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   December 2016

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
December 2016

Acknowledgements

And I would like to acknowledge ...

Abstract

This is where you write your abstract ...

Table of contents

List of figures

List of tables

Nomenclature

Roman Symbols

tick   Unit of time equal to 500 ns

Acronyms / Abbreviations

CRC   Cosmic Ray Counter

SSP   SiPM Signal Processor

MIP   Minimally Ionising Particle

MPV   Most Probable Value

PID   Particle IDentification

ADC   Analogue to Digital Converter

SiPM   Silicon Photo Multiplier

TPC   Time Projection Chamber

Chapter 1

Introduction

1.1 XXXXX

1.2 XXXXXX

Chapter 2

1

Theory

2

2.1 Theory of neutrino physics

3

2.2 Nucleon decay in Grand Unifying Theories

4

2.3 Exisiting and future experiments

5

2.4 How Liquid Argone Time Projection Chambers work

6

Chapter 3

The Deep Underground Neutrino Experiment

3.1 DUNE location and beamline

3.2 The DUNE detectors and schedule

3.3 Physics opportunities of DUNE

3.3.1 Neutrino physics

3.3.2 Nucleon decay and supernovae neutrinos

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.

3.5 The DUNE software

The software packed used by DUNE is called LArSoft [11] which is a simulation, reconstruction and analysis package for LArTPCs which is being used by many of the experiments in the US neutrino program. LArSoft has been developed to be detector agnostic, meaning 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 [9] and GEANT4 [2]

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 [4], Nuance [10] and CRY [15] already having been

incorporated. Recently the MUon Simulations UNderground (MUSUN) [17] generator which

takes the output of MUon SImulation Code (MUSIC) [5] has also been incorporated, see

Section 7.3. It is also possible to use an inbuilt single particle generation mode which is fully

tunable as particle type, momenta, positions and directions can all be varied.

The co-ordinates and angles in LArSoft, shown in Figure 3.2, are defined as;

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

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

    - In the far detector geometry $x = 0$ is defined as ....

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

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

    - In the far detector $y = 0$ is ....

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

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

    - In the far detector $z = 0$ is ....

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

- ϕ - The angle between the z axis and the point.
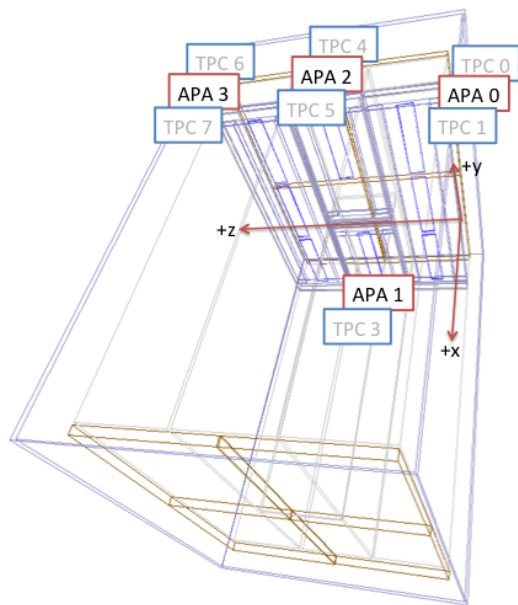
(a) The location of the origin of the 35 ton co-ordinate system in 3D.
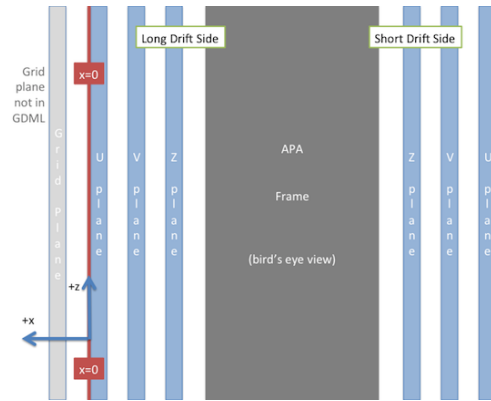
(b) The location of the origin of the 35 ton co-ordinate system in a 2D aerial view.

Fig. 3.2 Schematic of the LArSoft co-ordinate system.

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

- Generation.

- GEANT4.

- 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. 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 does not conserve pulse height and is not gauranteed to preserve the normalisation and is called deconvolution. 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 ditrbutions. For isolated hits this is typically acheived using only one Gaussian distrbution, however for large energy depositions over a large period time where many particles are involved multiple Guassian distributions are often required. Large energy depositions are also possible when the orientation of the particle alligns with a wire, this means that all of the deposited energy is collected on this single wire. The hits reconstructed due to one such deposition on a single wire is shown in Figure 3.3 along with examples of more simple energy depositions on each of the collection and induction planes.

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 occured 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 occured 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 dismabiguated, 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 requried 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 DUNEs scale. The inclination of wires in the 35 ton was $45° \pm 0.7°$ meaning that many wire triplets cross twice and some

(a) Collection plane depositions.

(b) Induction plane depositons.

(c) A large collection plane deposition over a large period of time.
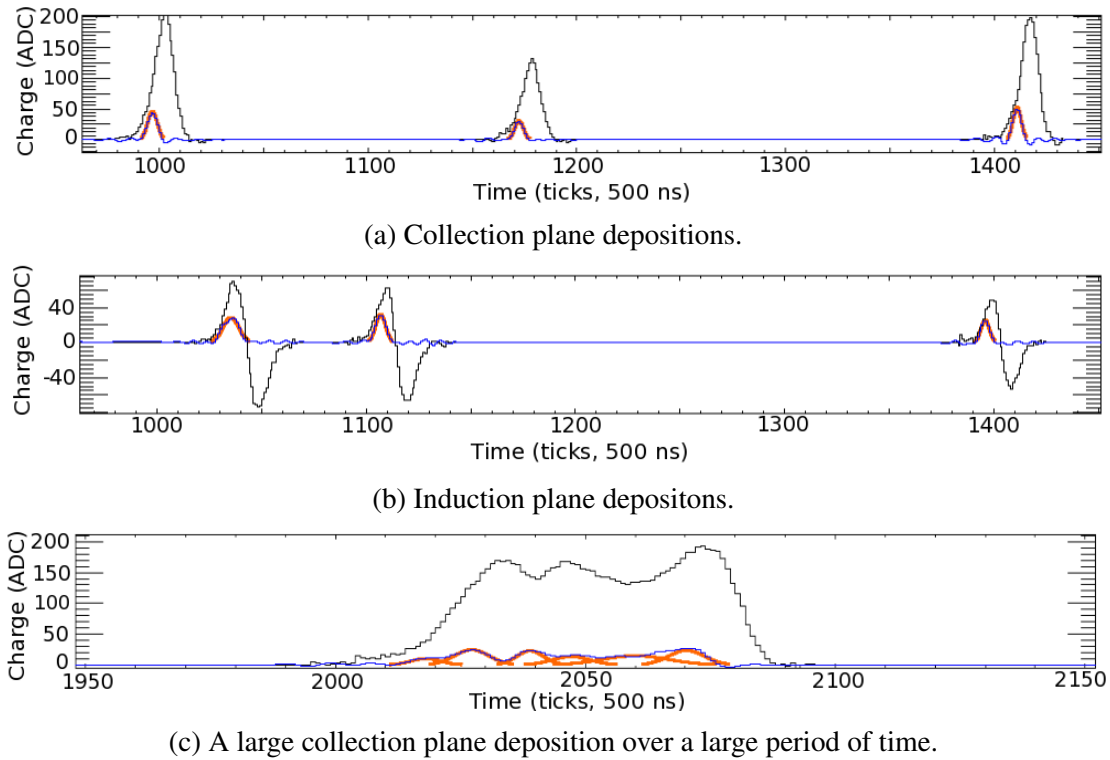
Fig. 3.3 The raw and deconvoluted signals with reconstructed hits on a single wire for a simulated energy deposition. The plot is shown with increasing charge on the *y* axis, and increasing time (in ticks) on the *x* axis. The black line shows the raw signal, the blue line shows the deconvoluted signal and the orange lines show the reconstructed hits.

1   wire pairs cross three times. When wire triplets cross multiple times the triplet which has
2   the smallest distance between the common intersection point and the two-wire intersection
3   points is chosen as the best intersection candidate. The different wire pitches are neccessary
4   so that one of the triple points can be evaluated to be a better candidate, as with a wire pitch
5   of $45°$ both triple points would be equally good fits. The inclination of wires in the FD
6   was chosen to be $36°$ to remove the possibility of multiple intersection points as given the
7   geometry of the APAs multiple intersection points are impossible and so disambiguation
8   is much simpler, but there are more wires on each of the induction planes making it more
9   expensive to instrument. This is shown in Figure 3.4.
10
11   Once the hits have been disambiguated they are combined to make clusters in each of the
12   three planes, before the clusters are merged to make reconstructed tracks or showers. The
13   clustering process is usually performed in wire-tick space on each plane separately with hits
14   from a single physical entitiy being grouped together. It is possible to help seed the start of
15   clusters by using imaging techniques such as a Harris transform, or to identify straight lines
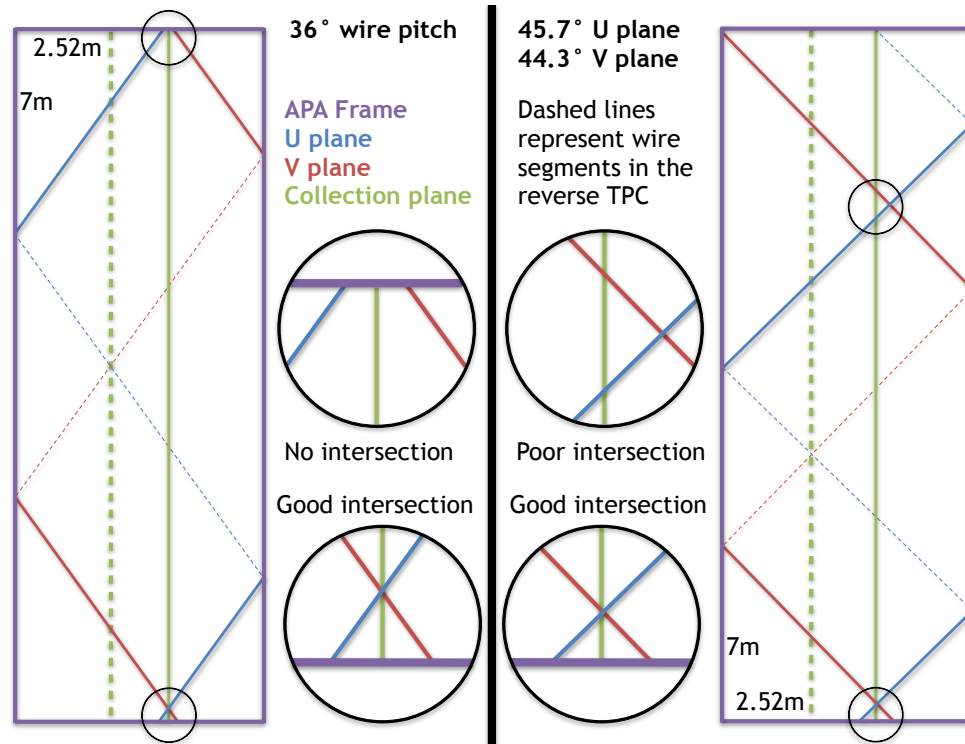
Fig. 3.4 The effect that different wire pitches have on the ability to perform disambiguation in APA with the far detector geometry. The left panel shows a wire pitch of 36°, which is the reference design for the far detector, whilst the right panel shows wire pitches of $45° \pm 0.7°$, as was used in the 35 ton.

by using Hough transforms. 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 PMA-Track!!citepPMATrack!! and Pandora!!citepPandora!!, and the two most common showering algorithms are EMShower3D!!citepEMShower3D!! and EMShower!!citepEMShower!!. 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 LArSoft, Birks model!!citepBirksModel!! and a modified box model!!citepModBox!!, though traditionally the later is used. Both models calculate the $\frac{dE}{dx}$ of a hit using the deposited charge (dQ) and the track pitch (dx) of the hit as well as the conversion of ADC value to number of electrons ($C_{GeV \to e^-}$), the LAr density (ρ), the electric field (E_{field}) and the tunable electron

recombination factors ($Recomb_X$). The series of equations used in Birks model is shown in Equation 3.1a, whilst those used in the modified box are shown in Equation 3.2a.

$$\frac{dE}{dx} = \frac{dQdx_e}{A-B} \tag{3.1a}$$

$$dQdx_e = \frac{dQ \times C_{lifetime}}{dx \times C_{ADC \to e^-}} \tag{3.1b}$$

$$A = \frac{Recomb_A}{C_{GeV \to e^-}} \tag{3.1c}$$

$$B = \frac{\frac{Recomb_B}{\rho}}{E_{field} \times dQdx_e} \tag{3.1d}$$

$$\frac{dE}{dx} = \frac{e^A - Recomb_A}{B} \tag{3.2a}$$

$$A = B \times C_{GeV \to e^-} \times \frac{dQ}{dx} \tag{3.2b}$$

$$B = \frac{Recomb_B}{\rho \times E_{field}} \tag{3.2c}$$

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.

