line joining any two blips in the events. The truth and reconstructed signal distributions are shown in (a) and (b). The background distribution is shown in (c).

5.9. Backgrounds

The main backgrounds in both search channels will be cosmic activity and beam neutrino interactions. Our single low-energy electron scatter channel has the same signature as electron-electron neutrino elastic scattering $\nu_e + e^- \rightarrow \nu_e + e^-$. Ultimately, it is the low energy of our signal which will help us remove most other neutrino backgrounds in the single scatter channel, at which point cosmic activity that was not correctly removed by Wire-Cell becomes the dominant remaining contribution. The reconstructed shower angles and energy distributions of these backgrounds are shown compared to truth information (from neutrinos) in figure 5.35. The angle reconstruction, shown in the top panels, shows a large population around the diagonal, indicating good agreement between truth and reconstruction. We also observe that the angular distribution is much more wide, compared to that of mCPs which are very focused on the angle of incidence from NuMI. Events away from the diagonal are likely cases where cosmic activity has been mistakenly included in the shower reconstruction, affecting the shower direction. In the bottom panel, the reconstructed neutrino energy is uncorrelated from the diagonal. This is expected, as neutrino interactions typically involve more complex topology than a single shower. Additional components, such as tracks or secondary showers not targeted in this selection, contribute to the total reconstructed energy, leading to discrepancies between these different axes.

In the multiple scatter channel, the condition of alignment with the beam target will remove most neutrino backgrounds, as even if a neutrino originates from the NuMI beam target, it is unlikely that the blips produced by it will retain this alignment. The NuMI beam has a significant ν_e component, which could be a potential background if other hadronic activity is not reconstructed correctly. Neutral current ν_μ interactions can also produce an electron in the final state. Photons initiated showers can be distinguished from electron showers thanks to the profile of energy deposited at the start of the shower.

To simulate neutrino interactions, we use the GENIE v3 event generator [100, 118]. This simulation uses a set of parameters to match data from neutrino interactions from the T2K experiment [119], commonly referred to as the "MicroBooNE tune" of GENIE. The simulation contains the full range of ν_e and ν_μ interactions expected from the NuMI meson flux, and has been used in numerous MicroBooNE NuMI based analyses[62–65].

5.10. Monte Carlo and data samples

For the single scatter search, we will generate samples of 12 different values for the mCP mass m_χ, between 15 and 400 MeV, which will go through the entire analysis chain. In chapter 6 we will describe the use of a boosted decision tree (BDT), which needs to be trained and validated over a large sample of signal and background. The size of the signal samples is of around 100 000 training events and 10 000 validation events, per mass point. All of these signal events are overlaid with beam-off data from cosmic rays, as described in section 5.5. We will use the same signal samples for all Runs and horn modes (FHC and RHC).

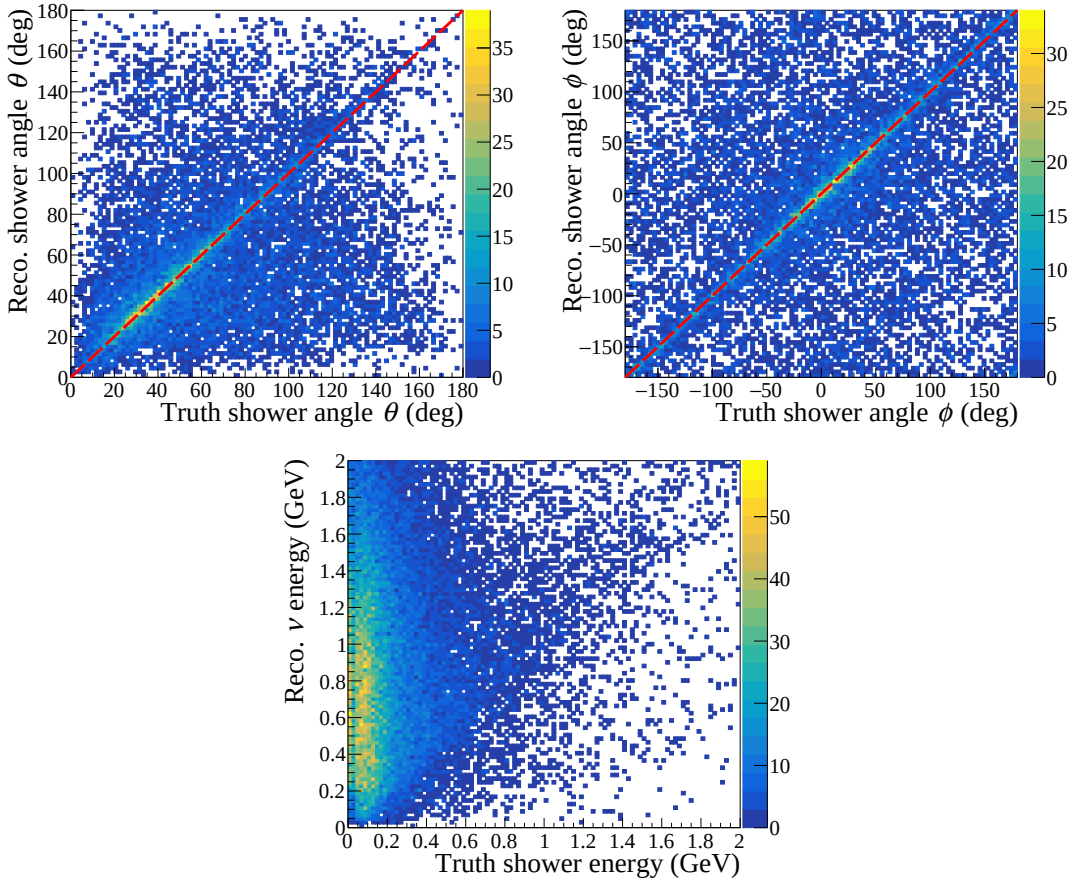


Figure 5.35.: Comparison between the true (simulated) and reconstructed shower angle orientations (top panels), and between true shower energy and the reconstructed neutrino energy (bottom panel), for the GENIE neutrino sample.

We will use beam-on data from MicroBooNE Runs 1, 2, and 3. Runs 1 and 2 can be divided into FHC and RHC components, while Run 3 only has data in RHC mode. The beam-on data POT for each Run is shown in table 5.1, where we also show the size of the GENIE MC simulation sample, which is a different sample for each Run/horn mode configuration. We also show the scaling factor applied for each Run/horn mode. The GENIE simulation is also overlaid with beam-off cosmic data, and when showing its distribution in histograms, we will separate it into "GENIE ν", which is neutrino interactions exclusively, and "Cosmics", which is due to the overlaid beam-off data. The "dirt" sample is a simulation of neutrino interactions which occur outside the cryostat but whose final state particles enter the TPC volume. This will only contribute a small fraction of the background, as we are selecting only events that are fully contained, and the neutrino vertex is outside the TPC which means that Wire-Cell will most of the time ignore them. We do not use the dirt samples for BDT training. As of the time of writing, we do not have a dirt sample for Run 1 RHC available.

Table 5.1.: Normalisation of Monte Carlo samples used for each Run period

	Sample	POT	Scaling
Run 1 FHC	Data	1.98×10^{20}	
	GENIE	1.12×10^{21}	1.78×10^{-1}
	Dirt	6.15×10^{20}	3.23×10^{-1}
Run 1 RHC	Data	5.96×10^{19}	
	GENIE	8.69×10^{20}	6.86×10^{-2}
Run 2 FHC	Data	1.28×10^{20}	
	GENIE	1.08×10^{21}	1.19×10^{-1}
	Dirt	1.37×10^{20}	9.38×10^{-1}
Run 2 RHC	Data	2.30×10^{20}	
	GENIE	9.97×10^{20}	2.31×10^{-1}
	Dirt	1.07×10^{20}	2.15
Run 3 RHC	Data	4.56×10^{20}	
	GENIE	2.70×10^{21}	1.69×10^{-1}
	Dirt	9.05×10^{20}	5.04×10^{-1}

In addition to the cosmic backgrounds overlaid in the GENIE sample, an extra "beam-off" sample, which contains only events triggered when the beam is off, is used (also known as the EXT trigger). This is to account for events in which the detector was triggered by cosmic activity, with no neutrino interactions present. These samples are only divided on a per-Run basis, and do not distinguish between horn current modes, as no beam activity was present. For the same reason, they cannot be normalised to POT; instead, they are normalised by dividing the number of triggers when the beam was on by the number of triggers when the beam was off.

For NuMI, only a fraction of the beam-off triggers are actually stored. This fraction (known as prescale factor) changes on a run by run (lowercase r, meaning a single data taking session of a few hours at most) basis, and it has to be accounted for when obtaining

the total number of EXT triggers. The number of EXT triggers is much larger on Runs 2 and 3, thanks to an increase in the available computing resources, which allowed MicroBooNE to collect more beam-off data.

For the multiple-scatter channel, our signal simulation samples contain approximately 30 000 events per mass point, and an additional 30 000 events for validation. Initial Monte Carlo studies indicated that our sensitivity is significantly lower than in the single-scatter channel; therefore, only a small subset of beam-on data, corresponding to 2.786×10^{19} POT, is used. As such, we present results for this channel as a sensitivity estimates only.

6. Selection and BDT training

In this chapter, we will describe the selection applied to the events. We use our Monte Carlo signal and background samples to train a boosted decision tree (BDT) in each search channel. After training, the BDT is applied to independent samples, including data, to obtain an output score, which has prediction power on whether an event is background- or signal-like. The resulting distributions of BDT scores are recorded as histograms, which alongside their uncertainties will be passed onto a limit setting framework to obtain a final result.

In section 6.1 we will describe the general principles behind BDTs as they are commonly used in experimental particle physics, along with the main parameters used in our training. We then take separate looks at both our signal channels. In section 6.2 we detail the event selection in the Wire-Cell single-scatter channel, and describe the BDT in section 6.3. After this, we move to the two-blip channel and describe the event selection in section 6.4, and its BDT in section 6.5. We end the chapter with a brief summary in section 6.6, which reiterates the main results obtained, which are histograms of high scoring events in each BDT which we will use to set exclusion limits.

6.1. Boosted decision trees (BDTs)

A boosted decision tree (BDT) is a machine learning tool used for classification, which has become commonplace in experimental high energy physics data analysis [120]. It consists of a series of algorithms which through an iterative process known as *training* can *learn* to distinguish between different types of data (in our case a binary classification between signal and background) based on multivariate information present in each event. BDTs are computationally efficient, can handle a large number of variables, and their results can be interpreted in terms of the physics variables used as inputs. We will use BDTs to predict whether the data events recorded by MicroBooNE correspond to a mCP signal or one of the background samples we consider in each of our two search channels.

BDTs are based on decision trees, an algorithm which uses the information from multiple variables of an event before deciding on a classification, rather than using a singular criterion [121]. A diagram of this concept is shown in figure 6.1. At each *node* of the tree, a binary decision is made based on the unbinned sorted values of one variable. Based on this decision, the event is split into two different *branches*, after which it may encounter more nodes, or reach a *leaf*, where a classification decision is made, usually in the form of a probability. In the case of binary classification this score s will be a value in the range $0.0 < s < 1.0$, although it is common to present these results using a logit

transformation

$$\text{logit}(s) = \log\left(\frac{s}{1-s}\right), \tag{6.1}$$

which has the effect of "spreading out" values that are very close to either 0.0 or 1.0. The maximum number of nodes that any event goes through before reaching a leaf is known as the *tree depth*.

In machine learning, *boosting* refers to the use of a collection of weak classifiers, which are more error-prone and may only perform slightly better than random guessing, but are easy to construct. These weak classifiers work together iteratively to build a strong classifier, which has better performance and a smaller error rate, but is harder to construct from fundamental principles [120]. In the case of BDTs, this means the use of multiple decision trees in an *ensemble*, with each new tree using the results from the tree ensemble as input, in addition to the multivariable data from the events. For example, the commonly used *AdaBoost* [122] (Adaptative Boosting) family of algorithms adds emphasis on events which have been misclassified by one tree by weighting these events more heavily when passed onto subsequent trees, with the aim of reducing the misclassification rate. In *gradient boosting* [123], the goal is to minimise the loss function in a gradient-like manner, by adding new decision trees in sequence which improve upon the errors of previous ones.

One issue commonly encountered in machine learning is overtraining. A BDT suffers from overtraining when it performs well on the training samples but is not capable of making accurate predictions on new ones. This may be due to the BDT having become too familiar with the training set, and basing its predictive power on very specific features which may not be present in more general data. Overtraining can occur if the model is too complex (for example if trees have too many nodes), there is insufficient training data, or too many iterations are used. One recent variant of gradient boosting, *DART* [124] (Dropouts meet Multiple Additive Regression Trees), attempts to alleviate overtraining by randomly dropping trees from the ensemble in each iteration. This results in the effects of the first trees in the ensemble not becoming overwhelming, and allowing newly added trees to contribute more significantly to predictions, even after many iterations. As a result, *DART* training can take longer, but it provides very powerful results.

6.1.1. XGBoost

XGBoost [125] (eXtreme Gradient Boosting) is an open-source gradient boosting library that is highly scalable, supports parallel computing, is capable of running on GPUs, and can be interfaced through many programming languages[1]. It gained attention in the high energy physics (HEP) community after its adoption by a number of teams participating in the 2014 "Higgs Machine Learning (ML) challenge" [126], which earned its creators the "HEP meets ML" award, in recognition of its potential benefit for HEP. Since then, XGBoost has become a widespread tool used in many particle physics searches, in particular BSM searches in MicroBooNE [61–64]. We use the DART algorithm of XGBoost version 2.1.2 to train a BDT in each of our channels.

[1] In this work we used the Python implementation.

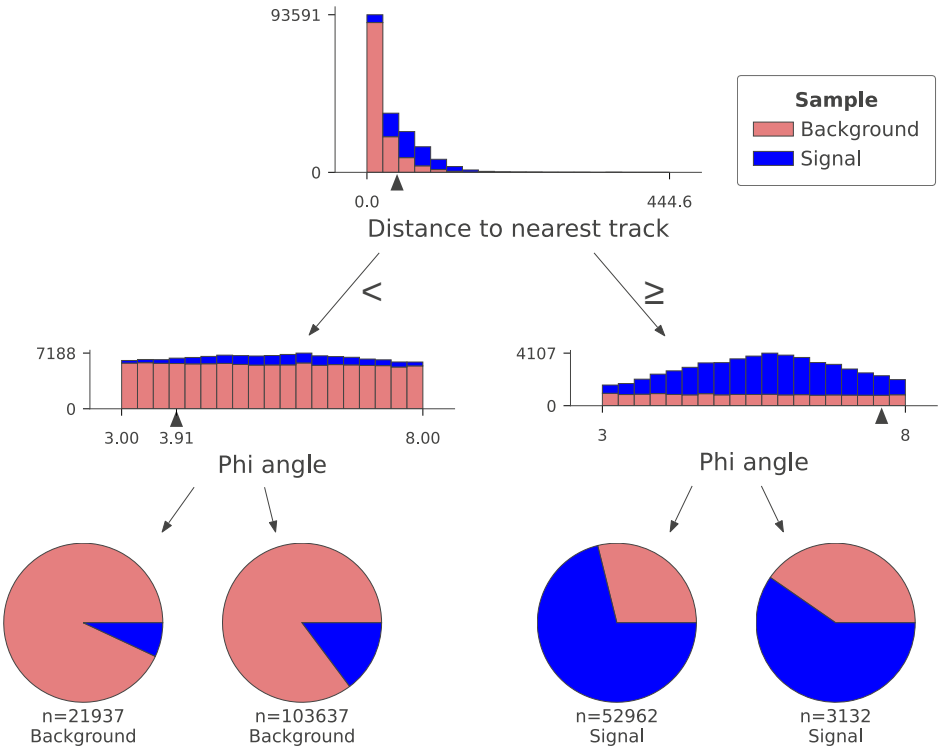


Figure 6.1.: Structure of a simple decision tree of depth two, illustrating a binary classification using two variables for a sample of background (red) and signal (blue) events. Events begin at the top node, where a split is performed based on the "Distance to nearest track" variable. Depending on whether the value is below or above a threshold (chosen to maximise signal-background separation), the event proceeds to one of two branches. Each branch then performs a further split using the ϕ angle. The final nodes, or leaves, are represented as pie charts showing the signal-to-background ratio within each leaf. This ratio determines the BDT score assigned to any event ending in that leaf.

Many aspects of the training can be configured via a set of parameters. The values we use in training only differ slightly between channels, so we present a brief description of the main ones which we have tuned in table 6.1. Any XGBoost parameters not shown in this table have been left in their default values in both channels. We aimed to keep the maximum tree depth small, preferring to do multiple iterations with smaller trees. One of the notable features of XGBoost is their use of an approximate algorithm for tree learning, which can handle very large datasets efficiently, and is described by the authors of XGBoost in Ref. [125]. We use this approximate method by setting the `tree_method` parameter to `hist`.

We use a fraction of our training samples to perform evaluation of the performance of the BDT during training (this subsample is then known as *testing* sample). Different metrics can be evaluated during training, which can be used to stop the iterations if no further improvements are seen. In practice we will consider multiple such metrics, which allows us to set the initial number of iterations to a very large value (say, 1000 iterations), and let the training run until reaching this point of diminishing returns, after which we select the best iteration as our model.

Table 6.1.: Main parameters used in XGBoost BDT training.

BDT parameter	Description
Tree parameters	
Maximum tree depth	Maximum depth of the decision trees
Learning rate (η)	Weight given to new trees added to the model, effectively slowing learning
Objective	Logistic regression for binary classification
Gamma	Minimum loss reduction to make a new partition at a node
Subsample	Fraction of the training data used in each step
Tree method	Tree construction algorithm
DART parameters	
Drop rate	Fraction of previous trees to drop
Skip drop	Probability of skipping the dropout procedure

6.2. Wire-Cell single-electron selection

Our event selection is based upon generic neutrino selection (GNS), which we have discussed in section 5.7.1; we will briefly summarise it here and discuss the modifications applied for our analysis. First, we require a light flash matched to a 3D charge cluster in time with the neutrino beam time. As we have seen many of our signal events fail this selection, and many events which pass it do so thanks to a "bad match" due to cosmic interactions instead of our signal. In real data we have no way of knowing whether a match is good or bad, thus in this section we will discuss selections only based on our "good matches", i.e. our mCP signal.

In GNS, the matched cluster must not be in a region with too many dead channels,

and it must not be tagged as either a through-going muon, stopped muon, or as a light-mismatched cluster. The cluster must not be tagged as low-energy. We make no changes in these selections, compared to GNS. In GNS, a minimum cluster length of 15 cm is required, however we will relax this restriction to allow for cluster lengths of any length greater than 0 cm, which only removes point-like clusters which we are targeting with our blip channel. We also require the candidate cluster to be fully contained. Although this is not a part of GNS, it is a selection criterion used in most other neutrino analyses to exclude events that would be poorly reconstructed. Finally, we require that a neutrino interaction is successfully reconstructed.

A summary of our event selection is shown table 6.2, alongside the generic neutrino selection, which also highlights our departures from it. The efficiency of each selection applied individually upon our millicharge signal is shown in figure 6.2. Note that the efficiencies shown here are applied over good beam matches. The "Event type" selection which we have not so far mentioned and shows an efficiency of 1 is a Wire-Cell consistency check, which never fails in our signal.

Table 6.2.: Selection criteria applied to our millicharged particle (mCP) signal in the Wire-Cell single electron scatter channel. The generic neutrino selection (GNS) is also shown, with items in red highlighting our departures from it. Selection criteria marked as "True" must fulfil the condition described, while those labelled "False" indicate that events are rejected if they fulfil it. A dash indicates a selection is not applied for this variable. We show a shorthand name which will be used in the labels of figures.

Shorthand	Description	mCP sel.	GNS
Beam match	Light-charge match in time with beam found	True	True
Dead channel	Candidate is in region with too many dead channels	False	False
Event type	Valid cluster classification	$\neq 0$	$\neq 0$
FC	Candidate is fully contained	True	-
TGM	Tagged as through-going muon	False	False
STM	Tagged as stopped muon	False	False
Low energy	Tagged as low-energy	False	False
Cluster length	Minimum cluster length (cm)	> 0	> 15
Light mismatch	Tagged as light-mismatched	False	False
ν reco.	Reconstructed neutrino energy (MeV)	> 0	-

We see in figure 6.2 that only three selections significantly reduce our efficiency: the low energy, cluster length, and ν reconstruction requirements. These three selections however are highly correlated, as shown in figure 6.3 for two examples of mCP masses of 100 MeV and 350 MeV. In these 2D histograms, we show the fraction of events passing the selections on the horizontal and vertical axes simultaneously, normalised to the separate efficiencies in each axis, E_i and E_j respectively, as $1/\sqrt{E_i E_j}$. We find that the correlations are similar across all simulated mCP mass points.

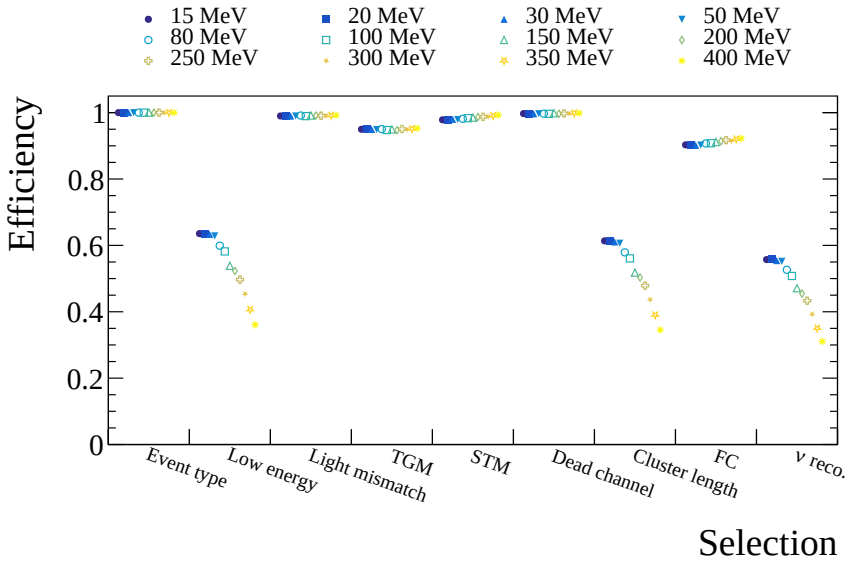


Figure 6.2.: Effect of individual selection criteria upon our signal efficiency for the different mCP mass simulations. The markers for different mass points are offset in the horizontal axis for easier visualisation.

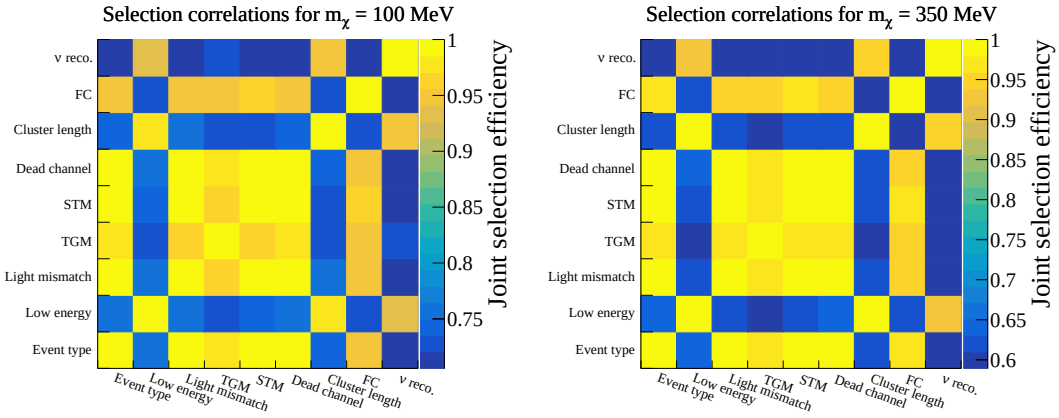


Figure 6.3.: Correlation among any two selection criteria for simulations of (left) a 100 MeV and (right) a 350 MeV mCP. The joint selection efficiency in the z axis is the fraction of events passing the selection criteria in the horizontal and vertical axes simulatenously, normalised to the product of their individual efficiencies.

6.3. Wire-Cell single-electron BDT

We apply the mCP selection to our signal and background training samples, and use the selected events to train a BDT in XGBoost. In the following subsections we will detail some aspects of the training, and the resulting BDT-score distributions, which will be used to set limits on the signal.

6.3.1. Wire-Cell single-electron BDT training

The signal used for training is a combination of all the mCP signal mass points. This was done to increase the robustness of our BDT via a bigger training sample, as we did not observe fundamental differences on the kinematics of the scatter electrons for different mCP masses. We separated the training samples in a 2:1 ratio for training and testing, which allows for a sufficiently large testing dataset while still having a robust training set, with a total of around 115 000 signal events are used for training, 57 000 for testing. In addition to the training and testing samples, we use a sample containing an independent set of Monte Carlo simulation events, not involved in BDT training, which we refer to as the validation sample. This sample helps prevent biases due to overtraining, and the histograms shown in this section are always derived from this validation sample. The size of the signal validation sample is of around 20 000 events, spread across the different mass points.

We also used 46 000 events from GENIE neutrino event generator as background, and around 3600 events of beam-off data to train against purely cosmic-induced backgrounds. For validation of the neutrino and cosmic samples, we used the simulation and beam-off data from different run periods — and thus are also independent samples, not involved in the training — and found no significant overtraining. The number of cosmic background events used for BDT training, which come from both the "EXT" beam-off sample and bad matches in the GENIE neutrino sample, is the lowest. This is because generic neutrino selection, which is the basis of our event selection, removes the vast majority of cosmic activity. The dirt sample was not used in BDT training.

We used 228 variables from Wire-Cell for training, comprising all variables containing physics information, with only bookkeeping variables (e.g. event number, run number) excluded. In addition to the final particle flow of a full neutrino interaction reconstruction, Wire-Cell also stores lower-level information from many of its "taggers" which were used in the neutrino vertex finding and full event reconstruction. Given the very large number of variables used in training, we reduce the complexity of the BDT by using a small maximum tree depth of 3, as an excessively complex model may lead to overtraining. We will only discuss here the 10 most important variables under two different metrics, total gain and weight, which are shown in figure 6.4. Total gain refers to the cumulative improvement in the loss function (i.e. how much the model's predictive accuracy improves) from all splits where the variable is used. Weight refers to the number of times that a variable is used to split the data across all trees in the ensemble. While gain reflects how useful a variable is in reducing error, weight captures how frequently it is used during training. The absolute numbers shown in figure 6.4 correspond to inter-

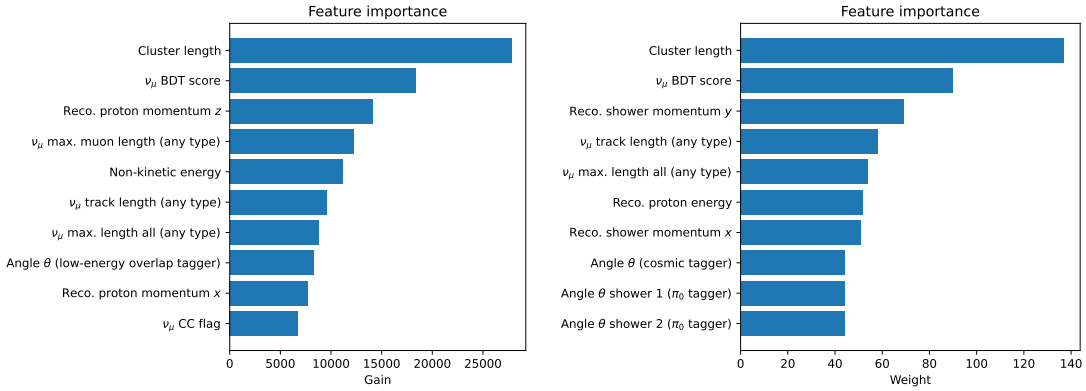


Figure 6.4.: Bar charts of the most important training features for the Wire-Cell BDT in terms of (left) total gain and (right) weight.

nal metrics used by *XGBoost* during training. While these values may not be directly interpretable, our primary interest lies in the relative ranking of the variables. We see that the the same variables appear at the top under both metrics, highlighting their predictive power.

The most important training variables across metrics are the cluster length and the ν_μ BDT score. As we have mentioned, generic neutrino selection involves a selection of cluster length above 15 cm only, which we have removed as our low-energy scatters are likely to result in shorter clusters. The ν_μ BDT score refers not to our BDT, but one that Wire-Cell uses to classify an event as originating from a muon neutrino interaction. These are characterised by the absence of electrons and presence of a muon in the main vertex. As our interactions feature an electron and no muons, our BDT will naturally reject any ν_μ-like interaction, which are most neutrino interactions from the NuMI beam. The distribution of signal, backgrounds, and data for Run 1 FHC are shown in figure 6.5. In this figure we see good agreement between data and Monte Carlo, except in a region close to a ν_μ BDT score near 0, which will only overlap with our signal region slightly.

Other variables related to the presence of tracks in ν_μ interactions which score highly are the maximum muon length, track length, and maximum track length. In figure 6.4 these are tagged as "(any type)", this is because the Wire-Cell ν_μ tagger has different categories, for example for short or longer muon tracks connected with the main vertex. These "any type" tracks is the broadest category. We show these variables in our samples in figure 6.6. We also observe good agreement between data and Monte Carlo backgrounds in these figures. In these histograms, most signal events have a length of 0 cm, indicating that no track was reconstructed for these events. The prominence of these variables in BDT training suggests that the BDT effectively rejects events in which the reconstructed "neutrino interaction" includes long muon tracks.

Many different Wire-Cell taggers use angular information in their respective tasks. We see that the angle θ information from the cosmic tagger (used to tag cosmic rays, which predominantly come from above), low-energy overlap tagger (which aids in sep-

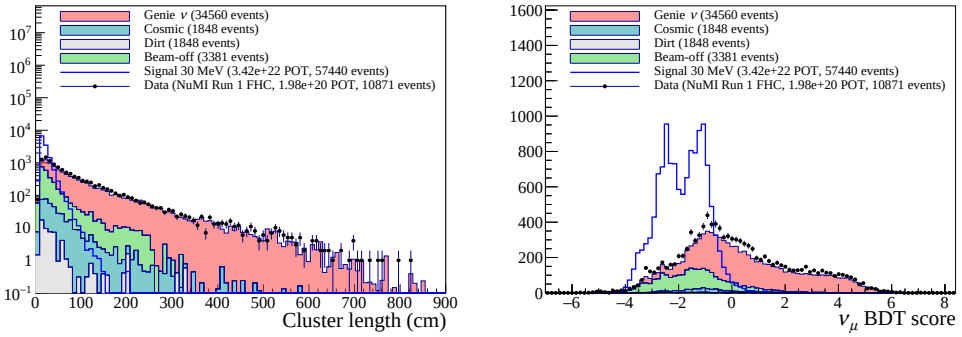


Figure 6.5.: Most important BDT training variables. The filled coloured stacked histograms represent the different backgrounds, the blue histogram is a mCP signal of mass 30 MeV, and the markers are data. The backgrounds are normalised to data POT, while the signal is scaled to a higher POT for better visualisation. Left: length of the main cluster. Right: ν_μ BDT score.

