

Writing your PhD thesis in **L^AT_EX2_ε** Using the CUED template



Krishna Kumar

Department of Engineering
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

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
October 2016

Acknowledgements

And I would like to acknowledge ...

Abstract

This is where you write your abstract ...

Table of contents

List of figures	8
List of tables	10
1 Introduction	11
1.1 XXXXX	11
1.2 XXXXXX	11
2 Theory	12
2.1 Theory of neutrino physics	12
2.2 Nucleon decay in Grand Unifying Theories	12
2.3 Existing and future experiments	12
2.4 How Liquid Argone Time Projection Chambers work	12
3 The Deep Underground Neutrino Experiment	13
3.1 DUNE location and beamline	13
3.2 The DUNE detectors and schedule	13
3.3 Physics opportunities of DUNE	13
3.3.1 Neutrino physics	13
3.3.2 Nucleon decay and supernovae neutrinos	13
3.4 Path to building DUNE - The 35 ton prototype	13
4 The 35 ton camera system	14
4.1 The need for cameras in a Liquid Argon Time Projection Chamber	14
4.2 Design of the camera system	14
4.3 Tabletop tests	14
4.4 Safety reviews and installation	14
4.5 Performance in the 35 ton	14

Table of contents	7
5 Simulations of the 35ton prototype	15
5.1 Structure of LArSoft	15
5.2 Making production event samples	15
5.3 Determination of interaction times	15
5.4 Calibrating calorimetric responses	15
5.5 Discerning reconstruction efficiencies	15
5.6 Performing particle identification	15
6 The 35 ton data sample	16
6.1 Organisation of the data structure	17
6.2 Reformatting the data to the offline structure	20
6.3 Observations on data quality and noise mitigation	21
6.4 Performance of reconstruction algorithms	25
6.5 Measuring interaction times using electron diffusion	29
6.5.1 Determining interaction times in 35 ton data	31
6.5.2 Determining interaction times in a low noise detector and differences with data	35
6.5.3 Discerning the impact of noise and electron lifetime in Monte Carlo when considering the limitations of the method	36
6.5.4 The limitations of and future improvements to the method of interac- tion time determination using diffusion	38
7 Simulations of the DUNE Far Detector	39
7.1 The MUSUN and MUSIC generators	39
7.2 Simulations of the LBNE surface detector	39
7.3 Incorporation of MUSUN into LArSoft	39
7.4 Simulation of background for DUNE	39
7.5 Cosmogenic background for nucleon decay channels in DUNE	39
References	40
Appendix A Something mildly interesting	41
Appendix B Something else mildly interesting	42

List of figures

6.1	The 35 ton data sample	17
6.2	The 35 ton data structure	19
6.3	Dropped TPC data in the 35 ton	22
6.4	Recovering stuck ADC codes in the 35 ton	23
6.5	Removing coherent noise in the 35 ton	23
6.6	Applying Wiener filters to the 35 ton data	24
6.7	The effect of noise removal algorithms in the 35 ton	24
6.8	dQ/dx in the 35 ton as a function of drift time	27
6.9	The dot product of track and counter coincidences	28
6.10	The alignment of tracks and counter coincidences	28
6.11	Reconstruction efficiencies of through going tracks in the 35 ton data	29
6.12	Schematic showing the process of diffusion	30
6.13	A simulated event display showing multiple tracks and flashes in the 35 ton	31
6.14	The effect of adding a noise baseline to a hit	32
6.15	The most probable values of the <i>RMS</i> and <i>RMS/Charge</i> distributions for tracks with a counter difference of 4 in the 35 ton data	33
6.16	The drift distance and angular dependence of diffusion in the 35 ton data . .	33
6.17	The predicted hit times for tracks with more than 100 'good' hit collection plane hits in the 35 ton data	34
6.18	The difference in predicted and reconstructed interaction times in the 35 ton data	34
6.19	The most probable values of the <i>RMS</i> and <i>RMS/Charge</i> distributions for tracks with a counter difference of 4 in a low noise 35 ton detector	36
6.20	Comparison of the drift distance and angular dependence of diffusion in a low noise 35 ton detector and the 35 ton dataset	36
6.21	A comparison of the accuracy of the interaction time prediction metrics for the low noise Monte Carlo and data 35 ton samples	36

- 6.22 The effect of increased noise in the accuracy of interaction time determination 37
- 6.23 The effect of increased noise in the accuracy of interaction time determination 37

List of tables

Chapter 1

1

Introduction

2

1.1 XXXXX

3

1.2 XXXXXXX

4

Chapter 2

Theory

2.1 Theory of neutrino physics

2.2 Nucleon decay in Grand Unifying Theories

2.3 Existing and future experiments

2.4 How Liquid Argone Time Projection Chambers work

Chapter 3

1

The Deep Underground Neutrino Experiment

2

3

3.1 DUNE location and beamline

4

3.2 The DUNE detectors and schedule

5

3.3 Physics opportunities of DUNE

6

3.3.1 Neutrino physics

7

3.3.2 Nucleon decay and supernovae neutrinos

8

3.4 Path to building DUNE - The 35 ton prototype

9

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

1

Simulations of the 35ton prototype

2

5.1 Structure of LArSoft

3

5.2 Making production event samples

4

5.3 Determination of interaction times

5

5.4 Calibrating calorimetric responses

6

5.5 Discerning reconstruction efficiencies

7

5.6 Performing particle identification

8

Chapter 6

The 35 ton data sample

THINGS I AM EXPECTING TO HAVE ALREADY DEFINED:

- The 35 ton geometry - TPC, SSP, CRC layout and structure.
- What a counter co-incidence is, with a nice picture from on my presentations.
- How reconstruction works - hit finding, clustering and tracking.