Chapter 4

The 35 ton camera system

4.1 The need for cameras in a Liquid Argon Time Projection Chamber

4.2 Design of the camera system

4.3 Tabletop tests

4.4 Safety reviews and installation

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 distances considered in simulations are small compared to the velocities which the particles are initially travelling meaning that interactions throughout the volume are less than 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 counters but one wold expect almost all of them to produce reconstructable scintillation light. Flashes of lights are reconstructed by using a prebuilt 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

library takes into account the expected quantum efficiencies of each photon detector.

When trying to produce an association metric a sample of 10,000 Anti-Muons with a cosmic-like distribution was used as then there there should only be one long track with which to match one reconstructed flash. A cosmic-like distribution is defined as a set of particles which have a $\cos^2$ angular distribution, no minimum or maximum energies and have a flat distrbution of initial positions in the *xz* plane and a uniform initial *y* position. When this sample was simulated it was clear that the photon detector reconstruction using the prebuilt libraries worked well as the reconstructed flash source normally lay very close to the track which caused it. It was found that a calculation of a Point Of Closest Approach (PoCA)!! citepPoCA!! of the reconstructed track to the flash source gave an effective metric by which the two could be combined. Other metrics such as the distance between the flash source and the track centre, and the perpendicular distance between the flash source and the line joining the start and end of track were investigated but found to provide a less reliable metrics. The latter of these metrics is less effective because the reconstructed tracks are rarely straight lines, due to particles scattering as they travel through the LAr and so the perpendicular distance at each hit must be calculated. A comparison of these metrics is shown in Figure 5.1.
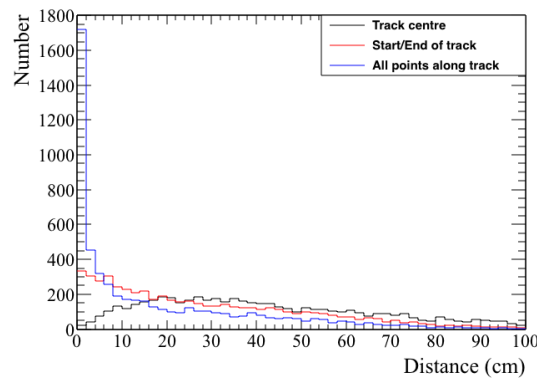
Fig. 5.1 A comparison of *yz* comparisons of reconstructed tracks and flashes in the 35 ton.

Another metric by which flashes could be assigned to reconstructed tracks is by utilising the relationship between the number of measured photoelectrons and the distance from the APAs at which they were produced. When considering two flashes of scintillation light that are produced at different distances from the APAs, it would be expected that more photo-electrons would be collected from the photons produced closer to the APAs. Utilising this relationship, shown in Figure 5.2a, means that the distance from the APAs can be predicted

1 from the number of photoelectrons which are measured. This predicted distance can then
2 be compared to the expected *x* position of a reconstructed track given the difference in flash
3 time and hit times, this is shown in Figure 5.2b. The difference in these two quantities is
4 used as the second metric as it gives an indication of how well a flash properties match the
5 reconstructed *x* position of the track, with a value of 0 representing an excellent match.
6
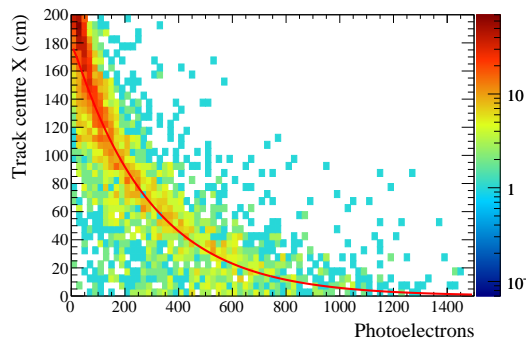
(a) How the number of photoelectrons measured changes with drift distance.
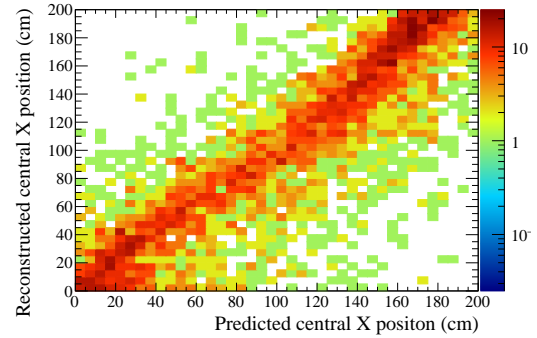
(b) The difference in *x* position using the relationship in Fig 5.2a and the difference in flash and hit times.

Fig. 5.2 How the number of reconstructed photoelectrons changes with increasing drift distance, and how this can be used to predict the interaction time of tracks. How consistent the predicted ineraction times using this method replicate the *x* positions one would expect given the drift times they correspond to, there is one entry for every track/flash pair.

7 Using these metrics it is possible to attempt to assign reconstructed flashes to recon-
8 structed tracks. Only flashes which are within one drift window of a given track are consid-
9 ered, as flashes outside of this time window cannot have been caused by the reconstructed
10 track. Once flashes are assigned to tracks it is possible to determine how well the matching
11 has performed by comparing the Monte Carlo truth interaction time with the photon detector
12 interaction time. When doing this it is more useful to use a long (16 ms, 32,000 tick) CRY
13 sample as then particles come at random timings as opposed to all at $T = 0$ as with the
14 Anti-Muon sampe initially considered. This comparison is shown in Figure 5.3, where there
15 is a clear peak at a time difference of 0 *ms* in the Monte Carlo truth and photon detector
16 interaction times. When zooming in on this peak it can be seen that there is a systematic
17 offset of 0.6 μs, this is due to an electronics offset applied in the simulation to the photon
18 detector system.
19

20 From Figure 5.3 it can clearly be seen that the metric using the proximity of the flash
21 centre to the track trajectory yields the best matches. This is likely caused by the large spread
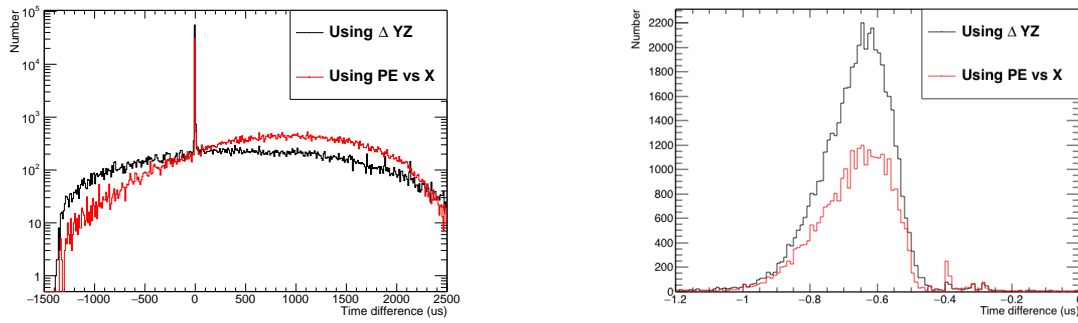
(a) The difference in interaction times.

(b) Zoomed in at low time differences.

Fig. 5.3 The difference in Monte Carlo interaction times and the predicted interaction times using the photon detectors.

in the number of photoelectrons collected at fixed drift distances, as shown by Figure 5.2a. 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 photonelectrons 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 doen correctly are the $Recomb_A$ and $Recomb_B$ of Equations 3.2b and 3.2c respectively. 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 2.1 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 Guassian distribution is then fitted

1   around the peaks for each of the planes to discern the Most Probable Value (MPV) of $\frac{dE}{dx}$ for

2   that plane. If the MPVs are not equal to 2.1 MeV cm^{-3} then the normalisation factors are

3   scaled through a process of trial and error until the correct MPVs are measured. An example

4   of the tuning being applied in shown in Figure 5.4. Tuning of the calorimetric constants is

5   requried whenever the electronics gains or signal shaping functions are changed.
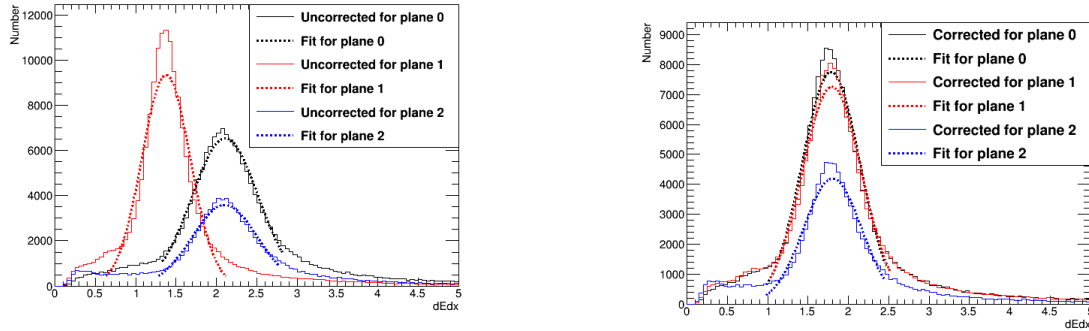
(a) Before a normalisation correction is applied.

(b) After a normalisation correction is applied.

Fig. 5.4 How the $\frac{dE}{dx}$ MPVs change for each plane when a change is made to the electronics
gains in the 35 ton. Figure 5.4a shows the MPVs after the change using the previous constants,
whilst Figure 5.4b shows the MPVs after a retuning of the constants.

6   5.3   Discerning reconstruction efficiencies

7   Knowledge of the strengths and weaknesses of different tracking algorithms is vital when

8   using them for physics analyses, to this end it is useful to develop a metric by which they can

9   be compared. In order to do this a series of conditions have to be applied to the reconstructed

10 tracks from a large set of simulated particles which are reconstructed using different tracking

11 algorithms. It is interesting to observe what the effect of event complexity has on the recon-

12 struction algorithms and so efficiencies will be calculated for both the Anti-Muon and CRY

13 samples used in Section 5.1.

14

15   The critera upon which to determine whether a particle is well reconstructed has to be

16 carefully chosen as every definition will have limitations. For example, consider a particle

17 that travels 100 cm in the active volume of the detector but is reconstructed as 2 separate

18 tracks (tracks 1 and 2), with lengths 77 cm and 23 cm respectively. Firstly, should these

19 tracks should be merged, or left separate? If the reconstruction algorithms have found them to

20 be separate tracks then it would be difficult to ascertain that they are from the sample particle

in real data, and so in considerations here they are not merged. One metric of efficiency would be to consider a track well reconstructed if it has a length between 75% and 125% of the Monte Carlo truth length that the particle traversed in the detector, in which case track 1 would be considered well reconstructed. Another metric however would be to consider a track well reconstructed if the Monte Carlo truth distance the particle traversed in the detector is between 75% and 125% of the reconstructed length, in which case neither track would be considered well matched. Both metrics have used exactly the same tracks and a seemingly identical method of evaluating whether a track is well reconstructed or not, but have got the opposite results. As such it is wrong to say which consideration gives the correct result, but instead the result of each should be considered equally. In discussions here the former definition of efficency will be used, such that a track is considered well reconstructed if:

- Reconstructed track length is more than or equal to 75% of the Monte Carlo track length.

- Reconstructed track length is less than or equal to 125% of the Monte Carlo track length.