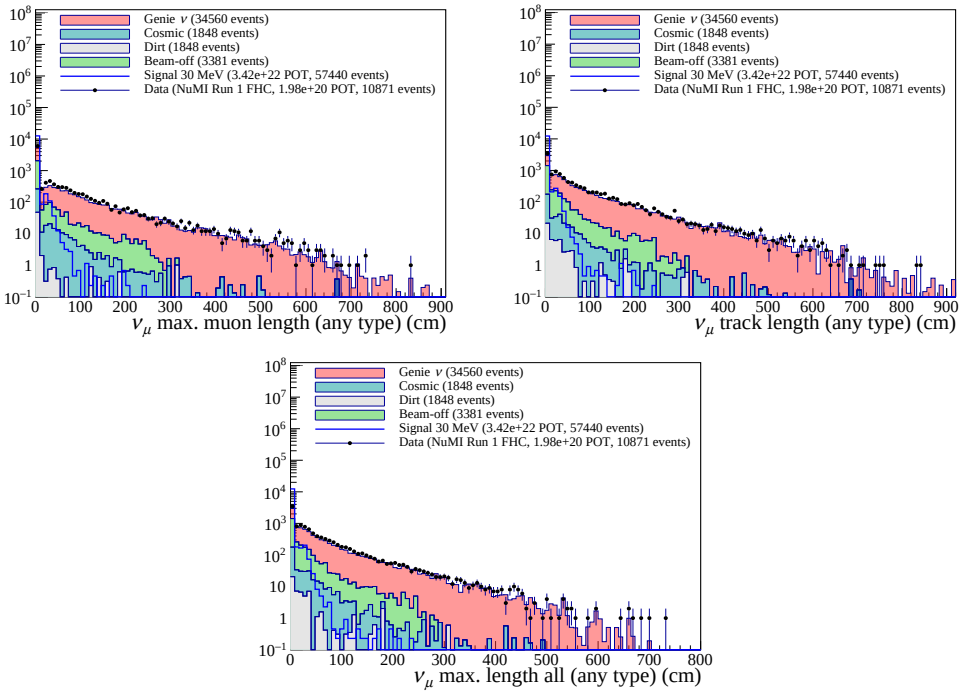


Figure 6.6.: Data, Monte Carlo signal and background comparison for the track length variables which are most important in classification. The filled coloured stacked histograms represent the different backgrounds, the blue histogram is a mCP signal of mass 30 MeV, and the markers are data. The backgrounds are normalised to data POT, while the signal is scaled to a higher POT for better visualisation.

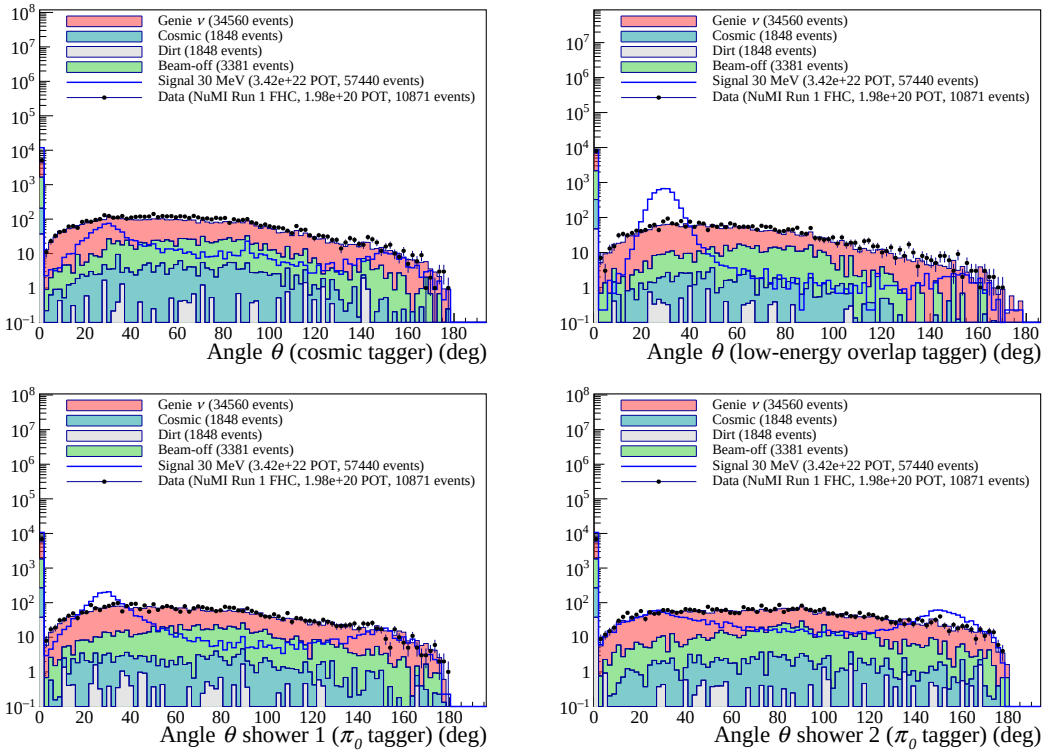


Figure 6.7.: Different Wire-Cell angle taggers. The filled coloured stacked histograms represent the different backgrounds, the blue histogram is a mCP signal of mass 30 MeV, and the markers are data. The backgrounds are normalised to data POT, while the signal is scaled to a higher POT for better visualisation.

arating highly overlapped low-energy clusters) and the π^0 tagger (which identifies final state neutral pions by fitting both its photon showers separately) are amongst the most important training variables of our BDT. This is expected, as our signal has a very characteristic angle which can be easily separated from the background. We show the values of our samples for these variables in figure 6.7. In these histograms we see good agreement between data and the Monte Carlo backgrounds. The signal shows a peak around 30° in all figures, which is close to the orientation of the direction from the NuMI target. This peak is especially pronounced in the case of the low-energy overlap tagger. A large population of signal and background events is concentrated on an angle of 0°, indicating that no shower was reconstructed.

Finally, we have variables related to the presence of protons and showers in our events. We see that proton energy and momenta in x and z appear in our top variables, as well as the shower momentum in x and y. We show some of these variables in figure 6.8. In all of these we see many events for signal, backgrounds, and data, which fall onto the default value of -1, which indicates that a proton or shower has not been reconstructed. Signal is vastly concentrated on this default value, indicating that it is rarely misidentified as

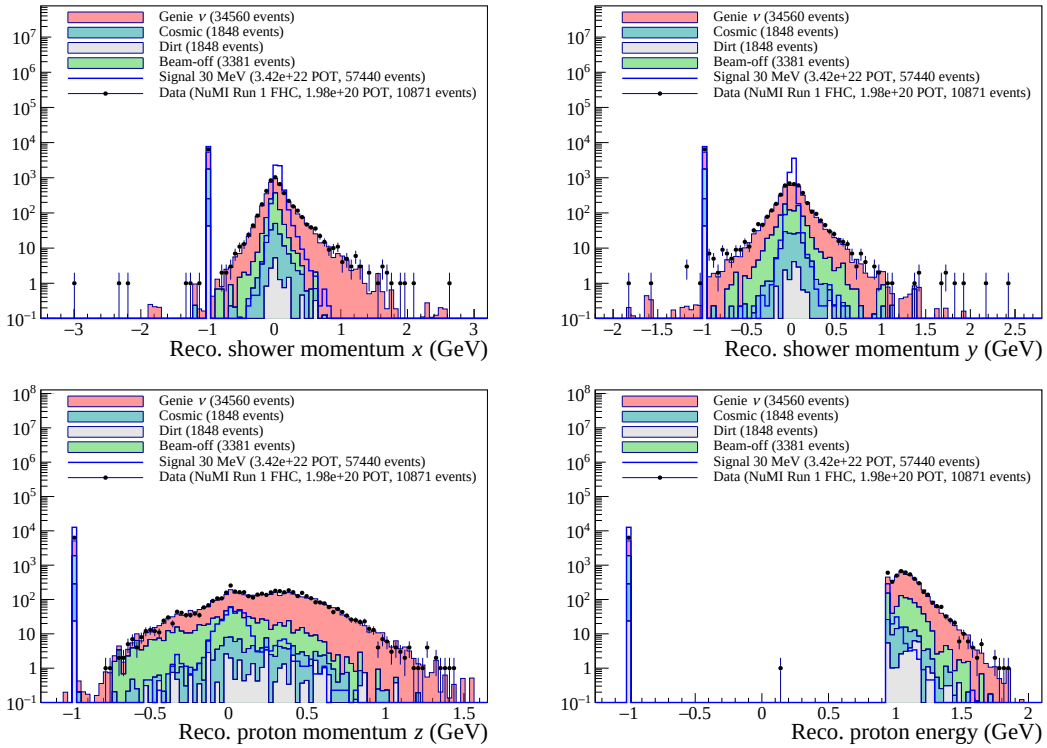


Figure 6.8.: Data and Monte Carlo signal and backgrounds comparison for proton and shower reconstruction variables. The filled coloured stacked histograms represent the different backgrounds, the blue histogram is a mCP signal of mass 30 MeV, and the markers are data. The backgrounds are normalised to data POT, while the signal is scaled to a higher POT for better visualisation.

a proton. The fraction of signal events in which a shower is reconstructed is larger, in which case it is of very low momentum in x or y. The absence of the z coordinate, or the x and y coordinates in the case of proton momentum, is simply due to our choice to focus only on the top 10 variables. Expanding our analysis to include the top 20 variables would reveal their presence as well, due to similar reasons, i.e. they encode information about showers or protons in the events. In the reconstructed proton energy shown in the bottom right histogram in figure 6.8, showing the reconstructed proton energy, we observe an anomalous data point at around 0.14 GeV. While this may be indicative of a bug, and will be investigated further, we do not believe it reflects a widespread issue or a cause for concern.

6.3.2. Wire-Cell BDT results

The receiver operating characteristic (ROC) curves for our BDT are shown in figure 6.9. ROC curves are a standard metric for evaluating the performance of binary classifiers, displaying the background rejection as a function of signal efficiency. In an ideal classifier,

the curve remains at a background rejection of 1.0 until the signal efficiency reaches 1.0, at which point it drops to 0.0. A common way to summarise the quality of a ROC curve is through its area under the curve (AUC), with a value close to 1.0 indicating a near-perfect classification. By contrast, a linearly decreasing ROC with an AUC of 0.5 — shown in figure 6.9 as a dashed red line — would mean that the classifier has no discriminating power, and is equivalent to random guessing.

In our case, the ROC curves shown in figure 6.9 are obtained by applying the BDT to our signal validation samples and the backgrounds from each respective Run period to obtain their BDT score distributions. We observe minimal variations across different Runs and horn current configurations, with the lowest AUC still reaching 0.978, indicating consistenly strong performance. As a result, we do not need to train separate BDTs for each MicroBooNE Run, saving computation time. The ROC-AUC metrics are also mCP-mass independent, only slightly favouring high masses, despite the signal efficiency of these being lower.

We show the (logit of) BDT scores in figure 6.10 for data and simulation of Runs 1, 2 and 3. In this figure, we distinguish between a "signal region" for scores between 0.0 and 6.5, and a "full region", showing scores from -10 to 6.5. We see that the values below 0.0 are vastly dominated by neutrino backgrounds. The Beam-off and cosmic components present a more uniform distribution, and represent the highest backgrounds in the signal region. The variable size binning is chosen to ensure sufficient statistics in the highest and lowest bins, especially when considering the systematic uncertainty samples which will be discussed in the next chapter. We see agreement between data and the sum of backgrounds, with a slight deficit of data events in the low-score regions and a slight excess in the high-score regions, but which is still mostly within the error margins. The error bars in figure 6.10 consist of the sum of statistical MC uncertainties, plus the systematic uncertainties which will be discussed in the next chapter, added in quadrature.

We find two events in the highest-scoring bin. Both events belong to Run 1, and Wire-Cell has reconstructed them as electrons of 64 and 170 MeV respectively. We show event displays of these in figure 6.11.

6.4. Two-blip selection

For our two-blip search channel, our strategy will be to study *blip-pairs* as an entity, and use this to train a BDT and classify each pair of blips in the event. Doing this allows us to use information which any single blip does not possess, such as orientation. However, if the number of blips in any event is of order ~100, then the number of unique blip-pairs in each event will be of order ~1000, by combinatorics. If our signal contains only one pair, this means a 1:1000 signal to background ratio. We will therefore perform a preselection on single blips, then construct the blip-pairs, and perform further selections on the blip pairs.

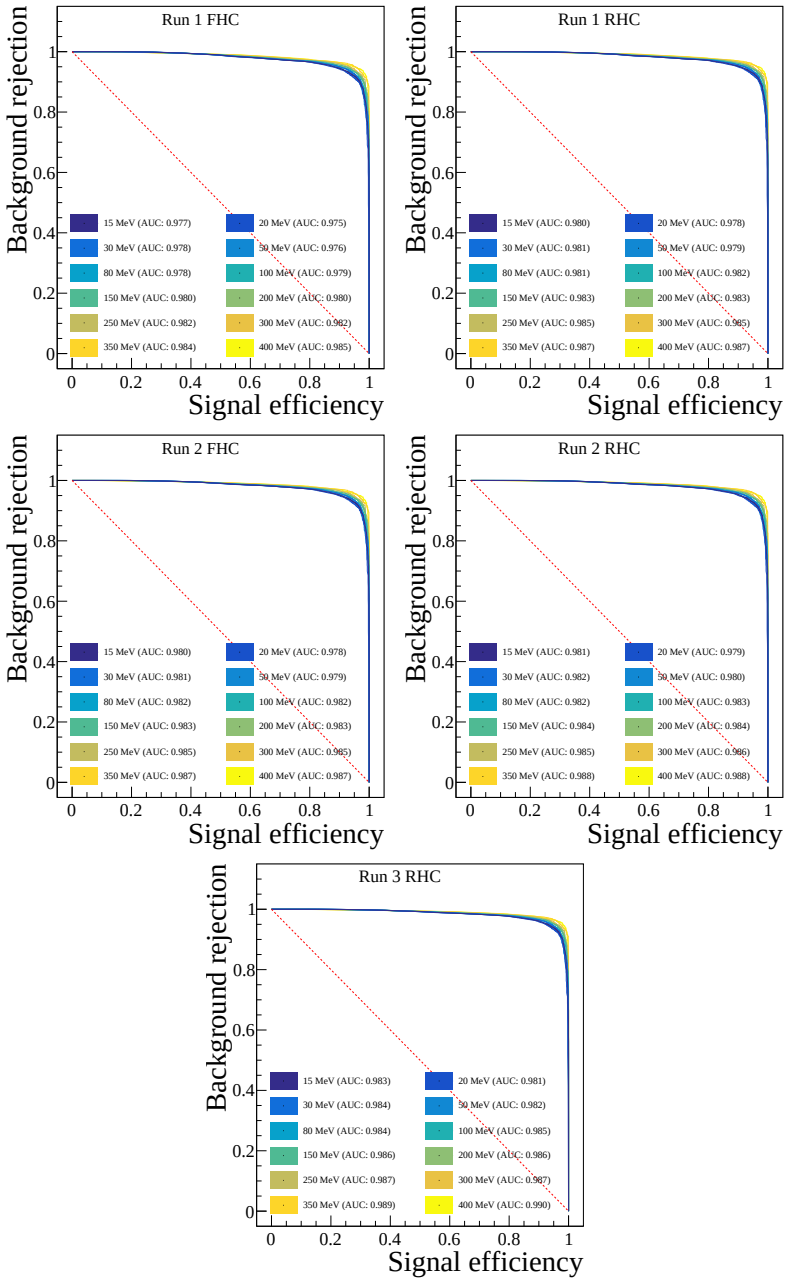
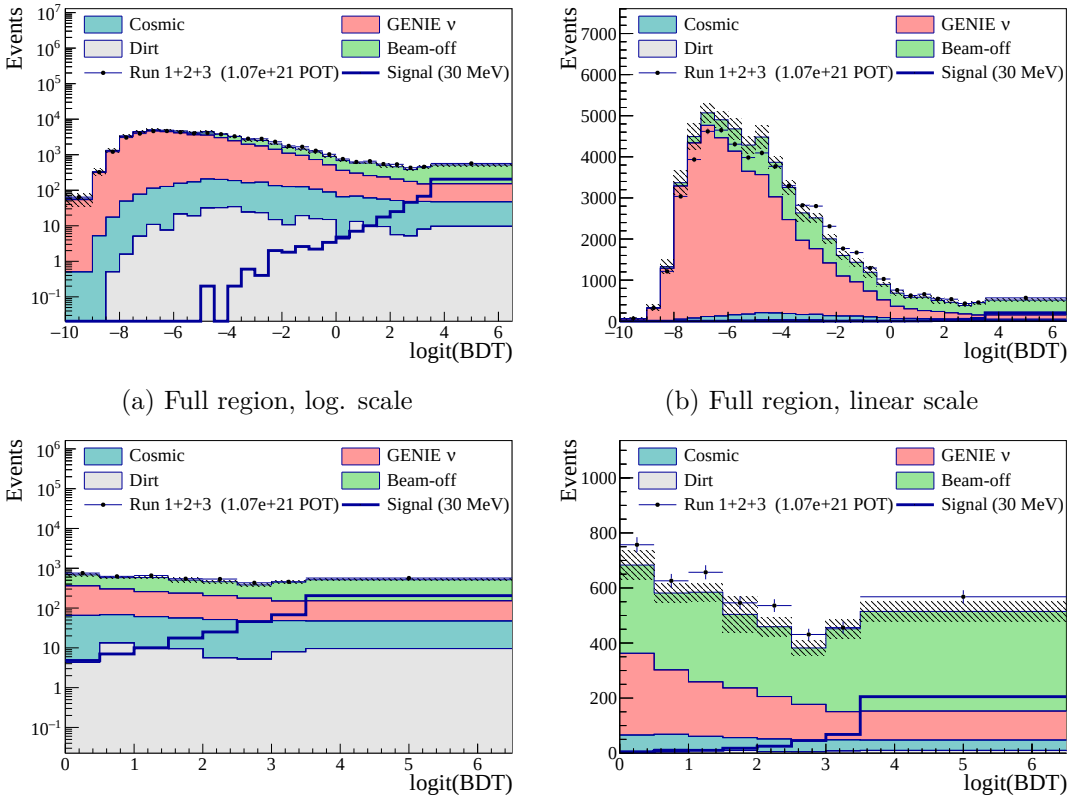


Figure 6.9.: ROC curves for different MicroBooNE Runs. The curves for different mCP masses are shown as lines following a colour palette. The legend shows the value of the area under the curve (AUC). The dashed red line indicates the ROC curve of a classifier with no discriminating power, shown for contrast.



(a) Full region, log. scale

(b) Full region, linear scale

(c) Signal region, log. scale

(d) Signal region, linear scale

Figure 6.10.: BDT scores for Runs 1, 2 and 3 data, shown alongside background prediction and 30 MeV signal.

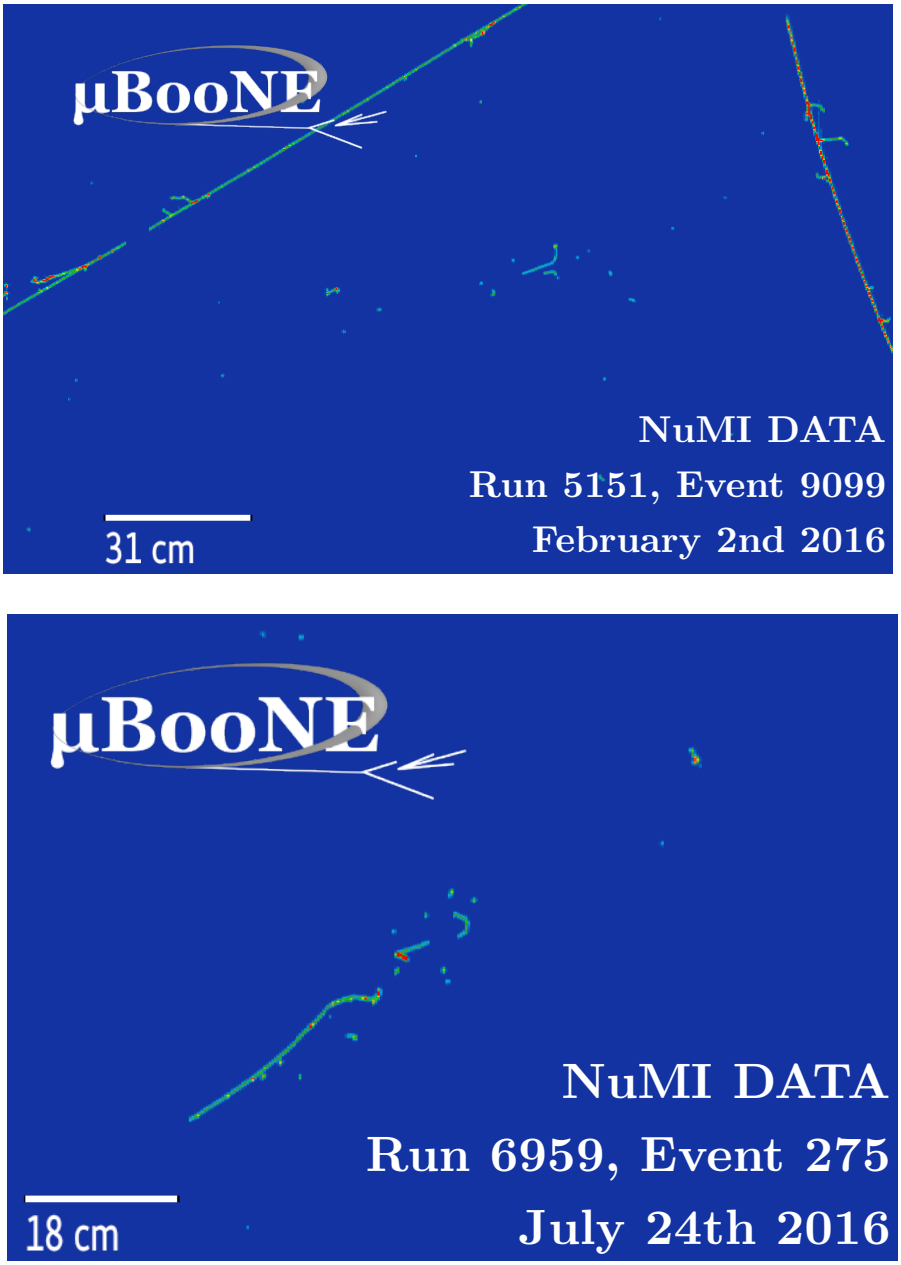


Figure 6.11.: Event displays showing the highest-scoring data events in the collection planed, as ranked by the Wire-Cell BDT. In the top panel, Wire-Cell reconstructed an electron with an energy of 64 MeV; in the bottom panel, an electron with 170 MeV.

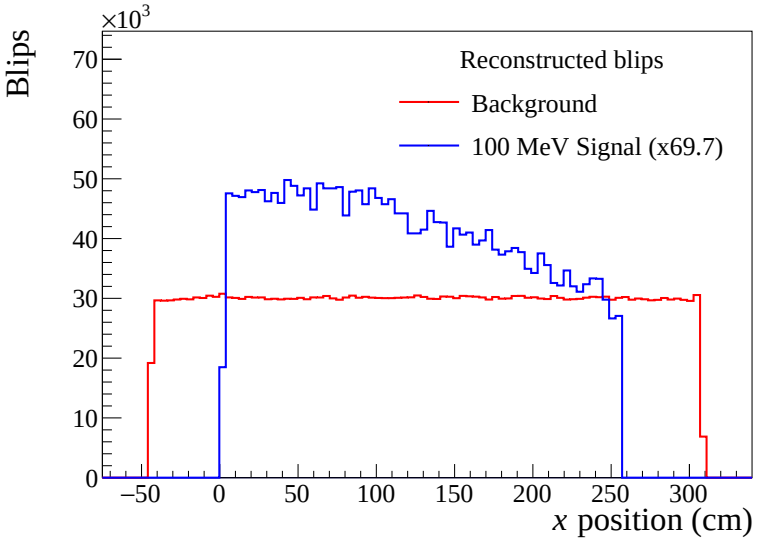


Figure 6.12.: Signal to background comparison for blip reconstruction of the x position, before any selections are applied. Blips originating from mCPs are shown in blue, while background blips are in red. Signal is normalised to have the same integral as background, for easier visualisation.

6.4.1. Single blip selection

As we have mentioned previously, the TPC does not have the ability to reconstruct the x position of a charge deposition by itself, it can only do so when matching to a flash from the light-detection system. For this reason the blip reconstruction uses the full time-window, as to not miss blips which may have been light-mismatched and thus lost. Blips by nature will have low energy, and thus low scintillation, and are more susceptible to this effect. In our simulation however, we observe that the x position of signal is matched correctly. In figure 6.12 we see a comparison of signal and background. It is safe for us then to select blips which are reconstructed within the nominal drift distance of $(0.0 \leq x \leq 260.0)$ cm in the TPC, without loss of useful signal. In figure 6.12 we also observe attenuation of the number of signal blips reconstructed far from the anode.

Previous studies of MeV-activity in MicroBooNE have found the existence of blip *hot-spots* coincident with the G-10 ribs used to support the field cage tubes [19, 106, 107]. We show the impact of removing these hotspots in figure 6.13. When comparing to no selection, the removal of the area around the G-10 ribs, and a harsher removal of an entire region in the YZ plane, 25 cm on each side of the Y axis and 40 cm on each side of the Z axis, the signal to background ratio increases.

At the end of these two selections we have effectively "fiducialised" the detector to a

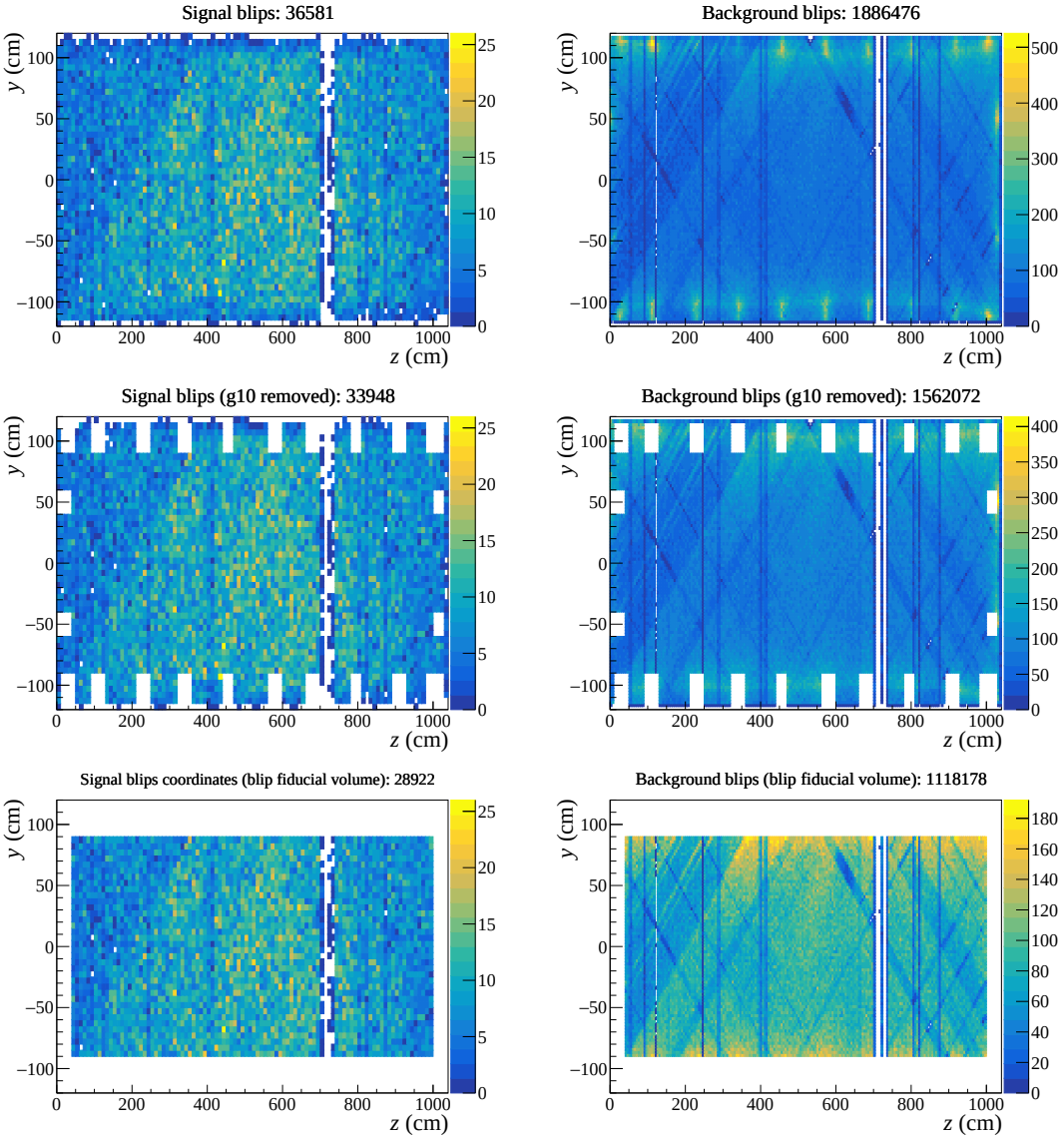


Figure 6.13.: Distribution of (left) signal and (right) background blips in the YZ plane for (top) no selection, (centre) cutting the G10 strut position and (bottom) removal of a slice in the YZ axis. The data shown corresponds to a sample of around 34,000 events of 100 MeV mCPs signal overlaid with beam-off data, after applying the previously described selection on the X axis.