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. This is explained further in Section 6.3.

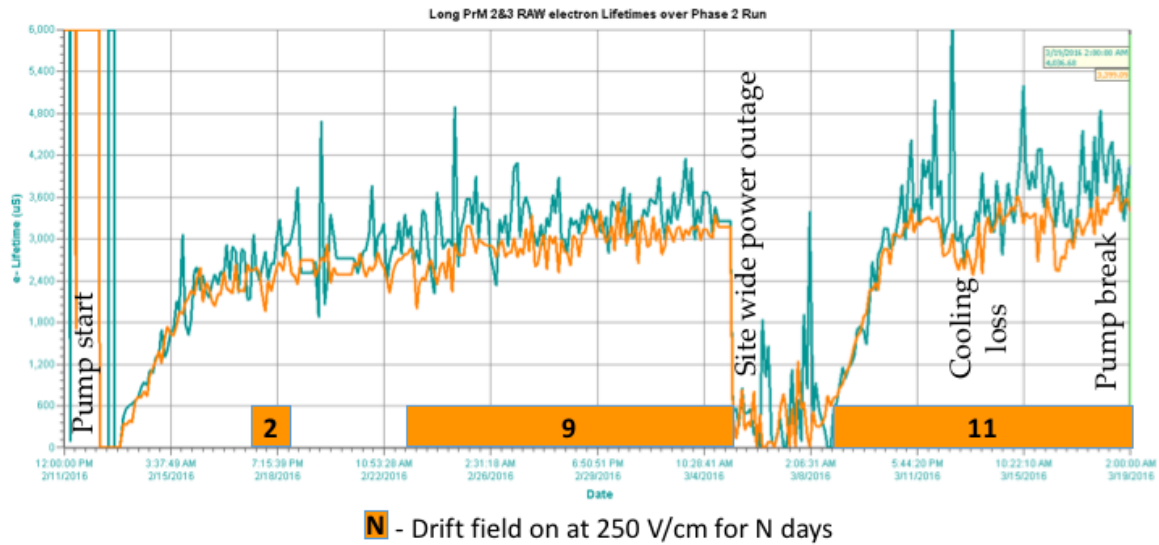


Fig. 6.1 Timeline showing the data collected during the 35 ton Phase II run once the purification pumps were turned on.

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 objects called `artdaq::Fragments`. These data objects would later have to be 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 triggered co-incidences on the CRCs on opposite 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 produce triggers on the N^{th} 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 elevated 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 XXXX 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 further 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

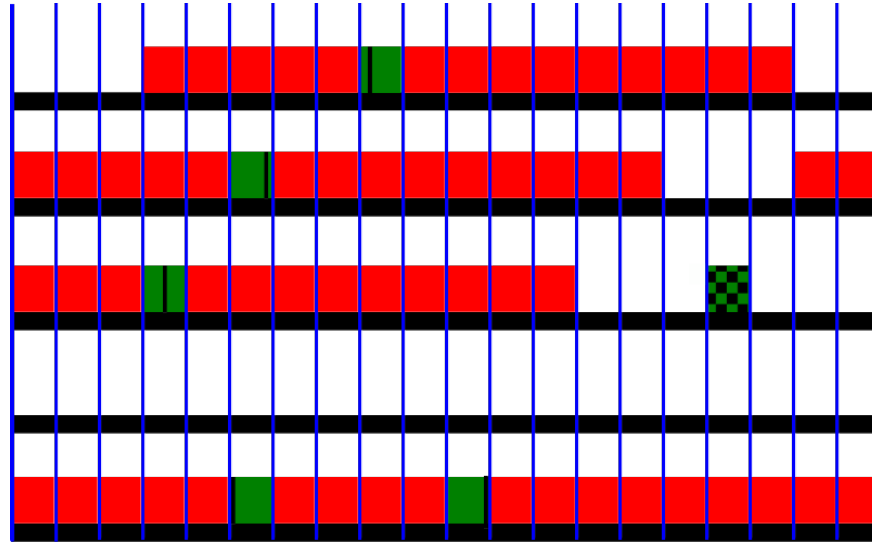


Fig. 6.2 A diagram of possible microslice structures for the TPC data recorded by the 35 ton. Each row represents a single DAQ event. The vertical blue lines delineate each microslice (0.5 ms, 1,000 ticks). Solid red boxes represent micro slices with TPC data in them. Green boxes represent triggers which were used with the black lines showing the time at which the trigger occurred, and green and black patterned boxes represent triggers which were ignored.

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.

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 millislices may only be accessed in whole. This means that a whole number of millislices must be loaded before the trigger so when a trigger is issued part way through a millislice the previous X millislices 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 millislice 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 raw data in the form of `artdaq::Fragments` to LArSoft objects such as `raw::RawDigits` (TPC), `raw::ExternalTrigger` (CRC) and `recob::OpDetWaveforms` (SSP) 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, check that the timing of each component was consistent and then 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 light or if there was a large change in the average TPC ADC value. Once a set of conditions is satisfied a user defined number of pre-condition ticks are gathered, 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 from and 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 Figure 6.3. If any of these issues were present in a given physics event then it is discarded as the integrity of the data cannot be guaranteed. It was decided that these events would be discarded as non-synchronous events would lead to hits in the detector being at incorrect times and padding empty events with pedestals could mean that tracks seem to disappear and reappear as they travel through the detector.

