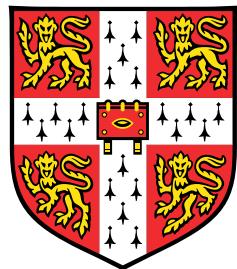


Writing your PhD thesis in **LATEX2e**

Using the CUED template



Krishna Kumar

Department of Engineering
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

King's College

January 2017

I would like to dedicate this thesis to my loving parents ...

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Krishna Kumar
January 2017

Acknowledgements

And I would like to acknowledge ...

Abstract

This is where you write your abstract ...

Table of contents

List of figures	8
List of tables	12
Nomenclature	13
1 Introduction	14
1.1 XXXXX	14
1.2 XXXXXX	14
2 Theory	15
2.1 Theory of neutrino physics	15
2.2 Nucleon decay in Grand Unifying Theories	15
2.3 Existing and future experiments	15
2.4 How Liquid Argone Time Projection Chambers work	15
3 The Deep Underground Neutrino Experiment	16
3.1 DUNE location and beamline	16
3.2 The DUNE detectors and schedule	16
3.3 Physics opportunities of DUNE	16
3.3.1 Neutrino physics	16
3.3.2 Nucleon decay and supernovae neutrinos	16
3.4 Path to building DUNE - The 35 ton prototype	16
3.5 The DUNE software	16
4 The 35 ton camera system	24
4.1 The need for cameras in a Liquid Argon Time Projection Chamber	24
4.2 Design of the camera system	24
4.3 Tabletop tests	24

Table of contents	7
4.4 Safety reviews and installation	24
4.5 Performance in the 35 ton	24
5 Simulations of the 35 ton prototype	25
5.1 Determination of interaction times	25
5.2 Calibrating calorimetric constants	29
5.3 Discerning reconstruction efficiencies	30
5.4 Performing particle identification	36
6 The 35 ton data sample	49
6.1 Organisation of the data structure	49
6.2 Reformatting the data to the offline structure	53
6.3 Observations on data quality and noise mitigation	54
6.4 Performance of reconstruction algorithms	59
6.5 Measuring interaction times using electron diffusion	65
6.5.1 Determining interaction times in 35 ton data	69
6.5.2 Determining interaction times in a low-noise detector using Monte Carlo, and differences with data	76
6.5.3 Discerning the impact of changing detector properties using Monte Carlo samples	81
6.5.4 The limitations of and future improvements to the method of interaction time determination using diffusion	85
7 Simulations of the DUNE Far Detector	89
7.1 Simulations of the LBNE surface detector	89
7.2 The use of MUSUN in LArSoft	89
7.3 Cosmogenic background for nucleon decay channels in DUNE	92
References	94
Appendix A Supporting figures to Monte Carlo studies concerning determining interaction times using the effects of diffusion	97
Appendix B Something else mildly interesting	103

List of figures

3.1	The wrapped wires of the 35 ton	16
3.2	A representation of the counter locations in the 35 ton	17
3.3	The LArSoft co-ordinate system as it is represented in the 35 ton.	19
3.4	Reconstructed hits from a simulated energy deposition	20
3.5	Performing disambiguation with different wire pitches.	22
5.1	Matching tracks and flashes in the 35 ton using positions in the yz plane . .	26
5.2	How the number of photoelectrons measured changes with drift distance . .	27
5.3	The predicted x positions of flashes using the relationship between photoelectron and drift distance	28
5.4	The difference in Monte Carlo interaction times and the predicted interaction times using the photon detectors	28
5.5	The tuning of the calorimetric constants in the 35 ton	29
5.6	The reconstruction efficiencies for simulated events as a function of Monte Carlo truth track length.	32
5.7	The reconstruction efficiencies for simulated events as a function of Monte Carlo truth deposited energy.	32
5.8	The reconstruction efficiencies for simulated events as a function of Monte Carlo truth track angle in theta.	33
5.9	The reconstruction efficiencies for simulated events as a function of Monte Carlo truth track angle in phi.	33
5.10	The reconstruction efficiencies for simulated events as a function of Monte Carlo truth track angle in theta and phi.	34
5.11	The medium and particle type dependence of the Bethe-Bloch equation . .	37
5.12	Defining the PIDA metric for particle identification.	38
5.13	The reconstruction efficiencies for protons in a sample generated using CRY. .	40
5.14	The reconstruction efficiencies for single muons and protons in the 35 ton. .	42
5.15	The calculated PIDA values for single muons and protons in the 35 ton. .	43

List of figures	9
5.16 The $\frac{dE}{dx}$ versus residual range plot for single muons and protons in the 35 ton.	44
5.17 A comparison between PIDA values calculated using truth and reconstructed information	44
5.18 The calculated PIDA values for muons and protons in a CRY sample through the 35 ton	48
6.1 The 35 ton data sample	50
6.2 The 35 ton data structure	52
6.3 Dropped TPC data in the 35 ton	55
6.4 Recovering stuck ADC codes in the 35 ton	56
6.5 Removing coherent noise in the 35 ton	56
6.6 Applying Wiener filters to the 35 ton data	57
6.7 The effect of noise removal algorithms in the 35 ton	58
6.8 dQ/dx in the 35 ton as a function of drift time	61
6.9 The numbering scheme for the East - West counters in the 35 ton	62
6.10 The dot product of the track and vector joining the centres of the coincidence counters in the yz plane	63
6.11 The alignment of reconstructed tracks with the vectors joining the centres of the coincidence counters	63
6.12 Reconstruction efficiencies of through going tracks in the 35 ton data	66
6.13 Schematic showing the process of diffusion	67
6.14 A simulated event display showing multiple tracks and flashes in the 35 ton	68
6.15 The effect of adding a noise baseline to a hit	70
6.16 The most probable values of the <i>RMS</i> and <i>RMS/Charge</i> distributions for tracks with a counter difference of 4 in the 35 ton data	71
6.17 The drift distance dependence of diffusion in the 35 ton dataset for coincidences with a counter difference of 4	72
6.18 The angular dependence of diffusion in the 35 ton dataset for hits within 10 cm of the APAs	72
6.19 The difference between the predicted and reconstructed hit times in the 35 ton dataset	73
6.20 The accuracy of the hit <i>RMS</i> method in the 35 ton dataset	74
6.21 The accuracy of the hit <i>RMS/Charge</i> method in the 35 ton dataset	75
6.22 The most probable values of the hit <i>RMS</i> and hit <i>RMS/Charge</i> distributions for tracks with a counter difference of 4 in a low noise 35 ton detector	77
6.23 The drift distance dependence of diffusion in the 35 ton dataset and Monte Carlo for coincidences with a counter difference of 4	77

6.24	The angular dependence of diffusion in the 35 ton dataset and Monte Carlo for hits within 10 cm of the APAs	78
6.25	Comparing the accuracy of the hit <i>RMS</i> method in the 35 ton dataset and a Monte Carlo simulation	79
6.26	Comparing the accuracy of the hit <i>RMS</i> method in the 35 ton dataset and a Monte Carlo simulation	80
6.27	Comparing the accuracy of the hit <i>RMS</i> method, as the electron lifetime changes	82
6.28	Comparing the accuracy of the hit <i>RMS</i> method, as the electron lifetime changes	83
6.29	Comparing the accuracy of the hit <i>RMS</i> method, as the electric field changes	84
6.30	Comparing the accuracy of the hit <i>RMS</i> method, as the electric field changes	85
6.31	Comparing the accuracy of the hit <i>RMS</i> method, as the constant of longitudinal diffusion changes	86
6.32	Comparing the accuracy of the hit <i>RMS</i> method, as the constant of longitudinal diffusion changes	87
7.1	The correlation between the surface profile and distribution of azimuthal angles at the DUNE far detector site	91
7.2	The distributions of some of the important quantities for muons generated by MUSUN in LArSoft	92
7.3	The initial positions of muons generated by MUSUN around a DUNE 10 kt module	93
A.1	The most probable values of the hit <i>RMS</i> and hit <i>RMS/Charge</i> distributions for tracks with a counter difference of 4, as the electron lifetime changes	98
A.2	The drift distance dependence of diffusion in the 35 ton dataset and Monte Carlo for coincidences with a counter difference of 4, as the electron lifetime changes	98
A.3	The angular dependence of diffusion in the 35 ton dataset and Monte Carlo for hits within 10 cm of the APAs, as the electron lifetime changes	99
A.4	The most probable values of the hit <i>RMS</i> and hit <i>RMS/Charge</i> distributions for tracks with a counter difference of 4, as the electric field changes	99
A.5	The drift distance dependence of diffusion in the 35 ton dataset and Monte Carlo for coincidences with a counter difference of 4, as the electric field changes	100

A.6 The angular dependence of diffusion in the 35 ton dataset and Monte Carlo for hits within 10 cm of the APAs, as the electric field changes	100
A.7 The most probable values of the hit <i>RMS</i> and hit <i>RMS/Charge</i> distributions for tracks with a counter difference of 4, as the constant of longitudinal diffusion changes	101
A.8 The drift distance dependence of diffusion in the 35 ton dataset and Monte Carlo for coincidences with a counter difference of 4, as the constant of longitudinal diffusion changes	101
A.9 The angular dependence of diffusion in the 35 ton dataset and Monte Carlo for hits within 10 cm of the APAs, as the constant of longitudinal changes .	102

List of tables

5.1	Stopping power parameterization for various particle types in liquid Argon .	38
5.2	The properties of initial particles simulated in the muon and proton samples.	41
5.3	A summary of the PIDA values calculated for the proton sample as sequential cuts are applied	46
5.4	A summary of the PIDA values calculated for the proton sample as sequential cuts are applied	46
5.5	A summary of the PIDA values calculated for the primary particles in the proton sample as sequential cuts are applied	46
6.1	The angles relative to the APAs corresponding to given counter differences	64
7.1	Muon flux parameters as calculated with MUSIC/MUSUN.	90

Nomenclature

Roman Symbols

tick Unit of time equal to 500 ns

Acronyms / Abbreviations

CRC Cosmic Ray Counter

FD Far Detector

MIP Minimally Ionising Particle

MPV Most Probable Value

PID Particle IDentification

PoCA Point of Closest Approach

ROI Region Of Interest

ADC Analogue to Digital Converter

SiPM Silicon Photo Multiplier

TPC Time Projection Chamber

¹ **Chapter 1**

² **Introduction**

³ **1.1 XXXXX**

⁴ **1.2 XXXXXX**

Chapter 2

Theory

2.1 Theory of neutrino physics

3

2.2 Nucleon decay in Grand Unifying Theories

4

2.3 Existing and future experiments

5

2.4 How Liquid Argone Time Projection Chambers work

6

¹ Chapter 3

² **The Deep Underground Neutrino ³ Experiment**

⁴ **3.1 DUNE location and beamline**

⁵ **3.2 The DUNE detectors and schedule**

⁶ **3.3 Physics opportunities of DUNE**

⁷ **3.3.1 Neutrino physics**

⁸ **3.3.2 Nucleon decay and supernovae neutrinos**

⁹ **3.4 Path to building DUNE - The 35 ton prototype**

Fig. 3.1 A schematic showing what the wrapped wire planes of the DUNE detector designs looked like in the 35 ton.

¹⁰ **3.5 The DUNE software**

¹¹ The software package used by DUNE is called LArSoft [1] [2] which is a simulation, recon-
¹² struction and analysis package for LArTPCs that is being used by many of the experiments
¹³ in the US neutrino program. LArSoft has been developed to be detector agnostic, meaning

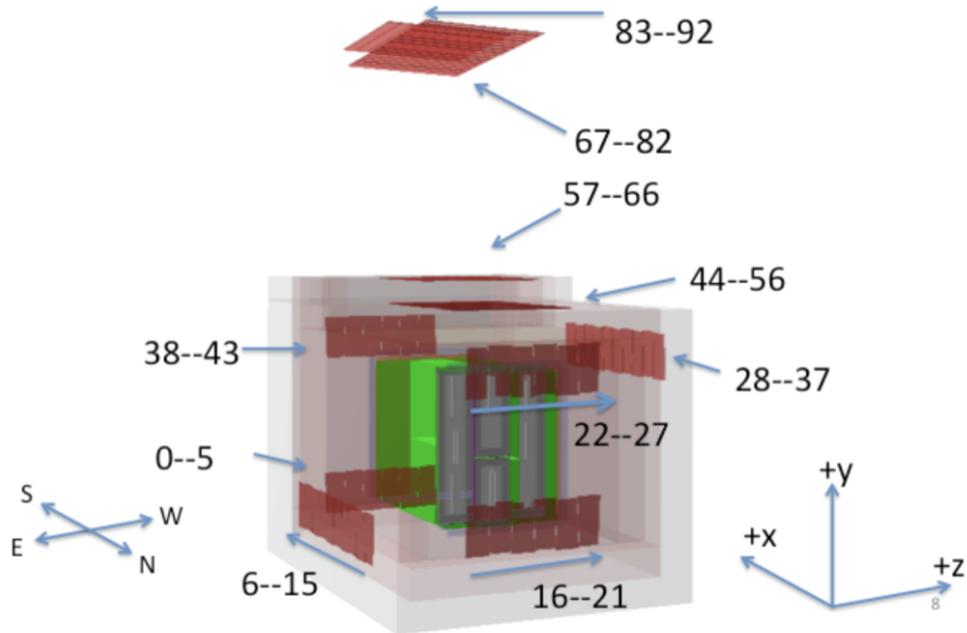


Fig. 3.2 A representation of the counter locations in the 35 ton, with the magnetic and LArSoft co-ordinate systems shown. The other detector components can be seen inside the cryostat, such that the counters on the North wall are behind the short drift volume. The East - West counters are numbered 6-15 and 28-37 respectively. The North Lower - South Upper counters are numbered 16-21 and 38 - 43 respectively. The North Upper - South Lower counters are numbered 22-27 and 0-5 respectively. The telescope triggers are numbered 44-92 and are split into four groups.

that much of the code is shared between experiments. To this end it is envisioned that it will be used as a platform for constant development in both existing experiments and those still in the planning phases such as DUNE. LArSoft is built around the Fermilab-supported analysis reconstruction framework (*art*). External packages such as ROOT [3] and GEANT4 [4] are incorporated into LArSoft meaning that the user does not have to co-ordinate specific versions of the packages as the newest versions are automatically incorporated.

There are numerous mechanisms by which particles can be generated within the software with external packages such as GENIE [5], Nuance [6] and CRY [7] already having been incorporated. Recently the MUon Simulations Underground (MUSUN) [8] [9] generator which takes the output of MUon SImulation Code (MUSIC) [8] [10] [11] has also been incorporated, see Section 7.2 for further details. It is also possible to use an inbuilt single particle generation mode which is fully tunable as particle type, momenta, positions and

3.5 The DUNE software

18

¹ directions can all be varied.

²

³ The co-ordinates and angles in LArSoft are defined as follows, and schematic representations of how this appears in the 35 ton are shown in Figure 3.3:

⁵ • x - The beam direction, with maximal x being where the beam enters the detector.

⁶ – In the 35 ton prototype where there is no beam positive x is in the opposite
⁷ direction to that which electrons drift in the large TPC, where $x = 0$ is the position
⁸ of the APA frames in the long drift volume.

⁹ – In the far detector geometry $x = 0$ is defined as

¹⁰ • y - The vertical direction, with maximal y being the most highest point.

¹¹ – In the 35 ton $y = 0$ is halfway between the gap created by the two centre APAs
¹² which are mounted one above the other.

¹³ – In the far detector $y = 0$ is

¹⁴ • z - Defined as such to have a right handed co-ordinate system.

¹⁵ – In the 35 ton $z = 0$ is at the edge of the leftmost APA frame when looking down
¹⁶ the long drift volume.

¹⁷ – In the far detector $z = 0$ is

¹⁸ • θ - The angle that a point makes from the x axis in the xy plane.

¹⁹ • ϕ - The angle between the z axis and the point.

²⁰ The simulation of particles is usually split into five separate distinct processes to reflect
²¹ the different stages in which development often progresses. The advantage of segmenting
²² the computational process in this way is that improvements can easily applied to a file
²³ without rerunning the entire chain. This is especially important when large Monte Carlo
²⁴ or data samples are produced for general use within collaborations so that users are able
²⁵ to concentrate on improving a specific part of the computational process. When these all
²⁶ purpose samples are produced the analysis performed provides users with any Monte Carlo
²⁷ truth information along with the reconstructed quantities for use in analyses performed
²⁸ outside LArSoft. The computational process is often broken down in the following way:

²⁹ • Generation.

³⁰ • GEANT4.

3.5 The DUNE software

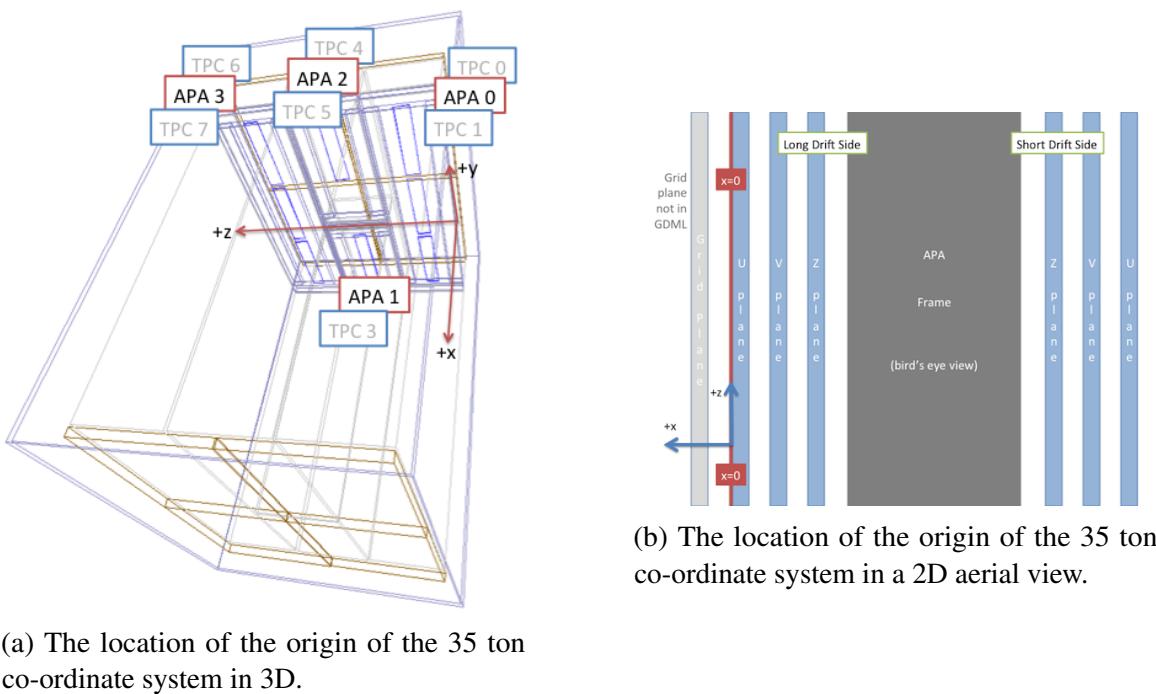


Fig. 3.3 The LArSoft co-ordinate system as it is represented in the 35 ton.

- Full detector simulation, including detector responses after which Monte Carlo is equivalent to collected data.
- Full detector reconstruction.
- Analysis.

Later significant focus will be given to the reconstruction of TPC data, and so it is necessary to briefly illustrate the mechanisms by which TPC data is reconstructed in LArSoft. Much of the information presented below is summarised in [12]. After the full detector simulation or data taking, detector effects such as the electronics response function and a pedestal offset have to removed. Once these effects are removed the signal is estimated using the optimal value of *signal/noise* which would produce the measured signal. This process, called deconvolution, does not conserve pulse height and is not guaranteed to preserve the normalisation. The deconvoluted signals are all unipolar distributions which means that Gaussian distributions can then be fitted to them when trying to reconstruct hits.

The deconvoluted signals are reconstructed into hits by identifying regions that are above a threshold value and then attempting to replicate the signal in these regions by introducing Gaussian distributions. For isolated hits this is typically achieved using only one Gaussian distribution, however for large energy depositions over a large period time where many particles

3.5 The DUNE software

are involved multiple Gaussian distributions are often required. Large energy depositions are also possible when the orientation of the particle aligns with a wire, this means that all of the deposited energy is collected on this single wire. Examples of reconstructed hits are shown in Figure 3.4, where Figures 3.4a and 3.4b show multiple time separated energy depositions on a collection and induction wire respectively. A more complex energy deposition on a collection plane wire is shown in Figure 3.4c where energy depositions from many particles at similar times have created a complicated energy deposition that requires many reconstructed hits to explain.

9

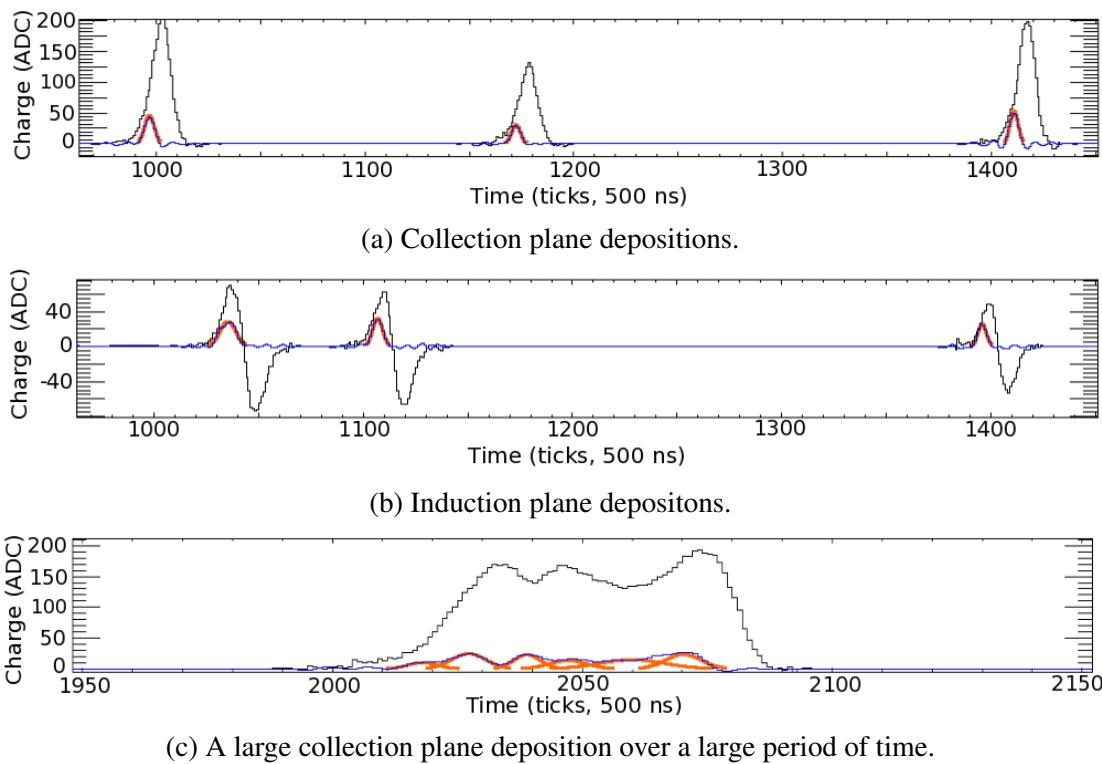


Fig. 3.4 The raw and deconvoluted signals with reconstructed hits on a single wire for a simulated energy deposition. The plot is shown with increasing charge on the y axis, and increasing time (in ticks) on the x axis. The black line shows the raw signal, the blue line shows the deconvoluted signal and the orange lines show the reconstructed hits.

As noted in Section 3.2 and Section 3.4 the DUNE FD and the 35 ton both have wrapped wires on the induction planes. A result of this is that the location of where the reconstructed hit occurred on an induction wire is ambiguous as a single wire has many wire segments, as shown in Figure 3.1. An important feature of this ambiguity is that the TPC in which the hit occurred cannot be identified unless it is combined with another hit. These ambiguities do not extend to the collection plane wires as they are not wrapped and so consist of only a

single wire segment in a single TPC. Hits are combined across the three planes by identifying wire segments on each plane which intersect and have hits at common times. In the traditional reconstruction process only hits that make these so-called 'triple points' are considered disambiguated, with other hits being identified as noise hits causing them to be discarded.

The inclination of the wire planes has to be carefully chosen so as to minise both the number of wires required and the number of times that wire triplets intersect. It is also important that all wires on a given APA are either read at the top or base of the APA due to the number of APAs required to build a detector of DUNE's scale. The inclination of wires in the 35 ton was $45^\circ \pm 0.7^\circ$ meaning that many wire triplets cross twice and some wire pairs cross three times. When wire triplets cross multiple times the triplet which has the smallest distance between the common intersection point and the two-wire intersection points is chosen as the best intersection candidate. The different wire pitches are neccessary so that one of the triple points can be evaluated to be a better candidate, as with a wire pitch of 45° it can be impossible to distinguish between different triple points. The inclination of wires in the FD was chosen to be 36° to remove the possibility of multiple intersection points as given the geometry of the APAs multiple intersection points are impossible and so disambiguation is much simpler, but there are more wires on each of the induction planes making it more expensive to instrument. This is shown in Figure 3.5.

Once the hits have been disambiguated they are combined to make clusters in each of the three planes, before the clusters are merged to make reconstructed tracks or showers. The clustering process is usually performed in wire-tick space on each plane separately with hits from a single physical entity being grouped together. It is possible to help seed the start of clusters by using imaging techniques such as a Harris transform [13], or to identify straight lines by using Hough transforms [14]. As hits from a physical entity are unlikely to remain on a single channel or all come at identical times, clusters are often spread out over many channels for a range of times especially when performing clustering for showers.

Once clusters have been identified in each plane they can then be merged into 3-dimensional tracks and showers. The two most common tracking algorithms are PM-Track [15] and Pandora [16], and the most common showering algorithm is EMShower [17]. Once 3D objects have been reconstructed the calorimetric quantities need to be determined, this is often done separately for each plane. Two models exist for calculating $\frac{dE}{dx}$ in LAr-Soft, Birks model [18] and a modified Box model [19] which uses a correction to the Box model [20] at low values of $\frac{dE}{dx}$. Normally the modified Box model is used as it holds for both

3.5 The DUNE software

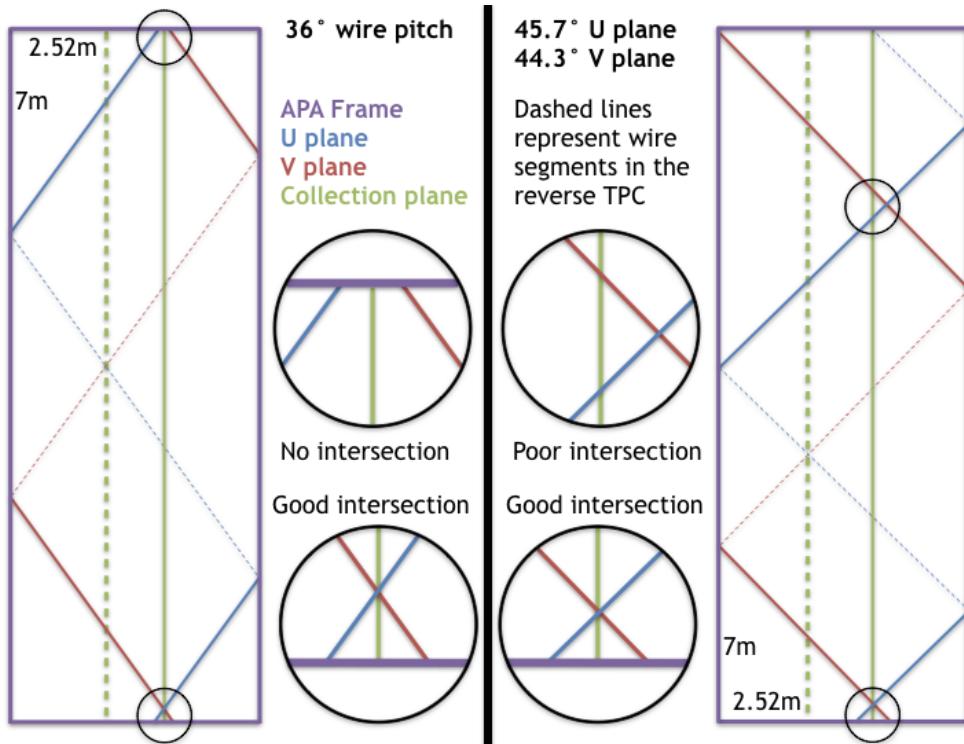


Fig. 3.5 The effect that different wire pitches have on the ability to perform disambiguation in APA with the far detector geometry. The left panel shows a wire pitch of 36° , which is the reference design for the far detector, whilst the right panel shows wire pitches of $45^\circ \pm 0.7^\circ$, as was used in the 35 ton.