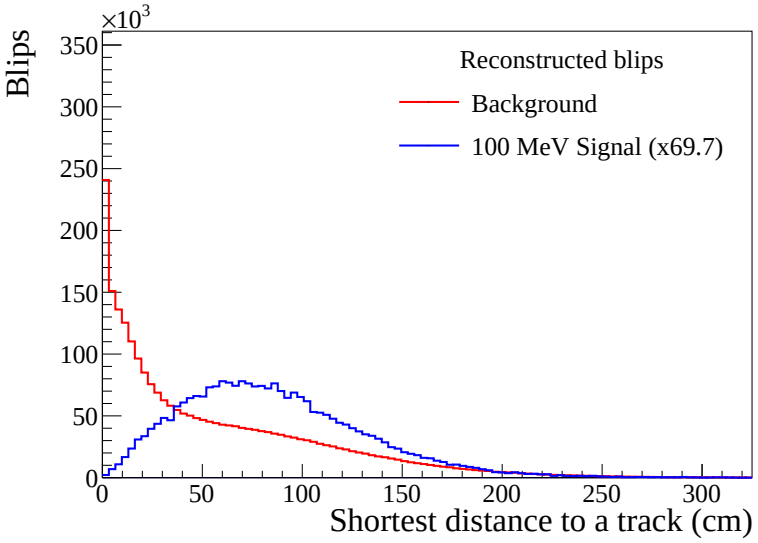


Figure 6.14.: Shortest distance of signal (blue) and background (red) blips to the nearest track in the event, after selections have been applied in the X direction and YZ plane. Signal is normalised to have the same integral as background, for easier visualisation.

volume contained within

$$\begin{aligned} 0.0 \text{ cm} < X < 260.0 \text{ cm}, \\ -90.0 \text{ cm} < Y < 90.0 \text{ cm}, \\ 40.0 \text{ cm} < Z < 1000.0 \text{ cm}. \end{aligned}$$

With these selections, the number of unique blip-pairs is significantly reduced, and the signal-to-background ratio increased.

A further selection we could perform is based on the distance between blips and the nearest track in the event. As cosmic muons traverse the detector, they tend to produce blips close to their trajectories. In contrast, blips originating from mCPs are not correlated to cosmic muon tracks; their distance to the nearest track is instead determined by the distribution and number of tracks in the event. This can be seen in figure 6.14, which shows the shortest distance of blips to the nearest track. A large fraction of background blips lie within 10 cm of a track, whereas signal blips most commonly fall in the50 cm to 100 cm range. This variable shows a clear separation between signal and background which we could use to perform a selection which optimised sensitivity. However we will not to apply this selection at this stage, and instead pass this information to the BDT. The main goal of the preselections described in this subsection was to optimise computation time by reducing the number of unique blip-pairs, which has now become sufficiently manageable. Further selections on blip-pairs will be much more powerful that what we can do at single-blip level.

6.4.2. Blip-pair selection

We construct blip-pairs by looping over each blip in the event and, for each one, looping over the remaining blips, while checking that they are both within the fiducial volume described in the previous section. The blips in each event are indexed, by which we avoid double counting a pair. Some of the training variables we will use (described in the next subsection) take into consideration a third blip, which is the nearest to the blip-pair in some capacity. For this, we also iterate over all blips not in the pair, without double counting.

Blip-pairs are formed from two blips, which we label blips 1 and 2. The pairs are formed such that the position in z of blip 2 is always greater than that of blip 1. This helps to keep the angular variables consistent, and gives us information on the lowest and highest z positions which would otherwise be randomly distributed.

We can separate the blip-pairs into three categories: signal, mixed and background, for the cases in which two, one, or zero blips in the pair are due to a mCP interaction. We will train our BDT using only the signal and background blip-pairs, as we will see that *mixed* pairs have only limited classification power. In this section, we will show the properties of all three of them, and show the results in the following section.

By far the biggest indicator of a mCP interacting multiple times with the detector is the orientation of the interactions, which in our case is the angles of our blip-pair. This is basically the strategy used in the ArgoNeuT search [17, 18], however in our case the angle corresponds to the position of the NuMI beam target with respect to MicroBooNE. We show the distributions of signal, mixed and background blip pairs in figure 6.15. We can see similarities between the background and mixed distributions, which possess features due to the geometry of the detector; a small ϕ angle indicates that both blips are on the same horizontal plane rather than vertical, while a small θ angle indicates small deviation from the Z axis, both of which are explained from the detector being much longer in the Z direction. The distribution of signal blips, on the other hand, is confined to only a few bins around the angle from the NuMI target, for which a preselection is applied according to

$$27.0^\circ < \theta < 28.5^\circ,$$
$$3.0^\circ < \phi < 8.0^\circ.$$

The results of this preselection, which is applied to the same sample shown in figure 6.15, is shown in figure 6.16. We see that we have removed a significant amount of background. Our mixed pair distribution in particular is nearly uniform and vastly reduced, but we have retained the region with the most signal.

As a final comment, we note that a natural consequence of considering blip-pairs is that there may be more than one blip-pair per event for both signal and background. As the POT information is given by the sample as a whole, we do not need to perform any further normalisation; in our final analysis we should however perform checks on the rate of events which contain more than one blip-pair, as the presence of an event with more than one blip-pair classified as signal would be indicative of a 3+ blip interaction.

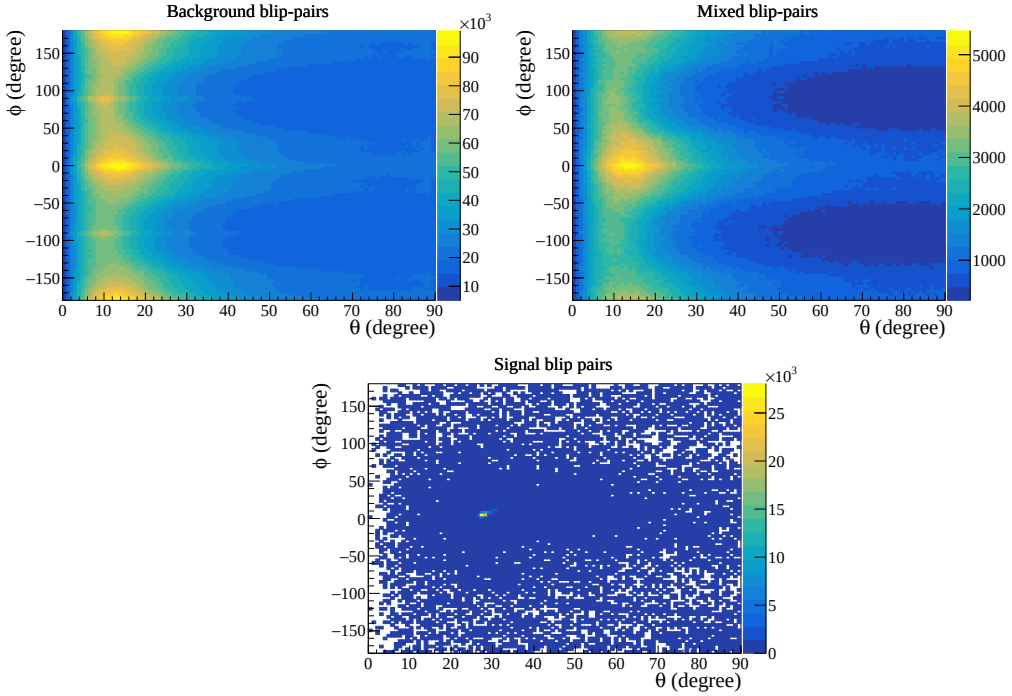


Figure 6.15.: Coordinates of the blip-pair in angular space in MicroBooNE coordinates.

6.5. Two-blip BDT

6.5.1. Blip-pair BDT input variables

We construct 28 variables from our blip-pair selection to train a BDT. A summary table of the variables can be found in figure 6.3, where they have been separated into categories. We will describe these categories and the variables in detail within in this section, however we will display figures with comparisons only for those with visible distinguishing features. The full set of signal-background comparison figures can be found in appendix A.

Per-blip variables, index-wise

We have a set of "per-blip" variables, in which we record information of the constituent blips 1 and 2 according to their index. Blips in the pair have indices 1 and 2, where by construction blip 2 is always downstream from blip 1 ($z_2 > z_1$). This makes these variables correlated to the geometry of the detector.

We use the six blip position coordinates ((x_i, y_i, z_i), where $i = 1, 2$) for training. The blip-position variables are shown in figure 6.17. In this figure, we observe a higher concentration of background blips towards the edges of the detector in the x and y axes, whereas the mCP signal blip distribution is more uniform in these axes. In the z axis we observe the gap in the region between $700 < z < 740\,\text{cm}$ due to dead wires

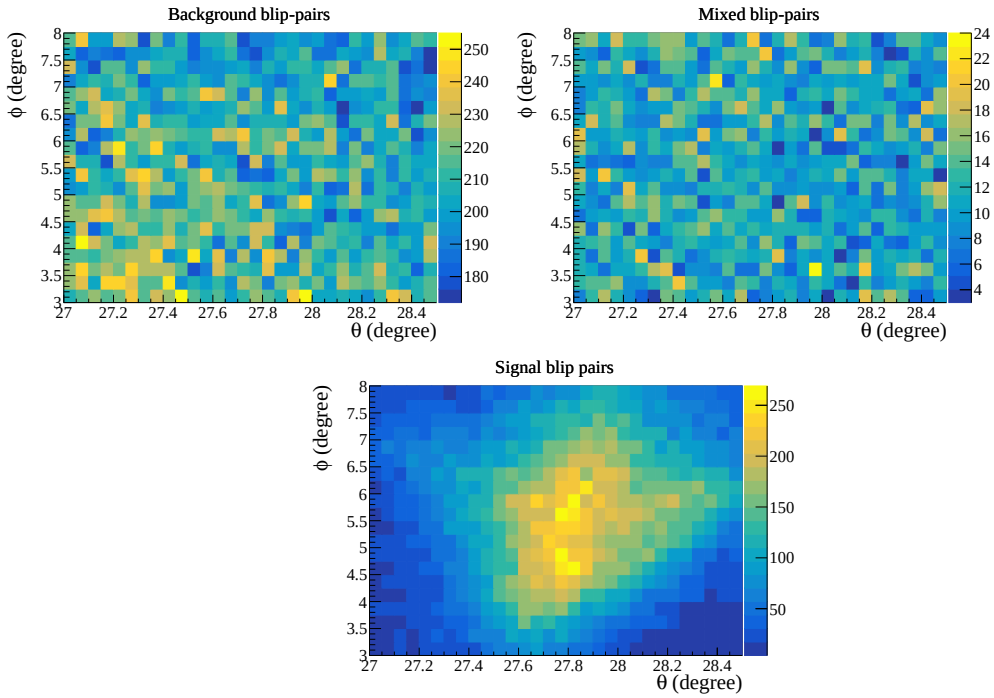


Figure 6.16.: Coordinates of the blip-pair in angular space in MicroBooNE coordinates after applying a narrow selection in their θ and ϕ angles.

Table 6.3.: Summary of blip-pair BDT training variables, separated in four categories.

Category	Variables	N. variables
Per-blip, index-wise	Blip i positions (x_i, y_i, z_i)	6
(blip 1 and blip 2)	Blip i extents Δx_i, $\Delta(yz)_i$	4
Per blip, magnitude-wise	Blip size	2
(blip min. and blip max.)	Blip energy	2
	Number of hits in the collection plane	2
	Sum of number of hits in induction planes	2
	Minimum distance to a track of each blip	2
Blip pair	Blip angles	2
	Blip separation	1
Third blip	Transverse distance of nearest 3rd blip	4
	Overall distance of nearest 3rd blip	1

in the collection plane, affecting signal and background equally. In this figure — and throughout all figures comparing signal and background training samples in this section — the signal and "mixed" distributions will be normalised to have the same integral as the background blip-pairs, to aid the visual comparison in the shapes of the distributions.

For signal, the electron scatter positions and energies are generated randomly and independently of each other; therefore the blip extent (which is correlated to energy) should not correlate to the index (correlated to position). However, this is not necessarily the case for background, as the electric field or other unaccounted effects may depend upon the spatial coordinates. In any case, we will include the blip size, which is correlated to the extents via $\sqrt{(\Delta x)^2 + (\Delta(yz))^2}$ as a non-indexed variable, so we will essentially study extent information in two ways: indexed and not indexed. These variables do not show strikingly different features between signal and background.

Per-blip variables, magnitude-wise

Per-blip information can also be separated index-independently according to their magnitude. These are the cases where we can define minimum and maximum magnitudes that do not necessarily correlate with the spatial position information stored in the indices.

Blip size, blip energy, and the number of hits in the collection plane are all examples of such variables, and we record the largest and smallest blip size, maximum and minimum blip energy, and maximum and minimum hits in the collection plane, for both blips in the blip-pair.

We use the sum of number of hits in the induction plane as blips may have matching hits from one or two of them (so called 2D or 3D blips). In the case of no matches, the number of hits on that plane is set to a default unphysical value. If we used four variables, for the minimum and maximum in each plane, many of the minima would be recorded with this default value; BDTs are able to handle this sort of feature, however summing the values in both planes also gives these discrete integer variables a wider usable range.

Finally, the blip reconstruction saves the distance from each blip to the nearest track. If for example blip 1 is closer to a track than blip 2, we record that distance as the minimum, and the distance of blip 2 to its closer track as the maximum. The maximum distance from any blip to its nearest track is notable, as it is the only variable in which the *mixed* sample is noticeably more similar to signal rather than background. This is due to the fact that background blips are more likely to be closer to a track (hence in the "minimum" distribution) whereas signal-blips are less likely to be close to a track. As mixed-pairs by construction have one signal and one background blip, this variable has differentiating power between signal, mixed, and background blip-pairs, and could be used for a targeted retraining, should we choose to consider the mixed channel as either signal or background. This classification power would have been lost had we performed a hard selection at the single-blip level. We display these features in figure 6.18.

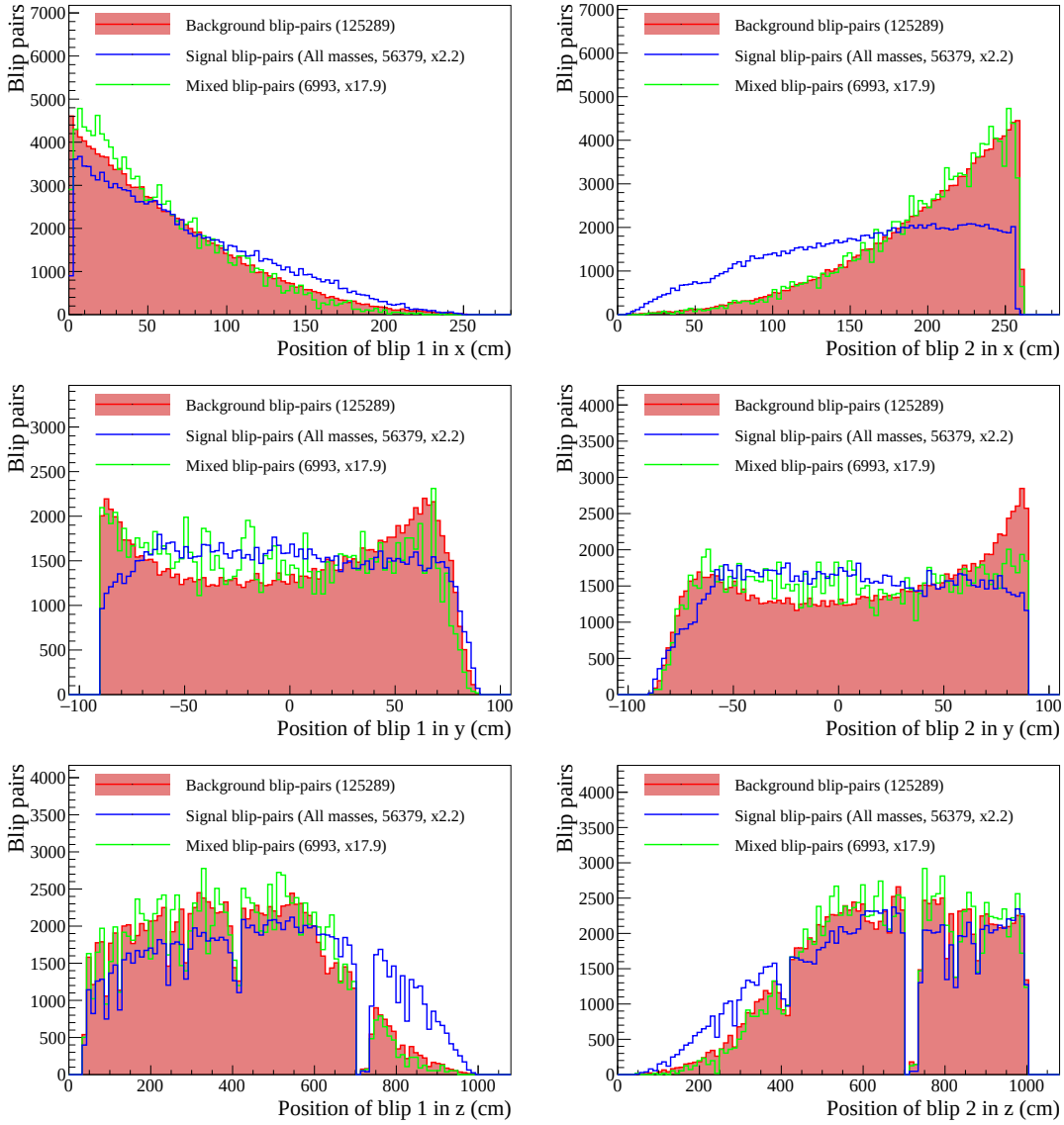


Figure 6.17.: Per-blip index-wise BDT training variables. The figures show the (x, y, z) positions of blips 1 and 2 for background (filled red histogram), signal (blue) and mixed (green).

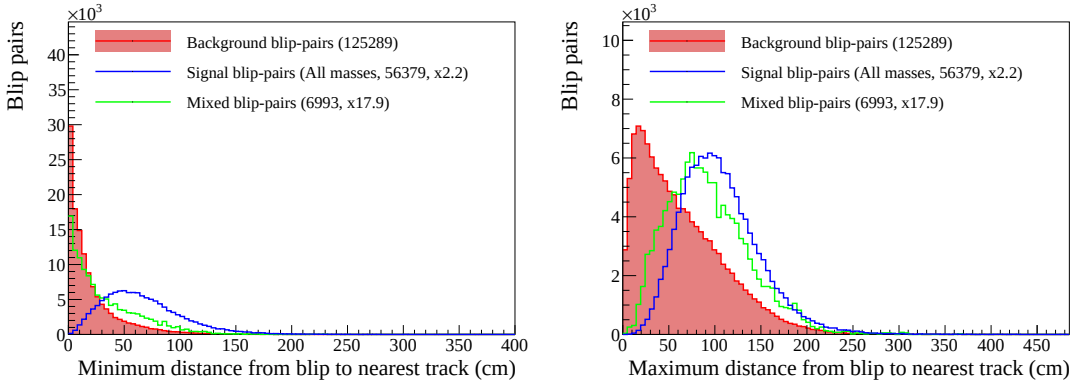


Figure 6.18.: Per-blip magnitude-wise variables. Left (right) shows the minimum (maximum) distance from any of the two blips to the nearest track.

Blip-pair variables

Blip-pair variables consider the blip-pair as its own entity. There are only three such variables, however they have by far the greatest distinguishing power between signal and background, and will be the most important for BDT training. Their distributions can be found in figure 6.19.

First, we have the angles θ and ϕ of the vector joining the pair, in MicroBooNE coordinates, which has been discussed in the previous selection section. The background and signal samples follow a flat distribution, while the signal samples have a marked peak at the centre; this is natural as we have made an angular preselection targeted at this.

The distance among the blips in the pair, or blip separation, is the magnitude of the vector joining the pair. In our generator, electron scatters are generated with a uniform separation along the trajectory of the mCP in the detector, which results in a linearly decreasing distribution when considering the finite detector length. The separation between background blips on the other hand will be dictated mostly by the rate of cosmic muons, which as we have discussed drives blip position.

Relation to third-blip variables

The final category of training variables considers the position of a third blip in relation to the blip-pair. We define three regions in the detector using two planes intersecting the blip-pair in consideration, and perpendicular the direction joining the blip-pair; we call these *behind*, *between* and *ahead*, according to the position relative to the blip-pair and (implicitly) the z direction of the detector. A diagram of these regions and the training variables is shown in figure 6.20.

For each blip-pair, we loop over all the remaining blips in the event and calculate their distance to either the nearest blip in the pair (when behind and ahead) or the centre of the line joining them (when in between). We record the minimum transverse distance $R_\perp$ to the line joining the blip-pair in all three regions, as well as the overall minimum

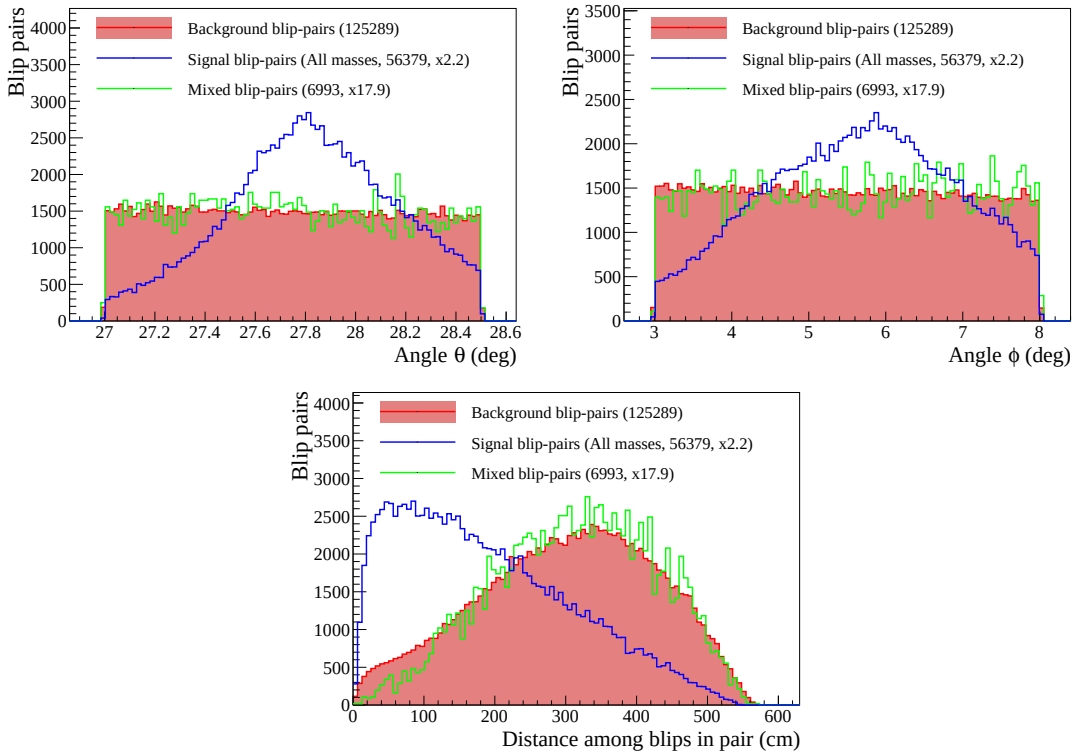


Figure 6.19.: Blip-pair BDT training variables. Top: angle variables θ and ϕ. Bottom: Blip separation. These variables show good separation between signal and background.

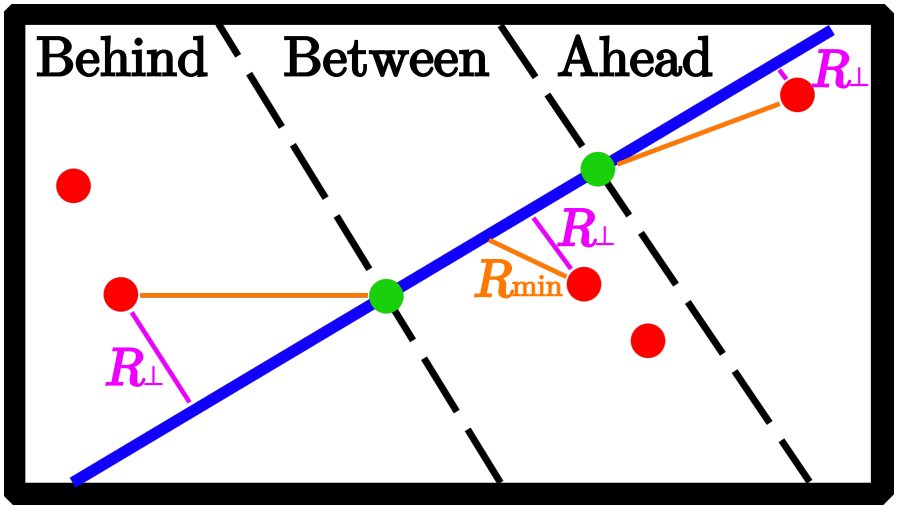


Figure 6.20.: Diagram showing the "third-blip" training variables. Green circles are the blip-pair in consideration, with the line joining them shown in blue. The dashed black-lines separate the remaining blips in regions behind, between, or ahead of the blip-pair. Red circles are other blips in the event, over which we loop to find the nearest blip in each region. The distances to the nearest blips are shown in orange, and their transverse distances $R_\perp$ in magenta. In this example the nearest "between" point is also the one with overall minimum distance $R_{\min}$, while the "ahead" blip has the minimum $R_\perp$. This is a 2D diagram, but all of these distances are calculated in 3 dimensions.

$R_\perp$ in the three regions. The reasoning behind these variables is that multiple scatters along the same line (low $R_\perp$) is a signature of scatters due to mCPs. We also record the magnitude of the minimum distance vector to any third blip in the event.

The distribution of these variables for the signal, background and mixed pairs is shown in figure 6.21. It should be noted that in each event, there will always be one blip furthest to the left and one furthest to the right, in which cases the respective regions would be empty. Additionally, there may be pairs for which there are no blips in between. In all these cases a default distance value of $-4\,\text{cm}$ is used, which is reflected in the horizontal axes in some panels in figure 6.21 not starting at zero. The "closest distance overall" could also have empty values if there is exactly one blip-pair after performing selection in the event, and thus no third-blips to compare to, but we did not observe this case in any of our samples.

6.5.2. Blip-pair BDT training

We train a BDT using a sample of combined masses of signal and background, with roughly the same number of events for each mass signal sample. The number of total signal pairs is around 60 000, and the number of background around 125 000 pairs. A fraction of 80% of the events is used for training, with the remaining 20% used for testing

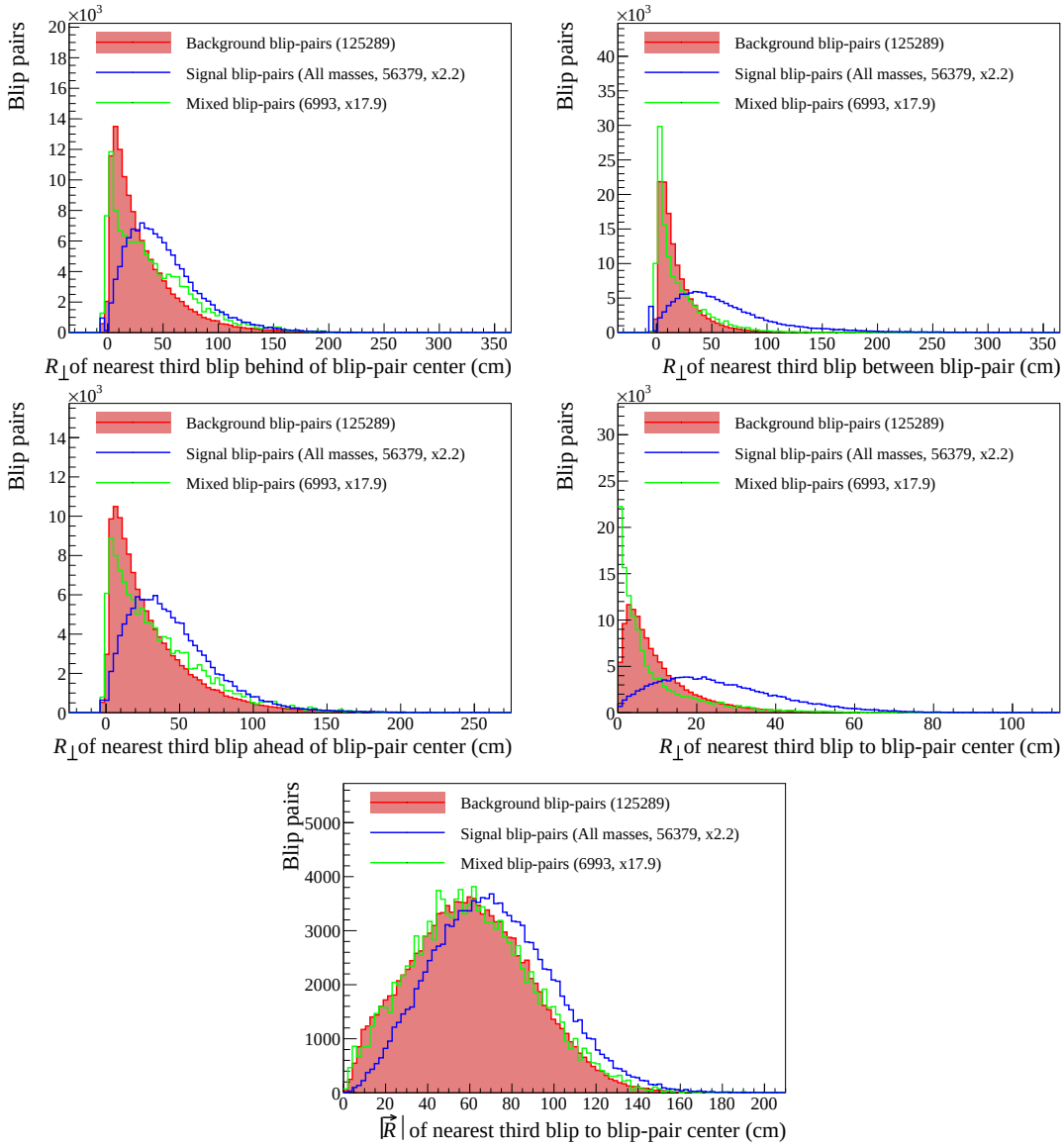


Figure 6.21.: Third-blip BDT training variables.

the performance of the BDT while training.

During training, the log loss, error and AUC metrics were evaluated for both training and testing samples. These metrics also determined whether early stopping would be triggered, based on the "Early stopping rounds" parameter of 20. Training stopped after 421 rounds, and the log loss, error, and AUC metrics measured during training are shown in the top panels in figure 6.22.

Overtraining would manifest as a clear separation in BDT score distribution between the training samples and independent validation samples. In the bottom panels of figure 6.22, we compare these distributions for two representative mass points of 30 MeV and 350 MeV. The validation samples have been scaled to match the number of entries in the training samples, to better visualise any shape differences. No significant shape differences are observed, indicating that overtraining is minimal. This behaviour is consistent across all other mass points, not shown in this figure.

The bottom panels of figure 6.22 also include the "mixed-pairs" validation sample. Mixed pairs were not used during training, and therefore can not be overtrained. We show their BDT score distribution to demonstrate that they closely resemble the background distribution, with only a slight bias towards higher scores. This slight separation may be leveraged in future analyses.

6.5.3. Blip-pair data-MC comparisons

We apply the BDT to a small beam-on sample of data equivalent to 2.786×10^{19} POT as described in section 5.10 to see the agreement between data and Monte Carlo (MC). We show the results for some of the most important blip-pair variables in figure 6.23, where we see good agreement between data and MC.