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

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

Though not directly affecting the noise issues stuck ADC codes were a feature of the data which had to be removed downstream. Stuck ADC codes were caused by the 6 least significant bits getting frozen to either 000000 or 111111, this was observed during the first stages of commissioning and an algorithm to remove them was developed and tested on Monte Carlo by Jonathan Insler of LSU. It was observed that in simulation the signal could be recovered with minimal losses, as shown in Figure 6.4 where the blue lines (after removal) are seen to closely match the black lines (before adding stuck codes).

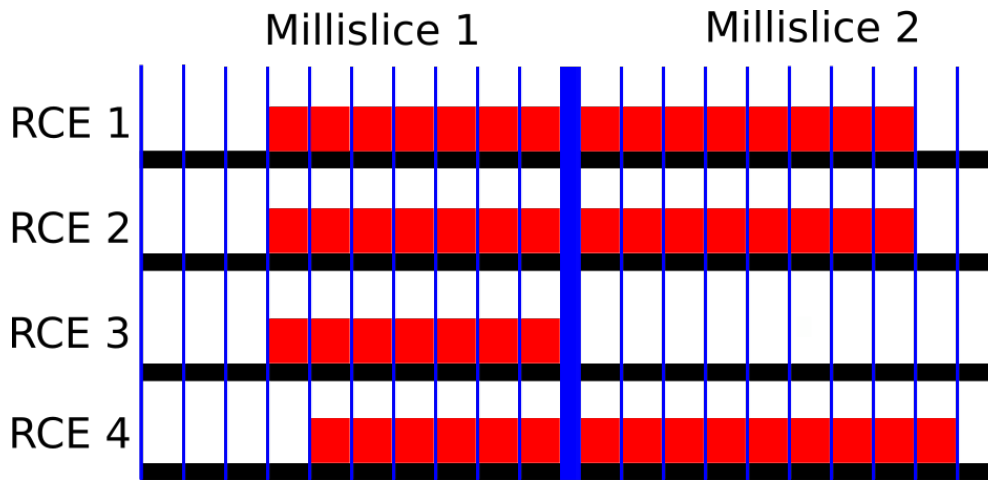


Fig. 6.3 A diagram of how TPC microslices can be saved in millislices in the 35 ton. Two millislices are shown, each containing 10 microslices. One physics event straddling the millislice boundaries is shown and 4 RCEs representing each row are read out. The vertical blue lines delineate each microslice (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,” whilst the data from RCE 4 is shifted by 1 microslice from RCEs 1 and 2 and is thus “Inconsistent.” As a result this physics would be discarded as data integrity cannot be guaranteed.

1 A significant portion of the noise was correlated between groups of 32 channels, where
2 the ADCs would coherently oscillate. To remove these coherent shifts ADC baselines were
3 calculated for the 32 channel groups at each tick and then subtracted from the measured
4 ADC values. This was also found to be an effective method of removing coherent noise in
5 MicroBooNE [8]. The effect of removing coherent noise is shown in Figure 6.5, where the
6 signal peak becomes much easier to discern after noise removal and a coherent noise hit at
7 tick ~ 6030 is removed. Subtracting increases in signal at a common time across adjacent
8 wires reduces the signal strength of events which are parallel to the APAs however as the
9 hits from these tracks will occur at roughly the same time. The only way to prevent this is to
10 “protect” potential signal regions from the coherent noise removal, as is done in MicroBooNE
11 [8].

12
13 When a Fast Fourier Transform (FFT) [2] is performed on the coherent noise subtracted
14 waveforms it can be seen that signals occur with specific frequencies. Some of these fre-
15 quencies are caused by real energy depositions, whilst others are due to the electronics noise.
16 It is possible to remove the noise frequencies by applying Wiener filters [12]. For each of
17 the three planes frequency spectrum’s are taken where a clear signal is both preserved and

6.3 Observations on data quality and noise mitigation

23

suppressed, the signal spectrum is then divided by the signal suppressed spectrum to produce a *signal/noise* frequency space. The signal regions which want to be excluded can then be found by fitting a combination of sigmoid functions to the frequency spaces, a demonstration of how this applied is shown in Figure 6.6. It is also possible to remove specific frequencies which are not removed by the filters, this was necessary for a 54 KHz noise component introduced by the fluorescent lights in the detector hall to be removed. After the run ended it was found that some of the high frequency noise components were introduced by a short on a warm power cable, the techniques used to find this cable will be used then commissioning future detectors [4].