- ¹ large and small ionizations, whereas Birks model experiences difficulties at large ionisations
- ² and the traditional Box model struggles at low $\frac{dE}{dx}$. Both models incorporated in LArSoft
- ³ calculate the $\frac{dE}{dx}$ of a hit using the deposited charge (dQ) and the track pitch (dx) of the hit
- ⁴ as well as the conversion of ADC value to number of electrons ($C_{GeV \rightarrow e^-}$), the LAr density
- ⁵ (ρ), the electric field (E_{field}) and tunable electron recombination factors ($Recomb_X$). The
- ⁶ series of equations used in Birks model are shown in Equation 3.1a, whilst those used in the

modified Box model are shown in Equation 3.2a.

$$\frac{dE}{dx} = \frac{dQdx_e}{A - B} \quad (3.1a) \quad 3$$

$$dQdx_e = \frac{dQ \times C_{lifetime}}{dx \times C_{ADC \rightarrow e^-}} \quad (3.1b) \quad 4$$

$$A = \frac{Recomb_A}{C_{GeV \rightarrow e^-}} \quad (3.1c) \quad 5$$

$$B = \frac{\frac{Recomb_B}{\rho}}{E_{field} \times dQdx_e} \quad (3.1d) \quad 6$$

$$\frac{dE}{dx} = \frac{e^A - Recomb_A}{B} \quad (3.2a) \quad 8$$

$$A = B \times C_{GeV \rightarrow e^-} \times \frac{dQ}{dx} \quad (3.2b) \quad 9$$

$$B = \frac{Recomb_B}{\rho \times E_{field}} \quad (3.2c) \quad 10$$

11

When performing calorimetry it is also important that the interaction time is known so that the x positions of hits can be corrected, as they will be reconstructed assuming an interaction time of 0 s. This assumption is made because when using beam events the beam trigger is placed at a time of $T = 0$. An unknown interaction time causes the hit and track positions to be calculated incorrectly, and will also skew the calorimetrics corrections, as recombination is a drift dependant effect.

12

13

14

15

16

17

¹ **Chapter 4**

² **The 35 ton camera system**

³ **4.1 The need for cameras in a Liquid Argon Time Projec-
⁴ tion Chamber**

⁵ **4.2 Design of the camera system**

⁶ **4.3 Tabletop tests**

⁷ **4.4 Safety reviews and installation**

⁸ **4.5 Performance in the 35 ton**

Chapter 5

Simulations of the 35 ton prototype

5.1 Determination of interaction times

As outlined at the end of Section 3.5 it is important to know the interaction time of a track when performing calorimetric reconstruction. When performing simulations the simplest interaction time to assign to a reconstructed object is the Monte Carlo truth time of when the particle was created. The creation time can be used as the time taken to travel the relatively modest distances considered in simulations are small when compared to the resolution of the detector (500 ns). When matching a reconstructed object with a GEANT4 particle the particle which contributed the most overall deposited charge to the whole track is chosen. This means that the energy deposited for each hit on the track is broken down into how much each particle contributed to the charge of the individual hit, with the energies summed over all hits. The ability to assign the true interaction times to 3D objects is vital when wanting to benchmark how well other determinations of interaction times perform or to determine the efficiency of the tracking algorithms as described in Section 5.3.

In the 35 ton detector, it was envisioned that there would be at least two ways in which interaction times could be assigned to tracks, one using the external cosmic ray counters and another using reconstructed scintillation light collected by the photon detectors. The cosmic ray counters were used extensively in the 35 ton data, as described in Section 6.4, however in simulation the scintillation light was used as this would have been more powerful during continuous running as not all particles would pass through the counters but one wold expect almost all of them to produce reconstructable scintillation light. The flashes of light are reconstructed using a pre-built library which models the expected number of photoelectrons to be measured on each photon detector given the 3D position of the source of the flash. This

5.1 Determination of interaction times

26

1 library takes into account the expected quantum efficiencies of each photon detector.
 2
 3 When trying to produce an association metric a sample of 10,000 Anti-Muons with a
 4 cosmic-like distribution was used as then there should only be one long track with
 5 which to match one reconstructed flash. A cosmic-like distribution is defined as a set of
 6 particles that have a \cos^2 angular distribution, no minimum or maximum energies and have
 7 a uniform y position with flat distributions of positions in x and z . When this sample was
 8 simulated it was clear that the photon detector reconstruction using the pre-built libraries
 9 worked well as the reconstructed flash source normally lay very close to the track which
 10 caused it. It was found that a calculation of a Point of Closest Approach (PoCA) of the
 11 reconstructed track to the flash source gave an effective metric by which the two could be
 12 combined. Other metrics such as the distance between the flash source and the track centre,
 13 and the perpendicular distance between the flash source and the line joining the start and
 14 end of track were investigated but found to provide less reliable metrics. The latter of these
 15 metrics is less effective because the reconstructed tracks are rarely straight lines, due to
 16 particles scattering as they travel through the LAr and so the perpendicular distance at each
 17 hit must be calculated. A comparison of these metrics is shown in Figure 5.1.

18

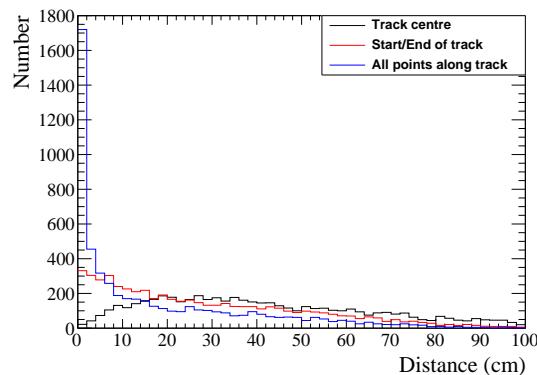


Fig. 5.1 A comparison of metrics to match tracks and flashes in the 35 ton detector using the proximity of tracks and flashes in the yz plane.

19 Another metric by which flashes could be assigned to reconstructed tracks is by utilising
 20 the relationship between the number of measured photoelectrons and the distance from the
 21 APAs at which they were produced. When considering two flashes of scintillation light
 22 that are produced at different distances from the APAs, it would be expected that more
 23 photoelectrons would be collected when the photons were produced closer to the APAs. This
 24 relationship is shown in Figure 5.2 where it can be seen that there is an exponential decay in

5.1 Determination of interaction times

27

the number of photoelectrons which are measured with increasing drift distances. Utilising this relationship, means that the distance from the APAs can be predicted from the number of photoelectrons which are measured. This predicted distance from the APA planes can then be compared to the expected x position of a reconstructed track given the difference in flash time and hit times, this is shown in Figure 5.3. The difference in these two quantities can then be used as the second metric as it gives an indication of how well the properties of a flash match the reconstructed x position of the track. If the predicted and reconstructed x positions are identical then the track and flash are well matched, this corresponds to the collection of points around the $y = x$ line in Figure 5.3.

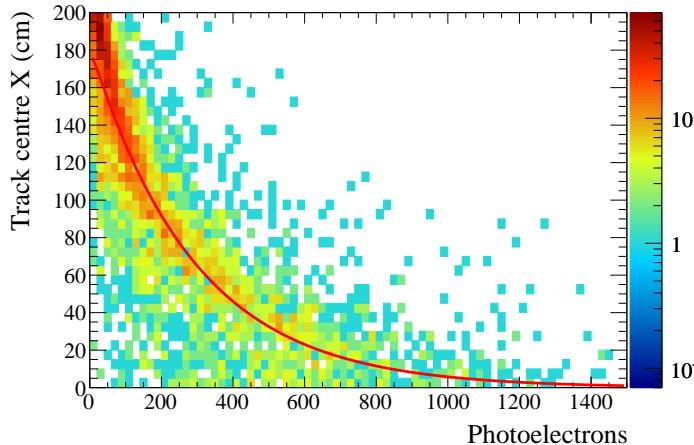


Fig. 5.2 How the number of photoelectrons measured changes with drift distance. The red line corresponds to a parameterisation of the distribution which can be used to predict the x position of a flash given the number of photoelectrons that are collected.

Using these metrics it is possible to attempt to assign reconstructed flashes to reconstructed tracks. Only flashes which are within one drift window of a given track are considered, as flashes outside of this time window cannot have been caused by the reconstructed track. Once flashes are assigned to tracks it is possible to determine how well the matching has performed by comparing the Monte Carlo truth interaction time with the photon detector interaction time. When doing this it is more useful to use a long (16 ms, 32,000 tick) CRY sample as then particles come at random timings as opposed to all at $T = 0$ in the Anti-Muon sample initially considered. This comparison is shown in Figure 5.4, where there is a clear peak at a time difference of 0 ms in the Monte Carlo truth and photon detector interaction times. When zooming in on this peak it can be seen that there is a systematic offset of $0.6 \mu\text{s}$, this is due to an electronics offset applied in the simulation to the photon detector system.

5.1 Determination of interaction times

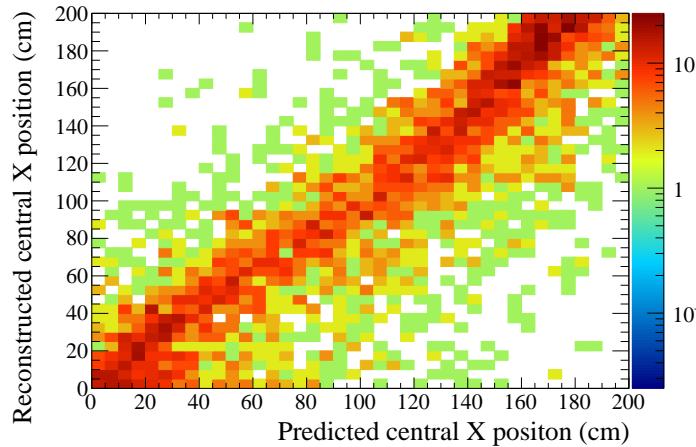
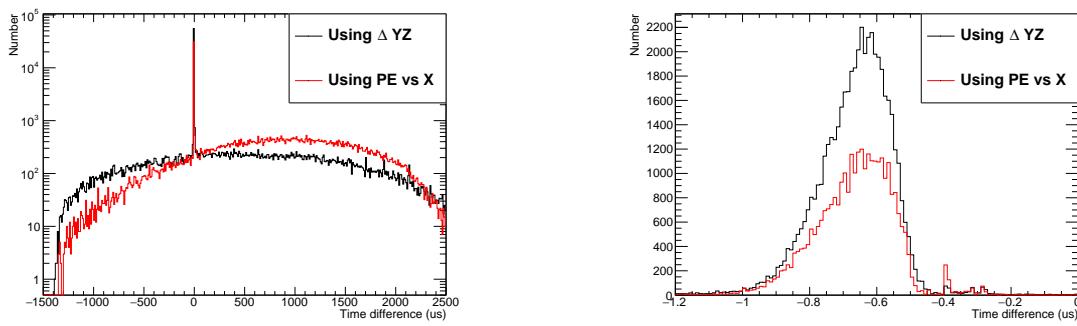


Fig. 5.3 A comparison of the x position predicted using the relationship in Fig 5.2 and the x position predicted by using the difference in flash and hit times.



(a) The difference in interaction times.

(b) Zoomed in at low time differences.

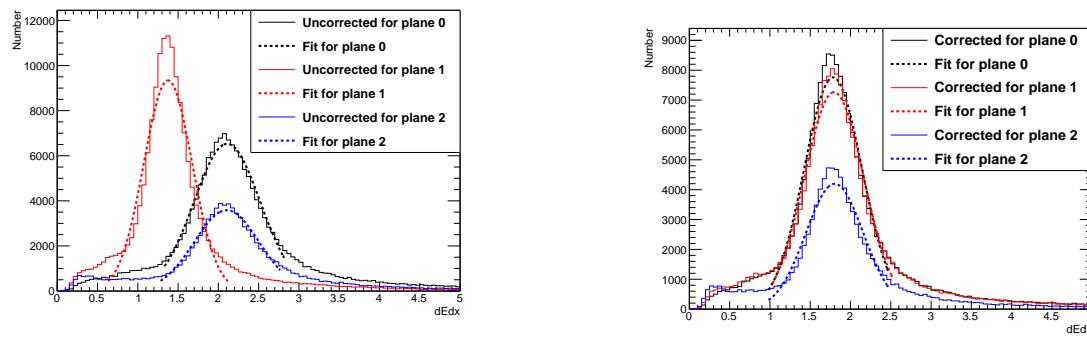
Fig. 5.4 The difference in Monte Carlo interaction times and the predicted interaction times using the photon detectors.

From Figure 5.4 it can clearly be seen that the metric using the proximity of the flash centre to the track trajectory yields the best matches. This is likely caused by the large spread in the number of photoelectrons collected at fixed drift distances, as shown by Figure 5.2. The two metrics can be combined to give a prediction for the interaction time, though given the increased sensitivity from the proximity metric this should be given greater weighting. In physics data the metric using the number of collected photoelectrons is particularly sensitive to the absolute light level in the detector as a high residual light level would reduce the proportional change in the number of photoelectrons collected for increasing drift distances. This metric also relies a sample of tracks with known x positions upon which it can be calibrated which may be difficult to obtain.

5.2 Calibrating calorimetric constants

Having the correct calorimetric responses is vital when trying to calculate $\frac{dE}{dx}$ as the measured change in charge has to be correctly converted to the change in energy. The parameters which need to be tuned in order to ensure that this is done correctly are the $Recomb_A$ and $Recomb_B$ of Equations 3.1c, 3.1d, 3.2a and 3.2c,. These parameters have to be tuned in such a way as to make a known particle energy deposition have the correct $\frac{dE}{dx}$, the easiest deposition to tune against is the Minimally Ionising Particle (MIP) peak which in LAr should have a value of 1.8 MeV cm^{-3} . To do this the sample of 10,000 Anti-Muons made to calibrate the photon detector track/flash assignment will be used as many of these particles will be MIPs.

To select the MIPs in the sample only tracks caused by through-going muons are used. The $\frac{dE}{dx}$ value for all hits in all tracks is then calculated, with the different planes separated out as each one will have its own normalisation factor. A Gaussian distribution is then fitted around the peaks for each of the planes to discern the Most Probable Value (MPV) of $\frac{dE}{dx}$ for that plane. If the MPVs are not equal to 1.8 MeV cm^{-3} then the normalisation factors are scaled through a process of trial and error until the correct MPVs are measured. An example of the tuning being applied is shown in Figure 5.5. Tuning of the calorimetric constants is required whenever the electronics gains or signal shaping functions are changed.



(a) Before a normalisation correction is applied.

(b) After a normalisation correction is applied.

Fig. 5.5 How the $\frac{dE}{dx}$ MPVs change for each plane when a change is made to the electronics gains in the 35 ton. Figure 5.5a shows the MPVs before the constants are corrected, whilst Figure 5.5b shows the MPVs after the constants are retuned.

¹ 5.3 Discerning reconstruction efficiencies

² Knowledge of the strengths and weaknesses of different tracking algorithms is vital when
³ using them for physics analyses, to this end it is useful to develop a metric by which they can
⁴ be compared. In order to do this a series of conditions have to be applied to the reconstructed
⁵ tracks from a large set of simulated particles which are reconstructed using different tracking
⁶ algorithms. It is interesting to observe what the effect of event complexity has on the recon-
⁷ struction algorithms and so efficiencies will be calculated for both the Anti-Muon and CRY
⁸ samples used in Section 5.1.

⁹

¹⁰ The criteria upon which to determine whether a particle is well reconstructed has to be
¹¹ carefully chosen as every definition will have limitations. For example, consider a particle
¹² that travels 100 cm in the active volume of the detector but is reconstructed as 2 separate
¹³ tracks (tracks 1 and 2), with lengths 77 cm and 23 cm respectively. Firstly, should these
¹⁴ tracks be merged, or left separate? If the reconstruction algorithms have found them to be
¹⁵ separate tracks then it is likely that it would be difficult to ascertain that they are from the
¹⁶ same particle in real data, and so in considerations here they are not merged. One metric of
¹⁷ efficiency would be to consider a track well reconstructed if it has a length between 75% and
¹⁸ 125% of the Monte Carlo truth length that the particle traversed in the detector, in which
¹⁹ case track 1 would be considered well reconstructed. Another metric however would be to
²⁰ consider a track well reconstructed if the Monte Carlo truth distance the particle traversed
²¹ in the detector is between 75% and 125% of the reconstructed length, in which case neither
²² track would be considered well matched. Both metrics have used exactly the same tracks
²³ and a seemingly identical method of evaluating whether a track is well reconstructed or not,
²⁴ but have got the opposite results. As such it is wrong to say which consideration gives the
²⁵ correct result, but instead the result of each should be considered equally. It should also be
²⁶ noted that these are just two of a wide range of definitions one could use to quantify a well
²⁷ reconstructed track. In discussions here the former definition of efficiency will be used, such
²⁸ that a track is considered well reconstructed if:

- ²⁹ • Reconstructed track length is more than or equal to 75% of the Monte Carlo track
³⁰ length.
- ³¹ • Reconstructed track length is less than or equal to 125% of the Monte Carlo track
³² length.
- ³³ • Only one reconstructed track can be matched per Monte Carlo particle.

When calculating efficiencies it is important to consider much more than just the ratio of reconstructed to true track length. To this end efficiencies with regards to many aspects of the tracks are calculated:

- Track length
- Energy deposited in the active volume of the detector
- The angle θ of the track
- The angle ϕ of the track

In all efficiency plots the Monte Carlo truth quantity, not the reconstructed quantity is shown so as to reflect how the variations of these quantities affect the reconstruction efficiencies. It is also useful to observe the effect on reconstruction of failed disambiguation and incorrect interaction time determination. To show this, two forms of reconstruction are ran on the particles. One reconstruction path uses no Monte Carlo information and so the interaction time is determined using the photon detectors as described in Section 5.1. The second reconstruction path uses cheated disambiguation and interaction time determination. Cheated disambiguation means using the Monte Carlo truth information of the energy deposition to correctly assign which wire segment the energy was deposited on.

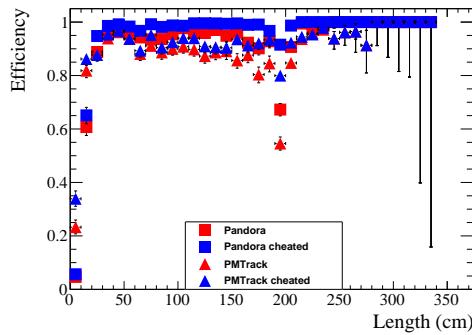
The calculation of reconstruction efficiencies also serves as an effective method upon which reconstruction algorithms can be further developed as it identifies aspects which do not work as expected. For example when the efficiencies for the CRY sample were initially calculated they were significantly lower than for the Anti-Muon sample, but only when disambiguation was not cheated. It transpired that this was because the disambiguation was only selecting the largest collection of hits on each plane for each TPC. This is not a problem when only 1 particle is simulated and will reduce the number of noise hits but in a CRY sample of 16 ms there will almost certainly be multiple particles in each TPC. Removing the hits from all but one of these multiple particles will cause them to have no reconstructed track, and thus cause the efficiency to drop significantly. Upon making the disambiguation algorithm no longer have this restriction the reconstruction efficiencies of the Anti-Muon and CRY samples were observed to become much more similar.

The reconstruction efficiencies given the current state of the most commonly used reconstruction algorithms are shown in Figures 5.6, 5.7, 5.8, 5.9 and 5.10. Efficiencies are shown for both the Anti-Muon and CRY samples, where it can be seen that the efficiency tends to be lower for the CRY sample. It is thought that this is due to the more complex event structure,

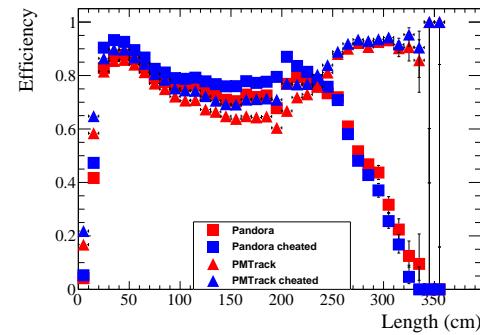
5.3 Discerning reconstruction efficiencies

as particles will have large interaction times and particles which have similar interaction times may cross causing reconstruction errors. The reconstruction efficiencies for the CRY sample are more realistic as events will rarely be isolated in the detector due to the large flux of cosmic particles on the Earth's surface.

5

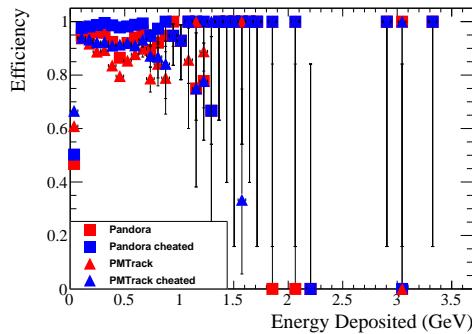


(a) Reconstruction efficiencies for an Anti-Muon sample.

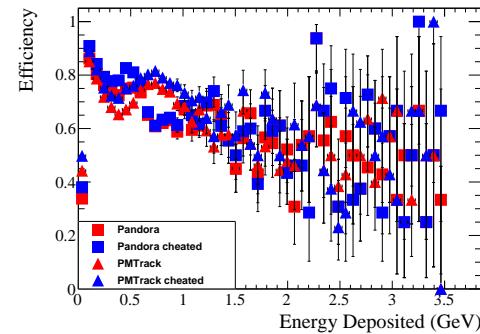


(b) Reconstruction efficiencies for a CRY sample.

Fig. 5.6 The reconstruction efficiencies for simulated events as a function of Monte Carlo truth track length. The efficiencies are shown for non-cheated reconstruction (red blocks) and cheated reconstruction (blue blocks) for both Pandora (square blocks) and PMTrack (triangle blocks).



(a) Reconstruction efficiencies for an Anti-Muon sample.

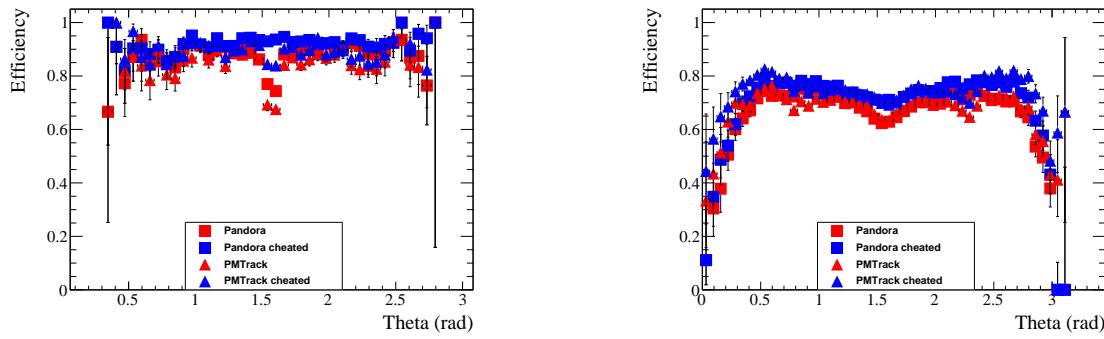


(b) Reconstruction efficiencies for a CRY sample.

Fig. 5.7 The reconstruction efficiencies for simulated events as a function of Monte Carlo truth deposited energy. The efficiencies are shown for non-cheated reconstruction (red blocks) and cheated reconstruction (blue blocks) for both Pandora (square blocks) and PMTrack (triangle blocks).

A striking feature of Figure 5.6 is the rapid decrease in reconstructed efficiency for the CRY sample for track lengths above 250 cm when using Pandora. The cause of this is that

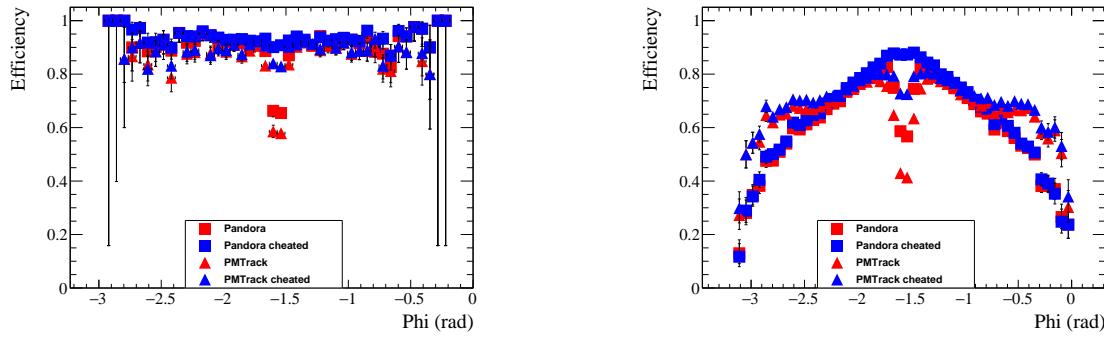
5.3 Discerning reconstruction efficiencies



(a) Reconstruction efficiencies for an Anti-Muon sample.

(b) Reconstruction efficiencies for a CRY sample.

Fig. 5.8 The reconstruction efficiencies for simulated events as a function of Monte Carlo truth track angle in theta. The efficiencies are shown for non-cheated reconstruction (red blocks) and cheated reconstruction (blue blocks) for both Pandora (square blocks) and PMTrack (triangle blocks).



(a) Reconstruction efficiencies for an Anti-Muon sample.

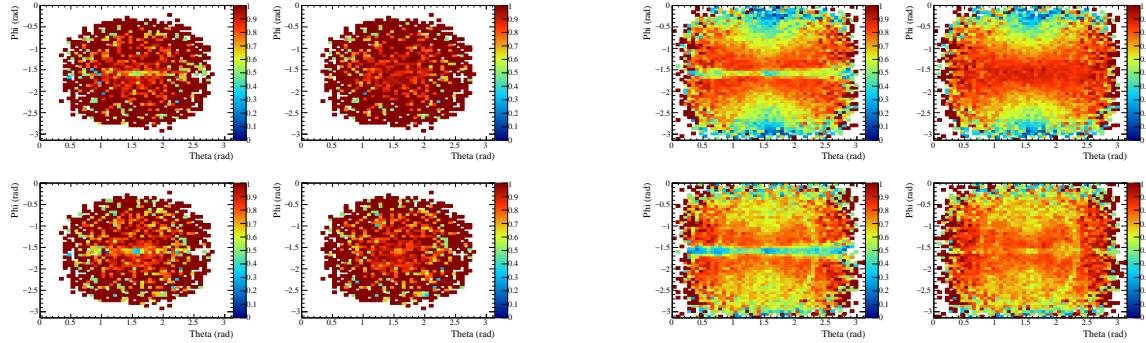
(b) Reconstruction efficiencies for a CRY sample.

Fig. 5.9 The reconstruction efficiencies for simulated events as a function of Monte Carlo truth track angle in phi. The efficiencies are shown for non-cheated reconstruction (red blocks) and cheated reconstruction (blue blocks) for both Pandora (square blocks) and PMTrack (triangle blocks).

tracks are reconstructed separately in the long and short drift volumes before being merged when they are found to be co-linear in the yz plane. This is not a problem in the Anti-Muon sample as the x position of the hits calculated using Equation 5.1a will be correct. However, when the same is done for hits in the CRY sample using particles with large interaction times the x positions will have offsets proportional to the interaction time unless the hit time is corrected by Equation 5.1b. The result of this is that merged tracks can have discontinuities in their x coordinates of more than 20 m. As the interaction time of the track is calculated using

1
2
3
4
5
6
7

5.3 Discerning reconstruction efficiencies



(a) Reconstruction efficiencies for an Anti-Muon sample.

(b) Reconstruction efficiencies for a CRY sample.

Fig. 5.10 The reconstruction efficiencies for simulated events as a function of Monte Carlo truth track angle in theta and phi. The efficiencies are shown for non-cheated reconstruction (plots on the left) and cheated reconstruction (plots on the right) for both Pandora (plots on the top) and PMTrack (plots on the bottom).

the output of the tracking algorithms it is not possible to directly correct for the interaction time at present. It is however possible to subtract this jump in x position from the track length quantity which is calculated when the stitched track is stored in the event, this will give the correct track length though the user will still have to correct individual hit positions in later analyses using the calculated interaction time. This is what is done by PMTrack, hence it not exhibiting this rapid decrease in reconstruction efficiency for long tracks.

7

$$x_{Hit} = T_{Hit} \times v_{Drift} \quad (5.1a)$$

$$T_{Hit} = T_{Measured} - T_{Interaction} \quad (5.1b)$$

It is clear from Figure 5.6 that tracks of lengths less than 30 cm are poorly reconstructed. The very low efficiency for tracks with lengths less than 10 cm can be partially attributed to particles with lengths of less than 1 cm as these particles, which represent 30% of the particles with lengths below 10 cm, are too short to be reconstructed using the current reconstruction process. These particles will need to be reconstructed when looking for supernovae bursts though special algorithms will be written to do this, as the traditional hit finding and clustering algorithms may discard them due to the isolated nature of the hits. Another issue is that the low energies of these particles may mean that the signals that they produce are below threshold and so will not even be reconstructed, or if hits are reconstructed they may be too close to a more energetic track and get absorbed into them. The reconstruction of

tracks is affected by the number of wires which they cross, though this should not matter for particles with lengths of more than 5 cm as they will have crossed roughly 10 wires in each plane which should produce enough unique hits for a cluster to be reliably constructed. This can be seen to be the case for PMTrack when considering the Anti-Muon sample, as the efficiency for track lengths between 10 and 20 cm is roughly the same as that for track lengths between 20 and 30 cm, however when considering the CRY sample there is still a significant decrease in efficiency. This is attributed to the more complex event structure in the CRY sample, where secondary particles are produced which are mis-reconstructed even though they travel reasonable distances. This is seen to be the case as when only primary muons in the CRY sample are considered the reconstruction efficiency is seen to be the same as that for the Anti-Muon sample.

The trend of increasing efficiency for longer track lengths from Figure 5.6 can also be seen in Figure 5.7 as the energy deposited increases. This is because particles which deposit more energy will tend to have travelled further in the detector. The amount of energy that particles deposit is limited by the size of the detector though as particles with an energy of more than 1 GeV are energetic enough to through-going MIPS. This results in few particles depositing more than 1 GeV in the detector causing the uncertainty in the reconstruction efficiency to increase above this energy. The increased statistics at high deposited energies in Figure 5.7b is due to the larger number of muons generated in the CRY which create large electromagnetic showers when they enter the LAr.