The GENIE and cosmic background samples in figure 6.23 have been normalised to the data sample POT. The Beam-off sample is normalised to the ratio of the number of beam-on versus beam-off triggers. The signal sample is normalised to have the same integral of the data, for better visualisation, otherwise it would not be visible. Its POT equivalent is shown in the legend.

We see that neutrinos from the GENIE simulation are not a significant contribution, and that the data can be explained very well only using the cosmic and beam-off samples, both representing the same type of activity from either cosmics or noise. This means that blip-pairs from neutrino interactions do not pass the selections we have applied.

The full figures of data-MC comparisons for the blip BDT can be found in appendix B.

6.5.4. Blip-pair BDT results

We apply the BDT to the samples shown in the previous section, to obtain a distribution of BDT scores for each pair, which are shown for in figure 6.24.

As we will not process the full dataset of Run 1, 2 and 3 for the blip-pair channel, we will scale the signal and backgrounds to 1.07×10^{21} POT and use these simulation-only samples to obtain a sensitivity for this channel. The scaled samples are shown in figure 6.25.

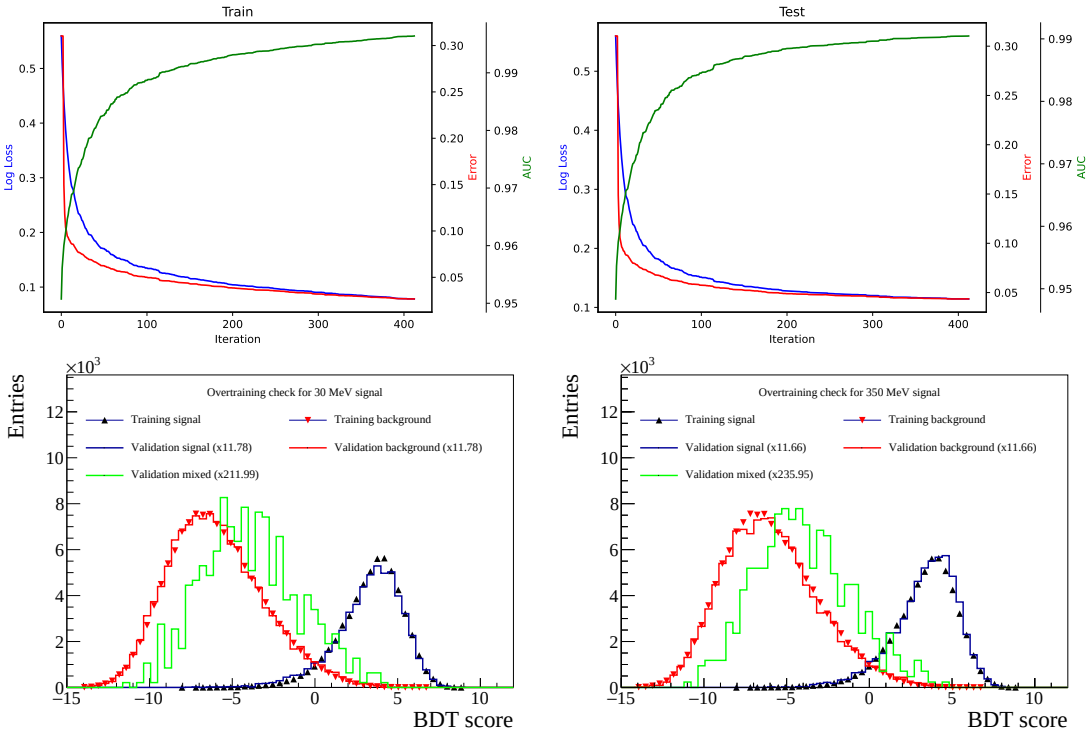


Figure 6.22.: Training evaluation metrics. The top panels show the log loss, error and AUC metrics during the training iterations, while the bottom panels show the training signal and background as blue and red markers, with the validation samples as lines.

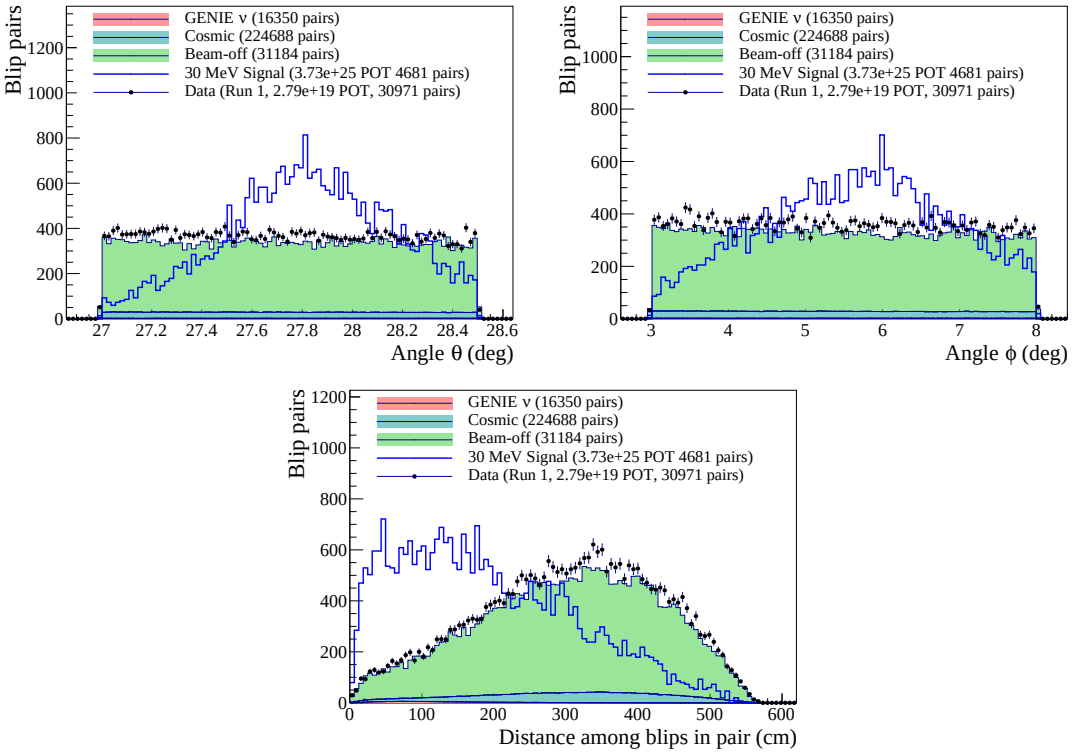
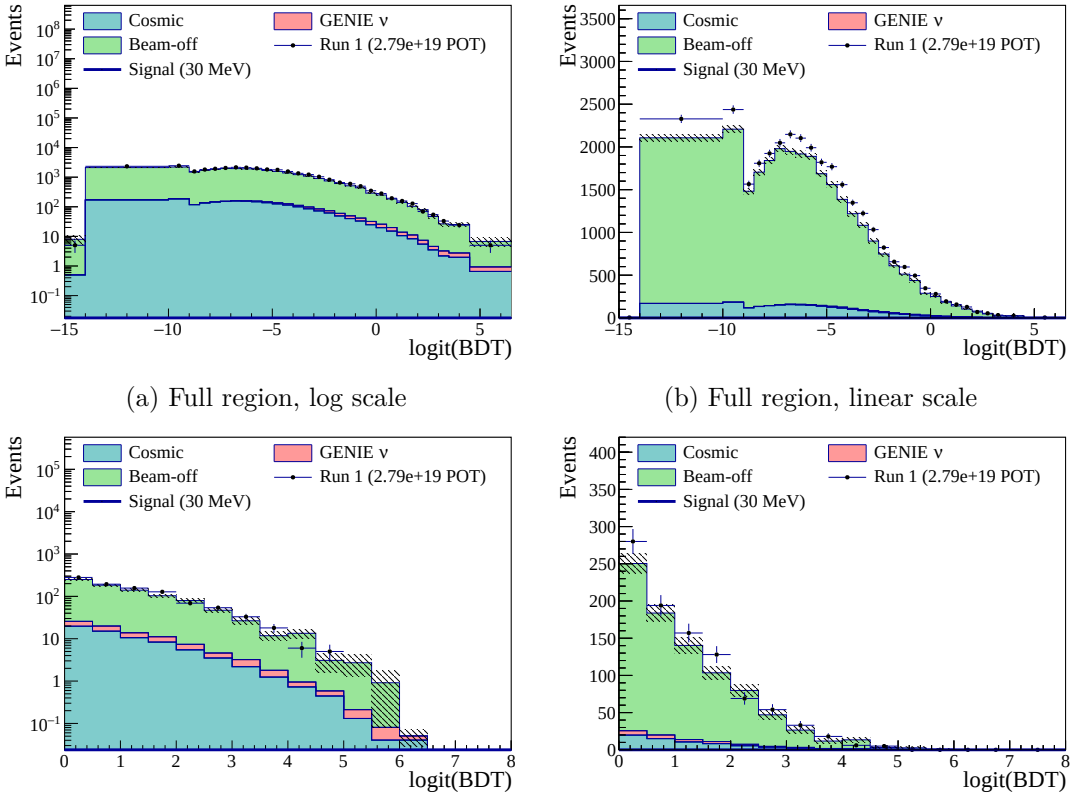


Figure 6.23.: Data-MC comparison for the blip-pair BDT training variables. The samples are normalised to data, except for the signal. The number of pairs in each histogram is shown in the legend. Top: angle variables θ and ϕ. Bottom: Blip separation. These variables show good separation between signal and background. We see good agreement between data and MC. The contribution from neutrino simulation is too small to be seen.



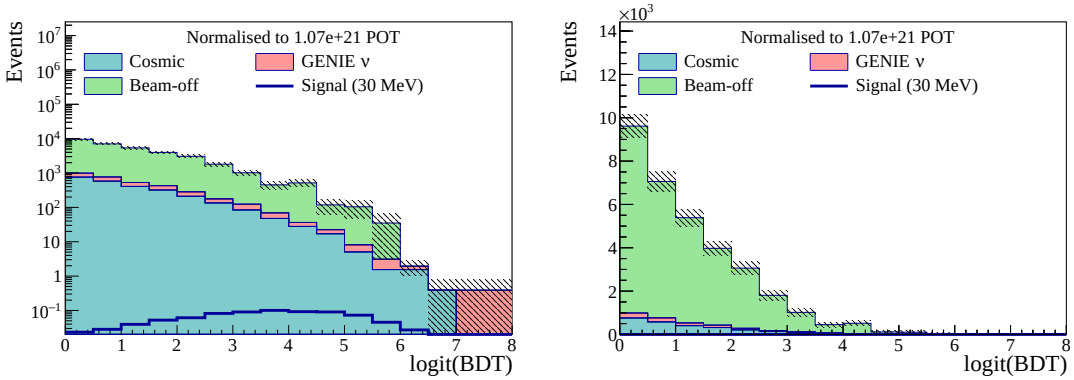
(a) Full region, log scale

(b) Full region, linear scale

(c) Signal region, log scale

(d) Signal region, linear scale

Figure 6.24.: BDT scores for a subset of Run 1 data, shown alongside background prediction and 30 MeV signal. All the samples are scaled to data. The top figures, 6.24(a) and 6.24(b), show the full range of BDT scores, while the bottom figures show scores above 0, which can be considered as our signal region. Figures on the left are shown on a logarithmic scale, while those on the right are on a linear scale in the vertical axis.



(a) Signal region, log. scale

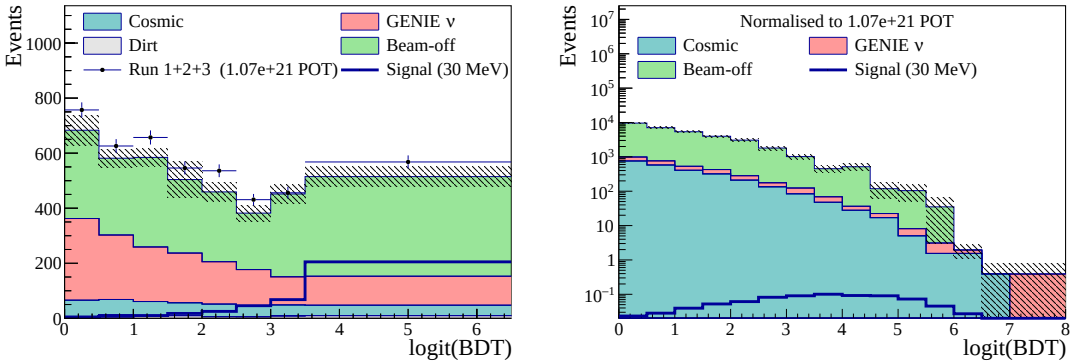
(b) Signal region, linear scale

Figure 6.25.: BDT scores for signal and background samples scaled to 1.07×10^{21} POT

6.6. BDT training summary

We have trained boosted decision trees for two channels (Wire-Cell single electron, and blip-pairs) using Monte Carlo simulation samples of signal and background, as well as beam-off data. We then used the BDT model to predict the BDT scores of independent samples of MC signal, MC background, beam-off data, and beam-on data. We obtained BDT score distributions and used the value of 0.0 to separate them into background and signal regions, with events below 0.0 completely dominated by background.

We observe good agreement between data and background prediction. In the case of the single electron scatter channel, we used data from the entirety of MicroBooNE Runs 1, 2, and 3, corresponding to 1.07×10^{21} POT, and we will use these to set limits on the mCP coupling. For the multiple-scatter channel, we only used a fraction of data corresponding to 2.79×10^{19} POT from Run 1. We then scaled the signal and background samples to that of an equivalent of 1.07×10^{21} POT, and we will use that to obtain a sensitivity for MicroBooNE in this channel. We will obtain this limit and sensitivity by passing the histogrammed distributions of BDT scores shown in figure 6.26 to a limit setting framework, which is described in chapter 8, following a discussion on systematic uncertainties.



(a) Wire-Cell BDT score distribution in the signal region for Runs 1, 2 and 3.

(b) Blip-pairs BDT score distribution for 1.07×10^{21} POT.

Figure 6.26.: BDT scores for signal and background samples scaled to 1.07×10^{21} POT.

7. Systematic uncertainties

In this chapter we will discuss the uncertainties which are applied to our samples. The uncertainty types considered are similar to those of previously published results from MicroBooNE, in particular, to those of previous BSM searches. Our search will be ultimately limited by the fact that we are looking for a very weak signal, and we have found that the uncertainties we chose do not affect the final results significantly.

The systematic uncertainties described here are only applied to the Wire-Cell search. We will present the *blip* channel as a sensitivity without systematic uncertainties, although preliminary studies show that the sensitivity is largely unaffected by different choices of uncertainties.

7.1. Flux uncertainties

The main uncertainty in the flux of millicharged particles and neutrinos is the hadron production at the NuMI target. In the case of the NuMI neutrino flux, these uncertainties are constrained using the `PPFX` software package [127], which uses data from fixed target experiments as a basis. The uncertainties for the NuMI neutrino flux on MicroBooNE have been found [36] to be of 22 %, which come from 21 % hadron production and 6 % beamline simulation added in quadrature.

The flux of millicharged particles, unlike that of neutrinos, originates from the decay of neutral mesons. Previous studies have shown [65] that the expected ratio of neutral meson production from `g4numi` when comparing neutral to charged mesons is what is expected from isospin symmetry, and therefore we are justified in using the same flux uncertainty of 22 %.

7.2. Monte Carlo statistical systematic uncertainties

Systematic uncertainties due to the limited number of Monte Carlo simulation events in our samples are considered as Poisson errors of $\sqrt{N_i}$ for N_i measurements in each bin i. They are calculated independently for each sample before performing POT normalisation, which is performed separately for each Run and horn current configuration. After POT normalisation, the scaled errors of each background sample are added in quadrature.

While the EXT sample does not correspond to Monte Carlo simulation but rather beam-off data, we mention here that its errors are also Poisson errors scaled by the number of triggers normalisation, which are added in quadrature to the other background samples (GENIE and Dirt).

7.3. Detector systematics

To evaluate the systematic uncertainties due to the detector response we perform Monte Carlo simulations under different detector configurations and with different physics parameters. These simulations of millicharged particle signal and neutrino background are a completely independent samples from those used in the main search. The same generated events are run through a modified version of the full analysis chain, which in the case of these "detector variations" now differs at one stage in which the special configuration is applied. After this point the events are treated as separate samples. The events in the initial sample are tagged, which allows us to remove events which are not present across all samples after the reconstruction stage, in order to avoid bias due to runtime errors. In addition to the variation samples, we also run a central value (CV) sample, which follows the unaltered event reconstruction chain and serves as a "nominal value" against which the variations will be compared.

The detector variations can be divided in three categories: *TPC systematics*, in which the interactions and propagation of particles within the detector argon volume are altered, *WireMod systematics*, which is a novel method developed by MicroBooNE [128] for evaluating variations in the position, time and charge as read out by the detector wires by directly modifying simulated waveforms and comparing to data, and *Light yield systematics*, which modify the propagation of light within the detector and parameters of the light detection system. Table 7.1 provides a brief summary of these 10 detector variations, which are described in Ref. [128] and in what follows.

Table 7.1.: Standard MicroBooNE detector variations, separated into categories.

Category	Systematic	Description
WireMod	X	Variations in x coordinate
	YZ	Variations in the $y-z$ plane
	θ_{XZ}	Variations in the θ_{XZ} angle
	θ_{YZ}	Variations in the θ_{YZ} angle
	dE/dx	Variations in the energy per unit length
TPC	SCE	Space charge effect
	Recombination	Recombination effect
Light yield	Down	25 % light yield reduction
	Rayleigh	Alternative liquid-argon Rayleigh scattering length
	Attenuation	Additional quenching and absorption

The TPC systematics are due to the space charge effect (SCE) and recombination effects, which have been briefly discussed in section 4.1. For the SCE systematic, an alternative sample is generated using a different SCE map. For the recombination effect, variations are performed on the two free parameters of the modified box model [86].

The WireMod variations are obtained by modifying the wire waveforms and comparing between beam-off data events and MC simulations. Variations are performed independently in the X direction and YZ plane, as Y and Z are highly correlated. Additionally,

there are variations in the θ_{XZ} and θ_{YZ} angles, which are cartesian angles with respect to the beam direction z rather than the spherical-coordinate based θ and ϕ angles we have considered. Finally, variations as a function of dE/dx are performed, although it has been found that this is highly correlated with the recombination variation.

The light yield "Down" variation simulates the effects of a 25 % reduction in overall light yield, as it was found through data/simulation comparisons that the simulation tends to overpredict the photo-electrons per centimetre of track length by this amount. The Rayleigh variation accounts for the uncertainty of the Rayleigh scattering length in liquid argon. MicroBooNE uses a nominal value of 60 cm, however other measurements have shown values as high as 100 cm. The variation used changes the light-yield in a position-dependent manner by scaling the ratio of the simulation between 120 cm compared to the nominal. Finally, the attenuation variation is used to account for a decline in light-yield which was observed for MicroBooNE in Run 3. The reasons for this decline are still under investigation, but it is measured in both anode and cathode, with a sharper decline at the cathode, which suggests both quenching and absorption components. This variation simulates a quenching of 20 and 40 %, and an absorption length of 8 and 13 m for the anode and cathode respectively. Despite the decline not being present in Run 1, we take a conservative approach and apply this variation to all Runs.

As the detector variations are computationally taxing, we used only two mass points (30 and 400 MeV) to test for differences between low and high masses. We find that the same detector variations are dominant in both samples, and no significant differences other than the fact that the 400 MeV sample requires more simulation due to the low reconstruction efficiency of this mass point.

The set of detector variations used in MicroBooNE have been chosen such that they have small correlations between them; thus we will add the detector variations in quadrature bin by bin, that is, for each bin i we obtain a value q given by

$$q_i = \sqrt{\sum_{\text{DV=det. vars}} \left(\frac{\text{CV}_i - \text{DV}_i}{\text{CV}_i} \right)^2}, \tag{7.1}$$

the term inside brackets $(\text{CV}_i - \text{DV}_i)/\text{CV}_i$ is a ratio representing the offset between each detector variation DV and the central value CV, in bin i. We show these ratios for each variation in each bin in figure 7.1 as coloured histograms. We see that most variations oscillate around a deviation of zero, with no clear pattern across bins. The exceptions to this are the light-yield attenuation and light-yield "down" variations, both of which consistenly result in a reduction of the reconstructed signal compared to the central value. This can be explained by the fact that our signals are very low energy electrons, which produce minimal scintillation light. Scintillation light is used to trigger the detector, hence a reduction in light-yield will reduce the trigger efficiency. Lowered light yield also impacts Wire-Cell selection, as discussed in section 5.7.1, reducing the selection efficiency.

Since their systematic effect is much larger than the other variations combined, and both result in a reduction of signal — which will result in a reduction of sensitivity

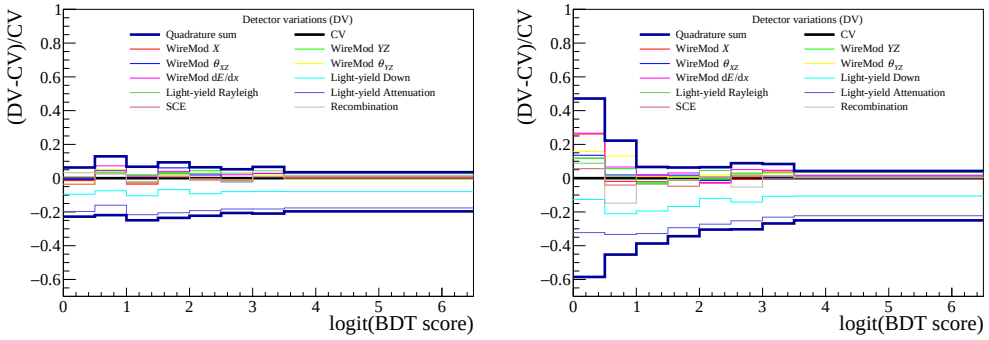


Figure 7.1.: Detector variations for millicharge signal samples of (left) 30 and (right) 400 MeV. Coloured histograms show the ratio of each variation (DV) compared to the central value (CV) sample. The bin-by-bin asymmetric quadrature sum as an envelope is shown as the outermost thick blue line.

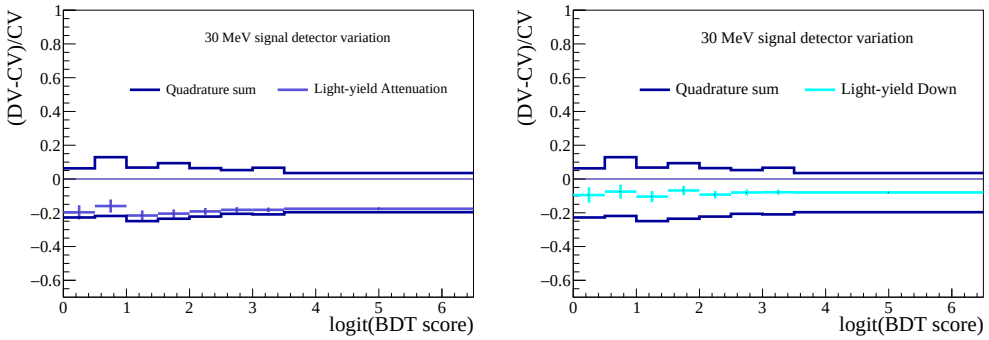


Figure 7.2.: Left: light-yield attenuation signal detector variation. Right: light-yield "down" signal detector variation. Both are shown for the 30 MeV sample. The asymmetric quadrature sum envelope is shown as the outermost blue line.

— we will use asymmetric detector systematics when adding in quadrature. In the negative direction, we add in quadrature all the detector variations, but on the positive side we exclude these two variations which have a clear downward trend. The resulting quadrature envelope can be seen in figure 7.1 as the outermost, thicker blue line. We show the detail of the light-yield attenuation and light-yield "down" variations in figure 7.2, where we also show the envelope for visual comparison.

Given that the quadrature sum obtained for the 400 MeV signal sample is bigger across all bins in the signal region, we will use that as our detector systematic uncertainty as a conservative estimate. This may seem excessive, given that the quadrature sum uncertainty is as high as 60 % in the lowest bin, and around 25 % at best. As previously mentioned, this choice does not affect the final result significantly when compared to using the quadrature sum from the 30 MeV sample, as will be shown in section 8.2.1.

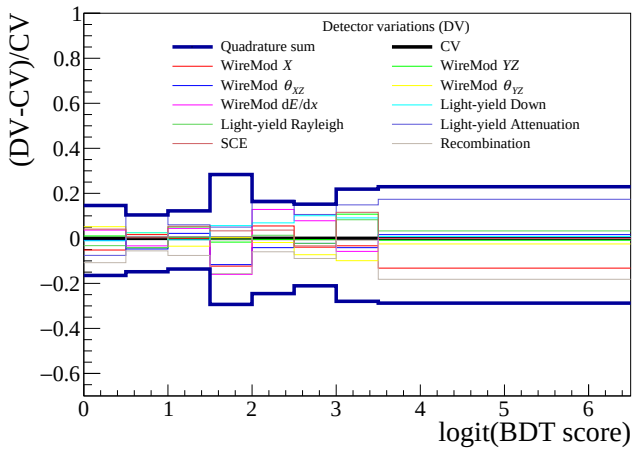


Figure 7.3.: Detector variations for the GENIE background neutrino sample. The ratio of each variation (DV) compared to the central value (CV) sample, as well as the bin-by-bin quadrature sum as an envelope are shown as coloured line histograms.

For the GENIE neutrino sample, the detector variations are shown in figure 7.3. There are no overwhelmingly dominating variations, and the uncertainty is driven by the limited statistics of neutrino interactions in the signal region of high BDT scores. For comparison we show the same light-yield variations which were dominant for signal in figure 7.4, where we now observe an increase compared to the CV. At first, it may seem counterintuitive that attenuation in light yield would lead to higher yields, but this is due to the fact that as we trained on a low-energy, low-scintillation signal the BDT has learned to classify the typically higher energy neutrino interactions as background. As the light yield on these events decreases, they may seem more signal-like. The uncertainties shown for each variation in figure 7.4 are due to the limited statistics in each sample.

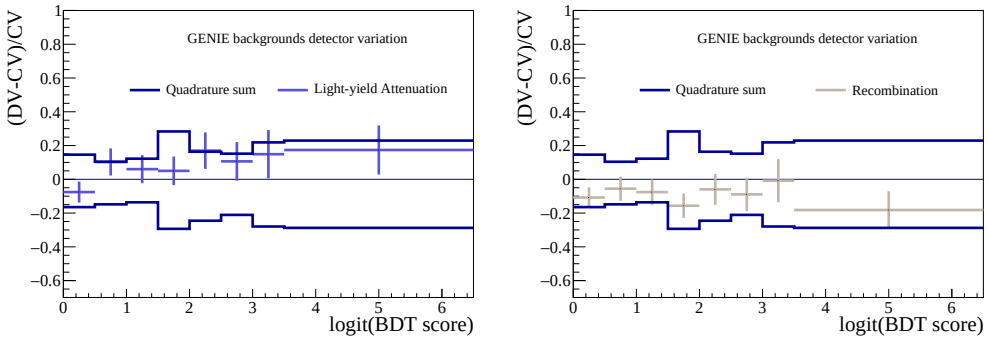


Figure 7.4.: The main contributors to background systematic uncertainty of the GENIE sample are (left) light-yield attenuation and (right) the recombination effect. Uncertainties are due to limited statistics of the samples.

8. Results

In previous chapters we have described the use of boosted decision trees (BDTs) to predict the probability that an event in data contains a millicharged particle (mCP) interaction. As we apply this BDT to our simulation and data samples, we obtain distributions of BDT-scores, from which we construct histograms. We have also considered the systematic uncertainties, constrained them, and used them to add error bands to our binned distributions.

In this chapter we will analyse these binned distributions using statistical analysis to test the hypothesis of whether we observe mCPs in MicroBooNE, and if not, what limits we are able to set upon their existence. For this we will use statistical analysis via specialised software which implements these mathematical tools, and apply it to our distributions.

8.1. Statistical data analysis

In terms of data analysis, the end-result of a particle physics experiment is a series of measurements, which usually take the form of a collection of real-valued variables per event, which can then be histogrammed. Our assumption is that, even if the specific values of the variables in our events are random, the underlying physics processes distribute the data in a specific manner, from which we can construct (or refine) a theoretical model. In statistics this is known as inference, and is illustrated in the diagram in figure 8.1. In experimental particle physics, the use of statistical inference is necessary due to the fundamentally random nature of quantum mechanics which underlies it.

There are numerous textbooks in which a detailed treatment can be found, such as Refs. [129, 130]. In this section we will give a brief overview of some fundamentals, with a focus on the methods which we will use in our analysis, particularly the concepts of hypothesis testing and limit setting.

8.1.1. Probability distributions, measurements, and parameters

Data can be distributed randomly with different probabilities in a discrete or continuous manner. An example of a discrete distribution would be the number of scatters of a millicharged particle in the MicroBooNE detector, which in equation (2.22) we have modelled using a Poisson distribution,

$$P\left(n|\nu\right) = \frac{\nu^n e^{-\nu}}{n!}. \tag{8.1}$$

A distinction can be made between the *measurements* themselves, which represent the physical quantities we measure directly (number of scatters), and the *parameters* (ν),

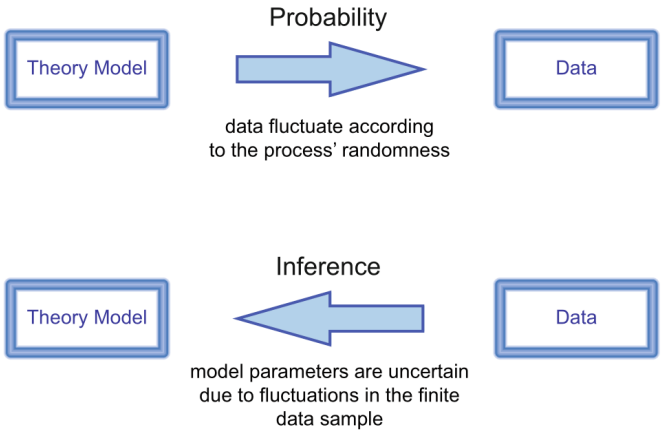


Figure 8.1.: Interaction between theory and data. We assume the distribution of the data arises from processes which can be modelled from theory, and the models can be refined with the study of more data. Figure from Ref. [129].

which are related to how the measurements are distributed. An example of inference would be to estimate the value of an unknown parameter ν (and its uncertainty $\Delta\nu$) given a number of measurements n and the knowledge that it must follow a Poisson distribution. We would assume that it is highly unlikely that consistently measuring a high n could be reconciled with a small ν. Formally, we define an *estimator* $\nu = \hat{\nu} + \Delta\hat{\nu}$, where the estimated values are denoted with a *hat*, and the non-hatted variable the true value of the parameter.