- 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 no Monte Carlo information is used and another where the disambiguation and interaction time are cheated. Cheated disambiguation means using the Monte Carlo truth information of the energy despoition 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 dismabiguation 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 5.5, 5.6, 5.7, 5.8 and 5.9. Efficiences 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, 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.
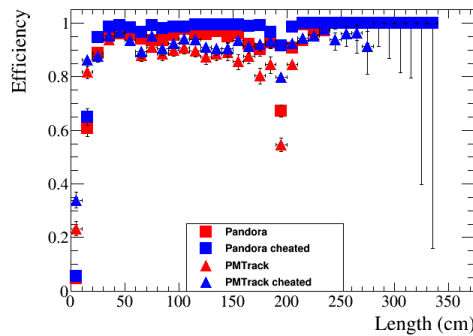
(a) Reconstruction efficiencies for an Anti-Muon sample.
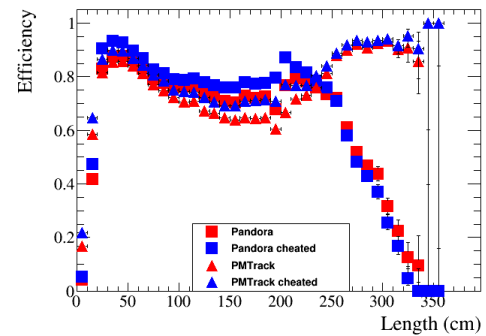
(b) Reconstruction efficiencies for a CRY sample.

Fig. 5.5 The reconstruction efficiencies for simulated events as a function of Monte Carlo truth track length. The efficiencies are shown for non-cheated reconstruction (square blocks) and cheated reconstruction (triangle blocks) for both PMTrack (black) and Pandora (blue).

(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 deposited energy. The efficiencies are shown for non-cheated reconstruction (square blocks) and cheated reconstruction (triangle blocks) for both PMTrack (black) and Pandora (blue).

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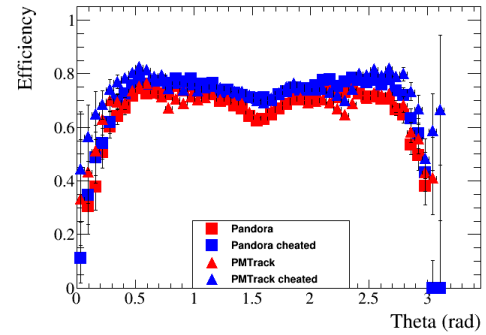
(b) Reconstruction efficiencies for a CRY sample.

Fig. 5.7 The reconstruction efficiencies for simulated events as a function of Monte Carlo truth track angle in theta. The efficiencies are shown for non-cheated reconstruction (square blocks) and cheated reconstruction (triangle blocks) for both PMTrack (black) and Pandora (blue).

A striking feature of Figure 5.5 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 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
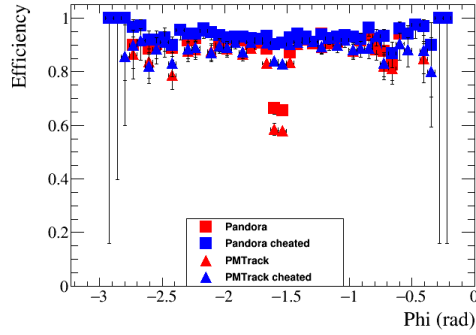
(a) Reconstruction efficiencies for an Anti-Muon sample.
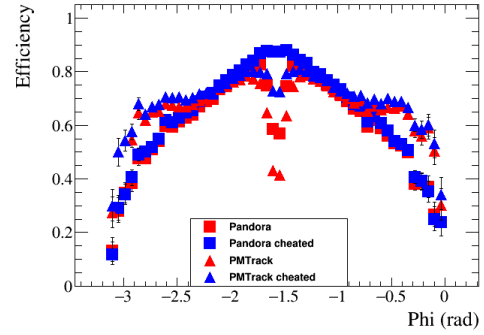
(b) Reconstruction efficiencies for a CRY sample.

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

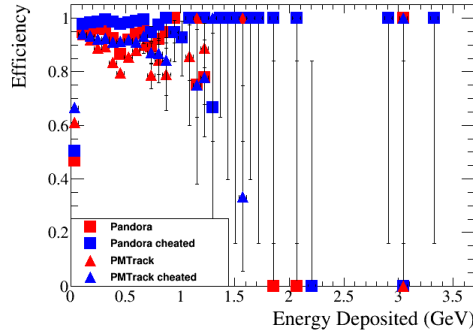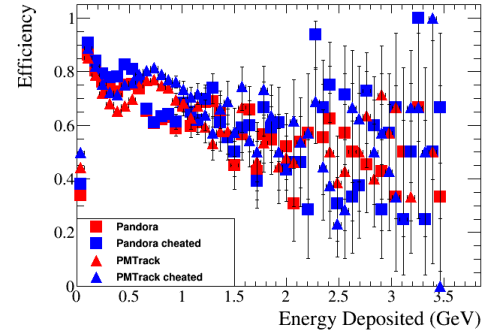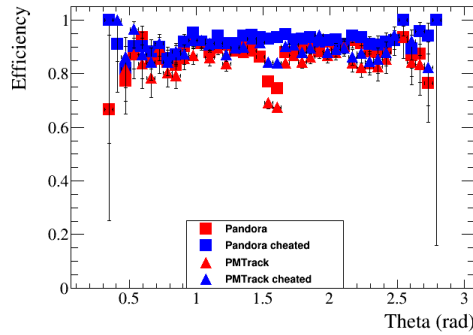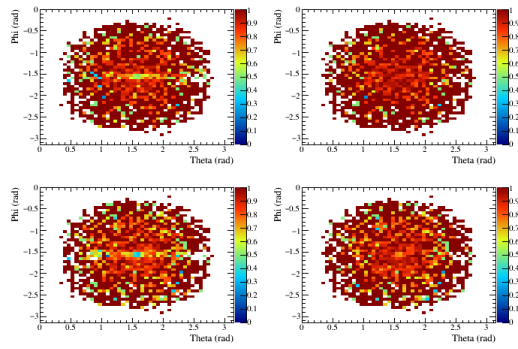(a) Reconstruction efficiencies for an Anti-Muon sample.
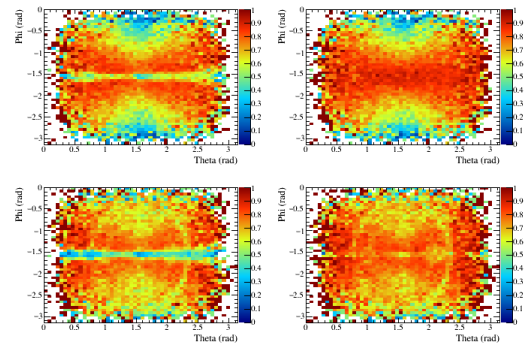
(b) Reconstruction efficiencies for a CRY sample.

Fig. 5.9 The reconstruction efficiencies for simulated events as a function of Monte Carlo truth track angle in 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).

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 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.

$$x_{Hit} = T_{Hit} \times v_{Drift} \tag{5.1a}$$

$$T_{Hit} = T_{Measured} - T_{Interaction} \tag{5.1b}$$

5.4   Performing particle identification

Being able to perform reliable particle identification (PID) is key deliverable for the DUNE experiment, and so efforts have been made to establish a metric by which this can be achieved. The predominant method of performing PID in LAr is to use the relationship between $\frac{dE}{dx}$ and the residiual range of the track, defined as being the distance between a point on the track and the stopping point of the track. The relationship between these particle is observed to be dependent on particle mass and is quantified by the Bethe-Bloch equation!!!citepBetheBloch!!! which is shown in Figure 5.10. The sharp increase in energy loss per unit length can be seen to occur at different energies for different particle masses meaning that the peak value of $\frac{dE}{dx}$ can change significantly between muons and protons.

Fig. 5.10 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}.

The particle mass dependance can be seen by plotting the $\frac{dE}{dx}$ against the residiual range of the particle on a log-log plot, as shown in Figure 5.11a. A power law dependence is found to describe the relationship [1], as shown in Equation 5.2. The dependence on b is found to be weak, and able to be set to -0.42 for all particle masses. This means that the main discriminant used is A which has a strong dependence on particle mass. The values for A and b calculated from Figure 5.11a are shown in Table 5.1.

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 found 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

Table 5.1 Values of the *A* and *b* parameters from Equation 5.2 for different particle masses from Figure 5.11a [1].

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

1 only be calculated for particles which stop in the detector as all other particles will have

2 very similar $\frac{dE}{dx}$ distributions. As shown in Figure 5.11a the average value of *A* is normally

3 calculated for the last 30 cm of the track.

4

5   The PIDA method was tested in [1],

(a) Stopping power for different particle masses.

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

Fig. 5.11 Defining the PIDA metric for particle identification and testing it on a Monte Carlo truth sample.

6
$$\left.\frac{dE}{dx}\right|_{calo} = AR^b \tag{5.2}$$

7
$$A_i = \left(\frac{dE}{dx}\right|_{calo})_i \times R_i^{0.42} \tag{5.3}$$

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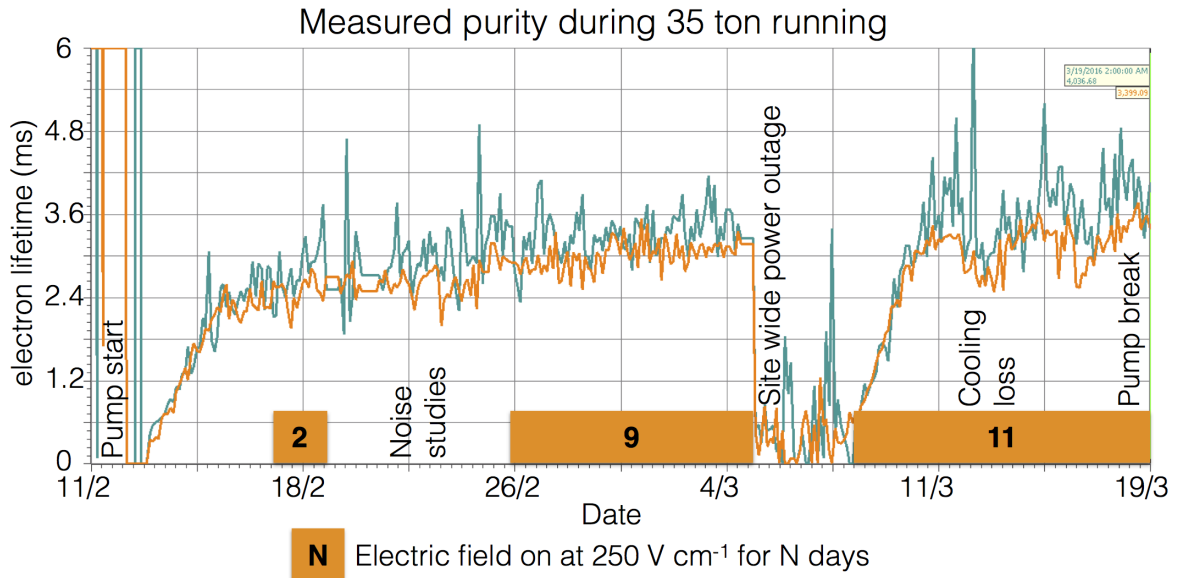
Fig. 6.1 Timeline showing the data collected during the 35 ton Phase II run once the purification pumps were turned on.

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

During operations the DAQ was configured to maximise data throughput and physics potential. This meant recording different lengths of times for each of the three sub-systems as the data volumes and length of physics information were significantly different. For example due to the emission of prompt light the physics information from the SSPs is of a much shorter length of time than from the RCEs where data has to be recorded whilst the electrons drift through the LAr. During the running period the recorded data was triggered by through-going muons which produced co-incidences on the CRCs on opposites side of the cryostat. The system used to collect the CRC data was also responsible for generating the triggers and so this meant that the production of triggers could be suppressed to approximately 1 Hz by only producing triggers on the Nth co-incidence. A trigger rate of 1 Hz was used as the maximum speed at which data could be written to disk was approximately 60 MB s^{-1}, which is roughly equal to the size of each triggered event when the entire detector is read-out in the configuration discussed below. The rate at which events were recorded could have been increased if zero-suppression of the TPC data had 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 co-incidence to be separated from other tracks it was necessary to have roughly one drift window both before and after the drift window around the co-incidence, meaning that data was recorded for 7.5 ms or 15,000 ticks around each co-incidence. Only the prompt light from through-going particles was collected and so only 200 μs 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, though the co-incidence triggers were only produced when a trigger was issued.

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. Removing these padded header objects was a remit of the online to offline converter discussed in Section 6.2. The length of the DAQ events was configurable and 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 DAQ events with headers between physics events introduced some peculiarities in the data recorded such as DAQ events containing two parts of non-continuous data as shown in Figure 6.2 where the second DAQ event has no information for the time between the end of physics event 2 and the start of physics event 3.