It is also interesting to note the pronounced decreases in reconstruction efficiencies for particular angles shown in Figure 5.8 and Figure 5.9. The decrease in efficiency at $\phi = \frac{\pi}{2}$ can be attributed to the drop in efficiency for tracks of 200 cm, as this corresponds to the vertical height of the detector meaning that few collection wires are hit and so determining the triple points needed by the disambiguation are difficult to find. This is verified by the large increase in efficiency achieved by cheating the disambiguation. Similarly the decrease in efficiency at $\theta = \frac{\pi}{2}$ can be attributed to particles which are perpendicular to the collection wires resulting in few collection wires being hit.

The information from Figures 5.8 and 5.9 is combined in Figure 5.10 where the sharp drops in efficiency for the CRY sample are particularly visible. The effect of cheated disambiguation is clear in Figure 5.10b where the dip in efficiency as a function of θ at fixed $\phi = \frac{\pi}{2}$ is completely removed. The same is not true for the dip in efficiency as a function ϕ at fixed $\theta = \frac{\pi}{2}$, though the reduction in efficiency was not uniform or as severe across all values

5.4 Performing particle identification

36

¹ of θ as it remains mainly confined to values of θ close to 0 or π , particularly when using
² Pandora . The observation that a significant improvement in the quality of reconstruction can
³ be made in improving the disambiguation is a driving force in the wire pitches being 36° for
⁴ the DUNE FD as opposed $45 \pm 0.7^\circ$ in the 35 ton, because as discussed in Section 3.5 the
⁵ shallower wire pitch makes disambiguation easier. Though disambiguation will be easier in
⁶ the different geometry, further efforts to improve disambiguation are still required, as are
⁷ continued efforts to reconstruct the shortest tracks.

⁸

⁹ 5.4 Performing particle identification

¹⁰ Being able to perform reliable particle identification (PID) is a key requirement for the DUNE
¹¹ experiment, and so efforts have been made to establish a metric by which this can be achieved.
¹² The predominant method of performing PID in LAr is to use the relationship between $\frac{dE}{dx}$
¹³ and the residual range of the track, defined as being the distance between a point on the track
¹⁴ and the stopping point of the track. This relationship is observed to be dependent on particle
¹⁵ mass and is quantified by the Bethe-Bloch equation [21] [22] which is shown in Figure 5.11.
¹⁶ The sharp increase in energy loss per unit length can be seen to occur at different momenta
¹⁷ for different particle masses meaning that the peak value of $\frac{dE}{dx}$ can change significantly. One
¹⁸ example of a large change in the peak value of $\frac{dE}{dx}$ can be seen by comparing muons and
¹⁹ protons, whilst muons and pions are very similar.

²⁰

²¹ The particle mass dependence can be seen by plotting the $\frac{dE}{dx}$ against the residual range of
²² the particle on a log-log plot, as shown in Figure 5.12a. A power law dependence is found to
²³ describe the relationship [19], as shown in Equation 5.2. The dependence on b is found to be
²⁴ weak, and so can be set to -0.42 for all particle masses. This means that the main discriminant
²⁵ used is the A parameter, which has a strong dependence on particle mass. The values for
²⁶ A and b calculated from Figure 5.12a are shown in Table 5.1. It is found that the error in-
²⁷ troduced by fixing the b parameter is small compared to the error from ionisation fluctuations.

²⁸

$$\frac{dE}{dx}_{calo} = AR^b \quad (5.2)$$

$$A_i = \left(\frac{dE}{dx}_{calo} \right)_i \times R_i^{0.42} \quad (5.3)$$

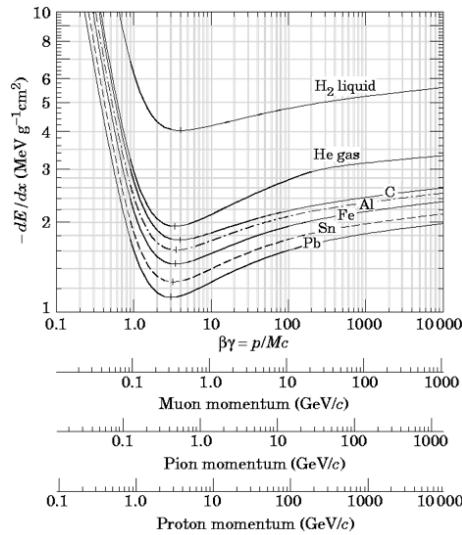


Fig. 5.11 The Bethe-Bloch equation describes energy loss per unit length as a function of energy in different mediums. The energy losses expected for different particle types is shown in different mediums. Liquid Argon with a density of 1.4 g cm^{-3} has a density slightly less than that of Carbon at 1.8 g cm^{-3} .

Once the b parameter is set to be constant for all particle types it is possible to calculate a value for the A parameter for each hit on the track using Equation 5.3, where R_i is the residual range of the track at that point. The particle type discriminant, called PIDA, can then be calculated for a track by finding the average value of A_i for the track. As the particle mass dependant increase in $\frac{dE}{dx}$ only occurs near the end of the track, the PIDA variable can only be calculated for particles which stop in the detector as all other particles will have MIP-like $\frac{dE}{dx}$ distributions and so cannot be identified in this way. As shown by the plotted range of Figure 5.12a the average value of A is normally calculated for the last 30 cm of the track.

The PIDA method was tested in [19], where the PIDA values were calculated for Monte Carlo particles which stopped in the detector using truth information over the last 30 cm of the particle lengths. This is shown in Figure 5.12b, where a clear separation can be seen between the peaks for Muons, Pions, Kaons and Protons. Though the Muon and Pion peaks

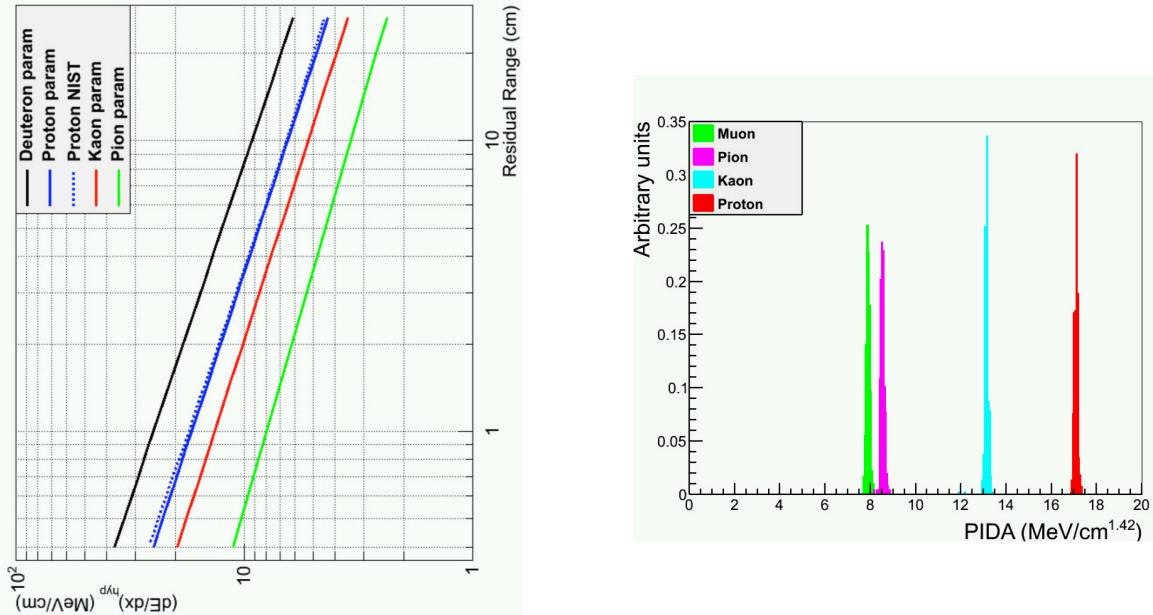
5.4 Performing particle identification

Table 5.1 Stopping power parameterization for various particle types in LAr [19].

Particle	$A \text{ MeV cm}^{-(1-b)}$	b
Pion	8	-0.37
Kaon	14	-0.41
Proton	17	-0.42
Deuteron	25	-0.43

- 1 are relatively close together they can still be resolved in the plot due to little overlap. It is
 2 interesting to note how tight the PIDA distributions found in the paper are, which allows
 3 the different particles types to cleanly separated in the truth study. The author notes that an
 4 incorrect tuning of the recombination effects will cause the distributions to become broader,
 5 and an incorrect calibration of the detector will introduce a systematic shift in the expected
 6 values of PIDA.

7



(a) Stopping power for different particle masses.

(b) Distribution of PIDA values for different particle masses.

Fig. 5.12 Defining the PIDA metric for particle identification and testing it on a Monte Carlo sample using truth information.

From Figure 5.12 it can be seen that the most distinct PIDA distributions are that of muons and protons, these are also two of the most common particle types in cosmic rays. For these reasons particle identification using the PIDA variable will be attempted on simulations of the 35 ton. As outlined in Sections 5.1 and 5.2 in order to do this the interaction times of particles have to be well known and the calibration constants must be tuned so as to ensure that the effects of recombination are properly accounted for. It is also useful to use the information found in Section 5.3 about the efficiency with which tracks are reconstructed. In this regard it is useful to produce additional figures showing the reconstruction efficiencies of protons in the CRY sample, these are shown in Figure 5.13.

Figure 5.13 shows that the average reconstruction efficiency for PMTrack is higher than that for Pandora when considering protons, as the efficiency for the former is roughly 10% higher for all angles as shown in Figure 5.13c, though it is much lower than the overall efficiency seen in Figure 5.8b. From Figure 5.13a it is evident that the efficiency for protons with track lengths of more than 10 cm is similar to that of the overall efficiency for the CRY sample, but the efficiency for the shortest tracks is significantly lower than that of the whole CRY sample. A review of the true path lengths of the simulated particles shows that 60% of the protons have path lengths of less than 1 cm and that none of these particles were reconstructed, it is this large number of very short particles which causes the overall reconstruction to be relatively low. When a minimum path length of 1 cm (10 cm) is required the reconstruction efficiency rises to 37% (58%), so when the shortest tracks are not counted the reconstruction performs reasonably well.

It is also useful to produce samples where the primary particle is a single muon or proton located in the active volume of the detector. This allows for a sample of isolated tracks to be made upon which the capabilities of the PIDA metric can be tested. It also allows the reconstruction efficiency to be found for particles in isolation. The properties of the generated particles are illustrated below in Table 5.2. The values of the simulated quantities were found by changing the given parameters by an amount taken from a random sampling of a Gaussian distribution of width equal to the error listed. These simulation parameters were chosen to produce samples which would contain both exiting and stopping particles whilst generating the particles in the LAr would ensure that there should always be a reconstructable track in the detector.

The reconstruction efficiencies when using the PMTrack reconstruction method are shown for the simulated particles in Figure 5.14. It should be noted that truth particles with track

5.4 Performing particle identification

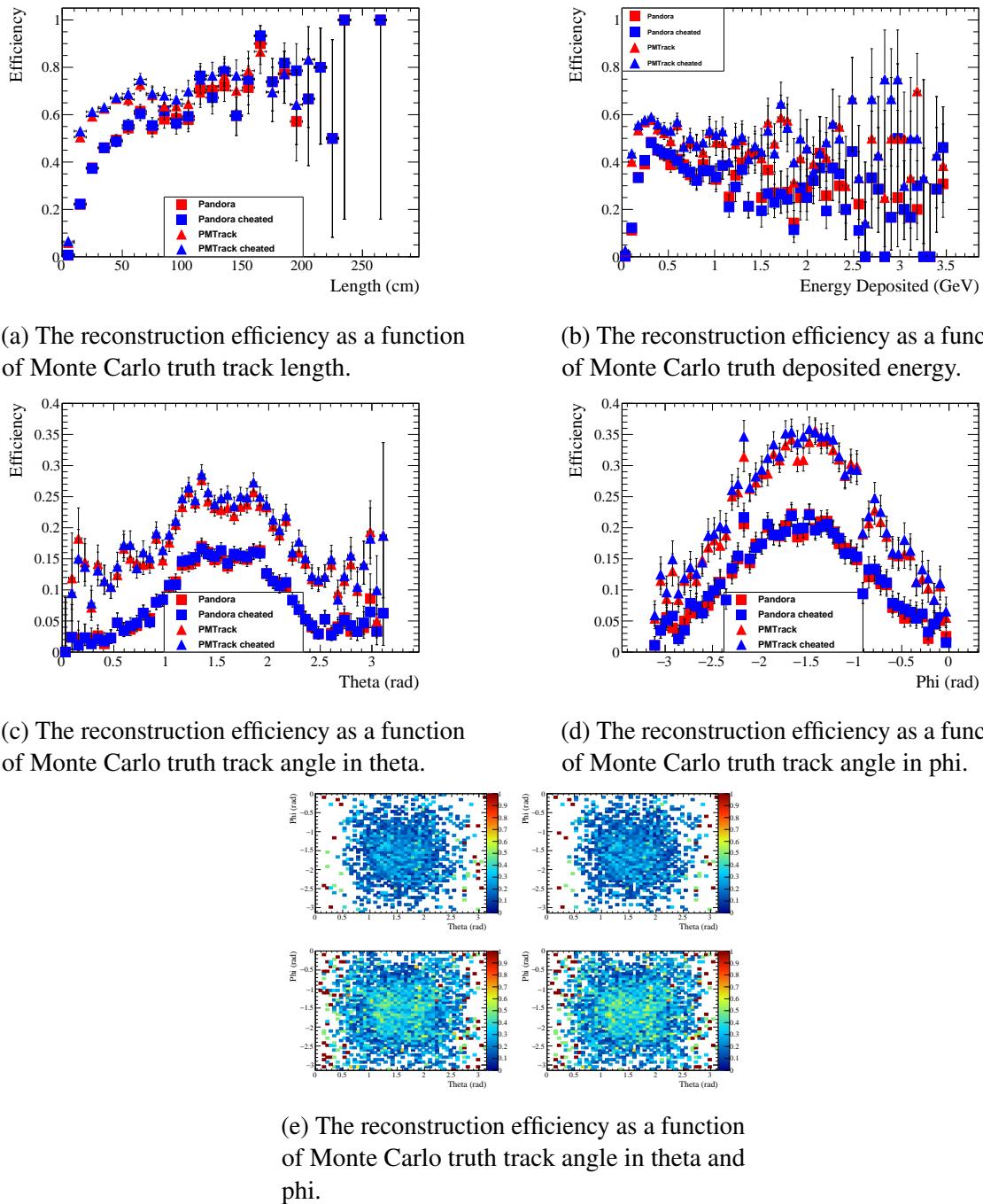


Fig. 5.13 The reconstruction efficiencies for protons in a sample generated using CRY. The efficiencies are shown for non-cheated reconstruction (red blocks) and cheated reconstruction (blue blocks) for both Pandora (square blocks) and PMTrack (triangle blocks).

¹ lengths of less than 1 cm have been excluded from these plots which is why the angular
² reconstruction efficiencies for protons in Figures 5.14c and 5.14d, are higher than those seen

5.4 Performing particle identification

41

Table 5.2 The properties of initial particles simulated in the muon and proton samples.

	Muon properties	Proton properties
Initial position (cm)	(100 ± 50, 0 ± 30, 80 ± 20)	(100 ± 50, 0 ± 30, 80 ± 20)
Initial momentum (GeV)	0.3 ± 0.1	0.8 ± 0.5
Initial θ_{XZ} (°)	0 ± 180	0 ± 180
Initial θ_{YZ} (°)	-45 ± 45	-45 ± 45

in Figures 5.13c and 5.13d. This was done as due to how the initial momenta and positions are sampled many of the primary simulated particles may travel very short distances that are contained in spaces between TPCs and including these particles would artificially reduce the efficiency presented. After discounting these very short particles the efficiencies generally follow similar patterns observed in the earlier efficiency plots, though there is a decrease in efficiencies for the longest track lengths which is not observed in other samples. This is attributed to the initial positions for the particles being within the detector volume, as this means that any particle travelling over 100 cm would have a very peculiar trajectory as the edge of the detector should never be more than 100 cm away from the starting position. The only exception to this is if a particle travelled along the x axis to the other end of the detector, which as discussed earlier is a very problematic orientation to reconstruct as all of the charge would be deposited over a large range of time on very few collection plane wires.

As the increase in $\frac{dE}{dx}$ is only visible when the particle stops in the detector it is necessary to remove exiting particles from the sample by applying a fiducial cut on the end point of the reconstructed track. It is important to only place this on the end point of the track, as one does not want to remove particles which enter the detector and then stop. When calorimetry is performed the end point of the track is determined using, among other metrics, the increase in $\frac{dE}{dx}$ and so the residual range of the track (a stored data member of the track object) should always refer to the distance to the end of the particles trajectory. For this study a fiducial cut of 5 cm is used, meaning that any track with hits within 5 cm of the edge of the detector volume is discarded and counted as an exiting particle. This should mean that very few tracks due to exiting particles are identified as stopping in the detector as it would require that a large section of the track would have to un-reconstructed. This will mean that some stopping particles are incorrectly assigned as exiting particles causing the identification efficiency to drop, but it is necessary to ensure that exiting particles are not included in the final distributions. A further cut that is applied is the requirement that there are a minimum of 5 continuous collection plane hits, this is to ensure that an adequate number of points are taken upon which to find an average value of PIDA for the track. Similar cuts are described

5.4 Performing particle identification

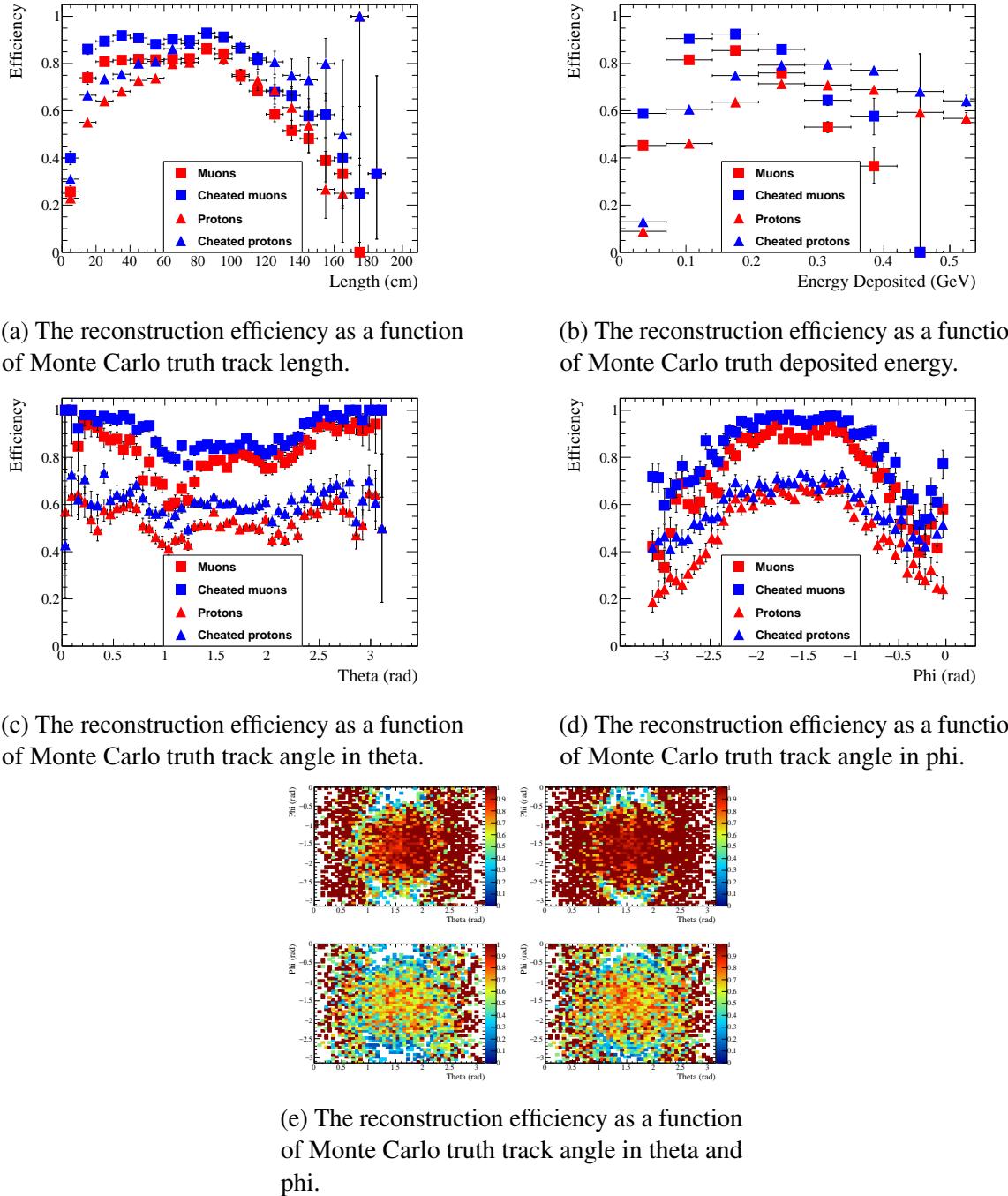
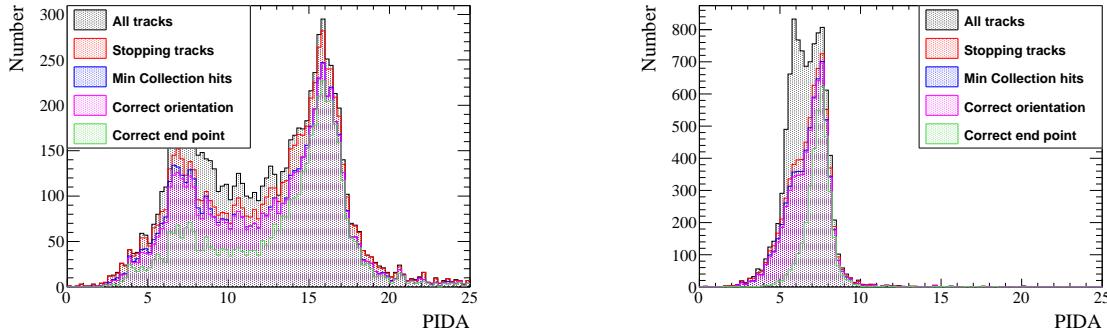


Fig. 5.14 The reconstruction efficiencies for single muons and protons in the 35 ton. The efficiencies are shown for non-cheated reconstruction (red blocks) and cheated reconstruction (blue blocks) for both muons (square blocks) and protons (triangle blocks).

¹ in [19], and the resulting distributions of PIDA values for the single proton and muon samples

5.4 Performing particle identification

are shown in Figure 5.15.



(a) The PIDA values calculated for the single proton sample.

(b) The PIDA values calculated for the single muon sample.

Fig. 5.15 The calculated PIDA values for single muons and protons in the 35 ton. A series of criteria designed to select only tracks due to stopping particles which have a required number of collection plane hits is applied. The tracks are then further refined using truth information such as the true end point of the particle.

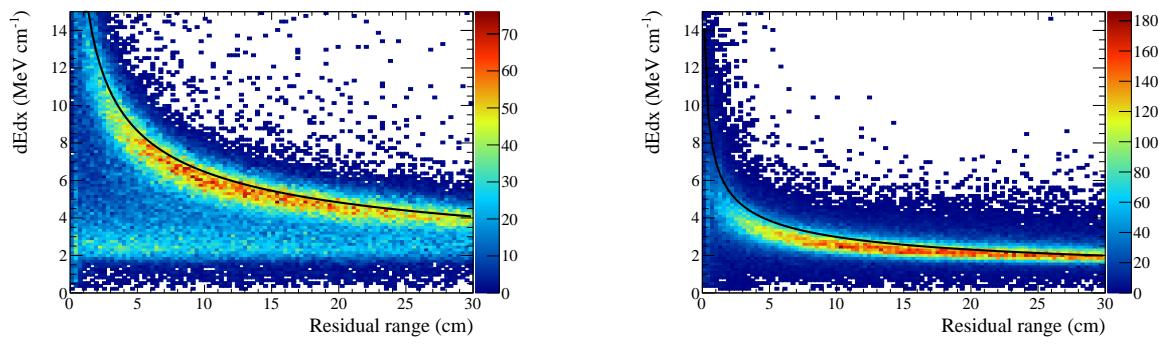
As can be seen from Figure 5.15 using truth information can make the distributions much cleaner, particularly when discounting particles for which the reconstruction algorithms do not track to their end point. A track is identified as having a correct end point if the reconstructed end point is within 2.5 cm of the true end point of the particle. It is reassuring to see that few tracks are reconstructed backwards, as if this were not the case then performing particle identification would be very difficult as it would indicate that the calorimetry and tracking algorithms are not performing well. Improvements can still be made though, as both plots in Figure 5.15 contain tracks which do not have the final energy depositions. This can be seen as when tracks which do not match with the true end points of the particles are removed the low tails of the PIDA distributions are significantly reduced. It is observed that the PIDA distributions are cleaner when information from all three wire planes are used as opposed to only using the collection plane and so that is what is presented here. This shows how important it is to calibrate the electronics responses of all three wire planes and how additional wire planes can improve calorimetry as well as the accuracy of reconstruction algorithms.

The relationship between the $\frac{dE}{dx}$ and residual range of a track is shown in Figure 5.16 for both protons and muons. The much steeper increase in $\frac{dE}{dx}$ at low residual range for protons compared to muons is clearly visible when comparing Figures 5.16a and 5.16b. The contamination in the proton sample at low PIDA can be seen in Figure 5.16a where

5.4 Performing particle identification

1 there is a clear sample of tracks for which the $\frac{dE}{dx}$ does not increase for low residual ranges.
 2 These plots are filled after tracks which do not correlate to the ends of the true trajectories
 3 are removed, and so the tail of low $\frac{dE}{dx}$ values is due to particles for which the simulated
 4 detector did not find increased energy depositions as the particle stopped. It is interesting
 5 to note that when a simple version of PIDA is calculated using the MC truth energy de-
 6 posits, shown in Figure 5.17, these particles are also found to have low PIDA values. It is
 7 therefore possible that at least some of these protons do not in fact stop, but interact inelasti-
 8 cally when they still have a significant amount of kinetic energy meaning that GEANT4 will
 9 create a new particle and the tracking algorithms are creating a new track after this interaction.

10



(a) The $\frac{dE}{dx}$ versus residual range plot for the single proton sample.

(b) The $\frac{dE}{dx}$ versus residual range plot for the single muon sample.

Fig. 5.16 The measured relationship between $\frac{dE}{dx}$ and residual range for single muons and protons in the 35 ton. The plots are made after applying all of the cuts outlined in Figure 5.15, meaning that the MIP peaks have been suppressed using truth information.

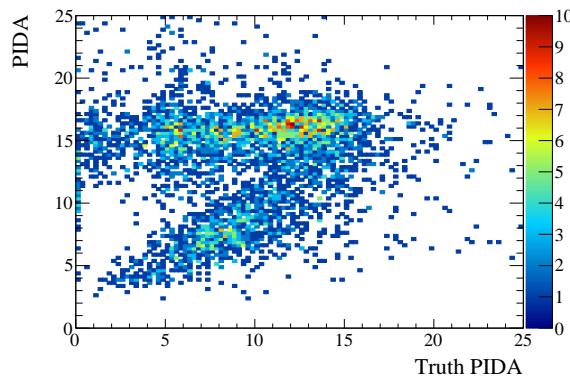


Fig. 5.17 A comparison between PIDA values calculated using truth and reconstructed information

It is useful to summarise the information shown in Figure 5.15 in a table so that an efficiency of identifying stopping particles can be found. This is shown in Table 5.3 for protons, and in Table 5.4 for muons. The efficiency shown in these tables is defined as the number of tracks in the PIDA range divided by the total number of stopping particles, this is why the 'efficiency' is more than 100% for the number of reconstructed tracks in Table 5.4. The purity shown in these tables is defined as the percentage of tracks in the PIDA range which are associated with particles which actually stop in the detector. As many of the reconstructed tracks shown in Table 5.4 are not due to stopping particles the purity is low. The PIDA ranges referred to are 14-18 and 5-9 for the protons and muons respectively, as these ranges cover the peaks of the distributions shown in Figure 5.16 and are centered on the peaks in Figure 5.12b.

As can be seen in Table 5.3 the efficiency upon which protons can be identified does not change significantly as the sequential criteria are applied, but as shown in Figure 5.15a the low PIDA peak decreases significantly. The same cannot be said for the muon sample however, as when the criteria that the tracking end point matches the true end point is applied a significant section of the tail within the PIDA range is removed. The resulting distribution is more similar to the distribution shown in Figure 5.12b though, showing that it preserves the stopping tracks which are reconstructed best. The cut to remove tracks that do not have the correct end points reduces both sets of efficiencies, but if all the tracks were reconstructed with the correct end points then one can imagine that the number of tracks within the PIDA ranges would increase and the distributions would become more symmetrical as shown in Figure 5.15b. Both tables also exhibit high purities which shows that the fiducial cut designed to removing exiting particles is effective, with only 2 exiting protons being mis-identified in the proton sample.

From Table 5.3 it can be seen that there are more stopping protons than primary protons as only 10,000 primary protons were generated. The effectiveness of the PIDA algorithm at identifying only primary protons is shown in Table 5.5. Comparing both tables it can be seen that the efficiency with which the primary protons can be identified is larger than the secondary protons as the efficiencies shown in Table 5.3 are lower than those in Table 5.5. It is thought that this is due to the low reconstruction efficiency for the very shortest tracks which many of the secondary protons have, as discussed in Section 5.3. A similar table is not produced for primary muons as there were no secondary muons produced in the muon sample, and so Table 5.4 is itself the efficiency with which the primary muons can be identified.