In general, we will work with continuous distributions, which when generalised are called probability density functions (PDFs). As functions denoting probabilities, they must be non-negative, and be normalised to 1. A PDF can have an arbitrary number of variables and parameters, which we denote as y and θ as in

$$f(y_1, y_2, \cdots, y_n|\theta_1, \theta_2, \cdots, \theta_n) = f\left(\vec{y}\,\middle|\,\vec{\theta}\right). \tag{8.2}$$

In general, we are not be interested in the variables $\vec{y}$ themselves, but rather in a set of observed values (which we will denote with $\vec{x} = x_1, \cdots, x_n$ instead) which are present in the data in each event.

The *likelihood* L is a function which, for a given set of parameters $\vec{\theta}$, returns the probability P of a simultaneous measurement of the n random variables $x_1, \cdots, x_n$, and is defined as

$$L\left(\vec{x}\,\middle|\,\vec{\theta}\right) = \frac{dP\left(\vec{x}\,\middle|\,\vec{\theta}\right)}{dx_1 \cdots dx_n}. \tag{8.3}$$

If we have N different observations of the variables $\vec{x} = (x_1, \cdots, x_n)$, we denote this *sample* as $\{\vec{x_1}, \cdots, \vec{x_N}\}$. If these are independent observations, as would be the case of particle physics events, the likelihood of the entire sample is the product of the PDFs of

each individual observation, as

$$L\left(\vec{x_1}, \cdots, \vec{x_N} \,\middle|\, \vec{\theta}\right) = \prod_{i=1}^{N} p\left(\vec{x_i} \,\middle|\, \vec{\theta}\right). \tag{8.4}$$

In practice, it is more computationally efficient to perform sums rather than products, for which we will take the logarithm of (8.4). It is also more efficient to minimise rather than maximise, for which we also add a minus sign. With these changes, equation (8.4) becomes

$$-\log L\left(\vec{x_1}, \cdots, \vec{x_N} \,|\vec{\theta}\right) = -\sum_{i=1}^{N} \log p\left(\vec{x_i} \,|\vec{\theta}\right), \tag{8.5}$$

and we say that the *maximum likelihood* is equivalent to the *minimum log-likelihood*.

8.1.2. Hypothesis testing

Having a sample does not tell us which underlying theory model would be the best fit. What we can do is study the PDF under different models with our sample, and establish which one would be the best fit. We refer to each of these competing models as an *hypothesis*. When two hypotheses are compared, one is said to be the *null hypothesis* H_0, and the other the *alternative hypothesis* H_1. For example, if we wanted to establish the discovery of a new particle, say, a millicharged particle, we would assume its existence as an alternative hypothesis to the Standard Model, which would act as the null hypothesis. In the blip channel, the number of aligned blip-pairs or single scatters coming from NuMI which score highly in our BDTs would be enhanced in the alternative hypothesis.

In general, the distribution of a PDF $f(\vec{x}\,)$ will be different under H_0 and H_1; instead of trying to use all the variables in $\vec{x}$, we use a *test statistic* t, which is a function of the measured values,

$$t = t(\vec{x}\,). \tag{8.6}$$

The Neyman-Pearson lemma [131] indicates that the test statistic with the largest signal selection for a fixed background misidentification probability is given by the likelihood ratio

$$\lambda(\vec{x}\,) = \frac{L(\vec{x}\,|H_1)}{L(\vec{x}\,|H_0)}. \tag{8.7}$$

In practice however, this requires the knowledge of the multidimensional PDFs of the hypotheses. The boosted decision trees which we have introduced in section 6.1 can be thought of as a numerical method which aims to approximate to this ratio.

In a search for new physics, we may only be interested in one or a few parameters (say, i parameters), which we refer to as *parameters of interest* (POI) and denote with μ_i. However, optimising our model with only these parameters may not give a good enough description of the data. For this reason, more parameters, known as *nuisance parameters* (denoted with the usual $\vec{\theta}$) are added. In practice, these are usually related to our expected backgrounds, whose values are constrained with the use of independent Monte Carlo simulation samples.

In our case, we search only for one signal, thus we only need a single POI μ, which we call *signal strength*. This parameter is defined such that the expected number of events ν in the signal plus background hypothesis is $\nu = \mu s + b$, with s and b the expected signal and background based on our Monte Carlo simulation, respectively; thus the background-only hypothesis is obtained by setting $\mu = 0$, resulting in $\nu = b$. From now on, unless we explicitly make use of μ in an expression (as in $\mu s + b$), we will use "$s + b$" and "b" to refer to the signal-plus-background, or background-only hypotheses.

When the number of parameters of one hypothesis is a subset of the parameters of another one, we can use an approximation to equation (8.7) known as the *profile likelihood*. This is given by

$$\lambda(\mu_0) = \frac{L\left(\vec{x}\,\middle|\,\mu_0, \hat{\hat{\vec{\theta}}}(\mu_0)\right)}{L\left(\vec{x}\,\middle|\,\hat{\mu}, \hat{\vec{\theta}}\right)}. \tag{8.8}$$

In this test statistic, the denominator is the (unconditional) likelihood function evaluated at the parameters which maximise it (i.e. both μ and $\vec{\theta}$ are unrestricted). In the numerator, the POI is fixed at a given value $\mu = \mu_0$, while the nuisance parameters are optimised to maximise the (conditional) likelihood [129, 132]. We say that, in the numerator, the nuisance parameters have been "profiled", which is what the double-hat notation indicates [101]. Note that this test statistic is no longer a ratio of different hypotheses as in equation (8.7), and has to be applied to the measurements under each hypothesis independently.

8.1.3. Discoveries and limits

Searches for new physics can be classified as either discoveries, or limits. A discovery occurs if an experiment sees evidence of new physics, in the form of disagreement with the background-only hypothesis. If, on the other hand, the results show agreement with the background-only hypothesis, they can be presented as an exclusion limit, meaning that the evidence of no new physics in the parameter space searched *excludes* that region of parameter space.

The standard of what constitutes that a result excludes an hypothesis "beyond statistical fluctuations" is ultimately dictated by scientific consensus, and whether a new signal is accepted as either "discovered" or "excluded" is a process which will involve discussion among the scientific community, going beyond any singular metric (not to mention differences in interpretation between the Bayesian and frequentist approaches); nonetheless there are quantifiable statistical measurements which are commonly used to aid in this discussion.

The probability p that a test statistic t has a value greater or equal than the observed one t_{obs} assuming that the null hypothesis is true, is called a p-value. A very small p-value would then mean that the observed data is very unlikely to occur under the null hypothesis due to statistical fluctuations only. This concept is illustrated in figure 8.2. The distributions depend on the value of the observations $\vec{x}$ which are more or less likely

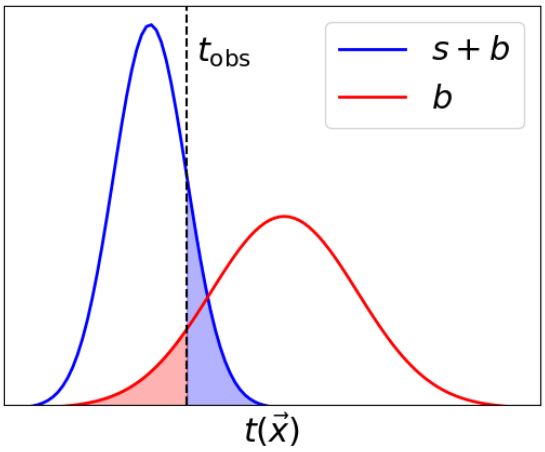


Figure 8.2.: Diagram illustrating the distribution of the signal-plus-background ($s+b$) hypothesis (blue) and the background-only (b) hypothesis (red) under a test statistic $t(\vec{x})$. The blue shaded area is the p-value of the $s+b$ hypothesis, and the red shaded area is 1 minus the p value of the background-only hypothesis.

to occur under each hypothesis. The value t_{obs} of the test statistic for the observations $\vec{x}_{\text{obs}}$ actually measured in an experiment is shown as a vertical dashed line. The shaded blue area under $s+b$ above this line is the p-value of the $s+b$ hypothesis if it were considered as the null hypothesis, as in the case of a limit search. If b was considered the null hypothesis, the shaded red area would correspond to 1 minus its p value, which is the correction term applied in the CL_s method.

In the case of a discovery search, the background-only hypothesis is taken as the null hypothesis, and very small p values of 1.35×10^{-3} and 2.87×10^{-7} are usually considered as *evidence* or *observation* of a new signal, respectively. These values are the equivalent p-values of the rightmost tail of a normal distribution either 3 or 5 standard deviations from its centre (σ); results in discovery searches are usually quoted in terms of this Z significance as $Z\sigma$ (so in these cases, 3σ or 5σ). In the case of exclusion limits, the hypotheses are inverted, and now the signal plus background hypothesis is considered the null hypothesis. The p-value requirements in this case are usually much less stringent, with $p < 0.05$ or $p < 0.10$, which correspond to 95 % confidence level (CL) or 90 % CL respectively, being the standard.

In the case of limit searches, a modified frequentist approach, also known as the CL_s method, has been adopted by the community in order to avoid cases in which an experiment with higher expected background may provide more stringent limits in the case of zero observed events. An example from Ref. [129] is very illustrative: consider two experiments A and B looking for a rare signal, expecting low backgrounds of $b = 0.01$ and $b = 0.1$ respectively. If both experiments observe zero counts, B will achieve a more stringent limit due to its higher background expectation, despite having a worse

expected performance. The correction proposed by the CL_s method is to consider the p values under both the $s+b$ and b hypotheses (denoted as p_{s+b} and p_b), and use the value of

$$\mathrm{CL}_s\left(\mu, \vec{\theta}\right) = \frac{p_{s+b}\left(\mu, \vec{\theta}\right)}{1 - p_b\left(\mu, \vec{\theta}\right)} \tag{8.9}$$

instead of simply p_{s+b} to set limits. Therefore, the denominator $1 - p_b$, which is shown graphically as the red shaded area in figure 8.2, acts as a correction term.

The test statistic we will use is a variation of the profile-likelihood shown in equation (8.8), which has become commonplace since the search for the Higgs boson. We define $\tilde{q}_\mu$ as

$$\tilde{q}_\mu = \begin{cases} -2\log\dfrac{L\left(\vec{x}\left|\mu_0, \hat{\hat{\vec{\theta}}}(\mu_0)\right.\right)}{L\left(\vec{x}\left|0, \hat{\hat{\vec{\theta}}}(0)\right.\right)} & \hat{\mu} < 0, \\ -2\log\dfrac{L\left(\vec{x}\left|\mu_0, \hat{\hat{\vec{\theta}}}(\mu_0)\right.\right)}{L\left(\vec{x}\left|\hat{\mu}, \hat{\vec{\theta}}\right.\right)} & 0 \leq \hat{\mu} \leq \mu_0, \\ 0 & \hat{\mu} > \mu_0. \end{cases} \tag{8.10}$$

Here, the first case is used to account for situations in which a negative signal would be a best fit of $\hat{\mu}$, as in the case of underfluctuations in data; it essentially forces the signal to have a physical, non-negative value by profiling the denominator to $\mu = 0$. The second case is the usual profile (negative log-) likelihood test statistic given in equation (8.8). The third case is meant to avoid the effects of upwards fluctuations in the data; this test statistic by construction follows $\tilde{q}_\mu > 0$, therefore a value of 0 will always be below the observed value and these cases will not form part of the $s+b$ p-value, preventing them from being counted against the $s+b$ hypothesis.

In general, obtaining the distributions of test statistics under different hypotheses to measure their p-values is not trivial. It is typically done using pseudo-experiments generated using Monte Carlo (often called Toy Monte Carlo). A collection of toy measurements $\vec{x}_{\mathrm{toy}}$ are generated around the expected values for $s+b$ and b, then the test statistic calculated, and the p-values obtained from the tails of the distributions. Naturally, few toys will be in the tails of these distributions, and thousands of pseudo-experiments need to be generated. Performing the parameter optimisation in the test statistic for each toy can be computationally expensive; for this reason, an asymptotic approximation to test statistics can be used, a detailed description of which can be found in Ref. [132]. Essentially, since we are only interested in the tails of the distributions, under certain conditions the distributions of the test statistic can be approximated using normal distributions, whose variance parameter σ can be obtained from a single data sample A (as opposed to thousands of toys) that has an estimated value of μ equal to the true value μ' (which in practice is set to its expected value). Such a sample A is called the Asimov data set.

8.1.4. Software implementation: `pyhf`

We will use the `pyhf` [133, 134] framework for limit setting. This is a Python implementation of `HistFactory` [135], a template for running statistical analysis which has been extensively used in particle physics searches. `pyhf` implements the variations to the test statistics and asymptotic formulas described in Refs. [129, 132], equation (8.10) in particular, and the p-values and CL_s. In addition to the asymptotic formulas it also supports the generation of Monte Carlo pseudo-experiments, and modern graph libraries such as PyTorch and TensorFlow, to take advantage of features such as GPU acceleration.

`HistFactory`, as its name implies, is based around histograms, organised in an XML file according to a determined hierarchy; for `pyhf` this has been translated into plain-text Python dictionaries. According to this hierarchy, the analysis is composed of disjoint binned distributions (*channels*), each of which comes from the sum of different physics processes (*samples*). The samples, in addition to having an identifier name and a histogram of nominal expected entries, are subject to parametrised variations which can be free ($\vec{\eta}$), or constrained ($\vec{\chi}$) parameters for systematic uncertainties, the POI being a subset of the free parameters. `HistFactory` and `pyhf` use a frequentist approach to constrain uncertainties using a set of *auxiliary measurements*, and *modifiers* for rate modifications. The modifier types differ on whether the uncertainties are constrained or unconstrained, applied to all samples or on a sample-by-sample basis (*shape* uncertainties), and in the case of binned modifiers whether the uncertainties are correlated across bins or not. The type of modifier we use for each uncertainty discussed in section 7 is shown in table 8.1.

Table 8.1.: `pyhf` modifiers used in both channels.

Uncertainty	Modifier
Flux	Normalisation Uncertainty
MC statistical	Uncorrelated shape
Detector variations	Correlated shape

8.2. Results of our two-channel millicharged particle search in MicroBooNE

We will now use `pyhf` to perform the limit search which we have outlined in the previous sections. We will show this for each of the two channels independently, going through all the steps in detail only in the first channel (Wire-Cell single scatter) starting with one mass point. Then, we will combine all mass points in the single-scatter channel, and finally apply the same strategy in the multiple-scatter channel.

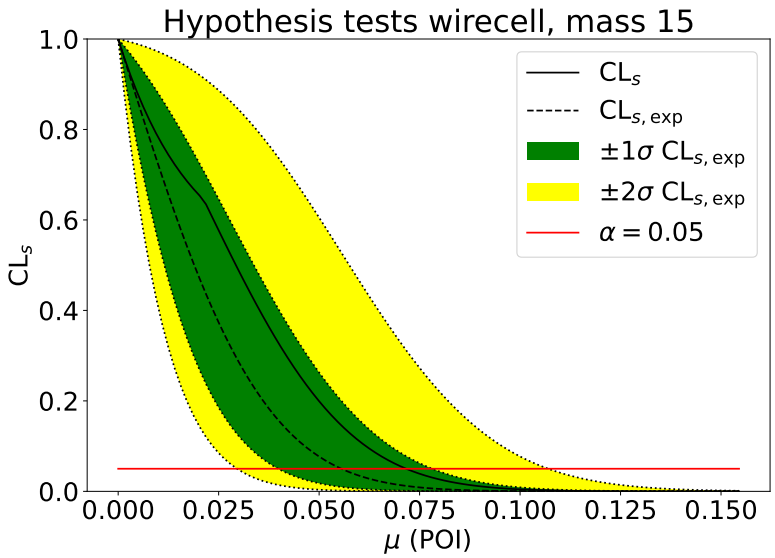


Figure 8.3.: Upper limits from a scan of fixed signal strength parameter values μ_0 from `pyhf` for the 15 MeV mass point. The dashed line indicates the expected value, with green and yellow bands indicating the $\pm1\sigma$ and $\pm2\sigma$ bands around it respectively. The solid line indicates the observed value. A red line shows the 95 % confidence level threshold from the CL$_s$ method, at which point we record expected and observed values of μ.

8.2.1. Wire-Cell single-scatter search

Single mass point upper limits

`pyhf` has a method for obtaining upper limits using hypothesis tests. It iterates over a range of values of the signal strength μ_0, computes the test statistic (8.10) for the $s+b$ and b hypotheses using an asymptotic approximation, then returns the CL$_s$ value for the expected and observed values. The expected values come from using the sum of background expectations as observations in the test statistic, while the observed values come from using the real observed data from the beam-on sample. The results are usually presented alongside the $\pm1\sigma, \pm2\sigma$ ranges for the expected values at each tested μ_0 POI point, in what is colloquially known as the "Brazil plot" due to the colours used for these bands. The application of this method for the 15 MeV mass sample is shown in figure 8.3.

Our normalised sample $\hat{S}_0$ was simulated with a fixed value of the millicharge coupling parameter $\epsilon_0 = 0.001$. As discussed in section 5.2, we can scale $\hat{S}_0$ to a sample $\hat{S}$ with an arbitrary parameter ϵ as shown in equation (5.6),

$$\epsilon = \epsilon_0 \left(\frac{\hat{S}}{\hat{S}_0} \right)^{\frac{1}{2+2n_{\mathrm{scatters}}}} .$$

We want to scale it to the sample in which a signal strength of 1 gives a 95 % confidence level of exclusion, in which case $\mu\hat{S}_0 = \hat{S}$, and the exclusion is set at the coupling parameter ϵ given by

$$\epsilon = \epsilon_0 \mu^{\frac{1}{2+2n_{\text{scatters}}}}. \tag{8.11}$$

For the Wire-Cell channel, the number of scatters is $n_{\text{scatters}} = 1$, and we use equation (8.11) to obtain the values of ϵ from the points of intersection of the μ bands along the red line in figure 8.3, which indicates 95 % confidence level. This is shown in table 8.2.

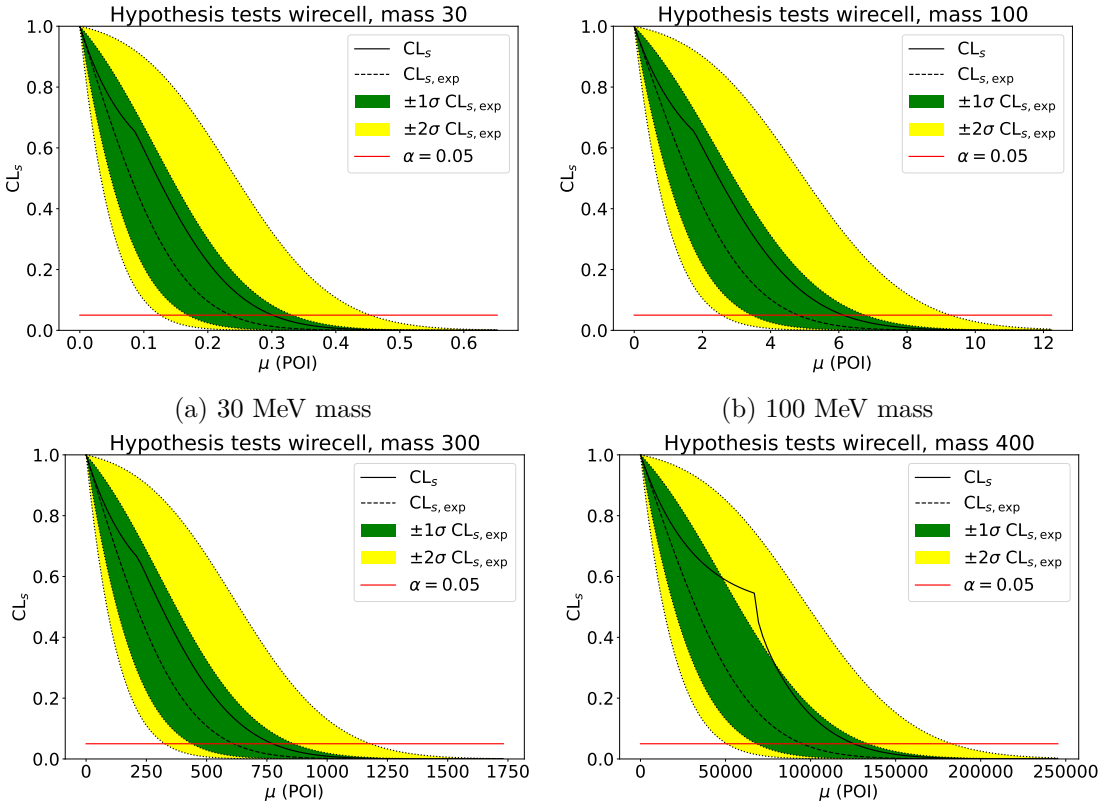
Table 8.2.: Values of the ϵ parameter obtained for the upper limits on the signal strength μ in `pyhf` for the 15 MeV signal mass point. The values of μ are those of the red line in figure 8.3.

Point	μ for $\epsilon_0 = 0.001$	Millicharge ϵ
Exp. -2σ	2.96×10^{-2}	4.15×10^{-4}
Exp. -1σ	3.98×10^{-2}	4.47×10^{-4}
Expected value	5.55×10^{-2}	4.85×10^{-4}
Exp. $+1\sigma$	7.81×10^{-2}	5.29×10^{-4}
Exp. $+2\sigma$	1.07×10^{-1}	5.72×10^{-4}
Observed value	7.17×10^{-2}	5.17×10^{-4}

Upper limits for all mass points

We repeat the above procedure for all the signal samples of different mCP mass to obtain the signal strength parameters, which are shown in table 8.3. We also show some examples of the *Brazil plots* for different mass points in figure 8.4. We then convert these values to the general millicharge parameter ϵ using equation (8.11) and show the results in table 8.4.

We can then combine the results from all masses to cover the mCP mass parameter space, as shown in figure 8.5. This is our final result. We exclude the region of parameter space for ϵ above the observed points at 95 % confidence level. Figure 8.5 shows the result as a continuous line, which is simply connecting the discrete values obtained for the 12 mass points simulated. We justify this on the basis tat, in the model we consider, the millicharge mass does not play any significant role — other than affecting the flux of millicharged particles produced (due to kinematics and phase-space suppression) and their selection efficiency, as shown in section 6.2 — both of which impact the statistical sample. That is, we do not expect any enhanced production of millicharged particles at any particular mass point, as would occur in the case of resonant production.



(a) 30 MeV mass

(b) 100 MeV mass

(c) 300 MeV mass

(d) 400 MeV mass

Figure 8.4.: Upper limits from a scan of fixed signal strength parameter values μ from `pyhf` for the different mass points. The dashed line indicates the expected value, with green and yellow bands indicating the $\pm1\sigma$ and $\pm2\sigma$ bands around it respectively. The solid line indicates the observed value. The red line shows the 95 % confidence level threshold.

Table 8.3.: Expected and observed limits for the signal strength parameter μ at 95 % confidence level at each mCP mass m_χ considered.

m_χ (MeV)	-2σ	-1σ	Exp. μ	$+1\sigma$	$+2\sigma$	Observed
15	2.96×10^{-2}	3.98×10^{-2}	5.55×10^{-2}	7.81×10^{-2}	1.07×10^{-1}	7.17×10^{-2}
20	5.07×10^{-2}	6.82×10^{-2}	9.51×10^{-2}	1.34×10^{-1}	1.83×10^{-1}	1.21×10^{-1}
30	1.26×10^{-1}	1.69×10^{-1}	2.36×10^{-1}	3.32×10^{-1}	4.54×10^{-1}	3.01×10^{-1}
50	5.43×10^{-1}	7.30×10^{-1}	1.02	1.43	1.96	1.31
80	1.42	1.91	2.67	3.76	5.15	3.50
100	2.56	3.45	4.81	6.77	9.28	6.13
150	1.14×10^{1}	1.54×10^{1}	2.14×10^{1}	3.02×10^{1}	4.13×10^{1}	2.72×10^{1}
200	5.06×10^{1}	6.80×10^{1}	9.49×10^{1}	1.33×10^{2}	1.83×10^{2}	1.20×10^{2}
250	1.32×10^{2}	1.78×10^{2}	2.48×10^{2}	3.49×10^{2}	4.78×10^{2}	3.15×10^{2}
300	3.27×10^{2}	4.39×10^{2}	6.12×10^{2}	8.62×10^{2}	1.18×10^{3}	7.74×10^{2}
350	1.24×10^{3}	1.67×10^{3}	2.33×10^{3}	3.28×10^{3}	4.50×10^{3}	2.99×10^{3}
400	5.08×10^{4}	6.83×10^{4}	9.51×10^{4}	1.34×10^{5}	1.83×10^{5}	1.25×10^{5}

Table 8.4.: Expected and observed limits for the millicharge parameter ϵ at 95 % confidence level at each mCP mass m_χ considered.

m_χ (MeV)	-2σ	-1σ	Exp. ϵ	$+1\sigma$	$+2\sigma$	Observed
15	4.15×10^{-4}	4.47×10^{-4}	4.85×10^{-4}	5.29×10^{-4}	5.72×10^{-4}	5.17×10^{-4}
20	4.74×10^{-4}	5.11×10^{-4}	5.55×10^{-4}	6.05×10^{-4}	6.54×10^{-4}	5.90×10^{-4}
30	5.95×10^{-4}	6.41×10^{-4}	6.97×10^{-4}	7.59×10^{-4}	8.21×10^{-4}	7.41×10^{-4}
50	8.58×10^{-4}	9.24×10^{-4}	1.00×10^{-3}	1.09×10^{-3}	1.18×10^{-3}	1.07×10^{-3}
80	1.09×10^{-3}	1.18×10^{-3}	1.28×10^{-3}	1.39×10^{-3}	1.51×10^{-3}	1.37×10^{-3}
100	1.27×10^{-3}	1.36×10^{-3}	1.48×10^{-3}	1.61×10^{-3}	1.75×10^{-3}	1.57×10^{-3}
150	1.84×10^{-3}	1.98×10^{-3}	2.15×10^{-3}	2.34×10^{-3}	2.53×10^{-3}	2.28×10^{-3}
200	2.67×10^{-3}	2.87×10^{-3}	3.12×10^{-3}	3.40×10^{-3}	3.68×10^{-3}	3.31×10^{-3}
250	3.39×10^{-3}	3.65×10^{-3}	3.97×10^{-3}	4.32×10^{-3}	4.68×10^{-3}	4.21×10^{-3}
300	4.25×10^{-3}	4.58×10^{-3}	4.97×10^{-3}	5.42×10^{-3}	5.86×10^{-3}	5.27×10^{-3}
350	5.94×10^{-3}	6.39×10^{-3}	6.95×10^{-3}	7.57×10^{-3}	8.19×10^{-3}	7.39×10^{-3}
400	1.50×10^{-2}	1.62×10^{-2}	1.76×10^{-2}	1.91×10^{-2}	2.07×10^{-2}	1.88×10^{-2}

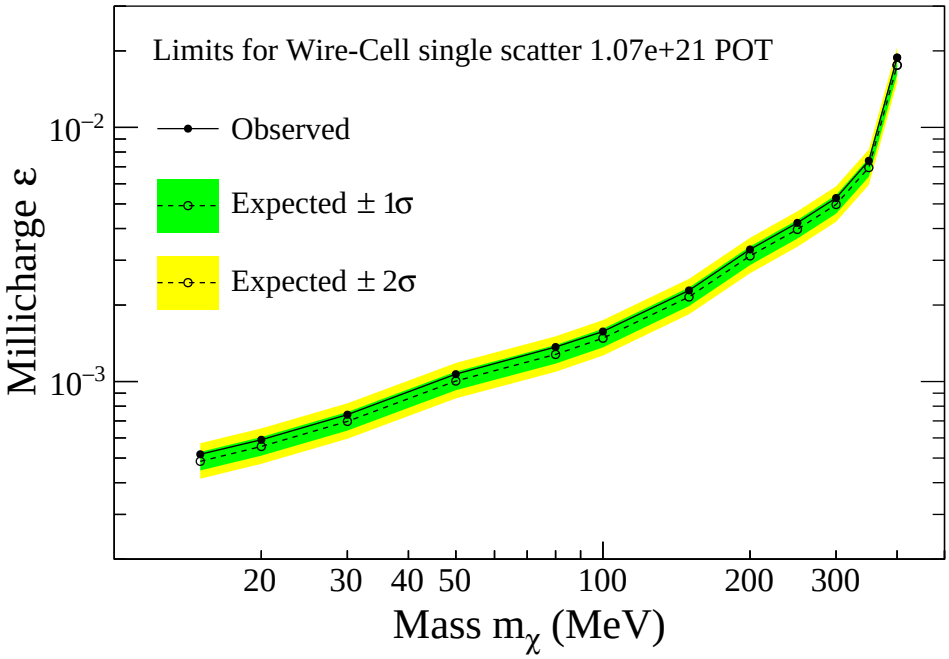


Figure 8.5.: Observed and expected limits for the Wire-Cell single-scatter channel using data from Runs 1, 2, and 3.

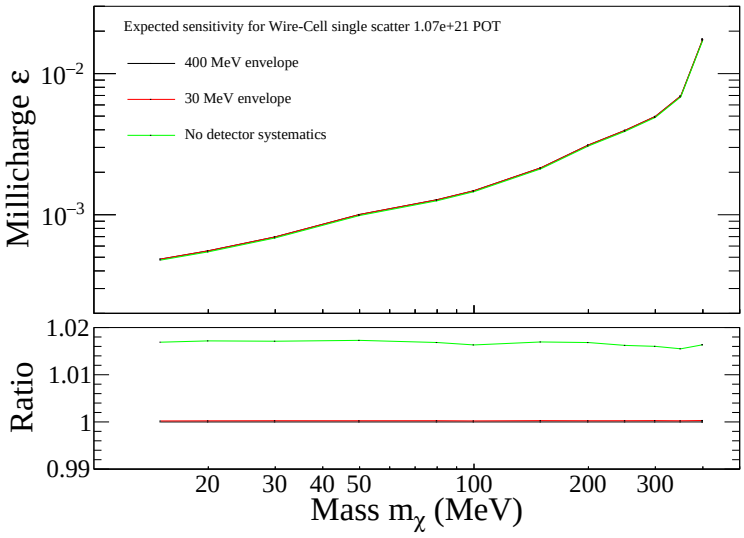


Figure 8.6.: Comparison of the expected sensitivity of the Wire-Cell two-scatter channel under different configurations of the detector systematics. The error bars have been omitted to simplify visualisation.

Comparison of expected sensitivity under different systematic uncertainties

In chapter 7 we discussed the systematic uncertainties applied to our signal and background Monte Carlo simulation samples. In particular, in section 7.3 we described the detector-related uncertainties applied to the mCP signal, which are derived from different detector variation samples. The quadrature sum of these variations defines an "envelope" that is applied to the signal samples as asymmetric systematic uncertainties in each bin.