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. Data is collected for every tick on each RCE, where each RCE controls 128 channels. This is called a nanoslice. A microslice is then made by combining N nanoslices such that it contains 0.5 ms (1,000 ticks) of data across all channels. Microslices are then combined to make millislices, where a millislice is synonymous with the DAQ events discussed earlier meaning that 20 microslices are combined to make a millislice. A lack of recorded data means that microslices consist of headers in the place of nanoslices with physics information.

During normal data taking the last N microslices are buffered in the RCEs so that if a trigger is issued the previous millislices can be accessed before they are deleted. As the data is buffered in the form of microslices previous microslices may only be accessed in whole. This means that a whole number of microslices must be loaded before the trigger so when
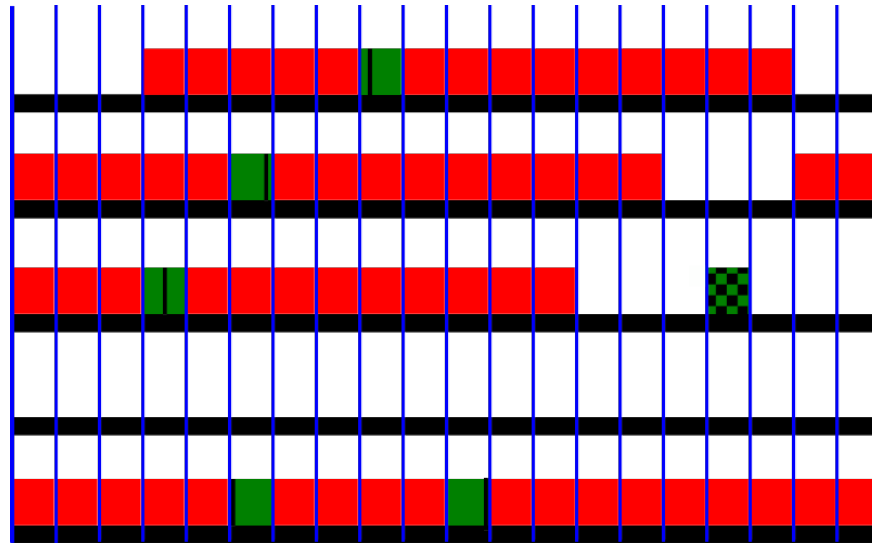
Fig. 6.2 A diagram of possible microslice structures for the TPC data recorded by the 35 ton. Each row represents a single DAQ event, also called a microslice. The vertical blue lines delineate each microslice (0.5 ms, 1,000 ticks). Solid red boxes represent micro slices with TPC data in them, with a group of red boxes being labelled "physics events". Green boxes represent triggers which were used with the black lines showing the time in the millislice at which the trigger occurred, and green and black patterned boxes represent triggers which were ignored.

a trigger is issued part way through a microslice the previous X microslices are sent to the event builders. As a result during running there are always a minimum number of ticks both before (5,000 ticks) and after the trigger (9,000 ticks) but the exact numbers can change by up to 1,000 ticks for a given event depending on where in a microslice the trigger came. This is shown in Figure 6.2 where the black lines representing triggers are seen to occur at different points within the microslices, for example physics event 1 will have more data after the trigger than physics event 2 as the trigger occurs earlier in the 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 following the unpacking of each
of the sub-systems the TPC ticks would be looped through to see if a user defined set
of conditions could be satisfied. These conditions were usually whether an East-West or
North-South counter co-incidence occurred at that time, or if the current 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 flight 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, clearly this 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 co-incidence to make an event a value of 300
pre-condition ticks is normally used. 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 co-incidence. 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 downstream 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
them around their chosen conditions and so the use of the unpacking was determined by the
interface depending on the format of the input file.

6.3   Observations on data quality and noise mitigation

Reformatting the online data to the offline format was an important step in maintaining
data quality as subsequently there was no access to the raw data due to how the DUNE
software is established. Some of the important checks which were performed are outlined in

1 Figure 6.3. If any of these issues were present in a given physics event then it is discarded
2 as the integrity of the data cannot be guaranteed. It was decided that these events would be
3 discarded as non-synchronous events would lead to hits in the detector being at incorrect
4 times and padding empty events with pedestals could mean that tracks seem to disappear and
5 later reappear as they travel through the detector.
6
7   Another example of inconsistent events is when the sub-systems are not synchronised
8 with each other, this is normally caused by one of the sub-systems missing a clock increment
9 from the master timing unit due to the data trigger being issued close to an increment from
10 the master unit. This misalignment causes an incorrect time sample being read out and so
11 the data from each sub-system within a millislice is not consistent meaning that it will fail
12 the timestamp check and so won't be added to the event record. To avoid incomplete events
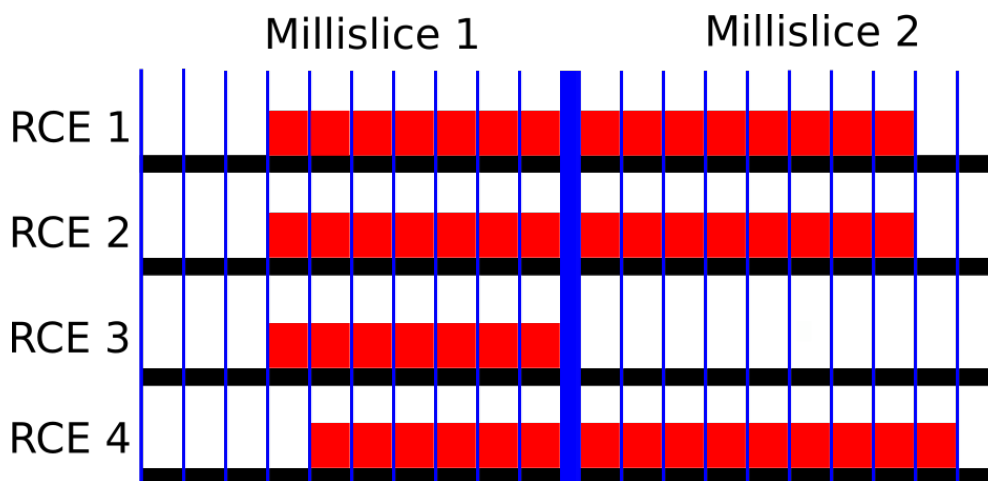13 these physics events are also discarded when observed.
14

Fig. 6.3 A diagram of how TPC microslices can be saved in millislices in the 35 ton. 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.

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

below.   1
                                                                                            2
    Though not directly affecting the noise issues stuck ADC codes were a feature of the data   3
which had to removed downstream. Stuck ADC codes were caused by the 6 least significant   4
bits getting frozen to either 000000 or 111111, this was observed during the first stages of   5
commissioning and an algorithm to remove them was developed and tested on Monte Carlo.   6
In simulations it was observed that the signal could be recovered with minimal losses, as   7
shown in Figure 6.4 where the blue lines (after removal) are seen to closely match the black   8
lines (before adding stuck codes).   9
                                                                                           10
    A significant portion of the noise was correlated between groups of 32 channels, where   11
the ADCs would coherently oscillate. To remove these coherent shifts ADC baselines were   12
calculated for these groups of 32 channels at each tick and then subtracted from the measured   13
ADC values. This was found to be an effective method of removing coherent noise in   14
MicroBooNE [19]. The effect of removing coherent noise is shown in Figure 6.5, where   15
the signal peak becomes much easier to discern after noise removal and a coherent noise   16
peak around tick 6030 is removed. An issue with removing coherent noise in this way is that   17
events which are parallel to the APAs will produce signals at common times across adjacent   18
wires and these signals may be removed along with the coherent noise causing a reduction in   19
the hit reconstruction efficiency. The only way to prevent this is to "protect" potential signal   20
regions from the coherent noise removal, as is done in MicroBooNE [19].   21
                                                                                           22
    When a Fast Fourier Transform (FFT) [12] is performed on the coherent noise subtracted   23
waveforms it can be seen that signals occur with specific frequencies. Some of these fre-   24
quencies are caused by real energy depositions, whilst others are due to the electronics noise.   25
It is possible to remove the noise frequencies by applying Wiener filters [22]. Frequency   26
spectrum's are taken for each of the three planes and a clear signal is both preserved and   27
suppressed. The raw signal spectra are then divided by the signal suppressed spectra to   28
produce *signal/noise* frequency spaces. The signal regions to be conserved can then be   29
found by fitting a combination of sigmoid functions to the frequency spaces around regions   30
of high *signal/noise*. A demonstration of how this applied is shown in Figure 6.6. It is   31
also possible to remove specific frequencies which are not removed by the filters, this was   32
necessary for a 54 KHz noise component introduced by the fluorescent lights in the detector   33
hall. After the run ended it was found that some of the high frequency noise components   34
were introduced by a short on a warm power cable, the techniques used to find this cable will   35

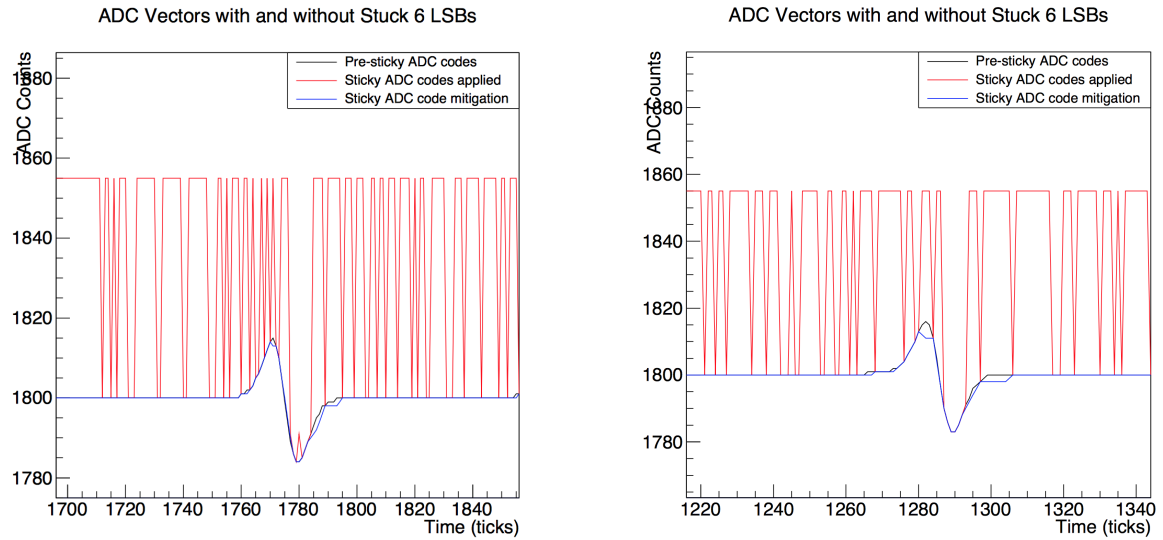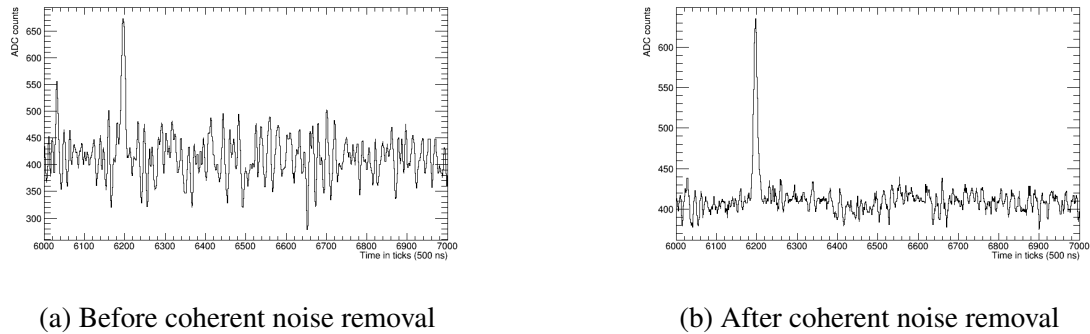be used when commissioning future detectors [13].

Fig. 6.4 Two Monte Carlo spectrum's showing the effect of the introduction and removal of stuck bits on a simulated signal [16].