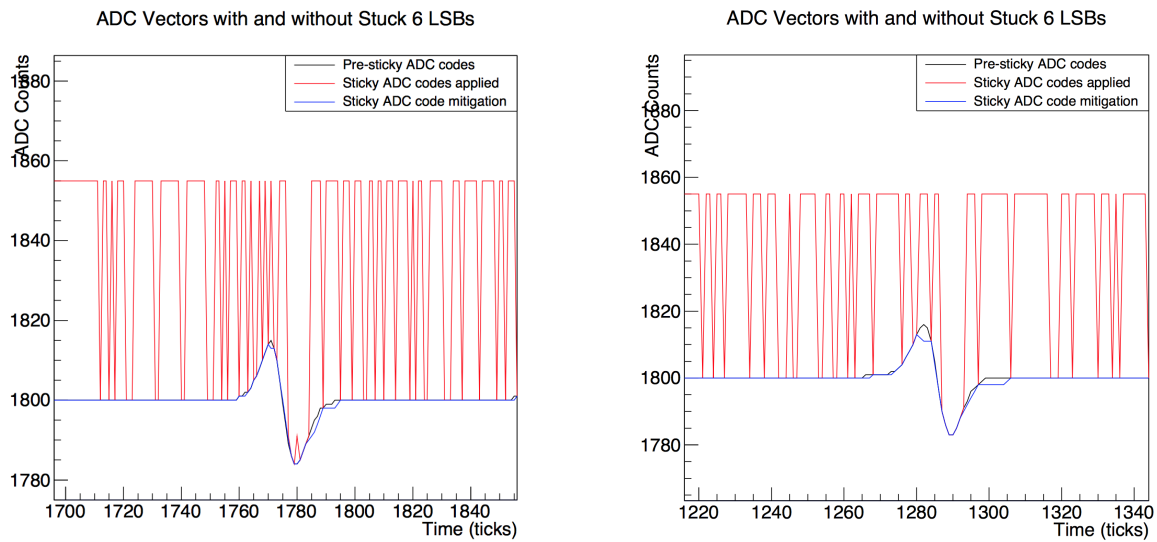
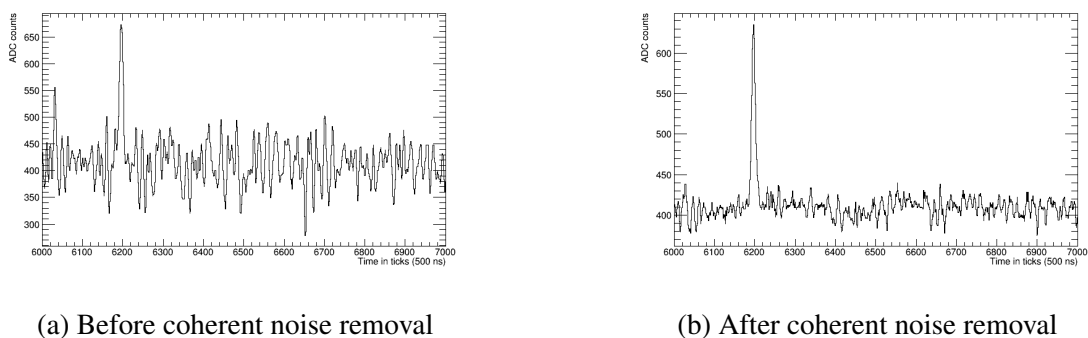


Fig. 6.4 Two Monte Carlo spectrum's showing the effect of the introduction and removal of stuck bits on a simulated signal [6].



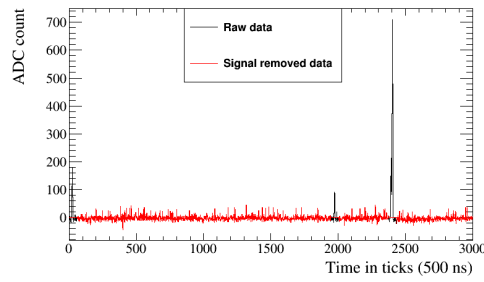
(a) Before coherent noise removal

(b) After coherent noise removal

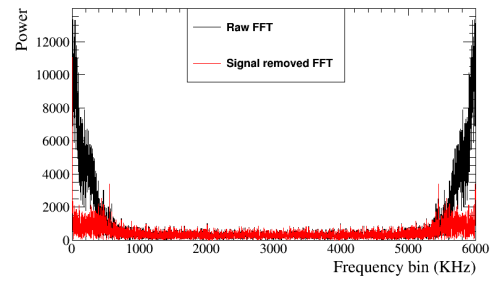
Fig. 6.5 The effect of coherent noise removal on a 35 ton signal event.

6.3 Observations on data quality and noise mitigation

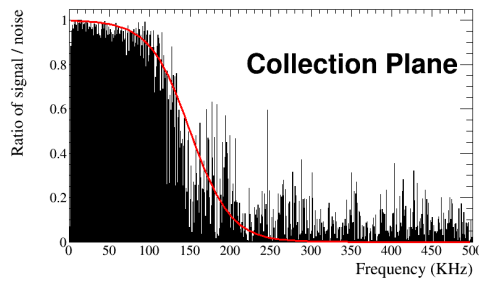
24



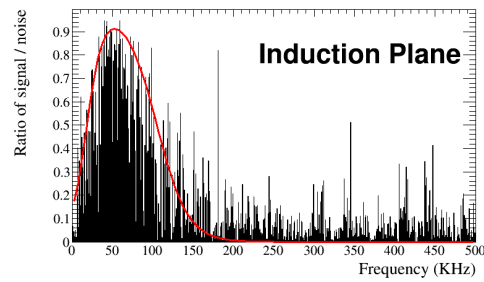
(a) A raw and signal subtracted waveform for a collection plane wire.



(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 signal strength of a frequency which is not filtered.



(d) The *signal/noise* ratio for an induction plane wire, the red line shows the signal strength of a frequency which is not filtered..

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

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

(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.