5.4 Performing particle identification

Table 5.3 A summary of the PIDA values calculated for the proton sample as sequential cuts are applied.

Applied cut	Proton sample			
	Tracks	In PIDA range	Efficiency	Purity
Total stopping particles	13295			
Reconstructed tracks	8761	3009	22.6%	98.7%
Survives 5 cm fiducial cut	7552	2894	21.8%	99.9%
Minimum of 10 collection plane hits	6186	2507	18.9%	99.9%
Correct track orientation	6022	2491	18.7%	99.9%
Correct tracking end point	4432	2288	17.2%	100%

Table 5.4 A summary of the PIDA values calculated for the proton sample as sequential cuts are applied.

Applied cut	Muon sample			
	Tracks	In PIDA range	Efficiency	Purity
Total stopping particles	6880			
Reconstructed tracks	9883	8907	129%	67.4%
Survives 5 cm fiducial cut	7126	6259	90.9%	90.2%
Minimum of 10 collection plane hits	6580	5876	85.4%	89.9%
Correct track orientation	6436	5767	83.8%	90.1%
Correct tracking end point	3676	3555	51.7%	100%

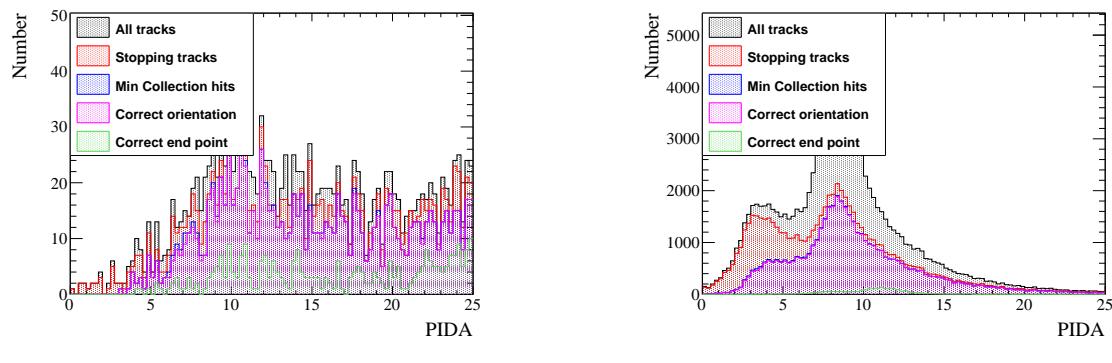
Table 5.5 A summary of the PIDA values calculated for the primary particles in the proton sample as sequential cuts are applied.

Applied cut	Proton sample			
	Tracks	In PIDA range	Efficiency	Purity
Total stopping particles	7798			
Reconstructed tracks	5920	1937	24.8%	98.9%
Survives 5 cm fiducial cut	5044	1878	24.1%	99.9%
Minimum of 10 collection plane hits	4485	1711	21.9%	99.9%
Correct track orientation	4363	1707	21.9%	99.9%
Correct tracking end point	3122	1565	20.1%	100%

Upon verifying that the PIDA metric can reliably determine particle type when they are simulated in isolation, the next step is to observe the accuracy upon which particles can be identified in a CRY sample. The sample used here differs from the CRY sample used earlier in that only events which contain a proton in the detector are reconstructed, this is done to reduce simulation time and storage space as this cut will still provide a substantial number of muons whilst ensuring that a large proton sample can be reconstructed. The process of calculating PIDA values for the tracks is identical in all samples, though as discussed in Section 5.3 the much more complicated event structure in the CRY sample affects the reconstruction efficiency and so will likely also affect the accuracy of the calorimetry. The calorimetry will be affected in two ways, firstly the reduced performance of the reconstruction algorithms will mean that some particles are not reconstructed at all, whilst those that are reconstructed may be more likely to have missing hits meaning that the end points may be less well reconstructed. This will cause the tail of low $\frac{dE}{dx}$ values seen in Figure 5.16a to be more pronounced. Secondly, as shown in Figure 5.4 though the photon detector time determination is very accurate for a large number of tracks it is also incorrect for a number of tracks, this will cause the recombination correction to be miscalculated which will in turn increase the calculated $\frac{dE}{dx}$ and hence PIDA values.

The PIDA values calculated for protons and muons in the CRY sample are shown in Figure 5.18. As can be seen from Figure 5.18b there is a tail of very high PIDA value muon tracks which contaminate the proton PIDA region of interest (ROI). This causes a serious problem when trying to identify protons from a cosmic sample as the number of muons present is significantly larger than the number of protons. The result of this will be a sample of tracks which will not be very pure, and so further cuts will have to be developed to enhance the purity of this sample whilst not reducing the efficiency upon which proton tracks are identified.

5.4 Performing particle identification



(a) The PIDA values calculated for protons.

(b) The PIDA values calculated for muons.

Fig. 5.18 The calculated PIDA values for muons and protons in a CRY sample through the 35 ton. A series of criteria designed to select only tracks due to stopping particles which have a required number of collection plane hits is applied. The tracks are then further refined using truth information such as the true end point of the particle.

Chapter 6

The 35 ton data sample

The data taking period for the 35 ton prototype was from November 2015 until March 2016. This included an extensive commissioning period before the detector was filled with LAr and the electric field was turned on. During this time many of the features of the data discussed below were first noticed and attempts to rectify these were pursued. A long commissioning period was also required because many of the DAQ sub-systems were still under active development in November.

A total of 22 days worth of data was collected with the electric field set at 250 V cm^{-1} , the breakdown of when these periods occurred is shown in Figure 6.1. It is clear that the analysable data is interspersed with data where the electric field was not turned on, this is both due to extenuating circumstances such as a site wide power outage in early March and a dedicated two week noise hunting exercise in February. The physics data taking period ended at 3am on 19th March 2016 when a filtration pump broke causing an unrecoverable loss of purity as air was pumped into the detector. Following this studies to understand the electronics noise and to test the high voltage systems continued but it was deemed too costly to acquire any more physics data. During this time the electric field was raised to the nominal value of 500 V cm^{-1} , and some of the causes of the higher than expected noise levels were discerned.

6.1 Organisation of the data structure

As previously noted the 35 ton consisted of three detector sub-systems: RCEs collecting TPC data, SSPs collecting photon detector data, and CRCs tagging cosmic rays. The DAQ combined these three data streams into synchronous events in time and saved them as LArSoft

6.1 Organisation of the data structure

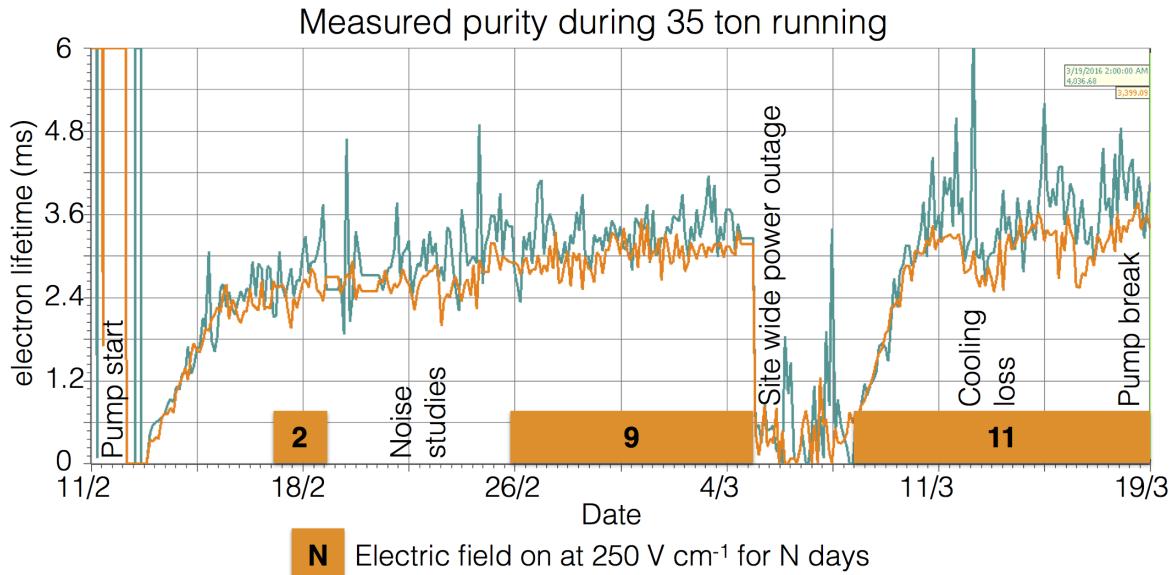


Fig. 6.1 Timeline showing the data collected during the 35 ton Phase II run once the purification pumps were turned on.

¹ data objects. These data objects would later have to be converted to the offline data products
² which the reconstruction tools developed on simulation used, this is discussed in Section 6.2.
³ This section describes the structure of the data objects in the raw form.

⁴
⁵ During operations the DAQ was configured to maximise data throughput and physics
⁶ potential. This meant recording different lengths of times for each of the three sub-systems
⁷ as the data volumes and length of physics information were significantly different. For
⁸ example due to the emission of prompt light the physics information from the SSPs is of a
⁹ much shorter length of time than from the RCEs where data has to be recorded whilst the
¹⁰ electrons drift through the LAr. During the running period the recorded data was triggered
¹¹ by through-going muons which produced coincidences on the CRCs on opposite side of
¹² the cryostat. A coincidence is defined as two CRC modules recording a hit within 30??? ns.
¹³ The system used to collect the CRC data was also responsible for generating the triggers
¹⁴ and so this meant that the trigger rate could be suppressed to approximately 1 Hz by only
¹⁵ producing triggers every N times a coincidence occurred, where N was a tuneable variable.
¹⁶ A trigger rate of 1 Hz was used as the maximum speed at which data could be written to disk
¹⁷ was approximately 60 MB s⁻¹, which is roughly equal to the size of each triggered event
¹⁸ when the entire detector is read-out in the configuration discussed below. The rate at which
¹⁹ events were recorded could have been increased if zero-suppression of the TPC data had

 6.1 Organisation of the data structure

51

been used, however the noise level meant that this was not feasible.

With an electric field of 250 V cm^{-1} and a drift of 2.25 m, the drift time for electrons at the long drift CPA was roughly 2.6 ms or 5200 ticks (where 1 tick is 500 ns). It was decided that in order for a track causing a counter coincidence to be separated from other tracks it was necessary to have roughly one drift window both before and after the drift window around the coincidence, meaning that data was recorded for 7.5 ms or 15,000 ticks around each coincidence. Only the prompt light from through-going particles was collected and so only $200 \mu\text{s}$ of SSP data was recorded for each event. The CRCs produced the least volume of data and so were able to be read out constantly.

As the run mode required accessing buffered data, it had to be discretised inside the components before being sent to the event builders in the DAQ. In the discussion of how this worked, focus will be given on the RCE data where some new terms need to be introduced. The smallest unit of data, called a nanoslice, is the data from one RCE for one tick, where each RCE controls 128 channels meaning that there were a total of 16 RCEs in the 35 ton. A microslice is then made by combining $1000 \times N$ nanoslices such that it contains 0.5 ms (1,000 ticks) of data across all channels, where N is the number of RCEs that are recorded in the run. Microslices are then combined to make millislices the length of which was configurable. Once produced these millislices were sent by the DAQ to the event builders to be stored as time synchronous LArSoft data objects.

The time synchronous events produced by the DAQ did not, however, correspond to the physics events, this is because the DAQ was originally designed to produce a continuous data stream. This meant that the DAQ was configured to pad events with headers when a sub-system provided no physics information, such as nanoslices in the case of the RCEs. Removing these padded header objects was a remit of the online to offline converter discussed in Section 6.2. As the length of the millislices was configurable it was chosen to be 10 ms (20,000 ticks) in order to best attempt to fully contain physics events and reduce the need for the online to offline converter to stitch DAQ events together. The padding of millislices with headers between physics events introduced some peculiarities in the data recorded such as millislices containing two parts of non-continuous data as shown in Figure 6.2 where the second millislice has no information for the time between the end of physics event 2 and the start of physics event 3.

6.1 Organisation of the data structure

52

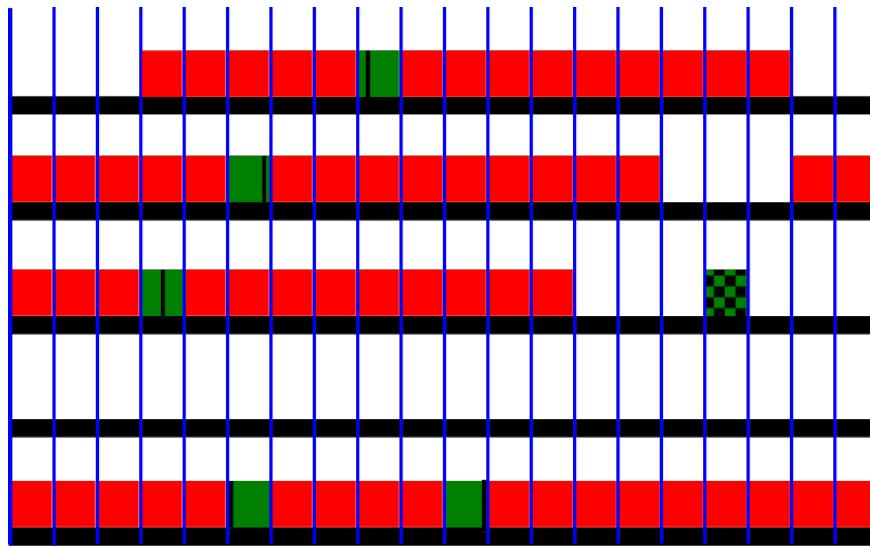


Fig. 6.2 A diagram of possible millislice structures for the TPC data recorded by the 35 ton. Each row represents a millislice, whilst each box represents a microslice. The vertical blue lines delineate each microslice, giving 20 microslices per millislice. Solid red and green boxes represent microslices with TPC data in them. A group of 15 continuous red and green boxes are the recorded “physics events”. Green boxes represent triggers which were used, with the black lines showing the time in the millislice at which the trigger occurred. Green and black patterned boxes represent coincidences of CRCs which were not issued as triggers due to their proximity to a previous coincidence.

1 During normal data taking the last N microslices are buffered in the RCEs so that if a
 2 trigger is issued the previous millislices can be accessed before they are deleted. As the data
 3 is buffered in the form of microslices, previous microslices may only be accessed in whole.
 4 This means that a whole number of microslices must be loaded before the trigger so, when a
 5 trigger is issued part way through a microslice, the previous X microslices are sent to the
 6 event builders. As a result during running there are always a minimum number of ticks both
 7 before (5,000 ticks) and after the trigger (9,000 ticks) but the exact numbers can change by
 8 up to 1,000 ticks for a given event depending on where in a microslice the trigger came. The
 9 result of this is that it is impossible to a priori know the number of ticks before/after a given
 10 counter coincidence. This is shown in Figure 6.2 where the black lines representing triggers,
 11 are seen to occur at different points within the microslices. For example, physics event 1
 12 will have more data after the trigger than physics event 2 as the trigger occurs earlier in the
 13 triggered microslice.

6.2 Reformatting the data to the offline structure

Conversion of the data objects stored in the raw data to the data objects used in simulation required a suite of unpacking services to be written, the specifics of which are not discussed here. These all required a common interface through which to access the data and check that the timing of each component was consistent, and then to produce a final LArSoft file for downstream use. This interface had the added role of producing complete physics events, meaning that it had to be able to combine multiple millislices and extract only the data containing the continuous physics events.

The format that the data reformatter followed was that upon unpacking each of the sub-systems, the TPC ticks would be looped through to see if a user defined set of conditions could be satisfied at that time. These conditions were usually whether an East-West or North-South counter coincidence occurred at that time, or if this millislice contained TPC data whilst the previous one did not. The latter was the default configuration as this gave the option of preserving all of the data gathered, for reasons discussed at the end of Section 6.1. Other conditions were available, though rarely used, such as if the SSPs observed a large flash of light, or if there was a large change in the average TPC ADC value. Once a set of conditions is satisfied a user defined number of pre-condition ticks are gathered. It is set to zero in the case of the previous millislice containing no TPC data as there is no previous data to load which would not have a gap in time, see Figure 6.2. In the case of using a counter coincidence to make an event, a value of 300 pre-condition ticks is normally used, with a maximum of 5000 ticks being able to reliably collected. Once the pre-conditions ticks are gathered a further N post-condition ticks are gathered, where N is defined by the user. Usually 15,000 ticks are gathered when the previous millislice is empty and 5,200 ticks are gathered when there is a coincidence, though a maximum of 9,000 ticks could be reliably gathered. Data from the other components is added to the event if its timestamp is within the timestamps of the first and last ticks in the event when no more TPC data is required or at the end of a millislice if stitching is required. All timestamps are corrected such that the event began at $t=0$ as the reconstruction assumes this and the timestamp of the start of the event is stored in the event record so that it can be accessed later if required.

At all points in this process it is important to integrate flexibility so that the user can choose the length of events, which sub-systems are in the events and what the conditions are for making events. It was also important for users to be able to run the service on already formatted events as the unpacking services were the major overhead in running the interface. It is also conceivable that users would want to reformat Monte Carlo events so as to centre

¹ them around their chosen conditions and so the use of the unpacking was determined by the
² interface depending on the format of the input file.

³ **6.3 Observations on data quality and noise mitigation**

⁴ Reformatting the online data to the offline format was an important step in maintaining
⁵ data quality as subsequently there was no access to the raw data due to the framework of
⁶ the 35 ton software. Some of the important checks which were performed are outlined in
⁷ Figure 6.3. If any of these issues were present in a given physics event then it is discarded
⁸ as the integrity of the data cannot be guaranteed. It was decided that these events would be
⁹ discarded as non-synchronous events would lead to hits in the detector being at incorrect
¹⁰ times and padding empty events with pedestals could mean that tracks seem to disappear and
¹¹ later reappear as they travel through the detector.

¹²

¹³ Another example of inconsistent events is when the sub-systems are not synchronised
¹⁴ with each other. This is normally caused by one of the sub-systems missing a clock increment
¹⁵ from the master timing unit due to the data trigger being issued close to an increment from
¹⁶ the master unit. This misalignment causes an incorrect time sample being read out and so
¹⁷ the data from each sub-system within a millislice is not consistent meaning that it will fail
¹⁸ the timestamp check and so won't be added to the event record. To avoid incomplete events
¹⁹ these physics events are also discarded when observed.

²⁰

²¹ The electronic noise in the 35 ton was higher than anticipated, with the RMS of the RCE
²² ADC being approximately 30 counts compared to an expected thermal noise of around 2.5
²³ ADC counts. Many sources contributed to this elevated noise, some of which are explained
²⁴ below.

²⁵

²⁶ Though not directly affecting the noise issues “stuck ADC codes” were a feature of the
²⁷ data which had to removed. “Stuck ADC codes” were caused by bit level corruption where
²⁸ lowest 6 bits in the ADC became frozen to either 0x0 or 0x3f. This was observed during the
²⁹ first stages of commissioning and an algorithm to remove them was developed and tested
³⁰ on Monte Carlo [23]. In simulations it was observed that the signal could be recovered with
³¹ minimal losses, as shown in Figure 6.4 where the blue lines (after removal) are seen to closely
³² match the black lines (before adding stuck codes).

³³

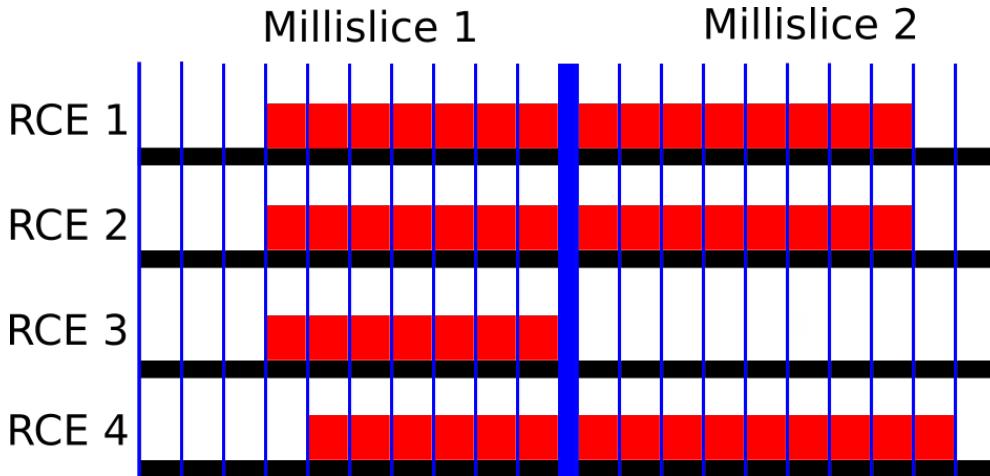


Fig. 6.3 A diagram of TPC microslices within millislices in the 35 ton data stream. Two millislices are shown, each containing 10 microslices. One physics event straddling the millislice boundaries is shown and 4 RCEs representing each row are read out. The vertical blue lines delineate each microslice (0.5 ms, 1,000 ticks), with the thick blue line showing the millislice boundary. Solid red boxes represent micro slices with TPC data in them. It can be seen that RCEs 1 and 2 contain data for the same interval, whilst the data from RCE 3 in millislice 2 has been “Dropped,” and the data from RCE 4 is shifted by 1 microslice from RCEs 1 and 2 and is thus “Inconsistent.” As a result of these issues this physics event would be discarded as data integrity cannot be guaranteed.

A significant portion of the noise was correlated between groups of 32 channels, where the ADCs would coherently oscillate. To remove these coherent shifts, ADC baselines were calculated for these groups of 32 channels at each tick and then subtracted from the measured ADC values. This was found to be an effective method of removing coherent noise in MicroBooNE [24]. The effect of removing coherent noise is shown in Figure 6.5, where the signal peak becomes much easier to discern after noise removal and a coherent noise peak around tick 6030 is removed. An issue with removing coherent noise in this way is that events which are parallel to the APAs will produce signals at common times across adjacent wires and these signals may be removed along with the coherent noise causing a reduction in the hit reconstruction efficiency. The only way to prevent this is to “protect” potential signal regions from the coherent noise removal, as is done in MicroBooNE [24].

When a Fast Fourier Transform (FFT) [25] is performed on the coherent noise subtracted waveforms, it can be seen that signals occur with specific frequencies. Some of these frequencies are caused by real energy depositions, whilst others are due to the electronics noise. It is possible to remove the noise frequencies by applying Wiener filters [26]. Frequency spectra are taken for each of the three planes and a clear signal is both preserved and suppressed. The

6.3 Observations on data quality and noise mitigation

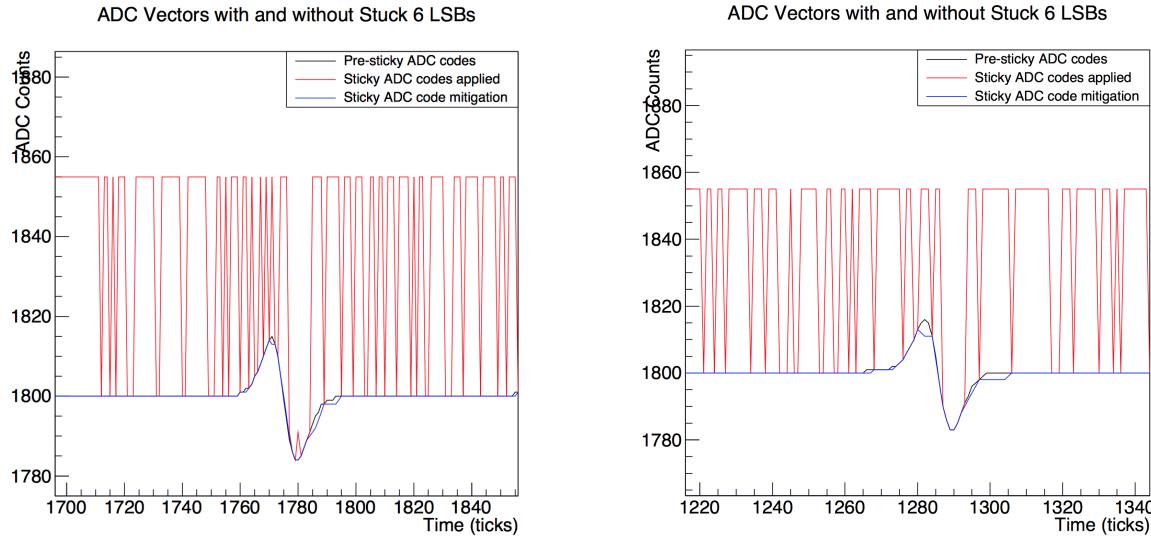


Fig. 6.4 Two Monte Carlo spectra showing the effect of the introduction and removal of stuck bits on a simulated signal. The black line shows the simulated signal on a wire, which is then modified by adding the effects of “stuck ADC codes,” shown by the red line. The “stuck ADC codes” are then removed, and the resulting signal is given by the blue line. It can be seen that the signal loss is minimal after the “stuck ADC codes” are removed. The figures were taken from [23].

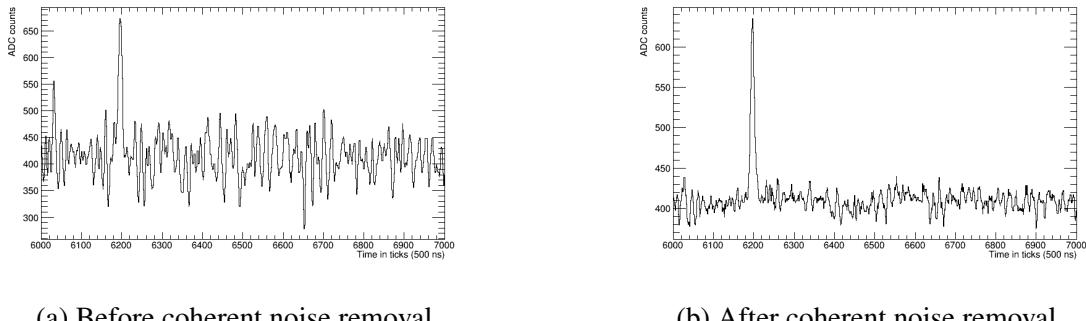
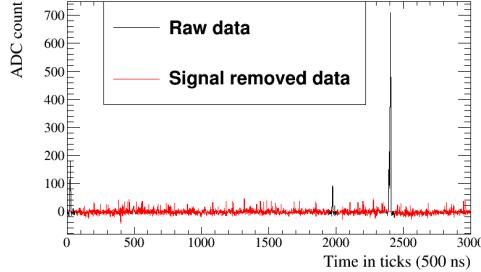


Fig. 6.5 The effect of coherent noise removal on a 35 ton signal event.

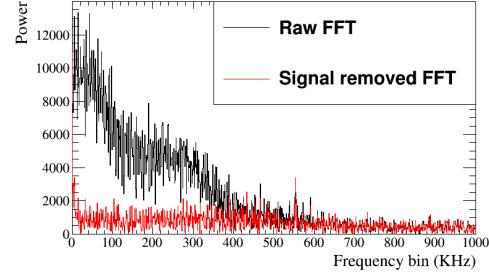
¹ raw signal spectra are then divided by the signal suppressed spectra to produce *signal/noise* frequency spaces. The signal regions to be conserved, can then be found by fitting a combination of sigmoid functions to the frequency spaces around regions of high *signal/noise*. A demonstration of how this was applied, is shown in Figure 6.6. It is also possible to remove specific frequencies which are not removed by the filters, this was necessary for a 54 KHz noise component introduced by the fluorescent lights in the detector hall. After the run ended it was found that some of the high frequency noise components were introduced by a short on a warm power cable, the techniques used to find this cable will be used when commissioning

6.3 Observations on data quality and noise mitigation

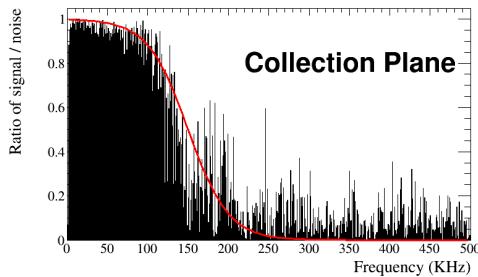
future detectors [27].



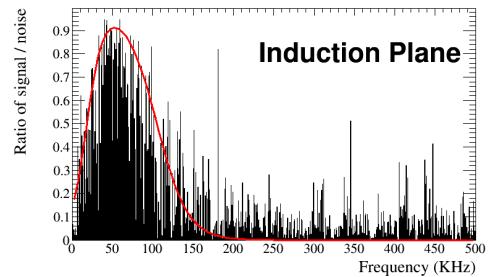
(a) A raw and signal subtracted waveform for a collection plane wire.



(b) The FFT of the raw and signal subtracted waveform for a collection plane wire.



(c) The *signal/noise* ratio for a collection plane wire, the red line shows the fraction of frequency power which passes the filter.



(d) The *signal/noise* ratio for an induction plane wire, the red line shows the fraction of frequency power which passes the filter.

Fig. 6.6 The application of Wiener filters to the 35 ton data.

An example of the effect of the noise mitigation steps is shown in Figure 6.7, where the left side shows the raw data and the right side shows the data after the stuck code unsticker, coherent noise removal and Wiener filter algorithms have been applied.