(a) Before coherent noise removal   (b) After coherent noise removal

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

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 after the removal of the most noisy wires.
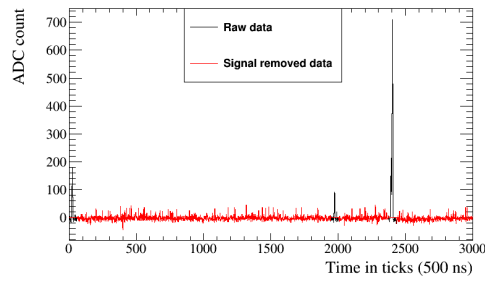
Transitions to a higher noise state associated with strong signals at high frequencies, between 400 and 650 KHz, were observed after cool down. The transitions would occur approximately every 2 hours and were occasionally observed to happen shortly after a saturation event across the whole detector [13]. Once the state was induced the only way to
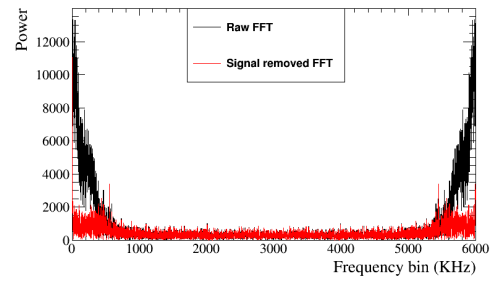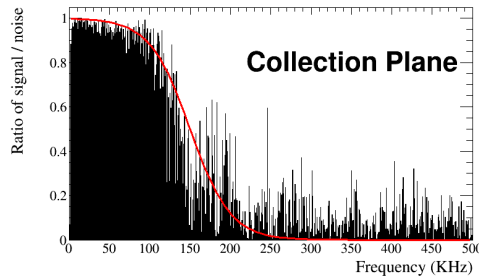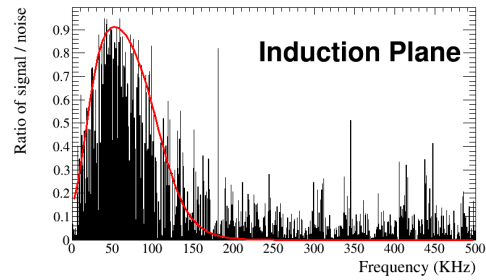
(a) A raw and signal subtracted waveform for a collection plane wire.

(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.
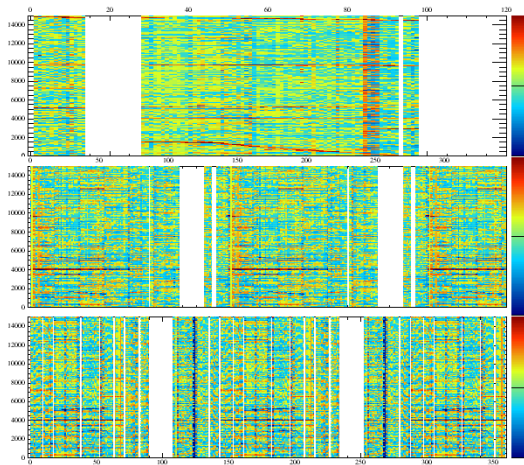
(a) Raw signal before noise removal
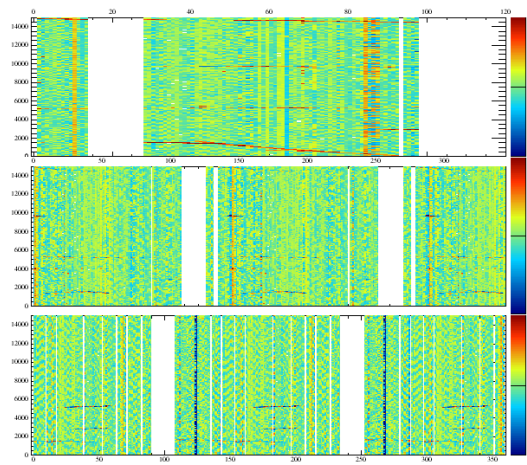
(b) Signal after noise removal

Fig. 6.7 Event displays showing the effect of the noise removal algorithms on data in the 35 ton.

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 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 [14].

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 co-incidences. When determining whether a reconstructed hit could have caused the counter co-incidence 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. 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?

    - This is because one would expect a track to deposit energy on multiple adjacent wire segments.

- Which hit lies closest to the line defined by unique collection plane hits in the XZ plane?

    - This is follows identical logic to the first criteria, but selects the hit which best matches the collection plane hits and attempts to remove the effect of noisy collection plane wires by only using wires which have one hit within the counter shadow. This would also hopefully improve the quality of the fit as there will not be numerous outlying hits.

    - This can be changed to consider the line defined by previously selected hits in the given TPC and plane where the hit choices are.

Following a re-optimisation of the clustering algorithms it was observed that the standard reconstruction could achieve track reconstruction to a similar efficiency as the counter shadowing and so the standard reconstruction has been used in the discussions to follow [24]. There has since been an effort to improve the counter shadowing hit disambiguation to remove the outlying collection plane hits using the MLESAC method [21] whereby points which are far away from a best fit are ignored, these studies are still on-going [20].

A symptom of the elevated noise state is that signals are often dropped on one of the induction planes, this means that the tracking algorithms often have to combine clusters in only two of the three planes. Reconstruction using two planes was shown to be effective by the ArgoNeuT collaboration [3] so the loss of signal in one of the three planes is not prohibitive to track reconstruction. Another consequence of the elevated noise level is that even when the counters are used to seed hit finding the hit finding threshold is too high to reconstruct the very lowest hits. This causes the plot of dQ/dx for muons, shown in Figure 6.8, to look flat due to a cutoff at 100 ADC cm^{-1} below which no hits are found. The inability to reconstruct the lowest energy hits means that calorimetry is all but impossible on the 35 ton dataset even though the tracking algorithms perform relatively well.

The inability to perform reliable calorimetry en masse means that the only particles which can be assuredly identified are the muons which triggered the counter co-incidences, making the analysis proposed in !!FIXME!!Section refPIDAnalysis!!FIXME!! extremely difficult if not impossible. The muons in the triggered sample will all traverse the detector
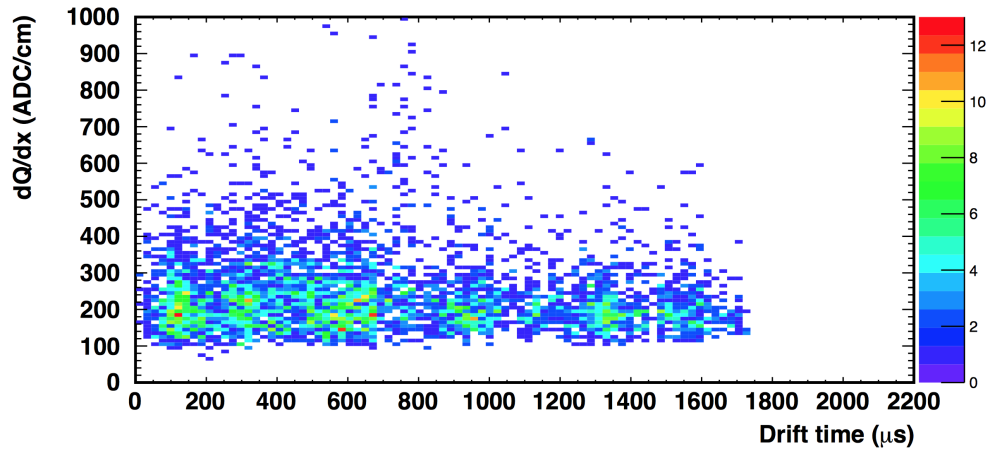
Fig. 6.8 The dQ/dx values for a sample of muon collection plane hits [23], note the cutoff at 100 ADC cm^{-1} due to the hit finding threshold.

but their orientations can be carefully selected by the user, for example one could easily select a sample of muons which cross the APAs at increasing angles, or are parallel to the wire planes at increasing drift distances. A reconstructed track is assigned to a given counter co-incidence if the dot product of the track and the co-incidence is more than 0.98 and the hit times are consistent with the x positions of the counters, as shown in Figure 6.9. This should result in 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 and so will not pass the dot product cut. 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 6.10.
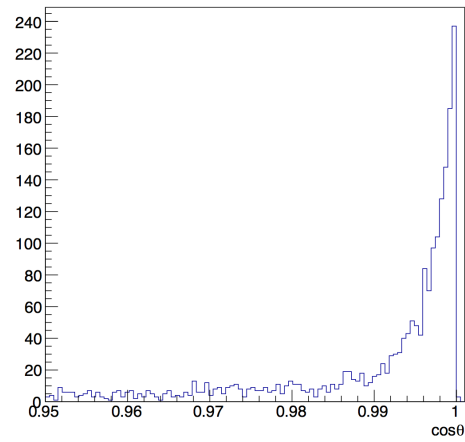
By matching tracks in this way it possible to evaluate the reconstruction efficiencies for these muons at increasing drift distances and track angles. If multiple tracks are aligned with the coincidence and are within the expected time region then their track lengths are summed when calculating reconstruction 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 track. A minimum reconstructed length of 50 cm is used to determine if the track was partially reconstructed as above this length the track is likely to have been stitched between TPCs, the verification of which was a design goal of the 35 ton. A reconstructed track that is 50 cm long is also likely to have a large number of hits on collection plane wires that are not noisy, these hits are needed when calculating purity or measuring the effect of diffusion as discussed in Section 6.5.
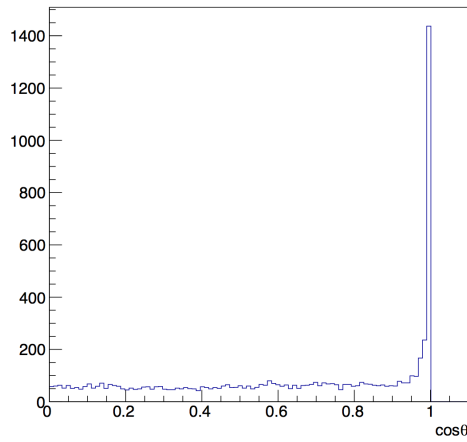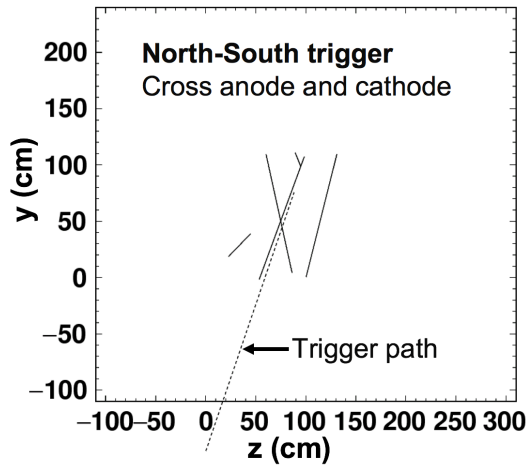
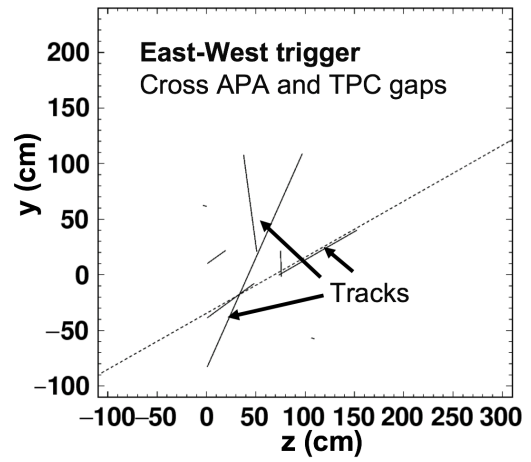(a) All dot product values.   (b) Dot product values close to 1.

Fig. 6.9 The dot product of track and counter coincidences.

(a) A North-South counter coincidence.   (b) An East-West counter coincidence.

Fig. 6.10 The alignment of reconstructed tracks and the counter coincidences they cause.

From Figure 6.11 it is evident that the reconstruction efficiency for tracks with shallow angles relative to the APAs is extremely poor, with the efficiency for tracks aligned with counter differences of 0 or 1 never rising above 10%. This is due to the coherent noise removal where hits which are correlated in time will be removed as they will be perceived as being noise as opposed to real signals. When considering tracks aligned with counter differences of between 3 and 5 the reconstruction is seen to be perform reasonably well for tracks whose centres are less than 140 cm away from the APAs, before decreasing for tracks which have a longer drift distance. A decrease in reconstruction efficiency for tracks

1　at large drift distances is due to the hits being below threshold and so not being recon-

2　structed. It should be noted that in order to observe a drift distance effect on reconstruction

3　efficiency the distance plotted is the average *x* position of the counters, so a track going

4　from 10 cm to 230 cm will have an average *x* position of 120 cm (assuming a straight tra-

5　jectory) as shown by the point for Counter difference 7 in Figure 6.11b with ∼23% efficiency.