6 Transitions to a higher noise level state were also observed after cool down. The transi-
 7 tions would seemingly occur at random, but were occasionally observed to happen shortly
 8 after a saturation event across the whole detector [4]. After the transition the only way to
 9 recover the normal noise state was to power down the TPC components on APA 3, this was
 10 the only APA which did not have a ground mesh between the two sets of wire planes which
 11 may offer some clue to the cause of the effect though it could not be induced during warm
 12 testing and has not been observed in other experiments such as MicroBooNE. The elevated

noise state was associated with a strong signals at high frequencies, between 400 and 650 KHz. The values of these frequencies changed during the run as shown in FigureXXXX, though once a transition at a given frequency was induced it did not change until the TPC components were powered down. Data taken during these events was unrecoverable as the noise was too large, as such upon observing a transition the TPC components were quickly power cycled.

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 6.1.

$$X_{Hit} = T_{Hit} \times V_{Drift} \quad (6.1)$$

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 means that each wire segment has to be considered individually and that multiple wire segments 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 the collection plane there should also be signals on the induction wires.
- Are there adjacent wires which have hits at a similar time?

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

12 DO I WANT SOME SORT OF FIGURE SHOWING THIS? LOTS OF HITS -> ONLY
 13 TRACK HITS? COL HITS USED IN ITEM 3, OVERLAYED WITH ALL INDUCTION
 14 PLANE HITS CONSIDERED, WITH THE ONES KEPT IN A DIFFERENT COLOUR?

15
 16 Following a re-optimisation of the clustering algorithms it was observed that the stan-
 17 dard reconstruction could achieve track reconstruction to a similar efficiency as the counter
 18 shadowing and so the standard reconstruction has been used in the discussions to follow
 19 [9]. There has since been an effort to improve the counter shadowing hit disambiguation to
 20 remove the outlying collection plane hits using the MLESAC method [10] whereby points
 21 which are far away from a best fit are ignored, these studies are still on-going.

22
 23 A symptom of the elevated noise state is that signals are often dropped on one of the
 24 induction planes, this means that the tracking algorithms often have to combine clusters in
 25 only two of the three planes. Reconstruction using two planes was shown to be effective
 26 by the ArgoNeuT collaboration [5] so the loss of signal in one of the three planes is not
 27 prohibitive to track reconstruction. Another consequence of the elevated noise level is that
 28 even when the counters are used to seed hit finding the hit finding threshold is too high
 29 to reconstruct the very lowest hits. This causes the plot of dQ/dx for muons, shown in
 30 Figure 6.8, to look flat due to a cutoff at 100 ADC cm^{-1} below which no hits are found. The
 31 inability to reconstruct the lowest energy hits means that calorimetry is all but impossible on
 32 the 35 ton dataset even though the tracking algorithms perform relatively well.

33
 34 The inability to perform reliable calorimetry en masse means that the only particles which
 35 can be assuredly identified are the muons which triggered the counter co-incidences, making

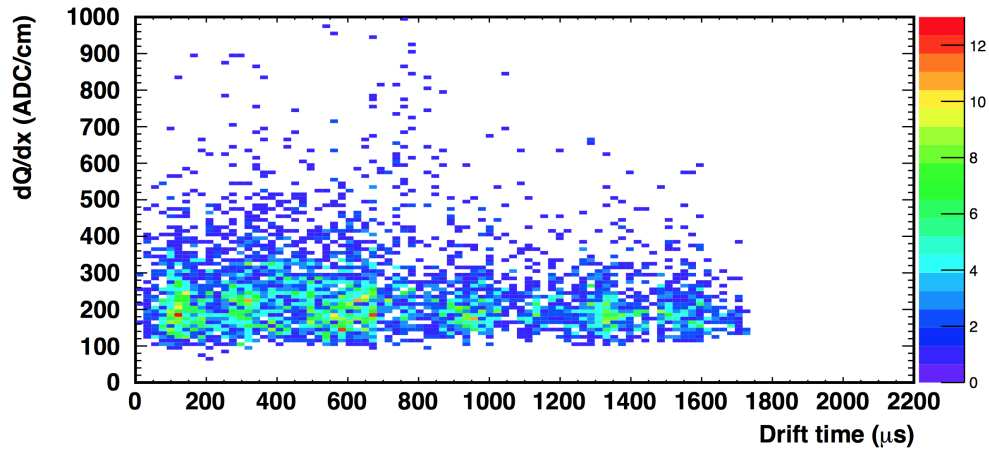


Fig. 6.8 The dQ/dx values for a sample of muon collection plane hits [11], note the cutoff at 100 ADC cm^{-1} due to the hit finding threshold.

the analysis proposed in !!FIXME!!Section refPIDAnalysis!!FIXME!! extremely difficult if not impossible. The muons in this sample will all traverse the detector 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