Transitions to a higher noise state associated with strong signals at high frequencies, between 400 and 650 KHz, were observed after cool down. The transitions would occur approximately every 2 hours and were occasionally observed to happen shortly after a saturation event across the whole detector [27]. Once the state was induced the only way to stop it was to power cycle the low voltage supplies. It was found that power cycling APA3 could both stop and induce the higher noise state, importantly this was the only APA with electronics located at the base of the TPC. The data taken during the elevated noise state was unrecoverable as the electronics noise was too large, and so upon the observation of a transition the low voltage supplies were power cycled. It was observed that the transitions occurred much less frequently when APA3 was not powered and so it was not used for

6.3 Observations on data quality and noise mitigation

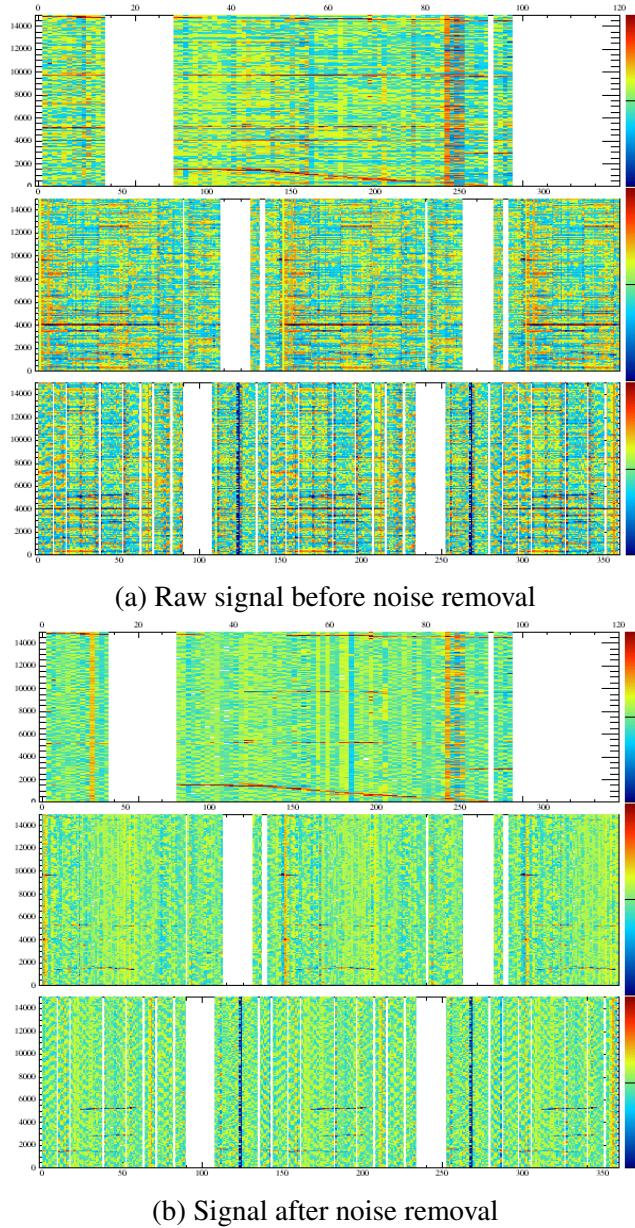


Fig. 6.7 Event displays showing the effect of the noise removal algorithms on data in the 35 ton. The event displays show the signals in the collection, U and V planes respectively. The plots show wire number, time in ticks and charge in ADC counts on the x , y and z axes respectively. The effect of the noise removal algorithms can clearly be seen, as large changes in charge due to the noise are no longer present after they have been applied. The application of the noise removal algorithms does however also remove real signals, as depositions across many channels at the same time which were present before their application can no longer be seen after they are applied.

significant portions of the data taking period. Despite efforts to study the transitions during warm testing they were unable to be induced and have not been observed in other experiments such as MicroBooNE despite using the same low voltage supplies. It is thought that the cause of the transitions is a feedback loop in the low voltage cable which was much longer in the 35 ton than in MicroBooNE, this would explain why APA3 was more susceptible as the cable is routed past its electronics [28].

6.4 Performance of reconstruction algorithms

Following the noise removal outlined above hit and track finding was still more difficult than in simulations due to the still elevated noise level. In order for a sensible number of hits to be reconstructed the hit finding threshold had to be substantially increased in data as compared to Monte Carlo, this meant that many of the low energy hits would not be reconstructed.

A potential solution to not reconstructing the low energy hits is to use the counter positions to select only hits which could have caused coincidences. When determining whether a reconstructed hit could have caused the counter coincidence a two-dimensional window around the counter edges in the yz plane is constructed and timing information is used to extend this to three dimensions. The x position of the hit can be calculated using the hit time and electron drift velocity using Equation 5.1a.

Determining whether collection plane hits are within the counter window is trivial as they have a constant z position and either cover the full detector height (tall APAs) or roughly half of the detector height (short APAs). The wrapping of the induction planes, however, means that each wire segment has to be considered individually and that multiple segments of a given wire could lie within the counter shadow. The 3-dimensional volume that is enclosed by connecting the edges of the counters which were hit to cause the counter coincidence, is called the “counter shadow,” and the wires which lie within the 2-dimensional projection of this volume onto the yz plane are considered here. Choosing between these potential wire segments is done by iterating through the following steps. If at any point only one segment satisfies the condition then this segment is chosen:

- Does the wire segment intersect any collection plane wires which record hits?
 - This is because when there is a signal on an induction plane there should also be signals on the collection wires.
- Are there adjacent wires which have hits at a similar time?

6.4 Performance of reconstruction algorithms

60

- 1 – This is because one would expect a track to deposit energy on multiple adjacent
2 wire segments.
- 3 • Which hit lies closest to the line defined by unique collection plane hits in the xz plane?
- 4 – This follows identical logic to the first criterion, but selects the hit which
5 best matches the collection plane hits and attempts to remove the effect of noisy
6 collection plane wires by only using wires which have one hit within the counter
7 shadow. This would also hopefully improve the quality of the fit as there will not
8 be numerous outlying hits.
- 9 – This can be changed to consider the line defined by previously selected hits in
10 the given TPC and plane where the hit choices are.

11 Following a re-optimisation of the clustering algorithms it was observed that the stan-
12 dard reconstruction could achieve track reconstruction to a similar efficiency as the counter
13 shadowing and so the standard reconstruction has been used in the discussions to follow
14 [29]. There has since been an effort to improve the counter shadowing hit disambiguation to
15 remove the outlying collection plane hits using the MLESAC method [30] whereby points
16 which are far away from a best fit are ignored. These studies are still on-going [31].

17
18 A symptom of the elevated noise state is that signals are often dropped on one of the
19 induction planes, this means that the tracking algorithms often have to combine clusters in
20 only two of the three planes. Reconstruction using two planes was shown to be effective
21 by the ArgoNeuT collaboration [32] so the loss of signal in one of the three planes is not
22 prohibitive to track reconstruction. Another consequence of the elevated noise level is that
23 even when the counters are used to seed hit finding, the hit finding threshold is too high
24 to reconstruct the very lowest hits. This causes the plot of dQ/dx for muons, shown in
25 Figure 6.8, to look flat due to a cutoff at 100 ADC cm⁻¹ below which no hits are found. The
26 inability to reconstruct the lowest energy hits means that calorimetry is all but impossible
27 on the 35 ton dataset even though the tracking algorithms perform relatively well. The
28 inability to perform reliable calorimetry en masse means that the only particles which can
29 be assuredly identified are the muons which triggered the counter coincidences, making the
30 analysis proposed in Section 5.4 extremely difficult, if not impossible.

31
32 The muons in the triggered sample will all traverse the detector but their orientations can
33 be carefully selected by the user, for example one could easily select a sample of muons
34 which cross the APAs at increasing angles, or are parallel to the wire planes at increasing
35 drift distances by matching through-going muons with counter coincidences. The process

6.4 Performance of reconstruction algorithms

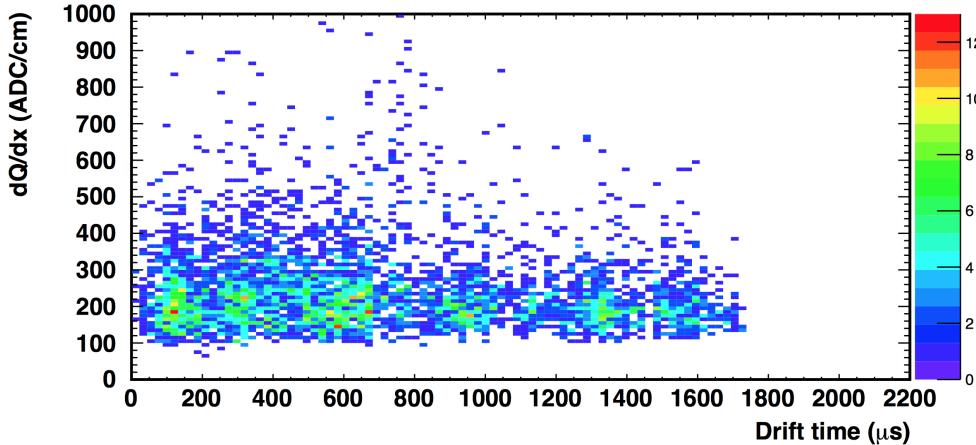


Fig. 6.8 The dQ/dx values for a sample of muon collection plane hits, note the cutoff at 100 ADC cm^{-1} due to the hit finding threshold. Figure taken from[33].

by which this is done is identical for both North-South and East-West coincidences, though more focus will be given to the later and so it will be presented in that regard. The same matching technique would also have been applied to vertical muons had the telescope triggered been utilised. For a reference as to the location of the counter positions around the cryostat see Figure 3.2, and for a representation of only the East-West counters see Figure 6.9.

It is possible to construct a line in the yz plane joining the centres of the two counters which were hit when a coincidence occurred, shown by the dashed line in Figure 6.9. This can then be compared with the trajectory of a track in the yz plane and a dot product of the two vectors calculated. A reconstructed track is assigned to a given counter coincidence if the dot product of the track and the coincidence is more than 0.98 and the hit times are consistent with the x positions of the counters. The results of the dot product calculation are shown in Figure 6.10. Matching only tracks which are well aligned with a counter coincidence should produce a pure sample of tracks, as parallel muons are unlikely to be highly correlated in time and any tracks reconstructed from the noise will have random directions. This is shown in data where if multiple tracks pass the dot product cut they are co-linear and are not randomly orientated, as shown in Figure 6.11.

By matching tracks in this way it is possible to evaluate the reconstruction efficiencies for these muons at increasing drift distances and track angles. If multiple tracks are aligned with the coincidence and are within the expected time region then their track lengths are summed when calculating reconstruction efficiencies as it is expected that the track was split by a region of the detector either being turned off or too noisy to reliably reconstruct a

6.4 Performance of reconstruction algorithms

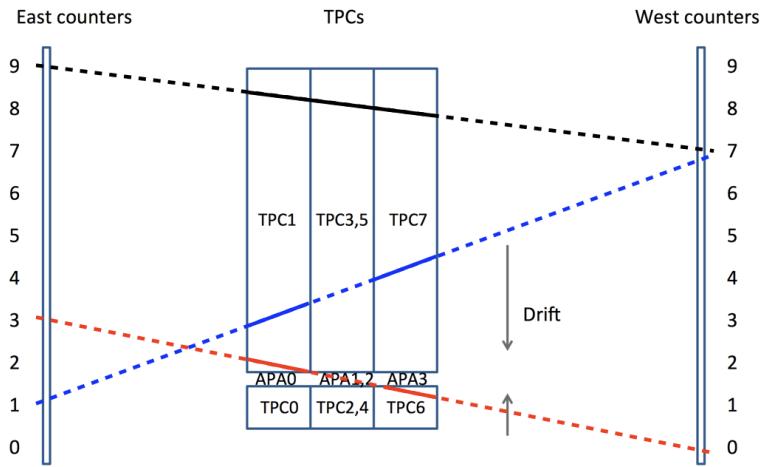


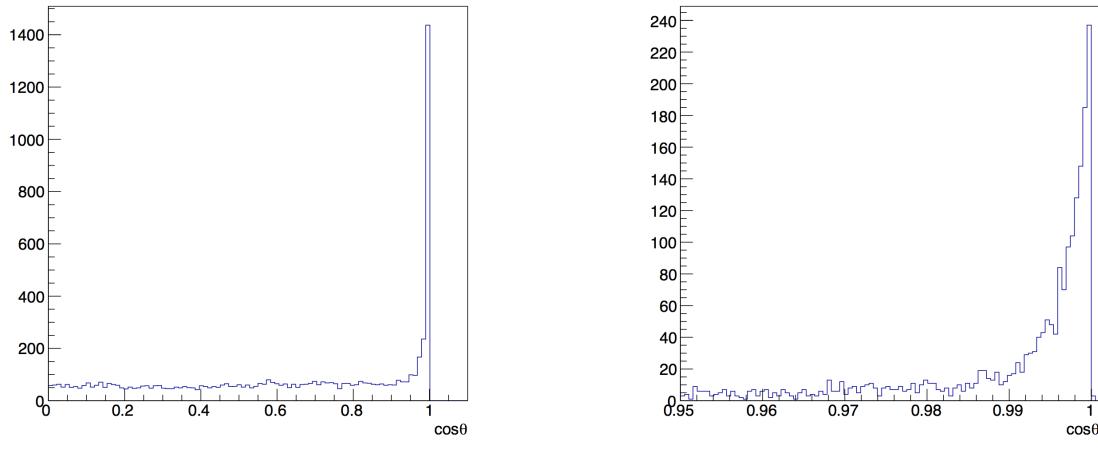
Fig. 6.9 The numbering scheme for the East - West counters in the 35 ton. The counters have been numbered from 0 to 10 depending on their position from the end of the short drift volume. This is different to the LArSoft numbering scheme shown in Figure 3.2 where they go from 6-15 and 28-37 for the East and West counters respectively. Three muons which would have caused coincidence triggers are shown as dashed lines, and the reconstructed tracks they produced are shown as solid lines. The red track would be an APA crossing event, and produced tracks in TPCs 1 and 6. The black muon was fully reconstructed as one continuous track, however the blue particle was not reconstructed in the middle TPCs and so was reconstructed as two separate tracks.

1 track. If these tracks have a combined track length of more than 50 cm then the coincidence
 2 is identified as having been successfully reconstructed. This threshold is much lower than
 3 the true track length which should be reconstructed, >150 cm, but few particles are fully
 4 reconstructed in the data and so a compromise is made to achieve a large enough sample of
 5 tracks upon which analyses can be performed. A reconstructed track that is 50 cm long is
 6 likely to have a large number of hits on collection plane wires that are not noisy, and it is
 7 these hits which are required when calculating purity or measuring the effect of diffusion
 8 as discussed in Section 6.5. A track with length more than 50 cm is also likely to have
 9 been stitched between TPCs due to the geometry of the 35 ton and track trajectories. The
 10 demonstration of stitching tracks between TPCs was a design goal of the 35 ton, and so
 11 identifying tracks where this was achieved satisfies that goal.

12

13 An important concept that must be introduced before these reconstruction efficiencies
 14 can be described is that of a “counter difference.” The “counter difference” of a coincidence
 15 and its associated tracks is defined as the absolute difference between the counter numbers

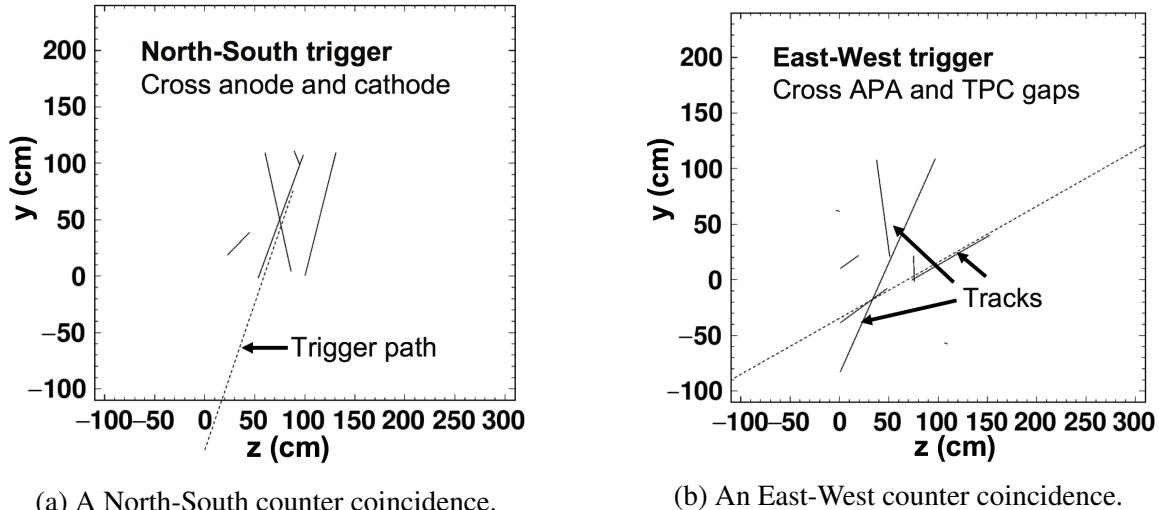
6.4 Performance of reconstruction algorithms



(a) All dot product values.

(b) Dot product values close to 1.

Fig. 6.10 The dot product of the track and vector joining the centres of the coincidence counters in the yz plane. A threshold value of 0.98 is required for a track to be considered to be due to the counter coincidence. It can be seen that many tracks are well aligned with counter coincidences, having dot products of more than 0.99.



(a) A North-South counter coincidence.

(b) An East-West counter coincidence.

Fig. 6.11 The alignment of reconstructed tracks with the vectors joining the centres of the coincidence counters. The dashed lines show the vectors joining the centres of counters hit in the coincidence, whilst the solid lines show the reconstructed tracks. Figures taken from [29].

of the East and West counters that were hit, as shown in Figure 6.10. As such, the “counter differences” of the coincidences shown in Figure 6.10 are 2, 3 and 6 for the black, red and blue coincidences respectively. Given the orientation of the counters, the rarest counter difference will be 9, as only particles which hit counters (E_0 and W_9) and (E_9 and W_0) will

6.4 Performance of reconstruction algorithms

64

Table 6.1 The angles relative to the APAs corresponding to given counter differences. Though the East and West counters have a width in the y (vertical) direction, this is much less than their extent in the z direction. The depth of the counters, their extent in x is negligible compared to the separation of the East-West counters. The counters have identical widths in both the y and z directions. The angles are calculated using the difference in the centres of the counters in the z direction divided by the separation of the East and West counters in z .

Absolute counter difference	Approximate angle ($^{\circ}$)
0	0 ± 2.1
1	4.2 ± 2.1
2	8.4 ± 2.0
3	12.5 ± 2.0
4	16.5 ± 2.0
5	20.3 ± 1.9
6	23.9 ± 1.8
7	27.3 ± 1.7
8	30.7 ± 1.6
9	33.5 ± 1.5

¹ have a counter difference of 9. In contrast to this, the most common value for the counter
² difference is 1, as there are many possible combinations of East and West counters being hit
³ to give this counter difference. In the discussions below “counter difference” is occasionally
⁴ referred to as “delta counter” or “ Δ counter.” The approximate angles which tracks with
⁵ given counter differences have relative to the APA frames is shown in Table 6.1.

⁶
⁷ Figure 6.12 shows a range of reconstruction efficiency plots for combinations of different
⁸ counter differences and different drift distances. As the counter coincidences with large
⁹ counter differences will have large variations in drift positions, the drift distance plotted here
¹⁰ is the average x position of the counter centres that were hit. For example, if the two counters
¹¹ that produced the coincidence are at 10 cm and 230 cm respectively, then the drift distance
¹² plotted would be 120 cm. This distance is called the “coincidence centre” in the following
¹³ discussion. It should be noted that only coincidences which would produce tracks that are
¹⁴ contained within the long drift volume are considered here, hence there being no negative x
¹⁵ positions.

¹⁶
¹⁷ From Figure 6.12a it is evident that the reconstruction efficiency for tracks with shallow
¹⁸ angles relative to the APAs is extremely poor, with the efficiency for tracks aligned with
¹⁹ counter differences of 0 or 1 never rising above 10%. This is due to the coherent noise
²⁰ removal where hits which are correlated in time will be removed as they will be perceived

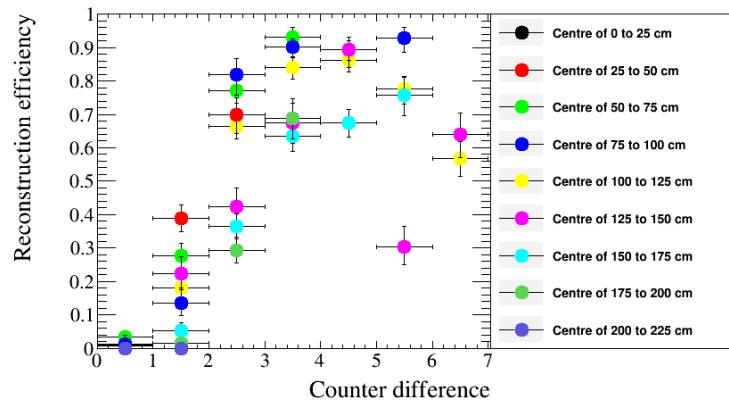
as being noise as opposed to real signals. As the difference in counter number increases the efficiency is seen to increase, though the rate of this increase is seen to depend on the “coincidence centre”. The effect of increasing “coincidence centre” can be seen more clearly in Figure 6.12b where the efficiency for each counter difference as a function of “coincidence centre” is plotted. Here it can be seen that the reconstruction efficiency decreases for coincidences that are centred further away from the APAs. This is due to the fact that when an energy deposition has further to drift, it will induce a smaller pulse on the wires meaning that it is more likely to be below the hit threshold. Figure 6.12c combines Figures 6.12a and 6.12b to show how the reconstruction efficiency for increasing “coincidence centre” changes with increasing counter difference. It can be seen that tracks with counter differences of between 3 and 5 where the “coincidence centre” is between 60 cm and 140 cm away from the APAs are the best reconstructed coincidences. Finally, Figure 6.12d shows how the frequency of coincidences of a given counter difference occurs compared to how many events contain reconstructed tracks which are aligned with the coincidence. It can be seen that as stated earlier, the most common counter difference is 1, with the least common being a counter difference of 9. However, given the low reconstruction efficiency seen for the lowest counter differences few tracks are reconstructed. This means that when considering the reconstructed tracks, most are due to coincidences with counter differences of either 3, 4 or 5.

6.5 Measuring interaction times using electron diffusion

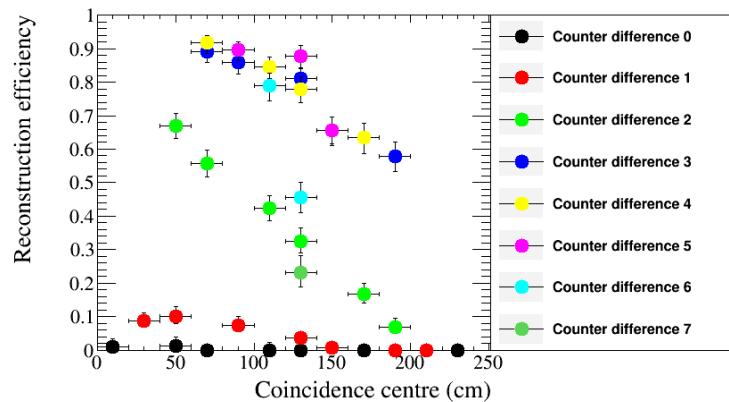
As electrons drift from the interaction point to the wire planes they become spread out in both time and space, this effect is known as diffusion and is an important property of electron transport in LAr which must be well understood. The mechanism by which diffusion occurs in LAr was first discussed by Atrazhev-Timoshkin [34], and has since been developed to consist of a complete set of measurements for electric fields between 100 and 2000 V cm⁻¹ [35]. The diffusion of electrons is rarely isotropic and so the component that is transverse to the drift field and the component that is parallel to the drift field are normally measured separately. Diffusion parallel to the drift field is called longitudinal diffusion and is generally smaller than the component of diffusion that is transverse to the drift field. Figure 6.13 shows how diffusion can smear the electrons collected on a set of wires when the electrons are initially highly correlated in time and space.

Longitudinal diffusion has the effect of spreading the drifting electrons out in time causing signals to become wider in time and smaller in height as the total charge is conserved.

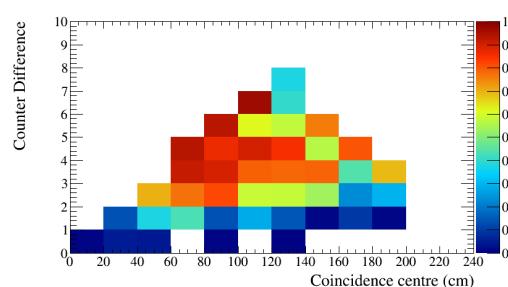
6.5 Measuring interaction times using electron diffusion



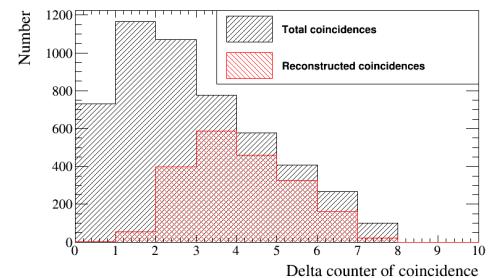
(a) The reconstruction efficiency as a function of counter difference for different coincidence centres.



(b) The reconstruction efficiency as a function of coincidence centres for different counter differences.



(c) How the reconstruction efficiency changes for increasing coincidence centres and counter differences.



(d) The number of events for each counter difference that were recorded in the data and the number of those which were successfully reconstructed.

Fig. 6.12 The reconstruction efficiencies for coincidences that trigger an East-West coincidence in the 35 ton data over a 2 day running period.

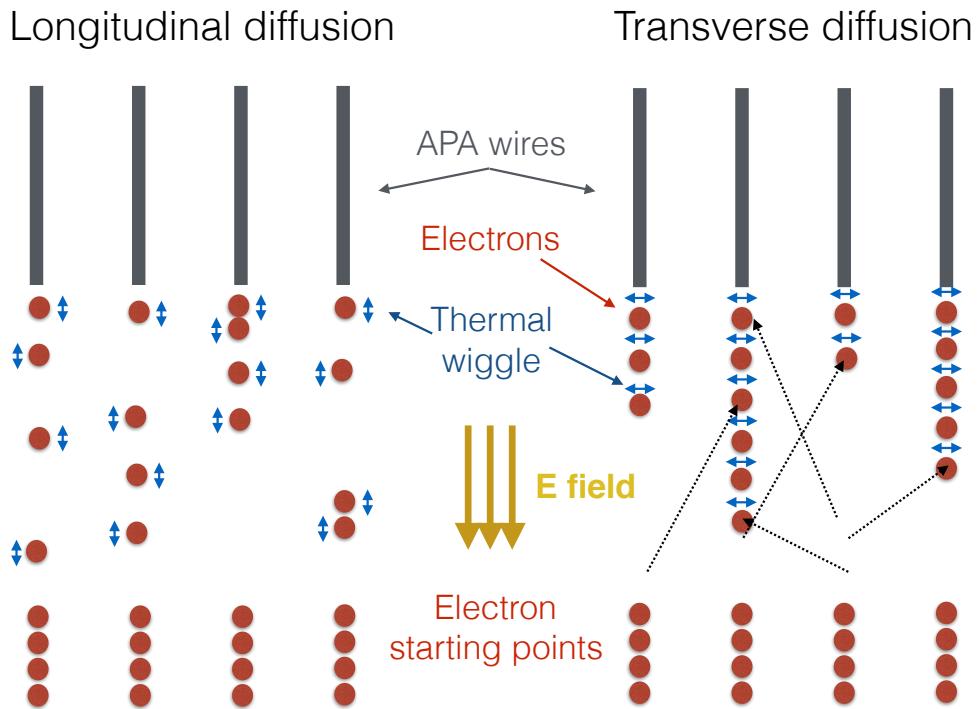


Fig. 6.13 A schematic showing the longitudinal diffusion (left) and transverse diffusion (right) of electrons. In both cases four electrons are initially shown below four wires, and are allowed to diffuse in either the drift direction or perpendicular to the drift direction in the longitudinal and transverse cases respectively. It can be seen that the effect of the diffusion is to make the electrons spread out in time in the case of longitudinal diffusion and to spread out in space in the case of transverse diffusion. Figure taken from [36].

The increasing hit width can be measured for increasing drift times (distances) provided the hits do not fall below a hit finding threshold. Transverse diffusion causes drifting electrons to spread out in space, changing the amount of charge deposited on a wire and reducing the charge resolution of the detector. Transverse diffusion is measured by discerning how the width of the hit charge distribution changes for increasing drift distances [35].

Through-going particles make ideal tracks to study diffusion as they are minimally ionising and so have roughly constant energy depositions along their tracks. The tracks they produce can also cover a wide range of drift distances if they are not parallel to the APAs. The drift distances of hits within a track can be determined by matching the track with a counter coincidence as discussed at the end of Section 6.4 and then correcting the x co-ordinates of the hits using the result of Equation 5.1b in Equation 5.1a.

6.5 Measuring interaction times using electron diffusion

Traditionally the only way to determine an interaction time for a track is to match it to either an external calibration source such as whether it aligns with an external counter coincidence, or to match it to a flash of scintillation light as in Section 5.1. These techniques are particularly crucial for neutrino detectors on the Earth's surface such as MicroBooNE where each neutrino interaction usually has a background of at least one cosmic muon. The reconstructed tracks from this muon background has to be distinguished from those due to the neutrino interaction in order correctly assign a scintillation flash to the reconstructed tracks. An example of an event that has many scintillation flashes and cosmic muons which need to be correctly associated, is shown in Figure 6.14. However, it may be possible that the change in hit width due to diffusion as a particle travels through the detector, could be used to determine the interaction time; though this has not been attempted before. To study whether this is possible, the effects of diffusion would have to be measured for a sample of tracks with known interaction times and orientations.