6
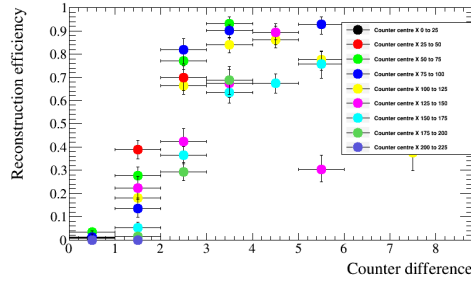
(a) The difference in efficiencies of different track angles relative to the APAs plotted for increasing distance from the APAs.
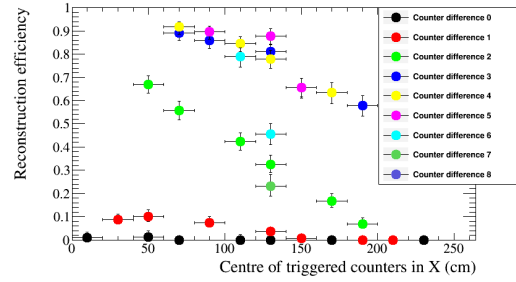
(b) The difference in efficiencies at increasing distances from the APAs plotted for increasing track angles relative to the APAs.
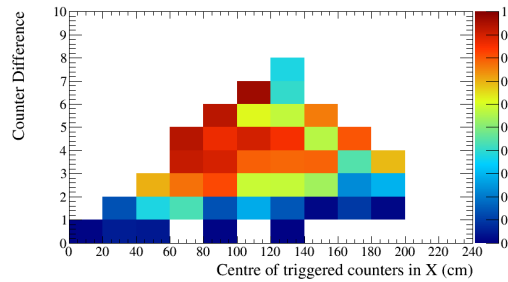
(c) The difference in efficiencies at increasing distance from the APAs against increasing track angles relative to the APAs.
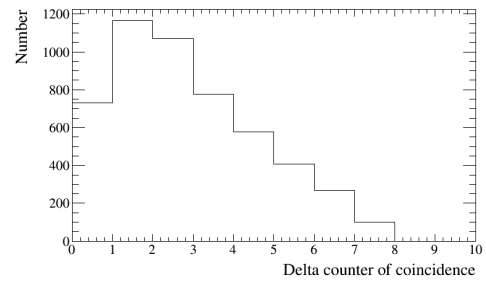
(d) The number of coincidences for each track angle.

Fig. 6.11 The reconstruction efficiencies for tracks which trigger an East-West coincidence in the 35 ton data and the number of events with each co-incidence over a 2 day running period.

7　# 6.5　Measuring interaction times using electron diffusion

8　As electrons drift from the interaction point to the wire planes they become spread out in

9　both time and space, this effect is known as diffusion and is an important property of electron

10　transport in LAr which must be well understood. The mechanism by which diffusion occurs

11　in LAr was first discussed by Atrazhev-Timoshkin [6], and has since been developed to con-

12　sist of a complete set of measurements for electric fields between 100 and 2000 V cm^{-1} [18].

The diffusion of electrons is rarely isotropic and so the component that is transverse to   1
the drift field and the component that is parallel to the drift field are normally measured   2
separately. Diffusion parallel to the drift field is called longitudinal diffusion and is generally   3
smaller than the component of diffusion that is transverse to the drift field. Figure 6.12 shows   4
how diffusion can smear the electrons collected on a set of wires when the electrons are   5
initially highly correlated in time and space.   6
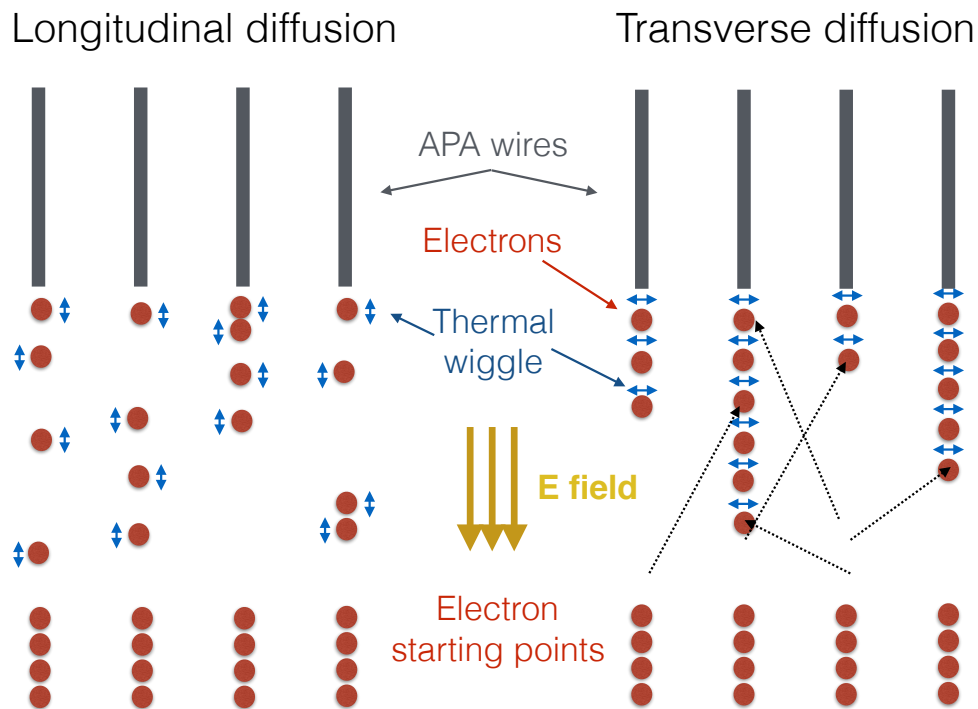   7

Fig. 6.12 A schematic showing the difference between longitudinal diffusion (left) and transverse diffusion (right). In both cases four electrons are initially shown below four wires, a thermal wiggle is then applied in the drift direction and perpendicular to the drift direction in the longitudinal and transverse cases respectively. It can be seen that the effect of the thermal wiggles 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 [7].

Longitudinal diffusion has the effect of spreading the drifting electrons out in time caus-   8
ing signals to become wider in time and smaller in height as the total charge is conserved.   9
The increasing hit width can be measured for increasing drift times (distances) provided the   10
hits do not fall below a hit finding threshold. Transverse diffusion causes drifting electrons to   11
spread out in space, changing the amount of charge deposited on a wire and reducing the   12
charge resolution of the detector. Transverse diffusion is measured by discerning how the   13

width of the hit charge distribution changes for increasing drift distances [7].

Through-going particles make ideal tracks to study diffusion as they are minimally ion-ising 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.

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 Earths surface such as MicroBooNE where each neutrino interaction has a background of many cosmic muons which need to be disentangled in order correctly assign a scintillation flash to a reconstructed track. An example of an event which has many scintillation flashes and cosmic muons which need to be correctly associated is shown in Figure 6.13. !!!I MAY HAVE A FIGURE FROM A PREVIOUS SECTION!!! 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.
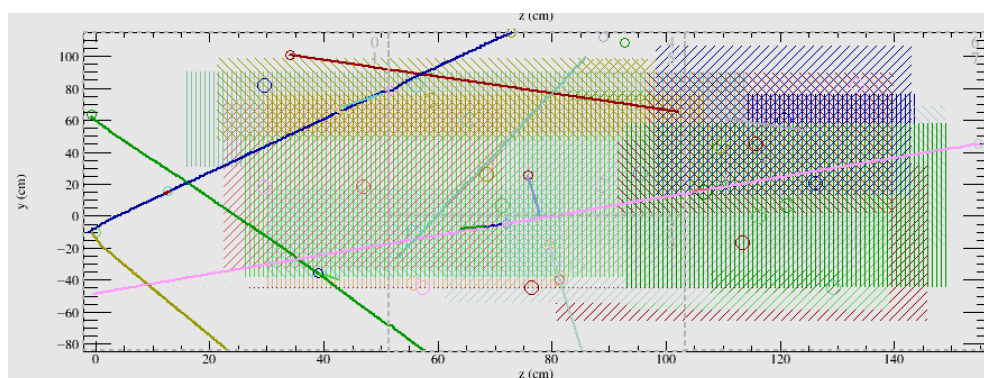

Fig. 6.13 A simulated event display showing multiple tracks and flashes to be assigned to each other in the 35 ton. The coloured lines represent reconstructed tracks, whilst the coloured dashed boxes represent flashes.

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 a baseline measurement on an ideal detector is also useful, Monte Carlo can provide this sample.

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 as shown by Figure 6.14 where when a baseline noise of 10 ADC counts was added to a simulated hit with a peak value of 50 ADC counts and RMS 10 ticks the width was seen to increase 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. In order 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. A further restriction applied is that 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 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. From Figure 6.11 it can be seen that tracks with a counter difference of 4 are the best reconstructed in data and so reference will be given to these
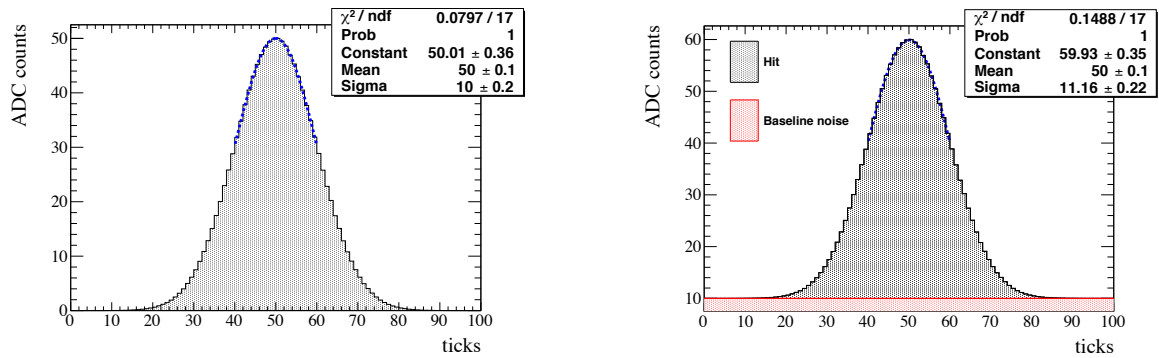
(a) A toy Gaussian hit with width 10 ticks, and amplitude 50 ADC counts.

(b) The reconstructed width when a noise baseline of 10 ADC counts is added.

Fig. 6.14 The width of a Gaussian fit is seen to increase by more than 10% when a noise baseline is added to the signal showing that noise can cause the measured width of a hit to increase [7].

tracks, though the same procedure is applied to all track angles. 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.

- The hit *RMS*/*Charge* - an attempt to incorporate the effect of impurities in the LAr for relatively low purity data which will have a drift distance dependence.

Fitting Gaussian's around the peaks of the distributions will yield the most probable values for the drift regions, as is shown in Figure 6.15.