Due to high computational cost of simulating all the mCP signals under multiple detector variations, we generated these detector variations only for two different mass points: a low mass point of 30 MeV and a high mass point of 400 MeV. We observed similar behaviour of the detector variations at both mass points, with the most noticeable differences stemming from statistical fluctuations. This is expected, as the 400 MeV signal has a lower flux and selection efficiency, which reduces the effective statistics of its detector variation samples.

Given this similarity, we use the 400 MeV envelope as a conservative approximation for all mass points. In figure 8.6, we show the expected sensitivities obtained using three approaches: the 400 MeV envelope (which is what we was used throughout the previous subsection 8.2.1), the 30 MeV envelope, and a scenario in which the detector systematics are excluded altogether (although we do retain all other systematic uncertainties described in chapter 7).

In figure 8.6, we observe that the expected limits under all three scenarios are extremely similar, and nearly indistinguishable in the top panel. The bottom panel shows the ratio of the 95 % confidence level limit on of the millicharge ϵ at each mass point,

$\epsilon_{\text{scenario}}$, relative to the case of 400 MeV uncertainties, $\epsilon_{400\,\text{MeV}}$. This ratio is defined as $\epsilon_{\text{scenario}}/\epsilon_{400\,\text{MeV}}$. Even in this case, there is virtually no difference between using the 30 MeV and 400 MeV uncertainties. We do observe a small deviation in the case of the "no detector systematics", where the sensitivity to ϵ is stronger by approximately 1.7 %.

These results behave as expected: excluding detector systematics gives a slightly higher sensitivity, as does using the 30 MeV envelope. Had this not been the case, it would have indicated a potential error in our limit-setting procedure.

8.2.2. Blip-pair two-scatter search

To obtain the sensitivity in the blip-pair two-scatter channel, we repeat the above procedure using the blip Monte-Carlo simulation samples. There are only a few differences with respect to the previous channel. First, we present these results as Monte Carlo-only expected sensitivities, meaning they are based exclusively on simulated signal and background predictions, with no use of beam-on data; therefore, we show only expected limits here. Second, when converting from signal strength μ for our simulation of $\epsilon_0 = 0.001$ towards the generalised millicharge parameter ϵ using equation (8.11), we now use $n_{\text{scatters}} = 2$.

The values obtained for the signal strength μ are shown in table 8.5, and their conversion to ϵ in table 8.6. The expected sensitivity obtained is shown in figure 8.7.

Table 8.5.: Expected value and uncertainty bands of the signal strength μ at which we obtain an exclusion at 95 % confidence level for each mCP mass m_χ considered.

m_χ (MeV)	-2σ	-1σ	Exp. μ	$+1\sigma$	$+2\sigma$
15	1.36×10^1	1.92×10^1	2.90×10^1	4.49×10^1	6.78×10^1
20	1.89×10^1	2.67×10^1	4.01×10^1	6.18×10^1	9.30×10^1
30	4.35×10^1	6.19×10^1	9.35×10^1	1.45×10^2	2.20×10^2
50	2.07×10^2	2.95×10^2	4.47×10^2	6.95×10^2	1.05×10^3
80	4.17×10^2	5.94×10^2	8.98×10^2	1.40×10^3	2.12×10^3
100	5.63×10^2	8.06×10^2	1.22×10^3	1.91×10^3	2.91×10^3
150	1.77×10^3	2.51×10^3	3.78×10^3	5.87×10^3	8.87×10^3
200	6.48×10^3	9.21×10^3	1.39×10^4	2.16×10^4	3.27×10^4
250	1.50×10^4	2.12×10^4	3.19×10^4	4.94×10^4	7.47×10^4
300	2.11×10^4	3.02×10^4	4.59×10^4	7.18×10^4	1.09×10^5
350	4.38×10^4	6.22×10^4	9.39×10^4	1.46×10^5	2.21×10^5
400	9.55×10^5	1.35×10^6	2.04×10^6	3.16×10^6	4.78×10^6

The points in figure 8.7 are obtained from table 8.6, and as in the single-scatter case the filled area is connected point-to-point between the 12 mass points simulated. As discussed, the sensitivity for this channel is not as stringent as the single-scatter channel. We present both results simultaneously in the next subsection.

Table 8.6.: Expected value and uncertainty bands of the millicharge parameter ϵ at which we obtain an at with 95 % confidence level for each mCP mass m_χ considered.

m_χ (MeV)	-2σ	-1σ	Exp. ϵ	$+1\sigma$	$+2\sigma$
15	1.54×10^{-3}	1.64×10^{-3}	1.75×10^{-3}	1.89×10^{-3}	2.02×10^{-3}
20	1.63×10^{-3}	1.73×10^{-3}	1.85×10^{-3}	1.99×10^{-3}	2.13×10^{-3}
30	1.88×10^{-3}	1.99×10^{-3}	2.13×10^{-3}	2.29×10^{-3}	2.46×10^{-3}
50	2.43×10^{-3}	2.58×10^{-3}	2.76×10^{-3}	2.98×10^{-3}	3.19×10^{-3}
80	2.73×10^{-3}	2.90×10^{-3}	3.11×10^{-3}	3.34×10^{-3}	3.58×10^{-3}
100	2.87×10^{-3}	3.05×10^{-3}	3.27×10^{-3}	3.52×10^{-3}	3.78×10^{-3}
150	3.48×10^{-3}	3.69×10^{-3}	3.95×10^{-3}	4.25×10^{-3}	4.55×10^{-3}
200	4.32×10^{-3}	4.58×10^{-3}	4.90×10^{-3}	5.28×10^{-3}	5.65×10^{-3}
250	4.96×10^{-3}	5.26×10^{-3}	5.63×10^{-3}	6.06×10^{-3}	6.49×10^{-3}
300	5.26×10^{-3}	5.58×10^{-3}	5.98×10^{-3}	6.45×10^{-3}	6.91×10^{-3}
350	5.94×10^{-3}	6.30×10^{-3}	6.74×10^{-3}	7.25×10^{-3}	7.77×10^{-3}
400	9.92×10^{-3}	1.05×10^{-2}	1.13×10^{-2}	1.21×10^{-2}	1.30×10^{-2}

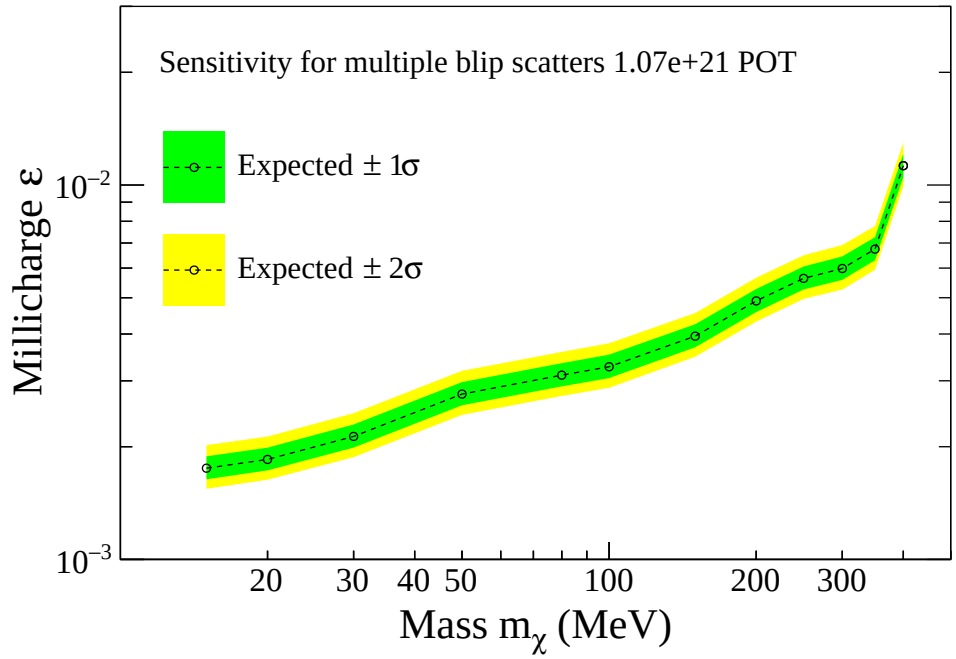


Figure 8.7.: Expected sensitivity for the multiple-scatter channel using blips, based on an exposure of 1.07×10^{21} POT. The empty circle markers indicate the expected limit, while the green and yellow bands represent the $\pm 1\sigma$ and $\pm 2\sigma$ confidence intervals, respectively.

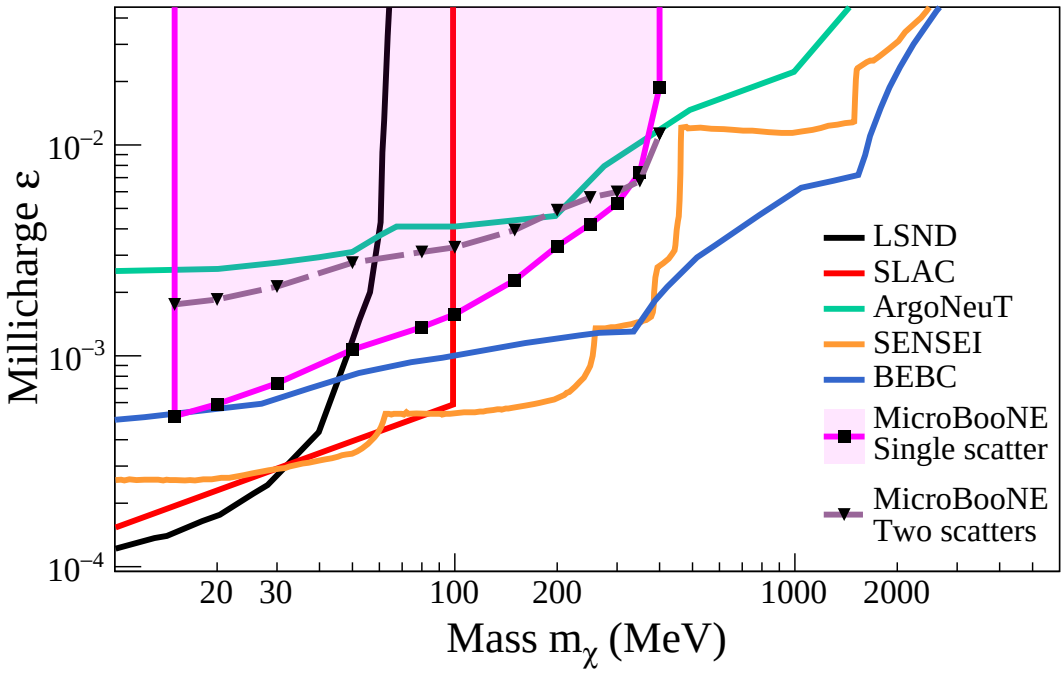


Figure 8.8.: Excluded parameter space from other experimental results compared to our own. Exclusion limits from the MicroBooNE Wire-Cell single-scatter channel are shown as a shaded magenta area, while the blip-pair scatter search is shown as a dashed line. Markers indicate the mass points used in simulation.

8.3. Comparison with other experiments

With our two main results, the limits for the single-scatter channel in figure 8.5, and the multiple-scatter sensitivity shown in figure 8.5, we now compare the results between both channels, as well as with the existing results from other experiments. Figure 8.8 shows our results alongside the current leading limits on the parameter space set by the experimental searches by SENSEI [25] and the milliQ experiment at SLAC [24], as well as reinterpretations of results from LSND [22, 23] and BEBC [27]. We also show results from the ArgoNeuT LArTPC [17].

We exclude values of the millicharge ϵ above 5.0×10^{-4} and 1.8×10^{-2} for masses of 15 and 400 MeV respectively. These limits are the most stringent of any LArTPC experiment, surpassing those of ArgoNeuT in the range 15 MeV to 350 MeV. We observe a rapid decrease in sensitivity for higher millicharge masses with the single-scatter channel, which makes our limit on a 400 MeV millicharge not as stringent. This is for three main reasons, which have been discussed in previous sections. First, the kinematics of momentum transfer in the elastic scattering of a millicharge with the electrons is limited by a maximum threshold, which highly overlaps with the minimum threshold in the single-scatter channel, which affects high masses especially. Second, the lower recon-

struction efficiency for low energy electron scatters, which is also accentuated at higher mass millicharged particles, which tend to have low energy themselves and produce less energetic scatters. Finally, we do not consider the decays of neutral mesons of a mass higher than η'. Doing so would increase the sensitivity at higher masses, and extend our sensitivity range beyond 400 MeV.

Comparing our two channels, the single-scatter search excludes a larger region of parameter space at all but the highest mass of 400 MeV considered. In general, the increased sensitivity in this channel is due to the higher probability of a millicharged particle generating a single scatter above detection threshold, as opposed to multiple scatters. Furthermore, the two-scatter result shown does not consider systematic uncertainties, which once applied will reduce the sensitivity.

Other experimental results show more stringent limits than ours. SENSEI [25] in particular, a highly sensitive pixelated dark matter detector, located underground on-axis with the NuMI beam, shows more stringent limits in all mass points. The MilliQ experiment at SLAC is the only dedicated millicharged particle experiment which has world-leading limits in a range 60 MeV to 85 MeV, however they did not have access to a beam capable of producing mCPs above 100 MeV.

9. Conclusions

In this thesis we showed the results of a search for millicharged particles in the MicroBooNE experiment using a single-electron scatter signature. We observe agreement between data and the expected backgrounds, and set limits on the production of millicharged particles between 15 and 400 MeV originating from the decay of light neutral mesons from the NuMI beam. We also show the sensitivity of MicroBooNE for a search using a different channel consisting of multiple low-energy electron scatters.

For this search we used the equivalent of 1.07×10^{21} POT of NuMI beam-on data interactions from Runs 1, 2 and 3 of MicroBooNE. We developed Monte Carlo simulations of signal from millicharged particles in the form of electron scatters within MicroBooNE, which we overlay alongside beam-off data of cosmic interactions. In the single-scatter channel, we reconstructed electron scatters as low as 20 MeV using the Wire-Cell reconstruction framework, which at this energy level only has a 1 % efficiency. For the multiple-scatter channel, we use a low-threshold of 500 keV for our signal, and use dedicated MeV-scale methods to reconstruct *blips*, which at this energy have a reconstruction efficiency of 20 %. As we look for signals with blip-pairs, the efficiency of our signal at this energy scale is reduced to 4 %.

We used Monte Carlo simulation samples of neutrino interactions from the GENIE event generator as a background, in addition to the cosmic interactions. We train different BDTs in both of our signal channels using *XGBoost*, and use them to perform a selection on the data. We find that the biggest background in both channels is due to cosmic activity, with neutrino interactions being a negligible background in the case of the multiple-scatter blip search. Systematic uncertainties are applied over the BDT score distribution of the single-scatter search by performing a full simulation under different physics parameters and detector configurations, which give as a result an uncertainty of around 25 % in the BDT-score region with most signal. We also consider a neutral meson flux uncertainty of 22 %, and a POT uncertainty of 2 %.

We use the signal region of our BDT to set limits on the millicharge parameter ϵ in a mass range 15 MeV to 400 MeV, for values of ϵ between 5.17×10^{-4} and 1.88×10^{-2}. Our results compare favourably to the ArgoNeuT liquid-argon time projection chamber (LArTPC), with more stringent limits in the range between 15 and 350 MeV mass point, making ours the most stringent limit on millicharged particles for LArTPCs in this mass range. However, other published experimental results place more stringent limits across all the mass range considered.

Our results show that LArTPCs can search for millicharged particles using a single-scatter signature. The only previous search for millicharged particles in a LArTPC is from ArgoNeuT, in which they used a two-blip signature. There are many key differences between the ArgoNeuT experiment and MicroBooNE, some of which with play in favour

of ArgoNeuT and some in favour of MicroBooNE, with the end result being that the sensitivity of the two-scatter search presented here is very close to that of ArgoNeuT. ArgoNeuT was placed 100 m underground, which provides shielding from cosmic rays thus reducing the number of background blips. They triggered on every NuMI beam spill, most of which did not produce a neutrino interaction, and thus had a vast neutrino-background free data sample, whereas we are restricted to using a "parasitic" trigger, envisioned for detecting neutrinos. Finally, the experiment was on-axis with the NuMI beam, which increased the flux of neutral mesons and their energy, and consequently of millicharge production. On the other hand, the advantages for MicroBooNE are having a bigger time-projection chamber, the use of cold electronics which reduces noise-induced blips, and around 10 times more POT recorded, not to mention the improvements in reconstruction techniques, many of which were pioneered by ArgoNeuT.

Future LArTPC experiments aiming to search for millicharged particles, such as SBND or DUNE, would benefit from considering a single-scatter channel alongside a multiple-scatter one. The biggest room for improvement in a future search is in bridging the gap between the energy reconstruction scales of 1 MeV to 20 MeV, the energy range in which neither blip reconstruction nor frameworks such as Wire-Cell have a good efficiency. This could be done in either a "bottom up" approach, starting from blip reconstruction and work towards higher energies, or a "top-down" approach, with targeted development of high-level frameworks towards the reconstruction of low-energy clusters. Collecting a sizeable unbiased trigger data sample in-time with the beam would also aid in searches for millicharged particles and many other BSM models, by reducing the neutrino backgrounds. In the case of searches for massive particles with a longer time of flight towards the detector, improvements in timing resolution using the photon detection system can also help to reduce neutrino backgrounds, as shown by MicroBooNE [69]. The nanosecond timing methods shown in Ref. [69] were not used in our search as at the time of writing they have only been implemented for the BNB beam, with work underway towards implementing them in NuMI. Finally, shielding from cosmic rays and improved understanding and modelling of the way in which they produce blips would be an advantage.

Bibliography

Bibliography

[1] J.J. Thomson. "Cathode Rays". In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 44.269 (1897), pp. 293–316. DOI: 10.1080/14786449708621070. URL: https://doi.org/10.1080/14786449708621070.

[2] J. J. Thomson. "On the masses of the ions in gases at low pressures." In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 48.295 (1899), pp. 547–567. DOI: 10.1080/14786449908621447. eprint: https://doi.org/10.1080/14786449908621447. URL: https://doi.org/10.1080/14786449908621447.

[3] R. A. Millikan. "The Isolation of an Ion, a Precision Measurement of Its Charge, and the Correction of Stokes's Law". In: *Science* 32.822 (1910), pp. 436–448. DOI: 10.1126/science.32.822.436. eprint: https://www.science.org/doi/pdf/10.1126/science.32.822.436. URL: https://www.science.org/doi/abs/10.1126/science.32.822.436.

[4] R. A. Millikan. "On the Elementary Electrical Charge and the Avogadro Constant". In: *Phys. Rev.* 2 (2 Aug. 1913), pp. 109–143. DOI: 10.1103/PhysRev.2.109. URL: https://link.aps.org/doi/10.1103/PhysRev.2.109.

[5] G. Bressi et al. "Testing the neutrality of matter by acoustic means in a spherical resonator". In: *Phys. Rev. A* 83 (5 May 2011), p. 052101. DOI: 10.1103/PhysRevA.83.052101. URL: https://link.aps.org/doi/10.1103/PhysRevA.83.052101.

[6] International Committee for Weights, on behalf of the General Conference on Weights Measures (CIPM), and Measures (CGPM), eds. *Resolution 1 of the 26th meeting of the CGPM*. 2018. DOI: 10.59161/CGPM2018RES1E. URL: https://www.bipm.org/en/committees/cg/cgpm/26-2018/resolution-1.

[7] Julian Schwinger. "Quantum Electrodynamics. I. A Covariant Formulation". In: *Phys. Rev.* 74 (10 Nov. 1948), pp. 1439–1461. DOI: 10.1103/PhysRev.74.1439. URL: https://link.aps.org/doi/10.1103/PhysRev.74.1439.

[8] K. Enqvist and D.V. Nanopoulos. "Proton decay". In: *Progress in Particle and Nuclear Physics* 16 (1986), pp. 1–33. ISSN: 0146-6410. DOI: https://doi.org/10.1016/0146-6410(86)90001-3. URL: https://www.sciencedirect.com/science/article/pii/0146641086900013.

[9] Paul Adrien Maurice Dirac. "Quantised singularities in the electromagnetic field". In: *Proc. Roy. Soc. Lond. A* 133.821 (1931), pp. 60–72. DOI: 10.1098/rspa.1931.0130.

[10] Acharya et al. "MoEDAL Search in the CMS Beam Pipe for Magnetic Monopoles Produced via the Schwinger Effect". In: *Phys. Rev. Lett.* 133 (7 Aug. 2024), p. 071803. DOI: 10.1103/PhysRevLett.133.071803. URL: https://link.aps.org/doi/10.1103/PhysRevLett.133.071803.

[11] Bob Holdom. "Two U(1)'s and ϵ charge shifts". In: *Physics Letters B* 166.2 (1986), pp. 196–198. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(86)91377-8. URL: https://www.sciencedirect.com/science/article/pii/0370269386913778.

[12] Judd D. Bowman et al. "An absorption profile centred at 78 megahertz in the sky-averaged spectrum". In: *Nature* 555.7694 (Mar. 2018), pp. 67–70. ISSN: 1476-4687. DOI: 10.1038/nature25792. URL: https://doi.org/10.1038/nature25792.

[13] Julian B. Muñoz and Abraham Loeb. "A small amount of mini-charged dark matter could cool the baryons in the early Universe". In: *Nature* 557.7707 (May 2018), pp. 684–686. ISSN: 1476-4687. DOI: 10.1038/s41586-018-0151-x. URL: https://doi.org/10.1038/s41586-018-0151-x.

[14] Kevin J. Kelly and Yu-Dai Tsai. "Proton fixed-target scintillation experiment to search for millicharged dark matter". In: *Phys. Rev. D* 100 (1 July 2019), p. 015043. DOI: 10.1103/PhysRevD.100.015043. URL: https://link.aps.org/doi/10.1103/PhysRevD.100.015043.

[15] Gabriel Magill et al. "Millicharged Particles in Neutrino Experiments". In: *Phys. Rev. Lett.* 122 (7 Feb. 2019), p. 071801. DOI: 10.1103/PhysRevLett.122.071801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.122.071801.

[16] R. Acciarri et al. "Demonstration of MeV-scale physics in liquid argon time projection chambers using ArgoNeuT". In: *Phys. Rev. D* 99 (1 Jan. 2019), p. 012002. DOI: 10.1103/PhysRevD.99.012002. URL: https://link.aps.org/doi/10.1103/PhysRevD.99.012002.

[17] R. Acciarri et al. "Improved Limits on Millicharged Particles Using the ArgoNeuT Experiment at Fermilab". In: *Phys. Rev. Lett.* 124 (13 Apr. 2020), p. 131801. DOI: 10.1103/PhysRevLett.124.131801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.124.131801.

[18] Roni Harnik, Zhen Liu, and Ornella Palamara. "Millicharged particles in liquid argon neutrino experiments". In: *Journal of High Energy Physics* 2019.7 (July 2019), p. 170. ISSN: 1029-8479. DOI: 10.1007/JHEP07(2019)170. URL: https://doi.org/10.1007/JHEP07(2019)170.

[19] P. Abratenko et al. "Demonstration of new MeV-scale capabilities in large neutrino LArTPCs using ambient radiogenic and cosmogenic activity in MicroBooNE". In: *Phys. Rev. D* 111 (3 Feb. 2025), p. 032005. DOI: 10.1103/PhysRevD.111.032005. URL: https://link.aps.org/doi/10.1103/PhysRevD.111.032005.

[20] Dmitry Gorbunov et al. "Double-hit signature of millicharged particles in 3D segmented neutrino detector". In: *Physics Letters B* 822 (2021), p. 136641. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2021.136641. URL: https://www.sciencedirect.com/science/article/pii/S0370269321005815.

[21] Claude Leroy and Pier-Giorgio Rancoita. *Principles of Radiation Interaction in Matter and Detection*. WORLD SCIENTIFIC, 2004. DOI: 10.1142/5578. eprint: https://www.worldscientific.com/doi/pdf/10.1142/5578. URL: https://www.worldscientific.com/doi/abs/10.1142/5578.

[22] Gabriel Magill et al. "Millicharged Particles in Neutrino Experiments". In: *Phys. Rev. Lett.* 122 (7 Feb. 2019), p. 071801. DOI: 10.1103/PhysRevLett.122.071801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.122.071801.

[23] L. B. Auerbach et al. "Measurement of electron-neutrino electron elastic scattering". In: *Phys. Rev. D* 63 (11 May 2001), p. 112001. DOI: 10.1103/PhysRevD.63.112001. URL: https://link.aps.org/doi/10.1103/PhysRevD.63.112001.

[24] A. A. Prinz et al. "Search for Millicharged Particles at SLAC". In: *Phys. Rev. Lett.* 81 (6 Aug. 1998), pp. 1175–1178. DOI: 10.1103/PhysRevLett.81.1175. URL: https://link.aps.org/doi/10.1103/PhysRevLett.81.1175.

[25] Liron Barak et al. "Search by the SENSEI Experiment for Millicharged Particles Produced in the NuMI Beam". In: *Phys. Rev. Lett.* 133 (7 Aug. 2024), p. 071801. DOI: 10.1103/PhysRevLett.133.071801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.133.071801.

[26] SENSEI Collaboration et al. *SENSEI: First Direct-Detection Results on sub-GeV Dark Matter from SENSEI at SNOLAB*. 2023. arXiv: 2312.13342 [astro-ph.CO]. URL: https://arxiv.org/abs/2312.13342.

[27] Giacomo Marocco and Subir Sarkar. "Blast from the past: Constraints on the dark sector from the BEBC WA66 beam dump experiment". In: *SciPost Phys.* 10 (2021), p. 043. DOI: 10.21468/SciPostPhys.10.2.043. URL: https://scipost.org/10.21468/SciPostPhys.10.2.043.

[28] A. Ball et al. "Search for millicharged particles in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Phys. Rev. D* 102 (3 Aug. 2020), p. 032002. DOI: 10.1103/PhysRevD.102.032002. URL: https://link.aps.org/doi/10.1103/PhysRevD.102.032002.

[29] Saeid Foroughi-Abari, Felix Kling, and Yu-Dai Tsai. "Looking forward to millicharged dark sectors at the LHC". In: *Physical Review D* 104.3 (Aug. 2021). ISSN: 2470-0029. DOI: 10.1103/physrevd.104.035014. URL: http://dx.doi.org/10.1103/PhysRevD.104.035014.

[30] Carlo Rubbia. *The liquid-argon time projection chamber: a new concept for neutrino detectors*. Tech. rep. Geneva: CERN, 1977. URL: https://cds.cern.ch/record/117852.

[31] P. Benetti et al. "A three-ton liquid argon time projection chamber". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 332.3 (1993), pp. 395–412. ISSN: 0168-9002. DOI: https://doi.org/10.1016/0168-9002(93)90297-U. URL: https://www.sciencedirect.com/science/article/pii/016890029390297U.

[32] R. Acciarri et al. "Design and construction of the MicroBooNE detector". In: *Journal of Instrumentation* 12.02 (Feb. 2017), P02017. DOI: 10.1088/1748-0221/12/02/P02017. URL: https://dx.doi.org/10.1088/1748-0221/12/02/P02017.

[33] J. Thomas and D. A. Imel. "Recombination of electron-ion pairs in liquid argon and liquid xenon". In: *Phys. Rev. A* 36 (2 July 1987), pp. 614–616. DOI: 10.1103/PhysRevA.36.614. URL: https://link.aps.org/doi/10.1103/PhysRevA.36.614.

[34] E. Aprile, K.L. Giboni, and C. Rubbia. "A study of ionization electrons drifting large distances in liquid and solid argon". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 241.1 (1985), pp. 62–71. ISSN: 0168-9002. DOI: https://doi.org/10.1016/0168-9002(85)90516-9. URL: https://www.sciencedirect.com/science/article/pii/0168900285905169.

[35] B. Abi et al. "First results on ProtoDUNE-SP liquid argon time projection chamber performance from a beam test at the CERN Neutrino Platform". In: *Journal of Instrumentation* 15.12 (Dec. 2020), P12004. DOI: 10.1088/1748-0221/15/12/P12004. URL: https://dx.doi.org/10.1088/1748-0221/15/12/P12004.

[36] P. Abratenko et al. "Measurement of the flux-averaged inclusive charged-current electron neutrino and antineutrino cross section on argon using the NuMI beam and the MicroBooNE detector". In: *Phys. Rev. D* 104 (5 Sept. 2021), p. 052002. DOI: 10.1103/PhysRevD.104.052002. URL: https://link.aps.org/doi/10.1103/PhysRevD.104.052002.

[37] D. Caratelli et al. "Low-Energy Physics in Neutrino LArTPCs". In: (2022). arXiv: 2203.00740 [physics.ins-det].

[38] A. A. Aguilar-Arevalo et al. "Significant Excess of Electronlike Events in the MiniBooNE Short-Baseline Neutrino Experiment". In: *Phys. Rev. Lett.* 121 (22 Nov. 2018), p. 221801. DOI: 10.1103/PhysRevLett.121.221801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.121.221801.

[39] A. Aguilar et al. "Evidence for neutrino oscillations from the observation of $\overline{\nu}_e$ appearance in a $\overline{\nu}_\mu$ beam". In: *Phys. Rev. D* 64 (11 Nov. 2001), p. 112007. DOI: 10.1103/PhysRevD.64.112007. URL: https://link.aps.org/doi/10.1103/PhysRevD.64.112007.

[40] P. Abratenko et al. "Search for an Excess of Electron Neutrino Interactions in MicroBooNE Using Multiple Final-State Topologies". In: *Phys. Rev. Lett.* 128 (24 June 2022), p. 241801. DOI: 10.1103/PhysRevLett.128.241801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.128.241801.

[41] P. Abratenko et al. "Search for an anomalous excess of charged-current ν_e interactions without pions in the final state with the MicroBooNE experiment". In: *Phys. Rev. D* 105 (11 June 2022), p. 112004. DOI: 10.1103/PhysRevD.105.112004. URL: https://link.aps.org/doi/10.1103/PhysRevD.105.112004.

[42] P. Abratenko et al. "Search for an anomalous excess of charged-current quasielastic ν_e interactions with the MicroBooNE experiment using Deep-Learning-based reconstruction". In: *Phys. Rev. D* 105 (11 June 2022), p. 112003. DOI: 10.1103/PhysRevD.105.112003. URL: https://link.aps.org/doi/10.1103/PhysRevD.105.112003.

[43] P. Abratenko et al. "Search for Neutrino-Induced Neutral-Current Radiative Decay in MicroBooNE and a First Test of the MiniBooNE Low Energy Excess under a Single-Photon Hypothesis". In: *Phys. Rev. Lett.* 128 (11 Mar. 2022), p. 111801. DOI: 10.1103/PhysRevLett.128.111801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.128.111801.

[44] P. Abratenko et al. "Search for an anomalous excess of inclusive charged-current ν_e interactions in the MicroBooNE experiment using Wire-Cell reconstruction". In: *Phys. Rev. D* 105 (11 June 2022), p. 112005. DOI: 10.1103/PhysRevD.105.112005. URL: https://link.aps.org/doi/10.1103/PhysRevLett.128.111801.