14

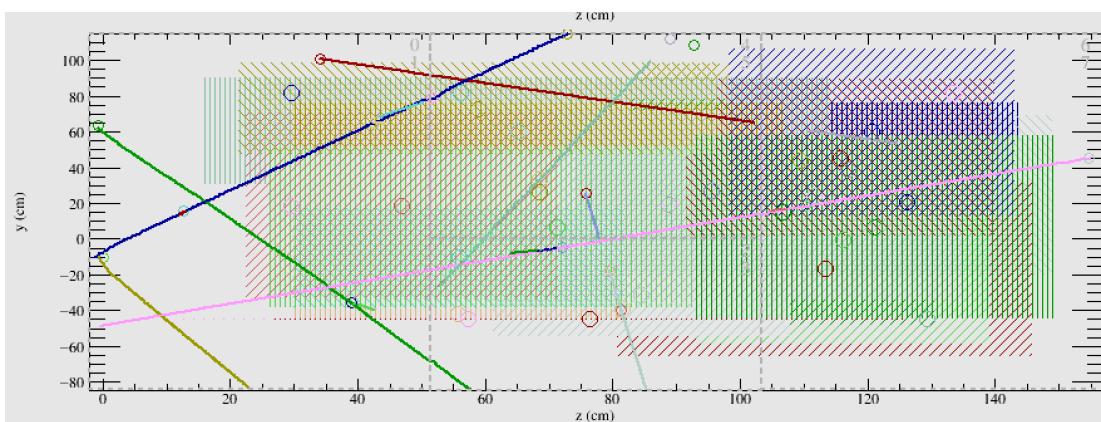


Fig. 6.14 A simulated event display showing multiple tracks and flashes to be assigned to each other in the 35 ton. The coloured lines represent reconstructed tracks, whilst the coloured dashed boxes represent flashes.

The 35 ton dataset is ideal for testing this hypothesis as the counters are able to provide a sample of tracks with known angles and interaction times which can be used to tune interaction time determination metrics. These metrics can then be applied to another sample of tracks where the interaction time is known but not used so that the accuracy of the calculated interaction times can be found. As longitudinal diffusion is the dominant effect that increases the hit width, transverse diffusion will not be directly considered further. However, as noted in Section 6.4, the noise level in the 35 ton data causes reconstruction issues and so it is also useful to compare the method against a low noise detector. Monte Carlo can provide this sample, and this comparison is shown in Section 6.5.2. It is also useful

6.5 Measuring interaction times using electron diffusion

69

to observe the effects that different detector conditions such as, the electric field, the electron lifetime, noise level and rate of diffusion have on the method. This is shown in Section 6.5.3.
 First though, the method is performed on the 35 ton dataset.

6.5.1 Determining interaction times in 35 ton data

When calculating the determination metrics, only hits on wires which are not noisy want to be considered. This is because wires with a high level of correlated noise observe hits with a wider RMS. This is shown in Figure 6.15, where, when a baseline noise of 10 ADC counts is added to a simulated hit with a peak value of 50 ADC counts and an RMS of 10 ticks, the width increases by over 10%. Hits with delta rays also need to be removed as the deposited energy will be larger and over a longer period of time than hits from the main track, this will make the RMS of the individual hit wider and also increase the width of the charge distribution for the track. To remove these hits only hits which satisfy the following cuts are used:

- No hit on the same wire within 50 ticks of the hit in question - removes delta rays.
- No more than 10 hits on the same wire in the whole 15,000 tick data sample - removes clearly noisy wires.

These cuts will clearly become much more restrictive as the noise level in the detector increases, but they are essential in order to collect a dataset which is not overpowered by noise. Only collection plane hits are used, as the charge resolution is better and the signals are unipolar as opposed to bipolar meaning that a Gaussian function can be easily fitted to the signals. Additionally the *signal/noise* ratio on the collection planes was much higher than on the induction planes for the 35 ton dataset and so the hits could be much more reliably reconstructed.

Diffusion is a track angle dependent property and so track angle ranges have to be considered independently. To minimise the number of figures presented, only graphs made for tracks which have a counter difference of 4 are shown, though the procedure is identical for tracks of all counter differences. Tracks with a counter difference of 4 were chosen as they were one of the angles for which tracks were well reconstruced in the data, see Figure 6.12. The tracks are considered en masse, and so the hits for every track are separated into 10 cm regions of increasing drift distance from the APAs. The following quantities are calculated for each 10 cm drift region:

- The hit *RMS* - the most direct way to measure transverse diffusion.

6.5 Measuring interaction times using electron diffusion

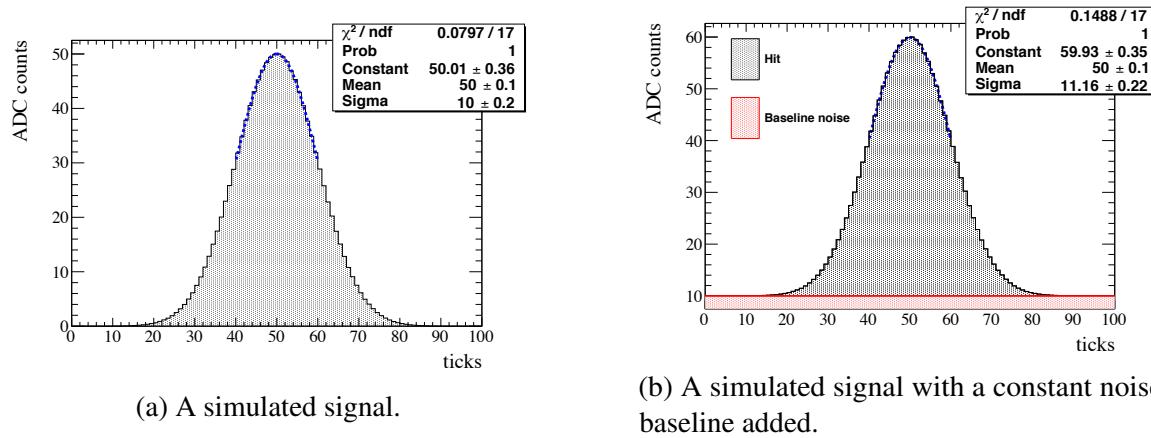
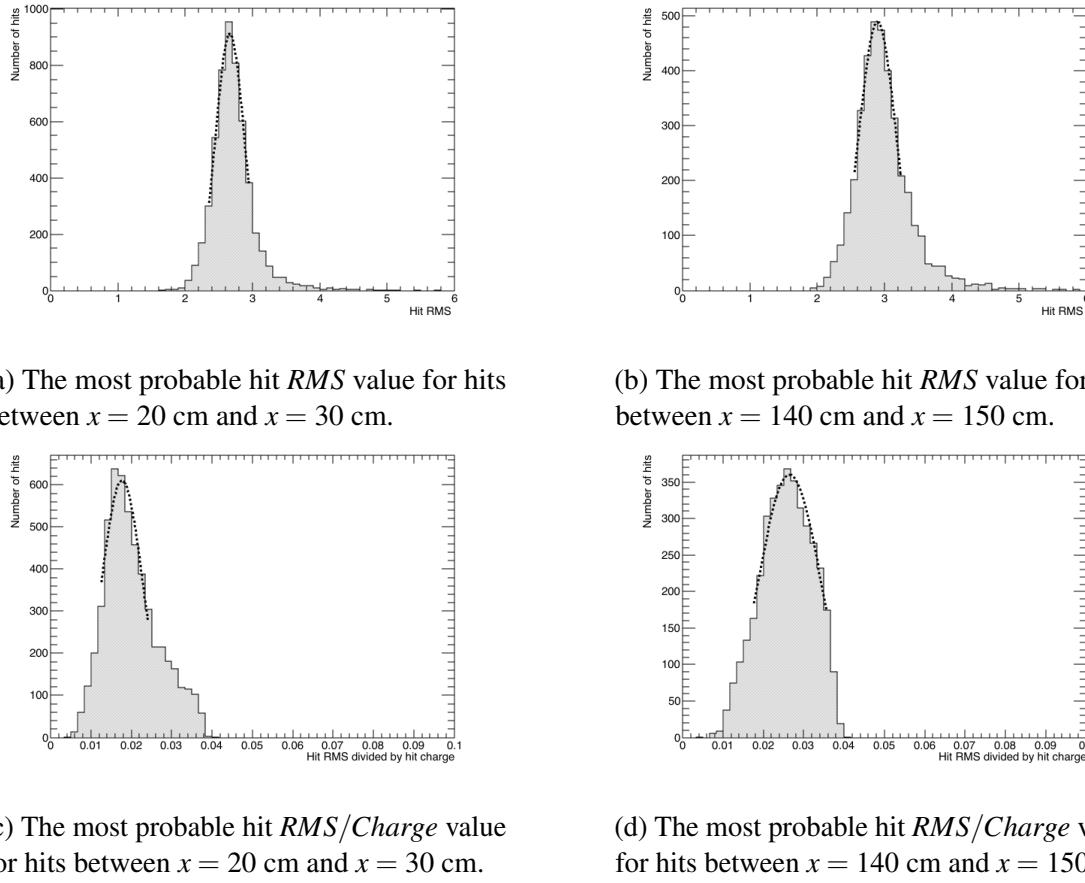


Fig. 6.15 A simulated signal with a width of 10 ticks and an amplitude of 50 ADC counts, both before and after a constant noise baseline, of 10 ADC counts, is added. In reality the noise would fluctuate with time. When a Gaussian function is fitted to each signal, it is seen to be more than 10% larger for the signal where the noise baseline is added. This shows that noise can cause the measured width of a hit to increase. Figure taken from [36].

- 1 • The hit *RMS/Charge* - an attempt to incorporate the effect of impurities in the LAr for
- 2 relatively low purity data which will have a drift distance dependence.
- 3 – The charge of a hit is calculated by integrating the ADCs of the reconstructed hit
- 4 over time.
- 5 Fitting Gaussian functions around the peaks of the distributions will yield the most probable
- 6 values for the drift regions, as is shown in Figure 6.16.
- 7
- 8 The drift distance effect of diffusion can then be observed by fitting these most probable
- 9 values as drift distance increases. This drift distance dependence is shown in Figure 6.17
- 10 for tracks that are associated with a coincidence which had a counter difference of 4. The
- 11 angular dependence can then be shown by observing how the most probable fit value at a
- 12 drift distance of 0 cm changes for increasing angles, this is shown in Figure 6.18. A drift
- 13 distance dependence can clearly be seen in the data as the most probable hit *RMS* is seen to
- 14 increase for hits which originate further from the APAs. It also clear that there is an angular
- 15 dependence on the hit width as the most probable hit widths next to the APAs is seen to rise
- 16 for tracks associated with coincidences with large counter differences. These dependencies
- 17 show that, when considering a large sample, diffusion can be separated into distance and
- 18 angular dependant dependencies, however whether this can be observed for individual tracks
- 19 has not yet been considered.

6.5 Measuring interaction times using electron diffusion



(a) The most probable hit RMS value for hits between $x = 20$ cm and $x = 30$ cm.

(b) The most probable hit RMS value for hits between $x = 140$ cm and $x = 150$ cm.

(c) The most probable hit $RMS/Charge$ value for hits between $x = 20$ cm and $x = 30$ cm.

(d) The most probable hit $RMS/Charge$ value for hits between $x = 140$ cm and $x = 150$ cm.

Fig. 6.16 Distributions of the most probable values of RMS of the hits (top) and $RMS/Charge$ of the hits (bottom), for points between 20 and 30 cm from the APAs (left) and points between 140 and 150 cm from the APAs (right), for tracks associated with a coincidences that have counter differences of 4.

To consider single tracks, the best line fits for the counter differences for a large sample of tracks, such as in Figure 6.17, need to be used to predict the position you would expect a hit to originate from given a value for its hit RMS and the angle of the track to which it belongs. The predicted positions can then be compared to the known position from the counter coincidence to determine the accuracy of the prediction. As the distributions shown in Figure 6.16 are roughly symmetric around the most probable value one would naively expect that if a track has a sufficient number of hits then the distribution of RMS values for those hits would match that found over a large sample. If this were to be the case then the difference in reconstructed and predicted hit times should be peaked around the track interaction time.

1
2
3
4
5
6
7
8
9
10
11

6.5 Measuring interaction times using electron diffusion

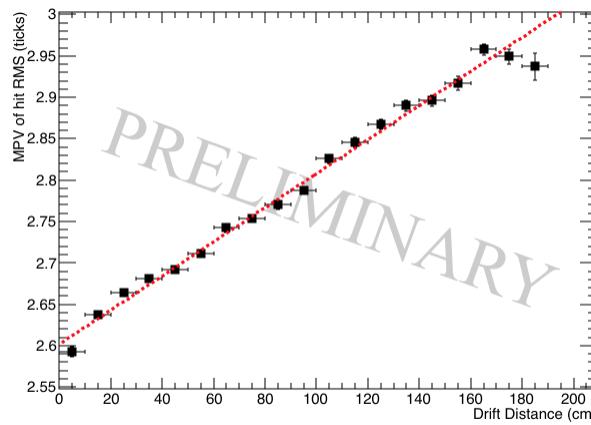


Fig. 6.17 The most probable values of hit *RMS* as a function of drift distance, for tracks associated with a coincidence that had a counter difference of 4.

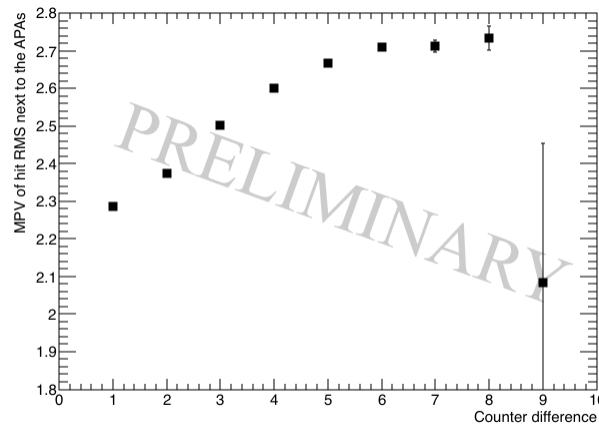


Fig. 6.18 The most probable values of hit *RMS* within 10 cm of the APAs, as a function of the counter difference of the coincidence, that the track was associated with.

¹ An intrinsic assumption in this method is that the track has a large number of collection plane hits that do not contain delta rays and are on wires which would not be identified as noisy. The tracks being considered will have crossed all z values in the detector meaning that a total of 336 collection hits could potentially be reconstructed. Given the reconstruction problems in the 35 ton detector, very few tracks will have hits on all of these collection wires. However, requiring at least 100 collection plane hits is not unreasonable and would correspond to a reconstructed track length of at least 50 cm. The difference between the predicted and reconstructed hit time for each hit is shown in Figure 6.19 for both the hit *RMS*

6.5 Measuring interaction times using electron diffusion

73

and hit *RMS/Charge* metrics.

1

2

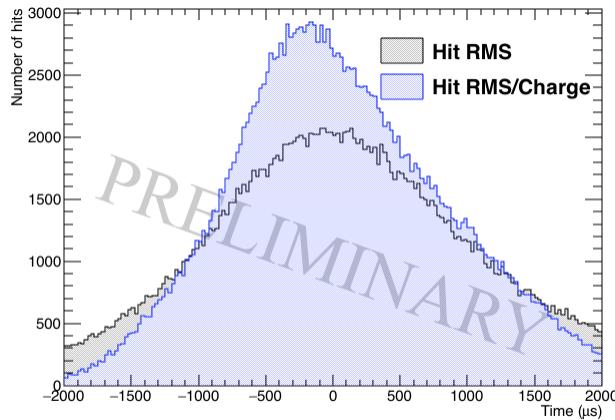


Fig. 6.19 The difference between the predicted and reconstructed hit times in the 35 ton dataset. The differences in time when the hit *RMS* metric is used are shown in black, whilst the differences in time when the hit *RMS/Charge* metric is used are shown in blue.

It can be seen from Figure 6.19, that in the 35 ton dataset both distributions are centred around a time difference of 0 μs which is encouraging as it shows that the method has potential. The width of the distribution for the *RMS/Charge* metric is smaller, and the peak larger, so it is expected that this will provide the more robust metric. This is because these features show that the hit times which are predicted, are more likely to be close to the reconstructed hit times. The peaks are centred around a time difference of 0 as the hit times had previously had the measured interaction time, from the counter coincidence, subtracted from them. This was done so as to avoid the uncertainty which would arise from allowing the coincidences to remain at random times between ticks 5000 and 6000, see the discussion concerning Figure 6.2 for an explanation as to why this occurs.

3

4

5

6

7

8

9

10

11

12

13

When evaluating interaction times the average difference in reconstructed and predicted hit times across every hit on the track must be considered. This is shown in Figures 6.20 and 6.21, where, as expected from Figure 6.19, the *RMS/Charge* metric provides a better estimation of the interaction time. The reason for this is that by utilising the charge information due to losses from impurities this metric gains an extra handle on the drift distance and hence the reconstructed time of the hits. The losses due to impurities are difficult to measure in high-purity LAr environments as the decrease in collected charge with increasing drift distances is small [37]. The effect of increasing LAr purity is shown in Section 6.5.3. Using the change in hit charge in the 35 ton may have a drawback though, because, as shown in

14

15

16

17

18

19

20

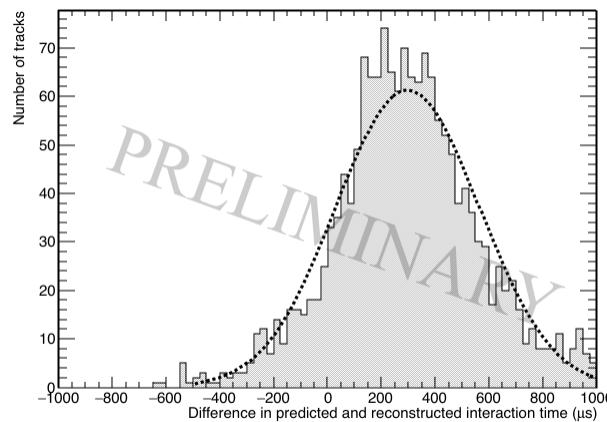
21

22

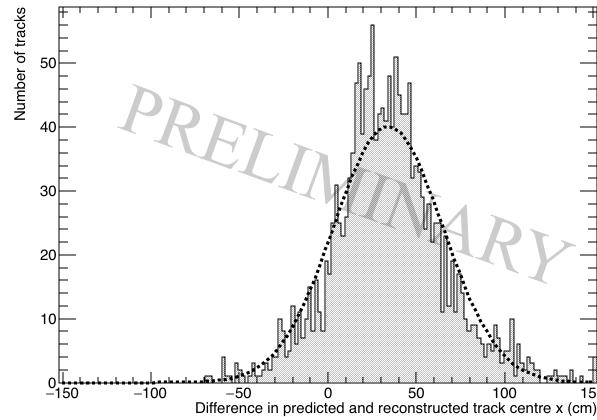
6.5 Measuring interaction times using electron diffusion

¹ Figure 6.8, there is a threshold effect for hits with large drift times. However, as the same
² threshold effect is present in all 35 ton data samples the limitation it introduces is mainly
³ in the efficiency with which 'good' collection plane hits will be reconstructed and so this
⁴ information can be confidently used.

⁵



(a) The average difference in interaction times using the hit *RMS* metric.



(b) The average difference in the central *x* position of a track using the hit *RMS* metric.

Fig. 6.20 The accuracy of the hit *RMS* method in the 35 ton dataset.

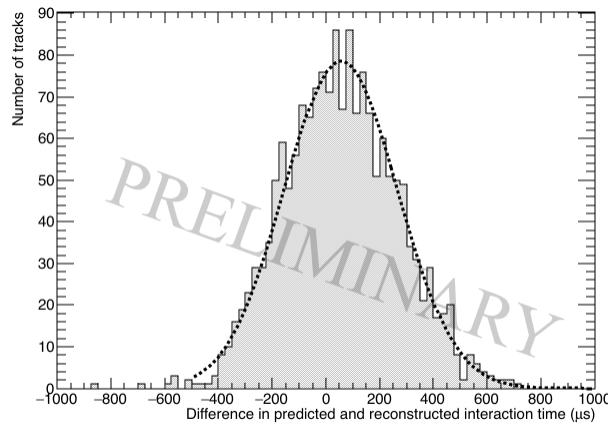
⁶ Figure 6.20 shows that using the effects of diffusion, and the hit *RMS*, the interaction
⁷ time and central *x* position of a track can be reliably predicted in the 35 ton dataset. The
⁸ accuracy in determining the interaction time is found to be 298 μ s, where the distribution has
⁹ a FWHM of 267 μ s. When this is converted into the difference in central *x* position of the

6.5 Measuring interaction times using electron diffusion

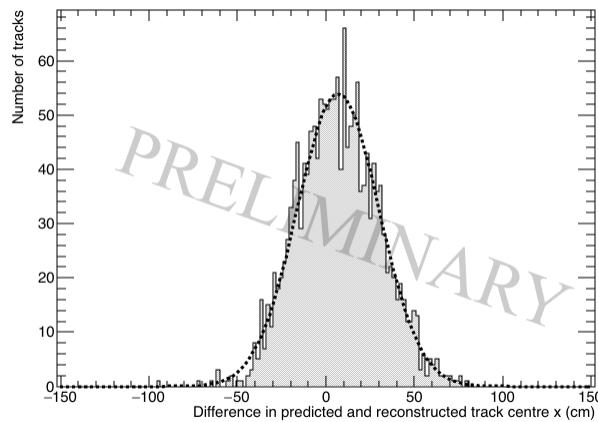
track the accuracy is found to be 32.2 cm with a FWHM of 28.8 cm.

1

2



(a) The average difference in interaction times using the hit *RMS* metric.



(b) The average difference in the central *x* position of a track using the hit *RMS/Charge* metric.

Fig. 6.21 The accuracy of the hit *RMS/Charge* method in the 35 ton dataset.

Figure 6.21 shows that using the effects of diffusion, and the hit *RMS/Charge*, the interaction time and central *x* position of a track can be reliably predicted in the 35 ton dataset. The accuracy in determining the interaction time is found to be $55.3 \mu\text{s}$, where the distribution has a FWHM of $212 \mu\text{s}$. When this is converted into the difference in central *x* position of the track the accuracy is found to be 6.88 cm with a FWHM of 23.1 cm.

3

4

5

6

7

8

1 The resolutions found are quite impressive as given that the total drift time for electrons
2 through the whole 35 ton detector volume of 250 cm is roughly 5200 ticks, it means that
3 tracks can be distinguished throughout the detector volume. As discussed earlier, the issues
4 with noise in the 35 ton dataset affect the accuracy with which tracking and calorimetry can
5 be performed, and so it is reasonable to expect that the effectiveness of the interaction time
6 determination was also affected. Therefore, it is prudent to repeat the study on a Monte Carlo
7 dataset where the detector noise is much lower. This is presented in Section 6.5.2.

8 **6.5.2 Determining interaction times in a low-noise detector using Monte
9 Carlo, and differences with data**

10 When determining interaction times in Monte Carlo simulations, exactly the same criteria are
11 applied to the hits, as δ -rays would still change the measured hit width and will be present in
12 any sample. In a low noise detector it is expected that few wires would be removed due to
13 being noisy but for consistency there is no danger in applying this cut. Imposing a minimum
14 number of collection plane hits is again important to ensure that the distribution of predicted
15 hit times is centred on the interaction time. In addition to the same criteria being imposed on
16 which wires are used, the same metrics are calculated. In all plots shown below the Monte
17 Carlo dataset has been normalised to the size of the 35 ton dataset.

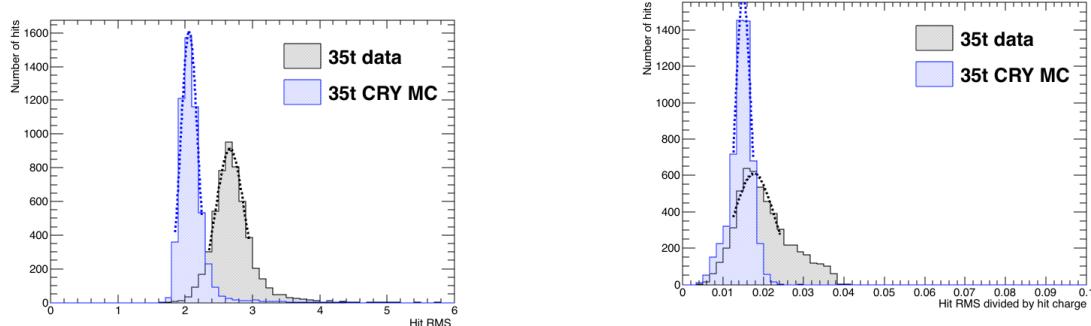
18
19 Figure 6.22 shows both the hit *RMS* and hit *RMS/Charge* distributions for hits that are
20 100 cm away from the APAs and are from tracks associated with a coincidence that has a
21 counter difference of 4. It can be seen that the distributions from the Monte Carlo simulation
22 are tighter than those from the 35 ton data and are also peaked at a lower values of hit *RMS*.
23 This is likely due to the fact that the coherent noise baseline seen in data can increase the
24 width of hits as shown in Figure 6.15 and a higher noise state will affect how well individual
25 hits can be reconstructed. In addition, the most probable values of hit *RMS* at increasing drift
26 distance are shown in Figure 6.23 where the Monte Carlo simulation is again shown with the
27 values from the data. The most probable value of hit *RMS* at a drift distance of 0 cm for a
28 range of counter differences is also shown in Figure 6.24. As was seen when considering
29 the distributions at specific distances and counter differences, the most probable values of
30 hit *RMS* in the Monte Carlo simulation is systematically lower than in the data due to the
31 elevated noise level seen in the data. Another difference between the Monte Carlo and the
32 data is that the gradient of the most probable values of hit *RMS* in data is roughly half of that
33 in the Monte Carlo, this could be due to an overestimation of longitudinal diffusion in the

6.5 Measuring interaction times using electron diffusion

Monte Carlo.

1

2



(a) The most probable hit *RMS* values for hits between $x = 20$ cm and $x = 30$ cm.

(b) The most probably hit *RMS/Charge* values for hits between $x = 20$ cm and $x = 30$ cm.

Fig. 6.22 The most probable values of the hit *RMS* and hit *RMS/Charge* distributions for hits between $x = 20$ cm and $x = 30$ cm, for tracks with a counter difference of 4. The distributions from the 35 ton dataset are shown in black, whilst the distributions from the Monte Carlo simulation are shown in blue.

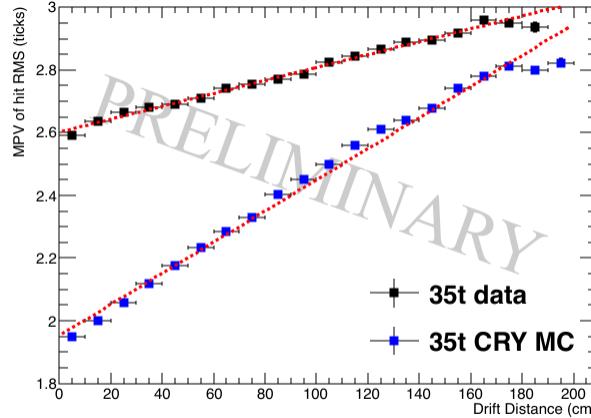


Fig. 6.23 The most probable values of hit *RMS* as a function of drift distance, for tracks associated with a coincidence that had a counter difference of 4. The distributions from the 35 ton dataset are shown in black, whilst the distribution for the Monte Carlo simulation is shown in blue.

Upon calculating the fit metrics in the low-noise Monte Carlo dataset it is then possible to use these to predict track interaction times, this is shown in Figures 6.25 and 6.25. Figure 6.25 compares how reliably the interaction time and central x position of a track can be predicted,

3

4

5

6.5 Measuring interaction times using electron diffusion

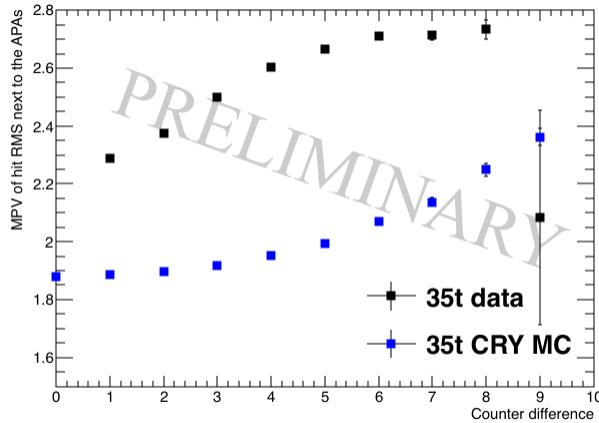
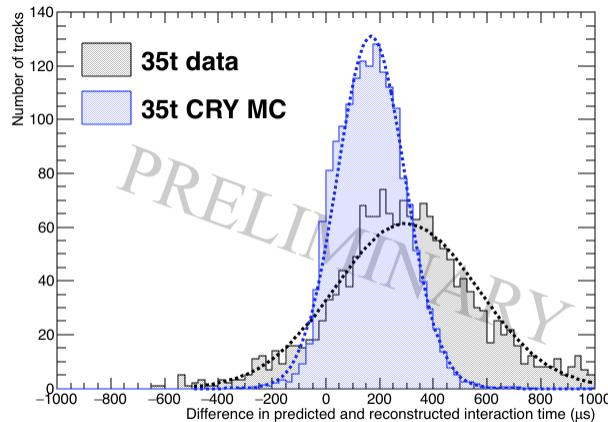


Fig. 6.24 The most probable values of hit *RMS* within 10 cm of the APAs, as a function of the counter difference of the coincidence, that the track was associated with. The distributions from the 35 ton dataset are shown in black, whilst the distribution for the Monte Carlo simulation is shown in blue.