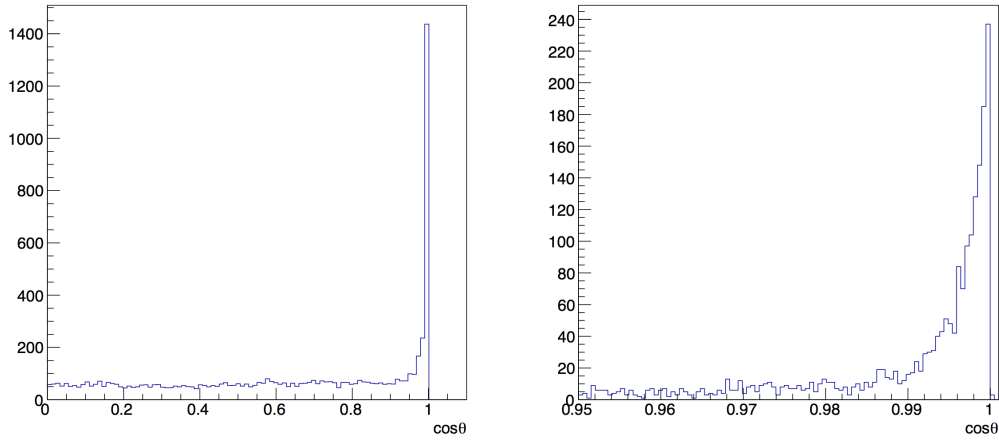


Fig. 6.9 The dot product of track and counter coincidences. The left plot shows all values of the dot product, whilst the right plot shows only the peak above 0.95.

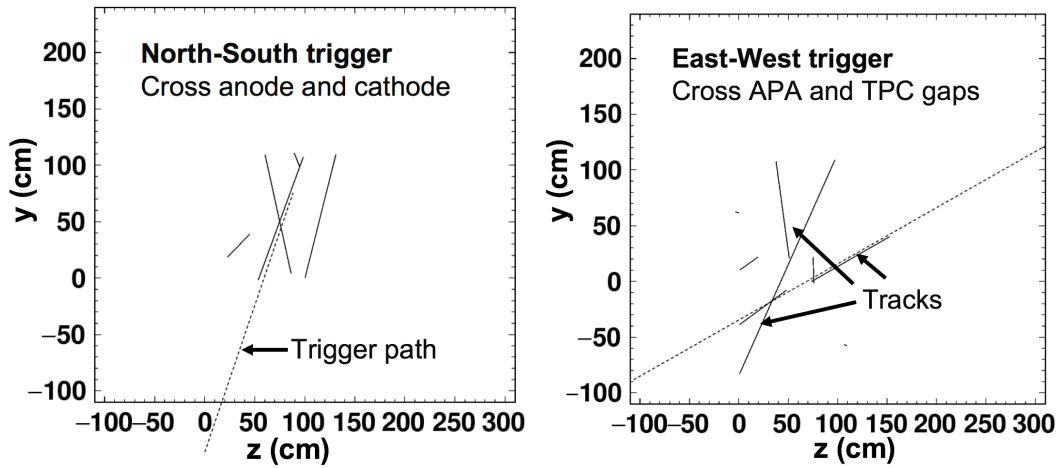


Fig. 6.10 The alignment of reconstructed tracks and the counter coincidences they cause. The left plot shows a North-South coincidence whilst the right plot shows an East-West coincidence.

1 discussed in Section 6.5.

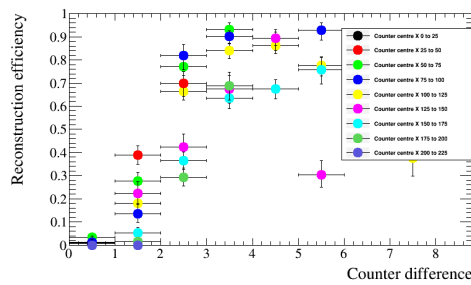
2

3 From Figure 6.11 it is evident that the reconstruction efficiency for tracks with shallow
 4 angles relative to the APAs is extremely poor, with the efficiency for tracks with angles less
 5 than 8° never rising above 10%. This is due to the coherent noise removal where hits which
 6 are correlated in time will be removed as they will be perceived as being noise as opposed to
 7 real signals. When considering tracks with angles in the 12° to 20° range the reconstruction
 8 is seen to be perform reasonably well for tracks whose centres are less than 140 cm away

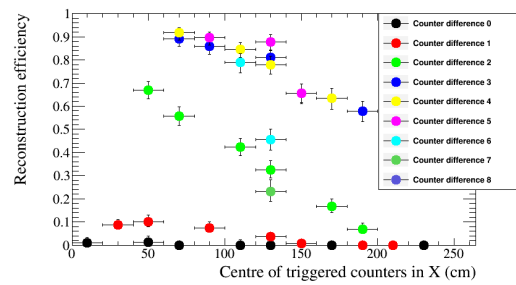
6.5 Measuring interaction times using electron diffusion

29

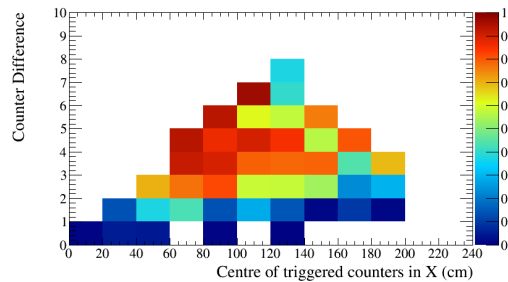
from the APAs, before decreasing for tracks which have a longer drift distance. A decrease in reconstruction efficiency for tracks which at large drift distances is due to the hits being below threshold and so not being reconstructed. It should be noted that in order to observe a drift distance effect on reconstruction efficiency the distance plotted is the average x position of the counters, so a track going from 10 cm to 230 cm will have an average x position of 120 cm (assuming a straight trajectory) as shown by the point for Counter difference 7 in Figure 6.11b with $\sim 23\%$ efficiency.