The drift distance effect of diffusion can then be observed for a given range of track angles by fitting these most probable values as drift distance increases. The angular dependence can then be shown by observing how the most probable fit value at a drift distance of 0 cm changes for increasing angles, this is shown in Figure 6.16. A drift distance dependence can clearly be seen in the data as the most probable hit *RMS* is seen to increase for hits which originate further from the APAs. It also clear that there is an angular dependence on the hit width as the most probable hit widths next to the APAs is seen to rise for tracks which are more inclined relative to the APAs. These dependencies show that when considering a large sample diffusion can be separated into distance and angular dependant dependencies, however whether this can be observed for individual tracks has not yet been considered.
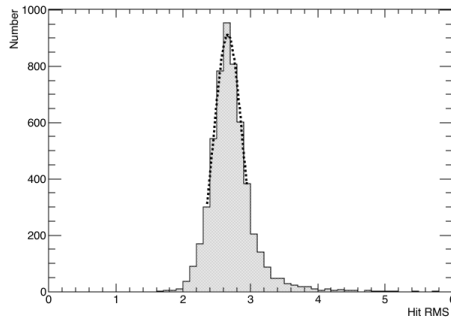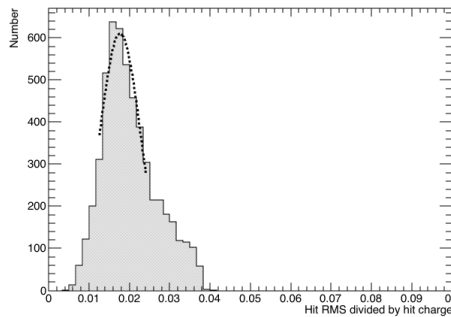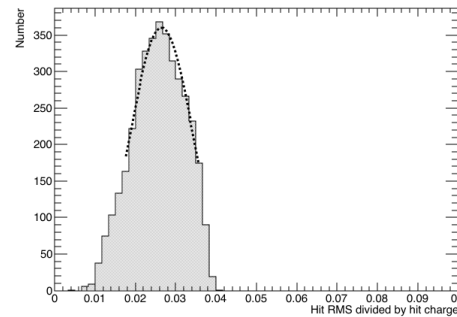
(a) The most probable hit *RMS* value at *x* = 20 cm for tracks with a counter difference of 4.
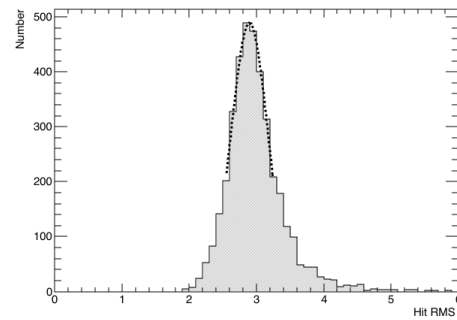
(b) The most probable hit *RMS* value at *x* = 140 cm for tracks with a counter difference of 4.

(c) The most probable hit *RMS/Charge* value at *x* = 20 cm for tracks with a counter difference of 4.

(d) The most probable hit *RMS/Charge* value at *x* = 140 cm for tracks with a counter difference of 4.

Fig. 6.15 The most probable values for hits from at 20 cm and 140 cm from tracks which triggered a counter difference of 4, when the *RMS* of the hit and the *RMS/charge* of the hit are considered.

In order to consider single tracks the best line fits for the counter differences for a large sample of tracks, such as in Figure 6.16a, need to be used to predict the drift position you would expect a hit to originate from given a value for its hit *RMS* and the angle of track to which it belongs. The predicted positions can then be compared to the known position from the counter co-incidence to determine the accuracy of the prediction. As the distributions shown in Figure 6.15 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.
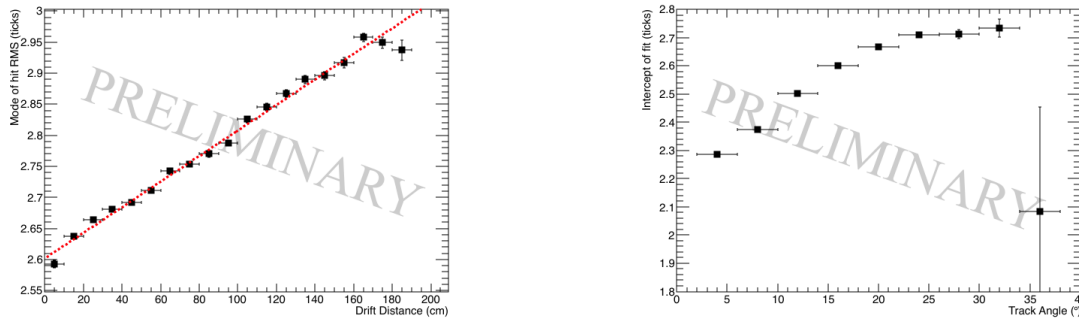
(a) The most probable hit *RMS* value for tracks with a counter difference of 4 for a range of drift distances.

(b) The most probable value of hit *RMS* at a drift distance of 0 cm at increasing track angles.

Fig. 6.16 Fits showing the drift distance and angular dependence of diffusion in the 35 ton dataset.

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 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 predicted hit times and the difference in predicted and reconstructed hit times are shown in Figure 6.17 for both prediction metrics. It can be seen that in the 35 ton dataset the *RMS/Charge* metric appears to provide a more robust measurement of the interaction time as the peak value for the predicted minus reconstructed hit time is more closely centred around 0 and is much larger than for the RMS metric. The peak is centred around a difference of 0 as the hits had previously been corrected so as to be centred around the measured interaction time from the counter coincidence.

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 Figure 6.18, where as expected from Figure 6.17 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 very high purity LAr environments though as the decrease in collected charge at increasing drift distances becomes small [8]. The effect of increasing LAr purity is shown in Section 6.5.3.
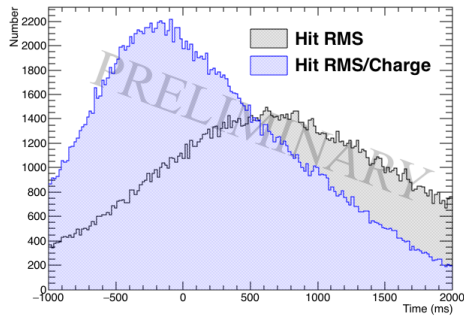
(a) The predicted hit time for each hit in all tracks with more than 100 'good' collection plane hits.
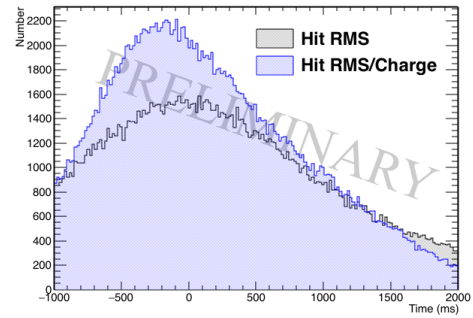
(b) The difference in reconstructed and predicted hit time for each hit in all tracks with more than 100 'good' collection plane hits.

Fig. 6.17 The predicted hit times for tracks with more than 100 'good' hit collection plane hits in the 35 ton data. The times predicted by the hit *RMS* metric are shown in black, whilst the times predicted by the hit *RMS/Charge* metric are shown in blue.

Using the change in hit charge in the 35 ton is not without its own dangers though, because as shown in Figure 6.8 there is a thresholding effect for hits with large drift times. However, as the same thresholding effect is present in all 35 ton data samples the limitation it introduces is mainly in the efficiency with which 'good' collection plane hits will be reconstructed and so this information can be confidently used.
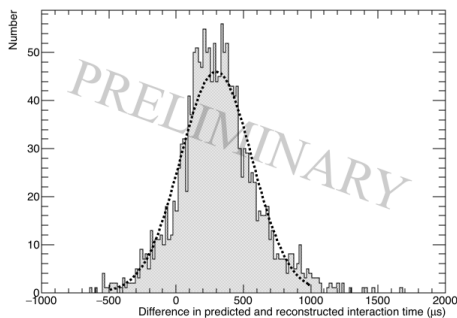
(a) The average difference in interaction times using the hit *RMS* metric.
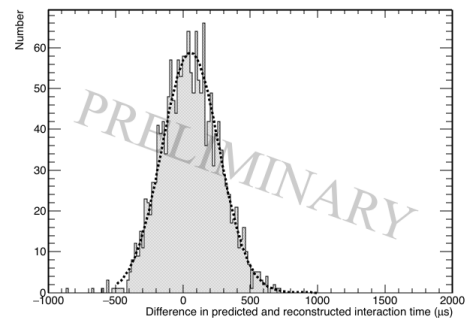
(b) The average difference in interaction times using the hit *RMS/Charge* metric.

Fig. 6.18 The difference in predicted and reconstructed interaction times in the 35 ton data.

Figure 6.18 shows that the interaction time of a reconstructed track in the 35 ton dataset can be determined using the effects of diffusion to an accuracy of $298.0\pm267.0\,\mu$s when using the hit *RMS* metric and to an accuracy of $55.6\pm210.4\,\mu$s when using the hit *RMS/Charge*

metric. This resolution is impressive as even with the issues with regards to data quality in the 35 ton tracks can be cleanly separated throughout a drift window of 5200 μs. The application of the method to a low noise environment would provide the true accuracy of the method, because one would expect that the problems with data quality would adversely affect the accuracy to which diffusion can be measured. For this reason the next section presents a complementary analysis performed on a low background sample of 35 ton Monte Carlo with the same electronics settings as the data presented here.

6.5.2 Determining interaction times in a low noise detector and differences with data

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

Figure 6.19 shows both the *RMS* and *RMS*/*Charge* distributions for hits that are 100 cm away from the APAs and are from tracks with a counter difference of 4. It can be seen that the distributions from the low noise Monte Carlo are tighter than those from the 35 ton data and are also peaked at a lower hit RMS. This is likely due to the fact that the coherent noise baseline seen in data can increase the width of hits as shown in Figure 6.14 and a higher noise state will affect how well individual hits can be reconstructed. In addition, the most probable values of hit *RMS* at increasing drift distance is shown in Figure 6.20a where the low noise Monte Carlo is again shown against the values from the data. The most probable value of hit *RMS* at a drift distance of 0 cm for a range of angles is also shown in Figure 6.20b. As was seen when considering the distributions at specific distances and angles the most probable hit *RMS* in the low noise Monte Carlo is systematically lower than in the data due to the elevated noise level in the data. Another difference between the Monte Carlo and the data is that the gradient of the most probable hit *RMS* values in data is roughly half of that in the Monte Carlo, this could be due to an overestimation of longitudinal diffusion in the Monte Carlo.
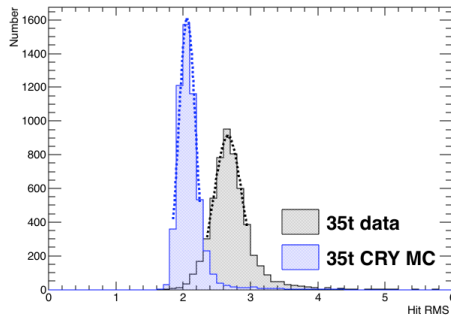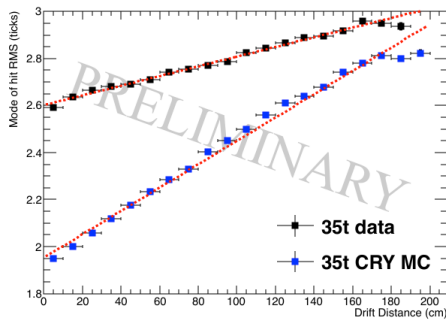
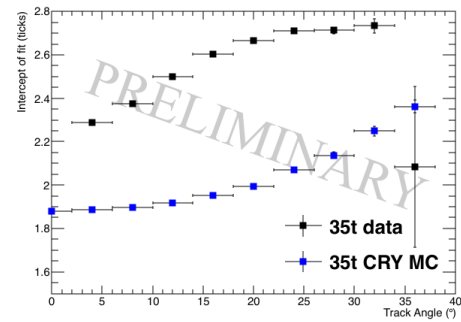(a) The distribution of hit *RMS* values at $x = 20$ cm.
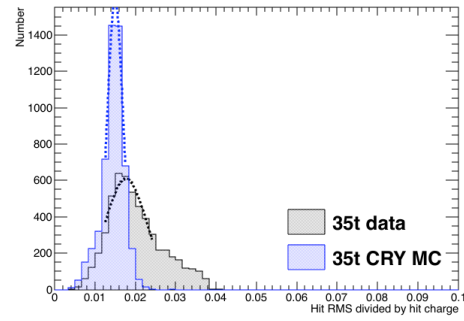
(b) The distribution of hit *RMS*/*Charge* values at $x = 20$ cm.

Fig. 6.19 The most probable values of the *RMS* and *RMS*/*Charge* distributions at $x = 20$ cm for tracks with a counter difference of 4. The distributions from the 35 ton data are shown in black, whilst the distributions from the low noise simulation are shown in blue.

(a) The most probable hit *RMS* values for tracks with a counter difference of 4.

(b) The most probable hit *RMS* values at $x = 0$ for a range of track angles.

Fig. 6.20 A comparison of the drift distance and angular dependence of diffusion in a low noise 35 ton detector and the 35 ton dataset. The distributions from the 35 ton data are shown in black, whilst the distributions from the low noise simulation are shown in blue.