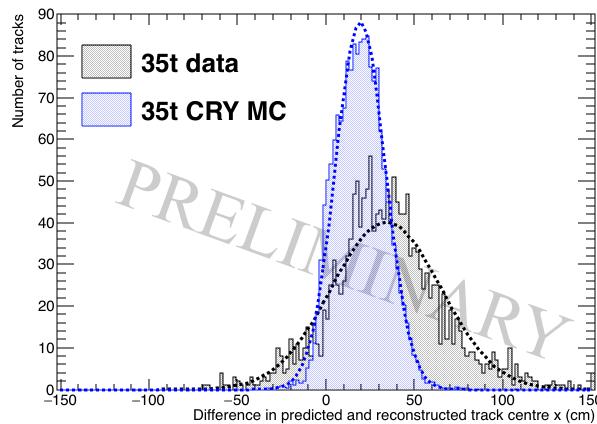
1 using the effect that diffusion has on the hit *RMS*, in the 35 ton dataset and a low-noise
 2 Monte Carlo sample. The accuracy in determining the interaction time in Monte Carlo (data)
 3 is found to be 168 (298) μs , where the distribution has a FWHM of 127 (267) μs . When
 4 this is converted into the difference in central *x* position of the track the accuracy is found
 5 to be 19.4 (32.2) cm with a FWHM of 14.0 (28.8) cm. Figure 6.25 compares how reliably
 6 the interaction time and central *x* position of a track can be predicted, using the effect that
 7 diffusion has on the hit *RMS/Charge*, in the 35 ton dataset and a low-noise Monte Carlo
 8 sample. The accuracy in determining the interaction time in Monte Carlo (data) is found to be
 9 -40.9 (55.3) μs , where the distribution has a FWHM of 110 (212) μs . When this is converted
 10 into the difference in central *x* position of the track the accuracy is found to be -3.5 (6.88) cm
 11 with a FWHM of 12.1 (23.1) cm.

12
 13 The hit *RMS/Charge* metric appears to be able to more accurately predict interaction
 14 times, as was seen in when considering the 35 ton dataset. This is again due to the ability
 15 to incorporate information about losses due to impurities which increase with drift distance.
 16 Also, as expected from the previous figures and the lower noise state in the Monte Carlo it
 17 is seen that the interaction times predicted in the Monte Carlo more closely match the true
 18 interaction times than in the data. An important feature to observe is that, as well as more
 19 accurately predicting the interaction times, the widths of the distributions in Monte Carlo
 20 are less than half of that in the data. This means that the resolution with which tracks can be
 21 distinguished in the Monte Carlo sample is much better than in the 35 ton dataset, again this

6.5 Measuring interaction times using electron diffusion



(a) The average difference in interaction times using the hit *RMS* metric.



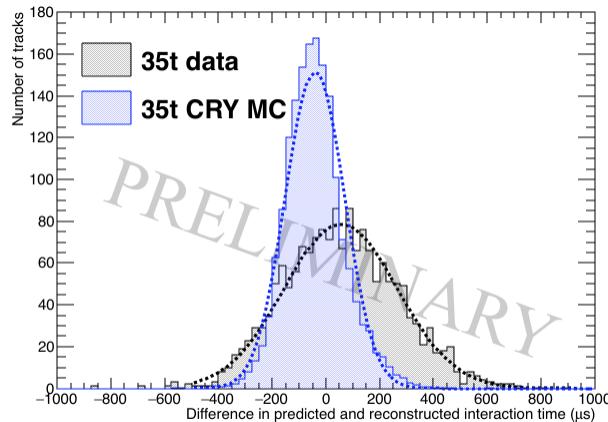
(b) The average difference in the central *x* position of a track using the hit *RMS* metric.

Fig. 6.25 The accuracy of the hit *RMS* method in the 35 ton dataset and a Monte Carlo simulation. The distributions from the 35 ton dataset are shown in black, whilst the distribution for the Monte Carlo simulation is shown in blue

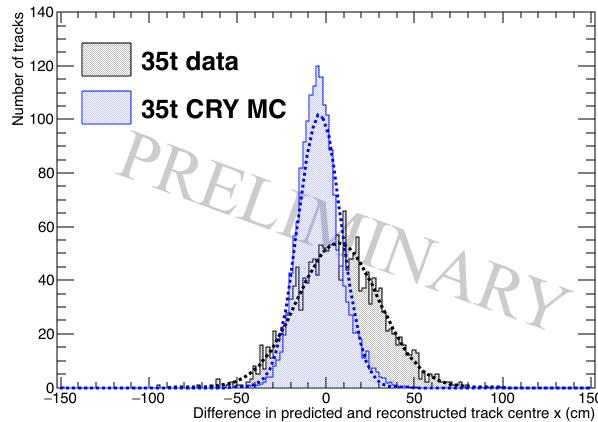
is attributed to the lower noise level in the Monte Carlo.

The calculation of interaction times is clearly much better in the low-noise Monte Carlo than in the 35 ton dataset, however, the distributions are still not centred around 0 implying that there is a systematic error in the method which has not been removed when considering a low-noise environment. Looking at Figure 6.22 the impact of δ -rays can still be seen where the hit *RMS* plot still has quite a significant tail above the most probable value. This will cause the predicted interaction times to be skewed towards larger times as the hits

6.5 Measuring interaction times using electron diffusion



(a) The average difference in interaction times using the hit *RMS* metric.



(b) The average difference in the central *x* position of a track using the hit *RMS/Charge* metric.

Fig. 6.26 The accuracy of the hit *RMS/Charge* method in the 35 ton dataset and a Monte Carlo simulation. The distributions from the 35 ton dataset are shown in black, whilst the distribution for the Monte Carlo simulation is shown in blue

- ¹ containing δ -rays will be wider and so appear to come at later times than they actually do.
- ² Hits containing undistinguishable δ -rays are difficult to remove without looking for slight
- ³ dips in the raw signals caused by the δ -ray beginning to separate from the main track. This
- ⁴ would be almost impossible in the 35 ton dataset given the oscillatory nature of the noise.
- ⁵ Δ -rays can also offer an explanation for the *RMS/Charge* plot underestimating interaction
- ⁶ time as hits containing δ -rays would deposit more charge and this increased charge would
- ⁷ likely be larger than the increased width, causing the *RMS/Charge* to decrease. This is seen
- ⁸ in Figure 6.22 where both the data and Monte Carlo samples have tails at small values of

hit $RMS/Charge$, this decrease in hit $RMS/Charge$ would lead to an underestimation of the interaction time. The 35 ton dataset as a whole overestimates the interaction time though, and this is due to the tail at large values of hit $RMS/Charge$ seen in the 35 ton dataset. It is thought that the collection of hits with large values of hit $RMS/Charge$ is correlated with the noise level in the detector, as it is not seen in the lower noise Monte Carlo sample.

6.5.3 Discerning the impact of changing detector properties using Monte Carlo samples

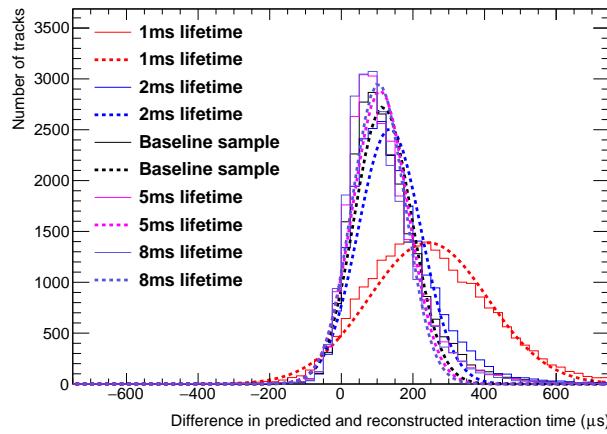
The difficulty of the reconstruction and analysis of the 35 ton data is due to the elevated noise level. We are now going to verify this statement with a study whereby the noise level in a Monte Carlo sample is changed from the low-noise state used in the previous section, to a level more similar to that which is seen in the 35 ton dataset. If the claim that the noise level made reconstruction difficult is correct, then the accuracy with which the interaction time can be determined, should be seen to anti-correlate with the noise level of the simulated detector. This study and the ones which follow have all been performed using the same muons, so as to ensure that the only difference between the different samples are the detector conditions, and only one detector condition is varied at a time. As such, there is one sample that is consistent to all samples, which is where the RMS of the noise is 2.5 ADCs, the electron lifetime is 3 ms, the electric field is 500 V cm^{-1} and the coefficient of longitudinal diffusion is 6.2×10^9 . When presenting the studies with changing detector conditions, the analogous figures presented in the previous section are repeated so as to show how the critical aspects of the methodology are affected.

Figures 6.27 and 6.28 show the accuracy to which the interaction time and central x position of a track can be determined using the effect that diffusion has on the hit RMS and hit $RMS/Charge$ for changing electron lifetimes respectively. Figure 6.27 shows that with an electron lifetime of 1 ms, the hit RMS metric is very inaccurate, this is likely due to hits which are a large distance away from the APAs being very difficult to reconstruct due to the extremely poor lifetime. For this reason, the accuracy to which the hit RMS metric predicts the interaction time improves as the electron lifetime increases, though this increase is small between the 3 ms, 5 ms and 8 ms samples. Figure 6.28 shows the opposite effect, the accuracy to which the interaction time can be determined decreases with increasing electron lifetime for the hit $RMS/Charge$ metric. This is because the decrease in hit charge is much greater when the electron lifetime is lower. This dependence is the corner stone of this metric, which

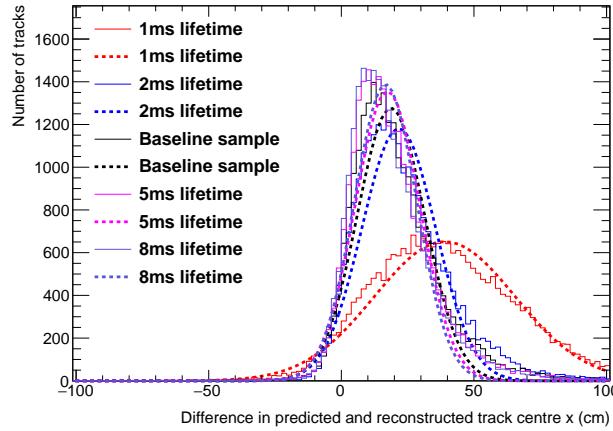
6.5 Measuring interaction times using electron diffusion

- ¹ is why it performs so well for low electron lifetimes, and so the decrease in its accuracy is an
² unavoidable consequence of increasing electron lifetime.

³



(a) The average difference in interaction times using the hit *RMS* metric.

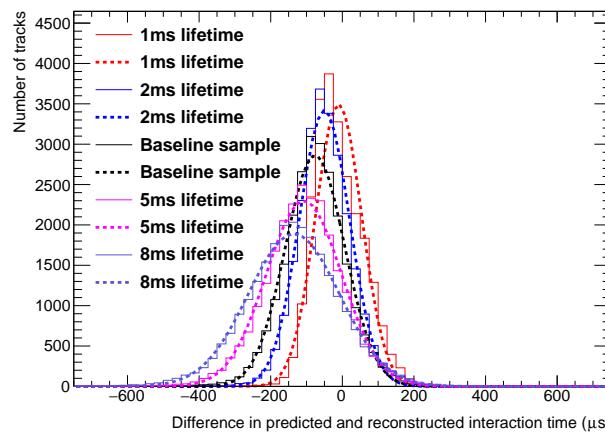


(b) The average difference in the central *x* position of a track using the hit *RMS* metric.

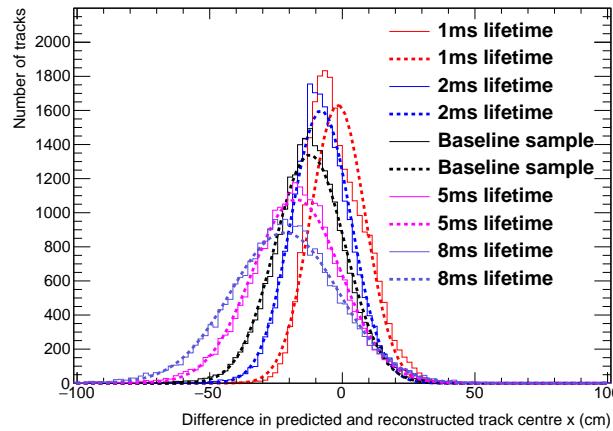
Fig. 6.27 The accuracy of the hit *RMS* method, for changing values of the electron lifetime.

- ⁴ Figures 6.29 and 6.30 show the accuracy to which the interaction time and central *x*
⁵ position of a track can be determined using the effect that diffusion has on the hit *RMS* and
⁶ hit *RMS/Charge* for changing electric field values respectively. Figures 6.29a and ?? show
⁷ that the accuracy to which the interaction time can be determined is relatively unaffected
⁸ by the electric field for both the hit *RMS* and hit *RMS/Charge* methods. However, when
⁹ these interaction times are used to predict the central *x* position the larger drift velocity in the

6.5 Measuring interaction times using electron diffusion



(a) The average difference in interaction times using the hit *RMS* metric.



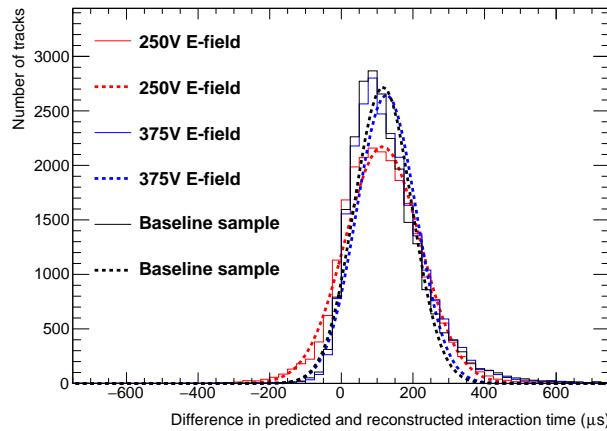
(b) The average difference in the central *x* position of a track using the hit *RMS/Charge* metric.

Fig. 6.28 The accuracy of the hit *RMS/Charge* method, for changing values of the electron lifetime.

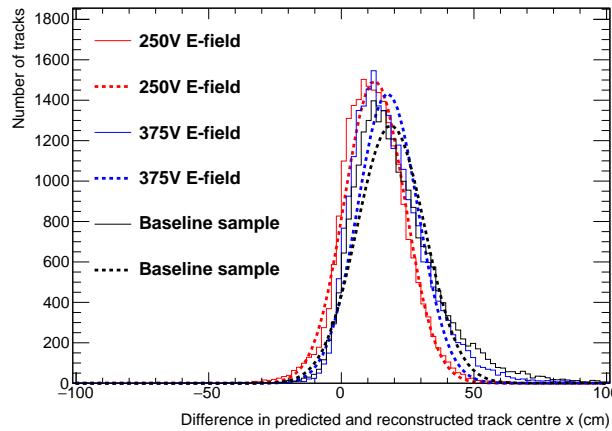
higher electric field samples causes the accuracy to decrease. This decrease in accuracy is because within the same time period electrons will have drifted further in a higher drift field, meaning that the same error in time will produce a larger *x* offset when compared a lower electric field. The increased field also causes the distributions of the difference in predicted and reconstructed central *x* position to become wider, particularly in Figure 6.30b, again this is due to the increase drift velocity as the widths of Figure 6.30a are consistent.

1
2
3
4
5
6
7

6.5 Measuring interaction times using electron diffusion



(a) The average difference in interaction times using the hit *RMS* metric.

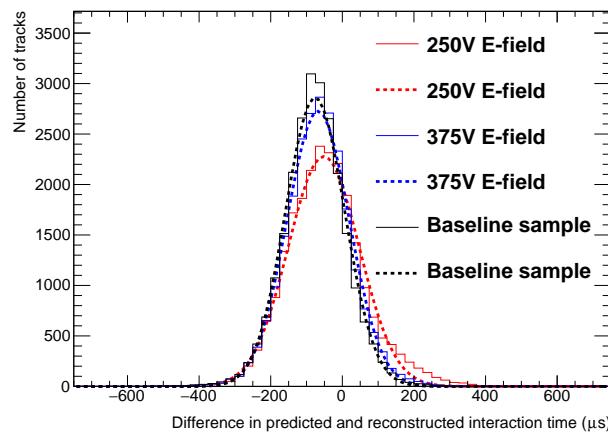


(b) The average difference in the central x position of a track using the hit *RMS* metric.

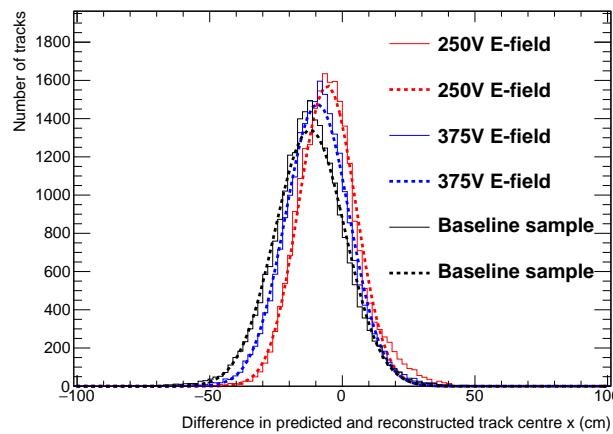
Fig. 6.29 The accuracy of the hit *RMS* method, for changing values of the electric field.

¹ Figures 6.31 and 6.32 show the accuracy to which the interaction time and central x
² position of a track can be determined using the effect that diffusion has on the hit *RMS* and
³ hit *RMS/Charge* for changing values of the longitudinal diffusion constant respectively. As
⁴ would be expected, both figures show that the accuracy to which the interaction time and
⁵ central x position can be predicted are highly dependant on the longitudinal diffusion constant.
⁶ It is interesting to note that the extremely poor resolution when there is no longitudinal seen
⁷ in Figure 6.31 is not present in Figure 6.32. It is thought that this is due to charge attenuation
⁸ which will still occur due to the finite electron lifetime.

6.5 Measuring interaction times using electron diffusion



(a) The average difference in interaction times using the hit *RMS* metric.



(b) The average difference in the central *x* position of a track using the hit *RMS/Charge* metric.

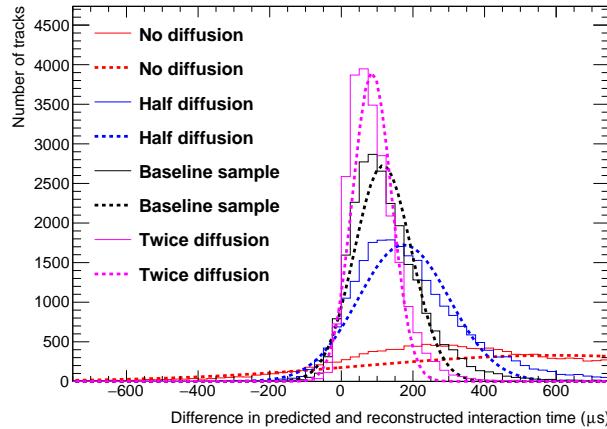
Fig. 6.30 The accuracy of the hit *RMS/Charge* method, for changing values of the electric field.

6.5.4 The limitations of and future improvements to the method of interaction time determination using diffusion

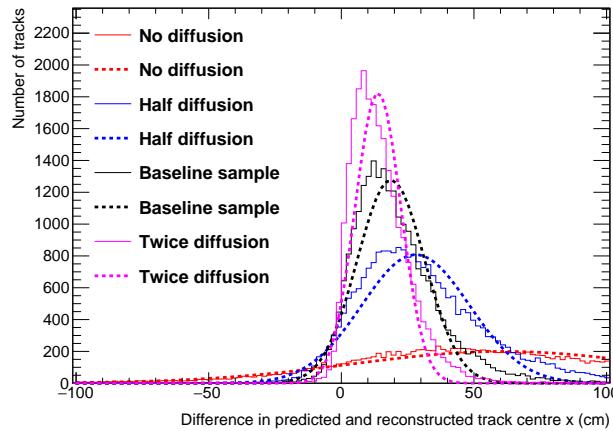
The comparison of the 35 ton data and Monte Carlo samples, as well as the Monte Carlo samples with differing detector conditions, show that there is potential in the ability to determine interaction times using the effects of diffusion. However, there are still some issues which need to be overcome. These will be discussed briefly below.

1
2
3
4
5
6
7

6.5 Measuring interaction times using electron diffusion



(a) The average difference in interaction times using the hit *RMS* metric.

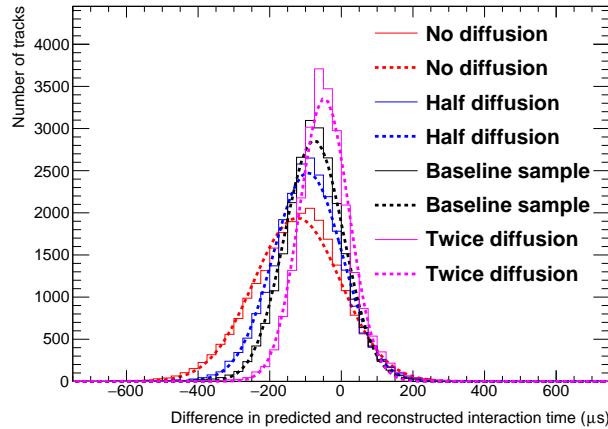


(b) The average difference in the central *x* position of a track using the hit *RMS* metric.

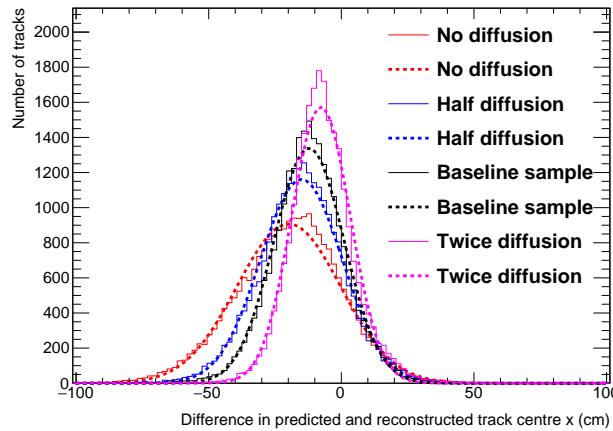
Fig. 6.31 The accuracy of the hit *RMS* method, for changing values of the constant of longitudinal diffusion.

All of the figures shown that had the difference in predicted and reconstructed interaction times were not centered around 0 μ s and also had large FWHMs. It is thought that this is due to interpreting distributions which are not Gaussian as Gaussian functions, such as Figure 6.22. This means that when the MPVs are calculated using only the values around the peaks, there are large tails at both high and low values which are not counted. As the hits which make up these tails are still used to predict the interaction times they will introduce the observed offsets as the times which they would predict will be far from that which the MPV would predict. The result of this is that the assumption made earlier, that over a large

6.5 Measuring interaction times using electron diffusion



(a) The average difference in interaction times using the hit *RMS* metric.



(b) The average difference in the central *x* position of a track using the hit *RMS/Charge* metric.

Fig. 6.32 The accuracy of the hit *RMS/Charge* method, for changing values of the constant of longitudinal diffusion.

number of hits the average predicted interaction time will be correct, could no longer hold. A potential solution to remedy this, is to refine the selection of hits that are used to predict the interaction time to exclude the extreme values of these tails. However, to do this accurately would require some knowledge of the hit location, as the MPV of hit *RMS* can change significantly over the drift length of the detector. This is shown in Figure 6.23, where the MPV of hit *RMS* in Monte Carlo changes from roughly 1.9 ticks near the APAs to roughly 2.8 ticks 200 cm away from the APAs. Looking at Figure 6.22a, this range encompasses the entire distribution of Monte Carlo hit *RMS* values, including the tail of hit *RMS* above 2.3

1
2
3
4
5
6
7
8

6.5 Measuring interaction times using electron diffusion**88**

¹ ticks which we would like to remove. Another solution to improving the accuracy of the
² predicted interaction time could be to weight the predicted interaction times from each hit in
³ some way which better represents the distribution of hit *RMS* and hit *RMS/Charge* seen in
⁴ Figure 6.22.

⁵

⁶ An important improvement to the method would be to expand it to include the induction
⁷ plane wires, as this will greatly increase both the number of wires which can be used, and
⁸ the range of track angles whose interaction times can be predicted. The angular range of
⁹ the method would increase as, when using only collection plane wires it is impossible to
¹⁰ reconstruct enough hits for nearly vertical muons as very few wires would be hit, this was
¹¹ discussed in Section 5.3. This was not attempted here as the electronics noise in the 35 ton
¹² data was too large to able to reliably reconstruct all hits on the induction planes without
¹³ reconstructing many noise hits, and so the hit threshold was very high.

¹⁴

Chapter 7

Simulations of the DUNE Far Detector

Previous work presented has been done concerning the 35 ton prototype, however it is also important to simulate the DUNE Far Detector (FD). Simulations in the FD have concentrated on cosmogenic background to neutrino oscillations, in Section 7.1, and the muon background to nucleon decay, in Section 7.3. The simulations shown in Section 7.1 are discussed in!!!! citepMartinsThesis!!!!, and were performed for the Long Baseline Neutrino Experiment (LBNE) which along with the Long Baseline Neutrino Oscillation (LBNO) experiment formed the basis for DUNE and so are included here for completeness. The other work presented was performed for the DUNE collaboration in conjunction with work done by Vityal Kudryavtsev and Matthew Robinson, both of the University of Sheffield.

7.1 Simulations of the LBNE surface detector

7.2 The use of MUSUN in LArSoft

The primary muons in the following discussions are all generated using MUSIC [8] [10] [11] and MUSUN [8] [9], and so a brief overview of them is required. MUSIC first propagates muons through a medium defined by the user for given initial energies, positions and direction cosines. A range of energies between 10^2 and 10^7 GeV are considered and their energy distributions are stored at depths of between 100 and 15,000 m w.e. Energy losses due to four processes are considered; ionisation, bremsstrahlung, electron-positron pair production and muon-nucleus inelastic scattering. The output of MUSIC is then used by MUSUN to generate a muon energy spectrum and angular distribution for a given detector location given

7.2 The use of MUSUN in LArSoft

90

¹ details about the local surface profile.

²

³ The location of the DUNE detector near the Ross shaft at SURF has global coordinates
⁴ of: latitude = $44^{\circ}20'45.21''$ N and longitude $103^{\circ}45'16.13''$ West. The rock composition
⁵ is assumed to be $\langle Z \rangle = 12.09$ and $\langle A \rangle = 24.17$, and the density is assumed to be 2.70
⁶ g cm⁻³ [38]. The flux calculated by MUSIC/MUSUN of 5.18×10^{-9} cm⁻² s⁻¹ sr⁻¹ is
⁷ well matched to the flux measured by the active veto system of the Davis' experiment which
⁸ was $(5.38 \pm 0.07) \times 10^{-9}$ cm⁻² s⁻¹ sr⁻¹ [39]. Given small differences in these values and
⁹ another measurement by the Majorana demonstrator, the systematic uncertainty in calculating
¹⁰ the muon flux is estimated to be 20% [40].

¹¹

¹² The surface profile around the proposed detector location is shown in Figure 7.1a, where
¹³ the proposed location is in the centre of the map. Each quadrant on the map has been
¹⁴ divided into 4 angles of 22° to help quide the eye when comparing to Figure 7.1b, where the
¹⁵ distribution of azimuth angles is plotted. The vertical lines in Figure 7.1b show the division
¹⁶ of the quadrants when the angle is calcualted from East. When moving from East to North it
¹⁷ is possible to discern how the peaks and troughs on the surface profile correspond to troughs
¹⁸ and peaks in the distribution of azimuthal angle.

¹⁹

²⁰ Given these parameters the muon flux when assuming a spherical detector geometry
²¹ without simulating a detector cavern is given by Table 7.1.

Table 7.1 Muon flux parameters as calculated with MUSIC/MUSUN.