[45] MicroBooNE collaboration et al. *Search for an Anomalous Production of Charged-Current ν_e Interactions Without Visible Pions Across Multiple Kinematic Observables*. 2024. arXiv: 2412.14407 [hep-ex]. URL: https://arxiv.org/abs/2412.14407.

[46] P. Abratenko et al. "First Measurement of Inclusive Muon Neutrino Charged Current Differential Cross Sections on Argon at $E_\nu \sim 0.8$ GeV with the MicroBooNE Detector". In: *Phys. Rev. Lett.* 123 (13 Sept. 2019), p. 131801. DOI: 10.1103/PhysRevLett.123.131801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.123.131801.

[47] P. Abratenko et al. "First Measurement of Differential Charged Current Quasielasticlike ν_μ-Argon Scattering Cross Sections with the MicroBooNE Detector". In: *Phys. Rev. Lett.* 125 (20 Nov. 2020), p. 201803. DOI: 10.1103/PhysRevLett.125.201803. URL: https://link.aps.org/doi/10.1103/PhysRevLett.125.201803.

[48] P. Abratenko et al. "Measurement of differential cross sections for ν_μ-Ar charged-current interactions with protons and no pions in the final state with the MicroBooNE detector". In: *Phys. Rev. D* 102 (11 Dec. 2020), p. 112013. DOI: 10.1103/PhysRevD.102.112013. URL: https://link.aps.org/doi/10.1103/PhysRevD.102.112013.

[49] P. Abratenko et al. "Measurement of the flux-averaged inclusive charged-current electron neutrino and antineutrino cross section on argon using the NuMI beam and the MicroBooNE detector". In: *Phys. Rev. D* 104 (5 Sept. 2021), p. 052002. DOI: 10.1103/PhysRevD.104.052002. URL: https://link.aps.org/doi/10.1103/PhysRevD.104.052002.

[50] P. Abratenko et al. "First measurement of inclusive electron-neutrino and antineutrino charged current differential cross sections in charged lepton energy on argon in MicroBooNE". In: *Phys. Rev. D* 105 (5 Mar. 2022), p. L051102. DOI: 10.1103/PhysRevD.105.L051102. URL: https://link.aps.org/doi/10.1103/PhysRevD.105.L051102.

[51] P. Abratenko et al. "First Measurement of Energy-Dependent Inclusive Muon Neutrino Charged-Current Cross Sections on Argon with the MicroBooNE Detector". In: *Phys. Rev. Lett.* 128 (15 Apr. 2022), p. 151801. DOI: 10.1103/PhysRevLett.128.151801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.128.151801.

[52] P. Abratenko et al. "Differential cross section measurement of charged current ν_e interactions without final-state pions in MicroBooNE". In: *Phys. Rev. D* 106 (5 Sept. 2022), p. L051102. DOI: 10.1103/PhysRevD.106.L051102. URL: https://link.aps.org/doi/10.1103/PhysRevD.106.L051102.

[53] MicroBooNE collaboration et al. *First Measurement of Differential Cross Sections for Muon Neutrino Charged Current Interactions on Argon with a Two-proton Final State in the MicroBooNE Detector*. 2023. arXiv: 2211.03734 [hep-ex]. URL: https://arxiv.org/abs/2211.03734.

[54] P. Abratenko et al. "Multidifferential cross section measurements of ν_μ-argon quasielasticlike reactions with the MicroBooNE detector". In: *Phys. Rev. D* 108 (5 Sept. 2023), p. 053002. DOI: 10.1103/PhysRevD.108.053002. URL: https://link.aps.org/doi/10.1103/PhysRevD.108.053002.

[55] MicroBooNE Collaboration et al. *Measurement of three-dimensional inclusive muon-neutrino charged-current cross sections on argon with the MicroBooNE detector*. 2024. arXiv: 2307.06413 [hep-ex]. URL: https://arxiv.org/abs/2307.06413.

[56] P. Abratenko et al. "Inclusive cross section measurements in final states with and without protons for charged-current ν_μ-Ar scattering in MicroBooNE". In: *Phys. Rev. D* 110 (1 July 2024), p. 013006. DOI: 10.1103/PhysRevD.110.013006. URL: https://link.aps.org/doi/10.1103/PhysRevD.110.013006.

[57] P. Abratenko et al. "First Simultaneous Measurement of Differential Muon-Neutrino Charged-Current Cross Sections on Argon for Final States with and without Protons Using MicroBooNE Data". In: *Phys. Rev. Lett.* 133 (4 July 2024), p. 041801. DOI: 10.1103/PhysRevLett.133.041801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.133.041801.

[58] MicroBooNE collaboration et al. *Measurement of double-differential cross sections for mesonless charged-current muon neutrino interactions on argon with final-state protons using the MicroBooNE detector*. 2024. arXiv: 2403.19574 [hep-ex]. URL: https://arxiv.org/abs/2403.19574.

[59] MicroBooNE collaboration et al. *Measurement of the differential cross section for neutral pion production in charged-current muon neutrino interactions on argon with the MicroBooNE detector*. 2024. arXiv: 2404.09949 [hep-ex]. URL: https://arxiv.org/abs/2404.09949.

[60] MicroBooNE collaboration et al. *First double-differential cross section measurement of neutral-current π^0 production in neutrino-argon scattering in the MicroBooNE detector*. 2024. arXiv: 2404.10948 [hep-ex]. URL: https://arxiv.org/abs/2404.10948.

[61] P. Abratenko et al. "Search for heavy neutral leptons decaying into muon-pion pairs in the MicroBooNE detector". In: *Phys. Rev. D* 101 (5 Mar. 2020), p. 052001. DOI: 10.1103/PhysRevD.101.052001. URL: https://link.aps.org/doi/10.1103/PhysRevD.101.052001.

[62] P. Abratenko et al. "Search for long-lived heavy neutral leptons and Higgs portal scalars decaying in the MicroBooNE detector". In: *Phys. Rev. D* 106 (9 Nov. 2022), p. 092006. DOI: 10.1103/PhysRevD.106.092006. URL: https://link.aps.org/doi/10.1103/PhysRevD.106.092006.

[63] P. Abratenko et al. "Search for a Higgs Portal Scalar Decaying to Electron-Positron Pairs in the MicroBooNE Detector". In: *Phys. Rev. Lett.* 127 (15 Oct. 2021), p. 151803. DOI: 10.1103/PhysRevLett.127.151803. URL: https://link.aps.org/doi/10.1103/PhysRevLett.127.151803.

[64] P. Abratenko et al. "Search for Heavy Neutral Leptons in Electron-Positron and Neutral-Pion Final States with the MicroBooNE Detector". In: *Phys. Rev. Lett.* 132 (4 Jan. 2024), p. 041801. DOI: 10.1103/PhysRevLett.132.041801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.132.041801.

[65] P. Abratenko et al. "First Search for Dark-Trident Processes Using the MicroBooNE Detector". In: *Phys. Rev. Lett.* 132 (24 June 2024), p. 241801. DOI: 10.1103/PhysRevLett.132.241801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.132.241801.

[66] MicroBooNE collaboration et al. *First application of a liquid argon time projection chamber for the search for intranuclear neutron-antineutron transitions and annihilation in ^{40}Ar using the MicroBooNE detector*. 2024. arXiv: 2308.03924 [hep-ex]. URL: https://arxiv.org/abs/2308.03924.

[67] R. Acciarri et al. "Noise Characterization and Filtering in the MicroBooNE Liquid Argon TPC". In: *Journal of Instrumentation* 12.08 (Aug. 2017), P08003. DOI: 10.1088/1748-0221/12/08/P08003. URL: https://dx.doi.org/10.1088/1748-0221/12/08/P08003.

[68] P. Abratenko et al. "Cosmic Ray Background Rejection with Wire-Cell LArTPC Event Reconstruction in the MicroBooNE Detector". In: *Phys. Rev. Appl.* 15 (6 June 2021), p. 064071. DOI: 10.1103/PhysRevApplied.15.064071. URL: https://link.aps.org/doi/10.1103/PhysRevApplied.15.064071.

[69] P. Abratenko et al. "First demonstration of $\mathcal{O}(1$ ns) timing resolution in the MicroBooNE liquid argon time projection chamber". In: *Phys. Rev. D* 108 (5 Sept. 2023), p. 052010. DOI: 10.1103/PhysRevD.108.052010. URL: https://link.aps.org/doi/10.1103/PhysRevD.108.052010.

[70] C. Adams et al. "Design and construction of the MicroBooNE Cosmic Ray Tagger system". In: *Journal of Instrumentation* 14.04 (Apr. 2019), P04004. DOI: 10.1088/1748-0221/14/04/P04004. URL: https://dx.doi.org/10.1088/1748-0221/14/04/P04004.

[71] Vladimir Shiltsev. "Fermilab proton accelerator complex status and improvement plans". In: *Modern Physics Letters A* 32.16 (May 2017), p. 1730012. ISSN: 1793-6632. DOI: 10.1142/s0217732317300129. URL: http://dx.doi.org/10.1142/S0217732317300129.

[72] S. Holmes et al. "Accelerator physics at the Tevatron Collider: Introduction". In: *Accelerator Physics at the Tevatron Collider*. Ed. by Valery Lebedev and Vladimir Shiltsev. 2014, pp. 1–28. DOI: 10.1007/978-1-4939-0885-1_1.

[73] P. Adamson et al. "The NuMI neutrino beam". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 806 (2016), pp. 279–306. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2015.08.063. URL: https://www.sciencedirect.com/science/article/pii/S016890021501027X.

[74] Krishan Vinay Jitesh Mistry. "First Measurement of the Flux-Averaged Differential Charged-Current Electron-Neutrino and Antineutrino Cross Section on Argon with the MicroBooNE Detector". PhD thesis. Manchester U., Manchester U., 2021. DOI: 10.1007/978-3-031-19572-3.

[75] D. G. Michael et al. "Observation of Muon Neutrino Disappearance with the MINOS Detectors in the NuMI Neutrino Beam". In: *Phys. Rev. Lett.* 97 (19 Nov. 2006), p. 191801. DOI: 10.1103/PhysRevLett.97.191801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.97.191801.

[76] L. Aliaga et al. "Design, calibration, and performance of the MINERvA detector". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 743 (2014), pp. 130–159. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2013.12.053. URL: https://www.sciencedirect.com/science/article/pii/S0168900214000035.

[77] C. Anderson et al. "First Measurements of Inclusive Muon Neutrino Charged Current Differential Cross Sections on Argon". In: *Phys. Rev. Lett.* 108 (16 Apr. 2012), p. 161802. DOI: 10.1103/PhysRevLett.108.161802. URL: https://link.aps.org/doi/10.1103/PhysRevLett.108.161802.

[78] Pedro A.N. Machado, Ornella Palamara, and David W. Schmitz. "The Short-Baseline Neutrino Program at Fermilab". In: *Annual Review of Nuclear and Particle Science* 69.1 (2019), pp. 363–387. DOI: 10.1146/annurev-nucl-101917-020949. eprint: https://doi.org/10.1146/annurev-nucl-101917-020949. URL: https://doi.org/10.1146/annurev-nucl-101917-020949.

[79] R. Acciarri et al. "Construction of precision wire readout planes for the Short-Baseline Near Detector (SBND)". In: *Journal of Instrumentation* 15.06 (June 2020), P06033. DOI: 10.1088/1748-0221/15/06/P06033. URL: https://dx.doi.org/10.1088/1748-0221/15/06/P06033.

[80] P. Abratenko et al. "Scintillation light in SBND: simulation, reconstruction, and expected performance of the photon detection system". In: *The European Physical Journal C* 84.10 (Oct. 2024), p. 1046. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-024-13306-3. URL: https://doi.org/10.1140/epjc/s10052-024-13306-3.

[81] A.A. Machado et al. "The X-ARAPUCA: an improvement of the ARAPUCA device". In: *Journal of Instrumentation* 13.04 (Apr. 2018), p. C04026. DOI: 10.1088/1748-0221/13/04/C04026. URL: https://dx.doi.org/10.1088/1748-0221/13/04/C04026.

[82] E. Buckley et al. "A study of ionization electrons drifting over large distances in liquid argon". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 275.2 (1989), pp. 364–372. ISSN: 0168-9002. DOI: https://doi.org/10.1016/0168-9002(89)90710-9. URL: https://www.sciencedirect.com/science/article/pii/0168900289907109.

[83] R Acciarri et al. "Oxygen contamination in liquid Argon: combined effects on ionization electron charge and scintillation light". In: *Journal of Instrumentation* 5.05 (May 2010), P05003. DOI: 10.1088/1748-0221/5/05/P05003. URL: https://dx.doi.org/10.1088/1748-0221/5/05/P05003.

[84] The MicroBooNE Collaboration. *Measurement of the Electronegative Contaminants and Drift Electron Lifetime in the MicroBooNE Experiment (MicroBooNE public note)*. 2016. URL: https://microboone.fnal.gov/wp-content/uploads/MICROBOONE-NOTE-1003-PUB.pdf.

[85] P. Abratenko et al. "Measurement of space charge effects in the MicroBooNE LArTPC using cosmic muons". In: *Journal of Instrumentation* 15.12 (Dec. 2020), P12037. DOI: 10.1088/1748-0221/15/12/P12037. URL: https://dx.doi.org/10.1088/1748-0221/15/12/P12037.

[86] R Acciarri et al. "A study of electron recombination using highly ionizing particles in the ArgoNeuT Liquid Argon TPC". In: *Journal of Instrumentation* 8.08 (Aug. 2013), P08005. DOI: 10.1088/1748-0221/8/08/P08005. URL: https://dx.doi.org/10.1088/1748-0221/8/08/P08005.

[87] C. Adams et al. "Calibration of the charge and energy loss per unit length of the MicroBooNE liquid argon time projection chamber using muons and protons". In: *Journal of Instrumentation* 15.03 (Mar. 2020), P03022. DOI: 10.1088/1748-0221/15/03/P03022. URL: https://dx.doi.org/10.1088/1748-0221/15/03/P03022.

[88] The MicroBooNE collaboration et al. "Measurement of the longitudinal diffusion of ionization electrons in the MicroBooNE detector". In: *Journal of Instrumentation* 16.09 (Sept. 2021), P09025. DOI: 10.1088/1748-0221/16/09/P09025. URL: https://dx.doi.org/10.1088/1748-0221/16/09/P09025.

[89] Lev Davidovich Landau. "On the energy loss of fast particles by ionization". In: *J. Phys.* 8.4 (1944), pp. 201–205. URL: https://cds.cern.ch/record/216256.

[90] The MicroBooNE Collaboration. *Detector calibration using through going and stopping muons in the MicroBooNE LArTPC (MicroBooNE public note)*. 2018. URL: https://microboone.fnal.gov/wp-content/uploads/MICROBOONE-NOTE-1048-PUB.pdf.

[91] D. Heck et al. *CORSIKA: a Monte Carlo code to simulate extensive air showers.* Feb. 1998.

[92] R. Acciarri et al. *A Proposal for a Three Detector Short-Baseline Neutrino Oscillation Program in the Fermilab Booster Neutrino Beam.* 2015. arXiv: 1503.01520 [physics.ins-det]. URL: https://arxiv.org/abs/1503.01520.

[93] Rene Brun and Fons Rademakers. "ROOT — An object oriented data analysis framework". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 389.1 (1997). New Computing Techniques in Physics Research V, pp. 81–86. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(97)00048-X. URL: https://www.sciencedirect.com/science/article/pii/S016890029700048X.

[94] Eric D. Church. *LArSoft: A Software Package for Liquid Argon Time Projection Drift Chambers.* 2014. arXiv: 1311.6774 [physics.ins-det]. URL: https://arxiv.org/abs/1311.6774.

[95] E. L. Snider and G. Petrillo. "LArSoft: Toolkit for Simulation, Reconstruction and Analysis of Liquid Argon TPC Neutrino Detectors". In: *J. Phys. Conf. Ser.* 898.4 (2017). Ed. by Richard Mount and Craig Tull, p. 042057. DOI: 10.1088/1742-6596/898/4/042057. URL: https://larsoft.org/.

[96] X. Qian et al. "Three-dimensional imaging for large LArTPCs". In: *Journal of Instrumentation* 13.05 (May 2018), P05032. DOI: 10.1088/1748-0221/13/05/P05032. URL: https://dx.doi.org/10.1088/1748-0221/13/05/P05032.

[97] The MicroBooNE collaboration et al. "Neutrino event selection in the MicroBooNE liquid argon time projection chamber using Wire-Cell 3D imaging, clustering, and charge-light matching". In: *Journal of Instrumentation* 16.06 (June 2021), P06043. DOI: 10.1088/1748-0221/16/06/P06043. URL: https://dx.doi.org/10.1088/1748-0221/16/06/P06043.

[98] L. Aliaga et al. "Neutrino flux predictions for the NuMI beam". In: *Phys. Rev. D* 94 (9 Nov. 2016), p. 092005. DOI: 10.1103/PhysRevD.94.092005. URL: https://link.aps.org/doi/10.1103/PhysRevD.94.092005.

[99] S. Agostinelli et al. "GEANT4: A simulation toolkit". In: *Nucl. Instrum. Meth.* A506 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8.

[100] C. Andreopoulos et al. "The GENIE Neutrino Monte Carlo Generator". In: *Nucl. Instrum. Meth. A* 614 (2010), pp. 87–104. DOI: 10.1016/j.nima.2009.12.009. arXiv: 0905.2517 [hep-ph].

[101] S. Navas et al. "Review of particle physics". In: *Phys. Rev. D* 110.3 (2024), p. 030001. DOI: 10.1103/PhysRevD.110.030001.

[102] C. Adams et al. "Ionization electron signal processing in single phase LArTPCs. Part I. Algorithm Description and quantitative evaluation with MicroBooNE simulation". In: *Journal of Instrumentation* 13.07 (July 2018), P07006. DOI: 10.1088/1748-0221/13/07/P07006. URL: https://dx.doi.org/10.1088/1748-0221/13/07/P07006.

[103] B. Baller. "Liquid argon TPC signal formation, signal processing and reconstruction techniques". In: *Journal of Instrumentation* 12.07 (July 2017), P07010. DOI: 10.1088/1748-0221/12/07/P07010. URL: https://dx.doi.org/10.1088/1748-0221/12/07/P07010.

[104] Emmanuel J Candes, Justin K Romberg, and Terence Tao. "Stable signal recovery from incomplete and inaccurate measurements". In: *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences* 59.8 (2006), pp. 1207–1223.

[105] The MicroBooNE collaboration. "Study of Reconstructed 39Ar Beta Decays at the MicroBooNE Detector". In: (June 2018). DOI: 10.2172/1573057. URL: https://www.osti.gov/biblio/1573057.

[106] The MicroBooNE Collaboration. *MeV-scale Physics in MicroBooNE*. 2020. URL: https://microboone.fnal.gov/wp-content/uploads/MICROBOONE-NOTE-1076-PUB.pdf.

[107] Avinay Bhat. "MeV Scale Physics in MicroBooNE". PhD thesis. Syracuse University, 2021. URL: https://surface.syr.edu/etd/1429.

[108] P. Abratenko, J. Anthony, and L. Arellano. "Observation of radon mitigation in MicroBooNE by a liquid argon filtration system". In: *Journal of Instrumentation* 17.11 (Nov. 2022), P11022. DOI: 10.1088/1748-0221/17/11/P11022. URL: https://dx.doi.org/10.1088/1748-0221/17/11/P11022.

[109] MicroBooNE collaboration et al. *Measurement of ambient radon daughter decay rates and energy spectra in liquid argon using the MicroBooNE detector*. 2023. arXiv: 2307.03102 [hep-ex].

[110] W. Castiglioni et al. "Benefits of MeV-scale reconstruction capabilities in large liquid argon time projection chambers". In: *Phys. Rev. D* 102 (9 Nov. 2020), p. 092010. DOI: 10.1103/PhysRevD.102.092010. URL: https://link.aps.org/doi/10.1103/PhysRevD.102.092010.

[111] P. Abratenko et al. "Demonstration of neutron identification in neutrino interactions in the MicroBooNE liquid argon time projection chamber". In: *The European Physical Journal C* 84.10 (Oct. 2024), p. 1052. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-024-13423-z. URL: https://doi.org/10.1140/epjc/s10052-024-13423-z.

[112] B. Abi et al. "Supernova neutrino burst detection with the Deep Underground Neutrino Experiment". In: *The European Physical Journal C* 81.5 (May 2021), p. 423. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-021-09166-w. URL: https://doi.org/10.1140/epjc/s10052-021-09166-w.

[113] DUNE Collaboration. *Supernova Pointing Capabilities of DUNE*. 2024. arXiv: 2407.10339 [hep-ex]. URL: https://arxiv.org/abs/2407.10339.

[114] S Andringa et al. "Low-energy physics in neutrino LArTPCs". In: *Journal of Physics G: Nuclear and Particle Physics* 50.3 (Feb. 2023), p. 033001. DOI: 10.1088/1361-6471/acad17. URL: https://dx.doi.org/10.1088/1361-6471/acad17.

[115] MicroBooNE Collaboration et al. "Wire-cell 3D pattern recognition techniques for neutrino event reconstruction in large LArTPCs: algorithm description and quantitative evaluation with MicroBooNE simulation". In: *Journal of Instrumentation* 17.01 (Jan. 2022), P01037. DOI: 10.1088/1748-0221/17/01/P01037. URL: https://dx.doi.org/10.1088/1748-0221/17/01/P01037.

[116] KOLMOGOROV A. L. "Sulla determinazione empirica di una legge di distribuzione". In: *G. Ist. Ital. Attuari* 4 (1933), pp. 83–91. URL: https://cir.nii.ac.jp/crid/1571135650766370304.

[117] N. Smirnov. "Table for Estimating the Goodness of Fit of Empirical Distributions". In: *The Annals of Mathematical Statistics* 19.2 (1948), pp. 279–281. DOI: 10.1214/aoms/1177730256. URL: https://doi.org/10.1214/aoms/1177730256.

[118] Luis Alvarez-Ruso et al. "Recent highlights from GENIE v3". In: *The European Physical Journal Special Topics* 230.24 (Dec. 2021), pp. 4449–4467. ISSN: 1951-6401. DOI: 10.1140/epjs/s11734-021-00295-7. URL: https://doi.org/10.1140/epjs/s11734-021-00295-7.

[119] P. Abratenko et al. "New CC0π GENIE model tune for MicroBooNE". In: *Phys. Rev. D* 105 (7 Apr. 2022), p. 072001. DOI: 10.1103/PhysRevD.105.072001. URL: https://link.aps.org/doi/10.1103/PhysRevD.105.072001.

[120] Yann Coadou. "Boosted Decision Trees". In: *Artificial Intelligence for High Energy Physics*. Chap. Chapter 2, pp. 9–58. DOI: 10.1142/9789811234033_0002. eprint: https://www.worldscientific.com/doi/pdf/10.1142/9789811234033_0002. URL: https://www.worldscientific.com/doi/abs/10.1142/9789811234033_0002.

[121] L. Breiman et al. "Classification and Regression Trees". In: 1984. DOI: 10.1201/9781315139470.

[122] Yoav Freund and Robert E. Schapire. "Experiments with a new boosting algorithm". In: *Proceedings of the Thirteenth International Conference on International Conference on Machine Learning*. ICML'96. Bari, Italy: Morgan Kaufmann Publishers Inc., 1996, pp. 148–156. ISBN: 1558604197.

[123] Jerome H. Friedman. "Greedy function approximation: A gradient boosting machine." In: *The Annals of Statistics* 29.5 (2001), pp. 1189–1232. DOI: 10.1214/aos/1013203451. URL: https://doi.org/10.1214/aos/1013203451.

[124] K. V. Rashmi and Ran Gilad-Bachrach. *DART: Dropouts meet Multiple Additive Regression Trees*. 2015. arXiv: 1505.01866 [cs.LG]. URL: https://arxiv.org/abs/1505.01866.

[125] Tianqi Chen and Carlos Guestrin. "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: Association for Computing Machinery, 2016, pp. 785–794. ISBN: 9781450342322. DOI: 10.1145/2939672.2939785. URL: https://doi.org/10.1145/2939672.2939785.

[126] C Adam-Bourdarios et al. "The Higgs Machine Learning Challenge". In: *Journal of Physics: Conference Series* 664.7 (Dec. 2015), p. 072015. DOI: 10.1088/1742-6596/664/7/072015. URL: https://dx.doi.org/10.1088/1742-6596/664/7/072015.

[127] Leonidas Aliaga Soplin. "Neutrino Flux Prediction for the NuMI Beamline". PhD thesis. William-Mary Coll., 2016. DOI: 10.2172/1250884.

[128] P. Abratenko et al. "Novel approach for evaluating detector-related uncertainties in a LArTPC using MicroBooNE data". In: *The European Physical Journal C* 82.5 (May 2022), p. 454. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-022-10270-8. URL: https://doi.org/10.1140/epjc/s10052-022-10270-8.

[129] L. Lista. *Statistical Methods for Data Analysis: With Applications in Particle Physics*. Lecture Notes in Physics. Springer International Publishing, 2023. ISBN: 9783031199349. URL: https://books.google.co.uk/books?id=PA28EAAAQBAJ.

[130] G. Cowan. *Statistical data analysis*. Oxford University Press, USA, 1998.

[131] Jerzy Neyman and Egon Sharpe Pearson. "On the Problem of the Most Efficient Tests of Statistical Hypotheses". In: *Phil. Trans. Roy. Soc. Lond. A* 231.694-706 (1933), pp. 289–337. DOI: 10.1098/rsta.1933.0009.

[132] Glen Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *The European Physical Journal C* 71.2 (Feb. 2011), p. 1554. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-011-1554-0. URL: https://doi.org/10.1140/epjc/s10052-011-1554-0.

[133] Lukas Heinrich, Matthew Feickert, and Giordon Stark. *pyhf: v0.7.6*. Version 0.7.6. https://github.com/scikit-hep/pyhf/releases/tag/v0.7.6. DOI: 10.5281/zenodo.1169739. URL: https://doi.org/10.5281/zenodo.1169739.

[134] Lukas Heinrich et al. "pyhf: pure-Python implementation of HistFactory statistical models". In: *Journal of Open Source Software* 6.58 (2021), p. 2823. DOI: 10.21105/joss.02823. URL: https://doi.org/10.21105/joss.02823.

[135] Kyle Cranmer et al. *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*. Tech. rep. New York: New York U., 2012. DOI: 10.17181/CERN-OPEN-2012-016. URL: https://cds.cern.ch/record/1456844.

Part III.

Appendix

A. Full blip-pair BDT training variables (signal vs. background)

Table A.1.: Summary of blip-pair BDT training variables, separated in four categories.