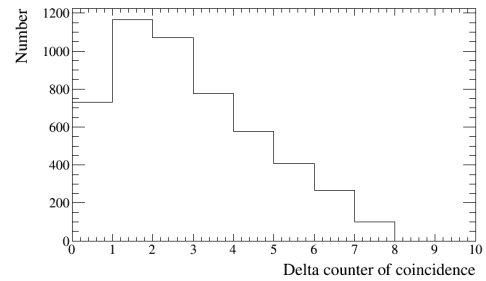
(a) The difference in efficiencies of different track angles relative to the APAs plotted for increasing distance from the APAs.



(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.

6.5 Measuring interaction times using electron diffusion

As electrons drift from the interaction point to the wire planes they become spread out in both time and space, this effect is known as diffusion and is an important property of electron transport in LAr which must be well understood. The mechanism by which diffusion

occurs in LAr was first discussed by Atrazhev-Timoshkin [1], and has since been developed to consist of a complete set of measurements for electric fields between 100 and 2000 V cm⁻¹ [7]. The diffusion of electrons is rarely isotropic and so the components that are in the transverse direction to the drift field and the component that is in the drift field direction are measured separately. Diffusion in the drift direction is called longitudinal diffusion and is generally smaller than the transverse diffusion. Figure 6.12 shows how diffusion can smear the electrons collected on a set of wires when the electrons are initially highly correlated in time and space.

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 direction of drift 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[3].

Longitudinal diffusion has the effect of spreading the drifting electrons out in time causing signals to become wider in time and smaller in height as the total charge is conserved. The increasing hit width can be measured for increasing drift times (distances) provided the hits do not fall below a hit finding threshold. Transverse diffusion causes drifting electrons to spread out in space changing the amount of charge deposited on a wire and reducing the charge resolution of the detector. Transverse diffusion is measured by discerning how the width of the hit charge distribution changes for increasing drift distances [3].

Through-going particles make ideal tracks to study diffusion as they are minimally ionising and so have roughly constant energy depositions along their tracks. The tracks they produce can also cover a wide range of drift distances if they are not parallel to the APAs. The drift distances of hits within a track can be determined by matching the track with a counter coincidence as discussed at the end of Section 6.4 and then correcting the x co-ordinates of the hits using Equation 6.1 where $T_{Hit} = T_{Measured} - T_{Interaction}$.

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. These techniques are particularly crucial for neutrino detectors on the Earth's 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. 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 35 ton dataset is ideal for testing this hypothesis as the counters are able to provide a sample with known track angles and interaction times which can be used to tune an interaction time determination metric. This metric 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 on wires with high noise 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 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. 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 XXXX (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 to the hits 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.

(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 [3].

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 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 region:

- The *RMS* of the hit - the most direct way to measure transverse diffusion.
- The *RMS/charge* of the hit - an attempt to remove the angular dependence and to incorporate an extra drift distance dependence on impurities for relatively low purity data.

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 intercept of the most probable fit values 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

6.5 Measuring interaction times using electron diffusion

33

- | | |
|--|--|
| <p>(a) The most probable hit <i>RMS</i> value for tracks with a counter difference of 4 at XXXX cm.</p> <p>(c) The most probable hit <i>RMS/Charge</i> value for tracks with a counter difference of 4 at XXXX cm.</p> | <p>(b) The most probable hit <i>RMS</i> value for tracks with a counter difference of 4 at XXXX cm.</p> <p>(d) The most probable hit <i>RMS/Charge</i> value for tracks with a counter difference of 4 at XXXX cm.</p> |
|--|--|

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

relative to the APAs. These dependencies show that over a large sample size diffusion can be separated into distance and angular dependant aspects, however whether this can be observed for individual tracks has not been considered here yet.

- | | |
|---|--|
| <p>(a) The most probable hit <i>RMS</i> value for tracks with a counter difference of 4 for a range of drift distances.</p> | <p>(b) The intercepts of the best fits of the most probable values for a range of drift distances for a range of angles.</p> |
|---|--|

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

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.

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

1 requiring at least 100 collection plane hits is not unreasonable and would correspond to a
 2 reconstructed track length of at least 50 cm. The predicted hit times and the difference in
 3 predicted and reconstructed hit times are shown in Figure 6.17 for both prediction metrics.
 4 It can be seen that in this dataset the *RMS/Charge* metric appears to provide a more robust
 5 measurement of the interaction time as the peak is more closely centred around 0 as the
 6 tracks were centred around the time of the counter coincidence which caused them.

(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.

7 When evaluating interaction times the average difference in reconstructed and predicted
 8 hit times for every hit on the track must be considered. This is shown in Figure 6.18, where
 9 as expected from Figure 6.17 the *RMS/Charge* metric provides a better estimation of the
 10 interaction time. The reason for this is that by utilising the charge information due to losses
 11 from impurity this metric gains an extra handle on the drift distance and hence the recon-
 12 structed time of the hits. The losses due to impurities are difficult to measure in very high
 13 purity LAr environments though, as illustrated by !!!!cite LongBo paper!!! where the LAr
 14 was so pure that it was difficult to measure the purity. The effect of increasing LAr purity
 15 is shown in Section 6.5.3. Using the change in hit charge in the 35 ton is not without its
 16 own dangers though, because as shown in Figure 6.8 there is a thresholding effect for hits
 17 with large drift times. However, as the same thresholding effect is present in all 35 ton data
 18 samples the limitation it introduces is mainly in the efficiency with which 'good' collection
 19 plane hits will be reconstructed and so this information can be confidently used.