Total flux (cm ⁻² s ⁻¹)	Mean E _μ (GeV)	Mean slant depth (m w.e)	Mean θ (°)
5.66×10^{-9}	283	4532	26

²²

²³ The muons simulated for DUNE are sampled on a the surface of a box surrounding the
²⁴ detector hall that also encompassed 7 m of rock above the cavern, and 5 m of rock on all
²⁵ other sides. This is to ensure that there is sufficient rock to induce cascades both above and
²⁶ around the detector hall, as it is mainly the secondaries produced in these interactions that
²⁷ enter the detector which are of concern to nucleon decay searches. This will be discussed
²⁸ in Section 7.3. The size of the box the muons are sampled from is $74.43 \times 29.54 \times 30.18$
²⁹ m³, compared to the simulated cryostat that has dimensions of $61.62 \times 14.94 \times 13.58$ m³,
³⁰ where these dimensions are (length × width × height). The muons are sampled randomly
³¹ according to their energy spectrum for a given zenith and azimuthal angle, using the angular

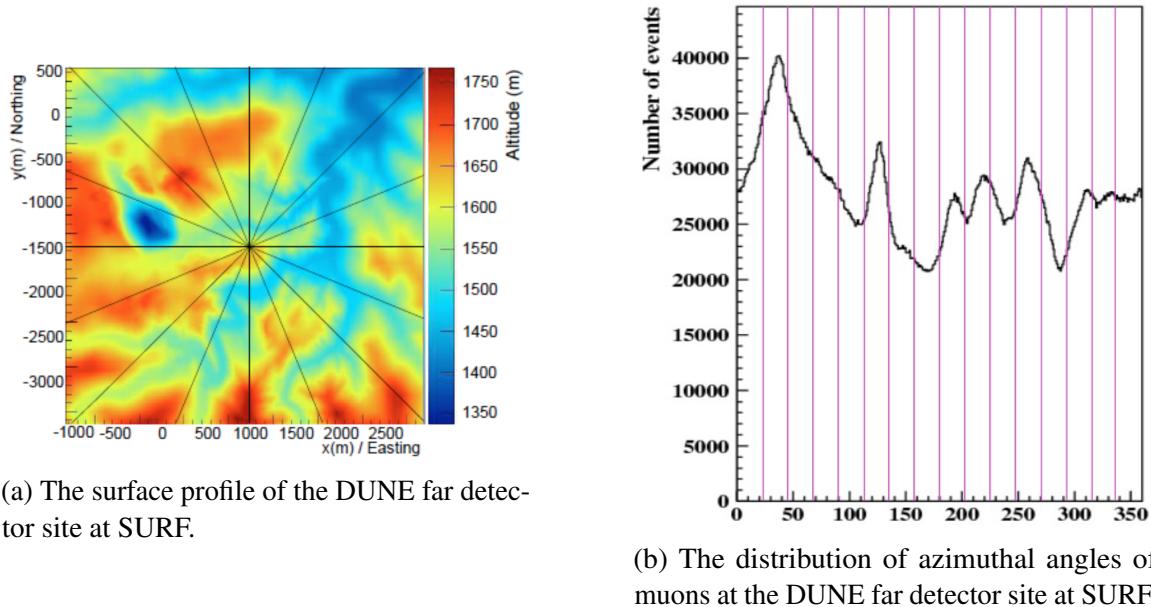


Fig. 7.1 The correlation between the surface profile and distribution of azimuthal angles at the DUNE far detector site. The quadrants have been divided into four angles of equal size. The azimuthal angle, calculated as the angle from East (pointing to the right in Fig. 7.1a), and increasing counterclockwise, is seen to follow the contours of the surface profile.

distribution obtained with MUSIC.

Before this could be done however, MUSUN had to be incorporated into the DUNE software framework as it has previously been maintained in FORTRAN as an external package. This involved building on the work done by the LZ collaboration in porting the code to C++ !!!!!citepKareem. The process by which this was done was to first reproduce the distributions produced by the LZ collaboration with the DUNE software framework using the LZ detector location, and then reproducing the muon distributions produced by the original FORTRAN code for the DUNE detector location. The distributions produced by the DUNE software framework are shown in Figure 7.2, these are seen to be consistent with the same distributions shown in [41]. The initial positions of 10,000 muons generated in LArSoft around the DUNE 10 kt module that is simulated is shown in Figure 7.3. The initial positions of the muons are shown as blue points, whilst the cryostat is a single black box and each TPC is a single red box.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15

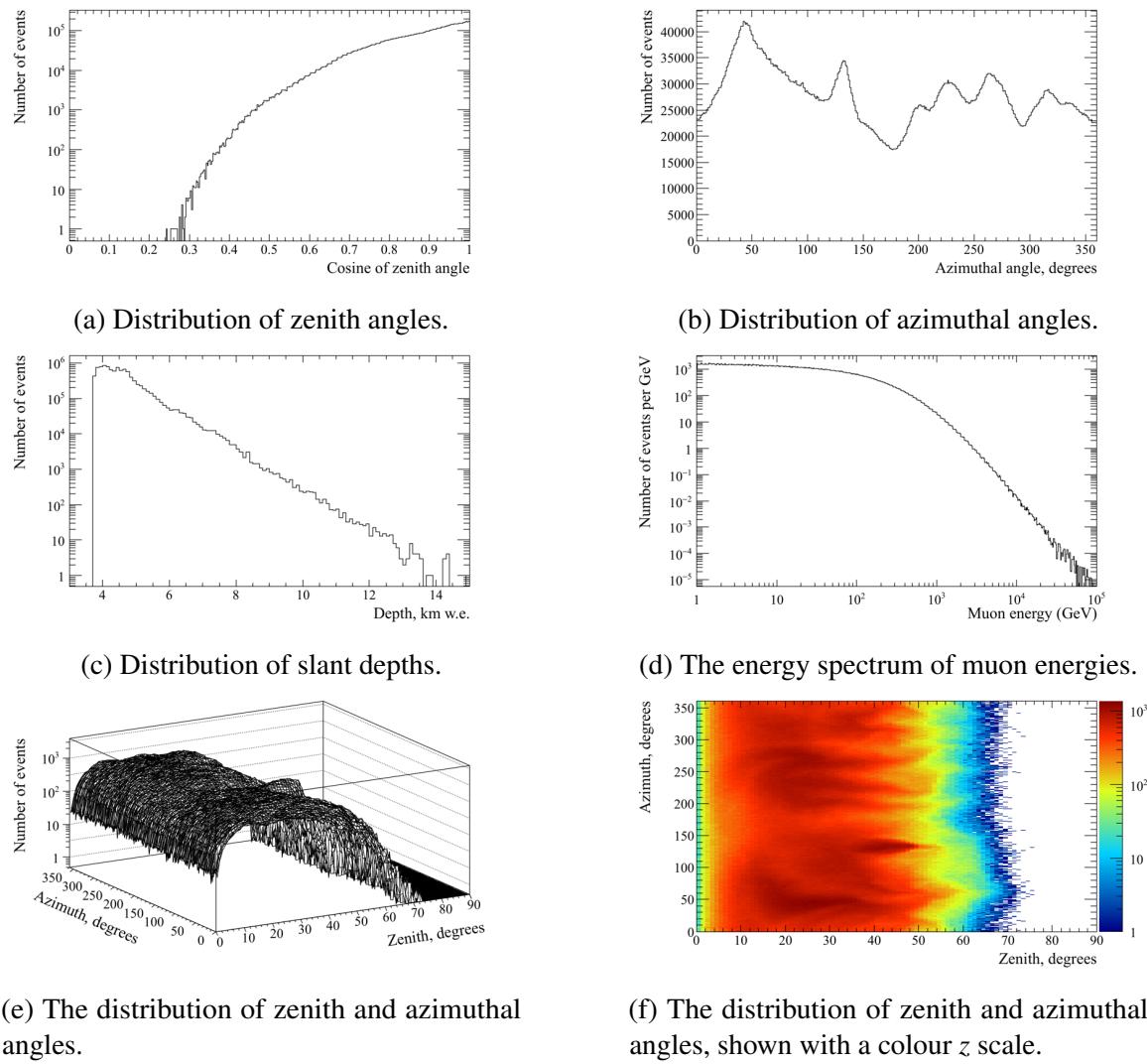


Fig. 7.2 The distributions of some of the important quantities for muons generated by MUSUN in LArSoft.

¹ 7.3 Cosmogenic background for nucleon decay channels in ² DUNE

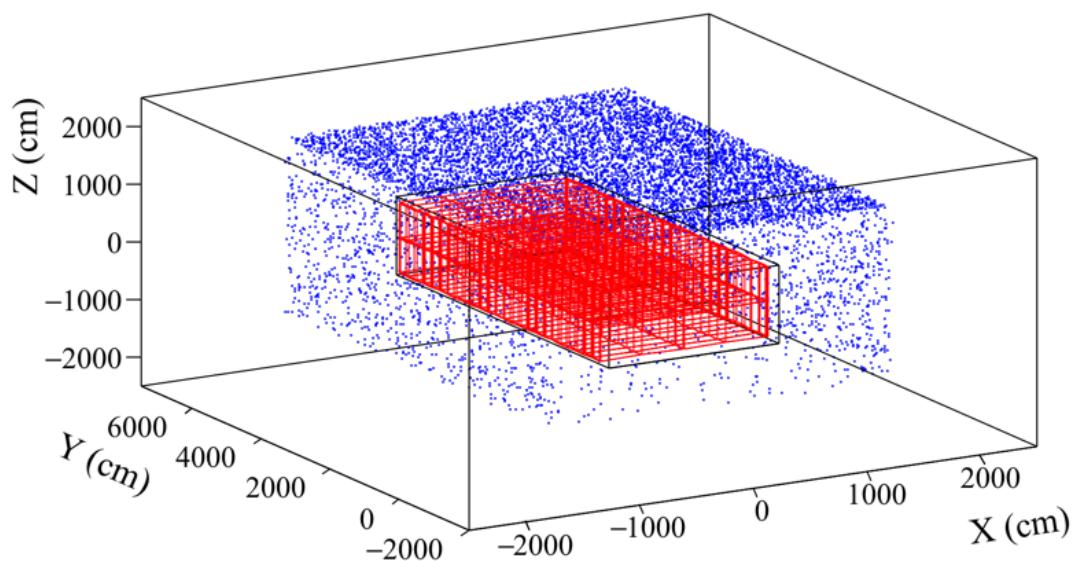


Fig. 7.3 The initial positions of muons generated by MUSUN around a DUNE 10 kt module. The initial positions of the muons are shown as blue points, whilst the cryostat is a single black box and each TPC is a single red box.

1 References

- 2 [1] E. D. Church, “LArSoft: A Software Package for Liquid Argon Time Projection Drift
3 Chambers,” 2013.
- 4 [2] “Larsoft,” 2017. [online] <http://larsoft.org>.
- 5 [3] R. Brun and F. Rademakers, “ROOT: An object oriented data analysis framework,”
6 *Nucl. Instrum. Meth.*, vol. A389, pp. 81–86, 1997.
- 7 [4] S. Agostinelli *et al.*, “Geant4—a simulation toolkit,” *Nuclear Instruments and Methods*
8 *in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated*
9 *Equipment*, vol. 506, no. 3, pp. 250 – 303, 2003.
- 10 [5] C. Andreopoulos *et al.*, “The {GENIE} neutrino monte carlo generator,” *Nuclear*
11 *Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers,*
12 *Detectors and Associated Equipment*, vol. 614, no. 1, pp. 87 – 104, 2010.
- 13 [6] D. Casper, “The Nuance neutrino physics simulation, and the future,” *Nucl. Phys. Proc.*
14 *Suppl.*, vol. 112, pp. 161–170, 2002. [,161(2002)].
- 15 [7] “Cosmic-ray shower library {CRY},” 2012. [online] http://nuclear.llnl.gov/simulation/doc_cry_v1.7/cry.pdf.
- 16 [8] V. Kudryavtsev, “Muon simulation codes {MUSIC} and {MUSUN} for underground
physics,” *Computer Physics Communications*, vol. 180, no. 3, pp. 339 – 346, 2009.
- 17 [9] V. A. Kudryavtsev, N. J. C. Spooner, and J. E. McMillan, “Simulations of muon induced
18 neutron flux at large depths underground,” *Nucl. Instrum. Meth.*, vol. A505, pp. 688–698,
21 2003.
- 22 [10] P. Antonioli, C. Ghetti, E. V. Korolkova, V. A. Kudryavtsev, and G. Sartorelli, “A
23 Three-dimensional code for muon propagation through the rock: Music,” *Astropart.*
24 *Phys.*, vol. 7, pp. 357–368, 1997.
- 25 [11] V. A. Kudryavtsev, E. V. Korolkova, and N. J. C. Spooner, “Narrow muon bundles from
26 muon pair production in rock,” *Phys. Lett.*, vol. B471, pp. 251–256, 1999.
- 27 [12] “The current status of simulation and reconstruction for the dune far detector,” 2016.
28 [online] <http://docs.dunescience.org:8080/cgi-bin>ShowDocument?docid=1689>.
- 29 [13] C. Harris and M. Stephens, “A combined corner and edge detector.” Citeseer, 1988.

-
- [14] P. V. C. Hough, “Machine Analysis Of Bubble Chamber Pictures,” in *Proceedings, 2nd International Conference on High-Energy Accelerators and Instrumentation, HEACC 1959: CERN, Geneva, Switzerland, September 14-19, 1959*, vol. C590914, pp. 554–558, 1959. 1
2
3
4
- [15] M. Antonello *et al.*, “Precise 3D track reconstruction algorithm for the ICARUS T600 liquid argon time projection chamber detector,” *Adv. High Energy Phys.*, vol. 2013, p. 260820, 2013. 5
6
7
- [16] J. S. Marshall and M. A. Thomson, “The Pandora Software Development Kit for Pattern Recognition,” *Eur. Phys. J.*, vol. C75, no. 9, p. 439, 2015. 8
9
- [17] “Emshower reconstruction,” 2016. [online] <http://docs.dunescience.org:8080/cgi-bin/ShowDocument?docid=1369>. 10
11
- [18] J. Birks, “The efficiency of organic scintillatois,” *ORGANIC*, p. 12, 1960. 12
- [19] R. Acciarri *et al.*, “A study of electron recombination using highly ionizing particles in the ArgoNeuT Liquid Argon TPC,” *JINST*, vol. 8, p. P08005, 2013. 13
14
- [20] J. Thomas and D. A. Imel, “Recombination of electron-ion pairs in liquid argon and liquid xenon,” *Phys. Rev. A*, vol. 36, pp. 614–616, Jul 1987. 15
16
- [21] H. Bethe, “Zur theorie des durchgangs schneller korpuskularstrahlen durch materie,” *Annalen der Physik*, vol. 397, no. 3, pp. 325–400, 1930. 17
18
- [22] F. Bloch, “Zur bremsung rasch bewegter teilchen beim durchgang durch materie,” *Annalen der Physik*, vol. 408, no. 3, pp. 285–320, 1933. 19
20
- [23] “Module to unstick adc codes,” 2016. [online] <https://indico.fnal.gov/getFile.py/access?contribId=2&resId=0&materialId=slides&confId=11627>. 21
22
- [24] “Noise characterization and filtering in the microboone tpc,” 2016. [online] <http://www-microboone.fnal.gov/publications/publicnotes/MICROBOONE-NOTE-1016-PUB.pdf>. 23
24
25
- [25] J. Cooley and J. Turkey, “An algorithm for the machine calculation of complex fourier series,” *Math. Comp.*, vol. 19, pp. 297–301, 1965. 26
27
- [26] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. [Cambridge]: Technology Press of the Massachusetts Institute of Technology., 1942. 28
29
- [27] “35t observations and measurements,” 2016. [online] <http://docs.dunescience.org:8080/cgi-bin/ShowDocument?docid=1704>. 30
31
- [28] “35t noise studies summary - apa wire readout noise,” 2016. [online] <http://docs.dunescience.org:8080/cgi-bin/ShowDocument>. 32
33
- [29] “Update on reconstruction of 35t data,” 2016. [online] <https://indico.fnal.gov/getFile.py/access?contribId=1&resId=0&materialId=slides&confId=12349>. 34
35

References

96

- ¹ [30] P. Torr and A. Zisserman, “Mlesac: A new robust estimator with application to estimating image geometry,” *Computer Vision and Image Understanding*, vol. 78, no. 1, pp. 138–156, 2000.
- ⁴ [31] “Something,” 2016. [online] /urlSOMETHING.
- ⁵ [32] C. Anderson *et al.*, “The ArgoNeuT Detector in the NuMI Low-Energy beam line at Fermilab,” *JINST*, vol. 7, p. P10019, 2012.
- ⁷ [33] “Update on 35 ton analysis,” 2016. [online] <https://indico.fnal.gov/getFile.py/access?contribId=1&resId=0&materialId=slides&confId=12396>.
- ⁹ [34] V. M. Atrazhev and I. V. Timoshkin, “Trasport of electrons in atomic liquids in high electric fields,” *IEEE Trans. Dielectrics and Electrical Insulation*, vol. 5, pp. 450–457, 1998.
- ¹² [35] Y. Li *et al.*, “Measurement of Longitudinal Electron Diffusion in Liquid Argon,” *Nucl. Instrum. Meth.*, vol. A816, pp. 160–170, 2016.
- ¹⁴ [36] “Looking at electron diffusion in 35t data and mc,” 2016. [online] <https://indico.fnal.gov/getFile.py/access?contribId=43&sessionId=19&resId=0&materialId=slides&confId=10613>.
- ¹⁷ [37] C. Bromberg *et al.*, “Design and operation of longbo: a 2 m long drift liquid argon tpc,” *Journal of Instrumentation*, vol. 10, no. 07, p. P07015, 2015.
- ¹⁹ [38] D. M. Mei, C. Zhang, K. Thomas, and F. Gray, “Early Results on Radioactive Background Characterization for Sanford Laboratory and DUSEL Experiments,” *Astropart. Phys.*, vol. 34, pp. 33–39, 2010.
- ²² [39] M. L. Cherry, M. Deakyne, K. Lande, C. K. Lee, R. I. Steinberg, B. Cleveland, and E. J. Fenyves, “Multiple muons in the homestake underground detector,” *Phys. Rev. D*, vol. 27, pp. 1444–1447, Apr 1983.
- ²⁵ [40] “September 2016 progress report from the atmospheric neutrinos, nucleon decay, and cosmogenics working groups,” 2016. [online] <http://docs.dunescience.org:8080/cgi-bin>ShowDocument?docid=1644>.
- ²⁸ [41] “Muon simulations for lbne using music and musun,” 2014. [online] <http://lbne2-docdb.fnal.gov/cgi-bin>ShowDocument?docid=9673>.

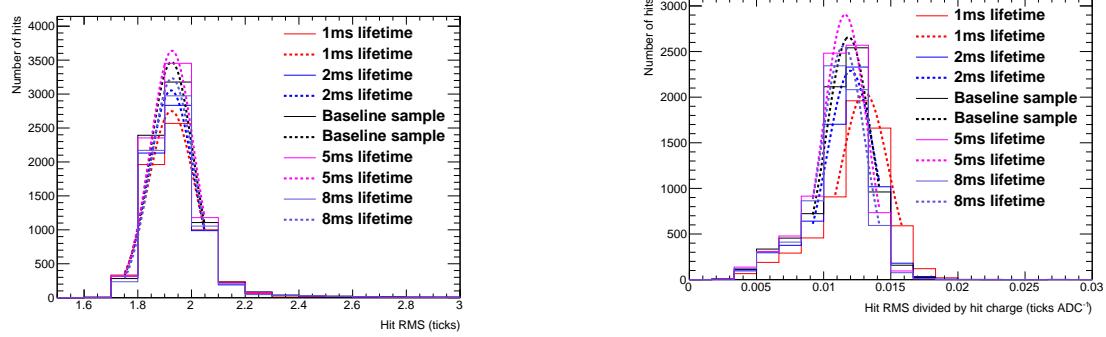
Appendix A

Supporting figures to Monte Carlo studies concerning determining interaction times using the effects of diffusion

Figure A.1, shows how the most probable values of the hit *RMS* and hit *RMS/Charge* change as the electron lifetime increases, for hits between 20 and 30 cm from the APAs. Figure A.2, shows how the most probable values of hit *RMS* changes as drift distance increases for track associated with counter differences of 4, for different values of the electron lifetime. Figure A.3, shows how the most probable value of hit *RMS* next to the APAs changes for increasing counter difference.

Figure A.4, shows how the most probable values of the hit *RMS* and hit *RMS/Charge* change as the electric field increases, for hits between 20 and 30 cm from the APAs. Figure A.5, shows how the most probable values of hit *RMS* changes as drift distance increases for track associated with counter differences of 4, for different values of the electric field. Figure A.6, shows how the most probable value of hit *RMS* next to the APAs changes for increasing counter difference.

Figure A.7, shows how the most probable values of the hit *RMS* and hit *RMS/Charge* change as the constant of longitudinal diffusion increases, for hits between 20 and 30 cm from the APAs. Figure A.8, shows how the most probable values of hit *RMS* changes as drift distance increases for track associated with counter differences of 4, for different values of the constant of longitudinal diffusion. Figure A.9, shows how the most probable value of hit



(a) The most probable hit *RMS* values for hits between $x = 20$ cm and $x = 30$ cm.

(b) The most probably hit *RMS/Charge* values for hits between $x = 20$ cm and $x = 30$ cm.

Fig. A.1 The most probable values of the hit *RMS* and hit *RMS/Charge* distributions for hits between $x = 20$ cm and $x = 30$ cm, for tracks with a counter difference of 4, as the electron lifetime changes.

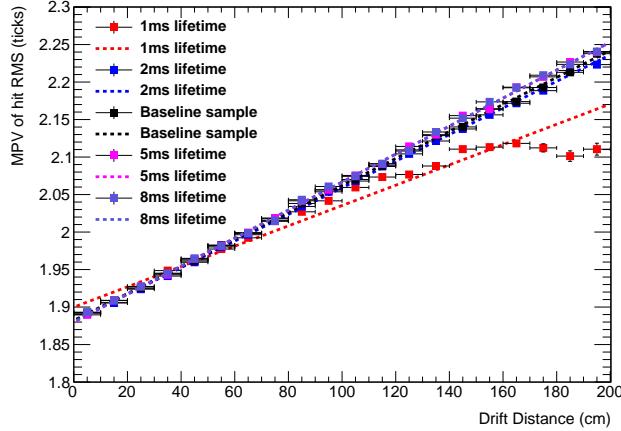


Fig. A.2 The most probable values of hit *RMS* as a function of drift distance, for tracks associated with a coincidence that had a counter difference of 4, as the electron lifetime changes.

² *RMS* next to the APAs changes for increasing counter difference.

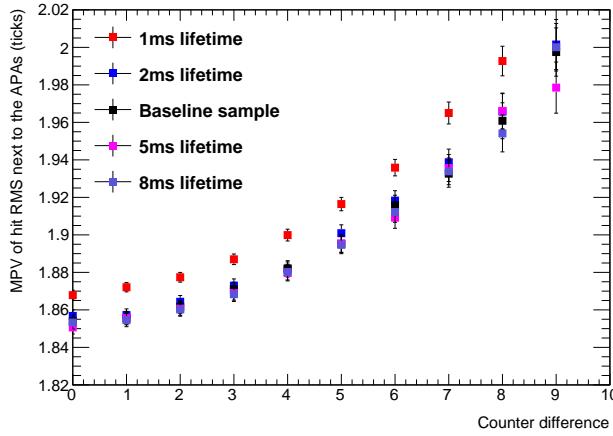
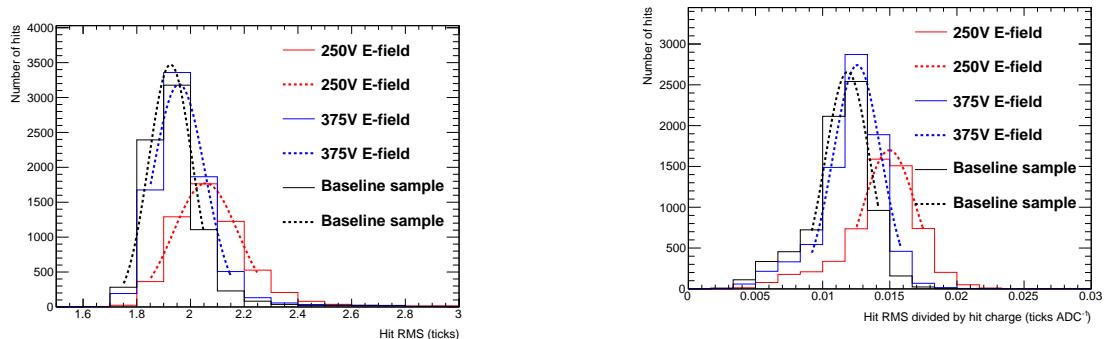


Fig. A.3 The most probable values of hit RMS within 10 cm of the APAs, as a function of the counter difference of the coincidence, that the track was associated with, as the electron lifetime changes.



(a) The most probable hit RMS values for hits between $x = 20$ cm and $x = 30$ cm.

(b) The most probable hit $RMS/Charge$ values for hits between $x = 20$ cm and $x = 30$ cm.

Fig. A.4 The most probable values of the hit RMS and hit $RMS/Charge$ distributions for hits between $x = 20$ cm and $x = 30$ cm, for tracks with a counter difference of 4, as the electric field changes.

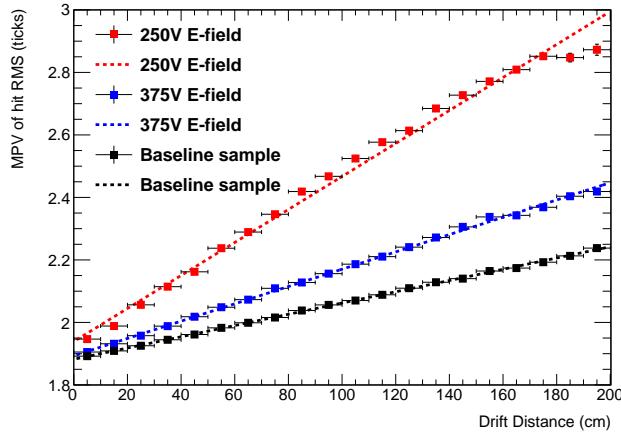


Fig. A.5 The most probable values of hit *RMS* as a function of drift distance, for tracks associated with a coincidence that had a counter difference of 4, as the electric field changes.

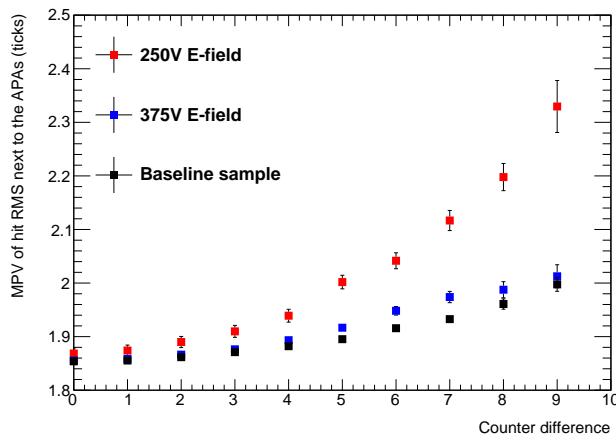
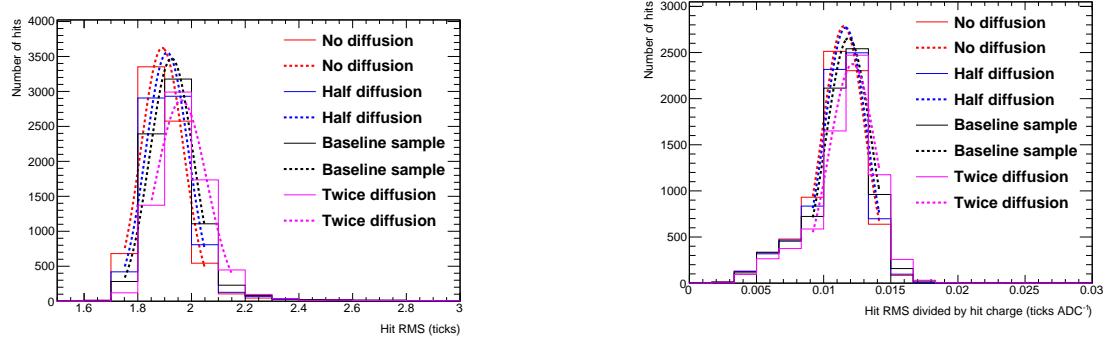


Fig. A.6 The most probable values of hit *RMS* within 10 cm of the APAs, as a function of the counter difference of the coincidence, that the track was associated with, as the electric field changes.



(a) The most probable hit *RMS* values for hits between $x = 20$ cm and $x = 30$ cm.

(b) The most probably hit *RMS/Charge* values for hits between $x = 20$ cm and $x = 30$ cm.

Fig. A.7 The most probable values of the hit *RMS* and hit *RMS/Charge* distributions for hits between $x = 20$ cm and $x = 30$ cm, for tracks with a counter difference of 4, as the constant of longitudinal diffusion changes.

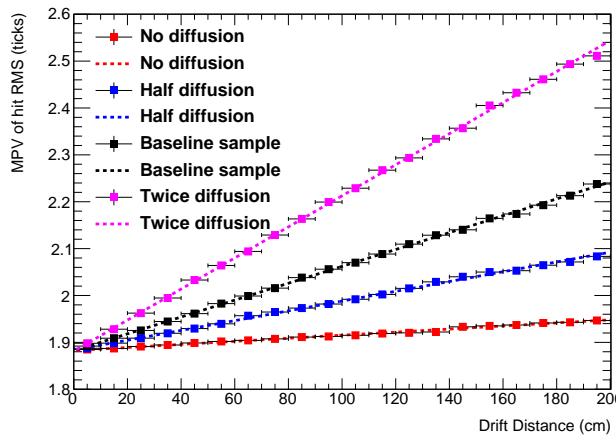


Fig. A.8 The most probable values of hit *RMS* as a function of drift distance, for tracks associated with a coincidence that had a counter difference of 4, as the constant of longitudinal diffusion changes.

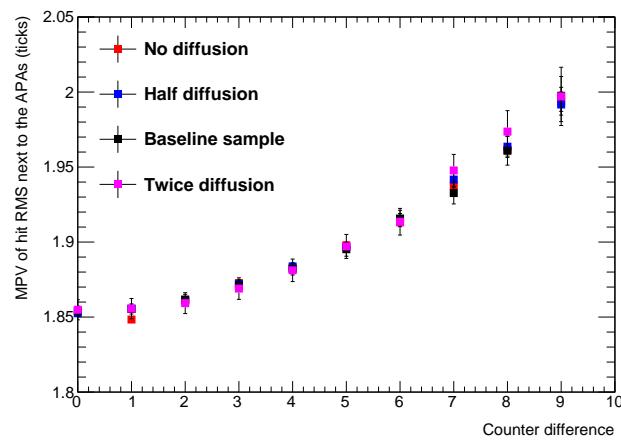


Fig. A.9 The most probable values of hit *RMS* within 10 cm of the APAs, as a function of the counter difference of the coincidence, that the track was associated with, as the constant of longitudinal diffusion changes.

Appendix B

2

Something else mildly interesting

1490