Upon calculating the fit metrics in the low noise Monte Carlo sample it is then possible to use these to predict track interaction times, this is shown in Figure 6.21. As observed when considering the data sample, the *RMS*/*Charge* metric appears to be better able to accurately predict interaction times, this is again due to the ability to incorporate information about loses due to impurities which increase with drift distance. Also, as expected from the previous figures and the lower noise state in the Monte Carlo it is seen that the interaction times predicted in the Monte Carlo more closely match the true interaction times than in the data. The accuracy of interaction time determination in Monte Carlo (data) is shown to be 168.1±126.5 (298.0±267.0) μs and -41.2±107.8 (55.6±210.4) μs for the hit *RMS* and

1   *RMS/Charge* metrics respectively. An important feature to observe is that the widths of the

2   distributions in Monte Carlo is less than half of that in the data, again this is attributed to the

3   lower noise level in the Monte Carlo.
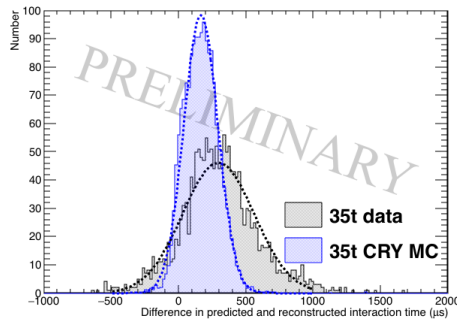
(a) The accuracy to which interaction times are determined using the hit *RMS* metric.
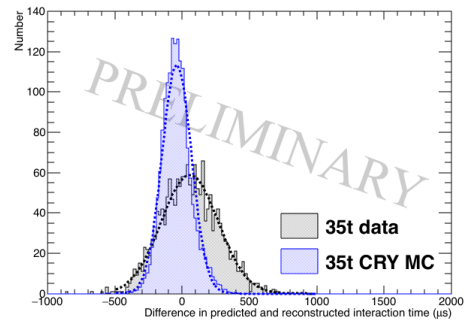
(b) The accuracy to which interaction times are determined using the hit *RMS/Charge* metric.

Fig. 6.21 A comparison of the accuracy of the interaction time prediction metrics for the low noise Monte Carlo and data 35 ton samples for both the *RMS* and *RMS/Charge* prediction metrics. The distributions from the 35 ton data are shown in black, whilst the distributions from the low noise simulation are shown in blue.

4   The calculation of interaction times is clearly much better in the low noise Monte Carlo

5   than in the 35 ton dataset, however the distributions are still not centred around 0 implying

6   that there is a systematic error in the method which has not been removed when considering

7   a low noise environment. Looking at Figure 6.19 the impact of delta rays can still be seen

8   where the *RMS* plot still has quite a significant tail above the most probable value. This will

9   cause the predicted interaction times to be skewed towards larger times as the hits containing

10   delta rays will be wider and so appear to come at later times than they actually do. Hits

11   containing unseparated delta rays are difficult to remove without looking for slight dips in the

12   raw signals caused by the delta ray beginning to separate from the main track. This would be

13   almost impossible in the 35 ton dataset given the oscillatory nature of the noise. Delta rays

14   can also offer an explanation for the *RMS/Charge* plot underestimating interaction time as

15   hits containing delta rays would deposit more charge and this increased charge would likely

16   be larger than the increased width, causing the *RMS/Charge* to decrease. This is seen in Fig-

17   ure 6.19 where both the data and Monte Carlo samples have consistent tails at small values of

18   *RMS/Charge*, this decrease would then lead to an underestimate of the interaction time. The

19   increased noise level in the 35 ton dataset still causes an overestimation of interaction time,

20   but the most probable difference in predicted and reconstructed interaction times is lower us-

ing the *RMS/Charge* metric than the *RMS* metric in data suggesting this effect is still present.

The widths of the distributions shows that whilst tracks can be disentangled over an entire drift window it is potentially difficult to do so over periods of around $100\,\mu$s, this width is likely due to interpreting distributions which are not Gaussian as Gaussian. This may mean that more sophisticated fits to the distributions used to make the prediction metrics are required which take into account the tails of the distributions as opposed to focusing only on the peaks. The non-Gaussian nature of the distributions will also affect the assumption that over a large number of hits the Gaussian nature of the hit distribution would tend to dominate as if this is not true then one would expect to measure a wide distribution of interaction hit times. It appears that this is the case and that the tails seen in Figure 6.19 are causing the predicted hit time distribution to be too broad. This feature could potentially be suppressed by performing a fit to the difference in predicted and reconstructed hit times for a track as opposed to taking an average.

6.5.3 Discerning the impact of noise and electron lifetime in Monte Carlo

Many of the difficulties with the reconstruction and analysis of the 35 ton data have been explained by stating that the noise level made studies difficult, it is important to verify these claims with a study showing the effect of increased noise on the accuracy of the interaction time predicted metrics. This study has been done using the same muons used in the previous Monte Carlo sample with only the noise level of the detector being increased and the inclusion of the noise mitigation algorithms to attempt to reduce the effects of noise as is done in the data sample. The results of this study are shown in Figure 6.22, where the metric distributions for tracks with a counter difference of 4 are shown at 20 cm in Figures 6.22a and 6.22b, the most probable values of hit *RMS* for tracks due to a counter difference of 4 at all drift distances and the most probable values of hit *RMS* at $x = 0$ cm for all counter differences are shown in Figures 6.22c and 6.22d respectively and the accuracy of the metrics are shown in Figures 6.22e and 6.22f.

!!!!!! EXPLANATIONS OF WHAT IT LOOKS LIKE !!!!!!!

Another important constraint on the effectiveness of interaction time determination is the electron lifetime of the liquid Argon. As the electron lifetime increase fewer electrons will recombine as they drift towards to the APAs meaning that the amount of charge collected

(a) The distribution of hit *RMS* values at $x =$ 20 cm.

(b) The distribution of hit *RMS*/*Charge* values at $x = 20$ cm.

(c) The most probable hit *RMS* values for tracks with a counter difference of 4.

(d) The most probable hit *RMS* values at $x = 0$ for a range of track angles.

(e) The accuracy to which interaction times are determined using the hit *RMS* metric.

(f) The accuracy to which interaction times are determined using the hit *RMS*/*Charge* metric.

Fig. 6.22 A study into how increased noise affects the accuracy to which interactions can be determined in the 35 ton.

will be less affected by the drift distance, potentially greatly affecting the accuracy of the hit *RMS*/*Charge* metric. The results of a study to observe the effect of increased electron lifetime are shown in Figure 6.23.

(a) The distribution of hit *RMS* values at $x =$ 20 cm.

(b) The distribution of hit *RMS*/*Charge* values at $x = 20$ cm.

(c) The most probable hit *RMS* values for tracks with a counter difference of 4.

(d) The most probable hit *RMS* values at $x = 0$ for a range of track angles.

(e) The accuracy to which interaction times are determined using the hit *RMS* metric.

(f) The accuracy to which interaction times are determined using the hit *RMS*/*Charge* metric.

Fig. 6.23 A study into how increased noise affects the accuracy to which interactions can be determined in the 35 ton.

!!!!!! EXPLANATIONS OF WHAT IT LOOKS LIKE !!!!!!!

The electric field in all of these studies has been 250 V cm^{-1}, however diffusion is known to be affected by the applied electric field. Many LArTPC detectors, including DUNE, are being designed to have an electric field of 500 V cm^{-1} and so it is important to observe the effect which a higher electric field has on the accuracy of the method. An intermediate field of 375 V cm^{-1} is considered as this was the next planned electric field for the 35 ton data taking before the filtration pump breakage stopped data taking. The results of a study to observe the effect of increased electric field are shown in Figure 6.24.

!!!!!! EXPLANATIONS OF WHAT IT LOOKS LIKE !!!!!!!

(a) The distribution of hit *RMS* values at $x = 20$ cm.

(b) The distribution of hit *RMS*/*Charge* values at $x = 20$ cm.

(c) The most probable hit *RMS* values for tracks with a counter difference of 4.

(d) The most probable hit *RMS* values at $x = 0$ for a range of track angles.

(e) The accuracy to which interaction times are determined using the hit *RMS* metric.

(f) The accuracy to which interaction times are determined using the hit *RMS*/*Charge* metric.

Fig. 6.24 A study into how increased noise affects the accuracy to which interactions can be determined in the 35 ton.

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

sdfsdsdas

Chapter 7

Simulations of the DUNE Far Detector

7.1   The MUSUN and MUSIC generators

7.2   Simulations of the LBNE surface detector

7.3   Incorporation of MUSUN into LArSoft

7.4   Simulation of background for DUNE

7.5   Cosmogenic background for nucleon decay channels in DUNE

References

[1] Acciarri, R. et al. (2013). A study of electron recombination using highly ionizing particles in the ArgoNeuT Liquid Argon TPC. *JINST*, 8:P08005.

[2] Agostinelli, S. et al. (2003). Geant4—a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250 – 303.

[3] Anderson, C. et al. (2012). The ArgoNeuT Detector in the NuMI Low-Energy beam line at Fermilab. *JINST*, 7:P10019.

[4] Andreopoulos, C. et al. (2010). The {GENIE} neutrino monte carlo generator. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 614(1):87 – 104.

[5] Antonioli, P., Ghetti, C., Korolkova, E. V., Kudryavtsev, V. A., and Sartorelli, G. (1997). A Three-dimensional code for muon propagation through the rock: Music. *Astropart. Phys.*, 7:357–368.

[6] Atrazhev, V. M. and Timoshkin, I. V. (1998). Trasport of electrons in atomic liquids in high electric fields. *IEEE Trans. Dielectrics and Electrical Insulation*, 5:450–457.

[7] Brailsford, D. Lancaster University. DUNE collaboration meeting (2016). Looking at electron diffusion in 35t data and mc. [online] https://indico.fnal.gov/getFile.py/access?contribId=43&sessionId=19&resId=0&materialId=slides&confId=10613.

[8] Bromberg, C. et al. (2015). Design and operation of longbo: a 2 m long drift liquid argon tpc. *Journal of Instrumentation*, 10(07):P07015.

[9] Brun, R. and Rademakers, F. (1997). ROOT: An object oriented data analysis framework. *Nucl. Instrum. Meth.*, A389:81–86.

[10] Casper, D. (2002). The Nuance neutrino physics simulation, and the future. *Nucl. Phys. Proc. Suppl.*, 112:161–170. [,161(2002)].

[11] Church, E. D. (2013). LArSoft: A Software Package for Liquid Argon Time Projection Drift Chambers.

[12] Cooley, J. and Turkey, J. (1965). An algorithm for the machine calculation of complex fourier series. *Math. Comp.*, 19:297–301.

[13] DUNE DocDB 1704 (2016). 35t observations and measurements. [online] http://docs.dunescience.org:8080/cgi-bin/ShowDocument?docid=1704.

[14] DUNE DocDB 1834 (2016). 35t noise studies summary - apa wire readout noise. [online] http://docs.dunescience.org:8080/cgi-bin/ShowDocument.

[15] Hagman, C. et al. (2012). Cosmic-ray shower library {CRY}. [online] http://nuclear. llnl.gov/simulation/doc_cry_v1.7/cry.pdf.

[16] Insler, J. LSU. 35 ton Meeting (2016). Module to unstick adc codes. [online] https://indico.fnal.gov/getFile.py/access?contribId=2&resId=0&materialId=slides& confId=11627.

[17] Kudryavtsev, V. (2009). Muon simulation codes {MUSIC} and {MUSUN} for underground physics. *Computer Physics Communications*, 180(3):339 – 346.

[18] Li, Y. et al. (2016). Measurement of Longitudinal Electron Diffusion in Liquid Argon. *Nucl. Instrum. Meth.*, A816:160–170.

[19] MicroBooNE Collaboration (2016). Noise characterization and filtering in the microboone tpc. [online] http://www-microboone.fnal.gov/publications/publicnotes/ MICROBOONE-NOTE-1016-PUB.pdf.

[20] Thiesse, M. Sheffield University. 35 ton Meeting (2016). Something. [online] /url-SOMETHING.

[21] Torr, P. and Zisserman, A. (2000). Mlesac: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, 78(1):138–156.

[22] Wiener, N. (1942). *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. [Cambridge]: Technology Press of the Massachusetts Institute of Technology.

[23] Yang, T. Fermilab. 35 ton Meeting (2016a). Update on 35 ton analysis. [online] https://indico.fnal.gov/getFile.py/access?contribId=1&resId=0&materialId=slides& confId=12396.

[24] Yang, T. Fermilab. 35 ton Meeting (2016b). Update on reconstruction of 35t data. [online] https://indico.fnal.gov/getFile.py/access?contribId=1&resId=0&materialId=slides& confId=12349.

Appendix A   1

Something mildly interesting   2

1 Appendix B

2 Something else mildly interesting