20

Fig. 6.18 The difference in predicted and reconstructed interaction times in the 35 ton data. The difference in interaction times using the hit *RMS* metric are shown in black, whilst the difference in interaction times using the hit *RMS/Charge* metric are shown in blue.

21 Figure 6.18 shows that the interaction time of a reconstructed track in the 35 ton dataset
 22 can be determined using the effects of diffusion to an accuracy of $298.0 \pm 267.0 \mu\text{s}$ when using
 23 the hit *RMS* metric and to an accuracy of $55.6 \pm 210.4 \mu\text{s}$ when using the hit *RMS/Charge*
 24 metric. This resolution is impressive as even with the issues with regards to data quality
 25 in the 35 ton tracks can be cleanly separated throughout a drift window of $5200 \mu\text{s}$. 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 one would expect that no 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 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 also peaked at a lower hit *RMS*. This is likely due to the fact that a coherent noise baseline can increase the width of hits as shown in Figure 6.14. !!!??? WILL THE PEAK BE NOTICEABLY LOWER??? CHECK THAT THEY ARE ???!!! 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, along with the most probable values of hit *RMS* at a range of angles 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, this is 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.

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

(a) The distribution of hit *RMS* values at $x = 100$ cm.

(b) The distribution of hit *RMS/Charge* values at $x = 100$ cm.

Fig. 6.19 The most probable values of the *RMS* and *RMS/Charge* distributions for tracks with a counter difference of 4 in a low noise 35 ton detector. 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 in a low noise Monte Carlo sample.

(b) The most probable values at $x = 0$ for a range of track angles in a low noise Monte Carlo sample.

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.

1 accurately predict interaction times, this is again due to the ability to incorporate information
 2 about losses due to impurities which increase with drift distance. Also, as expected from the
 3 previous figures and the lower noise state in the Monte Carlo it is seen that the interaction
 4 times predicted in the Monte Carlo more closely match the true interaction times than in
 5 the data. The accuracy of interaction time determination in Monte Carlo (data) is shown to
 6 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
 7 *RMS/Charge* metrics respectively. An important feature to observe is that the widths of the
 8 distributions in Monte Carlo is less than half of that in the data, again this is attributed to the
 9 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.

10 6.5.3 Discerning the impact of noise and electron lifetime in Monte 11 Carlo when considering the limitations of the method

12 Many of the difficulties with the reconstruction and analysis of the 35 ton data have been
 13 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 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 100 cm in Figures 6.22a and 6.22b, the most probable values for hit *RMS* and the most probable values of hit *RMS* 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.

(a) The .	(b) The .
(c) The .	(d) The .
(e) The .	(f) The .

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

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

(a) The .	(b) The .
(c) The .	(d) The .
(e) The .	(f) The .

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

!!!!!! EXPLANATIONINGS OF WHAT IT LOOKS LIKE !!!!!!!

- 1 ANY OTHER STUDIES? The effect of using less / more hits? The effect of using
2 coherent noise subtraction in Monte Carlo on low noise data?

3 **6.5.4 The limitations of and future improvements to the method of in-**
4 **teraction time determination using diffusion**

5 sdfsd

Chapter 7

1

Simulations of the DUNE Far Detector

2

7.1 The MUSUN and MUSIC generators

3

7.2 Simulations of the LBNE surface detector

4

7.3 Incorporation of MUSUN into LArSoft

5

7.4 Simulation of background for DUNE

6

7.5 Cosmogenic background for nucleon decay channels in DUNE

7

8

References

- [1] 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.
- [2] Cooley, J. and Turkey, J. (1965). An algorithm for the machine calculation of complex fourier series. *Math. Comp.*, 19:297–301.
- [3] D. Brailsford, 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>.
- [4] DUNE DocDB 1704 (2016). 35t observations and measurements. [online] <http://docs.dunescience.org:8080/cgi-bin/ShowDocument?docid=1704>.
- [5] f (2012). The argoneut detector in the numi low-energy beam line at fermilab. *JINST*, 7:10019.
- [6] J. Insler, 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>.
- [7] Li, Y. et al. (2016). Measurement of Longitudinal Electron Diffusion in Liquid Argon. *Nucl. Instrum. Meth.*, A816:160–170.
- [8] MicroBooNE Collaboration (2016). Noise characterization and filtering in the microboone tpc. [online] <http://www-microboone.fnal.gov/publications/publicnotes/MICROBOONE-NOTE-1016-PUB.pdf>.
- [9] T. Yang, FNAL, 35 ton Meeting (2016). Update on reconstruction of 35t data. [online] <https://indico.fnal.gov/getFile.py/access?contribId=1&resId=0&materialId=slides&confId=12349>.
- [10] Torr, P. and Zisserman, A. (2000). Mlesac: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, 17:138–156.
- [11] T. Yang, FNAL, 35 ton Meeting (2016). Update on 35 ton analysis. [online] <https://indico.fnal.gov/getFile.py/access?contribId=1&resId=0&materialId=slides&confId=12396>.
- [12] Wiener, N. (1942). *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. [Cambridge]: Technology Press of the Massachusetts Institute of Technology.

Appendix A

1

Something mildly interesting

2

¹ **Appendix B**

² **Something else mildly interesting**