Category	Variables	N. variables
Per-blip, index-wise	Blip i positions (x_i, y_i, z_i)	6
(blip 1 and blip 2)	Blip i extents Δx_i, $\Delta(yz)_i$	4
Per blip, magnitude-wise	Blip size	2
(blip min. and blip max.)	Blip energy	2
	Number of hits in the collection plane	2
	Sum of number of hits in induction planes	2
	Minimum distance to a track of each blip	2
Blip pair	Blip angles	2
	Blip separation	1
Third blip	Transverse distance of nearest 3rd blip	3
	Min. transverse distance to nearest 3rd blip	1
	Overall distance to nearest 3rd blip	1

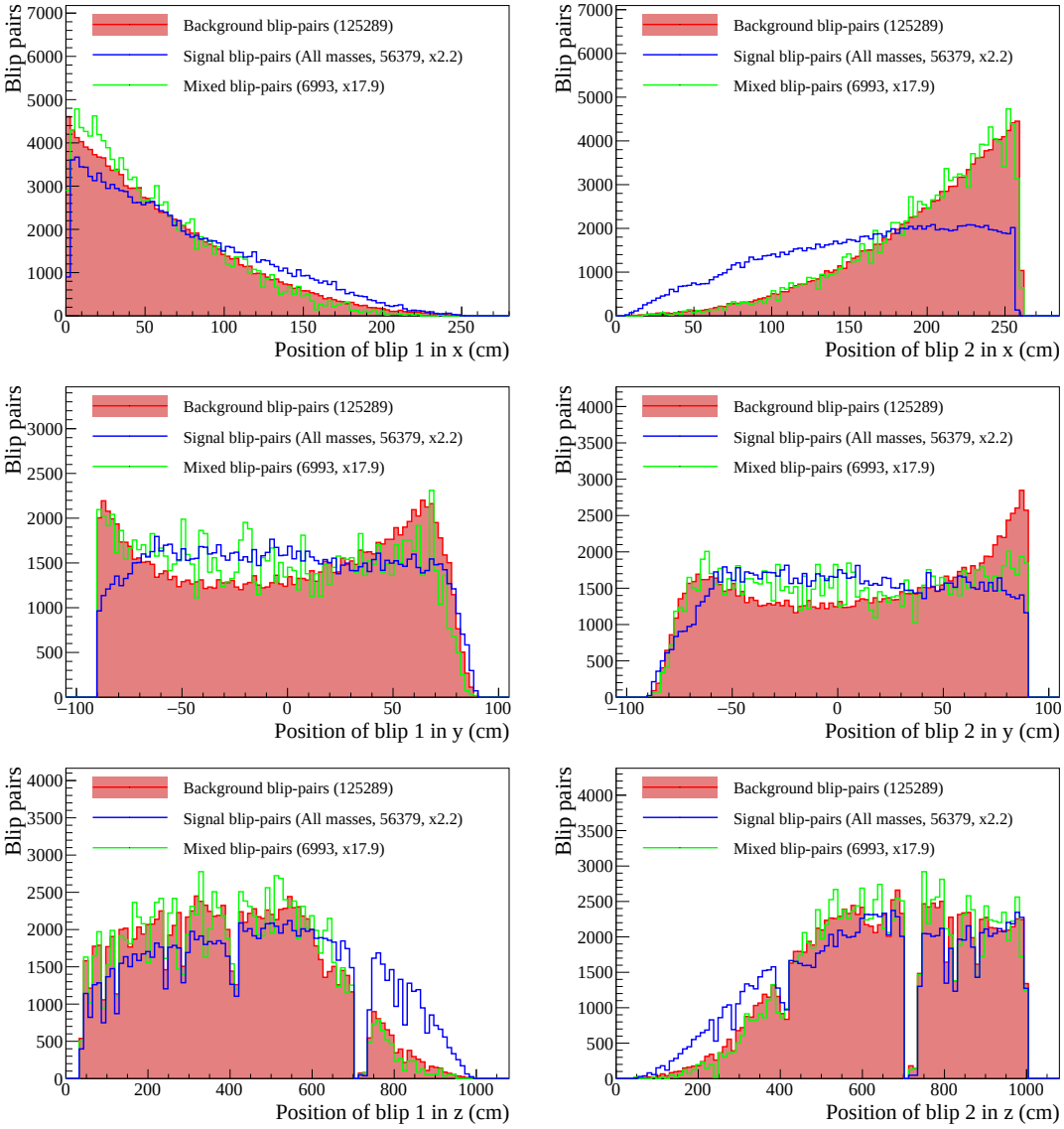


Figure A.1.: Per-blip, index-wise: blip position variables

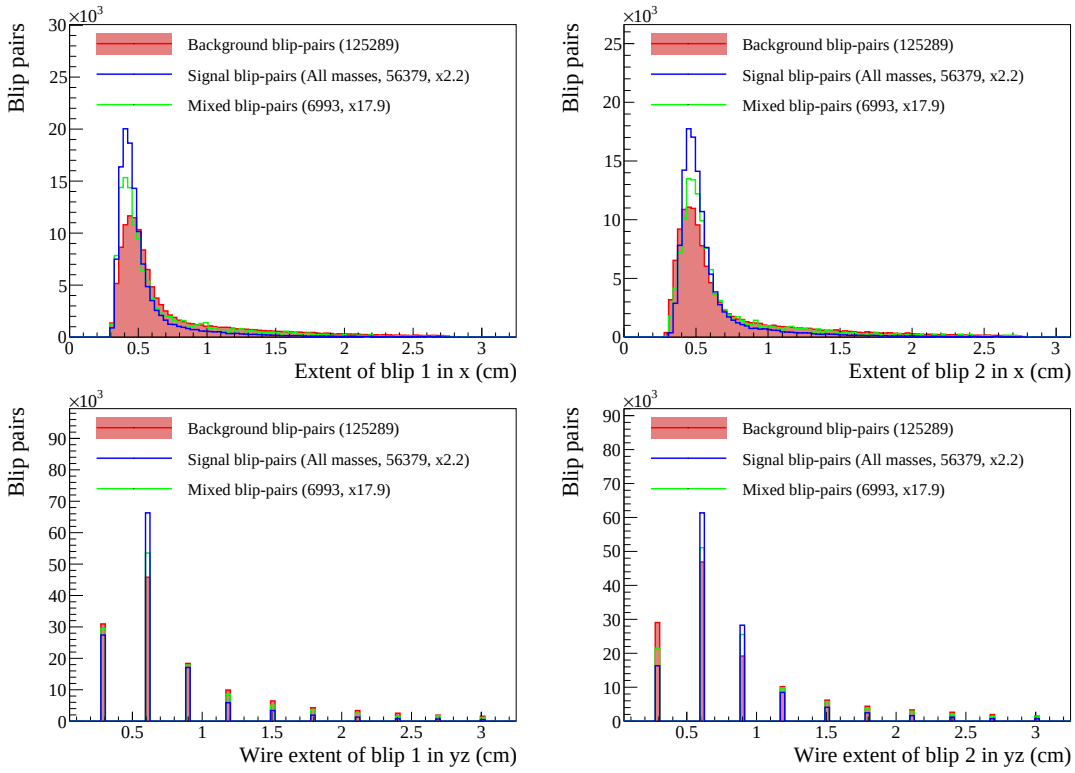


Figure A.2.: Per-blip, index-wise: blip extent variables

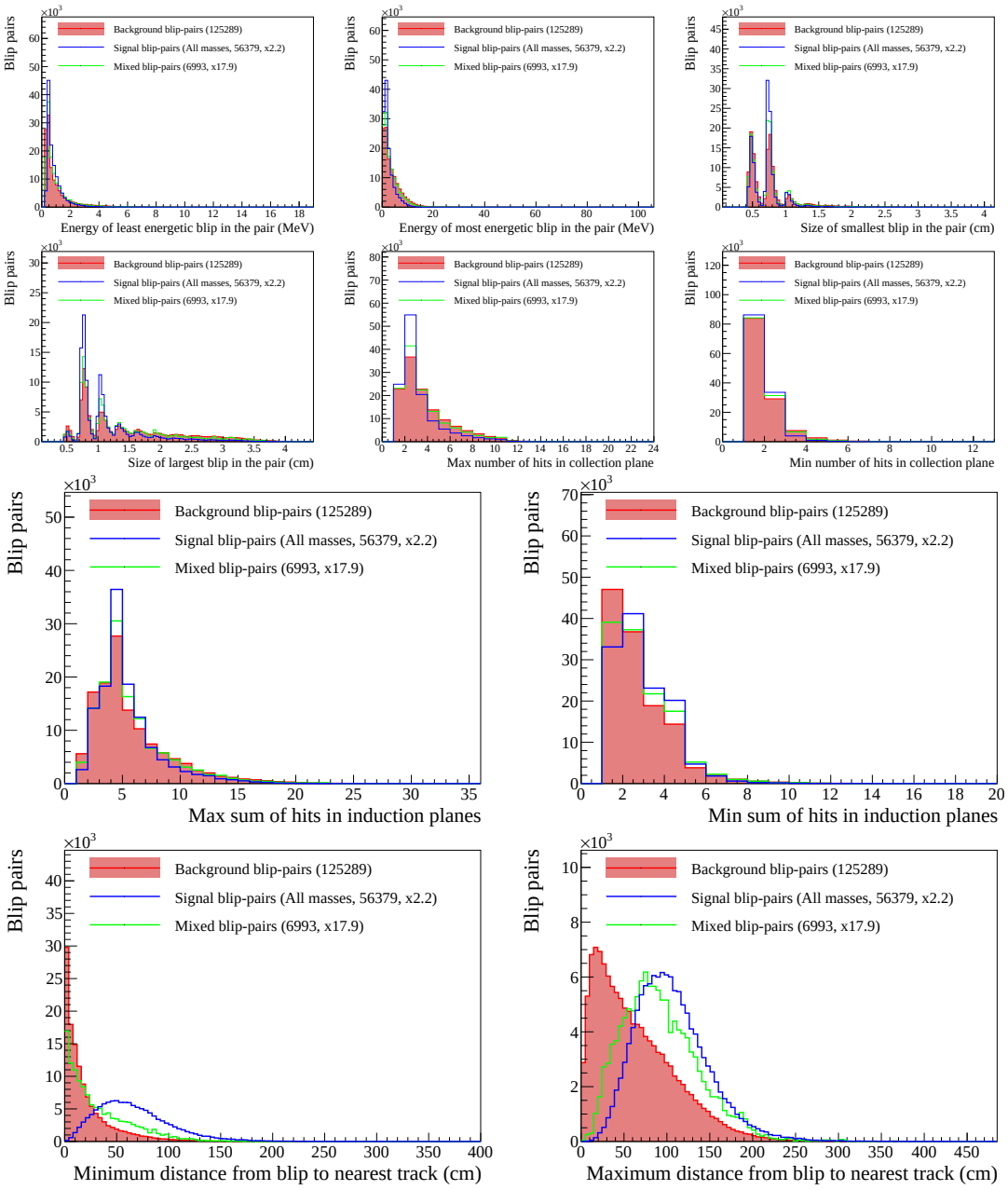


Figure A.3.: Per blip, magnitude-wise training variables.

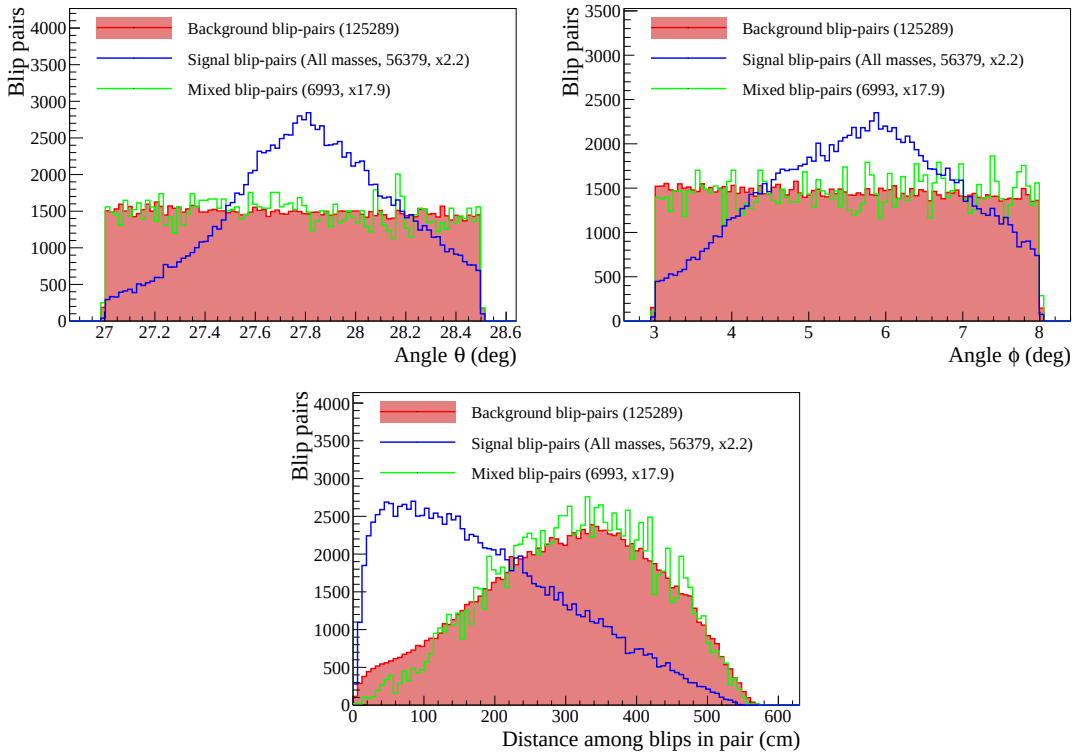


Figure A.4.: Blip-pair BDT training variables. Top: angle variables θ and ϕ. Bottom: Blip separation.

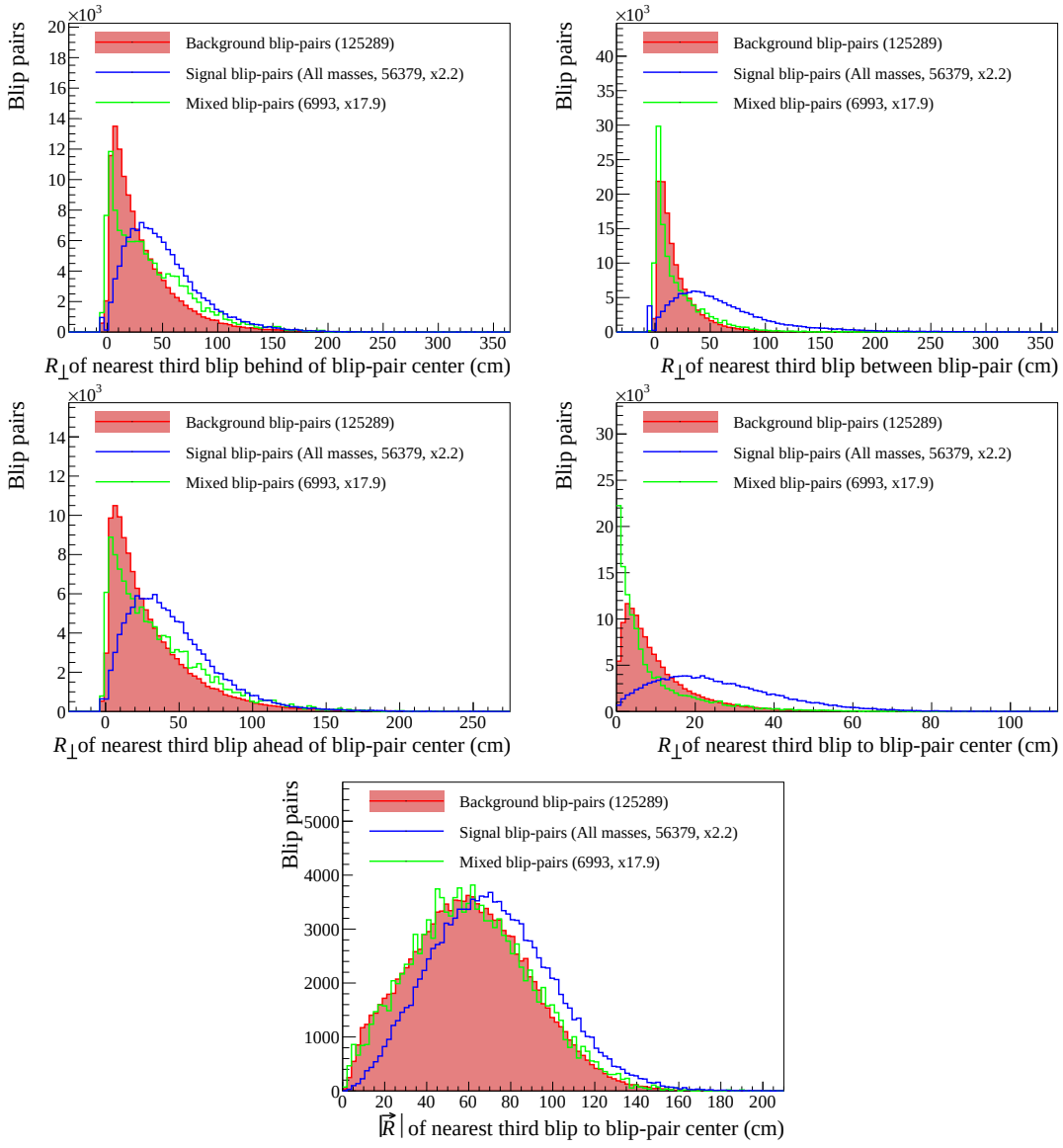


Figure A.5.: "Relation to third-blip" BDT training variables.

B. Full blip-pair BDT training variables (data vs. MC)

Table B.1.: Summary of blip-pair BDT training variables, separated in four categories.

Category	Variables	N. variables
Per-blip, index-wise	Blip i positions (x_i, y_i, z_i)	6
(blip 1 and blip 2)	Blip i extents Δx_i, $\Delta(yz)_i$	4
Per blip, magnitude-wise	Blip size	2
(blip min. and blip max.)	Blip energy	2
	Number of hits in the collection plane	2
	Sum of number of hits in induction planes	2
	Minimum distance to a track of each blip	2
Blip pair	Blip angles	2
	Blip separation	1
Third blip	Transverse distance of nearest 3rd blip	3
	Min. transverse distance to nearest 3rd blip	1
	Overall distance to nearest 3rd blip	1

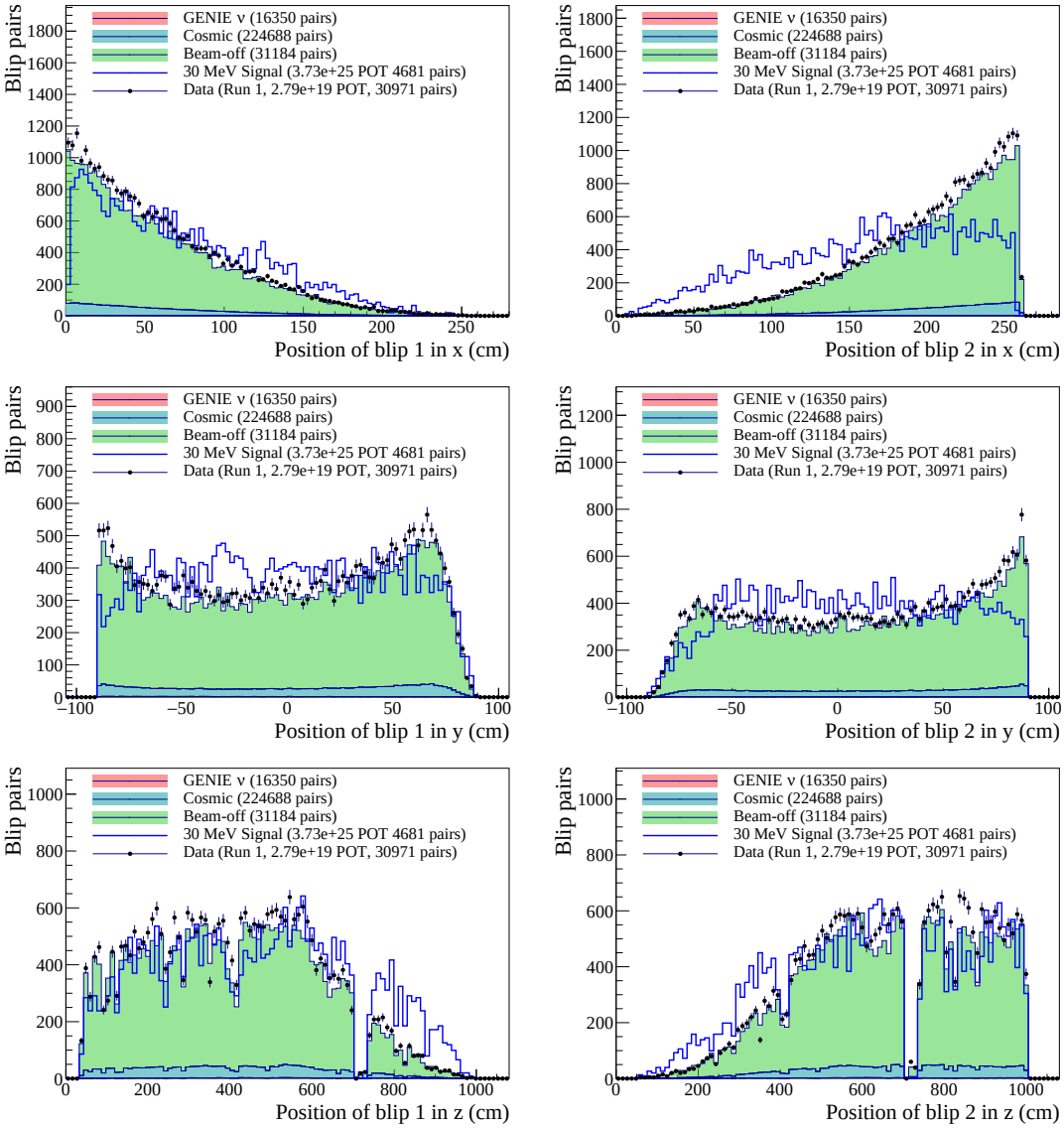


Figure B.1.: Per-blip, index-wise: blip position variables

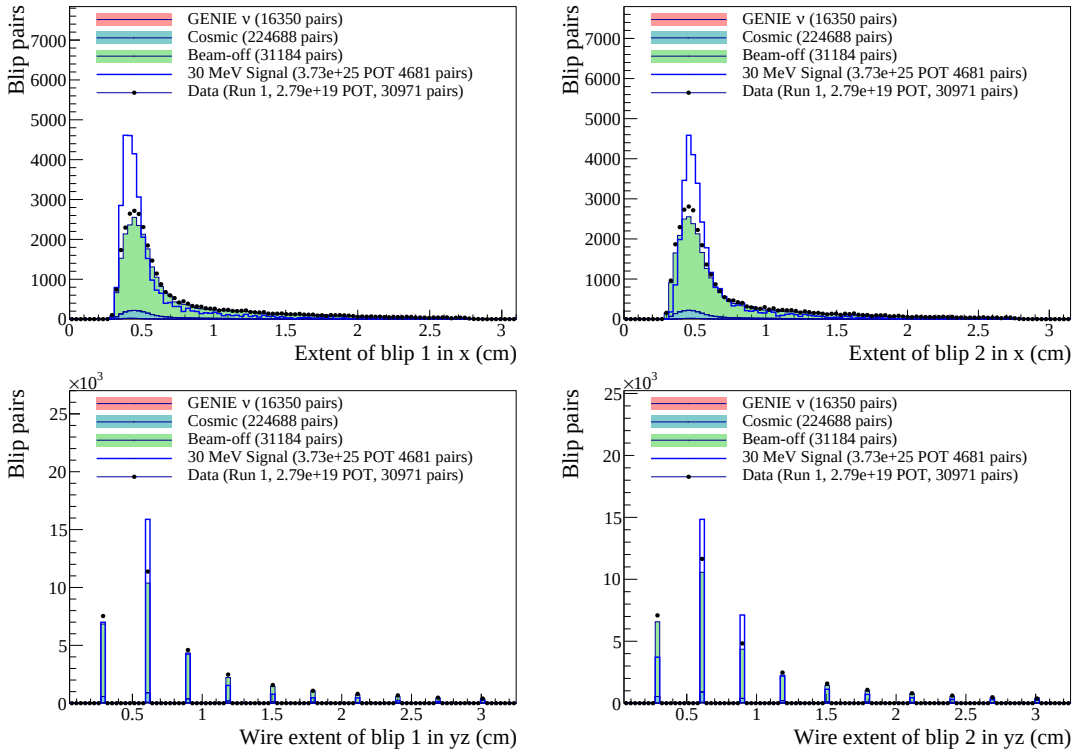


Figure B.2.: Per-blip, index-wise: blip extent variables

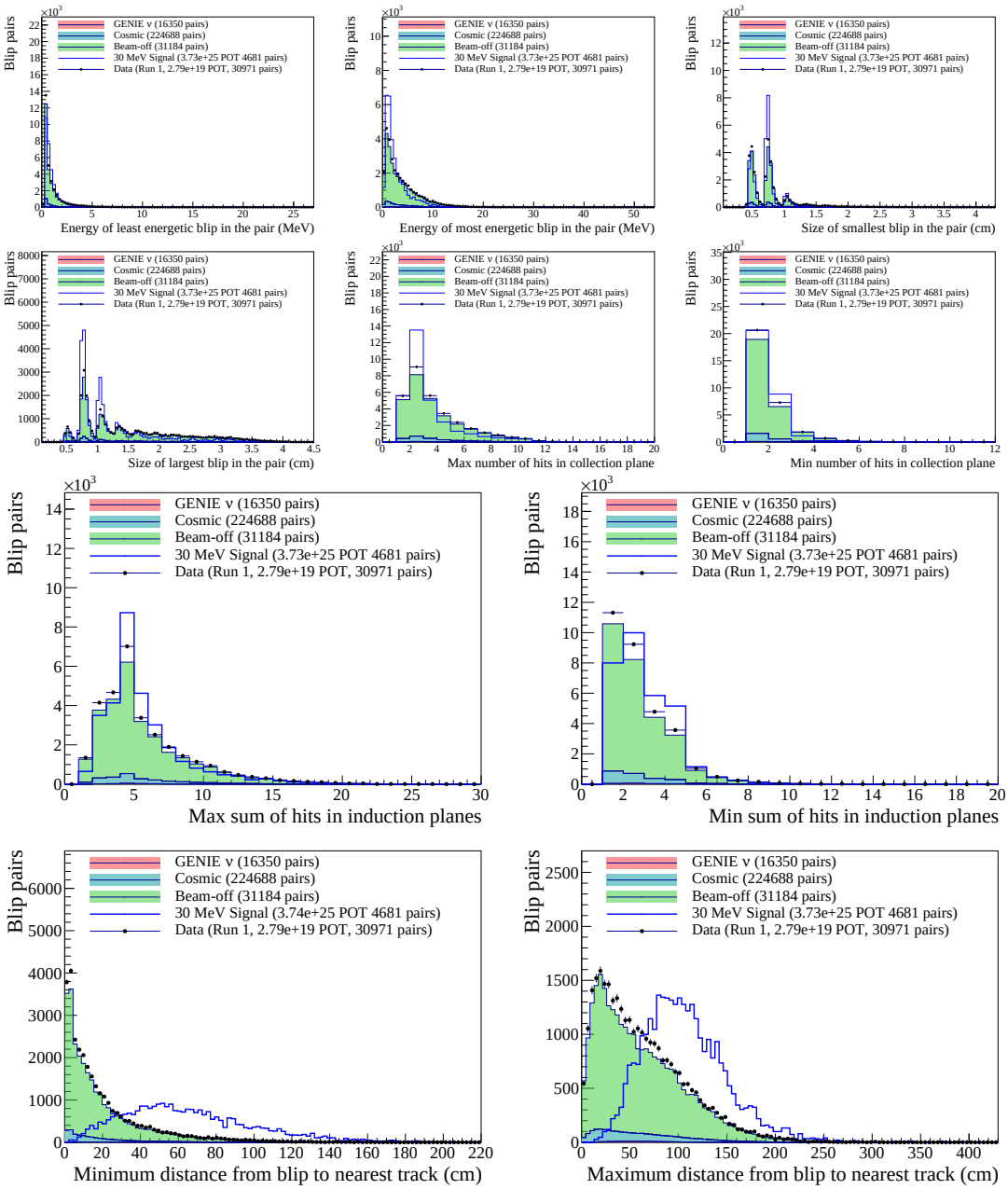


Figure B.3.: Per blip, magnitude-wise training variables.

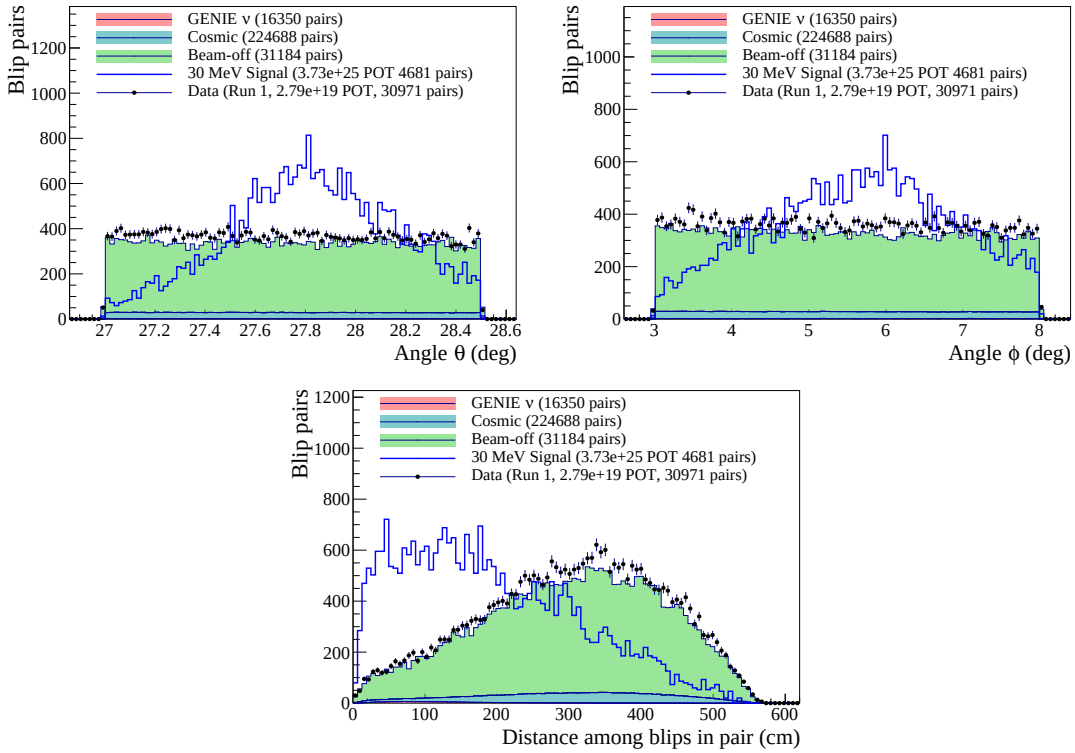


Figure B.4.: Blip-pair BDT training variables. Top: angle variables θ and ϕ. Bottom: Blip separation.

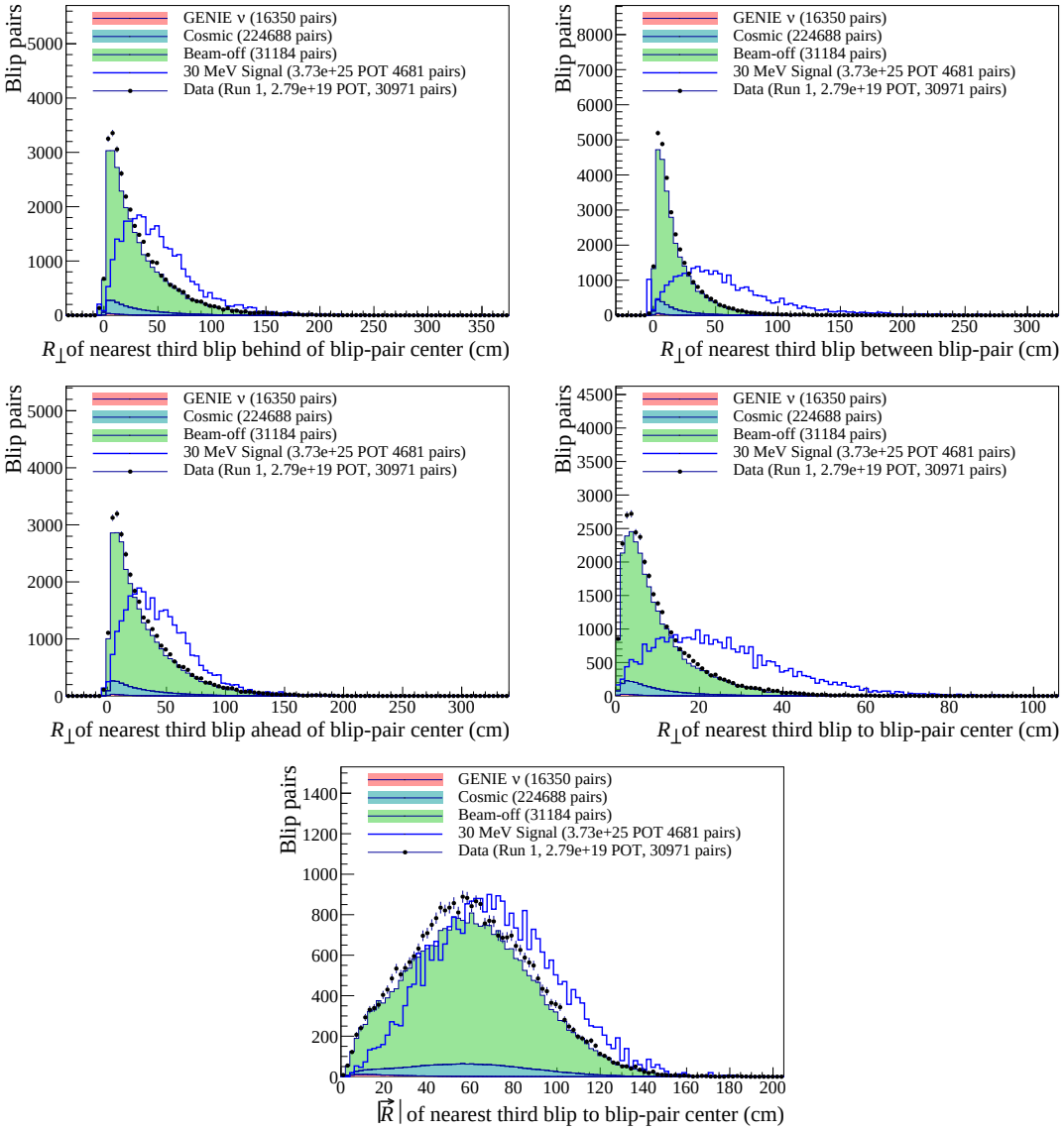


Figure B.5.: "Relation to third-blip" BDT training variables.